RANDOM WALKS IN THE HIGH-DIMENSIONAL LIMIT I

ZAKHAR KABLUCHKO AND ALEXANDER MARYNYCH

Abstract. We prove limit theorems for random walks with n steps in the d-dimensional Euclidean space as both n and d tend to infinity. One of our results states that the path of such a random walk, viewed as a random metric space with the induced Euclidean metric, converges in probability in the Gromov-Hausdorff sense to the Wiener spiral, as $d, n \to \infty$. Another group of results describes various possible limit distributions for the squared distance between the random walker at time n and the origin.

1. Introduction

The purpose of the present paper is to study asymptotic properties of random walks with n steps in the d-dimensional space $\mathbb{R}^d$ as both parameters, n and d, tend to infinity. To be more concrete, consider a d-dimensional random walk whose increments are independent identically distributed (i.i.d.) random vectors with the uniform distribution on the unit sphere $\mathbb{S}^{d-1}$. In the regime when the dimension d is fixed and the number of steps n tends to infinity, Donsker's invariance principle implies that such random walk converges, after appropriate normalization, to the d-dimensional Brownian motion. But how does the path of the random walk look like if d also tends to infinity? It is well known that, as $d \to \infty$, the angle between two independent random vectors sampled uniformly on the unit sphere $\mathbb{S}^{d-1}$ tends to $\pi/2$ in probability; see [19, Remark 3.2.5] or [18, Theorem 4] for stronger results. This suggests that, informally speaking, the high-dimensional scaling limit of the random walk should be a curve (in an infinite dimensional Hilbert space) obtained by gluing together infinitely many mutually orthogonal infinitesimal increments.

A well-known curve of this type is the *Wiener spiral* (or the *crinkled arc*) introduced by Kolmogorov [16]. It is defined as the set $\mathbb{W} := \{\mathbb{1}_{[0,t]} : 0 \le t \le 1\}$ of indicator functions of the intervals $[0,t]$, considered as a subset of the Hilbert space $L^2[0,1]$ and endowed with the induced L^2-metric. As a metric space, the Wiener spiral is isometric to the interval $[0,1]$ endowed with the distance $d(t,s) = \sqrt{|t-s|}$. The Wiener spiral can be thought of as a curve $(\gamma_t)_{t\in[0,1]} := (\mathbb{1}_{[0,t]})_{t\in[0,1]}$ in the Hilbert space $L^2[0,1]$. It is easy to check that any two "chords" $\gamma_y - \gamma_x$ and $\gamma_v - \gamma_u$ with $0 \le x < y \le u < v \le 1$ are orthogonal; see [11, Problems 5,6] and [13, 20] for results on the uniqueness of the curve having this property. If $(B_t)_{t\in[0,1]}$ is a standard Brownian motion defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$, then the set of random variables $\{B_t : t \in [0,1]\}$, considered as a deterministic subset of $L^2(\Omega,\mathcal{F},\mathbb{P})$, is isometric to the Wiener spiral.

In Section 2 we shall state conditions under which the path of the random walk, viewed as a random metric space, converges to the Wiener spiral in the Gromov-Hausdorff sense. In Section 3 we shall state results on the limit distribution of the distance between the random walker at time n and the origin. Proof are collected in Sections 4 and 5.

The present paper deals with random walks having finite second moments. The case of random walks with infinite second moment will be treated in the follow-up work [14].

2. Convergence in the Gromov-Hausdorff sense

2.1. Convergence of random walks to the Wiener spiral. For every $d \in \mathbb{N}$ we consider a random walk in $\mathbb{R}^d$ whose increments $X_1^{(d)}, X_2^{(d)}, \dots$ are independent copies of a d-dimensional random vector $X^{(d)}$. The random walk is denoted by

$$S_0^{(d)} := 0, \quad S_k^{(d)} := X_1^{(d)} + \dots + X_k^{(d)}, \quad k \in \mathbb{N}. \tag{1}$$

The components of the random vectors $X_i^{(d)}$ and $S_i^{(d)}$ are denoted by $X_i^{(d)} = (X_{i,1}^{(d)}, \dots, X_{i,d}^{(d)})$ and $S_i^{(d)} = (S_{i,1}^{(d)}, \dots, S_{i,d}^{(d)})$, respectively. In what follows, let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^d$ and let $\langle \cdot, \cdot \rangle$ be the standard inner product in $\mathbb{R}^d$. We impose the following conditions on the increments, which we assume to hold for all $d \in \mathbb{N}$.

