Electronic Journal of Probability

Electron. J. Probab. **22** (2017), no. 87, 1–30.
ISSN: 1083-6489 DOI: 10.1214/17-EJP106

Multivariate central limit theorems for Rademacher functionals with applications

Kai Krokowski* Christoph Thäle†

Abstract

Quantitative multivariate central limit theorems for general functionals of independent, possibly non-symmetric and non-homogeneous infinite Rademacher sequences are proved by combining discrete Malliavin calculus with the smart path method for normal approximation. In particular, a discrete multivariate second-order Poincaré inequality is developed. As a first application, the normal approximation of vectors of subgraph counting statistics in the Erdős-Rényi random graph is considered. In this context, we further specialize to the normal approximation of vectors of vertex degrees. In a second application we prove a quantitative multivariate central limit theorem for vectors of intrinsic volumes induced by random cubical complexes.

Keywords: discrete Malliavin calculus; intrinsic volume; multivariate central limit theorem; smart path method; subgraph count; random graph; random cubical complex; vertex degree.
AMS MSC 2010: Primary 60F05, Secondary 05C80; 60C05; 60D05; 60H07.
Submitted to EJP on January 25, 2017, final version accepted on September 11, 2017.

1 Introduction

Suppose that $X = (X_k)_{k\in\mathbb{N}}$ is an independent Rademacher sequence, that is, a sequence of independent random variables satisfying, for all $k \in \mathbb{N}$, $P(X_k = 1) = p_k$ and $P(X_k = -1) = q_k = 1 - p_k$ for some $p_k \in (0, 1)$. Further, fix a dimension parameter $d \in \mathbb{N}$ and let $F_1 = F_1(X), \dots, F_d = F_d(X)$ be d random variables depending on possibly infinite many members of the Rademacher sequence X. We shall refer to such random variables as Rademacher functionals in what follows. The goal of this paper is to derive handy conditions under which the random vector $\mathbf{F} = (F_1, \dots, F_d)$ consisting of d Rademacher functionals is close in distribution to a d-dimensional Gaussian random vector. In our paper the distributional closeness will be measured by means of a multivariate probability metric based on four times partially differentiable test functions. We will provide two versions of such a result. One is in the spirit of the Malliavin-Stein method and expresses the distributional closeness in terms of so-called discrete Malliavin

*Faculty of Mathematics, Ruhr University Bochum, Germany. E-mail: kai.krokowski@rub.de
†Faculty of Mathematics, Ruhr University Bochum, Germany. E-mail: christoph.thaele@rub.de

operators. The second one is a multivariate discrete second-order Poincaré inequality, a bound which only involves the first- and second-order discrete Malliavin derivatives of the Rademacher functionals $F_1, \dots, F_d$, or, more precisely, their moments up to order four. More formally, if $F = F(X)$ is a Rademacher functional, the discrete Malliavin derivative $D_k F$ in direction $k \in \mathbb{N}$ is defined as $D_k F = \sqrt{p_k q_k}(F_k^+ - F_k^-)$, where $F_k^\pm$ is the Rademacher functional for which the kth coordinate X_k of the Rademacher sequence X is conditioned to be ± 1. The second-order discrete derivative is iteratively given by $D_k D_\ell F$ for $k, \ell \in \mathbb{N}$. Such a bound is particularly attractive for concrete applications as demonstrated in the present text.

Let us describe the purpose and the content of our paper in some more detail.

(i) First of all, our aim is to provide a multivariate quantitative central limit theorem for vectors of Rademacher functionals by bringing together the discrete Malliavin calculus of variations with the so-called smart-path method for normal approximation. This leads to a limit theorem in the spirit of the Malliavin-Stein method and generalizes an earlier result from [10], where the underlying Rademacher sequence has been assumed to be homogeneous and symmetric, meaning that $p_k = q_k = 1/2$ for all $k \in \mathbb{N}$ in above notation.

(ii) From this result, a further aim of this text is to develop a discrete multivariate second-order Poincaré inequality, that is, a bound for the multivariate normal approximation that only involves the first- and second-order discrete Malliavin derivatives, or, more precisely, its moments up to order four. Such a result can be regarded as the multivariate analogue of the main theorem obtained in [11].

(iii) Finally, we want to demonstrate the flexibility and applicability of our discrete multivariate second-order Poincaré inequality by means of examples from the theory of random graphs and random topology. First, we are going to provide a bound of order $O(n^{-1})$ for the multivariate normal approximation of a vector of subgraph counts in the classical Erdős-Rényi random graph. This generalizes (in a different probability metric) a result of Reinert and Röllin [20], where vectors of the number of edges, 2-stars and triangles have been considered, and adds a rate of convergence to the related central limit theorem in the paper of Janson and Nowicki [7]. Moreover, for the same model we also provide a multivariate central limit theorem for the random vector of vertex degrees with a rate of convergence of order $O(n^{-1/2})$. This can be seen as a version of the result of Goldstein and Rinott [5] and is the multivariate analogue of a related Berry-Esseen bound proved by Goldstein [4] and Krokowski, Reichenbachs and Thäle [11]. Second, we consider the vector of intrinsic volumes determined by different models of random cubical complexes in $\mathbb{R}^d$ and derive bounds of order $O(n^{-d/2})$ on the error in their normal approximation. This constitutes a multivariate extension of the central limit theorem provided by Werman and Wright [25] and is in line with recent developments in the active field of random topology, see [1, 2, 8, 12] as well as the references cited therein.

Our results continue a recent line of research concerning limit theorems for Rademacher functionals. The field has been opened by Nourdin, Peccati and Reinert [13], who proved first limit theorems for a class of smooth probability metrics. Later, Krokowski, Reichenbachs and Thäle [10, 11] considered Berry-Esseen bounds and provided a first univariate discrete second-order Poincaré inequality. Zheng [26] has obtained a refined bound for the Wasserstein distance and also proved almost sure central limit theorems. Moreover, Privault and Torrisi [18] as well as Krokowski [9] also derived bounds for the Poisson approximation of Rademacher functionals.

This text is organized as follows. In Section 2 we briefly recall the basis of discrete Malliavin calculus in order to keep the paper reasonably self-contained. A first quantitative multivariate central limit theorem for functionals of an independent possibly non-symmetric and non-homogeneous infinite Rademacher sequence based on the discrete Malliavin-Stein method is presented in Section 3.1, while Section 3.2 contains the discrete multivariate second-order Poincaré inequality. The applications to subgraph and vertex degree counts in the Erdős-Rényi random graph and to the intrinsic volumes of random cubical complexes are discussed in the final Section 4.

2 Discrete Malliavin calculus

In this section we briefly recall the basis of discrete Malliavin calculus. We refer to the monograph [17] as well as to the papers [10, 11, 13] for details, proofs and further references.

Rademacher sequences. Let $p := (p_k)_{k\in\mathbb{N}}$ be a sequence of success probabilities $0 < p_k < 1$ and put $q := (q_k)_{k\in\mathbb{N}}$ with $q_k := 1 - p_k$. Furthermore, let $(\Omega, \mathcal{F}, P)$ be the following probability space: $\Omega := \{-1, +1\}^{\mathbb{N}}$, $\mathcal{F} := \mathrm{power}(\{-1, +1\})^{\otimes\mathbb{N}}$, where $\mathrm{power}(\,\cdot\,)$ denotes the power set of the argument set, and $P := \bigotimes_{k=1}^{\infty}(p_k\delta_{+1} + q_k\delta_{-1})$ with $\delta_{\pm1}$ being the unit-mass Dirac measure at ±1. We let $X := (X_k)_{k\in\mathbb{N}}$ be a sequence of independent random variables defined on $(\Omega, \mathcal{F}, P)$ by $X_k(\omega) := \omega_k$, for every $k \in \mathbb{N}$ and $\omega := (\omega_k)_{k\in\mathbb{N}} \in \Omega$. We refer to such a sequence X as independent (possibly non-symmetric and non-homogeneous infinite) Rademacher sequence. We also define the standardized sequence $Y := (Y_k)_{k\in\mathbb{N}}$ by putting $Y_k := (\mathrm{Var}(X_k))^{-1/2}(X_k - \mathbb{E}[X_k]) = (2\sqrt{p_kq_k})^{-1}(X_k - p_k + q_k)$ for every $k \in \mathbb{N}$. Note that $Y_k = X_k$, iff X is a homogeneous and symmetric Rademacher sequence, that is, if $p_k = q_k = 1/2$, for all $k \in \mathbb{N}$.

Discrete multiple stochastic integrals and chaos decomposition. Let us denote by κ the counting measure on $\mathbb{N}$. We put $\ell^2(\mathbb{N})^{\otimes n} := L^2(\mathbb{N}^n, \mathcal{P}(\mathbb{N})^{\otimes n}, \kappa^{\otimes n})$ for every $n \in \mathbb{N}$ and refer to the elements of that space as kernels. Let $\ell^2(\mathbb{N})^{\circ n}$ denote the subset of $\ell^2(\mathbb{N})^{\otimes n}$ consisting of symmetric kernels and let $\ell_0^2(\mathbb{N})^{\otimes n}$ be the subset of kernels vanishing on diagonals, that is, vanishing on the complement of the set $\Delta_n := \{(i_1, \dots, i_n) \in \mathbb{N}^n : i_j \neq i_k \text{ for } j \neq k\}$. We then put $\ell_0^2(\mathbb{N})^{\circ n} := \ell^2(\mathbb{N})^{\circ n} \cap \ell_0^2(\mathbb{N})^{\otimes n}$.

For $n \in \mathbb{N}$ and $f \in \ell_0^2(\mathbb{N})^{\circ n}$, we define the discrete multiple stochastic integral of order n of f by

$$\begin{aligned} J_n(f) &:= \sum_{(i_1,\dots,i_n)\in\mathbb{N}^n} f(i_1, \dots, i_n)Y_{i_1} \cdot \ldots \cdot Y_{i_n} = \sum_{(i_1,\dots,i_n)\in\Delta_n} f(i_1, \dots, i_n)Y_{i_1} \cdot \ldots \cdot Y_{i_n} \\ &= n! \sum_{1\le i_1<\ldots<i_n<\infty} f(i_1, \dots, i_n)Y_{i_1} \cdot \ldots \cdot Y_{i_n}\,. \end{aligned}$$

In addition, we put $\ell^2(\mathbb{N})^{\otimes 0} := \mathbb{R}$ and $J_0(c) := c$, for every $c \in \mathbb{R}$.

It is an important fact that every $F \in L^2(\Omega)$ admits a decomposition of the form

$$F = \mathbb{E}[F] + \sum_{n=1}^{\infty} J_n(f_n) \tag{2.1}$$

with uniquely determined kernels $f_n \in \ell_0^2(\mathbb{N})^{\circ n}$ (cf. Section 2.4 in [13] for the form of these uniquely determined kernels).

Discrete Malliavin derivative. For every $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ and $k \in \mathbb{N}$ we define the two sequences $\omega_\pm^k$ by putting $\omega_\pm^k := (\omega_1, \dots, \omega_{k-1}, \pm1, \omega_{k+1}, \dots)$. Furthermore, for every $F \in L^1(\Omega)$, $\omega \in \Omega$ and $k \in \mathbb{N}$ let $F_k^{\pm}(\omega) := F(\omega_{\pm}^k)$. For such an F the discrete Malliavin derivative is defined by $DF := (D_k F)_{k\in\mathbb{N}}$ with

$$D_k F := \sqrt{p_k q_k}\,(F_k^+ - F_k^-)\,, \qquad k \in \mathbb{N}\,. \tag{2.2}$$

Note that it immediately follows from (2.2) that, for every $k \in \mathbb{N}$, $D_k F$ is independent of X_k. In the following we state a product formula for the discrete Malliavin derivative. If $F, G \in L^1(\Omega)$, then

$$D_k(FG) = (D_k F)G + F(D_k G) - \frac{X_k}{\sqrt{p_k q_k}}(D_k F)(D_k G)\,, \qquad k \in \mathbb{N}\,. \tag{2.3}$$

For $m \in \mathbb{N}$ let us further define the iterated discrete Malliavin derivative of order m of F by $D^m F := (D^m_{k_1,\ldots,k_m} F)_{k_1,\ldots,k_m \in \mathbb{N}}$ with $D^m_{k_1,\ldots,k_m} F := D_{k_m}(D^{m-1}_{k_1,\ldots,k_{m-1}} F)$, for every $k_1, \ldots, k_m \in \mathbb{N}$, where $D^0 F := F$. Given $F \in L^2(\Omega)$ with chaos representation $F = \mathbb{E}[F] + \sum_{n=1}^{\infty} J_n(f_n)$ as in (2.1) and $m \in \mathbb{N}$, we will say that $F \in \mathrm{dom}(D^m)$, provided that

$$\mathbb{E}[\|D^m F\|^2_{\ell^2(\mathbb{N})^{\otimes m}}] = \sum_{n=m}^{\infty} \frac{n!}{(n-m)!} n! \|f_n\|^2_{\ell^2(\mathbb{N})^{\otimes n}} < \infty\,.$$

If $F \in \mathrm{dom}(D)$ with chaos decomposition (2.1), $D_k F$ can be P-almost surely be identified with the random variable given by

$$D_k F = \sum_{n=1}^{\infty} n J_{n-1}(f_n(\,\cdot\,, k))\,,$$

where $f_n(\,\cdot\,, k)$ stands for the kernel f_n with one of its variables fixed to be k (which one is irrelevant, since the kernels are symmetric).

