QUANTITATIVE NORMAL APPROXIMATION FOR SUMS OF RANDOM VARIABLES WITH MULTILEVEL LOCAL DEPENDENCE STRUCTURE

JULIAN FISCHER

ABSTRACT. We establish a quantitative normal approximation result for sums of random variables with multilevel local dependencies. As a corollary, we obtain a quantitative normal approximation result for integral functionals of random fields which may be approximated well by random fields with finite dependency range. Such random fields occur for example in the homogenization of linear elliptic equations with random coefficient fields. In particular, our result allows for the derivation of a quantitative normal approximation result for the approximation of effective coefficients in stochastic homogenization in the setting of coefficient fields with finite range of dependence. The proof of our normal approximation theorem is based on a suitable adaption of Stein's method and requires a different estimation strategy for terms originating from long-range dependencies as opposed to terms stemming from short-range dependencies.

1. Introduction

Stein's method of normal approximation [42, 43] – as well as Chatterjee's variant [10, 11] – are among the most widely used methods for the derivation of quantitative estimates on the deviation of a probability distribution from normality. Stein's method is remarkably flexible, being applicable to the multivariate setting [24, 23, 37, 12, 35, 13] and to the setting of sums of "locally dependent" random variables [43, 3, 4, 36, 14, 38, 13]. A survey of quantitative normal approximation by Stein's method may be found e. g. in [13, 39].

In recent years, Chatterjee's variant [10, 11] of Stein's method in the form of second-order Poincaré inequalities has found widespread use: For Gaussian random fields second-order Poincaré inequalities have been established in [11, 33], while for discrete probability distributions with product structure (like Poisson point processes) this has been accomplished in [10, 27]. Random geometries like random sequential adsorption processes or Voronoi and Delaunay tesselations for random (e. g. Poisson) point distributions have been considered in [34, 41, 28, 16].

In the present work, we derive a result on normal approximation for random variables that arise from random fields with finite range of dependence, featuring multilevel dependencies (with strong local dependencies and decaying dependencies on larger dependency ranges). The lack of a (hidden) product structure – as present

---

Key words and phrases. normal approximation, Stein's method, local dependence.

This project was initiated while the author enjoyed the hospitality of the Hausdorff Research Institute for Mathematics, Bonn, as a participant of the Trimester Program "Multiscale Problems: Algorithms, Numerical Analysis and Computation". The author would like to thank Mitia Duerinckx and Antoine Gloria for interesting discussions on the manuscript.

in the abovementioned applications of Chatterjee's variant of Stein's method – prevents us from using Chatterjee's variant of Stein's method. Instead, our approach is closer to the original version of Stein's method respectively the results on sums of random variables with local dependence structure [36, 14, 37].

The quantitative normal approximation results obtained in the present contribution have an interesting consequence for the homogenization of elliptic equations with random coefficient field: While so far the quantitative description of fluctuations had been limited to the setting of probability distributions of coefficient fields subject to a second-order Poincaré inequality [32, 21, 26, 30, 17], our result enables us to develop a first quantitative description of certain fluctuations in stochastic homogenization – namely, the fluctuations of the *effective conductivity* of random periodic materials – under the assumption of finite range of dependence [18]. It has also an interesting consequence for the numerical analysis of an algorithm by Le Bris, Legoll, and Minvielle [29] capable of increasing the accuracy of approximations for the effective conductivity, see [18].

1.1. **Random fields in stochastic homogenization.** Let us give a few remarks on random fields in stochastic homogenization, from which the present work draws its main motivation. The solution $u \in H^1_0(\mathbb{R}^d)$ to a linear elliptic equation with a random coefficient field $a : \mathbb{R}^d \times \Omega \to \mathbb{R}^{d\times d}$

$$-\nabla \cdot (a \nabla u) = f \tag{1}$$

(with $f \in L^2(\mathbb{R}^d)$) may – under the assumptions of uniform ellipticity and boundedness, stationarity, and finite range of dependence $\varepsilon \leq \frac{1}{2}$ of the random coefficient field a – be approximated by the solution to an *effective equation* of the form

$$-\nabla \cdot (a_{\mathsf{hom}} \nabla u_{\mathsf{hom}}) = f$$

with a constant (elliptic) effective coefficient $a_{\mathsf{hom}} \in \mathbb{R}^{d\times d}$ which depends on the law of a (but is invariant under spatial rescaling). Namely, one has the following quantitative estimate [1, 20, 22]: The difference of the solutions u and u_{hom} is bounded for a suitable $p(d) \geq 2$ by

$$||u - u_{\mathsf{hom}}||_{L^p(\mathbb{R}^d)} \leq \begin{cases} \mathcal{C}(a,f)||f||_{L^2(\mathbb{R}^d)}\varepsilon|\log\varepsilon| & \text{for } d = 2, \\ \mathcal{C}(a,f)||f||_{L^2(\mathbb{R}^d)}\varepsilon & \text{for } d \geq 3, \end{cases}$$

where $\mathcal{C}(a,f)$ denotes a random constant with almost Gaussian stochastic moments

$$\mathbb{E}\left[\exp\left(\frac{\mathcal{C}(a,f)^{2-\delta}}{C(d,\lambda,\delta)}\right)\right] \leq 2.$$

Note that the fluctuations of functionals of the solution u like $\int u\eta\,dx$ for smooth compactly supported η are expected to be of the order of the central limit theorem scaling $\varepsilon^{d/2}$. This suggests that for $d \geq 3$ it should be possible to achieve a higher-order deterministic approximation for u. As shown in [5, 7, 25], in weaker norms one may in fact obtain such a higher-order approximation for u in terms of solutions to deterministic equations, based on the concept of higher-order homogenization correctors (which also plays an important role in higher-order regularity theory for elliptic equations with periodic or random coefficient field, see [2, 19]).

However, to achieve a description of the fluctuations of the solution u a different approach is required. A description of the correlation structure of fluctuations of the homogenization corrector ϕ_i (see below for a definition) had been obtained in

[31]; subsequently, in [26] a description of the correlation structure of fluctuations in the solutions u to the equation (1) was given. The first quantitative normal approximation result in the context of stochastic homogenization was obtained for the effective conductivity on the torus in [32] and subsequently [21], following earlier qualitative results in [9, 40]; for linear functionals of the solution u of the form $\int u\eta\,dx$ it was obtained in [30].

A fundamental object in homogenization is the *homogenization corrector* ϕ_i, defined as the unique (up to addition of constants) solution with sublinear growth to the PDE

$$-\nabla\cdot(a(e_i+\nabla\phi_i))=0.$$

The effective coefficient a_{hom} is determined by the homogenization corrector as

$$a_{\mathsf{hom}}e_i := \mathbb{E}[a(e_i+\nabla\phi_i)]. \tag{2}$$

To faciliate the description of fluctuations in stochastic homogenization, in [17] the concept of the *homogenization commutator*

$$\Xi_{ij} := (a-a_{\mathsf{hom}})(e_i+\nabla\phi_i)\cdot e_j$$

has been introduced (bearing its name due to the similarity of the expression with a commutator). It allows for the description of fluctuations of solutions to equations of the form $-\nabla\cdot(a\nabla u)=\nabla\cdot g$ in the following way: Introducing the solutions u_{hom} to the homogenized equation $-\nabla\cdot(a_{\mathsf{hom}}\nabla u_{\mathsf{hom}})=\nabla\cdot g$ and the solution v_{hom} to the equation $-\nabla\cdot(a_{\mathsf{hom}}\nabla v_{\mathsf{hom}})=\nabla\cdot\eta$, one has

$$\int \eta\cdot(\nabla u-\mathbb{E}[\nabla u])\,dx \approx \int \Xi\nabla u_{\mathsf{hom}}\cdot\nabla v_{\mathsf{hom}}\,dx$$

up to an error of order $\varepsilon^{d/2+1}$ (which amounts to a relative error of order ε). As shown in [17] under the assumption of a second-order Poincaré inequality for the probability distribution of the coefficient field a, functionals of the homogenization commutator like $\int \Xi\eta\,dx$ display fluctuations on scale $\varepsilon^{d/2}$ while the distance of their probability distribution to a Gaussian is basically of order ε^{d} (e. g. in the 1-Wasserstein distance), amounting to a relative error of order $\varepsilon^{d/2}$. The results of our present work will allow for an analogous result in the setting of coefficient fields with finite range of dependence, see the upcoming work [15].

To obtain an approximation for the effective coefficient a_{hom} which is given analytically by (2), one typically considers a *periodization* of the random coefficient field, that is an $L\varepsilon$-periodic random coefficient field $a_{\mathsf{per},L}$ whose law on each cube of diameter $\frac{L\varepsilon}{2}$ coincides with the law of the original coefficient field a on the same cube. For this periodization, the homogenization corrector $\phi_{\mathsf{per},L,i}$ given by $-\nabla\cdot(a_{\mathsf{per},L}(e_i+\nabla\phi_{\mathsf{per},L,i}))=0$ is an $L\varepsilon$-periodic function and one may for a single realization compute an approximation for the effective coefficient a_{hom} according to

$$a^{\mathrm{RVE}}e_i := \fint_{[0,L\varepsilon]^d} a_{\mathsf{per},L}(e_i+\nabla\phi_{\mathsf{per},L,i})\,dx. \tag{3}$$

The error of this approximation is dominated by the fluctuations of a^{RVE}, which are of the order

$$\mathbb{E}\big[\big|a^{\mathrm{RVE}}-\mathbb{E}[a^{\mathrm{RVE}}]\big|^2\big]^{1/2} \leq CL^{-d/2},$$

while the systematic error is of higher oder

$$\left|\mathbb{E}[a^{\mathrm{RVE}}] - a_{\mathsf{hom}}\right| \leq CL^{-d}|\log L|^d.$$

As shown in [18], the quantitative result on normal approximation established in the present work has an interesting consequence: It facilitates a rigorous mathematical analysis of the *selection approach for representative volumes* introduces by Le Bris, Legoll, and Minvielle [29]. The selection approach of Le Bris, Legoll, and Minvielle is a remarkably successful numerical algorithm for increasing the accuracy of the approximations for effective coefficients: It basically proceeds by selecting not a random sample of the coefficient field a_{per} for the computation of the approximation a^{RVE} for the effective coefficient (3), but a sample of the coefficient field that is "particularly representative" in the sense that it captures certain statistical properties of the random coefficient field – like the spatial average $\fint_{[0,L\varepsilon)^d} a\,dx$ – exceptionally well. It has been observed that the method of Le Bris, Legoll, and Minvielle achieves its gain in accuracy by reducing the fluctuations of the approximations. In [18], it is shown that the approximate multivariate normality of the joint probability distribution of a^{RVE} and $\fint_{[0,L\varepsilon)^d} a\,dx$ – which is established using Theorem 4 below – allows for a rigorous analysis of the selection approach. In fact, it is this mathematical application that has dictated our choice of the distance between probability distributions in Definition 1.

Notation. For a vector $v \in \mathbb{R}^m$ we denote by $|v|$ its Euclidean norm; the vectors of the standard basis are denoted by e_i, $1 \leq i \leq m$. We denote the identity matrix in $\mathbb{R}^{N\times N}$ by Id or Id_N. For a matrix $A \in \mathbb{R}^{m\times m}$ we shall denote by $|A|$ its natural norm $|A| := \max_{v,w\in\mathbb{R}^m,|v|=|w|=1} |v\cdot Aw|$. Similarly, on the space of tensors $B \in \mathbb{R}^{m\times m\times m}$ we shall use the norm given by $|B| := \max_{u,v,w\in\mathbb{R}^m,|u|=|v|=|w|=1} \sum_{i,j,k=1}^m B_{ijk}u_iv_jw_k$. For $x \in \mathbb{R}^d$ we denote by $|x|_\infty = \max_i |x_i|$ its supremum norm. By $|x-y|_{\mathsf{per}} := \inf_{k\in\mathbb{Z}^d} |x-y-Lk|$ respectively (for sets) $\mathrm{dist}^{\mathsf{per}}(U,V) := \inf_{k\in\mathbb{Z}^d} \mathrm{dist}(U,k+V)$, we denote the periodicity-adjusted distance in the context of the torus $[0,L]^d$. By $|x-y|_{\mathsf{per},\infty}$ and $\mathrm{dist}^{\mathsf{per}}_\infty$, we denote the corresponding distances associated with the maximum norm.

Given a positive definite symmetric matrix $\Lambda \in \mathbb{R}^{N\times N}$, we denote the Gaussian with covariance matrix Λ by

$$\mathcal{N}_\Lambda(x) := \frac{1}{(2\pi)^{N/2}\sqrt{\det\Lambda}} \exp\bigg(-\frac{1}{2}\Lambda^{-1}x\cdot x\bigg).$$

For $\gamma > 0$, we equip the space of random variables X with stretched exponential moment $\mathbb{E}[\exp(|X|^\gamma/a)] < \infty$ for some $a = a(X) > 0$ with the norm $||X||_{\exp^\gamma} := \sup_{p\geq 1} p^{-1/\gamma}\mathbb{E}[|X|^p]^{1/p}$. For a discussion of this choice of norm, see Appendix C.

For a map $f : \mathbb{R}^N \to X$ into a normed vector space V, we denote for any $r > 0$ by $\mathrm{osc}_r f(x_0) := \sup_{x,y\in\{|x-x_0|\leq r\}} |f(x)-f(y)|_V$ its oscillation in the ball of radius r around x_0.

For two random variables X and Y (possibly defined on different probability spaces), we denote equality in law by $X \stackrel{d}{=} Y$. For a vector-valued random variable X, we denote by $\mathrm{Var}\, X$ the covariance matrix of its entries. Similarly, for two vectors X and Y we denote by $\mathrm{Cov}[X,Y]$ the matrix of covariances of the entries of X and the entries of Y. On the space of fields $v : \mathbb{R}^d \to \mathbb{R}^N$ we shall use the $L^p_{loc}(\mathbb{R}^d)$ topology.

For two subsets U and V of $\mathbb{R}^k$, we denote as usual by $U+V$ the Minkowski sum $U+V := \{y+z : y \in U, z \in V\}$. Similarly, for a subset $U \subset \mathbb{R}^k$, a vector $x \in \mathbb{R}^k$, and a scalar $\lambda > 0$, we denote by $x+U$ the translation of U by x and by λU the set $\lambda U := \{\lambda y : y \in U\}$. By B_r we denote the ball of radius r around 0.

2. Main Results

Throughout our present work, the following assumption of *finite range of dependence* is the central assumption on the random fields.

(A) Let $a : \mathbb{R}^d \times \Omega \to \mathbb{R}^n$ (with $d, n \in \mathbb{N}$) be a random field. We say that a has range of dependence 1 if for any two measurable sets $U, V \subset \mathbb{R}^d$ with $\operatorname{dist}(U,V) > 1$ the restrictions $a|_U$ and $a|_V$ are stochastically independent.

(A') Let $a : \mathbb{R}^d \times \Omega \to \mathbb{R}^n$ (with $d, n \in \mathbb{N}$) be an almost surely L-periodic random field for some $L \geq 1$.We say that a has range of dependence 1 if for any two measurable sets $U, V \subset \mathbb{R}^d$ with $\operatorname{dist}_{\mathsf{per}}(U,V) > 1$ the restrictions $a|_U$ and $a|_V$ are stochastically independent.

Our main result are a quantitative multivariate normal approximation result for integral functionals of random fields that admit a good approximation in terms of finite-range random fields as well as a quantitative multivariate normal approximation result for sums of random variables with multilevel local dependence. The distance of these probability distributions to a multivariate Gaussian will be quantified through the following notion of distance between probability measures. Note that this distance is a standard choice in the theory of multivariate normal approximation, see e. g. [13] and the references therein. Note also that our choice of distance $\mathcal{D}$ dominates the 1-Wasserstein distance.

Definition 1. *For a symmetric positive definite matrix $\Lambda \in \mathbb{R}^{N\times N}$ and $\bar{L} < \infty$, we consider the classes $\Phi_\Lambda^{\bar{L}}$ of functions $\phi : \mathbb{R}^N \to \mathbb{R}$ with the following properties:*

- *ϕ is smooth and its first derivative is bounded in the sense $|\nabla\phi(x)| \leq \bar{L}$ for all $x \in \mathbb{R}^N$.*
- *For any $r > 0$ and any $x_0 \in \mathbb{R}^N$, we have*

$$\int_{\mathbb{R}^N} \operatorname{osc}_r \phi(x)\ \mathcal{N}_\Lambda(x-x_0)\,dx \leq r, \tag{4}$$

where $\operatorname{osc}_r\phi(x)$ is the oscillation of ϕ defined as

$$\operatorname{osc}_r\phi(x) := \sup_{|z|\leq r} \phi(x+z) - \inf_{|z|\leq r} \phi(x+z)$$

and where

$$\mathcal{N}_\Lambda(x) := \frac{1}{(2\pi)^{N/2}\sqrt{\det\Lambda}} \exp\bigg(-\frac{1}{2}\Lambda^{-1}x\cdot x\bigg).$$

The class Φ_Λ is defined as

$$\Phi_\Lambda := \bigcup_{\bar{L}>0} \Phi_\Lambda^{\bar{L}}.$$

Furthermore, we introduce the distance $\mathcal{D}$ and the regularized distance $\mathcal{D}^{\bar{L}}$ between the law of an $\mathbb{R}^N$-valued random variable X and the N-variate Gaussian $\mathcal{N}_\Lambda$

as

$$\mathcal{D}(X,\mathcal{N}_\Lambda) := \sup_{\phi\in\Phi_\Lambda} \bigg(\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\bigg) \tag{5}$$

and

$$\mathcal{D}^{\bar L}(X,\mathcal{N}_\Lambda) := \sup_{\phi\in\Phi_\Lambda^{\bar L}} \bigg(\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\bigg). \tag{6}$$

By the definition of our distance $\mathcal{D}$, the regularization $|\nabla\phi| \leq L$ in the definition of the test functions for the distance $\mathcal{D}^{\bar L}$ may be removed by letting $\bar L \to \infty$. We shall prove most of our statements first in the regularized setting $\bar L < \infty$ – note that for random variables X with finite first moment the distance $\mathcal{D}^{\bar L}$ is guaranteed to be finite – and then extend them to $\mathcal{D}$ by passing to the limit $\bar L \to \infty$.

Note that for $\bar L > 1$ the distance $\mathcal{D}^{\bar L}$ is a stronger distance than the 1-Wasserstein distance (while for $L = 1$ it coincides with the 1-Wasserstein distance).

Let us also remark that (4) entails (by letting $r \to 0$) the bound

$$\int_{\mathbb{R}^N} |\nabla\phi|(x)\mathcal{N}_\Lambda(x-x_0)\,dx \leq 1 \tag{7}$$

for any $x_0 \in \mathbb{R}^N$.

For spatial averages of random fields v which may be approximated well by random fields v_r with finite range of dependence r the following normal approximation result holds true.

Theorem 2. *Let $a : \mathbb{R}^d \times \Omega \to \mathbb{R}^n$, $d, n \in \mathbb{N}$, be a random field with finite range of dependence 1 in the sense of (A). Let $k \in \mathbb{N}$ and let $v : \mathbb{R}^d \times \Omega \to \mathbb{R}^N$, $N \in \mathbb{N}$, be a random field with*

$$\fint_{\{|x-x_0|\leq 1\}} |v|\,dx \leq \mathcal{C}(a,x_0)$$

for some random constant $\mathcal{C}(a,x_0)$ with stretched exponential stochastic moments $||\mathcal{C}(a,x_0)||_{\exp^\gamma} \leq 1$ for some $\gamma > 0$ and any $x_0 \in \mathbb{R}^d$. Suppose that there exists a family of random fields $v_r : \mathbb{R}^d \times \Omega \to \mathbb{R}^N$, $1 \leq r < \infty$, such that v_r is an r-local function of a and an approximation for v in the following sense:

- *For any measurable $U \subset \mathbb{R}^d$ the restriction $v_r|_U$ is a measurable function of $a|_{U+B_r}$.*
- *There exist random constants $\mathcal{C}(a,r,x_0,\xi)$ with stretched exponential stochastic moments $||\mathcal{C}(a,r,x_0,\xi)||_{\exp^\gamma} \leq 1$ such that for any $L \geq r$, any $x_0 \in \mathbb{R}^d$, and any $\xi \in L^\infty(\mathbb{R}^d)$ with $\operatorname{supp}\xi \subset \{|x-x_0| \leq r\}$ the estimate*

$$\fint_{\{|x-x_0|\leq r\}} (v-v_r)\xi\,dx \leq \mathcal{C}(a,x_0,r,\xi)\sum_{l=0}^{k} r^l ||\nabla^l\xi||_{L^\infty} r^{-d} \tag{8}$$

  holds.

Then for any $\xi \in L^\infty(\mathbb{R}^d)$ with $\operatorname{supp}\xi \subset [0,L]^d$ and $|\nabla^k\xi| \leq (L+|x|)^{-d-1-k}$ for $0 \leq k \leq d$ the random variable

$$X := L^{-d}\int_{\mathbb{R}^d} \xi v\,dx$$

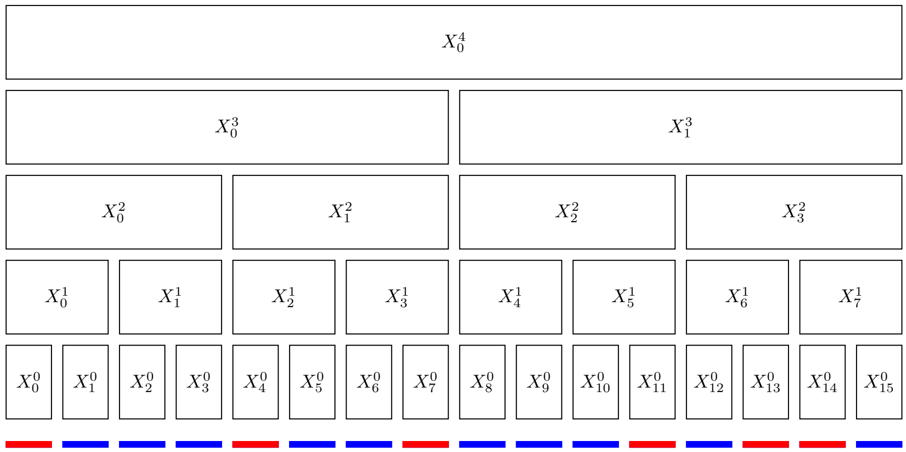

FIGURE 1. An illustration of the "multilevel local dependence structure" as introduced in Definition 3 (in a one-dimensional setting). At the bottom, a sample of the random field a is depicted; the X_y^k may depend not only on the values of the random field directly below their box, but on the random field in a region that is wider by a factor of $K \log L$.

admits the normal approximation

$$\mathcal{D}(X-\mathbb{E}[X],\mathcal{N}_{\mathrm{Var}\ X}) \leq C(d,\gamma)(\log L)^{C(d,\gamma)}\big(L^{-d}|\mathrm{Var}\ X^{1/2}||\mathrm{Var}\ X^{-1/2}|^3\big)L^{-d}.$$

In our second main theorem, we make use of the following notion of "multilevel local dependence decomposition". An illustration of this decomposition is provided in Figure 1.

Definition 3 (Sums of random variables with multilevel local dependence structure)**.** *Let $d \geq 1$ and $L \geq 2$. Consider a random field a on $\mathbb{R}^d$ subject to the assumption of finite range of dependence (A) or an L-periodic random field subject to the assumption of finite range of dependence (A'). Let $X = X(a)$ be a random variable depending on the random field.*

We then say that X is a sum of random variables with multilevel local dependence if there exist random variables $X_y^m = X_y^m(a)$, $0 \leq m \leq 1+\log_2 L$ and $y \in 2^m\mathbb{Z}^d \cap [0,L)^d$, and constants $K \geq 2$, $\gamma \in (0,2]$, and $B \geq 1$ with the following properties:

- *The random variable $X_y^m(a)$ is a measurable function of $a|_{y+K\log L\,[-2^m,2^m]^d}$.*
- *We have*

$$X = \sum_{m=0}^{1+\log_2 L} \sum_{y\in 2^m\mathbb{Z}^d\cap[0,L)^d} X_y^m.$$

- *The random variables X_y^m satisfy the bound*

$$||X_y^m||_{\exp^\gamma} \leq BL^{-d}. \tag{9}$$

It is well-known that Stein's method of normal approximation allows to establish a quantitative result on normal approximation for sums of random variables with local dependence structure, see e. g. [13, 14, 37] and the references therein. However, in many applications global dependencies arise naturally: For example, the approximation of the effective coefficient in the homogenization of linear elliptic equations with random coefficient field – that is, the random variable a^{RVE} as defined by (3) – features global dependencies. It is shown in the companion article [18] that a^{RVE} may nevertheless be approximated by such a sum of random variables with a *multilevel local dependence* structure.