(a) The increments are centered and normalized, that is

$$\mathbb{E} X^{(d)} = 0, \quad \mathbb{E}\|X^{(d)}\|^2 = 1. \tag{2}$$

(b) The components of $X^{(d)}$ are mutually uncorrelated, that is,

$$\mathbb{E}[X_{i,j}^{(d)} X_{i,k}^{(d)}] = 0, \quad j,k \in \{1,\dots,d\}, \quad j \neq k, \quad i \in \mathbb{N}. \tag{3}$$

(c) The sequence $(\|X^{(d)}\|^2)_{d\in\mathbb{N}}$ is uniformly integrable, that is

$$\lim_{A\to\infty} \sup_{d\in\mathbb{N}} \mathbb{E}\left[\|X^{(d)}\|^2 \mathbb{1}_{\{\|X^{(d)}\|^2 > A\}}\right] = 0. \tag{4}$$

(d) The individual components of $X^{(d)}$ are negligible in the following sense:

$$\lim_{d\to\infty} \max_{k=1,\dots,d} \mathbb{E}(X_{1,k}^{(d)})^2 = 0. \tag{5}$$

Example 2.1 (Increments with i.i.d. components)**.** Let $\xi_1, \xi_2, \dots$ be i.i.d. random variables with $\mathbb{E}\xi_1 = 0$, $\mathbb{E}\xi_1^2 = 1$. If we put $X^{(d)} := (\xi_1, \dots, \xi_d)/\sqrt{d}$, then conditions (a)–(d) are satisfied.

Example 2.2 (Rotationally invariant increments)**.** Let $X^{(d)}$ be a random vector in $\mathbb{R}^d$ with rotationally invariant distribution. This means that $X^{(d)} = R^{(d)}U^{(d)}$, where $U^{(d)}$ is uniformly distributed on the unit sphere in $\mathbb{R}^d$, and $R^{(d)} \geq 0$ is a random variable independent of $U^{(d)}$. If $\mathbb{E}(R^{(d)})^2 = 1$ for all $d \in \mathbb{N}$ and the sequence $((R^{(d)})^2)_{d\in\mathbb{N}}$ is uniformly integrable, then conditions (a)–(d) are satisfied. In particular, $(R^{(d)})_{d\in\mathbb{N}}$ are allowed to be identically distributed (with finite second moment).

Example 2.3 (Random walks jumping along the coordinate axes)**.** The following model generalizes the simple random walk on $\mathbb{Z}^d$. Let $e_1,\dots,e_d$ denote the standard orthonormal basis of $\mathbb{R}^d$ and let $V^{(d)}$ be a random vector distributed uniformly on the set $\{e_1,\dots,e_d\}$, that is $\mathbb{P}[V^{(d)} = e_j] = 1/d$ for all $j \in \{1,\dots,d\}$. Put $X^{(d)} := R^{(d)}V^{(d)}$, where $R^{(d)}$ is a random variable which is independent of $V^{(d)}$ and satisfies $\mathbb{E}R^{(d)} = 0$ and $\mathbb{E}(R^{(d)})^2 = 1$, for all $d \in \mathbb{N}$. If the sequence $((R^{(d)})^2)_{d\in\mathbb{N}}$ is uniformly integrable, then conditions (a)–(d) are satisfied. In particular, taking $R^{(d)}$ to be uniformly distributed on $\{+1,-1\}$, we recover the simple symmetric random walk on $\mathbb{Z}^d$.

Let $n = n(d)$ be an arbitrary sequence of positive integers such that $n(d) \to \infty$, as $d \to \infty$. By default, the notation $d \to \infty$ implies that also $n = n(d) \to \infty$. We regard the image of the random walk with n steps in $\mathbb{R}^d$ as a finite random metric space. More precisely, let $\mathbb{M}_d$ be the metric space consisting of the points

$$0, S_1^{(d)}/\sqrt{n}, \dots, S_n^{(d)}/\sqrt{n} \tag{6}$$

and endowed with the metric induced by the Euclidean metric on $\mathbb{R}^d$. Our first main result states that, with probability converging to 1 as $d \to \infty$, the random metric space $\mathbb{M}_d$ becomes close, in the sense of the Gromov-Hausdorff distance to be defined below, to the Wiener spiral $\mathbb{W}$ defined in Section 1. Note that $\mathbb{W}$ is a *deterministic* metric space meaning that, in the high-dimensional limit, the random walk "freezes" (i.e., loses its randomness).

The *Gromov-Hausdorff distance* $d_{GH}(E_1,E_2)$ between two compact metric spaces E_1 and E_2 is defined as the infimum of $d_H(\varphi_1(E_1),\varphi_2(E_2))$, where the infimum is taken over all metric spaces (M,ρ) and all isometric embeddings $\varphi_1 : E_1 \to M$ and $\varphi_2 : E_2 \to M$, and d_H denotes the Hausdorff distance between compact subsets of M defined by

$$d_H(A,B) = \inf\{r > 0 : A \subset U_r(B), B \subset U_r(A)\}.$$

Here, $U_r(A) = \{m \in M : \rho(A,m) < r\}$ is the r-neighborhood of A in M. For details, we refer to Chapter 7 of [3]. It is known that the set of isometry classes of compact metric spaces, endowed with the Gromov-Hausdorff distance, becomes a complete separable metric space, called the *Gromov-Hausdorff space*. We are now ready to state our first result.