Discrete divergence. We will now define the discrete divergence operator δ and its domain $\mathrm{dom}(\delta)$. Let $f_n \in \ell_0^2(\mathbb{N})^{\circ n-1} \otimes \ell^2(\mathbb{N})$, for every $n \in \mathbb{N}$, and consider the sequence $u := (u_k)_{k\in\mathbb{N}}$ given by $u_k := \sum_{n=1}^{\infty} J_{n-1}(f_n(\,\cdot\,, k))$ for every $k \in \mathbb{N}$. For such u, we say that $u \in \mathrm{dom}(\delta)$, if

$$\sum_{n=1}^{\infty} n! \|\widetilde{f}_n\, \mathbf{1}_{\Delta_n}\|^2_{\ell^2(\mathbb{N})^{\otimes n}} < \infty\,,$$

where $\widetilde{f}_n$ denotes the canonical symmetrization of f_n. For $u \in \mathrm{dom}(\delta)$, the discrete divergence operator δ is then defined by

$$\delta(u) := \sum_{n=1}^{\infty} J_n(\widetilde{f}_n\, \mathbf{1}_{\Delta_n})$$

(cf. Proposition 1.8.3 in [17] for a representation of $\delta(u)$ in terms of the elements of the sequence u). One can interpret δ as the operator that is adjoint to the discrete Malliavin derivative. Namely, if $F \in \mathrm{dom}(D)$ and $u \in \mathrm{dom}(\delta)$, then

$$\mathbb{E}[F\delta(u)] = \mathbb{E}[\langle DF, u\rangle_{\ell^2(\mathbb{N})}]\,. \tag{2.4}$$

Discrete Ornstein-Uhlenbeck operator and its inverse. Next, we define the discrete Ornstein-Uhlenbeck operator L and its (pseudo-)inverse L^{-1}. Given $F \in L^2(\Omega)$, again with chaos representation $F = \mathbb{E}[F] + \sum_{n=1}^{\infty} J_n(f_n)$ as above, we say that $F \in \mathrm{dom}(L)$, if

$$\sum_{n=1}^{\infty} n^2 n! \|f_n\|^2_{\ell^2(\mathbb{N})^{\otimes n}} < \infty \,.$$

For $F \in \mathrm{dom}(L)$, the discrete Ornstein-Uhlenbeck operator L is then defined by

$$LF := -\sum_{n=1}^{\infty} n J_n(f_n) \,.$$

For centered $F \in L^2(\Omega)$, its (pseudo-) inverse is given as follows:

$$L^{-1}F := -\sum_{n=1}^{\infty} \frac{1}{n} J_n(f_n) \,.$$

Discrete Ornstein-Uhlenbeck semigroup. Finally, we introduce the semigroup associated with the discrete Ornstein-Uhlenbeck operator L. The discrete Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$ is defined by

$$P_t F := \mathbb{E}[F] + \sum_{n=1}^{\infty} e^{-nt} J_n(f_n) \,, \qquad t \geq 0 \,.$$

The process associated with the discrete Ornstein-Uhlenbeck semigroup is given as follows. For every $k \in \mathbb{N}$, let X_k^* be an independent copy of X_k. Furthermore, let $(Z_k)_{k\in\mathbb{N}}$ be a sequence of independent and exponentially distributed random variables with mean 1, where Z_k is independent of X_k and X_k^*, for every $k \in \mathbb{N}$. For every real $t \geq 0$, let $X^t := (X_k^t)_{k\in\mathbb{N}}$ with

$$X_k^t := X_k^* \,\mathbf{1}_{\{Z_k \leq t\}} + X_k \,\mathbf{1}_{\{Z_k > t\}} \,, \qquad k \in \mathbb{N} \,.$$

Then, $(X^t)_{t\geq 0}$ is the discrete Ornstein-Uhlenbeck process associated with the Ornstein-Uhlenbeck semigroup $(P_t)_{t\geq 0}$. The relation of Ornstein-Uhlenbeck semigroup and process is exhibited in the following formula, known as Mehler's formula. If $F \in L^2(\Omega)$, then it P-almost surely holds that

$$P_t F = \mathbb{E}[F(X^t) \,|\, X] \,, \qquad t \geq 0 \,. \tag{2.5}$$

Integration by parts, integrated Mehler's formula and Poincaré inequality. We notice that the discrete Malliavin operators D, δ and L are related by the identity $L = -\delta D$. Moreover, the following discrete integration by parts formula is valid. If $F, G \in \mathrm{dom}(D)$, then

$$\mathbb{E}[(F - \mathbb{E}[F])G] = \mathbb{E}[\langle -DL^{-1}(F - \mathbb{E}[F]), DG\rangle_{\ell^2(\mathbb{N})}] \,. \tag{2.6}$$

Indeed, the relation $L = -\delta D$ and the adjointness of D and δ in (2.4) yield

$$\begin{aligned} \mathbb{E}[(F - \mathbb{E}[F])G] &= \mathbb{E}[LL^{-1}(F - \mathbb{E}[F])G] = \mathbb{E}[-\delta DL^{-1}(F - \mathbb{E}[F])G] \\ &= \mathbb{E}[\langle -DL^{-1}(F - \mathbb{E}[F]), DG\rangle_{\ell^2(\mathbb{N})}] \,. \end{aligned}$$

The following identity can be seen as an integrated version of Mehler's formula. If $m, k_1, \ldots, k_m \in \mathbb{N}$ and $F \in \mathrm{dom}(D^m)$ with $\mathbb{E}[F] = 0$, then it P-almost surely holds that

$$-D^m_{k_1,\ldots,k_m} L^{-1} F = \int_0^{\infty} e^{-mt} P_t D^m_{k_1,\ldots,k_m} F \, dt \,. \tag{2.7}$$

From this, one can immediately deduce the following important inequality. If $m, k_1, \ldots, k_m \in \mathbb{N}$, $\alpha \geq 1$ and $F \in \mathrm{dom}(D^m)$ with $\mathbb{E}[F] = 0$, then

$$\mathbb{E}[|D^m_{k_1,\ldots,k_m} L^{-1} F|^\alpha] \leq \mathbb{E}[|D^m_{k_1,\ldots,k_m} F|^\alpha] \,. \tag{2.8}$$

Finally, let us recall a discrete version of the classical Poincaré inequality. For every $F \in L^1(\Omega)$, it holds that

$$\mathrm{Var}(F) \leq \mathbb{E}[\|DF\|^2_{\ell^2(\mathbb{N})}] \,. \tag{2.9}$$

3 Multivariate central limit theorems

3.1 A discrete Malliavin-Stein bound

In the following, we will prove a bound on the error in the multivariate normal approximation of vectors of general functionals of independent possibly non-symmetric and non-homogeneous infinite Rademacher sequences. This way we generalize Theorem 5.1 in [10], where only functionals of independent symmetric Rademacher sequences have been considered. The proof proceeds along the lines of [10], but there are a number of subtleties arising in the more general case here that were not present before. In particular, in the non-symmetric case a new summand in the error bound becomes visible as further discussed in Remark 3.2 below. To make this and other phenomena transparent, we include the full details.

The distance between the law of a vector of Rademacher functionals and a multivariate normal distribution will be measured by the so-called d_4-distance that is defined as follows. Fix $d \in \mathbb{N}$ and let $n = 1, \ldots, d$. For an n times partially differentiable function $g : \mathbb{R}^d \to \mathbb{R}$ we put

$$M_k(g) := \max_{1 \leq i_1,\ldots,i_k \leq d} \Big\| \frac{\partial^k}{\partial x_{i_1} \ldots \partial x_{i_k}} g \Big\|_\infty$$

for every $k = 1, \ldots, n$, where $\| \cdot \|_\infty$ denotes the supremum norm of the argument function. The d_4-distance between the distributions of two $\mathbb{R}^d$-valued random vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined by

$$d_4(\mathbf{X}, \mathbf{Y}) := \sup_g |\mathbb{E}[g(\mathbf{X})] - \mathbb{E}[g(\mathbf{Y})]| \,,$$

where the supremum is running over all four times partially differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ with bounded partial derivatives fulfilling $M_1(g), M_2(g), M_3(g), M_4(g) \leq 1$. Recall that $p := (p_k)_{k\in\mathbb{N}}$ and $q := (q_k)_{k\in\mathbb{N}}$ are sequences, and thus, for a Rademacher functional $F \in \mathrm{dom}(D)$, a product of the form $(pq)^{1/4} DF = ((p_k q_k)^{1/4} D_k F)_{k\in\mathbb{N}}$ is understood coordinate-wise in what follows.

Theorem 3.1. *Fix $d \in \mathbb{N}$ and let $F_1, \ldots, F_d$ be Rademacher functionals with $F_i \in \mathrm{dom}(D)$, $\mathbb{E}[F_i] = 0$ and $\mathbb{E}[\|(pq)^{-1/4} DF_i\|^4_{\ell^4(\mathbb{N})}] < \infty$, for every $i = 1, \ldots, d$. Define $\mathbf{F} := (F_1, \ldots, F_d)$ and let $\mathbf{N} := (N_1, \ldots, N_d)$ be a centered Gaussian random vector with symmetric and positive semidefinite covariance matrix $\Sigma := (\Sigma_{ij})^d_{i,j=1}$. Further, let*

$$A_1 := \frac{1}{2} \sum_{i,j=1}^d \mathbb{E}[|\Sigma_{ij} - \langle DF_j, -DL^{-1} F_i \rangle_{\ell^2(\mathbb{N})}|] \,,$$

$$A_2 := \frac{1}{4} \mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}} \Big(\sum_{j=1}^d |DF_j| \Big)^2, \sum_{i=1}^d |-DL^{-1} F_i| \Big\rangle_{\ell^2(\mathbb{N})}\Big] \,,$$

$$A_3 := \frac{5}{24} \mathbb{E}\Big[\Big\langle \frac{1}{pq} \Big(\sum_{j=1}^d |DF_j| \Big)^3, \sum_{i=1}^d |-DL^{-1} F_i| \Big\rangle_{\ell^2(\mathbb{N})}\Big] \,.$$

Then,

$$d_4(\mathbf{F},\mathbf{N}) \le A_1 + A_2 + A_3\,.$$

Remark 3.2. A comparison of Theorem 3.1 with Theorem 5.1 in [10] shows that the extension to vectors of general functionals of independent, possibly non-symmetric and non-homogeneous infinite Rademacher sequences comes at the costs of an additional summand in the bound, namely

$$\mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}}\Big(\sum_{j=1}^d |DF_j|\Big)^2, \sum_{i=1}^d |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]\,.$$

However, resorting to the case where the underlying Rademacher sequence is symmetric, i.e., if $p_k = q_k = 1/2$, for every $k \in \mathbb{N}$, this additional summand vanishes and our bound in Theorem 3.1 coincides with the one from [10] with an improvement by a factor $1/2$ on the constant in front of the third term.

The proof of Theorem 3.1 relies on two multivariate integration by parts formulae, a Gaussian one and an approximate one from Malliavin calculus which combines (2.6) with a multivariate chain rule for the discrete gradient operator. We start by recalling the multivariate Gaussian integration by parts formula from Equation (A.41) in [24].

Lemma 3.3. *Fix $d \in \mathbb{N}$ and let $\mathbf{N} := (N_1,\dots,N_d)$ be a centered Gaussian random vector with symmetric covariance matrix $\Sigma := (\Sigma_{ij})_{i,j=1}^d$. Furthermore, let $g : \mathbb{R}^d \to \mathbb{R}$ be a partially differentiable function with bounded partial derivatives and $\mathbb{E}[|N_i g(\mathbf{N})|] < \infty$, for every $i = 1,\dots,d$. Then, for every $i = 1,\dots,d$,*

$$\mathbb{E}[N_i g(\mathbf{N})] = \sum_{j=1}^d \Sigma_{ij}\mathbb{E}\Big[\frac{\partial}{\partial x_j} g(\mathbf{N})\Big]\,.$$

The following lemma contains a multivariate chain rule for the discrete gradient operator, which is a generalization of Proposition 2.1 in [18] to the d-dimensional case. Also note that it not only generalizes Lemma 5.1 in [10] to the case where the underlying Rademacher sequence is non-symmetric and non-homogeneous, but also improves on the constants in the bound for the remainder term. For these reasons, we include a detailed proof.

Lemma 3.4. *Let $\mathbf{F}$ be a random vector of Rademacher functionals as in Theorem 3.1. Furthermore, let $f : \mathbb{R}^d \to \mathbb{R}$ be a thrice partially differentiable function. Then, for every $k \in \mathbb{N}$,*

$$\begin{aligned} D_k f(\mathbf{F}) = \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\mathbf{F}) D_k F_i - \frac{X_k}{4\sqrt{p_k q_k}} \sum_{i,j=1}^d \Big(&\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^+) \\ &+ \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^-)\Big)(D_k F_i)(D_k F_j) + R_k(\mathbf{F}) \end{aligned} \tag{3.1}$$

with $\mathbf{F}_k^\pm := \big((F_1)_k^\pm,\dots,(F_d)_k^\pm\big)$ and a remainder term $R_k(\mathbf{F})$ that fulfills

$$|R_k(\mathbf{F})| \le \frac{5}{12 p_k q_k} \sum_{i,j,\ell=1}^d \Big\|\frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f\Big\|_\infty |(D_k F_i)(D_k F_j)(D_k F_\ell)| \tag{3.2}$$

for every $k \in \mathbb{N}$.

Proof. Fix $k \in \mathbb{N}$ and observe that

$$\begin{aligned} D_k f(\mathbf{F}) &= \sqrt{p_k q_k}\,(f(\mathbf{F}_k^+) - f(\mathbf{F}_k^-)) \\ &= \sqrt{p_k q_k}\,(f(\mathbf{F}_k^+) - f(\mathbf{F})) - \sqrt{p_k q_k}\,(f(\mathbf{F}_k^-) - f(\mathbf{F}))\,. \end{aligned} \tag{3.3}$$

Now, a Taylor series expansion of f at $\mathbf{F}$ yields that, for every $\boldsymbol{x} := (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$\begin{aligned} f(\boldsymbol{x}) - f(\mathbf{F}) &= \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\mathbf{F})(x_i - F_i) + \frac{1}{2}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})(x_i - F_i)(x_j - F_j) \\ &\quad + \frac{1}{6}\sum_{i,j,\ell=1}^d \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f(\mathbf{F} + \theta(\boldsymbol{x} - \mathbf{F}))(x_i - F_i)(x_j - F_j)(x_\ell - F_\ell) \end{aligned}$$

with some $\theta := \theta(\boldsymbol{x}, \mathbf{F}) \in (0,1)$. By re-writing each of the quantities $f(\mathbf{F}_k^+) - f(\mathbf{F})$ and $f(\mathbf{F}_k^-) - f(\mathbf{F})$ in this way, it follows from (3.3) that, for every $k \in \mathbb{N}$,

$$\begin{aligned} &D_k f(\mathbf{F}) \\ &= \sqrt{p_k q_k}\sum_{i=1}^d \frac{\partial}{\partial x_i} f(\mathbf{F})((F_i)_k^+ - F_i) + \frac{1}{2}\sqrt{p_k q_k}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})((F_i)_k^+ - F_i)((F_j)_k^+ - F_j) \\ &\quad + R_1(\mathbf{F}, \mathbf{F}_k^+) - \sqrt{p_k q_k}\sum_{i=1}^d \frac{\partial}{\partial x_i} f(\mathbf{F})((F_i)_k^- - F_i) \\ &\quad - \frac{1}{2}\sqrt{p_k q_k}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})((F_i)_k^- - F_i)((F_j)_k^- - F_j) - R_2(\mathbf{F}, \mathbf{F}_k^-) \\ &= \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\mathbf{F}) D_k F_i + \frac{1}{2}\sqrt{p_k q_k}\sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})\big[((F_i)_k^+ - F_i)((F_j)_k^+ - F_j) \\ &\quad - ((F_i)_k^- - F_i)((F_j)_k^- - F_j)\big] + R_1(\mathbf{F}, \mathbf{F}_k^+) - R_2(\mathbf{F}, \mathbf{F}_k^-)\,, \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} R_1(\mathbf{F}, \mathbf{F}_k^+) &:= \frac{1}{6}\sqrt{p_k q_k}\sum_{i,j,\ell=1}^d \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f(\mathbf{F} + \theta_1(\boldsymbol{F}_k^+ - \mathbf{F})) \\ &\qquad \times ((F_i)_k^+ - F_i)((F_j)_k^+ - F_j)((F_\ell)_k^+ - F_\ell)\,, \\ R_2(\mathbf{F}, \mathbf{F}_k^-) &:= \frac{1}{6}\sqrt{p_k q_k}\sum_{i,j,\ell=1}^d \frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f(\mathbf{F} + \theta_2(\boldsymbol{F}_k^- - \mathbf{F})) \\ &\qquad \times ((F_i)_k^- - F_i)((F_j)_k^- - F_j)((F_\ell)_k^- - F_\ell) \end{aligned}$$

with $\theta_1 := \theta_1(\mathbf{F}, \mathbf{F}_k^+) \in (0,1)$ and $\theta_2 := \theta_2(\mathbf{F}, \mathbf{F}_k^-) \in (0,1)$. From the identities

$$F_k^+ - F = (F_k^+ - F_k^-)\,\mathbf{1}_{\{X_k=-1\}} = \frac{1}{\sqrt{p_k q_k}}(D_k F)\,\mathbf{1}_{\{X_k=-1\}} \tag{3.5}$$

and

$$F_k^- - F = (F_k^- - F_k^+)\,\mathbf{1}_{\{X_k=+1\}} = -\frac{1}{\sqrt{p_k q_k}}(D_k F)\,\mathbf{1}_{\{X_k=+1\}} \tag{3.6}$$

it follows that, for every $k \in \mathbb{N}$,

$$
\begin{aligned}
|R_1(\mathbf{F},\mathbf{F}_k^+)| &\le \frac{1}{6}\sqrt{p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f\Big\|_\infty |((F_i)_k^+ - F_i)((F_j)_k^+ - F_j)((F_\ell)_k^+ - F_\ell)| \\
&= \frac{1}{6 p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f\Big\|_\infty |(D_k F_i)(D_k F_j)(D_k F_\ell)|\,\mathbf{1}_{\{X_k=-1\}} \qquad (3.7)
\end{aligned}
$$

and

$$
\begin{aligned}
|R_2(\mathbf{F},\mathbf{F}_k^-)| &\le \frac{1}{6}\sqrt{p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f\Big\|_\infty |((F_i)_k^- - F_i)((F_j)_k^- - F_j)((F_\ell)_k^- - F_\ell)| \\
&= \frac{1}{6 p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i \partial x_j \partial x_\ell} f\Big\|_\infty |(D_k F_i)(D_k F_j)(D_k F_\ell)|\,\mathbf{1}_{\{X_k=+1\}}\,. \qquad (3.8)
\end{aligned}
$$