The main result of the present work is the following quantitative central limit theorem for sums of vector-valued random variables with a multilevel local dependence structure, which is not covered by the typical normal approximation results for sums of random variables with a given dependency graph.

Theorem 4. *Let $d \geq 1$ and $L \geq 2$. Consider a random field a on $\mathbb{R}^d$ subject to the assumption of finite range of dependence (A) or an L-periodic random field subject to the assumption of finite range of dependence (A'). Let $X = X(a)$ be a random variable that may be written as a sum of random variables with multilevel local dependence in the sense of Definition 3. Then the law of the random variable X is close to a multivariate Gaussian in the sense*

$$\mathcal{D}(X - \mathbb{E}[X], \mathcal{N}_\Lambda) \leq C(d,\gamma,N,K) B^3 (\log L)^{C(d,\gamma)} \big(L^{-d} |\Lambda^{1/2}| |\Lambda^{-1/2}|^3\big) L^{-d}, \tag{10}$$

where $\Lambda := \mathrm{Var}\, X$ and where the constant $C(d,\gamma,N,K)$ depends in a polynomial way on d, N, K, and B.

Furthermore, we have for any symmetric positive definite $\Lambda \in \mathbb{R}^{d\times d}$ with $\Lambda \geq \mathrm{Var}\, X$ and $|\Lambda - \mathrm{Var}\, X| \leq L^{-d}$

$$\begin{aligned} \mathcal{D}(X - \mathbb{E}[X], \mathcal{N}_\Lambda) \leq & C(d,\gamma,N,K) B^3 (\log L)^{C(d,\gamma)} \big(L^{-d} |\Lambda^{1/2}| |\Lambda^{-1/2}|^3\big) L^{-d} \\ & + C(d,N) (\log L)^{C(d,\gamma)} |\Lambda - \mathrm{Var}\, X|^{1/2}, \end{aligned} \tag{11}$$

providing a better bound in the case of degenerate covariance matrices $\mathrm{Var}\, X$.

For sums of random variables with multilevel local dependence structure, we also prove the following simple (and far from optimal) result on moderate deviations. Its proof makes use of the previous Theorem 4 and an auxiliary concentration estimate provided in Lemma 13 which is a consequence of Bennett's inequality.

Theorem 5. *Let $d \geq 1$ and $L \geq 2$. Consider a random field a on $\mathbb{R}^d$ subject to the assumption of finite range of dependence (A) or an L-periodic random field subject to the assumption of finite range of dependence (A'). Let $X = X(a)$ be a random variable that may be written as a sum of random variables with multilevel local dependence structure $X = \sum_{m=0}^{1+\log_2 L} \sum_{i \in 2^m\mathbb{Z}^d \cap [0,L)^d} X_i^m$ in the sense of Definition 3.*

Then there exists $\beta = \beta(d,\gamma) > 0$ and a positive definite symmetric matrix $\Lambda \in \mathbb{R}^{N\times N}$ with $|\Lambda - \mathrm{Var}\, X| \leq C(d,\gamma,N,K) B^2 L^{-2\beta} L^{-d}$ such that for any measurable $A \subset \mathbb{R}^N$ we have the estimate

$$\mathbb{P}\big[X \in A\big] \leq \int_{\{x \in \mathbb{R}^N : \mathrm{dist}(x,A) \leq L^{-\beta} L^{-d/2}\}} \mathcal{N}_\Lambda(x)\,dx + C(d,\gamma,N,K) \exp\Big(-\frac{c}{B^C} L^{2\beta}\Big).$$

3. Normal Approximation with an Abstract Multilevel Dependency Structure

We now establish a result on quantitative normal approximation for a sum of random variables with a more abstract dependence structure allowing for multiple dependency ranges. More precisely, for a finite index set I we consider a sum of random variables X_i

$$X := \sum_{i\in I} X_i$$

to each of which a "dependency level" $m(i) \in \mathbb{N}$ is assigned. In the application of our next result in the proof of Theorem 4, the dependency level $m(i)$ will correspond to a "range of dependence" $\sim 2^{m(i)}$, i. e. each random variable X_y^m from Definition 3 will be assigned the dependency level m. However, the dependence structure we introduce now is more general than Definition 3 (and in particular does not include any explicit reference to an underlying random field or even a spatial dimension).

The potential dependencies of the random variables X_i shall be encoded by a matrix $\chi_{ij} \in \{0,1\}$ and a tensor $\chi_{ijk} \in \{0,1\}$ with the following property:

- For all $i \in I$, the random variable X_i is independent from the collection of all random variables X_j with $\chi_{ij} = 0$, $j \in I$.
- The matrix χ_{ij} is symmetric, i. e. $\chi_{ij} = \chi_{ji}$.
- For all $i, j \in I$, the pair of random variables (X_i, X_j) is independent from the collection of all random variables X_k with $\chi_{ijk} = 0$, $k \in I$.
- The tensor χ_{ijk} is symmetric in its first two indices, i. e. $\chi_{ijk} = \chi_{jik}$.

Furthermore, we suppose that for any $j \in I$ and any $n \in \mathbb{N}$ with $n > m(j)$ there exists an assignment $i^n(j)$ which assigns the random variable X_j of level $m(j)$ to another random variable $X_{i^n(j)}$ of the (higher) level n with more possible dependencies:

- We have $m(i^n(j)) = n$.
- It must hold that $\chi_{ii^n(j)} \geq \chi_{ij}$, in other words $\chi_{ij} = 1$ implies $\chi_{ii^n(j)} = 1$.
- For all i and k the inequality $\chi_{ii^n(j)k} \geq \chi_{ijk}$ must be true, in other words $\chi_{ijk} = 1$ implies $\chi_{ii^n(j)k} = 1$.

Note that this dependence structure is reminiscent of the local dependence structure in the results of [43, 3, 4, 36, 14, 38, 13], but in addition features multiple dependency ranges.

Theorem 6. *Let I be a finite index set, let $N \in \mathbb{N}$, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For any $i \in I$, let X_i be an $\mathbb{R}^N$-valued random variable with vanishing expectation $\mathbb{E}[X_i] = 0$ and finite third moment. To each of the X_i, let a number $m(i) \in \mathbb{N}$ be assigned.*

Let $\chi_{ij} \in \{0,1\}$, $i, j \in I$, and $\chi_{ijk} \in \{0,1\}$, $i, j, k \in I$, be indicator functions for possible dependencies subject to the assumptions preceding the theorem. Let $m(i)$, $i \in I$, and $i^n(j)$, $i, j \in I$, $m(j) < n$, be as above.

Then the probability distribution of the sum

$$X := \sum_{i\in I} X_i \tag{12}$$

may be approximated by an N-dimensional Gaussian with covariance matrix

$$\Lambda := \sum_{i\in I} \sum_{j\in I} \chi_{ij} \mathbb{E}[X_i \otimes X_j] \tag{13}$$

in the distance $\mathcal{D}(X, \mathcal{N}_\Lambda)$ in the following sense:

Introduce the abbreviations

$$Z_{ij} := \sum_{k\in I:\chi_{ijk}=1} X_k, \tag{14a}$$

$$Z_i := \sum_{j\in I:\chi_{ij}=1} X_j, \tag{14b}$$

$$Y_{il} := \sum_{j\in I:m(j)<m(i),i^{m(i)}(j)=l,\chi_{ij}=1} \Big(X_i\otimes X_j - \mathbb{E}[X_i\otimes X_j]\Big), \tag{14c}$$

$$W_{ij} := X_i\otimes X_j - \mathbb{E}[X_i\otimes X_j], \tag{14d}$$

and let $\bar X_i, \bar Z_{ij}, \bar Z_i, \bar Y_{il}, \bar W_{ij} > 0$ *be arbitrary.*

Let $\ell\in\mathbb{N}$ *and* $\varepsilon\in(0,\frac{1}{2}]$ *be such that the condition*

$$\begin{aligned}
&\sum_{i\in I,m(i)\leq\ell} \frac{N^{9/2}|\Lambda^{-1/2}|^3}{\varepsilon}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar W_{ij}\bar Z_{ij} + 2\bar Y_{ij}\,\bar Z_{ij}\big) + \bar X_i \bar Z_i^2\Bigg)\\
&+\sum_{i\in I,m(i)>\ell} N^4|\Lambda^{-1}||\log\varepsilon|\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar W_{ij} + 2\bar Y_{ij}\big) + \bar X_i\bar Z_i\Bigg)\\
&\leq \frac{1}{C}
\end{aligned}\tag{15}$$

is satisfied, where the universal constant C *is given by the proof below.*

Then the normal approximation result

$$\mathcal{D}(X,\mathcal{N}_\Lambda) \leq C\sqrt{N}|\Lambda^{1/2}|\varepsilon + \mathcal{R}_{lowlevel} + \mathcal{R}_{alllevel} + \mathcal{R}_{tail} \tag{16}$$

holds true, where

(17a)

$$\mathcal{R}_{lowlevel} := \frac{CN^{9/2}|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{i\in I,m(i)\leq\ell}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar W_{ij}\bar Z_{ij}^2 + 2\bar Y_{ij}\,\bar Z_{ij}^2\big) + \bar X_i\bar Z_i^3\Bigg),$$

(17b)

$$\mathcal{R}_{alllevel} := CN^{9/2}|\Lambda^{-1/2}|^2|\log\varepsilon|\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar W_{ij}\bar Z_{ij} + 2\bar Y_{ij}\,\bar Z_{ij}\big) + \bar X_i\bar Z_i^2\Bigg),$$

(17c)

$$\begin{aligned}
\mathcal{R}_{tail} :=& C|\Lambda^{-1}|N^{3/2}\varepsilon^{-N}\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\mathbb{E}\Big[|W_{ij}||Z_{ij}|(\chi_{|Z_{ij}|>\bar Z_{ij}} + \chi_{|W_{ij}|>\bar W_{ij}})\\
&\qquad + \sum_{l\in I:m(l)=m(i),\chi_{il}=1}\mathbb{E}\Big[|Y_{il}|\,|Z_{il}|(\chi_{|Z_{il}|>\bar Z_{il}} + \chi_{|Y_{il}|>\bar Y_{il}})\Big]\\
&\qquad + \mathbb{E}\Big[|X_i||Z_i|^2(\chi_{|Z_i|>\bar Z_i} + \chi_{|X_i|>\bar X_i})\Big]\Bigg).
\end{aligned}$$

Concerning our theorem, a few remarks are in order:

- Note that our theorem is tailored towards an application to random variables X_k with stretched exponential moments: For typical dependence structures like our multilevel local dependence structure from Definition 3, by concentration estimates like in Lemma 17 the random variables Z_{ij}, Z_i,

and Y_{il} then also satisfy stretched exponential moment bounds. As a consequence, one may choose $\bar{Z}_{ij} \sim C||Z_{ij}||_{\exp^{\gamma_0}}(\log L)^{1/\gamma_0}$ and similar choices for the other variables.
- In our theorem $\varepsilon \in (0, \frac{1}{2}]$ is a free parameter that by (16) one would typically choose in such a way that $|\Lambda^{1/2}|\varepsilon$ is of the order of the normal approximation error.
- The dependence of our estimates on the dimension N is not optimal and we make no attempt to get close to optimality.

Before proving our theorem, let us state two auxiliary results that are required for the proof.

The majority of the following auxiliary results on the existence of solutions to the (smoothed) Stein's equation are standard, see e. g. [13]; however, in order to keep the paper self-contained and in order to keep the dependence on the dimension N explicit, we provide the detailed argument in the appendix.

Proposition 7. *Let $N \in \mathbb{N}$ and let $\Lambda \in \mathbb{R}^{N\times N}$ be a symmetric positive definite matrix. Let $\bar{L} > 0$.*

For any $\phi \in \Phi_\Lambda^{\bar{L}}$ and any $\varepsilon > 0$, there exists a function f_ε with the following properties:

a) *The functions ϕ and f_ε are related through the "mollified" Stein equation*

$$-(\nabla\cdot\Lambda\nabla f_\varepsilon)(x) + (x\cdot\nabla f_\varepsilon)(x) = \phi_\varepsilon(x) - \int_{\mathbb{R}^N}\phi_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dz \tag{18}$$

with

$$\phi_\varepsilon(x) := \int_{\mathbb{R}^N}\phi(\sqrt{1-\varepsilon^2}x - \varepsilon z)\mathcal{N}_\Lambda(z)\,dz. \tag{19}$$

b) *The third derivative of f_ε is subject to the uniform bound*

$$|\nabla^3 f_\varepsilon(x)| \le 15|\Lambda^{-1}|\varepsilon^{-N} \tag{20}$$

for all $x \in \mathbb{R}^N$.

c) *For any $\delta > 0$ and $K := 2\sqrt{N}+1$, the function*

$$H_\delta^\varepsilon(x) := 2(\mathcal{N}_{\delta^2\,\mathrm{Id}_N} * \mathrm{osc}_{K\delta}\nabla^2 f_\varepsilon)(x)$$

is an upper bound for the oscillation of $\nabla^2 f_\varepsilon$

$$(\mathrm{osc}_\delta\nabla^2 f_\varepsilon)(x) = \sup_{|z_a|\le\delta,|z_b|\le\delta}|\nabla^2 f_\varepsilon(x+z_a) - \nabla^2 f_\varepsilon(x+z_b)| \le H_\delta^\varepsilon(x) \tag{21}$$

and satisfies the estimates

$$\int_{\mathbb{R}^N} H_\delta^\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx \le 10^2 N^{3/2}|\Lambda^{-1}|\,|\log\varepsilon|\,\delta \tag{22}$$

and

$$\frac{1}{2\cdot 10^4 N^{5/2}|\Lambda^{-1}||\log\varepsilon|}H_\delta^\varepsilon \in \Phi_\Lambda^{\tilde{L}} \tag{23}$$

for any $\tilde{L} \ge 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1)$.

d) *For any $\delta > 0$, the function*

$$H'_{\varepsilon,\delta}(x) := |\nabla^3 f_\varepsilon(x)| + 2(\mathcal{N}_{\delta^2\,\mathrm{Id}_N} * \mathrm{osc}_{K\delta}\nabla^3 f_\varepsilon)(x)$$

with $K := 2\sqrt{N}+1$ is an upper bound for the supremum of $\nabla^3 f_\varepsilon$ in a δ-neighborhood

$$\sup_{|z|\leq\delta} |\nabla^3 f_\varepsilon|(x+z) \leq H'_{\varepsilon,\delta}(x) \tag{24}$$

and satisfies the estimates

$$\int_{\mathbb{R}^N} H'_{\varepsilon,\delta}(x)\mathcal{N}_\Lambda(x)\,dx \leq 10^2 N^3 |\Lambda^{-1/2}|^2 \Big(|\log\varepsilon| + \frac{|\Lambda^{-1/2}|\delta}{\varepsilon}\Big) \tag{25}$$

and

$$\frac{\varepsilon}{10^4 N^3 |\Lambda^{-1/2}|^3} H'_{\varepsilon,\delta} \in \Phi_\Lambda^{\tilde{L}} \tag{26}$$

for any $x_0 \in \mathbb{R}^N$ and any $\tilde{L} \geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1)+\varepsilon^{-N}$.

The following lemma enables us to replace the class of functions $\Phi_\Lambda^{\bar{L}}$ in the definition of the distance $\mathcal{D}^{\bar{L}}$ by a class of mollified functions ϕ_ε (with $\phi \in \Phi_\Lambda^{\bar{L}}$). This is a standard argument in the theory of normal approximation by Stein's method [8, 24]; however, we provide the proof of our version of the lemma in the appendix, as it keeps an explicit (though not optimal) dependence on the dimension N.

Lemma 8. *Given $\phi \in \Phi_\Lambda$, define*

$$\phi_\varepsilon(x) := \int_{\mathbb{R}^N} \phi(\sqrt{1-\varepsilon^2}x - \varepsilon z)\mathcal{N}_\Lambda(z)\,dz.$$

Introduce the "smoothed" distance

$$\mathcal{D}_\varepsilon^{\bar{L}}(X,\mathcal{N}_\Lambda) := \sup_{\phi\in\Phi_\Lambda^{\bar{L}}} \bigg(\mathbb{E}[\phi_\varepsilon(X)] - \int_{\mathbb{R}^N} \phi_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\bigg).$$

For any $0<\varepsilon\leq\frac{1}{2}$ and any $\bar{L} \geq 2\cdot 4^N\varepsilon^{-N}$ we then have

$$\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda) \leq 20\sqrt{N}|\Lambda^{1/2}|\varepsilon + 10^3 N^{3/2}\mathcal{D}_\varepsilon^{\bar{L}}(X,\mathcal{N}_\Lambda). \tag{27}$$

We now turn to the proof of our result on quantitative normal approximation for a sum X of random variables X_i with "multilevel local dependence structure".

Proof of Theorem 6. First we observe that by the fact

$$\mathcal{D}(X,\mathcal{N}_\Lambda) = \lim_{\bar{L}\to\infty} \mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda)$$

it suffices to establish the bound for $\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda)$ for all large enough but finite $\bar{L}<\infty$.

The proof proceeds using Stein's method of normal approximation. Proposition 7 provides for any $0<\varepsilon<1$ and for any $\phi \in \Phi_\Lambda^{\bar{L}}$ a function f_ε that solves the "smoothed" Stein's equation

$$-\nabla\cdot(\Lambda\nabla f_\varepsilon(x)) + x\cdot\nabla f_\varepsilon(x) = \phi_\varepsilon(x) - \int_{\mathbb{R}^N} \phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz. \tag{28}$$

The smoothing result of Lemma 8 allows us to control the distance $\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda)$ by

$$\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda) \leq 20\sqrt{N}|\Lambda^{1/2}|\varepsilon + 10^3 N^{3/2} \sup_{\phi\in\Phi_\Lambda^{L}} \bigg|\mathbb{E}[\phi_\varepsilon(X)] - \int_{\mathbb{R}^N} \phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\bigg|.$$

Thus, by the "mollified" Stein's equation (28) the estimate

(29)
$$\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda) \leq 20\sqrt{N}|\Lambda^{1/2}|\varepsilon + 10^3 N^{3/2} \sup_{\phi\in\Phi_\Lambda^{\bar{L}}} \Big|\mathbb{E}\Big[(-\nabla\cdot(\Lambda\nabla f_\varepsilon) + x\cdot\nabla f_\varepsilon)(X)\Big]\Big|$$

holds true. Hence, in order to obtain an estimate for $\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda)$ it suffices to bound

$$\mathbb{E}\big[(-\nabla\cdot(\Lambda\nabla f_\varepsilon) + x\cdot\nabla f_\varepsilon)(X)\big]$$

uniformly in $\phi \in \Phi_\Lambda^{\bar{L}}$.

In order to derive such an estimate, we may rewrite using the definition of X (see (12))

$$\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big] = \mathbb{E}\big[X\cdot\nabla f_\varepsilon(X)\big] = \sum_{i\in I}\mathbb{E}\big[X_i\cdot\nabla f_\varepsilon(X)\big].$$

As by definition of χ_{ij} the quantities X_i and $X - \sum_{j\in I}\chi_{ij}X_j$ are stochastically independent, we obtain by adding and subtracting $\mathbb{E}[X_i\cdot\nabla f_\varepsilon(X - \sum_{j\in I}\chi_{ij}X_j)]$ and using the fact that $\mathbb{E}[X_i] = 0$

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\\
&= \sum_{i\in I}\mathbb{E}[X_i]\cdot\mathbb{E}\Big[\nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big)\Big]\\
&\quad + \sum_{i\in I}\mathbb{E}\Big[X_i\cdot\Big(\nabla f_\varepsilon(X) - \nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big)\Big)\Big]\\
&= \sum_{i\in I}\mathbb{E}\Big[X_i\cdot\Big(\nabla f_\varepsilon(X) - \nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big)\Big)\Big].
\end{aligned}$$

Next, we infer by adding zero in order to obtain an expression reminiscent of Taylor expansion

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\\
&= \sum_{i\in I}\mathbb{E}\Big[X_i\otimes\Big(\sum_{j\in I}\chi_{ij}X_j\Big) : \nabla^2 f_\varepsilon(X)\Big]\\
&\quad + \sum_{i\in I}\mathbb{E}\Big[X_i\cdot\Big(\nabla f_\varepsilon(X) - \nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big) - \Big(\sum_{j\in I}\chi_{ij}X_j\Big)\cdot\nabla^2 f_\varepsilon(X)\Big)\Big].
\end{aligned}$$

Adding again zero, we deduce

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\\
&= \sum_{i\in I}\sum_{j\in I}\chi_{ij}\mathbb{E}[X_i\otimes X_j] : \mathbb{E}[(\nabla^2 f_\varepsilon)(X)]\\
&\quad + \sum_{i\in I}\sum_{j\in I}\chi_{ij}\mathbb{E}\Big[\Big(X_i\otimes X_j - \mathbb{E}[X_i\otimes X_j]\Big) : \nabla^2 f_\varepsilon(X)\Big]\\
&\quad + \sum_{i\in I}\mathbb{E}\Big[X_i\cdot\Big(\nabla f_\varepsilon(X) - \nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big) - \Big(\sum_{j\in I}\chi_{ij}X_j\Big)\cdot\nabla^2 f_\varepsilon(X)\Big)\Big].
\end{aligned}$$

We now observe that the double sum $\sum_{i\in I}\sum_{j\in I}\chi_{ij}\mathbb{E}[X_i\otimes X_j]$ in the first term on the right-hand side is equal to Λ. Splitting the second term on the right-hand side and using the symmetry with respect to i and j, we obtain

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]-\mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)]\\
&=\sum_{i\in I}\sum_{j\in I:m(j)=m(i)}\chi_{ij}\mathbb{E}\bigg[\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big):\nabla^2 f_\varepsilon(X)\bigg]\\
&\quad+2\sum_{i\in I}\sum_{j\in I:m(j)<m(i)}\chi_{ij}\mathbb{E}\bigg[\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big):\nabla^2 f_\varepsilon(X)\bigg]\\
&\quad+\sum_{i\in I}\mathbb{E}\bigg[X_i\cdot\bigg(\nabla f_\varepsilon(X)-\nabla f_\varepsilon\Big(X-\sum_{j\in I}\chi_{ij}X_j\Big)-\Big(\sum_{j\in I}\chi_{ij}X_j\Big)\cdot\nabla^2 f_\varepsilon(X)\bigg)\bigg].
\end{aligned}$$

Using the fact that for $m(i)\geq m(j)$ by definition of χ_{ijk} and $i^m(j)$ the quantities $X_i\otimes X_j$ and $X-\sum_{k\in I}\chi_{ii^{m(i)}(j)k}X_k$ are stochastically independent (recall that $\chi_{ijk}=1$ implies $\chi_{ii^{m(i)}(j)k}=1$ for $m(j)<m(i)$) and making use of the fact that $\chi_{ij}=\chi_{ij}\chi_{ii^{m(i)}(j)}$, we infer

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]-\mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)]\\
&=\sum_{i\in I}\sum_{j\in I:m(j)=m(i)}\chi_{ij}\mathbb{E}\bigg[\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big)\\
&\qquad\qquad:\bigg(\nabla^2 f_\varepsilon(X)-\nabla^2 f_\varepsilon\Big(X-\sum_{k\in I}\chi_{ijk}X_k\Big)\bigg)\bigg]\\
&\quad+2\sum_{i\in I}\sum_{j\in I:m(j)<m(i)}\chi_{ij}\chi_{ii^{m(i)}(j)}\mathbb{E}\bigg[\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big)\\
&\qquad\qquad:\bigg(\nabla^2 f_\varepsilon(X)-\nabla^2 f_\varepsilon\Big(X-\sum_{k\in I}\chi_{ii^{m(i)}(j)k}X_k\Big)\bigg)\bigg]\\
&\quad+\sum_{i\in I}\mathbb{E}\bigg[X_i\cdot\bigg(\nabla f_\varepsilon(X)-\nabla f_\varepsilon\Big(X-\sum_{j\in I}\chi_{ij}X_j\Big)-\Big(\sum_{j\in I}\chi_{ij}X_j\Big)\cdot\nabla^2 f_\varepsilon(X)\bigg)\bigg].
\end{aligned}$$