Theorem 2.4. *Let $n = n(d)$ be an arbitrary sequence of positive integers such that $n(d) \to \infty$, as $d \to \infty$. Suppose that conditions (a)–(d) are fulfilled. Then, as $d \to \infty$, the random metric space $\mathbb{M}_d$, considered as a random point in the Gromov-Hausdorff space, converges in probability to the Wiener spiral $\mathbb{W}$. That is to say, for every $\varepsilon > 0$,*

$$\lim_{d\to\infty} \mathbb{P}[d_{GH}(\mathbb{M}_d, \mathbb{W}) > \varepsilon] = 0.$$

The proof of Theorem 2.4 will be given in Sections 4.1 and 4.2. We shall also verify that the claim stays in force if $\mathbb{M}_d$ is replaced by the polygonal line interpolating consecutive points in (6), that is, for the metric space $\mathbb{M}_d^{cont}$ given by

$$\mathbb{M}_d^{cont} := \bigcup_{i=0}^{n-1} \left[\frac{S_i^{(d)}}{\sqrt{n}}, \frac{S_{i+1}^{(d)}}{\sqrt{n}} \right],$$

where $[a,b] \subset \mathbb{R}^d$ is the closed segment connecting $a, b \in \mathbb{R}^d$. As before, the space $\mathbb{M}_d^{cont}$ is endowed with the induced Euclidean metric on $\mathbb{R}^d$.

Corollary 2.5. *Under the same assumptions as in Theorem 2.4, for every $\varepsilon > 0$,*

$$\lim_{d\to\infty} \mathbb{P}[d_{GH}(\mathbb{M}_d^{cont}, \mathbb{W}) > \varepsilon] = 0.$$

2.2. **Convergence of high-dimensional stochastic processes.** In this section we state a result which is similar in spirit to Theorem 2.4 but applies to a different class of stochastic processes. Let K be an arbitrary index set and $X = (X(t))_{t\in K}$ be a real-valued stochastic process with $\mathbb{E}X(t) = 0$ and $\mathbb{E}(X(t))^2 < \infty$ for all $t \in K$. We suppose that $\rho(s,t) := \sqrt{\mathrm{Var}\,(X(s) - X(t))}$ defines a metric on K which turns K into a compact metric space and that the process X has a.s. continuous sample paths on (K, ρ). Finally, we suppose that $\mathbb{E}[\sup_{t\in K}(X(t))^2] < \infty$. Let $(X_1(t))_{t\in K}, (X_2(t))_{t\in K}, \ldots$ be independent copies of the process X. For every $d \in \mathbb{N}$ we consider the $\mathbb{R}^d$-valued stochastic process

$$\mathbb{X}_d(t) := \frac{(X_1(t), \ldots, X_d(t))}{\sqrt{d}} \in \mathbb{R}^d, \qquad t \in K.$$

Theorem 2.6. *The random metric space $\mathbb{K}_d := \{\mathbb{X}_d(t) : t \in K\} \subset \mathbb{R}^d$, endowed with the induced Euclidean metric, converges a.s. (as $d \to \infty$) to the deterministic metric space (K, ρ) in the Gromov-Hausdorff sense. That is to say,*

$$\mathbb{P}\left[\lim_{d\to\infty} d_{GH}(\mathbb{K}_d, K) = 0\right] = 1.$$

Example 2.7. Let $(X(t))_{t\in[0,1]}$ be the standard Brownian motion. Then, $(\mathbb{X}_d(t))_{t\in[0,1]}$ is a standard d-dimensional Brownian motion multiplied by $1/\sqrt{d}$. Theorem 2.6 implies that the random metric space $\{\mathbb{X}_d : t \in [0,1]\} \subset \mathbb{R}^d$, viewed as a random point in the Gromov-Hausdorff space, converges a.s. to the Wiener spiral $\mathbb{W}$.

3. Central limit theorems for the squared norm

In this section we state distributional limit theorems for the squared norm $\|S_n^{(d)}\|^2$, as $d \to \infty$, in three models presented in Examples 2.1, 2.2 and 2.3 of Section 2.1. Recall that all the corresponding random walks satisfy the assumptions (a)-(d) and, thus, converge to the Wiener spiral. However, the distributional behaviour is more sensitive to the details of each of the models and the corresponding distributional limit theorems are different. As before, $n = n(d)$ is an arbitrary sequence of positive integers such that $n(d) \to \infty$, as $d \to \infty$.