Again, by virtue of (3.5) and (3.6), for every $k \in \mathbb{N}$, the second summand on the right hand side of (3.4) can be rewritten as

$$
\begin{aligned}
&\frac{1}{2}\sqrt{p_k q_k}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})\big(((F_i)_k^+ - F_i)((F_j)_k^+ - F_j) - ((F_i)_k^- - F_i)((F_j)_k^- - F_j)\big) \\
&= \frac{1}{2\sqrt{p_k q_k}}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})(D_k F_i)(D_k F_j)(\mathbf{1}_{\{X_k=-1\}} - \mathbf{1}_{\{X_k=+1\}}) \\
&= -\frac{X_k}{2\sqrt{p_k q_k}}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F})(D_k F_i)(D_k F_j)\,. \qquad (3.9)
\end{aligned}
$$

Another Taylor series expansion of $\frac{\partial^2}{\partial x_i \partial x_j} f$ at $\mathbf{F}_k^+$ and $\mathbf{F}_k^-$, respectively, yields that, for every $i,j,k \in \mathbb{N}$,

$$
\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^+) + \sum_{\ell=1}^{d}\frac{\partial^3}{\partial x_\ell \partial x_i \partial x_j} f(\mathbf{F}_k^+ + \theta_3(\mathbf{F}-\mathbf{F}_k^+))(F_\ell - (F_\ell)_k^+)
$$

and

$$
\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}) = \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^-) + \sum_{\ell=1}^{d}\frac{\partial^3}{\partial x_\ell \partial x_i \partial x_j} f(\mathbf{F}_k^- + \theta_4(\mathbf{F}-\mathbf{F}_k^-))(F_\ell - (F_\ell)_k^-)\,,
$$

where $\theta_3 := \theta_3(\mathbf{F},\mathbf{F}_k^+) \in (0,1)$ and $\theta_4 := \theta_4(\mathbf{F},\mathbf{F}_k^-) \in (0,1)$. This adds up to

$$
\begin{aligned}
\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}) = \frac{1}{2}\Big(&\frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{F}_k^-)\Big) \\
&+ \frac{1}{2}\sum_{\ell=1}^{d}\frac{\partial^3}{\partial x_\ell \partial x_i \partial x_j} f(\mathbf{F}_k^+ + \theta_3(\mathbf{F}-\mathbf{F}_k^+))(F_\ell - (F_\ell)_k^+) \\
&+ \frac{1}{2}\sum_{\ell=1}^{d}\frac{\partial^3}{\partial x_\ell \partial x_i \partial x_j} f(\mathbf{F}_k^- + \theta_4(\mathbf{F}-\mathbf{F}_k^-))(F_\ell - (F_\ell)_k^-)
\end{aligned}
$$

for every $i,j,k \in \mathbb{N}$, and thus, it follows from (3.9) that for every $k \in \mathbb{N}$

$$\begin{aligned}
&\frac{1}{2}\sqrt{p_k q_k}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{F})(((F_i)_k^+ - F_i)((F_j)_k^+ - F_j) - ((F_i)_k^- - F_i)((F_j)_k^- - F_j))\\
&= -\frac{X_k}{4\sqrt{p_k q_k}}\sum_{i,j=1}^{d}\Big(\frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{F}_k^-)\Big)(D_kF_i)(D_kF_j) \qquad (3.10)\\
&\quad - R_3(\mathbf{F},\mathbf{F}_k^+) - R_4(\mathbf{F},\mathbf{F}_k^-)\,,
\end{aligned}$$

where

$$\begin{aligned}
R_3(\mathbf{F},\mathbf{F}_k^+) &:= \frac{X_k}{4\sqrt{p_k q_k}}\sum_{i,j,\ell=1}^{d}\frac{\partial^3}{\partial x_\ell\partial x_i\partial x_j}f(\mathbf{F}_k^+ + \theta_3(\mathbf{F}-\mathbf{F}_k^+))(D_kF_i)(D_kF_j)(F_\ell - (F_\ell)_k^+)\,,\\
R_4(\mathbf{F},\mathbf{F}_k^-) &:= \frac{X_k}{4\sqrt{p_k q_k}}\sum_{i,j,\ell=1}^{d}\frac{\partial^3}{\partial x_\ell\partial x_i\partial x_j}f(\mathbf{F}_k^- + \theta_4(\mathbf{F}-\mathbf{F}_k^-))(D_kF_i)(D_kF_j)(F_\ell - (F_\ell)_k^+)\,.
\end{aligned}$$

By the fact that $|X_k| \le 1$ for every $k \in \mathbb{N}$ and another application of (3.5) and (3.6) it holds that

$$\begin{aligned}
|R_3(\mathbf{F},\mathbf{F}_k^+)| &\le \frac{1}{4\sqrt{p_k q_k}}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i\partial x_j\partial x_\ell}f\Big\|_\infty |(D_kF_j)(D_kF_\ell)((F_i)_k^+ - F_i)|\\
&= \frac{1}{4p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i\partial x_j\partial x_\ell}f\Big\|_\infty |(D_kF_i)(D_kF_j)(D_kF_\ell)|\,\mathbf{1}_{\{X_k=-1\}} \qquad (3.11)
\end{aligned}$$

and

$$\begin{aligned}
|R_4(\mathbf{F},\mathbf{F}_k^-)| &\le \frac{1}{4\sqrt{p_k q_k}}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i\partial x_j\partial x_\ell}f\Big\|_\infty |(D_kF_j)(D_kF_\ell)((F_i)_k^- - F_i)|\\
&= \frac{1}{4p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i\partial x_j\partial x_\ell}f\Big\|_\infty |(D_kF_i)(D_kF_j)(D_kF_\ell)|\,\mathbf{1}_{\{X_k=+1\}}\,. \qquad (3.12)
\end{aligned}$$

Combining (3.4) and (3.10) finally yields that for every $k \in \mathbb{N}$

$$\begin{aligned}
&D_kf(\mathbf{F})\\
&= \sum_{i=1}^{d}\frac{\partial}{\partial x_i}f(\mathbf{F})D_kF_i - \frac{X_k}{4\sqrt{p_k q_k}}\sum_{i,j=1}^{d}\Big(\frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_i\partial x_j}f(\mathbf{F}_k^-)\Big)(D_kF_i)(D_kF_j)\\
&\quad + R_1(\mathbf{F},\mathbf{F}_k^+) - R_2(\mathbf{F},\mathbf{F}_k^-) - R_3(\mathbf{F},\mathbf{F}_k^+) - R_4(\mathbf{F},\mathbf{F}_k^-)\,,
\end{aligned}$$

where because of (3.7), (3.8), (3.11) and (3.12) we have that

$$\begin{aligned}
&|R_1(\mathbf{F},\mathbf{F}_k^+)| + |R_2(\mathbf{F},\mathbf{F}_k^-)| + |R_3(\mathbf{F},\mathbf{F}_k^+)| + |R_4(\mathbf{F},\mathbf{F}_k^-)|\\
&\quad \le \frac{5}{12p_k q_k}\sum_{i,j,\ell=1}^{d}\Big\|\frac{\partial^3}{\partial x_i\partial x_j\partial x_\ell}f\Big\|_\infty |(D_kF_i)(D_kF_j)(D_kF_\ell)|\,.
\end{aligned}$$

The proof is thus complete by taking $R_k(\mathbf{F}) := R_1(\mathbf{F},\mathbf{F}_k^+) - R_2(\mathbf{F},\mathbf{F}_k^-) - R_3(\mathbf{F},\mathbf{F}_k^+) - R_4(\mathbf{F},\mathbf{F}_k^-)$, for every $k \in \mathbb{N}$. $\square$

Let us now turn to the already announced multivariate approximate integration by parts formula. The next result not only generalizes Lemma 5.2 in [10] to the case in which the underlying Rademacher sequence is allowed to be non-symmetric and non-homogeneous, but also improves the constants in the bound for the remainder term. We emphasize that Lemma 3.5 is the first instance where the additional boundary term discussed in Remark 3.2 shows up.

Lemma 3.5. *Let $\mathbf{F}$ be a vector of Rademacher functionals as in Theorem 3.1. Furthermore, let $f : \mathbb{R}^d \to \mathbb{R}$ be a thrice partially differentiable function with bounded partial derivatives. Then, for every $i = 1, \dots, d$,*

$$\mathbb{E}[F_i f(\mathbf{F})] = \sum_{j=1}^{d} \mathbb{E}\Big[\frac{\partial}{\partial x_j} f(\mathbf{F}) \langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big] + \mathbb{E}[\langle R(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]$$

with a remainder $R(\mathbf{F})$ that satisfies the estimate

$$\begin{aligned} &|\mathbb{E}[\langle R(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]| \\ &\qquad \le \frac{1}{2} M_2(f) \mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}} \Big(\sum_{j=1}^{d} |DF_j|\Big)^2, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big] \\ &\qquad\quad + \frac{5}{12} M_3(f) \mathbb{E}\Big[\Big\langle \frac{1}{pq} \Big(\sum_{j=1}^{d} |DF_j|\Big)^3, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]. \end{aligned} \tag{3.13}$$

Proof. Fix $i = 1, \dots, d$. By the integration by parts formula (2.6) we have that

$$\mathbb{E}[F_i f(\mathbf{F})] = \mathbb{E}[\langle Df(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]. \tag{3.14}$$

Here, we implicitly used the fact that $f(\mathbf{F}) \in \mathrm{dom}(D)$, which can be verified as follows. At first, by the mean value theorem it holds that for every $k \in \mathbb{N}$

$$\begin{aligned} |D_k f(\mathbf{F})| &= \sqrt{p_k q_k}\, |f(\mathbf{F}_k^+) - f(\mathbf{F}_k^-)| = \sqrt{p_k q_k}\, \Big| \sum_{i=1}^{d} \frac{\partial}{\partial x_i} f(\mathbf{F}_k^- + \theta(\mathbf{F}_k^+ - \mathbf{F}_k^-))((F_i)_k^+ - (F_i)_k^-) \Big| \\ &\le \sqrt{p_k q_k}\, M_1(f) \sum_{i=1}^{d} |(F_i)_k^+ - (F_i)_k^-| = M_1(f) \sum_{i=1}^{d} |D_k F_i|, \end{aligned}$$

where $\theta \in (0,1)$. Thus, an application of the Cauchy-Schwarz inequality yields that

$$\begin{aligned} \mathbb{E}[\|Df(\mathbf{F})\|^2_{\ell^2(\mathbb{N})}] &= \mathbb{E}\Big[\sum_{k=1}^{\infty} (D_k f(\mathbf{F}))^2\Big] \\ &\le (M_1(f))^2\, \mathbb{E}\Big[\sum_{k=1}^{\infty} \Big(\sum_{i=1}^{d} |D_k F_i|\Big)^2\Big] \\ &\le d (M_1(f))^2\, \mathbb{E}\Big[\sum_{k=1}^{\infty} \sum_{i=1}^{d} (D_k F_i)^2\Big] \\ &= d (M_1(f))^2 \sum_{i=1}^{d} \mathbb{E}[\|DF_i\|^2_{\ell^2(\mathbb{N})}] \end{aligned} \tag{3.15}$$

and finiteness of the right hand side in (3.15) follows from the assumptions that, for every $i = 1, \dots, d$, $\frac{\partial}{\partial x_i} f$ is bounded and $F_i \in \mathrm{dom}(D)$. Now, by plugging (3.1) into (3.14) we immediately get

$$
\begin{aligned}
&\mathbb{E}[F_i f(\mathbf{F})]\\
&= \sum_{j=1}^d \mathbb{E}\Big[\frac{\partial}{\partial x_j} f(\mathbf{F})\langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big]\\
&\quad - \sum_{j,\ell=1}^d \mathbb{E}\Big[\Big\langle \frac{X}{4\sqrt{pq}}\Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^-)\Big)(DF_j)(DF_\ell), -DL^{-1}F_i\Big\rangle_{\ell^2(\mathbb{N})}\Big]\\
&\quad + \mathbb{E}[\langle R^{(1)}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}] \qquad (3.16)
\end{aligned}
$$

with $\mathbf{F}^+ := (\mathbf{F}_k^+)_{k\in\mathbb{N}}$ and $\mathbf{F}^- := (\mathbf{F}_k^-)_{k\in\mathbb{N}}$ as well as a remainder $R^{(1)}(\mathbf{F}) := (R_k(\mathbf{F}))_{k\in\mathbb{N}}$ which by (3.2) fulfills the estimate

$$
|\mathbb{E}[\langle R^{(1)}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]| \le \frac{5}{12} M_3(f)\mathbb{E}\Big[\Big\langle \frac{1}{pq}\Big(\sum_{j=1}^d |DF_j|\Big)^3, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]. \quad (3.17)
$$