We now split the sum in the second term on the right-hand side by introducing the additional variable $l:=i^{m(i)}(j)$. The reason for introducing this additional splitting is that the sum $\sum_{j\in I:m(j)<m(i),i^{m(i)}(j)=l}\chi_{ij}\big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\big)$ is subject to a better estimate than obtained by a standard triangle inequality, as one may exploit the stochastic independence of many terms in the sum. We deduce

$$\begin{aligned}
&\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]-\mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)]\\
&=\sum_{i\in I}\sum_{j\in I:m(j)=m(i)}\chi_{ij}\mathbb{E}\bigg[\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big)\\
&\qquad\qquad:\bigg(\nabla^2 f_\varepsilon(X)-\nabla^2 f_\varepsilon\Big(X-\sum_{k\in I}\chi_{ijk}X_k\Big)\bigg)\bigg]\\
&\quad+2\sum_{i\in I}\sum_{l\in I:m(l)=m(i)}\chi_{il}\mathbb{E}\Bigg[\sum_{j\in I:m(j)<m(i),i^{m(i)}(j)=l}\chi_{ij}\Big(X_i\otimes X_j-\mathbb{E}[X_i\otimes X_j]\Big)
\end{aligned}$$

$$: \Big(\nabla^2 f_\varepsilon(X) - \nabla^2 f_\varepsilon\Big(X - \sum_{k\in I}\chi_{ilk}X_k\Big)\Big)\Big]$$

$$+ \sum_{i\in I}\mathbb{E}\Big[X_i \cdot \Big(\nabla f_\varepsilon(X) - \nabla f_\varepsilon\Big(X - \sum_{j\in I}\chi_{ij}X_j\Big) - \Big(\sum_{j\in I}\chi_{ij}X_j\Big)\cdot \nabla^2 f_\varepsilon(X)\Big)\Big].$$

We intend to use the Taylor formula respectively the definition of the oscillation osc to bound each of the three terms on the right-hand side. For example, the terms in the first sum may either be estimated as

$$\Big|\Big(X_i \otimes X_j - \mathbb{E}[X_i\otimes X_j]\Big) : \Big(\nabla^2 f_\varepsilon(X) - \nabla^2 f_\varepsilon\Big(X - \sum_{k\in I}\chi_{ijk}X_k\Big)\Big)\Big|$$
$$\le \big|X_i\otimes X_j - \mathbb{E}[X_i\otimes X_j]\big| \cdot \Big|\sum_{k\in I}\chi_{ijk}X_k\Big| \cdot \sup_{|z|\le|\sum_{k\in I}\chi_{ijk}X_k|}|\nabla^3 f_\varepsilon(X+z)|$$

or as

$$\Big|\Big(X_i \otimes X_j - \mathbb{E}[X_i\otimes X_j]\Big) : \Big(\nabla^2 f_\varepsilon(X) - \nabla^2 f_\varepsilon\Big(X - \sum_{k\in I}\chi_{ijk}X_k\Big)\Big)\Big|$$
$$\le \big|X_i\otimes X_j - \mathbb{E}[X_i\otimes X_j]\big| \cdot \big(\mathrm{osc}_{|\sum_{k\in I}\chi_{ijk}X_k|}\nabla^2 f_\varepsilon\big)(X).$$

Using the abbreviations Z_{ij}, Z_i, Y_{il}, and W_{ij} from (14) and distinguishing the cases $|Z_{ij}| > \bar{Z}_{ij}$ and $|Z_{ij}| \le \bar{Z}_{ij}$ (and $|Z_i| > \bar{Z}_i$ and $|Z_i| \le \bar{Z}_i$, and so forth, for fixed but arbitrary constants $\bar{X}_i$, $\bar{Z}_{ij}$, $\bar{Z}_i$, $\bar{Y}_{il}$, $\bar{W}_{ij}$) as well as (in the latter cases) the cases $m(i) \le \ell$ and $m(i) > \ell$, we obtain by treating the other sums analogously

(30)

$$\begin{aligned}
&\Big| - \mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)] + \mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\Big| \\
&\le \sum_{i\in I, m(i)\le\ell}\ \sum_{j\in I: m(j)=m(i)} \chi_{ij}\mathbb{E}\Big[\bar{W}_{ij}\bar{Z}_{ij}\sup_{|z|\le\bar{Z}_{ij}}|\nabla^3 f_\varepsilon(X+z)|\Big] \\
&\quad + \sum_{i\in I, m(i)>\ell}\ \sum_{j\in I: m(j)=m(i)} \chi_{ij}\mathbb{E}\Big[\bar{W}_{ij}\,(\mathrm{osc}_{\bar{Z}_{ij}}\nabla^2 f_\varepsilon)(X)\Big] \\
&\quad + \sum_{i\in I}\ \sum_{j\in I: m(j)=m(i)} \chi_{ij}\mathbb{E}\Big[|W_{ij}||Z_{ij}|(\chi_{|Z_{ij}|>\bar{Z}_{ij}} + \chi_{|W_{ij}|>\bar{W}_{ij}})\sup_{z\in\mathbb{R}^N}|\nabla^3 f_\varepsilon(z)|\Big] \\
&\quad + 2\sum_{i\in I, m(i)\le\ell}\ \sum_{l\in I: m(l)=m(i)} \chi_{il}\mathbb{E}\Big[\bar{Y}_{il}\,\bar{Z}_{il}\sup_{|z|\le\bar{Z}_{il}}|\nabla^3 f_\varepsilon|(X+z)\Big] \\
&\quad + 2\sum_{i\in I, m(i)>\ell}\ \sum_{l\in I: m(l)=m(i)} \chi_{il}\mathbb{E}\Big[\bar{Y}_{il}\,(\mathrm{osc}_{\bar{Z}_{il}}\nabla^2 f_\varepsilon)(X)\Big] \\
&\quad + 2\sum_{i\in I}\ \sum_{l\in I: m(l)=m(i)} \chi_{il}\mathbb{E}\Big[|Y_{il}|\,|Z_{il}|(\chi_{|Z_{il}|>\bar{Z}_{il}} + \chi_{|Y_{il}|>\bar{Y}_{il}})\sup_{z\in\mathbb{R}^N}|\nabla^3 f_\varepsilon(z)|\Big] \\
&\quad + \sum_{i\in I, m(i)\le\ell}\mathbb{E}\Big[\bar{X}_i\bar{Z}_i^2\sup_{|z|\le\bar{Z}_i}|\nabla^3 f_\varepsilon|(X+z)\Big] \\
&\quad + \sum_{i\in I, m(i)>\ell}\mathbb{E}\Big[\bar{X}_i\bar{Z}_i(\mathrm{osc}_{\bar{Z}_i}\nabla^2 f_\varepsilon)(X)\Big]
\end{aligned}$$

$$+\sum_{i\in I}\mathbb{E}\Big[|X_i||Z_i|^2(\chi_{|Z_i|>\bar{Z}_i}+\chi_{|X_i|>\bar{X}_i})\sup_{z\in\mathbb{R}^N}|\nabla^3 f_\varepsilon(z)|\Big].$$

Reordering terms and using the definitions of H_δ^ε and $H'_{\varepsilon,\delta}$ in Proposition 7 as well as the bound (20), we deduce

$$\begin{aligned}
&\Big|-\mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)]+\mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\Big|\\
&\leq \sum_{i\in I,m(i)\leq\ell}\ \sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\bar{Z}_{ij}\mathbb{E}\big[H'_{\varepsilon,\bar{Z}_{ij}}(X)\big]\\
&\quad+2\sum_{i\in I,m(i)\leq\ell}\ \sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\,\bar{Z}_{il}\mathbb{E}\big[H'_{\varepsilon,\bar{Z}_{il}}(X)\big]\\
&\quad+\sum_{i\in I,m(i)\leq\ell}\bar{X}_i\bar{Z}_i^2\mathbb{E}\big[H'_{\varepsilon,\bar{Z}_i}(X)\big]\\
&\quad+\sum_{i\in I,m(i)>\ell}\ \sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\mathbb{E}\big[H^\varepsilon_{\bar{Z}_{ij}}(X)\big]\\
&\quad+2\sum_{i\in I,m(i)>\ell}\ \sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\mathbb{E}\big[H^\varepsilon_{\bar{Z}_{il}}(X)\big]\\
&\quad+\sum_{i\in I,m(i)>\ell}\bar{X}_i\bar{Z}_i\mathbb{E}\big[H^\varepsilon_{\bar{Z}_i}(X)\big]\\
&\quad+15\Lambda^{-1}\varepsilon^{-N}\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\mathbb{E}\Big[|W_{ij}||Z_{ij}|(\chi_{|Z_{ij}|>\bar{Z}_{ij}}+\chi_{|W_{ij}|>\bar{W}_{ij}})\\
&\qquad\qquad+2\sum_{l\in I:m(l)=m(i)}\chi_{il}\mathbb{E}\Big[|Y_{il}|\,|Z_{il}|(\chi_{|Z_{il}|>\bar{Z}_{il}}+\chi_{|Y_{il}|>\bar{Y}_{il}})\Big]\\
&\qquad\qquad+\mathbb{E}\Big[|X_i||Z_i|^2(\chi_{|Z_i|>\bar{Z}_i}+\chi_{|X_i|>\bar{X}_i})\Big]\Bigg).
\end{aligned}$$

For $L\geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1|)$, by Proposition 7 we may control

$$\begin{aligned}
&\mathbb{E}[H_\delta^\varepsilon(X)]\\
&=\mathbb{E}[H_\delta^\varepsilon(X)]-\int_{\mathbb{R}^N}H_\delta^\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz+\int_{\mathbb{R}^N}H_\delta^\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\
&\overset{(23)}{\leq} 2\cdot 10^4N^{5/2}|\Lambda^{-1}||\log\varepsilon|\mathcal{D}^L(X,\mathcal{N}_\Lambda)+\int_{\mathbb{R}^N}H_\delta^\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\
&\overset{(22)}{\leq} 2\cdot 10^4N^{5/2}|\Lambda^{-1}||\log\varepsilon|\mathcal{D}^L(X,\mathcal{N}_\Lambda)+10^2N^{3/2}|\Lambda^{-1}|\,|\log\varepsilon|\,\delta\\
&\leq CN^{5/2}|\Lambda^{-1}||\log\varepsilon|\mathcal{D}^L(X,\mathcal{N}_\Lambda)+CN^{3/2}|\Lambda^{-1}|\,|\log\varepsilon|\,\delta.
\end{aligned}$$

Similarly, for $L\geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1|)+\varepsilon^{-N}$ we have by Proposition 7

$$\begin{aligned}
&\mathbb{E}[H'_{\varepsilon,\delta}(X)]\\
&=\mathbb{E}[H'_{\varepsilon,\delta}(X)]-\int_{\mathbb{R}^N}H'_{\varepsilon,\delta}(z)\mathcal{N}_\Lambda(z)\,dz+\int_{\mathbb{R}^N}H'_{\varepsilon,\delta}(z)\mathcal{N}_\Lambda(z)\,dz\\
&\leq\frac{10^4N^3|\Lambda^{-1/2}|^3}{\varepsilon}\mathcal{D}^L(X,\mathcal{N}_\Lambda)+10^2N^3|\Lambda^{-1/2}|^2\Big(|\log\varepsilon|+\frac{|\Lambda^{-1/2}|\delta}{\varepsilon}\Big)
\end{aligned}$$

$$\leq \frac{CN^3|\Lambda^{-1/2}|^3}{\varepsilon}\mathcal{D}^L(X,\mathcal{N}_\Lambda) + CN^3|\Lambda^{-1/2}|^2\Big(|\log\varepsilon| + \frac{|\Lambda^{-1/2}|\delta}{\varepsilon}\Big).$$

Using these two estimates in the preceding bound and collecting terms, we obtain

$$\begin{aligned}
&\Big| -\mathbb{E}[(\nabla\cdot\Lambda\nabla f_\varepsilon)(X)] + \mathbb{E}\big[(x\cdot\nabla f_\varepsilon)(X)\big]\Big| \\
&\leq \mathcal{D}^L(X,\mathcal{N}_\Lambda)\frac{CN^3|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{i\in I, m(i)\leq\ell}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\bar{Z}_{ij} \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\,\bar{Z}_{il} + \bar{X}_i\bar{Z}_i^2\Bigg) \\
&\quad + \frac{CN^3|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{i\in I, m(i)\leq\ell}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\bar{Z}_{ij}^2 \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\,\bar{Z}_{il}^2 + \bar{X}_i\bar{Z}_i^3\Bigg) \\
&\quad + CN^3|\Lambda^{-1/2}|^2|\log\varepsilon|\sum_{i\in I, m(i)\leq\ell}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\bar{Z}_{ij} \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\,\bar{Z}_{il} + \bar{X}_i\bar{Z}_i^2\Bigg) \\
&\quad + \mathcal{D}^L(X,\mathcal{N}_\Lambda)\cdot CN^{5/2}|\Lambda^{-1}||\log\varepsilon|\sum_{i\in I, m(i)>\ell}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij} \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il} + \bar{X}_i\bar{Z}_i\Bigg) \\
&\quad + CN^{3/2}|\Lambda^{-1}||\log\varepsilon|\sum_{i\in I, m(i)>\ell}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar{W}_{ij}\bar{Z}_{ij} \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar{Y}_{il}\bar{Z}_{il} + \bar{X}_i\bar{Z}_i^2\Bigg) \\
&\quad + 15|\Lambda^{-1}|\varepsilon^{-N}\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\mathbb{E}\Big[|W_{ij}||Z_{ij}|(\chi_{|Z_{ij}|>\bar{Z}_{ij}} + \chi_{|W_{ij}|>\bar{W}_{ij}})\Big] \\
&\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\mathbb{E}\Big[|Y_{il}|\,|Z_{il}|(\chi_{|Z_{il}|>\bar{Z}_{il}} + \chi_{|Y_{il}|>\bar{Y}_{il}})\Big] \\
&\qquad\qquad + \mathbb{E}\Big[|X_i||Z_i|^2(\chi_{|Z_i|>\bar{Z}_i} + \chi_{|X_i|>\bar{X}_i})\Big]\Bigg).
\end{aligned}$$

Plugging in this bound into (29) and using the abbreviations (17) for a suitable choice of the constant C, we obtain

$$\begin{aligned}
&\mathcal{D}^{\bar{L}}(X,\mathcal{N}_\Lambda) \\
&\leq 20\sqrt{N}|\Lambda^{1/2}|\varepsilon + \frac{1}{2}\mathcal{R}_{low} + \frac{1}{2}\mathcal{R}_{alllevel} + \frac{1}{2}\mathcal{R}_{tail}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\mathcal{D}^{\bar L}(X,\mathcal{N}_\Lambda)\frac{CN^{9/2}|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{i\in I,m(i)\leq \ell}\bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar W_{ij}\bar Z_{ij}\\
&\qquad\qquad\qquad\qquad\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar Y_{il}\,\bar Z_{il}+\bar X_i\bar Z_i^2\bigg)\\
&+\frac{1}{2}\mathcal{D}^{\bar L}(X,\mathcal{N}_\Lambda)\cdot CN^4|\Lambda^{-1}||\log\varepsilon|\sum_{i\in I,m(i)>\ell}\bigg(\sum_{j\in I:m(j)=m(i)}\chi_{ij}\bar W_{ij}\\
&\qquad\qquad\qquad\qquad\qquad\qquad + 2\sum_{l\in I:m(l)=m(i)}\chi_{il}\bar Y_{il}+\bar X_i\bar Z_i\bigg).
\end{aligned}
$$

Defining the C in the condition (15) to be precisely the C in the previous estimate, we see that the condition (15) allows to absorb the last two terms on the right-hand side in the preceding estimate, thereby proving (16). □

4. Normal approximation for multilevel local dependence structures

We now proceed to the proof of our normal approximation result Theorem 4. The proof is essentially a reduction to the more abstract normal approximation result provided by Theorem 6.

Proof of Theorem 4. **Step 1: Proof of the estimate (10).** We first derive the normal approximation result in the case of nondegenerate Var X, i. e. estimate (10). Let X_y^m be the random variables from Definition 3. To derive our result on normal approximation, we apply Theorem 6 in the following way: We define the index set I to consist of all pairs $i=(m,y)$ with $0\leq m\leq 1+\log_2 L$ and $y\in 2^m\mathbb{Z}^d\cap[0,L)^d$, and set $X_i:=X_y^m$ as well as $m(i):=m$ for $i=(m,y)$. Furthermore, let us introduce the notation $y_i:=y$ for $i=(m,y)$. Finally, for $j\in I$ and $n>m(j)$ we set $i^n(j):=(n,\tilde y)$, where $\tilde y$ is given by $2^n\lfloor\frac{y_j}{2^n}\rfloor$ (where $\lfloor\cdot\rfloor$ denotes the component-wise floor).

We equip this collection of random variables X_i with the following dependence structure: We set

$$
\chi_{ij}=\begin{cases}1 & \text{if } \operatorname{dist}_\infty^{\mathsf{per}}(y_i+K\log L[-2^{m(i)},2^{m(i)}]^d,y_j+K\log L[-2^{m(j)},2^{m(j)}]^d)\\ & \qquad\qquad\leq 2\cdot 2^{\max\{m(i),m(j)\}}K\log L,\\ 0 & \text{otherwise,}\end{cases}
$$

and

$$
\chi_{ijk}=\begin{cases}1 & \text{if } \operatorname{dist}_\infty^{\mathsf{per}}(y_i+K\log L[-2^{m(i)},2^{m(i)}]^d,y_k+K\log L[-2^{m(k)},2^{m(k)}]^d)\\ & \qquad\qquad\leq 2\cdot 2^{\max\{m(i),m(j),m(k)\}}K\log L,\\ 1 & \text{if } \operatorname{dist}_\infty^{\mathsf{per}}(y_j+K\log L[-2^{m(j)},2^{m(j)}]^d,y_k+K\log L[-2^{m(k)},2^{m(k)}]^d)\\ & \qquad\qquad\leq 2\cdot 2^{\max\{m(i),m(j),m(k)\}}K\log L,\\ 0 & \text{otherwise,}\end{cases}
$$

where we recall that $\operatorname{dist}_\infty^{\mathsf{per}}(U,V)$ denotes the periodicity-adjusted maximum-norm distance, i. e. $\operatorname{dist}_\infty^{\mathsf{per}}(U,V):=\inf_{x\in U,y\in V}\inf_{k\in\mathbb{Z}^d}|x-y-Lk|$.

We readily verify by Definition 3 that $\chi_{ij}=0$ indeed entails independence of X_i and the collection of all X_j with $\chi_{ij}=0$ (recall that by Definition 3 each X_i only depends on $a|_{y_i+K\log L[-2^{m(i)},2^{m(i)}]^d}$ and that a has finite range of dependence

1). Similarly, the pair (X_i, X_j) is independent from the family of all X_k with $\chi_{ijk} = 0$. It is furthermore easy to verify that χ_{ij} and χ_{ijk} satisfy the further conditions on a dependence structure as defined in Section 3, namely the symmetry $\chi_{ij} = \chi_{ji}$ and $\chi_{ijk} = \chi_{jik}$ as well as the conditions $\chi_{ii^n(j)} \geq \chi_{ij}$ and $\chi_{ii^n(j)k} \geq \chi_{ijk}$ for $n > m(j)$ (the latter two properties follow from $y_j + K\log L[-2^{m(j)}, 2^{m(j)}] \subset y_{i^n(j)} + K\log L[-2^n, 2^n]$).

With these choices, we have

$$X = \sum_{i\in I} X_i.$$

Using the notation of Theorem 6 it is now the next goal to bound Z_{ij}, Z_i, Y_{il}, and W_{ij}. By (9), we trivially have for

$$W_{ij} = X_i \otimes X_j - \mathbb{E}[X_i \otimes X_j]$$

the bound

$$||W_{ij}||_{\exp^{\gamma/2}} \leq C(\gamma, N) B^2 L^{-2d}. \tag{31a}$$

Rewriting (with $p(m)$ denoting the smallest exponent with $2^{p(m)} > 4K\log L\, 2^m$, but at most $p(m) = \lfloor \log_2 L\rfloor + 1$)

$$\begin{aligned}
Z_i &= \sum_{j\in I:\chi_{ij}=1} X_j = \sum_{m=0}^{1+\log_2 L} \sum_{j\in I:\chi_{ij}=1, m(j)=m} X_j \\
&= \sum_{m=0}^{m(i)} \sum_{\substack{y\in 2^m\mathbb{Z}^d\cap[0,L)^d:\operatorname{dist}^{\mathsf{per}}_\infty(y+K\log L\cdot 2^m[-1,1]^d, y_i+K\log L\cdot 2^{m(i)}[-1,1]^d)\\ \leq 2\cdot 2^{m(i)}K\log L}} X^m_y \\
&\quad + \sum_{m=m(i)+1}^{1+\log_2 L} \sum_{\substack{y\in 2^m\mathbb{Z}^d\cap[0,L)^d:\operatorname{dist}^{\mathsf{per}}_\infty(y+K\log L\cdot 2^m[-1,1]^d, y_i+K\log L\cdot 2^{m(i)}[-1,1]^d)\\ \leq 2\cdot 2^{m}K\log L}} X^m_y \\
&= \sum_{m=0}^{m(i)} \sum_{z\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d} \sum_{\substack{y\in 2^{p(m)}\mathbb{Z}^d\cap[0,L)^d: y+z\in[0,L)^d,\\ \operatorname{dist}^{\mathsf{per}}_\infty(y+z+K\log L\cdot 2^m[-1,1]^d, y_i+K\log L\cdot 2^{m(i)}[-1,1]^d)\\ \leq 2\cdot 2^{m(i)}K\log L}} X^m_{y+z} \\
&\quad + \sum_{m=m(i)+1}^{1+\log_2 L} \sum_{\substack{y\in 2^m\mathbb{Z}^d\cap[0,L)^d:\operatorname{dist}^{\mathsf{per}}_\infty(y+K\log L\cdot 2^m[-1,1]^d, y_i+K\log L\cdot 2^{m(i)}[-1,1]^d)\\ \leq 2\cdot 2^{m}K\log L}} X^m_y,
\end{aligned}$$

we see that the sum on each level $m \leq m(i)$ may be written as the sum of $\leq C(d)(K\log L)^d$ sums of $\leq C(d)(2^{m(i)-m})^d$ independent random variables (to the latter sums we may apply a concentration estimate), while the sum on the levels $m > m(i)$ consists only of $\leq C(d)(K\log L)^d$ terms. By Lemma 17 and (9) we deduce for $\gamma_1 := \gamma/(\gamma+1)$

$$||Z_i||_{\exp^{\gamma_1}} \leq \sum_{m=0}^{m(i)} C(d,\gamma,N)(K\log L)^d \cdot C(\gamma)\big((2^{m(i)-m})^d\big)^{1/2} \cdot BL^{-d} + \sum_{m=m(i)+1}^{1+\log_2 L} C(d)(K\log L)^d BL^{-d}$$

and therefore

(31b)
$$||Z_i||_{\exp^{\gamma_1}} \leq C(d,\gamma,N)B(K\log L)^d(2^{m(i)})^{d/2}L^{-d} + C(d)BK^d(\log L)^{d+1}L^{-d}.$$

In a very similar way (note that for our choice of χ_{ijk} and χ_{ij} we have $\chi_{ijk} \leq \chi_{ik} + \chi_{jk}$), we can estimate

$$Z_{ij} = \sum_{k\in I:\chi_{ijk}=1} X_k$$

as

(31c)
$$||Z_{ij}||_{\exp^{\gamma_1}} \leq C(d,\gamma,N)B(K\log L)^d(2^{\max\{m(i),m(j)\}})^{d/2}L^{-d} + C(d)BK^d(\log L)^{d+1}L^{-d}.$$

It remains to bound

$$Y_{il} = \sum_{j\in I:m(j)<m(i),i^{m(i)}(j)=l,\chi_{ij}=1} \Big(X_i \otimes X_j - \mathbb{E}[X_i\otimes X_j]\Big).$$

Note that

$$Y_{il} = X_i \otimes \hat{Y}_{il} - \mathbb{E}[X_i \otimes \hat{Y}_{il}],$$

where

$$\hat{Y}_{il} := \sum_{j\in I:m(j)<m(i),i^{m(i)}(j)=l,\chi_{ij}=1} X_j = \sum_{m=0}^{m(i)-1} \sum_{j\in I:m(j)=m,i^{m(i)}(j)=l,\chi_{ij}=1} X_j.$$