3.1. **Model 1: Random walks whose increments have i.i.d. components.** Recall that in this model $(\xi_{i,j})_{i,j=1}^{\infty}$ are independent copies of a random variable ξ such that $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$, and for every $d \in \mathbb{N}$ the increments of a d-dimensional random walk (1) are given by

$$X_i^{(d)} := \frac{(\xi_{i,1}, \dots, \xi_{i,d})}{\sqrt{d}}, \quad i \in \mathbb{N}.$$

Theorem 3.1. *In the setting just described suppose additionally that $\mathbb{E}\xi^4 < \infty$. Then,*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

Here and in what follows, $\mathrm{N}(0,\sigma^2)$ denotes the centered normal distribution with variance σ^2, and $\xrightarrow{w}$ denotes weak convergence of probability measures (convergence in distribution). In the next theorem we treat the case when ξ^2 has infinite second moment. More precisely, we suppose that ξ^2 belongs to the domain of attraction of an α-stable distribution with $\alpha \in (1,2)$. This means that the independent copies of ξ, denoted by $(\xi_i)_{i=1}^{\infty}$, satisfy

$$\frac{\xi_1^2 + \dots + \xi_m^2 - m}{m^{1/\alpha} L(m)} \xrightarrow[m\to\infty]{w} \zeta_\alpha \tag{7}$$

for some slowly varying function L and a zero-mean random variable ζ_α having a spectrally positive α-stable distribution.

Theorem 3.2. *Suppose that* (7) *holds for some $\alpha \in (1,2)$.*

(a) *If $n > d^{\frac{2-\alpha}{2\alpha-2}+\delta}$ for some $\delta > 0$ and all sufficiently large d, then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

(b) *If* $n < d^{\frac{2-\alpha}{2\alpha-2}-\delta}$ *for some* $\delta > 0$ *and all sufficiently large* d, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{d^{-1}(nd)^{1/\alpha} L(nd)} \xrightarrow[d\to\infty]{w} \zeta_\alpha.$$

3.2. **Model 2: Random walks with rotationally invariant increments.** We shall further specialize Example 2.2 by assuming additionally that the distribution of $R^{(d)}$ is the same for all $d \in \mathbb{N}$. Thus, for every $d \in \mathbb{N}$ we consider a random walk (1) in $\mathbb{R}^d$ whose increments are given by

$$X_i^{(d)} := R_i U_i^{(d)}, \quad i \in \mathbb{N},$$

where

- the radial components $(R_i)_{i=1}^\infty$ are independent copies of a non-negative random variable R with $\mathbb{E}R^2 = 1$;
- the directional components $(U_i^{(d)})_{i=1}^\infty$ are i.i.d. random vectors uniformly distributed on the unit sphere in $\mathbb{R}^d$;
- $(R_i)_{i=1}^\infty$ and $(U_i^{(d)})_{i=1}^\infty$ are independent.

Theorem 3.3. *In the setting just described, suppose additionally that* $\mathbb{E}R^4 < \infty$.

(a) *If* $\lim_{d\to\infty} n/d = 0$ *and* R *is not deterministic, then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, \mathrm{Var}\,(R^2)).$$

(b) *If* $\lim_{d\to\infty} n/d = \infty$ *or* R *is deterministic, then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

(c) *If* $n \sim \gamma d$ *for some constant* $\gamma \in (0,\infty)$, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, 2\gamma + \mathrm{Var}\,(R^2)).$$

Remark 3.4. Let us mention known results for random walks with *fixed* number of steps. Stam [18, Theorem 4 and p. 227] showed that if $R = 1$ is deterministic and $m \in \mathbb{N}$ is fixed, then

$$\frac{\|S_m^{(d)}\|^2 - m}{\sqrt{2m(m-1)/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

On the other hand, if R is not deterministic and $m \in \mathbb{N}$ is fixed, then it follows from [18, Theorem 4] that

$$\|S_m^{(d)}\|^2 \xrightarrow[d\to\infty]{w} R_1^2 + \cdots + R_m^2.$$

Let us now consider the case when R^2 belongs to the domain of attraction of an α-stable distribution with $\alpha \in (1,2)$ meaning that

$$\frac{R_1^2 + \dots + R_n^2 - n}{n^{1/\alpha} L(n)} \xrightarrow[n\to\infty]{w} \zeta_\alpha \tag{8}$$

for some slowly varying function L and a zero-mean random variable ζ_α having a spectrally positive α-stable distribution.

Theorem 3.5. *Suppose that* (8) *holds for some* $\alpha \in (1,2)$.

(a) *If* $n > d^{\frac{\alpha}{2\alpha-2}+\delta}$ *for some* $\delta > 0$ *and all sufficiently large* d, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

(b) *If* $n < d^{\frac{\alpha}{2\alpha-2}-\delta}$ *for some* $\delta > 0$ *and all sufficiently large* d, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{n^{1/\alpha} L(n)} \xrightarrow[d\to\infty]{w} \zeta_\alpha.$$

We shall comment on the missing "critical" case of this theorem in Remark 5.5.