As a consequence, we only need to further bound the second term in (3.16). By virtue of the Cauchy-Schwarz inequality and (2.8) we see that, for every $j, \ell \in \mathbb{N}$,

$$
\begin{aligned}
&\mathbb{E}\Big[\sum_{k=1}^\infty \Big|\frac{X_k}{4\sqrt{p_k q_k}}\Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^-)\Big)(D_kF_j)(D_kF_\ell)(-D_kL^{-1}F_i)\Big|\Big]\\
&\le \frac{1}{2} M_2(f)\mathbb{E}\Big[\sum_{k=1}^\infty \frac{1}{\sqrt{p_k q_k}} |(D_kF_j)(D_kF_\ell)(-D_kL^{-1}F_i)|\Big]\\
&\le \frac{1}{2} M_2(f)\Big(\mathbb{E}\Big[\sum_{k=1}^\infty \frac{1}{p_k q_k}(D_kF_j)^2(D_kF_\ell)^2\Big]\Big)^{1/2}\Big(\mathbb{E}\Big[\sum_{k=1}^\infty (D_kL^{-1}F_i)^2\Big]\Big)^{1/2}\\
&\le \frac{1}{2} M_2(f)\Big(\mathbb{E}\Big[\sum_{k=1}^\infty \frac{1}{p_k q_k}(D_kF_j)^4\Big]\Big)^{1/4}\Big(\mathbb{E}\Big[\sum_{k=1}^\infty \frac{1}{p_k q_k}(D_kF_\ell)^4\Big]\Big)^{1/4}\Big(\mathbb{E}\Big[\sum_{k=1}^\infty (D_kF_i)^2\Big]\Big)^{1/2}\\
&= \frac{1}{2} M_2(f)(\mathbb{E}[\|(pq)^{-1/4}DF_j\|_{\ell^4(\mathbb{N})}^4])^{1/4}(\mathbb{E}[\|(pq)^{-1/4}DF_\ell\|_{\ell^4(\mathbb{N})}^4])^{1/4}(\mathbb{E}[\|DF_i\|_{\ell^2(\mathbb{N})}^2])^{1/2}
\end{aligned}
$$

and finiteness of this expression follows from the assumptions that $F_i \in \mathrm{dom}(D)$ and $\|(pq)^{-1/4}DF_i\|_{\ell^2(\mathbb{N})} < \infty$ for every $i = 1, \dots, d$. Thus, an exchange of expectation and summation is valid due to the Fubini-Tonelli theorem, and the independence of X_k and $(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^-))(D_kF_j)(D_kF_\ell)(-D_kL^{-1}F_i)$, for every $k \in \mathbb{N}$, yields that, for every $j, \ell \in \mathbb{N}$,

$$
\begin{aligned}
&\mathbb{E}\Big[\Big\langle \frac{X}{4\sqrt{pq}}\Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^-)\Big)(DF_j)(DF_\ell), -DL^{-1}F_i\Big\rangle_{\ell^2(\mathbb{N})}\Big]\\
&= \sum_{k=1}^\infty \frac{p_k - q_k}{4\sqrt{p_k q_k}}\mathbb{E}\Big[\Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}_k^-)\Big)(D_kF_j)(D_kF_\ell)(-D_kL^{-1}F_i)\Big]\\
&= \mathbb{E}\Big[\Big\langle \frac{p-q}{4\sqrt{pq}}\Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^-)\Big)(DF_j)(DF_\ell), -DL^{-1}F_i\Big\rangle_{\ell^2(\mathbb{N})}\Big]. \qquad (3.18)
\end{aligned}
$$

By plugging (3.18) into (3.16) we then get

$$
\begin{aligned}
\mathbb{E}[F_i f(\mathbf{F})] = \sum_{j=1}^d \mathbb{E}\Big[\frac{\partial}{\partial x_j} f(\mathbf{F})\langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big] + \mathbb{E}[\langle R^{(1)}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]\\
- \mathbb{E}[\langle R^{(2)}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}] \qquad (3.19)
\end{aligned}
$$

with a remainder term

$$R^{(2)}(\mathbf{F}) := \frac{p-q}{4\sqrt{pq}} \sum_{j,\ell=1}^{\infty} \Big(\frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^+) + \frac{\partial^2}{\partial x_j \partial x_\ell} f(\mathbf{F}^-)\Big)(DF_j)(DF_\ell)$$

satisfying

$$\begin{aligned} &|\mathbb{E}[\langle R^{(2)}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]| \\ &\qquad \le \frac{1}{2} M_2(f) \mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}} \Big(\sum_{j=1}^{d} |DF_j|\Big)^2, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]\,. \end{aligned} \tag{3.20}$$

Finally, the assertion follows from (3.19) upon putting $R(\mathbf{F}) := R^{(1)}(\mathbf{F}) - R^{(2)}(\mathbf{F})$ and using the bounds in (3.17) and (3.20). □

Remark 3.6. In the symmetric case where the underlying Rademacher sequence satisfies $p_k = q_k = 1/2$ for every $k \in \mathbb{N}$, the bound for the remainder term in (3.13) simplifies to

$$|\mathbb{E}[\langle R(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]| \le \frac{5}{3} M_3(f) \mathbb{E}\Big[\Big\langle \Big(\sum_{j=1}^{d} |DF_j|\Big)^3, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]\,,$$

since $p_k - q_k = 0$, for every $k \in \mathbb{N}$.

With both integration by parts formulae at hand we can now turn to the proof of Theorem 3.1. We will use an interpolation technique that is known as the 'smart path method' in the literature, cf. Section 2.4 in [24]. While the use of this interpolation technique within Stein's method goes back to [3], it has also already found applications within the Malliavin-Stein method for multivariate normal approximation in the framework considering non-linear functionals of Gaussian [14] and Poisson random measures [16]. We follow the lines of the proof of Theorem 4.2 in [16] until we have to use our discrete integration by parts formula developed in Lemma 3.5. Without loss of generality, we can and will from now on assume that $\mathbf{F}$ and $\mathbf{N}$ are independent.

Proof of Theorem 3.1. Let $g : \mathbb{R}^d \to \mathbb{R}$ be a four times partially differentiable function with bounded partial derivatives satisfying $M_1(g), M_2(g), M_3(g), M_4(g) \le 1$. Consider the function $\Psi : \mathbb{R} \to \mathbb{R}$ given by

$$\Psi(t) := \mathbb{E}[g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})]\,, \qquad t \in [0,1]\,. \tag{3.21}$$

Then, by the mean value theorem we have that

$$|\mathbb{E}[g(\mathbf{F})] - \mathbb{E}[g(\mathbf{N})]| = |\Psi(0) - \Psi(1)| \le \sup_{t\in(0,1)} |\Psi'(t)|\,, \tag{3.22}$$

where, for every $t \in (0,1)$, Ψ' is given by

$$\begin{aligned} \Psi'(t) &= \mathbb{E}\Big[\frac{d}{dt} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big] = \sum_{i=1}^{d} \mathbb{E}\Big[\frac{\partial}{\partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big(\frac{1}{2\sqrt{t}} N_i - \frac{1}{2\sqrt{1-t}} F_i\Big)\Big] \\ &= \frac{1}{2\sqrt{t}} A_t - \frac{1}{2\sqrt{1-t}} B_t \end{aligned} \tag{3.23}$$

with

$$A_t := \sum_{i=1}^{d} \mathbb{E}\Big[\frac{\partial}{\partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N}) N_i\Big] \quad \text{and} \quad B_t := \sum_{i=1}^{d} \mathbb{E}\Big[\frac{\partial}{\partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N}) F_i\Big]\,.$$

Now, by independence of $\mathbf{F}$ and $\mathbf{N}$ as well as by Fubini's theorem we have that, for every $t \in (0,1)$,

$$A_t = \sum_{i=1}^{d} \mathbb{E}_{\mathbf{N}}\Big[\mathbb{E}_{\mathbf{F}}\Big[\frac{\partial}{\partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big] N_i\Big]\,, \tag{3.24}$$

where $\mathbb{E}_{\mathbf{F}}$ and $\mathbb{E}_{\mathbf{N}}$ denote the expectations with respect to the distributions of $\mathbf{F}$ and $\mathbf{N}$, respectively. By the integration by parts formula in Lemma 3.3 we deduce that, for every $t \in (0,1)$,

$$A_t = \sqrt{t} \sum_{i,j=1}^{d} \Sigma_{ij} \mathbb{E}_{\mathbf{N}}\Big[\mathbb{E}_{\mathbf{F}}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big]\Big]\,, \tag{3.25}$$

where the exchange of differentiation and expectation in the last step of (3.25) is valid since $\frac{\partial}{\partial x_i} g$ is bounded. Again, using the independence of $\mathbf{F}$ and $\mathbf{N}$ together with Fubini's theorem yields that, for every $t \in (0,1)$,

$$A_t = \sqrt{t} \sum_{i,j=1}^{d} \Sigma_{ij} \mathbb{E}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big]\,. \tag{3.26}$$

Similarly as above, we deduce for the quantity B_t that, for every $t \in (0,1)$,

$$B_t = \sum_{i=1}^{d} \mathbb{E}_{\mathbf{F}}\Big[\mathbb{E}_{\mathbf{N}}\Big[\frac{\partial}{\partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big] F_i\Big]\,.$$

Using the integration by parts formula in Lemma 3.5, we see that, for every $t \in (0,1)$,

$$\begin{aligned} B_t = \sqrt{1-t} \sum_{i,j=1}^{d} &\mathbb{E}_{\mathbf{F}}\Big[\mathbb{E}_{\mathbf{N}}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N})\Big] \langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big] \\ &+ \sum_{i=1}^{d} \mathbb{E}_{\mathbf{F}}[\langle R_{i,t}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]\,, \end{aligned} \tag{3.27}$$

where we recall that the exchange of differentiation and expectation in the last step is valid since $\frac{\partial}{\partial x_i} g$ is bounded for every $i = 1, \ldots, d$, and $R_{i,t}(\mathbf{F})$ is a remainder which by (3.13) fulfills that for every $i = 1, \ldots, d$ and $t \in (0,1)$,

$$\begin{aligned} &|\mathbb{E}[\langle R_{i,t}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell(\mathbb{N})}]| \\ &\quad\le \frac{1}{2}(1-t) M_3(g) \mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}} \Big(\sum_{j=1}^{d} |DF_j|\Big)^2, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big] \\ &\qquad + \frac{5}{12}(1-t)^{3/2} M_4(g) \mathbb{E}\Big[\Big\langle \frac{1}{pq} \Big(\sum_{j=1}^{d} |DF_j|\Big)^3, |-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big]\,. \end{aligned} \tag{3.28}$$

Going back to (3.27), another application of the independence of $\mathbf{F}$ and $\mathbf{N}$ together with Fubini's theorem yields that for every $t \in (0,1)$,

$$\begin{aligned} B_t = \sqrt{1-t} \sum_{i,j=1}^{d} &\mathbb{E}_{\mathbf{F}}\Big[\mathbb{E}_{\mathbf{N}}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F} + \sqrt{t}\,\mathbf{N}) \langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big]\Big] \\ &+ \sum_{i=1}^{d} \mathbb{E}_{\mathbf{F}}[\mathbb{E}_{\mathbf{N}}[\langle R_{i,t}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]] \end{aligned}$$

$$= \sqrt{1-t}\sum_{i,j=1}^{d}\mathbb{E}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F}+\sqrt{t}\,\mathbf{N})\langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}\Big] + \sum_{i=1}^{d}\mathbb{E}[\langle R_{i,t}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]\,. \tag{3.29}$$

Hence, by combining (3.26) and (3.29) with (3.23) we deduce that for every $t \in (0,1)$,

$$\Psi'(t) = \frac{1}{2}\sum_{i,j=1}^{d}\mathbb{E}\Big[\frac{\partial^2}{\partial x_j \partial x_i} g(\sqrt{1-t}\,\mathbf{F}+\sqrt{t}\,\mathbf{N})(\Sigma_{ij} - \langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})})\Big] - \frac{1}{2\sqrt{1-t}}\sum_{i=1}^{d}\mathbb{E}[\langle R_{i,t}(\mathbf{F}), -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}]\,,$$

and by using (3.28) as well as the fact that $M_2(g), M_3(g), M_4(g) \le 1$ we thus conclude that, for every $t \in (0,1)$,

$$\begin{aligned}|\Psi'(t)| \le \frac{1}{2}\sum_{i,j=1}^{d}\mathbb{E}[|\Sigma_{ij} - \langle DF_j, -DL^{-1}F_i\rangle_{\ell^2(\mathbb{N})}|] &+ \frac{1}{4}\sqrt{1-t}\mathbb{E}\Big[\Big\langle \frac{|p-q|}{\sqrt{pq}}\Big(\sum_{j=1}^{d}|DF_j|\Big)^2, \sum_{i=1}^{d}|-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big] \\ &+ \frac{5}{24}(1-t)\mathbb{E}\Big[\Big\langle \frac{1}{pq}\Big(\sum_{j=1}^{d}|DF_j|\Big)^3, \sum_{i=1}^{d}|-DL^{-1}F_i|\Big\rangle_{\ell^2(\mathbb{N})}\Big].\end{aligned}$$

Plugging this into (3.22) where we take the supremum over all $t \in (0,1)$ completes the argument. □

3.2 A multivariate discrete second-order Poincaré inequality

In this section we use Theorem 3.1 to develop a discrete second-order Poincaré inequality for the normal approximation of vectors of Rademacher functionals. In comparison with Theorem 3.1 it has the advantage that it expresses the bound for $d_4(\mathbf{F},\mathbf{N})$ only in terms of discrete first- and second-order Malliavin derivatives and does not involve the operator L^{-1}. This in turn allows to apply the bound without specifying the chaos decomposition of the component random variables of the random vector $\mathbf{F}$. Our result can be seen as the natural multivariate extension of the main result from [11], where a univariate discrete second-order Poincaré inequality has been obtained for the Kolmogorov distance (see also Remark 3.2 [26] for a closely related bound for the Wasserstein distance).

Theorem 3.7. *Let the conditions of Theorem 3.1 prevail and assume additionally that* $F_i \in \mathrm{dom}(D^2)$ *for all* $i = 1,\dots,d$*. For* $i,j = 1,\dots,d$ *define*

$$B_1(i,j) := \Big(\frac{15}{4}\sum_{k,\ell,m=1}^{\infty}(\mathbb{E}[(D_kF_i)^2(D_\ell F_i)^2])^{1/2}(\mathbb{E}[(D_mD_kF_j)^2(D_mD_\ell F_j)^2])^{1/2}\Big)^{1/2},$$

$$B_2(i,j) := \Big(\frac{3}{4}\sum_{k,\ell,m=1}^{\infty}\frac{1}{p_mq_m}(\mathbb{E}[(D_mD_kF_i)^2(D_mD_\ell F_i)^2])^{1/2} \times (\mathbb{E}[(D_mD_kF_j)^2(D_mD_\ell F_j)^2])^{1/2}\Big)^{1/2},$$

$$B_3(i,j) := \frac{1}{2}d^{3/2}\sum_{k=1}^{\infty}\frac{|p_k-q_k|}{\sqrt{p_kq_k}}(\mathbb{E}[(D_kF_i)^2])^{1/2}(\mathbb{E}[(D_kF_j)^4])^{1/2}\,,$$

$$B_4(i,j) := \frac{5}{12} d^2 \sum_{k=1}^{\infty} \frac{1}{p_k q_k} (\mathbb{E}[(D_k F_i)^4])^{1/4} (\mathbb{E}[(D_k F_j)^4])^{3/4} .$$

Then,

$$d_4(\mathbf{F}, \mathbf{N}) \le \frac{1}{2} \sum_{i,j=1}^{d} \left[|\Sigma_{ij} - \mathrm{cov}(F_i, F_j)| + B_1(i,j) + B_2(i,j) + B_3(i,j) + B_4(i,j) \right] .$$

Proof. Let A_1, A_2 and A_3 be the three terms defined in Theorem 3.1. We start with A_1. An application of the triangle inequality yields

$$\begin{aligned} &\mathbb{E}[|\Sigma_{ij} - \langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}|] \\ &\quad \le |\Sigma_{ij} - \mathrm{cov}(F_i, F_j)| + \mathbb{E}[|\mathrm{cov}(F_i, F_j) - \langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}|]. \end{aligned} \tag{3.30}$$

Let us further consider the second summand on the right hand side of (3.30). By the integration by parts formula in (2.6) we see that

$$\mathrm{cov}(F_i, F_j) = \mathbb{E}[\langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}],$$

and thus, by the Cauchy-Schwarz inequality and the Poincaré inequality in (2.9) we have that

$$\begin{aligned} &\mathbb{E}[|\mathrm{cov}(F_i, F_j) - \langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}|] \\ &\le (\mathrm{Var}(\langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}))^{1/2} \\ &\le (\mathbb{E}[\|D(\langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})})\|^2_{\ell^2(\mathbb{N})}])^{1/2} \\ &= \Big(\mathbb{E}\Big[\sum_{\ell=1}^{\infty} \Big(\sum_{k=1}^{\infty} D_\ell((D_k F_j)(-D_k L^{-1} F_i)) \Big)^2 \Big] \Big)^{1/2}. \end{aligned} \tag{3.31}$$