For each level m, the inner sum may again be written as a sum of $C(d)(K\log L)^d$ sums of $\lesssim \frac{(2^{m(i)-m})^d}{(K\log L)^d}$ independent random variables (recall that $y_{i^m(j)} = 2^m\lfloor\frac{y_j}{2^m}\rfloor$). We therefore get by the concentration estimate in Lemma 17 and (9) for $\gamma_1 := \gamma/(\gamma+1)$

$$\begin{aligned}||\hat{Y}_{il}||_{\exp^{\gamma_1}} &\leq \sum_{m=0}^{m(i)-1} C(d)(K\log L)^d C(\gamma,N)\sqrt{(2^{m(i)-m})^d}\cdot BL^{-d}\\ &\leq C(d,\gamma,N)B(K\log L)^d(2^{m(i)})^{d/2}L^{-d}.\end{aligned}$$

As a consequence, by (9) and Lemma 16a we obtain for $\gamma_2 = 1/(1/\gamma+1/\gamma_1) = \gamma/(\gamma+2)$

$$||Y_{il}||_{\exp^{\gamma_2}} \leq C(d,\gamma,N)B^2(K\log L)^d(2^{m(i)})^{d/2}L^{-2d}. \tag{31d}$$

Choosing now the constants $\bar{X}_i$, $\bar{Z}_{ij}$, $\bar{Z}_i$, $\bar{Y}_{il}$, $\bar{W}_{il}$, for some $S \geq 1$ as

$$\bar{X}_i := C(d,\gamma,N)B(S\log L)^{1/\gamma}L^{-d}, \tag{32a}$$

$$\bar{W}_{ij} := C(d,\gamma,N)B^2(S\log L)^{2/\gamma}L^{-2d}, \tag{32b}$$

$$\bar{Z}_i := C(d,\gamma,N)B(S\log L)^{1/\gamma_1}(K\log L)^{d+1}(2^{m(i)})^{d/2}L^{-d}, \tag{32c}$$

$$\bar{Z}_{ij} := C(d,\gamma,N)B(S\log L)^{1/\gamma_1}(K\log L)^{d+1}(2^{\max\{m(i),m(j)\}})^{d/2}L^{-d}, \tag{32d}$$

$$\bar{Y}_{il} := C(d,\gamma,N)B^2(S\log L)^{1/\gamma_2}(K\log L)^{d}(2^{m(i)})^{d/2}L^{-2d}, \tag{32e}$$

we obtain that the condition (15) is certainly satisfied if

$$\begin{aligned}
&\sum_{i\in I, m(i)\leq \ell} \frac{N^{9/2}|\Lambda^{-1/2}|^3}{\varepsilon}\bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1} C(d,\gamma,N)B^3(S\log L)^{1/\gamma_2+1/\gamma_1}(K\log L)^{2d+1}(2^{m(i)})^d L^{-3d}\\
&\qquad\qquad + C(d,\gamma,N)B^3(S\log L)^{1/\gamma+2/\gamma_1}(K\log L)^{2d+2}(2^{m(i)})^d L^{-3d}\bigg)\\
&\quad+\sum_{i\in I, m(i)>\ell} N^4|\Lambda^{-1}||\log\varepsilon|\bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1} C(d,\gamma,N)B^2(S\log L)^{1/\gamma_2}(K\log L)^{d}(2^{m(i)})^{d/2}L^{-2d}\\
&\qquad\qquad + C(d,\gamma,N)B^2(S\log L)^{1/\gamma+1/\gamma_1}(K\log L)^{d+1}(2^{m(i)})^{d/2}L^{-2d}\bigg)\\
&\quad\leq \frac{1}{C}.
\end{aligned}$$

Note that the sum $\sum_{i\in I,m(i)=m}$ consists of the order of $L^d(2^m)^{-d}$ terms and the sum $\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}$ consists of the order of $C(d)(K\log L)^d$ terms. Thus, the condition (15) is satisfied if

$$\begin{aligned}
&\sum_{m=0}^{\ell} C(d,\gamma,N)B^3S^{1/\gamma_2+1/\gamma_1}\frac{K^{3d+2}(\log L)^{3d+1+1/\gamma_2+1/\gamma_1}|\Lambda^{-1/2}|^3L^{-2d}}{\varepsilon}\\
&\quad+\sum_{m=\ell}^{1+\log_2 L} C(d,\gamma,N)B^2S^{1/\gamma_2}K^{2d}|\Lambda^{-1}||\log\varepsilon|(\log L)^{2d+1/\gamma_2}(2^m)^{-d/2}L^{-d}\\
&\quad\leq \frac{1}{C}.
\end{aligned}$$

This condition is satisfied for the choice

$$\varepsilon := C(d,\gamma,N)B^3S^{1/\gamma_2+1/\gamma_1}K^{3d+2}(\log L)^{1+3d+1+1/\gamma_2+1/\gamma_1}|\Lambda^{-1/2}|^3L^{-2d} \tag{33}$$

and ℓ as the smallest nonnegative integer with

$$(2^\ell)^{d/2} \geq C(d,\gamma,N)B^2S^{1/\gamma_2}K^{2d}|\Lambda^{-1}||\log(B^3|\Lambda^{-1/2}|^3L^{-2d})|(\log L)^{2d+1/\gamma_2}L^{-d}. \tag{34}$$

Note that the choice of ε entails $\varepsilon \geq B^3|\Lambda^{-1/2}|^3L^{-2d}$, which in turn implies by the bound $|\Lambda| \leq C(d,\gamma,N,K)B^2L^{-d}|\log L|^{C(\gamma)}$ from Lemma 9 that the lower bound $\varepsilon \geq c(d,\gamma,N,K)L^{-d/2}|\log L|^{C(d)}$ holds.

Let us now estimate the terms $\mathcal{R}_{lowlevel}$, $\mathcal{R}_{alllevel}$, and $\mathcal{R}_{tail}$ in Theorem 6. We have by Lemma 16b and our choices (32) as well as our bounds (31a), (31b), (31c), and (31d) (note that the inner sums in the next two lines contain at most

$C(d)(K\log L)^d$ summands each, while the outer sum consists of $\leq L^d(2^m)^{-d}$ summands of level $m(i)=m$)

$$
\begin{aligned}
\mathcal{R}_{tail} &= C|\Lambda^{-1}|N^{3/2}\varepsilon^{-N}\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\mathbb{E}\Big[|W_{ij}||Z_{ij}|(\chi_{|Z_{ij}|>\bar{Z}_{ij}}+\chi_{|W_{ij}|>\bar{W}_{ij}})\Big]\\
&\qquad+\sum_{l\in I:m(l)=m(i),\chi_{il}=1}\mathbb{E}\Big[|Y_{il}|\,|Z_{il}|(\chi_{|Z_{il}|>\bar{Z}_{il}}+\chi_{|Y_{il}|>\bar{Y}_{il}})\Big]\\
&\qquad+\mathbb{E}\Big[|X_i||Z_i|^2(\chi_{|Z_i|>\bar{Z}_i}+\chi_{|X_i|>\bar{X}_i})\Big]\Bigg)\\
&\leq C|\Lambda^{-1}|N^{3/2}\varepsilon^{-N}C(d)L^d(K\log L)^d\\
&\qquad\times\Bigg(\max_{i,j}\mathbb{E}[\chi_{|Z_{ij}|>\bar{Z}_{ij}}+\chi_{|W_{ij}|>\bar{W}_{ij}}]^{1/2}\max_{i,j}\mathbb{E}[|W_{ij}|^2|Z_{ij}|^2]^{1/2}\\
&\qquad\quad+\max_{i,j}\mathbb{E}[\chi_{|Z_{ij}|>\bar{Z}_{ij}}+\chi_{|Y_{ij}|>\bar{Y}_{ij}}]^{1/2}\mathbb{E}[|Y_{ij}|^2|Z_{ij}|^2]^{1/2}\\
&\qquad\quad+\max_{i}\mathbb{E}[\chi_{|Z_i|>\bar{Z}_i}+\chi_{|X_i|>\bar{X}_i}]^{1/2}\mathbb{E}[|X_i|^2|Z_i|^4]^{1/2}\Bigg)\\
&\leq C|\Lambda^{-1}|N^{3/2}\varepsilon^{-N}C(d)L^d(K\log L)^d\times L^{-c(d)S/2}(K\log L)^{C(d,\gamma)}B^3L^{-3d}\\
&\leq C(d,\gamma,K,N)|\Lambda^{-1}|B^3L^{-5d}
\end{aligned}
$$

for S chosen large enough (depending only on d, γ, and N).

Using also (33), we obtain first in case $\ell\geq 1$ (note that S has now been chosen as a constant depending only on d, γ, and N)

$$
\begin{aligned}
\mathcal{R}_{lowlevel} &= \frac{CN^{9/2}|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{i\in I,m(i)\leq\ell}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar{W}_{ij}\bar{Z}_{ij}^2+2\bar{Y}_{ij}\,\bar{Z}_{ij}^2\big)+\bar{X}_i\bar{Z}_i^3\Bigg)\\
&\leq\frac{C(d,\gamma,K,N)|\Lambda^{-1/2}|^3}{\varepsilon}\sum_{m=0}^{\ell}\left(\frac{L}{2^m}\right)^d\cdot(K\log L)^d\cdot B^4(\log L)^{C(d,\gamma)}(2^m)^{3d/2}L^{-4d}\\
&\leq C(d,\gamma,K,N)B\sum_{m=0}^{\ell}\left(\frac{L}{2^m}\right)^d\cdot(K\log L)^d\cdot(\log L)^{C(d,\gamma)}(2^m)^{3d/2}L^{-2d}\\
&\leq C(d,\gamma,K,N)B\cdot(\log L)^{C(d,\gamma)}(2^\ell)^{d/2}L^{-d}\\
&\overset{(34)}{\leq} C(d,\gamma,K,N)B^3\cdot|\log(B^3|\Lambda^{-1/2}|^3L^{-2d})|(\log L)^{C(d,\gamma)}|\Lambda^{-1}|L^{-2d}.
\end{aligned}
$$

In the case $\ell=0$, the same overall estimate holds true by $|\Lambda^{-1}|\geq|\Lambda|^{-1}\geq cB^{-2}L^d|\log L|^{-d}$ (the second inequality being a consequence of Lemma 9).

Using again (33), we deduce

$$
\begin{aligned}
\mathcal{R}_{alllevel} &= CN^{9/2}|\Lambda^{-1/2}|^2|\log\varepsilon|\sum_{i\in I}\Bigg(\sum_{j\in I:m(j)=m(i),\chi_{ij}=1}\big(\bar{W}_{ij}\bar{Z}_{ij}+2\bar{Y}_{ij}\,\bar{Z}_{ij}\big)+\bar{X}_i\bar{Z}_i^2\Bigg)\\
&\leq C(d,\gamma,K,N)B^3|\Lambda^{-1/2}|^2|\log(B^3|\Lambda^{-1/2}|^3L^{-2d})|\\
&\qquad\times\sum_{m=0}^{1+\log_2 L}\left(\frac{L}{2^m}\right)^d\cdot(K\log L)^d\cdot(\log L)^{C(d,\gamma)}(2^m)^dL^{-3d}
\end{aligned}
$$

$$\leq C(d,\gamma,K,N)B^3|\Lambda^{-1/2}|^2|\log(B^3|\Lambda^{-1/2}|^3L^{-2d})|(\log L)^{C(d,\gamma)}L^{-2d}.$$

As a consequence, we deduce by Theorem 6 and (33)

$$\begin{aligned}&\mathcal{D}(X-\mathbb{E}[X],\mathcal{N}_\Lambda)\\&\quad\leq C(d,\gamma,N,K)B^3|\log(B^3|\Lambda^{-1/2}|^3L^{-2d})|\cdot(\log L)^{C(d,\gamma)}|\Lambda^{1/2}||\Lambda^{-1/2}|^3L^{-2d}.\end{aligned}$$

Using the bound $|\Lambda^{-1/2}|\geq|\Lambda|^{-1/2}\geq(c(d,\gamma,N,K)B^{-2}L^{-d}(\log L)^d)^{-1/2}$ (the last inequality being a consequence of Lemma 9), we infer the first estimate of our theorem.

Step 2: Proof of the estimate (11)**.** To obtain the second estimate in our theorem which provides a better bound for degenerate covariance matrices Var X, we repeat the preceding proof, however now adding $Q=L^d$ additional independent multivariate Gaussian random variables $G_1,\ldots,G_Q$ with zero expectation $\mathbb{E}[G_q]=0$ and variance Var $G_q=\frac{1}{Q}(\Lambda-\text{Var }X)$. This yields by an argument analogous to the above one (exploiting that the new additional random variables are independent from all others and using the fact that Var $G_q\leq\frac{1}{Q}(\Lambda-\text{Var }X)\leq L^{-2d}$)

$$\begin{aligned}&\mathcal{D}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)\\&\leq C(d,\gamma,N,K)B^3(\log L)^{C(d,\gamma)}\big(L^{-d}|\Lambda^{1/2}||\Lambda^{-1/2}|^3\big)L^{-d},\end{aligned}\tag{35}$$

where $G:=\sum_{q=1}^Q G_q$ and $\Lambda=\text{Var }X+\kappa L^{-d}\,\text{Id}_N$. Note that G has Gaussian moments with

$$||G||_{\exp^2}\leq C(N)\sqrt{|\Lambda-\text{Var }X|}.\tag{36}$$

Let $\phi\in\Phi_\Lambda$. Fixing $\bar G\geq|\Lambda-\text{Var }X|^{1/2}$, we may rewrite for ϕ_ε as defined in (19)

$$\begin{aligned}&\mathbb{E}[\phi_\varepsilon(X-\mathbb{E}[X])]-\int_{\mathbb{R}^N}\phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\&\leq\mathbb{E}[\phi_\varepsilon(X+G-\mathbb{E}[X])]-\int_{\mathbb{R}^N}\phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\&\quad+\mathbb{E}[|\phi_\varepsilon(X+G-\mathbb{E}[X])-\phi_\varepsilon(X-\mathbb{E}[X])|]\\&\leq\mathcal{D}^{\tilde L}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)+\mathbb{E}[\text{osc}_{\bar G}\phi_\varepsilon(X+G-\mathbb{E}[X])]\\&\quad+\sup_z|\nabla\phi_\varepsilon(z)|\;\mathbb{E}[|G|\chi_{|G|\geq\bar G}].\end{aligned}$$

An application of Lemma 10 c) and d) to the function

$$\hat h(x):=(\mathcal{N}_{\bar G^2\,\text{Id}}*\text{osc}_{(2\sqrt{N}+1)\bar G}\phi_\varepsilon)(x)$$

yields $\frac{1}{2\sqrt{N}+2}\cdot\frac{1}{40N}\hat h\in\Phi_\Lambda^{\tilde L}$ for $\tilde L$ large enough as well as $\text{osc}_{\bar G}\phi_\varepsilon(x)\leq 2\hat h(x)$ and thus

$$\begin{aligned}&\mathbb{E}[\phi_\varepsilon(X-\mathbb{E}[X])]-\int_{\mathbb{R}^N}\phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\&\leq\mathcal{D}^{\tilde L}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)+\mathbb{E}[2\hat h(X+G-\mathbb{E}[X])]+\sup_z|\nabla\phi_\varepsilon(z)|\;\mathbb{E}[|G|\chi_{|G|\geq\bar G}]\\&\leq C(N)\mathcal{D}^{\tilde L}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)+\int_{\mathbb{R}^N}2\hat h(z)\mathcal{N}_\Lambda(z)\,dz+\sup_z|\nabla\phi_\varepsilon(z)|\;\mathbb{E}[|G|\chi_{|G|\geq\bar G}]\\&\leq C(N)\mathcal{D}^{\tilde L}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)+C(N)\bar G+\sup_z|\nabla\phi_\varepsilon(z)|\;\mathbb{E}[|G|\chi_{|G|\geq\bar G}],\end{aligned}$$

where in the last step we have used the bound

$$\begin{aligned}
&\int_{\mathbb{R}^N} \hat{h}(z)\mathcal{N}_\Lambda(z)\,dz\\
&\leq \int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \mathcal{N}_{\bar{G}^2\,\mathrm{Id}}(z-x)\mathcal{N}_{\varepsilon^2\Lambda}(y)\operatorname{osc}_{\sqrt{1-\varepsilon^2}(2\sqrt{N}+1)\bar{G}}\phi(\sqrt{1-\varepsilon^2}x-y)\mathcal{N}_\Lambda(z)\,dx\,dy\,dz\\
&\overset{(4)}{\leq} (2\sqrt{N}+1)\bar{G}.
\end{aligned}$$

Choosing $\bar{G} := S|\log L||(\Lambda - \mathrm{Var}\ X)^{1/2}|$ and $\varepsilon := L^{-d}$ and using the estimate

$$\begin{aligned}
|\nabla\phi_\varepsilon(z)| &\leq \int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-\varepsilon^2}z-y)\mathcal{N}_{\varepsilon^2\Lambda}(y)\,dy\\
&\leq \varepsilon^{-N}\int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-\varepsilon^2}z-y)\mathcal{N}_\Lambda(y)\,dy \overset{(4)}{\leq} \varepsilon^{-N} \leq L^{C(d,N)}
\end{aligned}$$

as well as Lemma 16b and (36), we deduce for all $\phi\in\Phi_\Lambda$ that

$$\begin{aligned}
&\mathbb{E}[\phi_\varepsilon(X-\mathbb{E}[X])] - \int_{\mathbb{R}^N}\phi_\varepsilon(z)\mathcal{N}_\Lambda(z)\,dz\\
&\leq C(N)\mathcal{D}^{\tilde{L}}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda) + C(N)S|\log L||(\Lambda-\mathrm{Var}\ X)^{1/2}|\\
&\quad + L^{C(d,N)}\cdot|\Lambda^{1/2}|\cdot\exp(-cS\log L).
\end{aligned}$$

Lemma 9 and our assumption $|\Lambda - \mathrm{Var}\ X| \leq L^{-d}$ imply the upper bound $|\Lambda| \leq C(d,\gamma,N,K)B^2L^{-d}|\log L|^d$. As a consequence, choosing S large enough and using the notation from Lemma 8 we obtain

$$\begin{aligned}
\mathcal{D}^{\tilde{L}}_\varepsilon(X-\mathbb{E}[X],\mathcal{N}_\Lambda) &\leq C(N)\mathcal{D}^{\tilde{L}}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda) + C(d,N)|\log L||(\Lambda-\mathrm{Var}\ X)^{1/2}|\\
&\quad + C(d,\gamma,N,K)B|\log L|^C L^{-d}.
\end{aligned}$$

Using Lemma 8, we conclude that

$$\begin{aligned}
&\mathcal{D}^{\tilde{L}}(X-\mathbb{E}[X],\mathcal{N}_\Lambda)\\
&\leq C(N)|\Lambda^{1/2}|\varepsilon + C(N)\mathcal{D}^{\tilde{L}}_\varepsilon(X-\mathbb{E}[X],\mathcal{N}_\Lambda)\\
&\leq C(d,\gamma,N,K)B(\log L)^{d/2}L^{-d/2}\cdot L^{-d} + C(N)\mathcal{D}^{\tilde{L}}(X+G-\mathbb{E}[X],\mathcal{N}_\Lambda)\\
&\quad + C(d,N)|\log L||(\Lambda-\mathrm{Var}\ X)^{1/2}| + C(d,\gamma,N,K)B|\log L|^C L^{-d}\\
&\overset{(35)}{\leq} C(d,\gamma,N,K)|\log L|^{C(d,\gamma)}(B+B^3L^{-d}|\Lambda^{1/2}||\Lambda^{-1/2}|^3)L^{-d}\\
&\quad + C(d,N)|\log L||(\Lambda-\mathrm{Var}\ X)^{1/2}|.
\end{aligned}$$

This yields the estimate (11) upon noticing that $B \leq CB^3|\log L|^C L^{-d}|\Lambda^{1/2}||\Lambda^{-1/2}|^3$ (by Lemma 9 and our assumption $|\Lambda - \mathrm{Var}\ X| \leq L^{-d}$). □

We next prove our normal approximation result for integral functionals of random fields which may be approximated well by random fields with finite dependency range; the proof is a simple reduction to the statement of Theorem 4.

Proof of Theorem 2. By rescaling and translation we may restrict ourselves to random fields v and v_r that vanish identically outside of the cube $[0,L]^d$.

In order to establish our result for random field v and v_r which are supported on the cube $[0,L]^d$, we shall reduce it to the normal approximation result of Theorem 4. For each $0\leq m\leq \log_2 L+1$, introduce a partition of unity η_y, $y\in 2^m\mathbb{Z}^d\cap[0,L)^d$, with $\operatorname{supp}\eta_y \subset y + [0, 2^m]$ and $|\nabla^l \eta_y| \leq C(2^m)^{-l}$ for $0 \leq l \leq k$. We may then rewrite

$$\begin{aligned}
X &= L^{-d} \int_{\mathbb{R}^d} v_1 \xi \, dx + \sum_{m=0}^{\log_2 L} L^{-d} \int_{\mathbb{R}^d} (v_{2^{m+1}} - v_{2^m}) \xi \, dx + L^{-d} \int_{\mathbb{R}^d} (v - v_{2^{\log_2 L+1}}) \xi \, dx \\
&= \sum_{y \in \mathbb{Z}^d} \underbrace{L^{-d} \int_{\mathbb{R}^d} v_1 \xi \eta_y^0 \, dx}_{=: X_y^0} + \sum_{m=0}^{\log_2 L} \sum_{y \in 2^{m+1}\mathbb{Z}^d} \underbrace{L^{-d} \int_{\mathbb{R}^d} (v_{2^{m+1}} - v_{2^m}) \xi \eta_y^{m+1} \, dx}_{=: X_y^{m+1}} \\
&\quad + \underbrace{L^{-d} \int_{\mathbb{R}^d} (v - v_{2^{\log_2 L+1}}) \xi \, dx}_{=: X_0^{\log L+1}} .
\end{aligned}$$

The X_y^m give rise to a multilevel local dependence structure in the sense of Definition 3. In particular, we easily check that by the bound on v and the assumption (8) we have $||X_y^m||_{\exp^\gamma} \leq CL^{-d}$ for all m and all $y \in 2^m\mathbb{Z}^d \cap [0, L)^d$ (as we have $|\nabla^l (\xi \eta_y^m)| \leq C(2^m)^{-l}$). The statement of our theorem is then a direct consequence of the normal approximation result from Theorem 4. □

5. Proof of the Result on Moderate Deviations

We now provide the proof of our moderate deviations result for sums of random variables with multilevel local dependence structure.

Proof of Theorem 5. To simplify notation, we only consider the case $L = 2^{\hat{k}}$ for some $\hat{k} \in \mathbb{N}$; the proof for the general case is similar.

Step 1 (Decomposition of X).
We first decompose X into groups of terms $\mathcal{G}_i$ and "remainder terms" $\mathcal{R}$. The groups $\mathcal{G}_i$ are stochastically independent from each other and heuristically consist of the random variables summed over a cube of diameter ℓ, where $1 \ll \ell \ll L$ is an intermediate length scale that we are going to choose. The fact that the groups $\mathcal{G}_i$ sum up all random variables over an intermediate length scale ℓ allows us to apply normal approximation to the groups $\mathcal{G}_i$. The rest $\mathcal{R}$ basically corresponds to the random variables "between the groups" (to achieve independence of the groups) and the random variables with long-range dependencies. We shall prove that these terms are small in a suitable sense.