3.3. **Model 3: Random walks jumping along the coordinate axes.** As we did in the previous model, here we also impose an additional assumption in the setting of Example 2.3 and suppose that the distribution of $R^{(d)}$ is the same for all $d \in \mathbb{N}$. Thus, for every $d \in \mathbb{N}$ we consider a random walk (1) in $\mathbb{R}^d$ whose increments are given by

$$X_i^{(d)} := R_i V_i^{(d)}, \quad i \in \mathbb{N},$$

where

- $(R_i)_{i=1}^\infty$ are independent copies of a random variable R with $\mathbb{E}R = 0$ and $\mathbb{E}R^2 = 1$.
- $(V_i^{(d)})_{i=1}^\infty$ are i.i.d. random vectors uniformly distributed on $\{e_1, \dots, e_d\}$, the standard orthonormal basis of $\mathbb{R}^d$. That is to say,

$$\mathbb{P}[V_i^{(d)} = e_j] = 1/d, \quad i \in \mathbb{N}, \quad j \in \{1, \dots, d\}.$$

- $(R_i)_{i=1}^\infty$ and $(V_i^{(d)})_{i=1}^\infty$ are independent.

This model is related to an experiment in which n balls are independently placed into d equiprobable boxes. If the i-th ball is placed into box j, then the i-th increment of the random walk is equal to $R_i e_j$.

Theorem 3.6. *In the setting just described suppose that* $\mathbb{E}R^4 < \infty$.

(a) *If* $\lim_{d\to\infty} n/d = 0$, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, \mathrm{Var}\,(R^2)).$$

(b) *If* $\lim_{d\to\infty} n/d = \infty$, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, 1).$$

(c) *If* $n \sim \gamma d$ *for some constant* $\gamma \in (0,\infty)$, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, 2\gamma + \mathrm{Var}\,(R^2)).$$

In the case when R^2 belongs to the domain of attraction of an α-stable distribution with $\alpha \in (1,2)$, the conclusion is identical to that of Theorem 3.5.

Theorem 3.7. *If* (8) *holds in the setting of Model 3, then the same conclusions as in Theorem 3.5 apply.*

Note that the conclusions of Theorems 3.3 and 3.6 are almost identical, the only difference being that the latter does not provide a precise answer in the case of deterministic R in the regime $\lim_{d\to\infty} n/d = 0$, since the limit in Part (a) is then degenerate. The next theorem gives a more precise result in this case. Without loss of generality, we assume that $R^2 = 1$. The latter in conjunction with $\mathbb{E}R = 0$ implies that $(S_i^{(d)})_{i=0}^\infty$ must be the simple symmetric random walk.

Theorem 3.8. *Let* $(S_i^{(d)})_{i=0}^\infty$ *be the simple symmetric random walk on* $\mathbb{Z}^d$ *starting at* 0.

(a) *If* $n = o(\sqrt{d})$, *then* $\lim_{d\to\infty} \mathbb{P}[\|S_n^{(d)}\|^2 = n] = 1$.

(b) *If* $n \sim c\sqrt{d}$ *for some constant* $c \in (0,\infty)$, *then*

$$\|S_n^{(d)}\|^2 - n \xrightarrow[d\to\infty]{w} 3P' - P'',$$

where P' *and* P'' *are independent Poisson random variables with mean* $c^2/4$.

(c) *If* $\lim_{d\to\infty} n/\sqrt{d} = \infty$, *then*

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, 1).$$

4. Proofs of the Gromov-Hausdorff convergence

The remaining part of the paper is devoted to proofs. In this section we prove Theorems 2.4 and 2.6.

4.1. **Functional law of large numbers for the norm.** We begin with a result whose proof contains the main idea of the proof of Theorem 2.4.

Theorem 4.1. *Let $n = n(d)$ be an arbitrary sequence of positive integers such that $n(d) \to \infty$, as $d \to \infty$. Under the assumptions (a)–(d) of Section 2.1,*

$$\sup_{t\in[0,1]} \left| \frac{\|S_{\lfloor nt\rfloor}^{(d)}\|^2}{n} - t \right| \underset{d\to\infty}{\overset{P}{\longrightarrow}} 0, \tag{9}$$

where $\overset{P}{\longrightarrow}$ denotes convergence in probability.

Before giving the proof of Theorem 4.1 some preparatory work has to be done. First, observe that, for every $k \in \mathbb{N}_0$,

$$\|S_k^{(d)}\|^2 = \langle S_k^{(d)}, S_k^{(d)}\rangle = \langle X_1^{(d)} + \cdots + X_k^{(d)}, X_1^{(d)} + \cdots + X_k^{(d)}\rangle = T_k^{(d)} + Q_k^{(d)}, \tag{10}$$

where

$$T_k^{(d)} := \sum_{i=1}^{k} \|X_i^{(d)}\|^2, \quad Q_k^{(d)} := \sum_{\substack{i,j\in\{1,\dots,k\}\\ i\neq j}} \langle X_i^{(d)}, X_j^{(d)}\rangle, \quad k \in \mathbb{N}, \tag{11}$$

and $T_0^{(d)} := 0$, $Q_0^{(d)} := 0$. Further, note that

$$Q_k^{(d)} = 2\sum_{i=1}^{k} Y_i^{(d)}, \quad Y_i^{(d)} := \langle X_i^{(d)}, S_{i-1}^{(d)}\rangle. \tag{12}$$

It will be of major importance for what follows that $(Q_n^{(d)})_{n\in\mathbb{N}_0}$ is a martingale. More precisely, the following holds true.