By the product formula for the discrete Malliavin derivative in (2.3) and the triangle inequality we get that, for every $k, \ell \in \mathbb{N}$,

$$\begin{aligned} &|D_\ell((D_k F_j)(-D_k L^{-1} F_i))| \\ &\le |(D_\ell D_k F_j)(-D_k L^{-1} F_i)| + |(D_k F_j)(-D_\ell D_k L^{-1} F_i)| + \frac{1}{\sqrt{p_\ell q_\ell}} |(D_\ell D_k F_j)(-D_\ell D_k L^{-1} F_i)| \, . \end{aligned}$$

Plugging this into (3.31) and using the Cauchy-Schwarz inequality then yields

$$\mathbb{E}[|\mathrm{cov}(F_i, F_j) - \langle DF_j, -DL^{-1}F_i \rangle_{\ell^2(\mathbb{N})}|] \le \big(3\big[T_1(i,j) + T_2(i,j) + T_3(i,j) \big] \big)^{1/2} \tag{3.32}$$

with

$$\begin{aligned} T_1(i,j) &:= \mathbb{E}\Big[\sum_{\ell=1}^{\infty} \Big(\sum_{k=1}^{\infty} |(D_\ell D_k F_j)(-D_k L^{-1} F_i)| \Big)^2 \Big] , \\ T_2(i,j) &:= \mathbb{E}\Big[\sum_{\ell=1}^{\infty} \Big(\sum_{k=1}^{\infty} |(D_k F_j)(-D_\ell D_k L^{-1} F_i)| \Big)^2 \Big] , \\ T_3(i,j) &:= \mathbb{E}\Big[\sum_{\ell=1}^{\infty} \frac{1}{p_\ell q_\ell} \Big(\sum_{k=1}^{\infty} |(D_\ell D_k F_j)(-D_\ell D_k L^{-1} F_i)| \Big)^2 \Big] . \end{aligned}$$

Each of these quantities will now be further bounded from above. Considering T_1, an application of (2.7) and (2.5) as well as the triangle inequality yields that, for every $\ell \in \mathbb{N}$,

$$\Big(\sum_{k=1}^\infty |(D_\ell D_k F_j)(-D_k L^{-1}F_i)|\Big)^2 = \Big(\sum_{k=1}^\infty \Big|(D_\ell D_k F_j)\int_0^\infty e^{-t}P_t D_k F_i\,dt\Big|\Big)^2$$
$$= \Big(\sum_{k=1}^\infty \Big|(D_\ell D_k F_j)\int_0^\infty e^{-t}\mathbb{E}[D_k F_i(X^t)\,|\,X]\,dt\Big|\Big)^2$$
$$\le \Big(\sum_{k=1}^\infty |D_\ell D_k F_j|\int_0^\infty e^{-t}\mathbb{E}[|D_k F_i(X^t)|\,|\,X]\,dt\Big)^2.$$

Furthermore, by virtue of the monotone convergence theorem we get that, for every $\ell \in \mathbb{N}$,

$$\Big(\sum_{k=1}^\infty |D_\ell D_k F_j|\int_0^\infty e^{-t}\mathbb{E}[|D_k F_i(X^t)|\,|\,X]\,dt\Big)^2$$
$$= \Big(\int_0^\infty e^{-t}\sum_{k=1}^\infty |D_\ell D_k F_j|\mathbb{E}[|D_k F_i(X^t)|\,|\,X]\,dt\Big)^2$$
$$= \Big(\int_0^\infty e^{-t}\mathbb{E}\Big[\sum_{k=1}^\infty |(D_\ell D_k F_j)(D_k F_i(X^t))|\,\Big|\,X\Big]\,dt\Big)^2.$$

By using Jensen's inequality as well as the Cauchy-Schwarz inequality we then conclude that, for every $\ell \in \mathbb{N}$,

$$\Big(\int_0^\infty e^{-t}\mathbb{E}\Big[\sum_{k=1}^\infty |(D_\ell D_k F_j)(D_k F_i(X^t))|\,\Big|\,X\Big]\,dt\Big)^2$$
$$\le \int_0^\infty e^{-t}\mathbb{E}\Big[\Big(\sum_{k=1}^\infty |(D_\ell D_k F_j)(D_k F_i(X^t))|\Big)^2\,\Big|\,X\Big]\,dt$$
$$= \int_0^\infty e^{-t}\mathbb{E}\Big[\sum_{m,k=1}^\infty |(D_\ell D_m F_j)(D_m F_i(X^t))(D_\ell D_k F_j)(D_k F_i(X^t))|\,\Big|\,X\Big]\,dt$$
$$= \sum_{m,k=1}^\infty |(D_\ell D_m F_j)(D_\ell D_k F_j)|\int_0^\infty e^{-t}\mathbb{E}[|(D_m F_i(X^t))(D_k F_i(X^t))|\,|\,X]\,dt$$
$$\le \sum_{m,k=1}^\infty |(D_\ell D_m F_j)(D_\ell D_k F_j)|\int_0^\infty e^{-t}(\mathbb{E}[(D_m F_i(X^t))^2(D_k F_i(X^t))^2\,|\,X])^{1/2}\,dt$$
$$\le \sum_{m,k=1}^\infty |(D_\ell D_m F_j)(D_\ell D_k F_j)|\Big(\int_0^\infty e^{-t}\mathbb{E}[(D_m F_i(X^t))^2(D_k F_i(X^t))^2\,|\,X]\,dt\Big)^{1/2}.$$

Thus, another application of the Cauchy-Schwarz inequality leads to the bound

$$T_1(i,j)$$
$$\le \mathbb{E}\Big[\sum_{m,k,\ell=1}^\infty |(D_\ell D_m F_j)(D_\ell D_k F_j)|\Big(\int_0^\infty e^{-t}\mathbb{E}[(D_m F_i(X^t))^2(D_k F_i(X^t))^2\,|\,X]\,dt\Big)^{1/2}\Big]$$
$$\le \sum_{m,k,\ell=1}^\infty (\mathbb{E}[(D_\ell D_m F_j)^2(D_\ell D_k F_j)^2])^{1/2}$$
$$\times\Big(\mathbb{E}\Big[\int_0^\infty e^{-t}\mathbb{E}[(D_m F_i(X^t))^2(D_k F_i(X^t))^2\,|\,X]\,dt\Big]\Big)^{1/2}$$

$$= \sum_{m,k,\ell=1}^{\infty} (\mathbb{E}[(D_\ell D_m F_j)^2 (D_\ell D_k F_j)^2])^{1/2} \Big(\int_0^\infty e^{-t} \mathbb{E}[(D_m F_i)^2 (D_k F_i)^2] \, dt \Big)^{1/2}$$
$$= \sum_{m,k,\ell=1}^{\infty} (\mathbb{E}[(D_\ell D_m F_j)^2 (D_\ell D_k F_j)^2])^{1/2} (\mathbb{E}[(D_m F_i)^2 (D_k F_i)^2])^{1/2} .$$

The quantities T_2 and T_3 can be treated in the same manner as T_1, and thus, it holds that

$$T_2(i,j) \le \frac{1}{4} \sum_{m,k,\ell=1}^{\infty} (\mathbb{E}[(D_\ell D_m F_i)^2 (D_\ell D_k F_i)^2])^{1/2} (\mathbb{E}[(D_m F_j)^2 (D_k F_j)^2])^{1/2},$$
$$T_3(i,j) \le \frac{1}{4} \sum_{m,k,\ell=1}^{\infty} \frac{1}{p_\ell q_\ell} (\mathbb{E}[(D_\ell D_m F_i)^2 (D_\ell D_k F_i)^2])^{1/2} (\mathbb{E}[(D_\ell D_m F_j)^2 (D_\ell D_k F_j)^2])^{1/2} .$$

Therefore, combining the bounds for T_1, T_2 and T_3 with (3.32) and (3.30) yields

$$A_1 \le \frac{1}{2} \sum_{i,j=1}^{d} \big[|\Sigma_{ij} - \mathrm{cov}(F_i, F_j)| + B_1(i,j) + B_2(i,j) \big] .$$

Turning to the term A_2, by several applications of the Cauchy-Schwarz inequality and due to (2.8) we see that

$$A_2 = \frac{1}{4} \sum_{k=1}^{\infty} \frac{|p_k - q_k|}{\sqrt{p_k q_k}} \sum_{i=1}^{d} \mathbb{E}\Big[\Big(\sum_{j=1}^{d} |D_k F_j| \Big)^2 |-D_k L^{-1} F_i| \Big]$$
$$\le \frac{1}{4} \sum_{k=1}^{\infty} \frac{|p_k - q_k|}{\sqrt{p_k q_k}} \Big(\mathbb{E}\Big[\Big(\sum_{j=1}^{d} |D_k F_j| \Big)^4 \Big] \Big)^{1/2} \sum_{i=1}^{d} (\mathbb{E}[(D_k F_i)^2])^{1/2}$$
$$\le \frac{1}{4} d^{3/2} \sum_{i,j=1}^{d} \sum_{k=1}^{\infty} \frac{|p_k - q_k|}{\sqrt{p_k q_k}} (\mathbb{E}[(D_k F_j)^4])^{1/2} (\mathbb{E}[(D_k F_i)^2])^{1/2} .$$

For the third and last term A_3 we similarly see that, by using Hölder's inequality with Hölder conjugates 4 and $4/3$ as well as (2.8),

$$A_3 = \frac{5}{24} \sum_{k=1}^{\infty} \frac{1}{p_k q_k} \sum_{i=1}^{d} \mathbb{E}\Big[\Big(\sum_{j=1}^{d} |D_k F_j| \Big)^3 |-D_k L^{-1} F_i| \Big]$$
$$\le \frac{5}{24} d^2 \sum_{i,j=1}^{d} \sum_{k=1}^{\infty} \frac{1}{p_k q_k} \mathbb{E}[|D_k F_j|^3 \cdot |-D_k L^{-1} F_i|]$$
$$\le \frac{5}{24} d^2 \sum_{i,j=1}^{d} \sum_{k=1}^{\infty} \frac{1}{p_k q_k} (\mathbb{E}[(D_k F_j)^4])^{3/4} (\mathbb{E}[(D_k L^{-1} F_i)^4])^{1/4}$$
$$\le \frac{5}{24} d^2 \sum_{i,j=1}^{d} \sum_{k=1}^{\infty} \frac{1}{p_k q_k} (\mathbb{E}[(D_k F_j)^4])^{3/4} (\mathbb{E}[(D_k F_i)^4])^{1/4} .$$

This completes the argument. □

4 Applications to random graphs and random cubical complexes

4.1 Subgraph and degree counts in the Erdős-Rényi random graph

In this section we consider an application of the discrete second-order Poincaré inequality developed above to subgraph counts in the Erdős-Rényi random graph. To describe the model, let K_n be the complete graph on $n \in \mathbb{N}$ vertices and fix $p \in (0,1)$. In what follows we implicitly assume that n is sufficiently large (for instance, we assume $n \geq 2$ when we consider subgraphs with at least one edge, or we assume $n \geq 6$ in the proof of Theorem 4.2 where we make a case distinction that is illustrated in Figure 1). We number the $\binom{n}{2}$ edges of K_n in a fixed but arbitrary way and denote them by $e_1, \dots, e_{\binom{n}{2}}$. Now, to each edge e_k of K_n a Rademacher random variable X_k with success probability p is assigned and we remove e_k from K_n if $X_k = -1$ and keep e_k otherwise. This gives rise to the Erdős-Rényi random graph denoted by $\mathcal{G}(n,p)$, which has n vertices and a binomially distributed number of edges with parameters $\binom{n}{2}$ and p. In what follows we assume p to be independent of n.

Let Γ be a fixed (finite, simple) graph and denote by X_Γ the number of subgraphs of $\mathcal{G}(n,p)$ that are isomorphic to Γ (we assume here that all graphs we consider have at least one edge). To represent this counting statistic formally, we denote by v_Γ the number of vertices and by e_Γ the number of edges of Γ. Moreover, we shall denote by $\mathrm{aut}(\Gamma)$ the (finite) group of graph-automorphisms of Γ and by $|\mathrm{aut}(\Gamma)|$ its cardinality. Using this notation, X_Γ may be written as

$$X_\Gamma = \sum_{\Gamma'} \mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}\,, \tag{4.1}$$

where the sum is running over all $\binom{n}{v_\Gamma}\frac{v_\Gamma!}{|\mathrm{aut}(\Gamma)|}$ copies Γ' of Γ in K_n and where $\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}$ is the indicator function of the event that Γ' is a subgraph of $\mathcal{G}(n,p)$. Since $\mathbb{E}[\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}] = P(\Gamma' \subset \mathcal{G}(n,p)) = p^{e_\Gamma}$, it readily follows that

$$\mathbb{E}[X_\Gamma] = \binom{n}{v_\Gamma}\frac{v_\Gamma!}{|\mathrm{aut}(\Gamma)|}p^{e_\Gamma}\,.$$

To proceed, we also need information about the covariance between X_Γ and X_Φ for two graphs Γ and Φ. Before we state the result, let us introduce our asymptotic notation. We shall write $a_n = O(b_n)$ for two sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ if $\limsup_{n\to\infty}|a_n/b_n| < \infty$. Moreover, $a_n \asymp b_n$ will indicate that $|a_n/b_n| \to 1$, as $n \to \infty$.

Lemma 4.1. *Let Γ and Φ be two graphs and define X_Γ and X_Φ as above. Then,*

$$\begin{aligned}\mathrm{cov}(X_\Gamma, X_\Phi) \asymp 2\,&\frac{n^{v_\Gamma+v_\Phi-2}}{|\mathrm{aut}(\Gamma)||\mathrm{aut}(\Phi)|}\,e_\Gamma e_\Phi\,p^{e_\Gamma+e_\Phi-1}(1-p)\\&+c(\Gamma,\Phi)\,n^{v_\Gamma+v_\Phi-3}\,p^{e_\Gamma+e_\Phi-2}(1-p)+O(n^{v_\Gamma+v_\Phi-3})\end{aligned}$$

with a constant $c(\Gamma,\Phi) > 0$ only depending on Γ and Φ.