More precisely, we introduce an intermediate scale $\ell = 2^{\tilde{k}}$ and define $m_0 := \lfloor \log_2 \frac{\ell}{4K \log L} \rfloor$. This enables us to rewrite the random variable X as

$$\begin{aligned}
X &= \sum_{m=0}^{1+\log_2 L} \sum_{i \in 2^m\mathbb{Z}^d \cap [0,L)^d} X_i^m \\
&= \sum_{m=0}^{m_0} \sum_{i \in \ell\mathbb{Z}^d \cap [0,L)^d} \sum_{j \in 2^m\mathbb{Z}^d \cap [0,\ell)^d} X_{i+j}^m + \sum_{m=m_0+1}^{1+\log_2 L} \underbrace{\sum_{i \in 2^m\mathbb{Z}^d \cap [0,L)^d} X_i^m}_{=: \mathcal{R}^m}
\end{aligned}$$

which gives (defining $p(m)$ to be the smallest integer with $2^{p(m)} \geq 2^{m+2} K \log L$)

$$\begin{aligned} X = &\sum_{m=0}^{m_0} \sum_{i \in \ell\mathbb{Z}^d \cap [0,L)^d} \sum_{j \in 2^m \mathbb{Z}^d \cap [2^{p(m)}, \ell - 2^{p(m)})^d} X_{i+j}^m \\ &+ \sum_{m=0}^{m_0} \underbrace{\sum_{i \in \ell\mathbb{Z}^d \cap [0,L)^d} \sum_{j \in 2^m \mathbb{Z}^d \cap [0,\ell)^d \setminus [2^{p(m)}, \ell - 2^{p(m)})^d} X_{i+j}^m}_{=:\mathcal{R}^m} \\ &+ \sum_{m=m_0+1}^{1+\log_2 L} \underbrace{\sum_{i \in 2^m \mathbb{Z}^d \cap [0,L)^d} X_i^m}_{=:\mathcal{R}^m}. \end{aligned}$$

Exchanging the order of summation in the first term and defining

$$\mathcal{G}_i := \sum_{m=0}^{m_0} \sum_{j \in 2^m \mathbb{Z}^d \cap [2^{p(m)}, \ell - 2^{p(m)})^d} X_{i+j}^m, \tag{37}$$

we obtain

$$X = \sum_{i \in \ell\mathbb{Z}^d \cap [0,L)^d} \mathcal{G}_i + \sum_{m=0}^{m_0} \mathcal{R}^m + \sum_{m=m_0+1}^{1+\log_2 L} \mathcal{R}^m.$$

Step 2 (Estimate on the terms $\mathcal{R}^m$ for $m > m_0$).
The remaining terms on level m for $m_0 + 1 \leq m \leq 1 + \log_2 L$

$$\mathcal{R}^m := \sum_{i \in 2^m \mathbb{Z}^d \cap [0,L)^d} X_i^m$$

(observe that the sum consists of $(\frac{L}{2^m})^d$ terms) may be grouped into $\sim (K \log L)^d$ groups, each only consisting of $\sim \frac{(\frac{L}{2^m})^d}{(K \log L)^d}$ independent random variables: Choosing $p(m)$ as before to be the smallest integer with $2^{p(m)} \geq 2^{m+2} K \log L$ (but choosing $p(m) = \log_2 L$ if this integer were larger than $\log_2 L$), we have

$$\mathcal{R}^m = \sum_{j \in 2^m \mathbb{Z}^d \cap [0, 2^{p(m)})^d} \underbrace{\sum_{i \in 2^{p(m)} \mathbb{Z}^d \cap [0,L)^d} X_{i+j}^m}_{=:\mathcal{R}^{m,j}}.$$

The random variables in each $\mathcal{R}^{m,j}$ are now independent and we deduce from Lemma 17 (with $\tilde{\gamma} := \gamma/(\gamma+1)$)

$$\big|\big|\mathcal{R}^{m,j}\big|\big|_{\exp^{\tilde{\gamma}}} \leq C(\gamma) \sqrt{\Big(\frac{L}{2^{p(m)}}\Big)^d} \max_i ||X_i^m||_{\exp^\gamma}$$

and as a consequence

$$\begin{aligned}
&\Big|\Big|\sum_{m=m_0+1}^{1+\log_2 L}\mathcal{R}^m\Big|\Big|_{\exp^{\tilde\gamma}} = \Big|\Big|\sum_{m=m_0+1}^{1+\log_2 L}\sum_{j\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\mathcal{R}^{m,j}\Big|\Big|_{\exp^{\tilde\gamma}}\\
&\le \sum_{m=m_0+1}^{1+\log_2 L}\sum_{j\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d} C(\gamma)\sqrt{\Big(\frac{L}{2^{p(m)}}\Big)^d}\max_i ||X_i^m||_{\exp^\gamma}\\
&\le \sum_{m=m_0+1}^{1+\log_2 L}\Big(\frac{2^{p(m)}}{2^m}\Big)^d C(\gamma)\sqrt{\Big(\frac{L}{2^{p(m)}}\Big)^d}\max_i ||X_i^m||_{\exp^\gamma}\\
&\le \sum_{m=m_0+1}^{1+\log_2 L} C(d)(K\log L)^d C(\gamma)\sqrt{\Big(\frac{L}{2^m K\log L}\Big)^d}BL^{-d}\\
&\le C(d)C(\gamma)(K\log L)^{d/2}(2^{m_0})^{-d/2}BL^{-d/2}
\end{aligned}$$

which yields

$$\Big|\Big|\sum_{m=m_0+1}^{\log_2 L}\mathcal{R}^m\Big|\Big|_{\exp^{\tilde\gamma}} \le C(d,\gamma)B(K\log L)^d\ell^{-d/2}L^{-d/2}. \tag{38}$$

Step 3 (Estimate on the terms $\mathcal{R}^m$ for $m\le m_0$).
The remaining terms on level m for $0\le m\le m_0$

$$\mathcal{R}^m = \sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\sum_{j\in 2^m\mathbb{Z}^d\cap[0,\ell)^d\setminus[2^{p(m)},\ell-2^{p(m)})^d} X_{i+j}^m$$

may be grouped as follows into a sum of sums of independent random variables: We have

$$\begin{aligned}
\mathcal{R}^m &= \sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\sum_{j\in 2^{p(m)}\mathbb{Z}^d\cap[0,\ell)^d\setminus[2^{p(m)},\ell-2^{p(m)})^d}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d} X_{i+j+k}^m\\
&= \sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\underbrace{\sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\sum_{j\in 2^{p(m)}\mathbb{Z}^d\cap[0,\ell)^d\setminus[2^{p(m)},\ell-2^{p(m)})^d} X_{i+j+k}^m}_{=:\mathcal{R}^{m,k}}.
\end{aligned}$$

Note that $\mathcal{R}^{m,k}$ is a sum of $\le (\frac{L}{\ell})^d\cdot\frac{C(d)\ell^{d-1}2^{p(m)}}{(2^{p(m)})^d}$ independent random variables. An application of Lemma 17 yields

$$\big|\big|\mathcal{R}^{m,k}\big|\big|_{\exp^{\tilde\gamma}} \le C(\gamma)\sqrt{\Big(\frac{L}{\ell}\Big)^d\cdot\frac{C(d)\ell^{d-1}2^{p(m)}}{(2^{p(m)})^d}}\max_i ||X_i^m||_{\exp^\gamma}$$

which entails

$$\begin{aligned}
\Big|\Big|\sum_{m=0}^{m_0}\mathcal{R}^m\Big|\Big|_{\exp^{\tilde\gamma}} &= \Big|\Big|\sum_{m=0}^{m_0}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\mathcal{R}^{m,k}\Big|\Big|_{\exp^{\tilde\gamma}}\\
&\leq \sum_{m=0}^{m_0}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d} C(\gamma)\sqrt{\Big(\frac{L}{\ell}\Big)^d\cdot\frac{C(d)\ell^{d-1}2^{p(m)}}{(2^{p(m)})^d}}\max_i ||X_i^m||_{\exp^\gamma}\\
&\leq \sum_{m=0}^{m_0}\Big(\frac{2^{p(m)}}{2^m}\Big)^d C(\gamma)\sqrt{\Big(\frac{L}{\ell}\Big)^d\cdot\frac{C(d)\ell^{d-1}2^{p(m)}}{(2^{p(m)})^d}}\max_i ||X_i^m||_{\exp^\gamma}\\
&\leq \sum_{m=0}^{m_0} C(d)(K\log L)^d C(\gamma)\sqrt{\Big(\frac{L}{\ell}\Big)^d\cdot\frac{C(d)\ell^{d-1}}{(2^m)^{d-1}(K\log L)^{d-1}}}BL^{-d}\\
&\leq \sum_{m=0}^{m_0} C(d)C(\gamma)(K\log L)^{(d+1)/2}\ell^{-1/2}(2^m)^{-(d-1)/2}BL^{-d/2}
\end{aligned}$$

and as a consequence

$$\Big|\Big|\sum_{m=0}^{m_0}\mathcal{R}^m\Big|\Big|_{\exp^{\tilde\gamma}} \leq C(d,\gamma)B(K\log L)^{(d+1)/2}\ell^{-1/2}L^{-d/2}. \tag{39}$$

Step 4 (Estimate on the "bulk contributions" $\mathcal{G}_i$).
The formula (37) may be rewritten as (recalling that $p(m)$ is the smallest integer with $2^{p(m)} \geq 2^{m+2}K\log L$)

$$\mathcal{G}_i = \sum_{m=0}^{m_0}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\sum_{j\in 2^{p(m)}\mathbb{Z}^d\cap[2^{p(m)},\ell-2^{p(m)})^d} X_{i+j+k}^m,$$

which yields upon applying Lemma 17 to the inner sum

$$\begin{aligned}
||\mathcal{G}_i||_{\exp^{\tilde\gamma}} &\leq \sum_{m=0}^{m_0}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d} C(d,\gamma)\Big(\frac{\ell}{2^{p(m)}}\Big)^{d/2}BL^{-d}\\
&\leq \sum_{m=0}^{m_0} C(d,\gamma)B\Big(\frac{\ell}{2^m K\log L}\Big)^{d/2}(K\log L)^d L^{-d}\\
&\leq C(d,\gamma)B(K\log L)^{d/2}\ell^{d/2}L^{-d}. \qquad (40)
\end{aligned}$$

Next, to each of the groups

$$\mathcal{G}_i := \sum_{m=0}^{m_0}\sum_{j\in 2^m\mathbb{Z}^d\cap[2^{p(m)},\ell-2^{p(m)})^d} X_{i+j}^m,$$

we apply Theorem 4 with L replaced by ℓ; note that we may rescale our variables X_{i+j}^m in the group $\mathcal{G}_i$ by a factor of $(\frac{L}{\ell})^d$, as we have the bound $||X_{i+j}^m||_{\exp^\gamma} \leq BL^{-d}$ while to apply Theorem 4 to the group $\mathcal{G}_i$ we only need the estimate $||X_{i+j}^m||_{\exp^\gamma} \leq B\ell^{-d}$. We then obtain for any positive matrix $\Lambda_i \geq \mathrm{Var}\ \mathcal{G}_i$ with $|\Lambda_i - \mathrm{Var}\ \mathcal{G}_i| \leq$

$\ell^d L^{-2d}$

$$\begin{aligned}
&\mathcal{D}\Big[\Big(\frac{L}{\ell}\Big)^d \mathcal{G}_i, \mathcal{N}_{(L/\ell)^{2d}\Lambda_i}\Big] \\
&\quad\leq C(d,\gamma,N,K) B^3 (\log \ell)^{C(d,\gamma)} \ell^{-d} \Big(\frac{L}{\ell}\Big)^d |\Lambda_i^{1/2}| \Big(\frac{L}{\ell}\Big)^{-3d} |\Lambda_i^{-1/2}|^3 \ell^{-d} \\
&\qquad + C(d,N)(\log \ell)^{C(d,\gamma)} \Big(\frac{L}{\ell}\Big)^d |\Lambda_i - \mathrm{Var}\ \mathcal{G}_i|^{1/2}.
\end{aligned}$$

Rescaling and using the fact that the 1-Wasserstein distance $\mathcal{W}_1$ is bounded by our distance $\mathcal{D}$, we deduce

$$\begin{aligned}
\mathcal{W}_1[\mathcal{G}_i, \mathcal{N}_{\Lambda_i}] \leq & C(d,\gamma,N,K) B^3 (\log L)^{C(d,\gamma)} |\Lambda_i^{1/2}| |\Lambda_i^{-1/2}|^3 \ell^d L^{-3d} \\
& + C(d,N)(\log L)^{C(d,\gamma)} |\Lambda_i - \mathrm{Var}\ \mathcal{G}_i|^{1/2}.
\end{aligned}$$

We choose ℓ with $L^{1/2} \leq \ell \leq 2L^{1/2}$ and set

$$\Lambda_i := \mathrm{Var}\ \mathcal{G}_i + \ell^{d-1/4} L^{-2d}\, \mathrm{Id}\,. \tag{41}$$

Note that this choice entails

$$|\Lambda_i^{-1/2}| \leq \ell^{-d/2+1/8} L^d.$$

As a consequence of these estimates, the choice of ℓ, and (40), we obtain

$$\mathcal{W}_1[\mathcal{G}_i, \mathcal{N}_{\Lambda_i}] \leq C(d,\gamma,N,K) B^4 (\log L)^{C(d,\gamma)} \ell^{d/2-1/8} L^{-d}. \tag{42}$$

We are now in position to apply Lemma 14 to the sum $\sum_{i\in \ell\mathbb{Z}^d\cap[0,L)^d} \mathcal{G}_i$ (recall that the $\mathcal{G}_i$ form a collection of independent random variables). Note that in our setting we have $M = (L/\ell)^d$. By our estimates (42) and (40), in our application of Lemma 14 we may choose $b := C(d,\gamma,K) B^4 (\log L)^{d/2} \ell^{d/2} L^{-d}$ and any $\tau \leq \frac{1}{2}$ with

$$\tau \geq C(d,\gamma,N,K)(\log L)^{C(d,\gamma)} \ell^{-1/8}.$$

With this choice, Lemma 14 yields

$$\sum_{i\in \ell\mathbb{Z}^d\cap[0,L)^d} \mathcal{G}_i \overset{d}{=} Y + Z \tag{43}$$

where Y is a multivariate Gaussian random variable with covariance matrix

$$\tilde{\Lambda} := \sum_{i\in \ell\mathbb{Z}^d\cap[0,L)^d} \Lambda_i \tag{44}$$

and Z satisfies the estimate

$$\mathbb{P}[|Z| \geq r] \leq 3N \exp\bigg(-\frac{r^2}{C(d,\gamma,K,N) B^8 \tau |\log \tau|^{C(d,\gamma)} L^{-d}}\bigg)$$

for any

$$\begin{aligned}
r \leq & \sqrt{\tau |\log \tau|^{1/\tilde{\gamma}}}\, C (\log L)^{C(d,\gamma)} L^{-d/2} \\
& \times \min\bigg\{ \frac{\sqrt{(L/\ell)^d \tau |\log \tau|^{1/\gamma}}}{(2\log(2(L/\ell)^d))^{1/\tilde{\gamma}}}, \big(\tau |\log \tau|^{1/\tilde{\gamma}} (L/\ell)^d\big)^{\tilde{\gamma}/(4+2\tilde{\gamma})} \bigg\}.
\end{aligned}$$

A crude estimate on $|\log\tau|$ of the form $|\log\tau|\leq C(d,\gamma,N,K)(\log L)^{C(d,\gamma)}$ yields using also $(L/\ell)^d\tau\geq c$ (which holds by the lower bound on τ and the choice $\ell\sim L^{1/2}$)

$$\mathbb{P}[|Z|\geq r]\leq 3N\exp\bigg(-\frac{r^2}{\tau\cdot C(d,\gamma,K,N)B^8(\log L)^{C(d,\gamma)}L^{-d}}\bigg)$$

for any

$$r\leq L^{-d/2}\cdot\Big(\frac{L}{\ell}\Big)^{d\tilde\gamma/(4+2\tilde\gamma)}\cdot c(d,\gamma,N,K)(\log L)^{-C(d,\gamma)}\tau^{1/2+\tilde\gamma/(4+2\tilde\gamma)}.$$

We now set $\tau:=c(d,\gamma,N,K)L^{-\min\{d\tilde\gamma/2(2+2\tilde\gamma),1/32\}}$. The previous estimate then yields

$$\mathbb{P}[|Z|\geq r]\leq 3N\exp\bigg(-\frac{r^2}{L^{-\beta_1}\cdot C(d,\gamma,K,N)B^8(\log L)^{C(d,\gamma)}L^{-d}}\bigg)$$

for some $\beta_1>0$ as long as

$$r\leq L^{-d/2}\cdot c(d,\gamma,N,K)(\log L)^{-C(d,\gamma)}.$$

Choosing $r=\frac{1}{2}L^{-d/2}L^{-\beta_1/4}$ and using $L\geq C$ (note that upon changing the constants, our theorem is trivially true for $L\leq C$), we get

$$\mathbb{P}\bigg[|Z|\geq\frac{1}{2}L^{-d/2}L^{-\beta_1/4}\bigg]\leq C(N)\exp\bigg(-\frac{L^{\beta_1/2}}{C(d,\gamma,K,N)B^8(\log L)^{C(d,\gamma)}}\bigg).\tag{45}$$

Step 5 (Conclusion).
By (43), we know that the law of the sum of the main groups

$$\sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\mathcal{G}_i$$

is equal to the law of $Y+Z$, where Y is a Gaussian random variable with covariance matrix $\tilde\Lambda=\sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\tilde\Lambda_i$ and where Z satisfies the smallness estimate (45). By (44), (41), and (39) as well as (38) and (40) and the choice $\ell\sim L^{1/2}$, we see that

$$\begin{aligned}|\tilde\Lambda-\mathrm{Var}\;X|&\leq\bigg|\tilde\Lambda-\sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\mathrm{Var}\;\mathcal{G}_i\bigg|+\bigg|\mathrm{Var}\;X-\mathrm{Var}\sum_{i\in\ell\mathbb{Z}^d\cap[0,L)^d}\mathcal{G}_i\bigg|\\&\leq CL^{-d-1/8}+CB^2L^{-d-1/4}.\end{aligned}$$

The estimates (39) and (38) imply using Lemma 16b that

$$\mathbb{P}\bigg[\bigg|\sum_{m=0}^{\log_2 L+1}\mathcal{R}^m\bigg|\geq r\bigg]\leq C\exp\bigg(-\Big(\frac{r}{C(d,\gamma,N,K)B(\log L)^{C(d,\gamma)}\ell^{-1/2}L^{-d/2}}\Big)^{\tilde\gamma}\bigg)$$

and therefore for our choice $\ell:=L^{1/2}$

$$\mathbb{P}\bigg[\bigg|\sum_{m=0}^{\log_2 L+1}\mathcal{R}^m\bigg|\geq\frac{1}{2}L^{-1/8}\cdot L^{-d/2}\bigg]\leq C\exp\bigg(-\frac{L^{\tilde\gamma/8}}{C(d,\gamma,N,K)B}\bigg).$$

Combining this estimate with (45) and

$$X-\sum_{m=0}^{1+\log_2 L}\mathcal{R}^m\overset{d}{=}Y+Z,$$

we see that there exists $\beta=\beta(d,\gamma)>0$ and $\Lambda=\tilde\Lambda\in\mathbb{R}^{N\times N}_{\mathrm{sym}}$ with $\Lambda>0$ and

$$|\Lambda-\mathrm{Var}\ X|\leq CB^2L^{-1/8}L^{-d}$$

such that for any measurable $A\subset\mathbb{R}^N$ we have

$$\mathbb{P}\big[X\in A\big]\leq\int_{\{x\in\mathbb{R}^N:\mathrm{dist}(x,A)\leq L^{-\beta}L^{-d/2}\}}\mathcal{N}_\Lambda(x)\,dx+C\exp\Big(-\frac{c}{B^8}L^{2\beta}\Big).$$

□

We have made use of the following elementary lemma.

Lemma 9. *Let $d\geq 1$ and $L\geq 2$. Consider a random field a on $\mathbb{R}^d$ subject to the assumption of finite range of dependence (A) or an L-periodic random field subject to the assumption of finite range of dependence (A'). Let $X=X(a)$ be a random variable that is a sum of random variables with multilevel local dependence in the sense of Definition 3. Then for $\tilde\gamma:=\gamma/(\gamma+1)$ the concentration estimate*

$$||X-\mathbb{E}[X]||_{\exp^{\tilde\gamma}}\leq C(d,\gamma,K)B|\log L|^{d/2}L^{-d/2}$$

holds true.

Proof. Defining $p(m)$ to be the smallest integer with $2^{p(m)}\geq 2^{m+2}K|\log L|$ (but defining $p(m)=\log_2 L$ if this integer were larger than $\log_2 L$), we rewrite

$$X=\sum_{m=0}^{\log_2 L+1}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\ \sum_{j\in 2^{p(m)}\mathbb{Z}^d\cap[0,L)^d,j+k\in[0,L)^d}X^m_{j+k}.$$

By Definition 3, the inner sum is now a sum of independent random variables. Applying Lemma 17 to this sum, we obtain

$$\begin{aligned}
&||X-\mathbb{E}[X]||_{\exp^{\tilde\gamma}}\\
&\leq\sum_{m=0}^{\log_2 L+1}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}\Bigg|\Bigg|\sum_{j\in 2^{p(m)}\mathbb{Z}^d\cap[0,L)^d,j+k\in[0,L)^d}(X^m_{j+k}-\mathbb{E}[X^m_{j+k}])\Bigg|\Bigg|_{\exp^{\tilde\gamma}}\\
&\leq\sum_{m=0}^{\log_2 L+1}\sum_{k\in 2^m\mathbb{Z}^d\cap[0,2^{p(m)})^d}C(\gamma)\bigg(\frac{L^d}{(2^{p(m)})^d}\bigg)^{1/2}\max_i||X^m_i-\mathbb{E}[X^m_i]||_{\exp^\gamma}\\
&\leq\sum_{m=0}^{\log_2 L+1}\frac{(2^{p(m)})^d}{(2^m)^d}\cdot C(d,\gamma)\bigg(\frac{L^d}{(2^{p(m)})^d}\bigg)^{1/2}BL^{-d}\\
&\leq\sum_{m=0}^{\log_2 L+1}\frac{(2^m)^d(K\log L)^d}{(2^m)^d}\cdot C(d,\gamma)\bigg(\frac{L^d}{(2^m)^d(K\log L)^d}\bigg)^{1/2}BL^{-d}\\
&\leq C(d,\gamma,K)B(\log L)^{d/2}L^{-d/2}.
\end{aligned}$$

□

6. Proof of the main result

Proof of Theorem 2. We choose a function η with $0 \leq \eta \leq 1$, $\operatorname{supp} \eta \subset B_d(0)$, $|\nabla^n \eta| \leq C(n)$ for all $n \in \mathbb{N}$, and $\sum_{k\in\mathbb{Z}^d} \eta(x-k) = 1$ for all $x \in \mathbb{R}^d$. We next define $m_L := \lfloor \log_2 L \rfloor$ and $\tilde{L} := 2^{m_L}$ and write

$$v = (v - v_{\tilde{L}}) + v_1 + \sum_{m=2}^{m_L} (v_{2^m} - v_{2^{m-1}}).$$

Next, define

$$X_0^{m_L} := \int_{\mathbb{R}^d} (v - v_{\tilde{L}})\psi \, dx,$$

and

$$X_k^0 := \int_{\mathbb{R}^d} v_1 \eta_k \psi \, dx \qquad \text{for any } k \in \mathbb{Z}^d$$

and

$$X_i^m := \int_{\mathbb{R}^d} (v_{2^m} - v_{2^{m-1}})\eta()\psi \, dx$$

In a second step, we simply group the together, yielding .

□

Appendix A. Auxiliary results for Stein's method in the multivariate setting

In this appendix, we provide the proof of the bounds on the solutions to the "smoothed" Stein's equation stated in Proposition 7 and the estimate on the distance $\mathcal{D}^{\bar{L}}$ in terms of the smoothed distance $\mathcal{D}_\varepsilon^{\bar{L}}$ stated in Lemma 8.