Lemma 4.2. *For any d-dimensional random walk with i.i.d. zero-mean increments $X_1^{(d)}, \dots, X_n^{(d)}$, the random variables $Y_1^{(d)}, \dots, Y_n^{(d)}$ form a triangular array of martingale differences with respect to the natural filtration $\mathcal{F}_1^{(d)} \subset \cdots \subset \mathcal{F}_n^{(d)}$, where $\mathcal{F}_i^{(d)}$ is the σ-algebra generated by $X_1^{(d)}, \dots, X_i^{(d)}$, for all $i \in \{1, \dots, n\}$.*

Proof. To prove the martingale difference property observe that $Y_i^{(d)}$ is $\mathcal{F}_i^{(d)}$-measurable and

$$\mathbb{E}\left[Y_i^{(d)} \middle| \mathcal{F}_{i-1}^{(d)}\right] = \mathbb{E}\left[\langle X_i^{(d)}, S_{i-1}^{(d)}\rangle \middle| \mathcal{F}_{i-1}^{(d)}\right] = \sum_{j=1}^{d} \mathbb{E}\left[X_{i,j}^{(d)} S_{i-1,j}^{(d)} \middle| \mathcal{F}_{i-1}^{(d)}\right] = 0,$$

for all $i = 1, \dots, n$, where we used that $S_{i-1,j}^{(d)}$ is $\mathcal{F}_{i-1}^{(d)}$-measurable and that $X_{i,j}^{(d)}$ is independent of $\mathcal{F}_{i-1}^{(d)}$ and has zero mean. $\square$

Proof of Theorem 4.1. To prove (9), it suffices to show that

$$\sup_{t\in[0,1]}\left|\frac{T^{(d)}_{\lfloor nt\rfloor}}{n}-t\right| \overset{P}{\underset{d\to\infty}{\longrightarrow}} 0, \tag{13}$$

and

$$\frac{\sup_{t\in[0,1]}|Q^{(d)}_{\lfloor nt\rfloor}|}{n} \overset{P}{\underset{d\to\infty}{\longrightarrow}} 0. \tag{14}$$

Proof of (13). According to a version of the law of large numbers stated in Lemma 6.1,

$$f_d(t) := n^{-1}T^{(d)}_{\lfloor nt\rfloor} = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor}\|X_i^{(d)}\|^2 \overset{P}{\underset{d\to\infty}{\longrightarrow}} t,$$

for every $t\geq 0$. Since the functions $t\mapsto f_d(t)$ and $t\mapsto t$ are monotone in $t\in[0,1]$ and the latter function is continuous, this convergence in probability is in fact uniform by Pólya's extension of Dini's theorem. Indeed, for every $m\in\mathbb{N}$ the union bound yields

$$\max_{i=0,\dots,m}|f_d(i/m)-(i/m)| \overset{P}{\underset{d\to\infty}{\longrightarrow}} 0.$$

The monotonicity of $t\mapsto f_d(t)$ implies that

$$\sup_{t\in[0,1]}|f_d(t)-t|\leq \max_{i=0,\dots,m}|f_d(i/m)-(i/m)|+(1/m).$$

Given $\varepsilon>0$ we choose $m\in\mathbb{N}$ such that $1/m<\varepsilon/2$. Then,

$$\mathbb{P}\left[\sup_{t\in[0,1]}|f_d(t)-t|>\varepsilon\right]\leq \mathbb{P}\left[\max_{i=0,\dots,m}|f_d(i/m)-(i/m)|>\varepsilon/2\right] \underset{d\to\infty}{\longrightarrow} 0.$$

It follows that $\sup_{t\in[0,1]}|f_d(t)-t|$ converges in probability to 0, thus proving (13).

Proof of (14). Since $(Q^{(d)}_\ell)_{\ell\in\mathbb{N}_0}$ is a martingale, for every fixed $d\in\mathbb{N}$, Doob's martingale inequality entails that

$$\mathbb{P}\Big[\sup_{t\in[0,1]}|Q^{(d)}_{\lfloor nt\rfloor}|\geq n\varepsilon\Big]\leq \frac{\mathbb{E}(Q_n^{(d)})^2}{n^2\varepsilon^2}.$$

Hence, to prove (14), it suffices to check that

$$\frac{\mathbb{E}(Q_n^{(d)})^2}{n^2}=\frac{1}{n^2}\mathbb{E}\Bigg(\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}}\langle X_i^{(d)},X_j^{(d)}\rangle\Bigg)^2=\frac{1}{n^2}\mathbb{E}\Bigg(\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}}\sum_{k=1}^{d}X_{i,k}^{(d)}X_{j,k}^{(d)}\Bigg)^2 \underset{d\to\infty}{\longrightarrow} 0, \tag{15}$$