Proof. Recalling (4.1), we see that

$$\mathrm{cov}(X_\Gamma, X_\Phi) = \sum_{\Gamma',\Phi'}\mathrm{cov}(\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}, \mathbf{1}\{\Phi' \subset \mathcal{G}(n,p)\})\,.$$

By the independence properties of the construction of $\mathcal{G}(n,p)$ we have that $\mathrm{cov}(\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}, \mathbf{1}\{\Phi' \subset \mathcal{G}(n,p)\}) \neq 0$ if and only if Γ' and Φ' have at least one common edge. In what follows we shall write $e_{\Gamma',\Phi'}$ for the number of edges that Γ' and Φ' have in common. Thus,

$$\begin{aligned}&\mathrm{cov}(X_\Gamma, X_\Phi)\\&= \sum_{\Gamma',\Phi':e_{\Gamma',\Phi'}\geq 1}\mathrm{cov}(\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}, \mathbf{1}\{\Phi' \subset \mathcal{G}(n,p)\})\\&= \sum_{\Gamma',\Phi':e_{\Gamma',\Phi'}\geq 1}\big(\mathbb{E}[\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p), \Phi' \subset \mathcal{G}(n,p)\}] - \mathbb{E}[\mathbf{1}\{\Gamma' \subset \mathcal{G}(n,p)\}]\mathbb{E}[\mathbf{1}\{\Phi' \subset \mathcal{G}(n,p)\}]\big)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{\Gamma',\Phi':e_{\Gamma',\Phi'}\geq 1} \left(p^{e_\Gamma+e_\Phi-e_{\Gamma',\Phi'}} - p^{e_\Gamma}p^{e_\Phi}\right) \\
&= \sum_{\Gamma',\Phi':e_{\Gamma',\Phi'}\geq 1} p^{e_\Gamma+e_\Phi-e_{\Gamma',\Phi'}}(1-p^{e_{\Gamma',\Phi'}}) \\
&= \sum_{i=1}^{\min(e_\Gamma,e_\Phi)} \sum_{\Gamma'} \sum_{\Phi':e_{\Gamma',\Phi'}=i} p^{e_\Gamma+e_\Phi-i}(1-p^i)\,.
\end{aligned}
$$

Now, we notice that the second sum is running over $\binom{n}{v_\Gamma}\frac{v_\Gamma!}{|\mathrm{aut}(\Gamma)|} \asymp \frac{n^{v_\Gamma}}{|\mathrm{aut}(\Gamma)|}$ terms. By choosing $i=1$ in the first sum (a choice that leads to the asymptotically dominating term), we see that the third sum is running over $\asymp \frac{n^{v_\Phi-2}}{|\mathrm{aut}(\Phi)|}$ terms, since Γ' and Φ' have precisely one edge in common and there are $\binom{n-v_\Gamma}{v_\Phi-2} \asymp n^{v_\Phi-2}$ possible choices for the $e_\Phi - 1$ missing vertices to build a copy Φ' of Φ in $\mathcal{G}(n,p)$. Moreover, taking into account all possible choices and orientations for this common edge gives rise to another factor $2e_\Gamma e_\Phi$. Summarizing, the term with $i=1$ yields the asymptotic contribution

$$
2\,\frac{n^{v_\Gamma+v_\Phi-2}}{|\mathrm{aut}(\Gamma)||\mathrm{aut}(\Phi)|}\,e_\Gamma e_\Phi\,p^{e_\Gamma+e_\Phi-1}(1-p)\,.
$$

Choosing $i=2$ we see that there are two possible situations. Namely, the two common edges of Γ' and Φ' can or cannot have a common vertex. In the first situation and by the same reasoning as above, the asymptotic contribution is $\asymp c_1(\Gamma,\Phi)n^{v_\Gamma+v_\Phi-3}p^{e_\Gamma+e_\Phi-2}(1-p^2)$, while in the second case we have the asymptotic contribution

$$
\asymp c_2(\Gamma,\Phi)n^{v_\Gamma+v_\Phi-4}p^{e_\Gamma+e_\Phi-2}(1-p^2)
$$

with constants $c_1(\Gamma,\Phi), c_2(\Gamma,\Phi) > 0$ only depending on Γ and Φ. Moreover, it is clear from this discussion that for all $i \geq 3$ the corresponding terms in the above sum are of order $O(n^{v_\Gamma+v_\Phi-3})$. This proves the claim. □

Now, let us turn to the multivariate central limit theorem for the subgraph counting statistics X_Γ. For this, fix some $d \in \mathbb{N}$ and let $\Gamma_1,\ldots,\Gamma_d$ be d fixed (finite, simple) graphs with associated counting statistics $X_{\Gamma_1},\ldots,X_{\Gamma_d}$. For $i \in \{1,\ldots,d\}$ define the normalized random variables $\widetilde{X}_{\Gamma_i} := n^{1-v_{\Gamma_i}}(X_{\Gamma_i} - \mathbb{E}[X_{\Gamma_i}])$ and the random vector $\mathbf{X_\Gamma} := (\widetilde{X}_{\Gamma_1},\ldots,\widetilde{X}_{\Gamma_d})$. Our next result is the announced multivariate central limit theorem for $\mathbf{X_\Gamma}$, which adds a rate of convergence to the related result in the paper of Janson and Nowicki [7].

Theorem 4.2. *Let $\Sigma = (\Sigma_{ij})_{i,j=1}^d$ be the matrix given by*

$$
\Sigma_{ij} := \sigma_i\sigma_j \qquad \textit{with} \qquad \sigma_i := \sqrt{2p(1-p)}\,\frac{e_{\Gamma_i}}{|\mathrm{aut}(\Gamma_i)|}p^{e_{\Gamma_i}-1}\,, \qquad i \in \{1,\ldots,d\}
$$

and let $\mathbf{N}_\Sigma$ denote a d-dimensional centered Gaussian vector with covariance matrix Σ. Then, there exists a constant $c := c(\Gamma_1,\ldots,\Gamma_d,p) > 0$ only depending on $\Gamma_1,\ldots,\Gamma_d$ and on p such that

$$
d_4(\mathbf{X_\Gamma},\mathbf{N}_\Sigma) \leq \frac{c}{n}
$$

for all sufficiently large n.

Proof. It readily follows from Lemma 4.1 and the definition of the constants σ_i in the statement of the theorem that, for all $i,j = 1,\ldots,d$, $\mathrm{cov}(\widetilde{X}_{\Gamma_i},\widetilde{X}_{\Gamma_j}) = \Sigma_{ij} + O(n^{-1})$ and hence $|\mathrm{cov}(\widetilde{X}_{\Gamma_i},\widetilde{X}_{\Gamma_j}) - \Sigma_{ij}| = O(n^{-1})$.

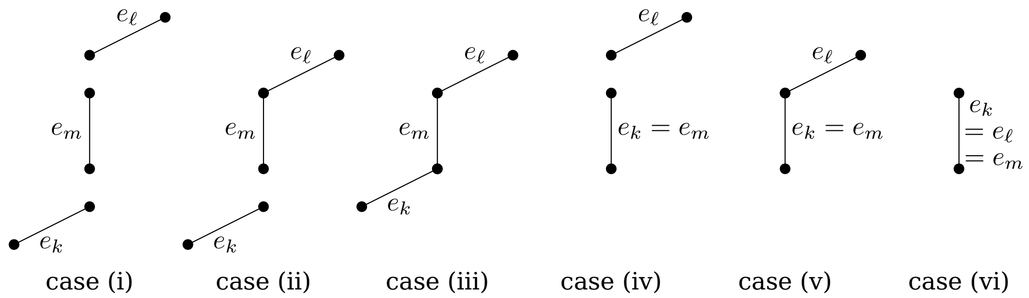

Figure 1: The different cases arising in the proof of Theorem 4.2. We distinguish according to the number of vertices that edges e_m has in common with e_k and e_ℓ, respectively. For example, the illustration in case (ii) means that $|e_k \cap e_m| = 0$ and $|e_\ell \cap e_m| = 1$ or, vice versa, $|e_k \cap e_m| = 1$ and $|e_\ell \cap e_m| = 0$. This allows both situations, $|e_k \cap e_\ell| = 0$ or $|e_k \cap e_\ell| = 1$ with $|e_k \cap e_\ell \cap e_m| = 0$.

To evaluate the other terms in the bound provided by the discrete second-order Poincaré inequality in Theorem 3.7, we first consider for each $i = 1, \ldots, d$ and $k, \ell = 1, \ldots, \binom{n}{2}$ the first-and second-order discrete Malliavin derivatives $D_k \widetilde{X}_{\Gamma_i}$ and $D_k D_\ell \widetilde{X}_{\Gamma_i}$. From the very definition it follows that

$$D_k \widetilde{X}_{\Gamma_i} = \frac{\sqrt{p(1-p)}}{n^{v_{\Gamma_i}-1}}((X_{\Gamma_i})_k^+ - (X_{\Gamma_i})_k^-)$$

and the difference $(X_{\Gamma_i})_k^+ - (X_{\Gamma_i})_k^-$ is just the number of copies of Γ_i that contain edge e_k. Since there are $O(n^{v_{\Gamma_i}-2})$ possible choices for the $v_{\Gamma_i} - 2$ missing vertices to build such a copy, it follows that $(X_{\Gamma_i})_k^+ - (X_{\Gamma_i})_k^- = O(n^{v_{\Gamma_i}-2})$ and thus

$$D_k \widetilde{X}_{\Gamma_i} = O(n^{-1})\,. \tag{4.2}$$

For the same reason we conclude that

$$D_k D_\ell \widetilde{X}_{\Gamma_i} = \begin{cases} O(n^{-1}) & : |e_k \cap e_\ell| = 2\,, \\ O(n^{-2}) & : |e_k \cap e_\ell| = 1\,, \\ O(n^{-3}) & : |e_k \cap e_\ell| = 0\,, \end{cases} \tag{4.3}$$

where $|e_k \cap e_\ell|$ denotes the number of vertices that e_k and e_ℓ have in common.

We can now start to bound, for each $i, j = 1, \ldots, d$, the term $B_1(i,j)$. Using the Cauchy-Schwarz inequality, it first follows that

$$B_1(i,j)^2 = \frac{15}{4} \sum_{k,\ell,m=1}^{\binom{n}{2}} (\mathbb{E}[(D_k X_{\Gamma_i})^4]\mathbb{E}[(D_\ell X_{\Gamma_i})^4])^{1/4} (\mathbb{E}[(D_m D_k X_{\Gamma_j})^4]\mathbb{E}[(D_m D_\ell X_{\Gamma_j})^4])^{1/4}\,.$$

Now, we have to distinguish different cases that are illustrated in Figure 1 (up to permutation of the indices k, ℓ and m). In case (i), we have $O(\binom{n}{2}) = O(n^2)$ possibilities to choose each of the three edges and by (4.2) each first-order discrete Malliavin derivative contributes $O(n^{-1})$, while each second-order derivatives contribute $O(n^{-3})$ according to (4.3). Thus, in case (i) the sum is of order $O(n^6 \cdot n^{-2} \cdot n^{-6}) = O(n^{-2})$. In case (ii), we have $O(\binom{n}{2}) = O(n^2)$ possibilities to choose each of the edges e_k and e_m, while there are only $O(n)$ possibilities for e_ℓ. Moreover, in view of (4.2) and (4.3) the first-order discrete Malliavin derivatives contribute again $O(n^{-2})$, but the second-order derivatives contribute only $O(n^{-5})$. Thus, the terms corresponding to case (ii) in the above sum are of order $O(n^5 \cdot n^{-2} \cdot n^{-5}) = O(n^{-2})$. The same behavior is also valid for cases (iii), (iv), (v) and (vi), which shows that $B_1(i,j)^2 = O(n^{-2})$. Similarly we see that $B_2(i,j)^2 = O(n^{-2})$.

We are thus left with the terms $B_3(i,j)$ and $B_4(i,j)$ given by

$$B_3(i,j) := \frac{1}{2} d^{3/2} \sum_{k=1}^{\binom{n}{2}} \frac{|p_k - q_k|}{\sqrt{p_k q_k}} (\mathbb{E}[(D_k X_{\Gamma_i})^2])^{1/2} (\mathbb{E}[(D_k X_{\Gamma_j})^4])^{1/2} ,$$

$$B_4(i,j) := \frac{5}{12} d^2 \sum_{k=1}^{\binom{n}{2}} \frac{1}{p_k q_k} (\mathbb{E}[(D_k X_{\Gamma_i})^4])^{1/4} (\mathbb{E}[(D_k X_{\Gamma_j})^4])^{3/4} .$$

In $B_3(i,j)$ there are $\binom{n}{2} \asymp n^2$ choices for k and the first-order discrete Malliavin derivatives are of order $O(n^{-1})$ by (4.2), which shows that $B_3(i,j) = O(n^{-1})$ for all choices of i and j. Finally, in $B_4(i,j)$ there are again $\binom{n}{2} \asymp n^2$ choices for k and once again by (4.2) the derivatives are of order $O(n^{-1})$. Hence, $B_4(i,j) = O(n^{-2})$ for all possible choices of i and j. Summarizing we conclude that

$$\begin{aligned} d_4(\mathbf{X}_\mathbf{\Gamma}, \mathbf{N}_\Sigma) &\le \frac{1}{2} \sum_{i,j=1}^{d} \left[|\Sigma_{ij} - \mathrm{cov}(X_{\Gamma_i}, X_{\Gamma_j})| + B_1(i,j) + B_2(i,j) + B_3(i,j) + B_4(i,j) \right] \\ &= O(n^{-1}) , \end{aligned}$$

where the constant hidden in the O-notation only depends on p as well as on the graphs $\Gamma_1, \ldots, \Gamma_d$. This completes the proof of the theorem. □

Remark 4.3. The structure of the asymptotic covariance matrix Σ in the previous theorem implies that Σ has rank 1. Thus, Σ cannot be positive definite, but it clearly is positive semidefinite.

Remark 4.4. We believe that there are also other methods available in the existing literature that allow to prove results similar to Theorem 4.2 and even with respect to a stronger probability metric. For example, the multivariate exchangeable pairs approach used in [20] might be generalized to subgraph counts of arbitrary graphs. On the other hand, this might require serious technical efforts, while our proof of the quantitative multivariate central limit theorem for subgraph counts basically only requires simple (asymptotic) counting arguments. A similar comment also applies to the random cubical complexes treated in the next section.

We continue our study of the Erdős-Rényi random graph $\mathcal{G}(n,p)$ by establishing a central limit theorem for the vertex degree statistic in the the case that $p = \theta/(n-1)$ for a $\theta \in (0,1)$. Although the number of vertices of a given degree is a special case of a subgraph counting statistic as considered above, the significant difference here is that we allow the success probability p to vary with n.

For $i \ge 0$ we denote by V_i the number of vertices of degree i in the Erdős-Rényi random graph $\mathcal{G}(n,p)$ for a $p \in (0,1)$ and where we assume that n is sufficiently large so that all quantities we deal with are well-defined. More formally, if we denote by $v_1, \ldots, v_n$ the n vertices of the complete graph K_n, then

$$V_i = \sum_{k=1}^{n} \mathbf{1}\{\deg(v_k) = i\} ,$$

where $\deg(v_k)$ is the degree of v_k in $\mathcal{G}(n,p)$, that is, the (random) number of edges emanating from v_k. Since for each $k \in \{1, \ldots, n\}$, $\mathbb{E}[\mathbf{1}\{\deg(v_k) = i\}] = P(\deg(v_k) = i) = \binom{n-1}{i} p^i (1-p)^{n-1-i}$, it follows that

$$\mathbb{E}[V_i] = n \binom{n-1}{i} p^i (1-p)^{n-1-i} .$$

The covariance between V_i and V_j for $i,j \geq 0$ under the choice $p = \theta/(n-1)$ for a $\theta \in (0,1)$ has been investigated in [5] and we recall from Theorem 4.2 there that

$$\mathrm{cov}(V_i, V_j) = \frac{1}{n}\mathbb{E}[V_i]\mathbb{E}[V_j]\Big(\frac{(i-\theta)(j-\theta)}{\theta(1-\theta/(n-1))} - 1\Big) + \mathbf{1}\{i=j\}\mathbb{E}[V_i]\,. \tag{4.4}$$

We define $F_i := (V_i - \mathbb{E}[V_i])/\sqrt{n}$, fix $d \geq 1$ as well as $1 \leq i_1 < \ldots < i_d$, put $\mathbf{D} := (i_1, \ldots, i_d)$ and define the random vector $\mathbf{F_D} := (F_{i_1}, \ldots, F_{i_d})$. From now on and for the rest of this subsection, we assume that the success probability p is of the form $p = \theta/(n-1)$ for some fixed $\theta \in (0,1)$. Then, it is easily seen from the expression for $\mathrm{cov}(V_i, V_j)$ in (4.4) that

$$\mathrm{cov}(F_i, F_j) = \frac{1}{n}\mathrm{cov}(V_i, V_j) \to \frac{\theta^{i+j}}{i!\,j!}e^{-2\theta}\Big(\frac{(i-\theta)(j-\theta)}{\theta} - 1\Big) + \mathbf{1}\{i=j\}\frac{\theta^i}{i!}e^{-\theta}\,, \tag{4.5}$$

as $n \to \infty$. Our next result is a multivariate central limit theorem for the vertex degree vector $\mathbf{F_D}$. It is a weaker version of [5, Theorem 4.2] for which, using a slightly smoother probability metric, we can give a quick proof based on our multivariate discrete second-order Poincaré inequality. For Berry-Essen-type rates of convergence in the one-dimensional case we refer to [4, Theorem 2.1], [11, Theorem 1.3] and [19, Theorem 2.1 and Equation (3.3)].