Proof of Proposition 7. **Proof of a).** Following the argument of [24] (see also [13]), we consider the function

(46)

$$f_\varepsilon(x) = \frac{1}{2} \int_{\varepsilon^2}^1 \Big(\int_{\mathbb{R}^N} \phi(\sqrt{1-s}x - \sqrt{s}z) \mathcal{N}_\Lambda(z) \, dz - \int_{\mathbb{R}^N} \mathcal{N}_\Lambda(z)\phi(z) \, dz \Big) \frac{1}{1-s} \, ds$$

which in the case of smooth ϕ with compactly supported derivative $\nabla\phi$ satisfies

$$\begin{aligned}
&-(\nabla\cdot\Lambda\nabla f_\varepsilon)(x)+(x\cdot\nabla f_\varepsilon)(x)\\
&=-\frac{1}{2}\int_{\varepsilon^2}^1\int_{\mathbb{R}^N}\Lambda:(\nabla^2\phi)(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_\Lambda(z)\,dz\,ds\\
&\quad+\frac{1}{2}\int_{\varepsilon^2}^1\int_{\mathbb{R}^N}x\cdot(\nabla\phi)(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_\Lambda(z)\frac{1}{\sqrt{1-s}}\,dz\,ds\\
&=-\frac{1}{2}\int_{\varepsilon^2}^1\int_{\mathbb{R}^N}\Lambda:(\nabla^2\phi)(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_\Lambda(z)\,dz\,ds\\
&\quad-\frac{1}{2}\int_{\varepsilon^2}^1\int_{\mathbb{R}^N}z\cdot(\nabla\phi)(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_\Lambda(z)\frac{1}{\sqrt{s}}\,dz\,ds\\
&\quad+\int_{\varepsilon^2}^1-\frac{d}{ds}\int_{\mathbb{R}^N}\phi(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_\Lambda(z)\,dz\,ds\\
&=-\frac{1}{2}\int_{\varepsilon^2}^1\int_{\mathbb{R}^N}\phi(\sqrt{1-s}x-\sqrt{s}z)\frac{1}{s}\Big(\big(\nabla\cdot(\Lambda\nabla\mathcal{N}_\Lambda)\big)(z)+\nabla\cdot(z\mathcal{N}_\Lambda(z))\Big)\,dz\,ds\\
&\quad+\int_{\mathbb{R}^N}\phi(\sqrt{1-\varepsilon^2}x-\varepsilon z)\mathcal{N}_\Lambda(z)\,dz-\int_{\mathbb{R}^N}\phi(z)\mathcal{N}_\Lambda(z)\,dz\\
&=\int_{\mathbb{R}^N}\phi(\sqrt{1-\varepsilon^2}x-\varepsilon z)\mathcal{N}_\Lambda(z)\,dz-\int_{\mathbb{R}^N}\phi(z)\mathcal{N}_\Lambda(z)\,dz.
\end{aligned}$$

For general $\phi\in\Phi_\Lambda^{\bar{L}}$, this equation follows by approximation. Hence, (18) follows from the additional computation

$$\begin{aligned}
&\int_{\mathbb{R}^N}\phi_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dz\\
&=\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\phi(\sqrt{1-\varepsilon^2}x-\varepsilon z)\mathcal{N}_\Lambda(x)\mathcal{N}_\Lambda(z)\,dz\,dx\\
&\overset{(64)}{=}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\phi(\sqrt{1-\varepsilon^2}x-\varepsilon z)\mathcal{N}_\Lambda(\sqrt{1-\varepsilon^2}x-\varepsilon z)\mathcal{N}_\Lambda(\sqrt{1-\varepsilon^2}z+\varepsilon x)\,dz\,dx\\
&=\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\phi(\tilde{x})\mathcal{N}_\Lambda(\tilde{x})\mathcal{N}_\Lambda(\tilde{z})\,d\tilde{z}\,d\tilde{x}\\
&=\int_{\mathbb{R}^N}\phi(x)\mathcal{N}_\Lambda(x)\,dx
\end{aligned}$$

where we have used the fact that $(\tilde{x},\tilde{z})=(\sqrt{1-\varepsilon^2}x-\varepsilon z,\sqrt{1-\varepsilon^2}z+\varepsilon x)$ is an orthogonal linear transformation. This finishes the proof of Proposition 7a.

Proof of b). In order to establish (20), we derive from (46) the following representation for the third spatial derivative of f_ε:

$$\nabla^3 f_\varepsilon(x)=\frac{1}{2}\int_{\varepsilon^2}^1\frac{\sqrt{1-s}}{s^{3/2}}\int_{\mathbb{R}^N}\phi(\sqrt{1-s}x-\sqrt{s}z)\nabla^3\mathcal{N}_\Lambda(z)\,dz\,ds. \tag{47}$$

This entails

$$\begin{aligned}
&|\nabla^3 f_\varepsilon(x)| \\
&\leq \frac{1}{2}\int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{s}\int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)|\nabla^2\mathcal{N}_\Lambda|(z)\,dz\,ds \\
&\leq \frac{1}{2}\int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{s}\int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)\frac{|\nabla^2\mathcal{N}_\Lambda(z)|}{\mathcal{N}_{(1+\tau)\Lambda}(z)}\mathcal{N}_{(1+\tau)\Lambda}(z)\,dz\,ds \\
&\overset{(62)}{\leq} \frac{3}{2}(1+\tau)^{(N+2)/2}\tau^{-1}|\Lambda^{-1}| \\
&\qquad\times\int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{s}\int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_{(1+\tau)\Lambda}(z)\,dz\,ds.
\end{aligned}$$

Combining this estimate with the bound

$$\begin{aligned}
&\int_{\mathbb{R}^N} |\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_{(1+\tau)\Lambda}(z)\,dz \\
&= \int_{\mathbb{R}^N} |\nabla\phi|(\tilde{z})\mathcal{N}_{(1+\tau)\Lambda}\Big(\frac{1}{\sqrt{s}}(\sqrt{1-s}x-\tilde{z})\Big)\sqrt{s}^{-N}\,d\tilde{z} \\
&\leq \int_{\mathbb{R}^N} |\nabla\phi|(\tilde{z})\mathcal{N}_{(1+\tau)\Lambda}\big(\sqrt{1-s}x-\tilde{z}\big)s^{-N/2}\,d\tilde{z} \\
&\overset{(61)}{=} \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} |\nabla\phi|(\tilde{z})\mathcal{N}_\Lambda(\tilde{z}-\sqrt{1-s}x+w)\mathcal{N}_{\tau\Lambda}(w)s^{-N/2}\,d\tilde{z}\,dw \\
&\overset{(7)}{\leq} s^{-N/2}\int_{\mathbb{R}^N}\mathcal{N}_{\tau\Lambda}(w)\,dw \\
&= s^{-N/2},
\end{aligned} \tag{48}$$

we infer choosing $\tau := 2/(N+2)$

$$\begin{aligned}
|\nabla^3 f_\varepsilon(x)| &\leq \frac{3}{2}(1+\tau)^{(N+2)/2}\tau^{-1}|\Lambda^{-1}|\int_{\varepsilon^2}^1 s^{-(N+2)/2}\,ds \\
&\leq \frac{3}{2}\cdot e\cdot\frac{N+2}{2}\cdot|\Lambda^{-1}|\cdot\frac{2}{N}(\varepsilon^{-N}-1).
\end{aligned}$$

This proves (20).

Proof of c). We now turn to the proof of Proposition 7c. The bound (21) is immediate from Lemma 10c.

Computing the second spatial derivative of f_ε as defined by (46), we infer

$$\nabla^2 f_\varepsilon(x) = \frac{1}{2}\int_{\varepsilon^2}^1 \frac{1}{s}\int_{\mathbb{R}^N}\phi(\sqrt{1-s}x-\sqrt{s}z)\nabla^2\mathcal{N}_\Lambda(z)\,dz\,ds. \tag{49}$$

In order to derive the estimates (22) and (23), we estimate for $r > 0$ and $\tau := 2/(N+2)$

$$\begin{aligned}
&\int_{\mathbb{R}^N} (\mathrm{osc}_r \nabla^2 f_\varepsilon)(x) \mathcal{N}_\Lambda(x-x_0)\,dx \\
&\leq \frac{1}{2} \int_{\mathbb{R}^N} \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} (\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x - \sqrt{s}z) |\nabla^2 \mathcal{N}_\Lambda|(z)\,dz\,ds\,\mathcal{N}_\Lambda(x-x_0)\,dx \\
&\overset{(62)}{\leq} \frac{1}{2} \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x - \sqrt{s}z) \\
&\qquad\qquad \times 3(1+\tau)^{(N+2)/2}\tau^{-1}|\Lambda^{-1}| \mathcal{N}_{(1+\tau)\Lambda}(z)\mathcal{N}_\Lambda(x-x_0)\,dz\,dx\,ds \\
&\leq \frac{3e(N+2)}{4}|\Lambda^{-1}| \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x - \sqrt{s}z) \\
&\qquad\qquad \times (\mathcal{N}_{\tau\Lambda} * \mathcal{N}_\Lambda)(z) \mathcal{N}_\Lambda(x-x_0)\,dz\,dx\,ds \\
&= \frac{3e(N+2)}{4}|\Lambda^{-1}| \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\mathcal{N}_{s\tau\Lambda} * \mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}(x+x_0) - \sqrt{s}z) \\
&\qquad\qquad \times \mathcal{N}_\Lambda(z)\mathcal{N}_\Lambda(x)\,dz\,dx\,ds.
\end{aligned}$$

Invoking the change of variables $(\tilde{x}, \tilde{z}) := (\sqrt{1-s}x - \sqrt{s}z, \sqrt{1-s}z + \sqrt{s}x)$ (note that this is a linear orthogonal transformation) as well as the multiplication property (64), we deduce

$$\begin{aligned}
&\int_{\mathbb{R}^N} (\mathrm{osc}_r \nabla^2 f_\varepsilon)(x) \mathcal{N}_\Lambda(x-x_0)\,dx \\
&\leq \frac{3e(N+2)}{4}|\Lambda^{-1}| \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\mathcal{N}_{s\tau\Lambda} * \mathrm{osc}_{\sqrt{1-s}r}\phi)(\tilde{x} + \sqrt{1-s}x_0) \\
&\qquad\qquad \times \mathcal{N}_\Lambda(\tilde{x})\mathcal{N}_\Lambda(\tilde{z})\,d\tilde{x}\,d\tilde{z}\,ds \\
&= \frac{3e(N+2)}{4}|\Lambda^{-1}| \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} (\mathrm{osc}_{\sqrt{1-s}r}\phi)(\tilde{x} + \sqrt{1-s}x_0 - w) \\
&\qquad\qquad \times \mathcal{N}_\Lambda(\tilde{x})\mathcal{N}_{s\tau\Lambda}(w)\mathcal{N}_\Lambda(\tilde{z})\,dw\,d\tilde{x}\,d\tilde{z}\,ds \\
&\overset{(4)}{\leq} \frac{9(N+2)}{4}|\Lambda^{-1}| \int_{\varepsilon^2}^1 \frac{1}{s} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \sqrt{1-s}r \mathcal{N}_{s\tau\Lambda}(w)\mathcal{N}_\Lambda(\tilde{z})\,dw\,d\tilde{z}\,ds
\end{aligned}$$

and hence

$$\int_{\mathbb{R}^N} (\mathrm{osc}_r \nabla^2 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \leq 16N|\Lambda^{-1}||\log\varepsilon| r \tag{50}$$

for any $r > 0$ and any $x_0 \in \mathbb{R}^N$. As a consequence, we get

$$\begin{aligned}
&\int_{\mathbb{R}^N} 2(\mathcal{N}_{\delta^2 \mathrm{Id}_N} * \mathrm{osc}_{K\delta}\nabla^2 f_\varepsilon)(x)\mathcal{N}_\Lambda(x)\,dx \\
&= 2\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} (\mathrm{osc}_{K\delta}\nabla^2 f_\varepsilon)(x-w)\mathcal{N}_{\delta^2 \mathrm{Id}_N}(w)\mathcal{N}_\Lambda(x)\,dx\,dw \\
&\leq 2\int_{\mathbb{R}^N} 16N|\Lambda^{-1}||\log\varepsilon|K\delta\mathcal{N}_{\delta^2 \mathrm{Id}_N}(w)\,dx,
\end{aligned}$$

which (by our choice of K) immediately proves (22).

Furthermore, (50) entails that

$$h(x) := \frac{1}{32N|\Lambda^{-1}||\log\varepsilon|K}(\mathrm{osc}_{K\delta}\nabla^2 f_\varepsilon)(x)$$

satisfies the estimates

$$\int_{\mathbb{R}^N} |h(x)|\mathcal{N}_\Lambda(x-x_0)\,dx \leq \delta$$

for any $x_0 \in \mathbb{R}^N$ and (by the inequality $(\mathrm{osc}_r \mathrm{osc}_{K\delta} f)(x) \leq \mathrm{osc}_{r+K\delta} f(x)$)

$$\begin{aligned}
&\int_{\mathbb{R}^N} \mathrm{osc}_r h(x)\mathcal{N}_\Lambda(x-x_0)\,dx\\
&\leq \frac{1}{32N|\Lambda^{-1}||\log\varepsilon|K}\int_{\mathbb{R}^N}(\mathrm{osc}_{r+K\delta}\nabla^2 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx\\
&\leq \frac{1}{2K}(r+K\delta)\\
&\leq r
\end{aligned}$$

for any $x_0 \in \mathbb{R}^N$ and any $r \geq \delta$. In conclusion, Lemma 10d implies

$$\frac{1}{40N} \times \frac{1}{32N|\Lambda^{-1}||\log\varepsilon|K}(\mathcal{N}_{\delta^2\,\mathrm{Id}_N} * \mathrm{osc}_{K\delta}\nabla^2 f_\varepsilon) \in \Phi_\Lambda^{\tilde{L}}$$

for any $\tilde{L} \geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1)$, which proves (23).

Proof of d), Part 1. We now establish Proposition 7d. The estimate (24) is immediate by Lemma 10c and the inequality $\sup_{|y|\leq\delta}|\nabla^3 f_\varepsilon(x+y)| \leq |\nabla^3 f_\varepsilon(x)| + \mathrm{osc}_\delta\nabla^3 f_\varepsilon(x)$.

Before proceeding, let us first prove the following auxiliary result: The third derivative of f_ε satisfies the estimate

$$\int_{\mathbb{R}^N} |\nabla^3 f_\varepsilon|(x)\mathcal{N}_\Lambda(x)\,dx \leq 16N|\Lambda^{-1}||\log\varepsilon|, \tag{51}$$

and has the property

$$\frac{\varepsilon}{10N^{3/2}|\Lambda^{-1/2}|^3}|\nabla^3 f_\varepsilon| \in \Phi_\Lambda^{\tilde{L}} \tag{52}$$

for any $\tilde{L} \geq \varepsilon^{-N}$.

The estimate (51) is simply a consequence of (50) in the limit $r \to 0$. To show (52) we first establish a uniform bound for $\nabla^4 f_\varepsilon$. This is done in an analogous way to the proof of assertion b) of our proposition: Using (63) instead of (62), we deduce

$$\begin{aligned}
&|\nabla^4 f_\varepsilon(x)|\\
&\leq \frac{1}{2}\int_{\varepsilon^2}^1 \frac{1-s}{s^{3/2}}\int_{\mathbb{R}^N}|\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)|\nabla^3\mathcal{N}_\Lambda|(z)\,dz\,ds\\
&\overset{(63)}{\leq} \frac{5}{2}(1+\tau)^{(N+3)/2}\tau^{-3/2}|\Lambda^{-1/2}|^3\\
&\qquad\times\int_{\varepsilon^2}^1 \frac{1}{s^{3/2}}\int_{\mathbb{R}^N}|\nabla\phi|(\sqrt{1-s}x-\sqrt{s}z)\mathcal{N}_{(1+\tau)\Lambda}(z)\,dz\,ds
\end{aligned}$$

and choosing $\tau := 2/(N+3)$ and applying (48), we get

$$|\nabla^4 f_\varepsilon(x)| \leq \frac{5}{2}\cdot e\cdot \frac{(N+3)^{3/2}}{2^{3/2}}|\Lambda^{-1/2}|^3 \int_{\varepsilon^2}^1 \frac{1}{s^{3/2}}\cdot s^{-N/2}\,ds$$
$$\leq \frac{5}{2}\cdot e\cdot \frac{(N+3)^{3/2}}{2^{3/2}}|\Lambda^{-1/2}|^3 \cdot \frac{2}{N+1}\cdot \frac{1}{(\varepsilon^2)^{(N+1)/2}}$$
$$\leq 10\sqrt{N+3}|\Lambda^{-1/2}|^3\varepsilon^{-N-1}. \tag{53}$$

In order to show (52), for any $r>0$ we estimate starting with (47) and choosing $\tau := 2/(N+3)$

$$\int_{\mathbb{R}^N} (\mathrm{osc}_r \nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx$$
$$\leq \frac{1}{2}\int_{\varepsilon^2}^1 \frac{\sqrt{1-s}}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x-\sqrt{s}z)|\nabla^3\mathcal{N}_\Lambda|(z)\mathcal{N}_\Lambda(x-x_0)\,dz\,dx\,ds$$
$$\overset{(63)}{\leq} \frac{1}{2}\int_{\varepsilon^2}^1 \frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x-\sqrt{s}z)$$
$$\times 5(1+\tau)^{(N+3)/2}\tau^{-3/2}|\Lambda^{-1/2}|^3\mathcal{N}_{(1+\tau)\Lambda}(z)\mathcal{N}_\Lambda(x-x_0)\,dz\,dx\,ds$$
$$\leq \frac{5e}{2}\cdot\frac{(N+3)^{3/2}}{2^{3/2}}|\Lambda^{-1/2}|^3\int_{\varepsilon^2}^1\frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}x-\sqrt{s}z)$$
$$\times(\mathcal{N}_{\tau\Lambda}*\mathcal{N}_\Lambda)(z)\mathcal{N}_\Lambda(x-x_0)\,dz\,dx\,ds$$
$$= \frac{5e(N+3)^{3/2}}{2\cdot\sqrt{2}^3}|\Lambda^{-1/2}|^3\int_{\varepsilon^2}^1\frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathcal{N}_{s\tau\Lambda}*\mathrm{osc}_{\sqrt{1-s}r}\phi)(\sqrt{1-s}(x+x_0)-\sqrt{s}z)$$
$$\times\mathcal{N}_\Lambda(z)\mathcal{N}_\Lambda(x)\,dz\,dx\,ds.$$

Invoking the change of variables $(\tilde{x},\tilde{z}) := (\sqrt{1-s}x-\sqrt{s}z, \sqrt{1-s}z+\sqrt{s}x)$ (note that this is a linear orthogonal transformation) as well as the multiplication property (64), we deduce

$$\int_{\mathbb{R}^N} (\mathrm{osc}_r \nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx$$
$$\leq \frac{5(N+3)^{3/2}}{2}|\Lambda^{-1/2}|^3\int_{\varepsilon^2}^1\frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathcal{N}_{s\tau\Lambda}*\mathrm{osc}_{\sqrt{1-s}r}\phi)(\tilde{x}+\sqrt{1-s}x_0)$$
$$\times\mathcal{N}_\Lambda(\tilde{x})\mathcal{N}_\Lambda(\tilde{z})\,d\tilde{x}\,d\tilde{z}\,ds$$
$$= \frac{5(N+3)^{3/2}}{2}|\Lambda^{-1/2}|^3\int_{\varepsilon^2}^1\frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}(\mathrm{osc}_{\sqrt{1-s}r}\phi)(\tilde{x}+\sqrt{1-s}x_0-w)$$
$$\times\mathcal{N}_\Lambda(\tilde{x})\mathcal{N}_{s\tau\Lambda}(w)\mathcal{N}_\Lambda(\tilde{z})\,dw\,d\tilde{x}\,d\tilde{z}\,ds$$
$$\overset{(4)}{\leq} \frac{5(N+3)^{3/2}}{2}|\Lambda^{-1/2}|^3\int_{\varepsilon^2}^1\frac{1}{s^{3/2}}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\sqrt{1-s}r\mathcal{N}_{s\tau\Lambda}(w)\mathcal{N}_\Lambda(\tilde{z})\,dw\,d\tilde{z}\,ds$$

and hence

$$\int_{\mathbb{R}^N}(\mathrm{osc}_r\nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \leq 10N^{3/2}|\Lambda^{-1/2}|^3\frac{r}{\varepsilon} \tag{54}$$

for any $r>0$ and any $x_0\in\mathbb{R}^N$. Combining this estimate with (53), we infer (52).

Proof of d), Part 2. We now turn to the last part of the proof of our proposition. As a consequence of (54), we get

$$\begin{aligned}
&\int_{\mathbb{R}^N} 2(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * \operatorname{osc}_{K\delta}\nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x)\,dx\\
&= 2\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} (\operatorname{osc}_{K\delta}\nabla^3 f_\varepsilon)(x-w)\mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)\mathcal{N}_\Lambda(x)\,dx\,dw\\
&\leq 2\int_{\mathbb{R}^N} 10N^{3/2}|\Lambda^{-1/2}|^3\frac{K\delta}{\varepsilon}\mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)\,dx,
\end{aligned}$$

which gives

$$\int_{\mathbb{R}^N} 2(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * \operatorname{osc}_{K\delta}\nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x)\,dx \leq 20N^{3/2}K|\Lambda^{-1/2}|^3\,\frac{\delta}{\varepsilon}.$$

Using $K = 2\sqrt{N}+1$ and combining this estimate with (51), we infer (25).

Furthermore, we know by (54) that

$$h(x) := \frac{\varepsilon}{20N^{3/2}|\Lambda^{-1/2}|^3 K}\operatorname{osc}_{K\delta}\nabla^3 f_\varepsilon(x)$$

satisfies the estimate

$$\int_{\mathbb{R}^N} h(x)\mathcal{N}_\Lambda(x-x_0)\,dx \leq \delta$$

for any $x_0 \in \mathbb{R}^N$; by (54), it also satisfies the bound

$$\begin{aligned}
&\int_{\mathbb{R}^N} \operatorname{osc}_r h(x)\mathcal{N}_\Lambda(x-x_0)\,dx\\
&\leq \frac{\varepsilon}{20N^{3/2}|\Lambda^{-1/2}|^3 K}\int_{\mathbb{R}^N} (\operatorname{osc}_{r+K\delta}\nabla^3 f_\varepsilon)(x)\mathcal{N}_\Lambda(x-x_0)\,dx\\
&\leq \frac{1}{2K}(r+K\delta)\\
&\leq r
\end{aligned}$$

for any $x_0 \in \mathbb{R}^N$ and any $r \geq \delta$. In conclusion, Lemma 10d implies

$$\frac{1}{40N}\times\frac{\varepsilon}{20N^{3/2}|\Lambda^{-1/2}|^3 K}(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * \operatorname{osc}_{K\delta}\nabla^3 f_\varepsilon) \in \Phi_\Lambda^{\tilde{L}}$$

for any $\tilde{L} \geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1)$. In conjunction with (52) and $K = 2\sqrt{N}+1$, we infer (26). □

Proof of Lemma 8. Without loss of generality (and in order to simplify notation) we may assume

$$|\Lambda| \leq 1,$$

that is the maximal eigenvalue of Λ is bounded by 1; the general case then follows upon rescaling $\hat{X} := \frac{1}{|\Lambda^{1/2}|}X$, $\hat{\phi}(x) := \frac{1}{|\Lambda^{1/2}|}\phi(|\Lambda^{1/2}|x)$, $\hat{\Lambda} := \Lambda/|\Lambda|$ (note that then $\hat{\phi}_\varepsilon(x) = \frac{1}{|\Lambda^{1/2}|}\phi_\varepsilon(|\Lambda^{1/2}|x)$).

In order to establish (27), by the very definition (5) of $\mathcal{D}^{\bar{L}}$ and upon replacing ε by $\frac{1}{2}\varepsilon$ it suffices to prove

$$\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N}\phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \leq 10\sqrt{N}\varepsilon + 10^3 N^{3/2}\mathcal{D}^{\bar{L}}_{\varepsilon/2}(X,\mathcal{N}_\Lambda) \tag{55}$$

for all $\phi \in \Phi_\Lambda^{\bar{L}}$ and all $0 < \varepsilon \leq 1$.