An alternative way to see this sufficiency is to apply Corollary 2 on p. 1888 in [6] with $\Phi(x) = x^2$ and $f_{i,j}(u,v) = \langle u, v\rangle$, leading to the estimate

$$\mathbb{E}\left(\sup_{t\in[0,1]} |Q^{(d)}_{\lfloor nt\rfloor}|\right)^2 = \mathbb{E}\left(\max_{m=1,\dots,n}\left|\sum_{\substack{i,j\in\{1,\dots,m\}\\ i\neq j}} \langle X_i^{(d)}, X_j^{(d)}\rangle\right|\right)^2 \leq 2048\cdot\mathbb{E}\left(\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}} \langle X_i^{(d)}, X_j^{(d)}\rangle\right)^2.$$

In order to prove (15) we write

$$\begin{aligned}\mathbb{E}\left(\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}}\sum_{k=1}^d X_{i,k}^{(d)}X_{j,k}^{(d)}\right)^2 &= \sum_{k=1}^d\sum_{k'=1}^d\sum_{\substack{i,j\in\{1,\dots,n\}\\ i\neq j}}\sum_{\substack{i',j'\in\{1,\dots,n\}\\ i'\neq j'}}\mathbb{E}\left[X_{i,k}^{(d)}X_{i',k'}^{(d)}X_{j,k}^{(d)}X_{j',k'}^{(d)}\right]\\ &= 2\sum_{k=1}^d\sum_{i\neq j}\mathbb{E}(X_{i,k}^{(d)})^2\mathbb{E}(X_{j,k}^{(d)})^2 = 2n(n-1)\sum_{k=1}^d\mathbb{E}(X_{1,k}^{(d)})^2\mathbb{E}(X_{1,k}^{(d)})^2,\end{aligned}$$

where for the second equality we used that by independence, uncorrelatedness and $\mathbb{E}X_{i,k}^{(d)} = 0$, the expectation $\mathbb{E}[X_{i,k}^{(d)}X_{i',k'}^{(d)}X_{j,k}^{(d)}X_{j',k'}^{(d)}]$ vanishes unless $k = k'$ and $\{i,j\} = \{i',j'\}$. It remains to note that

$$\sum_{k=1}^d\mathbb{E}(X_{1,k}^{(d)})^2\mathbb{E}(X_{1,k}^{(d)})^2 \leq \max_{k=1,\dots,d}\mathbb{E}(X_{1,k}^{(d)})^2\sum_{k=1}^d\mathbb{E}(X_{1,k}^{(d)})^2 = \max_{k=1,\dots,d}\mathbb{E}(X_{1,k}^{(d)})^2 \underset{d\to\infty}{\longrightarrow} 0,$$

where (5) has been utilized on the last step. The proof of (15) is complete. □

4.2. **Proof of Theorem 2.4.** We identify the Wiener spiral $\mathbb{W}$ with the interval $[0,1]$ equipped with the metric $d(t,s) = \sqrt{|t-s|}$. Define a surjective map $\varphi_n : [0,1] \to \mathbb{M}_d$ by $\varphi_n(t) := S^{(d)}_{\lfloor nt\rfloor}/\sqrt{n}$. By Corollary 7.3.28 on page 258 of [3], the Gromov-Hausdorff distance between $\mathbb{W}$ and $\mathbb{M}_d$ is bounded above by twice the distortion of the map φ_n, that is

$$d_{\mathrm{GH}}(\mathbb{W},\mathbb{M}_d) \leq 2\sup_{0\leq s\leq t\leq 1}\left|\frac{\|S^{(d)}_{\lfloor nt\rfloor} - S^{(d)}_{\lfloor ns\rfloor}\|}{\sqrt{n}} - \sqrt{t-s}\right|.$$

To prove the theorem, it suffices to verify that

$$\sup_{0\leq s\leq t\leq 1}\left|\frac{\|S^{(d)}_{\lfloor nt\rfloor} - S^{(d)}_{\lfloor ns\rfloor}\|}{\sqrt{n}} - \sqrt{t-s}\right| \overset{P}{\underset{d\to\infty}{\longrightarrow}} 0.$$

Take some $m \in \mathbb{N}$. We know from Theorem 4.1 that, for every $i = 0,\dots,m-1$,

$$\frac{\|S^{(d)}_{\lfloor (i/m)\cdot n\rfloor}\|}{\sqrt{n}} \overset{P}{\underset{d\to\infty}{\longrightarrow}} \sqrt{\frac{i}{m}}.$$

Moreover, for every integer $0 \le i \le j \le m$,

$$\frac{\|S^{(d)}_{\lfloor (j/m)\cdot n\rfloor} - S^{(d)}_{\lfloor (i/m)\cdot n\rfloor}\|}{\sqrt{n}} \underset{d\to\infty}{\overset{P}{\longrightarrow}} \sqrt{\frac{j-i}{m}}.$$