Theorem 4.5. *Let $\Sigma = (\Sigma_{ij})_{i,j=1}^d$ be the matrix given by*

$$\Sigma_{ij} = \frac{\theta^{i+j}}{i!\,j!}e^{-2\theta}\Big(\frac{(i-\theta)(j-\theta)}{\theta} - 1\Big) + \mathbf{1}\{i=j\}\frac{\theta^i}{i!}e^{-\theta}\,, \qquad i \in \{1, \ldots, d\}\,,$$

and let $\mathbf{N}_\Sigma$ be a d-dimensional centered Gaussian random vector with covariance matrix Σ. Then, there exists a constant $c = c(i_1, \ldots, i_d, \theta) > 0$ only depending on $i_1, \ldots, i_d$ and θ such that

$$d_4(\mathbf{F_D}, \mathbf{N}_\Sigma) \leq \frac{c}{\sqrt{n}}$$

for all sufficiently large n.

Proof. From (4.5) we infer that $\mathrm{cov}(F_i, F_j) \to \Sigma_{ij}$, as $n \to \infty$. Moreover, from the structure of the covariance (4.4) we also conclude that $|\mathrm{cov}(F_i, F_j) - \Sigma_{ij}| = O(n^{-1})$.

Next, we fix $i \in \{1, \ldots, d\}$ and $k, \ell \in \{1, \ldots, \binom{n}{2}\}$. As in the proof of Theorem 1.3 in [11] we notice that adding or removing an edge from $\mathcal{G}(n,p)$ results in a change of at most 2 for the number of vertices of degree i. In other words,

$$|D_k F_i| \leq \frac{2\sqrt{pq}}{\sqrt{n}}\,.$$

For the second-order discrete Malliavin derivative we observe that $D_k D_\ell F_i$ is zero whenever $k = \ell$ or the edges e_k and e_ℓ corresponding to k and ℓ, respectively, do not have a common vertex. Thus, it follows that

$$|D_k D_\ell F_i| \leq \frac{2pq}{\sqrt{n}}\,\mathbf{1}\{|e_k \cap e_\ell| = 1\}\,.$$

We can now evaluate the terms $B_1(i,j)$ to $B_4(i,j)$ in Theorem 3.7. Using the Cauchy-Schwarz inequality we first conclude that

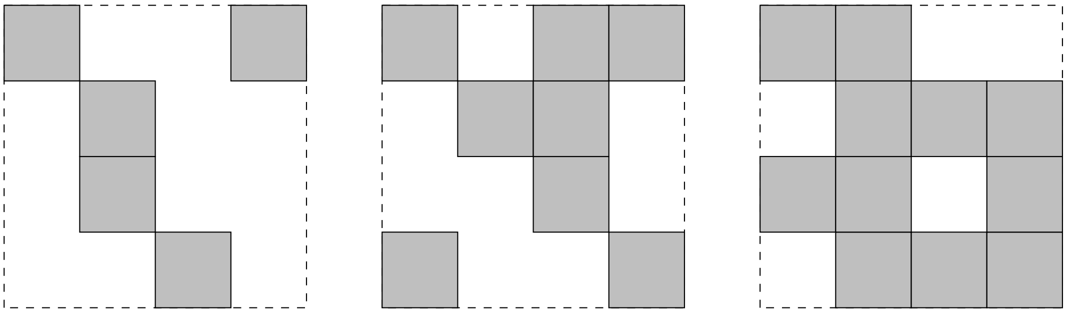

Figure 2: Illustrations of the voxel model $\mathcal{C}$ of a random cubical complex with $d = 2$ and $n = 4$ for increasing values of p.

$$
\begin{aligned}
B_1(i,j)^2 &= \frac{15}{4} \sum_{k,\ell,m=1}^{\binom{n}{2}} (\mathbb{E}[(D_k F_i)^4]\mathbb{E}[(D_\ell F_i)^4])^{1/4} (\mathbb{E}[(D_m D_k F_j)^4]\mathbb{E}[(D_m D_\ell F_j)^4])^{1/4} \\
&\le \frac{60(pq)^3}{n^2} \sum_{k,\ell,m=1}^{\binom{n}{2}} \mathbf{1}\{|e_m \cap e_k| = 1, |e_m \cap e_\ell| = 1\} \\
&= \frac{60(pq)^3}{n^2} \binom{n}{2} (2(n-2))^2 = O(n^{-1}),
\end{aligned}
$$

since $p = \theta/(n-1)$. Similarly, we have that $B_2(i,j)^2 = O(n^{-1})$. For the remaining terms $B_3(i,j)$ and $B_4(i,j)$ we see that

$$
\begin{aligned}
B_3(i,j) &= \frac{1}{2} d^{3/2} \sum_{k=1}^{\binom{n}{2}} \frac{|p_k - q_k|}{\sqrt{p_k q_k}} (\mathbb{E}[(D_k F_i)^2])^{1/2} (\mathbb{E}[(D_k F_j)^4])^{1/2} \\
&\le \binom{n}{2} \frac{4pq}{n^{3/2}} d^{3/2} = O(n^{-1/2}), \\
B_4(i,j) &= \frac{5}{12} d^2 \sum_{k=1}^{\binom{n}{2}} \frac{1}{p_k q_k} (\mathbb{E}[(D_k F_i)^4])^{1/4} (\mathbb{E}[(D_k F_j)^4])^{3/4} \\
&\le \binom{n}{2} \frac{20pq}{3n^2} d^2 = O(n^{-1}).
\end{aligned}
$$

By Theorem 3.7 we have thus proved the result. □

4.2 Intrinsic volumes of random cubical complexes

Fix a space dimension $d \ge 1$, $n \ge 3$ and consider the lattice $\mathcal{L} := \{[0,1]^d + z : z \in \{0,\dots,n-1\}^d\}$ consisting of n^d unit cubes $C_1,\dots,C_{n^d}$ of dimension d. To avoid boundary effects, we identify opposite faces in $\mathcal{L}$, a convention which supplies $\mathcal{L}$ with the topology of a d-dimensional torus. Now, we number the cubes in $\mathcal{L}$ in a fixed but arbitrary way and assign to each cube $C_k \in \mathcal{L}$ a Rademacher random variable X_k such that $P(X_k = 1) = p$ and $P(X_k = -1) = 1 - p =: q$ for some fixed parameter $p \in (0,1)$. Following the paper of Werman and Wright [25] the voxel model for a so-called random cubical complex $\mathcal{C}$ arises from $\mathcal{L}$ when each cube C_k is removed from $\mathcal{L}$ for which $X_k = -1$, see Figure 2. It should be clear that the random cubical complexes arising in this way may be represented as finite unions of disjoint open cubes of dimensions $0, 1, \dots, d$, corresponding to the vertices, edges, etc.

We are interested in the intrinsic volumes $V_j(\mathcal{C})$ of the random cubical complex $\mathcal{C}$ for all $j \in \{0, 1, \dots, d\}$. Let us point out that intrinsic volumes of convex sets (and their additive extensions to finite unions of convex sets) are of fundamental importance in geometry. This is demonstrated, for example, by Hadwiger's celebrated theorem saying that the intrinsic volumes form a basis of the vector space of continuous and motion-invariant valuations on compact, convex subsets of $\mathbb{R}^d$ (cf. Chapter 4 in [21]). Formally, the intrinsic volumes $V_0(K), \dots, V_d(K)$ of a convex body $K \subset \mathbb{R}^d$ arise as the coefficients of the Steiner polynomial, that is,

$$\mathrm{vol}_d(K \oplus rB^d) = \sum_{j=0}^{d} V_j(K)\kappa_{d-j} r^{d-j} , \qquad r \geq 0 ,$$

where vol_d means the d-dimensional Lebesgue measure, B^d is the d-dimensional unit ball, $\oplus$ means Minkowski addition and, for any $k \in \mathbb{N}$, κ_k is the k-volume of the k-dimensional unit ball. For example, $V_d(K)$ is the volume, $2V_{d-1}(K)$ is the surface area, $V_1(K)$ is a constant multiple of the mean width and $V_0(K)$ is the Euler characteristic of K. Let us also point out that the intrinsic volumes are highly relevant in image analysis and stereology, see [15, 22, 23]. In the context of (random) cubical complexes, the intrinsic volumes were first studied in [25] in order to understand noise in digital images.

In our context, we also need to introduce the intrinsic volumes for the relative interior of a convex body, a concept which goes back to Groemer [6]. Since we are interested only in finite unions of cubes, we go the more direct way already used in [25] and do not present the general definition. Namely, for a $\delta \in \mathbb{N}$, by a closed δ-cube we understand any translate of $[0, 1]^\delta$, while an open δ-cube refers to a translate of $(0, 1)^\delta$. The intrinsic volume $V_j(C)$ of order $j \in \{0, 1, \dots, i\}$ of a closed δ-cube C is given by $V_j(C) = \binom{\delta}{j}$, while the jth intrinsic volume of an open δ-cube D is $V_j(D) = (-1)^{\delta-j}\binom{\delta}{j} =: V_j(\delta)$. Finally, for the random cubical complex $\mathcal{C}$ as defined above we have the following representation for $V_j(\mathcal{C})$ from [25]:

$$V_j(\mathcal{C}) = \sum_{D \text{ open cube in } \mathcal{L}} \xi_{D,j} , \qquad \xi_{D,j} := (-1)^{\dim(D)-j}\binom{\dim(D)}{j} \mathbf{1}\{D \text{ belongs to } \mathcal{C}\} . \tag{4.6}$$

From this representation it readily follows that

$$\mathbb{E}[V_j(\mathcal{C})] = \sum_{D \text{ open cube in } \mathcal{L}} \mathbb{E}[\xi_{D,j}] = \sum_{\delta=j}^{d} N_\delta \, P_\delta \, V_j(\delta) ,$$

where $N_\delta = \binom{d}{\delta} n^d$ denotes the number of δ-cubes in $\mathcal{L}$ and $P_\delta = 1 - q^{2^{d-\delta}}$ is the probability that an arbitrary δ-cube is included in $\mathcal{C}$, see [25].

Although the variance of $V_j(\mathcal{C})$ has been computed in [25], in our context we will also need information about the covariance structure between $V_i(\mathcal{C})$ and $V_j(\mathcal{C})$. This is provided by the next lemma.

Lemma 4.6. *Let $i, j \in \{0, 1, \dots, d\}$. Then,*

$$\mathrm{cov}(V_i(\mathcal{C}), V_j(\mathcal{C})) = c(i, j) \, n^d$$

with the constant $c(i, j)$ given by

$$c(i, j) = \sum_{a=0}^{d} \sum_{b=0}^{d} \sum_{\delta=0}^{d} V_i(a) V_j(b) \binom{d}{\delta} N_{a,b,\delta} \, q^{2^{d-a}+2^{d-b}} (q^{-2^{d-\delta}} - 1) ,$$

where $N_{a,b,\delta} = \sum_{\ell=0}^{\delta}(-1)^{\delta-\ell}\binom{\delta}{\ell}\binom{\ell}{a}\binom{\ell}{b}2^{\delta+\ell-a-b}$.

Proof. We first notice that for two open cubes D and D' in $\mathcal{L}$ (possibly having different dimensions) the random variables ξ_D and $\xi_{D'}$ are independent whenever D and D' are not faces of a common d-dimensional cube from $\mathcal{L}$. Thus, using (4.6) we conclude that

$$\mathrm{cov}(V_i(\mathcal{C}), V_j(\mathcal{C})) = \sum_{D,D'} \mathrm{cov}(\xi_{D,i}, \xi_{D',j}) = \sum_{D,D'} \left(\mathbb{E}[\xi_{D,i}\xi_{D',j}] - \mathbb{E}[\xi_{D,i}]\mathbb{E}[\xi_{D',j}]\right)$$

with the sum running over all open cubes D, D' in $\mathcal{L}$ that are faces of a common d-cube. To evaluate this sum, we observe that for each pair of cubes D, D' there is a unique cube $C(D, D')$ of which D and D' are common faces and which has the smallest dimension among all such cubes (in fact, the existence of such a cube is the reason why $n \geq 3$ is assumed in this section). On the contrary, if C is a cube of dimension $\delta \in \{0, 1, \ldots, d\}$, we let $N_{a,b,\delta}$ be the number of pairs of cubes D and D' of dimensions a and b, respectively, for which $C(D, D') = C$. We notice that the value of $N_{a,b,\delta}$ is independent of the particular choice of C and given by

$$N_{a,b,\delta} = \sum_{\ell=0}^{\delta}(-1)^{\delta-\ell}\binom{\delta}{\ell}\binom{\ell}{a}\binom{\ell}{b}2^{\delta+\ell-a-b}$$

according to Equation (18) in [25]. Especially, $N_{a,b,\delta}$ is independent of n. Moreover, following Equation (20) in [25] we denote by

$$P_{a,b,\delta} = (1 - q^{2^{d-\delta}}) + q^{2^{d-\delta}}(1 - q^{2^{d-a}-2^{d-\delta}})(1 - q^{2^{d-b}-2^{d-\delta}})$$

the probability that both D and D' are included in the cubical complex $\mathcal{C}$. Then, we conclude that

$$\mathrm{cov}(V_i(\mathcal{C}), V_j(\mathcal{C})) = \sum_{a=0}^{d}\sum_{b=0}^{d}\sum_{\delta=0}^{d} N_\delta N_{a,b,\delta}\left(\mathbb{E}[\xi_{D,i}\xi_{D',j}] - \mathbb{E}[\xi_{D,i}]\mathbb{E}[\xi_{D',j}]\right).$$

According to our above discussion, the two expectations $\mathbb{E}[\xi_{D,i}]$ and $\mathbb{E}[\xi_{D',j}]$ are given by $\mathbb{E}[\xi_{D,i}] = P_a V_i(a)$ and $\mathbb{E}[\xi_{D',j}] = P_b V_j(b)$. Finally, $\mathbb{E}[\xi_{D,i}\xi_{D',j}]$ equals $P_{a,b,\delta}V_i(a)V_j(b)$, which implies that

$$\mathrm{cov}(V_i(\mathcal{C}), V_j(\mathcal{C})) = \sum_{a=0}^{d}\sum_{b=0}^{d}\sum_{\delta=0}^{d} V_i(a)V_j(b)N_\delta N_{a,b,\delta}(P_{a,b,\delta} - P_a P_b).$$

Since $P_{a,b,\delta} - P_a P_b = q^{2^{d-a}+2^{d-b}}(q^{-2^{d-\delta}} - 1)$, the proof is complete. □

Now, define for $j \in \{0, 1, \ldots, d\}$ the centered, normalized random variables $\widetilde{V}_j(\mathcal{C}) := n^{-d/2}(V_j(\mathcal{C}) - \mathbb{E}[V_j(\mathcal{C})])$ and the random vector $\mathbf{V} := (\widetilde{V}_0(\mathcal{C}), \widetilde{V}_1(\mathcal{C}), \ldots, \widetilde{V}_d(\mathcal{C}))$. Our next theorem provides a bound for the multivariate normal approximation of $\mathbf{V}$ and this way extends Theorem 4 in [25].