Introducing the function

$$\tilde{\phi}_\varepsilon(x) := \phi_\varepsilon(x/\sqrt{1-\varepsilon^2}),$$

the error stemming from the smoothing of ϕ may be estimated as

$$\begin{aligned}|\tilde{\phi}_\varepsilon(x) - \phi(x)| &= |\phi_\varepsilon(x/\sqrt{1-\varepsilon^2}) - \phi(x)| \\ &\leq \left| \int_{\mathbb{R}^N} \big(\phi(x-\varepsilon z) - \phi(x)\big)\mathcal{N}_\Lambda(z)\,dz \right| \\ &\leq \int_{\mathbb{R}^N} \operatorname{osc}_{\varepsilon|z|}\phi(x)\mathcal{N}_\Lambda(z)\,dz.\end{aligned}$$

As a consequence, we deduce

$$\begin{aligned}&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\ &\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| + \mathbb{E}\big[\big|\tilde{\phi}_\varepsilon - \phi\big|(X)\big] \\ &\quad + \left|\int_{\mathbb{R}^N} (\tilde{\phi}_\varepsilon - \phi)(x)\mathcal{N}_\Lambda(x)\,dx\right| \\ &\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| + \int_{\mathbb{R}^N} \mathbb{E}\big[\operatorname{osc}_{\varepsilon|z|}\phi(X)\big]\mathcal{N}_\Lambda(z)\,dz \\ &\quad + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \operatorname{osc}_{\varepsilon|z|}\phi(x)\mathcal{N}_\Lambda(z)\,dz\;\mathcal{N}_\Lambda(x)\,dx.\end{aligned}$$

For $K := 2\sqrt{N}$, by Lemma 10c we infer

$$\begin{aligned}&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\ &\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| \\ &\quad + 2\int_{\mathbb{R}^N} \mathbb{E}\big[\big(\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N} * \operatorname{osc}_{(|z|+K)\varepsilon}\phi\big)(X)\big]\mathcal{N}_\Lambda(z)\,dz \\ &\quad + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \operatorname{osc}_{\varepsilon|z|}\phi(x)\mathcal{N}_\Lambda(z)\,dz\;\mathcal{N}_\Lambda(x)\,dx\end{aligned}$$

Introducing the definition

$$\psi^z(x) := \big(\mathcal{N}_{\varepsilon^2(2\,\mathrm{Id}_N-\Lambda)} * \operatorname{osc}_{(|z|+K)\varepsilon}\phi\big)(x), \tag{56}$$

we see that we may rewrite (recall that $\mathcal{N}_{\varepsilon^2\Lambda} * \mathcal{N}_{\varepsilon^2(2\,\mathrm{Id}_N-\Lambda)} = \mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N}$)

$$\begin{aligned}(\tilde{\psi^z})_\varepsilon(x) :&= (\psi^z)_\varepsilon(x/\sqrt{1-\varepsilon^2}) := \int_{\mathbb{R}^N} \psi^z(x-\varepsilon w)\mathcal{N}_\Lambda(w)\,dw \\ &= \int_{\mathbb{R}^N} \psi^z(x-w)\mathcal{N}_{\varepsilon^2\Lambda}(w)\,dw = (\mathcal{N}_{\varepsilon^2\Lambda} * \psi^z)(x) \\ &= \big(\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N} * \operatorname{osc}_{(|z|+K)\varepsilon}\phi\big)(x).\end{aligned}$$

As a consequence, we deduce the bound

$$\begin{aligned}
&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\quad + 2\int_{\mathbb{R}^N} \left(\mathbb{E}\big[(\tilde{\psi}^z)_\varepsilon(X)\big] - \int_{\mathbb{R}^N} (\tilde{\psi}^z)_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + 2\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \big(\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N} * \mathrm{osc}_{(|z|+K)\varepsilon}\phi\big)(x)\mathcal{N}_\Lambda(x)\,dx\;\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \mathrm{osc}_{\varepsilon|z|}\phi(x)\mathcal{N}_\Lambda(z)\,dz\;\mathcal{N}_\Lambda(x)\,dx
\end{aligned}$$

and therefore

$$\begin{aligned}
&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\quad + 2\int_{\mathbb{R}^N} \left(\mathbb{E}\big[(\tilde{\psi}^z)_\varepsilon(X)\big] - \int_{\mathbb{R}^N} (\tilde{\psi}^z)_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + 2\int_{\mathbb{R}^N}\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \mathrm{osc}_{(|z|+K)\varepsilon}\phi(x-w)\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N}(w)\mathcal{N}_\Lambda(x)\,dx\,dw\;\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \mathrm{osc}_{\varepsilon|z|}\phi(x)\mathcal{N}_\Lambda(z)\,dz\;\mathcal{N}_\Lambda(x)\,dx \\
&\overset{(4)}{\leq} \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\quad + 2\int_{\mathbb{R}^N} \left(\mathbb{E}\big[(\tilde{\psi}^z)_\varepsilon(X)\big] - \int_{\mathbb{R}^N} (\tilde{\psi}^z)_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + 2\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} (|z|+K)\varepsilon\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N}(w)\,dw\;\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \int_{\mathbb{R}^N} \varepsilon|z|\mathcal{N}_\Lambda(z)\,dz \\
&\leq \left|\mathbb{E}[\tilde{\phi}_\varepsilon(X)] - \int_{\mathbb{R}^N} \tilde{\phi}_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\quad + 2\int_{\mathbb{R}^N} \left(\mathbb{E}\big[(\tilde{\psi}^z)_\varepsilon(X)\big] - \int_{\mathbb{R}^N} (\tilde{\psi}^z)_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \big(2\sqrt{N} + 2K\big)\varepsilon \\
&\quad + \sqrt{N}\varepsilon,
\end{aligned} \tag{57}$$

where in the last step we have used Hölder's inequality and the estimate $\int |z|^2\mathcal{N}_\Lambda(z)\,dz \leq \operatorname{tr}\Lambda \leq N$.

Given $\psi \in \Phi_\Lambda^{\bar{L}}$, introducing the notation

$$\hat{\psi}(x) := \psi(x/\sqrt{1-\varepsilon^2/4})$$

and the notation

$$\tilde{\psi}_\varepsilon(x) := \int_{\mathbb{R}^N} \psi(x-\varepsilon z)\mathcal{N}_\Lambda(z)\,dz$$

we may write

$$\begin{aligned}
\tilde{\psi}_\varepsilon(x) &= \int_{\mathbb{R}^N} \psi(x-\varepsilon z)\mathcal{N}_\Lambda(z)\,dz = \int_{\mathbb{R}^N} \hat{\psi}(\sqrt{1-\varepsilon^2/4}\,x - \sqrt{1-\varepsilon^2/4}\,\varepsilon z)\mathcal{N}_\Lambda(z)\,dz \\
&= \int_{\mathbb{R}^N} \hat{\psi}(\sqrt{1-\varepsilon^2/4}\,x - z)\mathcal{N}_{(1-\varepsilon^2/4)\varepsilon^2\Lambda}(z)\,dz \\
&= \int_{\mathbb{R}^N} (\mathcal{N}_{(3/4-\varepsilon^2/4)\varepsilon^2\Lambda} * \hat{\psi})(\sqrt{1-\varepsilon^2/4}\,x - z)\mathcal{N}_{\frac{1}{4}\varepsilon^2\Lambda}(z)\,dz \\
&= \theta_{\varepsilon/2}(x) \qquad (58)
\end{aligned}$$

for $\theta_{\varepsilon/2}(x) := \int_{\mathbb{R}^N} \theta(\sqrt{1-\varepsilon^2/4}\,x - z)\mathcal{N}_{\frac{1}{4}\varepsilon^2\Lambda}(z)\,dz$ with

$$\theta := \mathcal{N}_{(3/4-\varepsilon^2/4)\varepsilon^2\Lambda} * \hat{\psi}.$$

Note that by Lemma 10 b) and a), from $\psi \in \Phi_\Lambda^{\bar{L}}$ it follows that $\frac{1}{2}\hat{\psi} \in \Phi_\Lambda^{\bar{L}}$ and therefore also $\frac{1}{2}\theta \in \Phi_\Lambda^{\bar{L}}$. Applying these considerations for $\psi := \phi$, we deduce from estimate (57)

$$\begin{aligned}
&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \qquad (59)\\
&\leq 2\mathcal{D}^{\bar{L}}_{\varepsilon/2}(X,\mathcal{N}_\Lambda) \\
&\quad + 2\int_{\mathbb{R}^N} \left(\mathbb{E}\big[(\tilde{\psi^z})_\varepsilon(X)\big] - \int_{\mathbb{R}^N} (\tilde{\psi^z})_\varepsilon(x)\mathcal{N}_\Lambda(x)\,dx\right)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \big(2K + 3\sqrt{N}\big)\varepsilon.
\end{aligned}$$

Finally, notice that Lemma 10d entails (with $h(x) := \frac{1}{|z|+K+1}\mathrm{osc}_{(|z|+K)\varepsilon}\phi$ and $\delta := \varepsilon$; note that $\mathrm{osc}_r h(x) \leq \frac{1}{|z|+K+1}\mathrm{osc}_{(|z|+K)\varepsilon+r}\phi(x)$)

$$\frac{1}{40N(|z|+K+1)}(\mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N} * \mathrm{osc}_{(|z|+K)\varepsilon}\phi) \in \Phi_\Lambda^{\tilde{L}}$$

for any $\tilde{L} \geq 4^N(|\Lambda^{1/2}|^N\varepsilon^{-N}+1)$, i.e. in particular for $\tilde{L} = \bar{L}$. Using $\psi^z = \mathcal{N}_{\varepsilon^2(\mathrm{Id}_N-\Lambda)} * (\mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N} * \mathrm{osc}_{(|z|+K)\varepsilon}\phi)$ (by definition (56) and (61)) and Lemma 10a this yields

$$\frac{1}{40N(|z|+K+1)}\psi^z \in \Phi_\Lambda^{\bar{L}}.$$

Consequently, the function $\hat{\psi}^z(x) := \psi^z(x/\sqrt{1-\varepsilon^2/4})$ satisfies by Lemma 10b

$$\frac{1}{2\times 40N(|z|+K+1)}\hat{\psi}^z \in \Phi_\Lambda^{\bar{L}}.$$

Applying the considerations around (58) to $\psi := \psi^z/(2\times 40N(|z|+K+1))$, we deduce from (59)

$$\begin{aligned}
&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\leq 2\mathcal{D}^{\bar L}_{\varepsilon/2}(X,\mathcal{N}_\Lambda) \\
&\quad + 2\int_{\mathbb{R}^N} 2\times 40N(|z|+K+1)\mathcal{D}^{\bar L}_{\varepsilon/2}(X,\mathcal{N}_\Lambda)\mathcal{N}_\Lambda(z)\,dz \\
&\quad + \big(2K+3\sqrt{N}\big)\varepsilon
\end{aligned}$$

which enables us to conclude

$$\begin{aligned}
&\left|\mathbb{E}[\phi(X)] - \int_{\mathbb{R}^N} \phi(x)\mathcal{N}_\Lambda(x)\,dx\right| \\
&\leq 2\mathcal{D}^{\bar L}_{\varepsilon/2}(X,\mathcal{N}_\Lambda) \\
&\quad + 2\mathcal{D}^{\bar L}_{\varepsilon/2}(X,\mathcal{N}_\Lambda)\times 2\times 40N\times(\sqrt{N}+K+1) \\
&\quad + \big(2K+3\sqrt{N}\big)\varepsilon.
\end{aligned}$$

Recalling that we have assumed $|\Lambda|\leq 1$ and that $K=2\sqrt{N}$, we deduce (55). □

Lemma 10. *The function classes $\Phi^{\bar L}_\Lambda$ are subject to the following properties:*

a) *Given any $b\in L^1(\mathbb{R}^N)$ with $\int_{\mathbb{R}^N}|b(x)|\,dx\leq 1$ and any $\psi\in\Phi^{\bar L}_\Lambda$, the convolution $b*\psi$ satisfies*
$$b*\psi\in\Phi^{\bar L}_\Lambda.$$
b) *Given any $\psi\in\Phi^{\bar L}_\Lambda$ and any $\tau\geq 1$, the rescaled function*
$$\psi_\tau(x) := \frac{1}{\tau}\psi(\tau x)$$
satisfies $\psi_\tau\in\Phi^{\bar L}_\Lambda$.

c) *For $K\geq 2\sqrt{N}$, $\delta>0$, $\varepsilon>0$, $r>0$, and any function $\psi\in L^\infty_{loc}(\mathbb{R}^N;V)$ for any normed vector space V, the estimates*
$$\mathrm{osc}_\delta\psi(x)\leq 2(\mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N}*\mathrm{osc}_{K\varepsilon+\delta}\psi)(x)$$
and
$$\mathrm{osc}_\delta\psi(x)\leq 2(\mathcal{N}_{2\varepsilon^2\,\mathrm{Id}_N}*\mathrm{osc}_{K\varepsilon+\delta}\psi)(x)$$
hold.

d) *Let $\delta>0$ and let $h\in L^\infty_{loc}(\mathbb{R}^N)$ be any function subject to the properties*
$$\int_{\mathbb{R}^N}|h|(x)\ \mathcal{N}_\Lambda(x-x_0)\,dx\leq\delta$$
for any $x_0\in\mathbb{R}^N$ and
$$\int_{\mathbb{R}^N}\mathrm{osc}_r h(x)\ \mathcal{N}_\Lambda(x-x_0)\,dx\leq r$$
for any $r\geq\delta$ and any $x_0\in\mathbb{R}^N$. We then have
$$\frac{1}{40N}\mathcal{N}_{\delta^2\,\mathrm{Id}_N}*h\in\Phi^{\tilde L}_\Lambda$$

for any $\tilde{L} \geq 4^N(|\Lambda^{1/2}|^N\delta^{-N}+1)$.

Proof. To establish d), we choose $\tau := 2/N$ and compute for $r \geq \delta$

$$\begin{aligned}
&\int_{\mathbb{R}^N} \operatorname{osc}_r(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * h)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \\
&= \int_{\mathbb{R}^N} \operatorname{osc}_r\bigg(\int_{\mathbb{R}^N} \mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)h(\cdot - w)\,dw\bigg)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \\
&\leq \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)\operatorname{osc}_r h(x-w)\,dw\;\mathcal{N}_\Lambda(x-x_0)\,dx \\
&= \int_{\mathbb{R}^N} \mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)\int_{\mathbb{R}^N} \operatorname{osc}_r h(x)\mathcal{N}_\Lambda(x+w-x_0)\,dx\,dw \\
&\leq \int_{\mathbb{R}^N} \mathcal{N}_{\delta^2 \operatorname{Id}_N}(w) r\,dw \\
&\leq r.
\end{aligned}$$

For $r \in [\frac{1}{2N}\delta, \delta]$, upon replacing r by δ we obtain

$$\int_{\mathbb{R}^N} \operatorname{osc}_r(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * h)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \leq \delta \leq 2Nr.$$

In contrast, for $r \leq \frac{1}{2N}\delta$ we estimate

$$\begin{aligned}
&\int_{\mathbb{R}^N} \operatorname{osc}_r(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * h)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \\
&= \int_{\mathbb{R}^N} \operatorname{osc}_r\bigg(\int_{\mathbb{R}^N} \mathcal{N}_{\delta^2 \operatorname{Id}_N}(\cdot - w)h(w)\,dw\bigg)(x)\mathcal{N}_\Lambda(x-x_0)\,dx \\
&\leq \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \operatorname{osc}_r\mathcal{N}_{\delta^2 \operatorname{Id}_N}(x-w)|h(w)|\,dw\;\mathcal{N}_\Lambda(x-x_0)\,dx \\
&\overset{(66)}{\leq} \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \frac{r}{\delta}\,40\sqrt{N}\mathcal{N}_{(1+\tau)\delta^2 \operatorname{Id}_N}(x-w)|h(w)|\,dw\;\mathcal{N}_\Lambda(x-x_0)\,dx \\
&= \frac{r}{\delta}\,40\sqrt{N}\int_{\mathbb{R}^N} \mathcal{N}_{(1+\tau)\delta^2 \operatorname{Id}_N}(w)\int_{\mathbb{R}^N} |h(x-w)|\;\mathcal{N}_\Lambda(x-x_0)\,dx\,dw \\
&\leq \frac{r}{\delta}\,40\sqrt{N}\int_{\mathbb{R}^N} \mathcal{N}_{(1+\tau)\delta^2 \operatorname{Id}_N}(w)\delta\,dw \\
&\leq 40\sqrt{N}\,r.
\end{aligned}$$

To complete the proof of d), it only remains to show the estimate on the Lipschitz constant of $\mathcal{N}_{\delta^2 \operatorname{Id}_N} * h$. To do so, we estimate

$$\begin{aligned}
&|\nabla(\mathcal{N}_{\delta^2 \operatorname{Id}_N} * h)(x)| \\
&\leq \int_{\mathbb{R}^N} |\nabla\mathcal{N}_{\delta^2 \operatorname{Id}_N}(x-w)||h(w)|\,dw \\
&= \int_{\mathbb{R}^N} |\nabla\mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)||h(x-w)|\,dw \\
&\leq \int_{\mathbb{R}^N} |\delta^{-2}w|\,\mathcal{N}_{\delta^2 \operatorname{Id}_N}(w)\,|h(x-w)|\,dw \\
&\leq \sqrt{2}^N\int_{\mathbb{R}^N} |\delta^{-2}w|\exp\bigg(-\frac{1}{4}\delta^{-2}|w|^2\bigg)\mathcal{N}_{2\delta^2 \operatorname{Id}_N}(w)\,|h(x-w)|\,dw
\end{aligned}$$

$$
\begin{aligned}
&\leq \sqrt{2}^N \delta^{-1} \int_{\mathbb{R}^N} \mathcal{N}_{2\delta^2 \operatorname{Id}_N}(w)\, |h(x-w)|\, dw \\
&\leq \sqrt{2}^N \delta^{-1} \bigg(\frac{\sqrt{2}\delta + |\Lambda^{1/2}|}{\sqrt{2}\delta}\bigg)^N \int_{\mathbb{R}^N} \mathcal{N}_{2\delta^2 \operatorname{Id}_N + \Lambda}(w)\, |h(x-w)|\, dw \\
&\leq \sqrt{2}^N \delta^{-1} \bigg(\frac{\sqrt{2}\delta + |\Lambda^{1/2}|}{\sqrt{2}\delta}\bigg)^N \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \mathcal{N}_{2\delta \operatorname{Id}_N}(z) \mathcal{N}_\Lambda(w-z)\, |h(x-w)|\, dw\, dz \\
&\leq \sqrt{2}^N \delta^{-1} \bigg(\frac{\sqrt{2}\delta + |\Lambda^{1/2}|}{\sqrt{2}\delta}\bigg)^N \int_{\mathbb{R}^N} \delta \mathcal{N}_{2\delta \operatorname{Id}_N}(z)\, dz \\
&\leq \sqrt{2}^N \bigg(\frac{\sqrt{2}\delta + |\Lambda^{1/2}|}{\sqrt{2}\delta}\bigg)^N .
\end{aligned}
$$

This establishes assertion d).

Assertion a) is a direct consequence of the estimates

$$
|\nabla(b * \psi)(x)| = \bigg| \int_{\mathbb{R}^N} b(w) \nabla\psi(x-w)\, dw \bigg| \leq \int_{\mathbb{R}^N} |b(w)| \bar{L}\, dw \leq \bar{L}
$$

and

$$
\begin{aligned}
&\int_{\mathbb{R}^N} \operatorname{osc}_r (b * \psi)(x) \mathcal{N}_\Lambda(x-x_0)\, dx \\
&= \int_{\mathbb{R}^N} \operatorname{osc}_r \bigg(\int_{\mathbb{R}^N} b(w) \psi(\cdot - w)\, dw \bigg)(x) \mathcal{N}_\Lambda(x-x_0)\, dx \\
&\leq \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} |b(w)| \operatorname{osc}_r \psi(x-w)\, dw\; \mathcal{N}_\Lambda(x-x_0)\, dx \\
&= \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \operatorname{osc}_r \psi(x)\, \mathcal{N}_\Lambda(x+w-x_0)\, dx\; |b(w)|\, dw \\
&\leq \int_{\mathbb{R}^N} r |b(w)|\, dw \\
&\leq r.
\end{aligned}
$$

Concerning b), we directly verify $|\nabla\psi_\tau(x)| = |\nabla\psi(\tau x)| \leq \bar{L}$. On the other hand, we have for any $r > 0$ using the convolution property $\mathcal{N}_{\tau^2\Lambda} = \mathcal{N}_\Lambda * \mathcal{N}_{(\tau^2-1)\Lambda}$

$$
\begin{aligned}
&\int_{\mathbb{R}^N} \operatorname{osc}_r \psi_\tau(x) \mathcal{N}_\Lambda(x-x_0)\, dx \\
&= \int_{\mathbb{R}^N} \frac{1}{\tau} \operatorname{osc}_{\tau r} \psi(\tau x) \mathcal{N}_\Lambda(x-x_0)\, dx \\
&= \int_{\mathbb{R}^N} \frac{1}{\tau} \operatorname{osc}_{\tau r} \psi(\tilde{x}) \mathcal{N}_{\tau^2\Lambda}(\tilde{x} - \tau x_0)\, d\tilde{x} \\
&= \frac{1}{\tau} \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} \operatorname{osc}_{\tau r} \psi(\tilde{x}) \mathcal{N}_{(\tau^2-1)\Lambda}(w) \mathcal{N}_\Lambda(\tilde{x} - w - \tau x_0)\, d\tilde{x}\, dw \\
&\leq \frac{1}{\tau} \int_{\mathbb{R}^N} \tau r\, \mathcal{N}_{(\tau^2-1)\Lambda}(w)\, dw \\
&\leq r.
\end{aligned}
$$

This establishes b).

Regarding c), we have by $\mathrm{osc}_\delta\psi(x) \le \mathrm{osc}_{K\varepsilon+\delta}\psi(x-z)$ for any $|z| \le K\varepsilon$

$$\begin{aligned}
&\mathrm{osc}_\delta\psi(x)\\
&\le \frac{1}{\int_{\{|z|\le K\varepsilon\}} \mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N}(z)\,dz} \int_{\mathbb{R}^N} \mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N}(z)\mathrm{osc}_{K\varepsilon+\delta}\psi(x-z)\,dz\\
&= \frac{1}{\int_{\{|z|\le K\}} \mathcal{N}_{\mathrm{Id}_N}(z)\,dz}(\mathcal{N}_{\varepsilon^2\,\mathrm{Id}_N} * \mathrm{osc}_{K\varepsilon+\delta}\psi)(x)
\end{aligned}$$

which for $K \ge \sqrt{2N}$ entails the desired bound using

$$\int_{\{|z|\le\sqrt{2N}\}} \mathcal{N}_{\mathrm{Id}_N}(z)\,dz \ge \frac{1}{2}. \tag{60}$$

The latter estimate follows from

$$\int_{\{|z|\ge r\}} \mathcal{N}_{\mathrm{Id}_N}(z)\,dz \le \frac{1}{r^2}\int_{\{|z|\ge r\}} |z|^2\mathcal{N}_{\mathrm{Id}_N}(z)\,dz \le \frac{N}{r^2}.$$

□

In the following lemma, we collect a number of standard properties of Gaussians.

Lemma 11. *Let $\Lambda \in \mathbb{R}^{N\times N}$ be a symmetric positive definite matrix.*

For any symmetric positive definite matrix $\tilde\Lambda \in \mathbb{R}^{N\times N}$, we have

$$\mathcal{N}_\Lambda * \mathcal{N}_{\tilde\Lambda} = \mathcal{N}_{\Lambda+\tilde\Lambda}. \tag{61}$$

For any $0 < \tau \le 1$ the second and the third derivative of the Gaussian $\mathcal{N}_\Lambda$ satisfy the bounds

$$\frac{|\nabla^2\mathcal{N}_\Lambda(z)|}{\mathcal{N}_{(1+\tau)\Lambda}(z)} \le 3(1+\tau)^{(N+2)/2}\tau^{-1}|\Lambda^{-1}| \tag{62}$$

and

$$\frac{|\nabla^3\mathcal{N}_\Lambda(z)|}{\mathcal{N}_{(1+\tau)\Lambda}(z)} \le 5(1+\tau)^{(N+3)/2}\tau^{-3/2}|\Lambda^{-1/2}|^3 \tag{63}$$

for all $z \in \mathbb{R}^N$.

For any $0 \le s \le 1$ we have the multiplication property

$$\mathcal{N}_\Lambda(z)\mathcal{N}_\Lambda(x) = \mathcal{N}_\Lambda(\sqrt{1-s}z + \sqrt{s}x)\mathcal{N}_\Lambda(\sqrt{1-s}x - \sqrt{s}z) \tag{64}$$

for all $x, z \in \mathbb{R}^N$.