Theorem 4.7. *Let $\Sigma := (\Sigma_{ij})_{i,j=0}^{d}$ be the matrix $\Sigma_{ij} := c(i, j)$ with the constants $c(i, j)$ given by Lemma 4.6. Then, there exists a constant $C = C(p, d)$ only depending on p and on d such that*

$$d_4(\mathbf{V}, \mathbf{N}_\Sigma) \leq \frac{C}{n^{d/2}},$$

where $\mathbf{N}_\Sigma$ is a $(d+1)$-dimensional centered Gaussian vector with covariance matrix Σ.

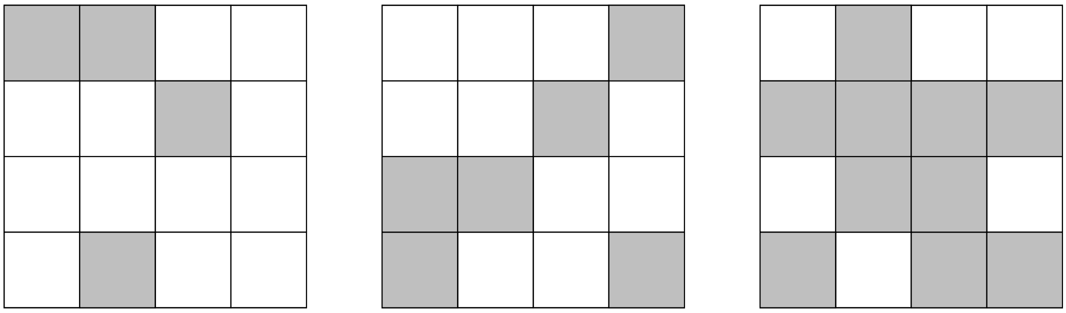

Figure 3: Illustrations of the plaquette model $\mathcal{P}$ of a random cubical complex with $d = 2$ and $n = 4$ for increasing values of p. The grey cubes are included, while the white cubes are not included in $\mathcal{P}$.

Proof. By Lemma 4.6 it follows that, for all $i, j \in \{0, 1, \ldots, d\}$, $\mathrm{cov}(\widetilde{V}_i(\mathcal{C}), \widetilde{V}_j(\mathcal{C})) = \Sigma_{ij}$. Thus, it only remains to bound the terms $B_1(i, j)$ to $B_4(i, j)$ in Theorem 3.7. To this end, we need appropriate estimates for the first- and second-order discrete Malliavin derivatives $D_k\widetilde{V}_i(\mathcal{C})$ and $D_kD_\ell\widetilde{V}_i(\mathcal{C})$ for all $i \in \{0, 1, \ldots, d\}$, respectively. For this, we recall the representation (4.6) and observe that for each $k \in \{1, \ldots, n^d\}$, $D_k\widetilde{V}_i(\mathcal{C})$ can be written as $\sqrt{pq}/n^{d/2}$ times a sum of at most 6^d summands, where each of them is bounded independently of n. Here, $6^d \geq 2^{d-\delta} \cdot 3^d$ for any $\delta \in \{0, 1, \ldots, d\}$ and $2^{d-\delta}$ is the number of d-dimensional cubes of which a fixed δ-dimensional cube is a face of, while $3^d = \sum_{\delta=0}^{d} \binom{d}{\delta} 2^{d-\delta}$ is the total number of faces of a d-dimensional cube. As a consequence, we find that

$$D_k\widetilde{V}_i(\mathcal{C}) = O(n^{-d/2})$$

and by the triangle inequality also

$$D_kD_\ell\widetilde{V}_i(\mathcal{C}) = O(n^{-d/2})\,,$$

where the hidden constants only depend on d and on p. Now, it is crucial to observe that for any fixed $k \in \{1, \ldots, n^d\}$ the second-order discrete Malliavin derivative $D_kD_\ell\widetilde{V}_i(\mathcal{C})$ is even identically zero whenever the cubes corresponding to k and ℓ are not neighbors of each other. Since any cube in $\mathcal{L}$ has only a finite number of neighbors, independently of n, we conclude that in the term $B_1(i, j)$ provided by the multivariate discrete second-order Poincaré inequality in Theorem 3.7 there are exactly n^d choices for m and only a constant number of choices for k and ℓ for which the corresponding summand is non-vanishing. As a consequence, $B_1(i, j)^2$ is of order $O(n^d \cdot n^{-d} \cdot n^{-d}) = O(n^{-d})$, implying that $B_1(i, j) = O(n^{-d/2})$ for any choice of $i, j \in \{0, 1, \ldots, d\}$. Since the same behavior can also be observed for the remaining terms $B_2(i, j), B_3(i, j)$ and $B_4(i, j)$, the claim follows. □

Besides of the voxel model, the authors of [25] also consider three further models for random cubical complexes: the plaquette model, the closed faces model and the independent faces model. For each of these models our method can be used to derive a multivariate central limit theorem for the random vector of their intrinsic volumes and to obtain bounds on the d_4-distance of order $O(n^{-d/2})$ in each case. We present the result only in the case of the plaquette model, since it is close in spirit to the celebrated random simplicial complexes introduced by Linial and Meshulam [12] that have been object of intensive studies. To introduce the model formally, we fix $d \geq 1$, $n \geq 3$, and define the set $\mathcal{G} := \{\partial[0, 1]^d + z : z \in \{0, \ldots, n-1\}^d\}$, where $\partial[0, 1]^d$ stands for the boundary of the unit d-cube $[0,1]^d$. The open cubes $C_1,\dots,C_{n^d}$ in $\{(0,1)^d+z : z\in\{0,\dots,n-1\}^d\}$ are assumed to be numbered in a fixed but arbitrary way and we assign to each cube C_k a Rademacher random variable X_k with $P(X_k=1)=p$ and $P(X_k=-1)=1-p=: q$. The plaquette model now arises if those open cubes C_k are joint with the set $\mathcal{G}$ for which the associated Rademacher random variable X_k takes the value 1, see Figure 3.

The construction just described gives rise to a random set $\mathcal{P}$ and as in the case of the voxel model $\mathcal{C}$ we are interested in its intrinsic volumes $V_j(\mathcal{P})$, $j\in\{0,1,\dots,d\}$. Using the same notation as in the previous example, we formally have that $V_j(\mathcal{P})=\sum_D \xi_D$ with the sum running over all open cubes in $\mathcal{P}$ and hence $\mathbb{E}[V_j(\mathcal{P})]=\sum_{\delta=j}^d N_\delta P_\delta V_j(\delta)$. However, in the plaquette model we have that the probabilities $P_0,P_1\dots,P_d$ satisfy $P_d=p$ and $P_\delta=1$ for $\delta\in\{0,\dots,d-1\}$, which implies (after some simplifications) that

$$\mathbb{E}[V_j(\mathcal{P})]=\begin{cases} pn^d & : j=d\,,\\ (-1)^{d-j}\binom{d}{j}(p-1)n^d & : j\in\{0,\dots,d-1\}\,,\end{cases}$$

see also Equation (27) in [25]. The covariance structure of the intrinsic volumes for the plaquette model is described in the next lemma.

Lemma 4.8. *Let $i,j\in\{0,1,\dots,d\}$. Then,*

$$\mathrm{cov}(V_i(\mathcal{P}),V_j(\mathcal{P}))=\binom{d}{i}\binom{d}{j}p(1-p)\,n^d\,.$$

Proof. By definition it follows that

$$\mathrm{cov}(V_i(\mathcal{P}),V_j(\mathcal{P}))=\sum_{D,D'}\mathrm{cov}(\xi_{D,i},\xi_{D',j})\,,$$

again with the sum running over all open cubes D and D' in $\mathcal{P}$. We notice that in this model the random variables $\xi_{D,i}$ and $\xi_{D',j}$ are independent except if $D=D'$ and $\dim(D)=\dim(D')=d$. In this case, we clearly have

$$\begin{aligned}\mathrm{cov}(\xi_{D,i},\xi_{D',j})&=\mathbb{E}[\xi_{D,i}\xi_{D,j}]-\mathbb{E}[\xi_{D,i}]\mathbb{E}[\xi_{D,j}]\\ &=\binom{d}{i}\binom{d}{j}p-\binom{d}{i}p\cdot\binom{d}{j}p=\binom{d}{i}\binom{d}{j}p(1-p)\end{aligned}$$

and the result follows. □

Now, we define the centered and normalized random variables $\widehat{V}_j(\mathcal{P}):=n^{-d/2}(V_j(\mathcal{P})-\mathbb{E}[V_j(\mathcal{P})])$ and the $(d+1)$-dimensional random vector $\mathbf{W}:=(\widehat{V}_0(\mathcal{P}),\widehat{V}_1(\mathcal{P}),\dots,\widehat{V}_d(\mathcal{P}))$. The next result is a multivariate central limit theorem for the random vector $\mathbf{W}$. Since the arguments are the same as in the proof of Theorem 4.7, we have decided not to present the details.

Theorem 4.9. *Let $\Sigma:=(\Sigma_{ij})_{i,j=0}^d$ be the matrix given by $\Sigma_{ij}=\binom{d}{i}\binom{d}{j}p(1-p)$. Then, there exists a constant $C=C(p,d)$ only depending on p and on d such that*

$$d_4(\mathbf{W},\mathbf{N}_\Sigma)\le\frac{C}{n^{d/2}}$$

with a $(d+1)$-dimensional centered Gaussian random vector $\mathbf{N}_\Sigma$ having covariance matrix Σ.

Remark 4.10. As for subgraph counting statistics it follows from the structure of the asymptotic covariance matrices Σ in Theorems 4.7 and 4.9 that Σ only has rank 1 and is hence only positive semidefinite rather than positive definite.

References

[1] Adler, R.J., Bobrowski, O., Borman, M.S., Subag, E. and Weinberger, S.: Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications - a Festschrift for Lawrence D. Brown*, IMS Collections **6**, 124–143, 2010. MR-2798515

[2] Bobrowski, O. and Kahle, M.: Topology of random geometric complexes: a survey. To appear in *Topology in Statistical Inference, the Proceedings of Symposia in Applied Mathematics*, 2014+.

[3] Chatterjee, S.: Fluctuations of eigenvalues and second order Poincaré inequalities. *Probab. Theory Relat. Fields* **143**, (2009), 1–40. MR-2449121

[4] Goldstein, L.: A Berry-Esseen bound with applications to vertex degree counts in the Erdős-Rényi random graph. *Ann. Appl. Probab.* **23**, (2013), 617–636. MR-3059270

[5] Goldstein, L. and Rinott, Y.: Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33**, (1996), 1–17. MR-1371949

[6] Groemer, H.: Eulersche Charakteristik, Projektionen und Quermaßintegrale. *Math. Ann.* **198**, (1972), 23–56. MR-0328774

[7] Janson, S. and Nowicki, K.: The asymptotic distributions of generalized U-statistics with applications to random graphs. *Probab. Theory Relat. Fields* **90**, (1991), 341–375. MR-1133371

[8] Kahle, M.: Topology of random simplicial complexes: a survey. In *Algebraic Topology: Applications and New Directions*, Contemporary Mathematics Volume **620**, American Mathematical Society, Providence, 201–221, 2014. MR-3290093

[9] Krokowski, K.: Poisson approximation of Rademacher functionals by the Chen-Stein method and Malliavin calculus. *Commun. Stoch. Anal.* **11**, (2017), 195–222.

[10] Krokowski, K., Reichenbachs, A. and Thäle, C.: Berry-Esseen bounds and multivariate limit theorems for functionals of Rademacher sequences. *Ann. Inst. H. Poincaré Probab. Stat.* **52**, (2016), 763–803. MR-3498009

[11] Krokowski, K., Reichenbachs, A. and Thäle, C.: Discrete Malliavin-Stein method: Berry-Esseen bounds for random graphs and percolation. *Ann. Probab.* **45**, (2017), 1071–1109. MR-3630293

[12] Linial, N. and Meshulam, R.: Homological connectivity of random 2-complexes. *Combinatorica* **26**, (2006), 475–487. MR-2260850

[13] Nourdin, I., Peccati, G. and Reinert, G.: Stein's method and stochastic analysis of Rademacher functionals. *Electron. J. Probab.* **15**, (2010), 1703–1742. MR-2735379

[14] Nourdin, I., Peccati, G. and Réveillac, A.: Multivariate normal approximation using Stein's method and Malliavin calculus. *Ann. Inst. H. Poincaré Probab. Stat.* **46**, (2010), 45–58. MR-2641769

[15] Ohser, J. and Schladitz, K.: 3D Images of Materials Structures: Processing and Analysis. Wiley-VCH Verlag, Weinheim, 2009.

[16] Peccati, G. and Zheng, C.: Multi-dimensional Gaussian fluctuations on the Poisson space. *Electron. J. Probab.* **15**, (2010), 1487–1527. MR-2727319

[17] Privault, N.: Stochastic Analysis in Discrete and Continuous Settings with Normal Martingales. Lecture Notes in Mathematics **1982**, Springer-Verlag, Berlin, 2009. MR-2531026

[18] Privault, N. and Torrisi, G.L.: The Stein and Chen-Stein methods for functionals of non-symmetric Bernoulli processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **12**, (2015), 309–356. MR-3357130

[19] Raič, M.: CLT-related large deviation bounds based on Stein's method. *Adv. Appl. Probab.* **39**, (2007), 731–752. MR-2357379

[20] Reinert, G. and Röllin, A.: Random subgraph counts and U-statistics: multivariate normal approximation via exchangeable pairs and embedding. *J. Appl. Probab.* **47**, (2010), 378–393. MR-2668495

[21] Schneider, R.: Convex Bodies: The Brunn-Minkowski Theory. 2nd Edition, Cambridge University Press, Cambridge, 2013. MR-3155183

[22] Svane, A.M.: Estimation of intrinsic volumes from digital grey-scale images. *J. Math. Imaging Vision* **49**, (2016), 352–376. MR-3197345

[23] Svane, A.M.: Valuations in image analysis. In *Tensor Valuations and their Applications in Stochastic Geometry and Imaging*, edited by E.B. Vedel-Jensen and M. Kiderlen, Lecture Notes in Mathematics **2177**, Springer-Verlag, Berlin, 435–454, 2017.

[24] Talagrand; M.: Spin Glasses: A Challenge for Mathematicians. Springer-Verlag, Berlin, 2003. MR-1993891

[25] Werman, M. and Wright, M.L.: Intrinsic volumes of random cubical complexes. *Discrete Comput. Geom.* **56**, (2016), 93–113. MR-3509032

[26] Zheng, G.: Normal approximation and almost sure central limit theorem for non-symmetric Rademacher functionals. *Stochastic Process. Appl.* **127**, (2017), 1622–1636. MR-3630238

Acknowledgments. We would like to thank the referee for a number of comments and useful suggestions that helped us to improve the paper.