Let $0 < \tau \le 1$ and $r \le \frac{1}{4}\tau\delta|\Lambda^{-1}|^{-1/2}$. Then the Gaussian $\mathcal{N}_{\delta^2\Lambda}$ satisfies the estimate

$$\mathrm{osc}_r\mathcal{N}_{\delta^2\Lambda}(z) \le \frac{r}{\delta} \times \frac{20}{\tau^{1/2}}(1+\tau)^{N/2}|\Lambda^{-1/2}|\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z). \tag{65}$$

In particular, for $\tau := 2/N$ and $r \le \frac{1}{2N}\delta|\Lambda^{-1}|^{-1/2}$ we have

$$\mathrm{osc}_r\mathcal{N}_{\delta^2\Lambda}(z) \le \frac{r}{\delta} \times 40\sqrt{N}|\Lambda^{-1/2}|\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z). \tag{66}$$

Proof. Equation (64) is immediate from the definition

$$\mathcal{N}_\Lambda(z) := \frac{1}{(2\pi)^{N/2}\sqrt{\det\Lambda}}\exp\Big(-\frac{1}{2}\Lambda^{-1}z\cdot z\Big).$$

The estimate (62) follows from

$$\nabla^2\mathcal{N}_\Lambda(z) = \Big(\Lambda^{-1} + \Lambda^{-1}z\otimes\Lambda^{-1}z\Big)\mathcal{N}_\Lambda(z)$$

and

$$\frac{\mathcal{N}_\Lambda(z)}{\mathcal{N}_{(1+\tau)\Lambda}(z)} \leq (1+\tau)^{N/2}\exp\bigg(-\frac{1}{2}\frac{\tau}{1+\tau}\Lambda^{-1}z\cdot z\bigg) \tag{67}$$

as well as the bound for all $b\geq 0$

$$b^2\exp\bigg(-\frac{1}{2}\frac{\tau}{1+\tau}b^2\bigg) \leq \frac{1+\tau}{\tau}\cdot\frac{2}{e}.$$

Similarly, we deduce (63) by applying (67) to

$$\nabla^3\mathcal{N}_\Lambda(z) = \Big(3\big(\Lambda^{-1}\otimes\Lambda^{-1}z\big)_{sym} + \Lambda^{-1}z\otimes\Lambda^{-1}z\otimes\Lambda^{-1}z\Big)\mathcal{N}_\Lambda(z)$$

and using the estimate (valid for all $b\geq 0$)

$$\begin{aligned}(3b+b^3)\exp\bigg(-\frac{1}{2}\frac{\tau}{1+\tau}b^2\bigg) &\leq 3\frac{(1+\tau)^{1/2}}{\tau^{1/2}}\cdot\frac{1}{\sqrt{e}} + \frac{(1+\tau)^{3/2}}{\tau^{3/2}}\cdot\frac{3\cdot\sqrt{3}}{e^{3/2}}\\ &\leq 5\frac{(1+\tau)^{3/2}}{\tau^{3/2}}.\end{aligned}$$

Finally, we turn to the proof of (65), from which (66) follows as an immediate consequence. By the estimate

$$|\exp(-a)-\exp(-(a+b))| \leq b\exp(-a)$$

(valid for any $a,b\geq 0$) we deduce for $\tau\leq 1$ the bound

$$\begin{aligned}&\mathrm{osc}_r\mathcal{N}_{\delta^2\Lambda}(z)\\ &\leq \frac{1}{(2\pi)^{N/2}\sqrt{\det\delta^2\Lambda}}\big(2\delta^{-2}|\Lambda^{-1/2}z||\Lambda^{-1/2}|r + \delta^{-2}|\Lambda^{-1}|r^2\big)\\ &\quad\times\exp\bigg(-\frac{1}{2}\delta^{-2}\Lambda^{-1}z\cdot z + \delta^{-2}|\Lambda^{-1/2}z||\Lambda^{-1/2}|r + \frac{1}{2}\delta^{-2}|\Lambda^{-1}|r^2\bigg)\\ &\leq \delta^{-2}(1+\tau)^{N/2}\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z)\ \sup_x\big(2|\Lambda^{-1/2}x||\Lambda^{-1/2}|r + |\Lambda^{-1}|r^2\big)\\ &\qquad\times\exp\bigg(-\frac{1}{2}\frac{\tau}{1+\tau}\delta^{-2}|\Lambda^{-1/2}x|^2 + \delta^{-2}|\Lambda^{-1/2}x||\Lambda^{-1/2}|r + \frac{1}{2}\delta^{-2}|\Lambda^{-1}|r^2\bigg)\\ &\leq \delta^{-2}(1+\tau)^{N/2}\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z)\ \sup_x\big(2|\Lambda^{-1/2}x||\Lambda^{-1/2}|r + |\Lambda^{-1}|r^2\big)\\ &\qquad\times\exp\bigg(-\frac{\tau}{8}\delta^{-2}|\Lambda^{-1/2}x|^2\bigg)\times\exp\bigg(\frac{1+4\tau^{-1}}{2}\delta^{-2}|\Lambda^{-1}|r^2\bigg)\end{aligned}$$

which for $r\leq\frac{1}{4}\tau\delta|\Lambda^{-1}|^{-1/2}$ entails

$$\begin{aligned}&\mathrm{osc}_r\mathcal{N}_{\delta^2\Lambda}(z)\\ &\leq \delta^{-2}(1+\tau)^{N/2}\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z)\bigg(2|\Lambda^{-1/2}|r\sup_{b\geq 0}b\exp\Big(-\frac{\tau}{8}\delta^{-2}b^2\Big) + \delta|\Lambda^{-1/2}|r\bigg)\times 2\\ &\leq \delta^{-2}(1+\tau)^{N/2}\mathcal{N}_{(1+\tau)\delta^2\Lambda}(z)\big(8\delta\tau^{-1/2}|\Lambda^{-1/2}|r + \delta|\Lambda^{-1/2}|r\big)\times 2.\end{aligned}$$

As a consequence, we deduce (65). □

Appendix B. Concentration inequalities for independent random variables

A concentration estimate for the sum of bounded independent random variables is provided by Bennett's inequality.

Lemma 12 (Bennett's inequality)**.** *Let $X_1, \ldots, X_M$ be independent random variables with $\mathbb{E}[X_i] = 0$ for all $1 \leq i \leq M$. Suppose that for some $A > 0$ the X_i satisfy the uniform bound $|X_i| \leq A$ almost surely. Defining*

$$\sigma^2 := \sum_{i=1}^{M} \mathrm{Var}\ X_i,$$

and using the notation $h(x) := (1+x)\log(1+x) - x$, we have for any $r \geq 0$ the estimate

$$\mathbb{P}\bigg[\sum_{i=1}^{M} X_i \geq r\bigg] \leq \exp\bigg(-\frac{\sigma^2}{A^2} h\Big(\frac{Ar}{\sigma^2}\Big)\bigg).$$

In particular, we have

$$\mathbb{P}\bigg[\sum_{i=1}^{M} X_i \geq r\bigg] \leq \exp\bigg(-\min\Big\{\frac{r^2}{3\sigma^2}, \frac{r}{3A}\Big\}\bigg).$$

Proof. For the proof of Bennett's inequality – the first inequality of the lemma – see [6]. The second inequality follows by distinguishing the cases $r \leq \frac{\sigma^2}{A}$, in which Bennett's inequality gives

$$\mathbb{P}\bigg[\sum_{i=1}^{M} X_i \geq r\bigg] \leq \exp\bigg(-\frac{r^2}{3\sigma^2}\bigg),$$

and $r \geq \frac{\sigma^2}{A}$, in which case we deduce

$$\mathbb{P}\bigg[\sum_{i=1}^{M} X_i \geq r\bigg] \leq \exp\bigg(-\frac{r}{3A}\bigg).$$

□

As a corollary of Bennett's inequality, we deduce the following concentration inequality for random variables with stretched exponential moments.

Lemma 13. *Let $X_1, \ldots, X_M$ be independent random variables with zero mean and uniformly bounded stretched exponential moments*

$$\mathbb{E}\bigg[\exp\bigg(\frac{|X_i|^{\gamma_0}}{b^{\gamma_0}}\bigg)\bigg] \leq 2$$

for some $\gamma_0 > 0$ and some $b > 0$. The sum

$$X := \sum_{i=1}^{M} X_i$$

then satisfies for any $V \geq \mathrm{Var}\ X$ the estimate

$$\mathbb{P}\bigg[\bigg|\sum_{i=1}^{M} X_i\bigg| \geq r\bigg] \leq 3\exp\bigg(-\frac{r^2}{3V}\bigg)$$

for any

$$r \leq \sqrt{V} \min\left\{\frac{\sqrt{V}}{b(2\log(2M))^{1/\gamma_0}}, \left(\frac{\sqrt{V}}{b}\right)^{\gamma_0/(2+\gamma_0)}\right\}. \tag{68}$$

Proof. Fixing $A > 0$, we split each random variable X_i according to

$$X_i = X_i^{bulk} + X_i^{tail}$$

with

$$X_i^{bulk} := X_i \chi_{|X_i| \leq A},$$
$$X_i^{tail} := X_i \chi_{|X_i| > A}.$$

By Lemma 12, we get

$$\mathbb{P}\left[\left|\sum_{i=1}^M X_i^{bulk}\right| \geq r\right] \leq 2\exp\left(-\min\left\{\frac{r^2}{3\mathrm{Var}\ X}, \frac{r}{3A}\right\}\right).$$

Furthermore, we have the bound

$$\mathbb{P}\left[\exists i : X_i^{tail} \neq 0\right] \leq 2M\exp\left(-\frac{A^{\gamma_0}}{b^{\gamma_0}}\right).$$

Choosing $A := \frac{V}{r}$ for some $V \geq \mathrm{Var}\ X$, we deduce

$$\mathbb{P}\left[\left|\sum_{i=1}^M X_i\right| \geq r\right] \leq 3\exp\left(-\frac{r^2}{3V}\right)$$

as long as $\log(2M) \leq \frac{A^{\gamma_0}}{2b^{\gamma_0}} = \frac{V^{\gamma_0}}{2(rb)^{\gamma_0}}$ and

$$-\frac{V^{\gamma_0}}{2r^{\gamma_0}b^{\gamma_0}} \leq -\frac{r^2}{2V}.$$

Note that the latter condition may be rewritten as $r^{2+\gamma_0}b^{\gamma_0} \leq V^{\gamma_0+1}$ which is satisfied in case (68), while the former condition is

$$r \leq \frac{V}{b(2\log(2M))^{1/\gamma_0}}$$

which is also satisfied under our assumption (68). □

For the sum of independent random variables, each of which is approximately a multivariate Gaussian, the following simple concentration estimate towards a Gaussian holds true.

Lemma 14. *Let $X_1, \ldots, X_M$ be independent $\mathbb{R}^N$-valued random variables with zero mean and uniformly bounded stretched exponential moments*

$$\mathbb{E}\left[\exp\left(\frac{|X_i|^{\gamma_0}}{b^{\gamma_0}}\right)\right] \leq 2$$

for some $\gamma_0 > 0$ and some $b > 0$. Let $0 < \tau \leq 1$ and suppose that the probability distribution of each X_m is close to a Gaussian $\mathcal{N}_{\Lambda_m}$ with $||\mathcal{N}_{\Lambda_m}||_{\exp^{\gamma_0}} \leq b$ in the 1-Wasserstein distance

$$\mathcal{W}_1(X_m, \mathcal{N}_{\Lambda_m}) \leq \tau b.$$

Then there exists a probability space and random variables Y and Z such that the law of the sum

$$X := \sum_{m=1}^{M} X_m$$

coincides with the law of the sum

$$Y + Z,$$

where Y is a multivariate Gaussian with covariance matrix $\Lambda := \sum_{m=1}^{M} \Lambda_m$ and where Z is a random variable subject to the estimate

$$\mathbb{P}[|Z| \geq r] \leq 3N \exp\bigg(- \frac{r^2}{3N\tau|\log\tau|^{1/\gamma_0} M b^2} \bigg)$$

for any

$$r \leq \sqrt{\tau|\log\tau|^{1/\gamma_0}}\sqrt{M} b \min\bigg\{ \frac{\sqrt{M\tau|\log\tau|^{1/\gamma_0}}}{(2\log(2M))^{1/\gamma_0}}, \big(\tau|\log\tau|^{1/\gamma_0} M\big)^{\gamma_0/(4+2\gamma_0)} \bigg\}.$$

Proof. By the stochastic independence of the X_m and the definition of the 1-Wasserstein distance $\mathcal{W}_1(X_m, \mathcal{N}_{\Lambda_m})$, there exists a probability space and independent triples of random variables $(\tilde{X}_m, Y_m, Z_m)$ with $\tilde{X}_m \stackrel{d}{=} X_m$ and

$$\tilde{X}_m = Y_m + Z_m$$

where Y_m is a Gaussian random variable with covariance matrix Λ_m and where Z_m satisfies

$$\mathbb{E}[|Z_m|] = \mathcal{W}_1(X_m, \mathcal{N}_{\Lambda_m}).$$

By writing $Z_m = \tilde{X}_m - Y_m$, we deduce using Lemma 15

$$||Z_m||_{\exp^{\gamma_0}} \leq ||X_m||_{\exp^{\gamma_0}} + ||Y_m||_{\exp^2} \leq Cb + Cb$$

and therefore for A with $z^2/b^2 \leq \exp(z^{\gamma_0}/2(Cb)^{\gamma_0})$ for all $z \geq A$ (using again Lemma 15)

$$\begin{aligned}
\mathbb{E}[|Z_m|^2] &\leq A\mathbb{E}[|Z_m|] + \mathbb{E}[|Z_m|^2 \chi_{|Z_m| \geq A}] \\
&\leq A\mathbb{E}[|Z_m|] + b^2 \exp(-A^{\gamma_0}/2b^{\gamma_0})\mathbb{E}[\exp(|Z_m|^{\gamma_0}/(Cb)^{\gamma_0})] \\
&\leq A \cdot Cb\tau + 2b^2 \exp(-A^{\gamma_0}/2b^{\gamma_0}).
\end{aligned}$$

Choosing $A := Cb|\log\tau|^{1/\gamma_0}$ (which satisfies the previous assumption) we get

$$\mathbb{E}[|Z_m|^2] \leq C\tau b^2 |\log\tau|^{1/\gamma_0}.$$

As the sum of independent Gaussian random variables is again a Gaussian random variable, we get that

$$Y = \sum_{m=1}^{M} Y_m$$

is a multivariate Gaussian with covariance matrix $\Lambda = \sum_{m=1}^{M} \Lambda_m$. On the other hand, applying Lemma 13 to

$$Z := \sum_{m=1}^{M} Z_m$$

with the choice $V := M\tau b^2|\log\tau|^{1/\gamma_0}$, we obtain

$$\mathbb{P}[|Z|\geq r]\leq 3N\exp\bigg(-\frac{r^2}{3N\tau|\log\tau|^{1/\gamma_0}Mb^2}\bigg)$$

for any

$$r\leq\sqrt{\tau|\log\tau|^{1/\gamma_0}}\sqrt{M}b\min\bigg\{\frac{\sqrt{M\tau|\log\tau|^{1/\gamma_0}}}{(2\log(2M))^{1/\gamma_0}},\big(\tau|\log\tau|^{1/\gamma_0}M\big)^{\gamma_0/(4+2\gamma_0)}\bigg\}.$$

□

Appendix C. Calculus for random variables with stretched exponential moments

Throughout this paper, we have equipped the space of random variables X with stretched exponential moments in the sense

$$\mathbb{E}\bigg[\exp\bigg(\frac{|X|^\gamma}{C}\bigg)\bigg]\leq 2$$

for some $\gamma>0$ and some $C>0$ with the norm

$$||X||_{\exp^\gamma}:=\sup_{p\geq 1}\frac{1}{p^{1/\gamma}}\mathbb{E}\big[|X|^p\big]^{1/p}.$$

In the setting of exponential or higher moments $\gamma\geq 1$, this norm is equivalent to the Luxemburg norm associated with the convex function $\exp(x^\gamma)-1$. However, it has two advantages over the Luxemburg norm: First, it simplifies calculus when considering the integrability of products of random variables or the concentration properties of independent random variables. Secondly and more importantly, it is also a well-defined norm on the space of random variables X with subexponential stretched exponential moments $\gamma\in(0,1)$.

Lemma 15. *Let $\gamma>0$. Consider a random variable X on some probability space. Define the quasinorm*

$$||X||_{\exp^\gamma,\mathrm{quasi}}:=\inf\bigg\{s>0:\mathbb{E}\bigg[\exp\bigg(\frac{|X|^\gamma}{s^\gamma}\bigg)\bigg]\leq 2\bigg\}.$$

Then we have $||X||_{\exp^\gamma,\mathrm{quasi}}<\infty$ if and only if $||X||_{\exp^\gamma}<\infty$ and there exist constants $c(\gamma),C(\gamma)$ such that the estimate

$$c(\gamma)||X||_{\exp^\gamma}\leq||X||_{\exp^\gamma,\mathrm{quasi}}\leq C(\gamma)||X||_{\exp^\gamma}$$

is satisfied.

We omit the (elementary) proof of this lemma and the next lemma; the proofs may be found in the companion article [18].

Lemma 16 (Calculus for random variables with stretched exponential moments)**.** *Let X, Y be random variables with stretched exponential moments in the sense $||X||_{\exp^\gamma}<\infty$ and $||Y||_{\exp^\beta}<\infty$ for some $\gamma,\beta>0$.*

a) *The product XY has stretched exponential moments with exponent α given by $\frac{1}{\alpha}=\frac{1}{\gamma}+\frac{1}{\beta}$ and satisfies the bound*

$$||XY||_{\exp^\alpha}\leq C(\beta,\gamma)||X||_{\exp^\gamma}||Y||_{\exp^\beta}.$$

b) *There exists constants* $c = c(\gamma) > 0$, $C = C(\gamma) < \infty$, *with the following property: For any* $K \geq 0$, *we have the estimate*

$$\mathbb{P}\big[|X| \geq K||X||_{\exp^\gamma}\big] \leq C\exp(-cK^\gamma).$$

For independent random variables with stretched exponential moments, the following simple concentration estimate holds. Again, we omit the proof and refer the reader to [18].

Lemma 17. *Let* $X_1, \ldots, X_M$ *be independent random variables with vanishing expectation and uniformly bounded stretched exponential moments*

$$||X_m||_{\exp^{\gamma_0}} \leq b$$

for some $\gamma_0 > 0$ *and some* $b > 0$. *Then the sum*

$$X := \sum_{m=1}^{M} X_m$$

has uniformly bounded stretched exponential moments

$$||X||_{\exp^{\tilde\gamma}} \leq C(\gamma_0)\sqrt{M}b$$

for $\tilde\gamma := \gamma_0/(\gamma_0+1)$.

References

[1] S. Armstrong, T. Kuusi, and J.-C. Mourrat. The additive structure of elliptic homogenization. *Invent. Math.*, 208(3):999–1154, 2017.

[2] M. Avellaneda and F.-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, 40(6):803–847, 1987.

[3] P. Baldi and Y. Rinott. On normal approximations of distributions in terms of dependency graphs. *Ann. Probab.*, 17(4):1646–1650, 1989.

[4] P. Baldi, Y. Rinott, and C. Stein. A normal approximation for the number of local maxima of a random function on a graph. In *Probability, statistics, and mathematics*, pages 59–81. Academic Press, Boston, MA, 1989.

[5] P. Bella, B. Fehrman, J. Fischer, and F. Otto. Stochastic homogenization of linear elliptic equations: Higher-order error estimates in weak norms via second-order correctors. *SIAM J. Math. Anal.*, 49(6):4658–4703, 2017.

[6] G. Bennett. Probability inequalities for the sum of independent random variables. *J. Amer. Stat. Assoc.*, 57(297):33–45, 1962.

[7] A. Benoit and A. Gloria. Long-time homogenization and asymptotic ballistic transport of classical waves. *Preprint*, 2017. arXiv:1701.08600.

[8] R. N. Bhattacharya and R. Ranga Rao. *Normal approximation and asymptotic expansions.* Robert E. Krieger Publishing Co., Inc., Melbourne, FL, 1986. Reprint of the 1976 original.

[9] M. Biskup, M. Salvi, and T. Wolff. A central limit theorem for the effective conductance: linear boundary data and small ellipticity contrasts. *Comm. Math. Phys.*, 328(2):701–731, 2014.

[10] S. Chatterjee. A new method of normal approximation. *Ann. Probab.*, 36(4):1584–1610, 2008.

[11] S. Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Related Fields*, 143(1-2):1–40, 2009.

[12] S. Chatterjee and E. Meckes. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008.

[13] L. H. Y. Chen, L. Goldstein, and Q.-M. Shao. *Normal approximation by Stein's method.* Probability and its Applications (New York). Springer, Heidelberg, 2011.

[14] L. H. Y. Chen and Q.-M. Shao. Normal approximation under local dependence. *Ann. Probab.*, 32(3A):1985–2028, 2004.

[15] M. Duerinckx, J. Fischer, A. Gloria, and F. Otto. The structure of fluctuations in stochastic homogenization: The case of finite range of dependence. *in preparation*, 2019.

[16] M. Duerinckx and A. Gloria. Weighted second-order Poincaré inequalities: Application to RSA models. *Preprint*, 2017. arXiv:1711.03158.

[17] M. Duerinckx, A. Gloria, and F. Otto. The structure of fluctuations in stochastic homogenization. *Preprint*, 2016. arXiv:1602.01717.

[18] J. Fischer. The choice of representative volumes in the approximation of effective properties of random materials. *Preprint*, 2018. arXiv:1807.00834.

[19] J. Fischer and F. Otto. A higher-order large-scale regularity theory for random elliptic operators. *Comm. Partial Differential Equations*, 41(7):1108–1148, 2016.

[20] A. Gloria, S. Neukamm, and F. Otto. An optimal quantitative two-scale expansion in stochastic homogenization of discrete elliptic equations. *ESAIM Math. Model. Numer. Anal.*, 48(2):325–346, 2014.

[21] A. Gloria and J. Nolen. A quantitative central limit theorem for the effective conductance on the discrete torus. *Comm. Pure Appl. Math.*, 69(12):2304–2348, 2016.

[22] A. Gloria and F. Otto. The corrector in stochastic homogenization: optimal rates, stochastic integrability, and fluctuations. *Preprint*, 2015. arXiv:1510.08290.

[23] L. Goldstein and Y. Rinott. Multivariate normal approximations by stein's method and size bias couplings. *Journal of Applied Probability*, 33(1):1–17, 1996.

[24] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991.

[25] Y. Gu. High order correctors and two-scale expansions in stochastic homogenization. *Probab. Theory Related Fields*, 169(3):1221–1259, 2017.

[26] Y. Gu and J.-C. Mourrat. Scaling limit of fluctuations in stochastic homogenization. *Multiscale Model. Simul.*, 14(1):452–481, 2016.

[27] R. Lachièze-Rey and G. Peccati. New berry-esseen bounds for functionals of binomial point processes. *Ann. Appl. Probab.*, 27(4):1992–2031, 2017.

[28] G. Last, G. Peccati, and M. Schulte. Normal approximation on Poisson spaces: Mehler's formula, second order Poincaré inequalities and stabilization. *Probab. Theory Related Fields*, 165(3-4):667–723, 2016.

[29] C. Le Bris, F. Legoll, and W. Minvielle. Special quasirandom structures: a selection approach for stochastic homogenization. *Monte Carlo Methods Appl.*, 22(1):25–54, 2016.

[30] J.-C. Mourrat and J. Nolen. Scaling limit of the corrector in stochastic homogenization. *Ann. Appl. Probab.*, 27(2):944–959, 2017.

[31] J.-C. Mourrat and F. Otto. Correlation structure of the corrector in stochastic homogenization. *Ann. Probab.*, 44(5):3207–3233, 2016.

[32] J. Nolen. Normal approximation for the net flux through a random conductor. *Stoch. Partial Differ. Equ. Anal. Comput.*, 4(3):439–476, 2016.

[33] I. Nourdin, G. Peccati, and G. Reinert. Second order Poincaré inequalities and CLTs on Wiener space. *J. Funct. Anal.*, 257(2):593–609, 2009.

[34] M. D. Penrose and J. E. Yukich. Normal approximation in geometric probability. In *Stein's method and applications*, volume 5 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 37–58. Singapore Univ. Press, Singapore, 2005.

[35] G. Reinert and A. Röllin. Multivariate normal approximation with stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.*, 37(6):2150–2173, 2009.

[36] Y. Rinott. On normal approximation rates for certain sums of dependent random variables. *J. Comput. Appl. Math.*, 55(2):135–143, 1994.

[37] Y. Rinott and V. Rotar. A multivariate CLT for local dependence with $n^{-1/2}\log n$ rate and applications to multivariate graph related statistics. *J. Multivariate Anal.*, 56(2):333–350, 1996.

[38] Y. Rinott and V. Rotar. On edgeworth expansions for dependency-neighborhoods chain structures and stein's method. *Probab. Theory Related Fields*, 126(4):528–570, 2003.

[39] N. Ross. Fundamentals of Stein's method. *Probab. Surv.*, 8:210–293, 2011.

[40] R. Rossignol. Noise-stability and central limit theorems for effective resistance of random electric networks. *Ann. Probab.*, 44(2):1053–1106, 2016.

[41] T. Schreiber, M. D. Penrose, and J. E. Yukich. Gaussian limits for multidimensional random sequential packing at saturation. *Comm. Math. Phys.*, 272(1):167–183, 2007.

[42] C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. pages 583–602, 1972.

[43] C. Stein. *Approximate computation of expectations*, volume 7 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1986.

Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria, E-Mail: julian.fischer@ist.ac.at