By the union bound, it follows that, for every fixed $m \in \mathbb{N}$,

$$\max_{0\le i\le j\le m}\left|\frac{\|S^{(d)}_{\lfloor (j/m)\cdot n\rfloor} - S^{(d)}_{\lfloor (i/m)\cdot n\rfloor}\|}{\sqrt{n}} - \sqrt{\frac{j-i}{m}}\right| \underset{d\to\infty}{\overset{P}{\longrightarrow}} 0.$$

If $0 \le s \le t \le 1$ are such that $s \in [\frac{i}{m}, \frac{i+1}{m})$ and $t \in [\frac{j}{m}, \frac{j+1}{m})$, then, by the triangle inequality,

$$\left|\frac{\|S^{(d)}_{\lfloor nt\rfloor} - S^{(d)}_{\lfloor ns\rfloor}\|}{\sqrt{n}} - \frac{\|S^{(d)}_{\lfloor (j/m)\cdot n\rfloor} - S^{(d)}_{\lfloor (i/m)\cdot n\rfloor}\|}{\sqrt{n}}\right| \le \sup_{z\in[\frac{i}{m},\frac{i+1}{m}]}\frac{\|S^{(d)}_{\lfloor nz\rfloor} - S^{(d)}_{\lfloor n\cdot(i/m)\rfloor}\|}{\sqrt{n}} + \sup_{z\in[\frac{j}{m},\frac{j+1}{m}]}\frac{\|S^{(d)}_{\lfloor nz\rfloor} - S^{(d)}_{\lfloor n\cdot(j/m)\rfloor}\|}{\sqrt{n}}.$$

Consider the random variable

$$\omega_{n,m} := \max_{i\in\{0,\dots,m-1\}}\sup_{z\in[\frac{i}{m},\frac{i+1}{m}]}\frac{\|S^{(d)}_{\lfloor nz\rfloor} - S^{(d)}_{\lfloor n\cdot(i/m)\rfloor}\|}{\sqrt{n}}.$$

To complete the proof, it suffices to show that for every $\varepsilon > 0$,

$$\lim_{m\to\infty}\limsup_{d\to\infty}\mathbb{P}[\omega_{n,m} \ge \varepsilon] = 0.$$

Applying the union bound and recalling that $X_1^{(d)},\dots,X_n^{(d)}$ are i.i.d. we can write

$$\mathbb{P}[\omega_{n,m} \ge \varepsilon] \le m\mathbb{P}\left[\frac{1}{n}\sup_{t\in[0,\frac{1}{m}]}\|S^{(d)}_{\lfloor nt\rfloor}\|^2 \ge \varepsilon^2\right].$$

Recalling decomposition (11), observe that

$$\|S^{(d)}_{\lfloor nt\rfloor}\|^2 = T^{(d)}_{\lfloor nt\rfloor} + Q^{(d)}_{\lfloor nt\rfloor}.$$

To complete the proof, it suffices to verify that

$$\lim_{m\to\infty}\limsup_{d\to\infty} m\,\mathbb{P}\left[T^{(d)}_{\lfloor n/m\rfloor} \ge n\varepsilon^2/2\right] = 0, \tag{16}$$

and

$$\lim_{m\to\infty}\limsup_{d\to\infty} m\,\mathbb{P}\left[\sup_{t\in[0,\frac{1}{m}]} Q^{(d)}_{\lfloor nt\rfloor} \ge n\varepsilon^2/2\right] = 0. \tag{17}$$

Proof of (16). We observe that, for every fixed $m \in \mathbb{N}$, $T^{(d)}_{\lfloor n/m \rfloor}/n$ converges in probability to $1/m$ by the version of the law of large numbers stated in Lemma 6.1. This implies that for every $m > 2/\varepsilon^2$, the $\limsup_{d\to\infty}$ in (16) equals 0.

Proof of (17). By yet another appeal to Doob's martingale inequality we obtain

$$m\,\mathbb{P}\left[\sup_{t\in[0,\frac{1}{m}]} Q^{(d)}_{\lfloor nt \rfloor} \geq n\varepsilon^2/2\right] \leq \frac{m}{n^2(\varepsilon^2/2)^2}\mathbb{E}(Q^{(d)}_{\lfloor n/m \rfloor})^2.$$

As we have already shown in (15), for every $m \in \mathbb{N}$,

$$\lim_{d\to\infty} \frac{\mathbb{E}(Q^{(d)}_{\lfloor n/m \rfloor})^2}{n^2} = 0.$$

It follows that the $\limsup_{d\to\infty}$ in (17) equals 0 for every $m \in \mathbb{N}$. □

Proof of Corollary 2.5. Note that $\mathbb{M}_d \subset \mathbb{M}_d^{cont}$ and

$$d_{GH}(\mathbb{M}_d^{cont}, \mathbb{M}_d) \leq d_H(\mathbb{M}_d^{cont}, \mathbb{M}_d) \leq \max_{i\in\{1,\dots,n\}} \frac{\|X_i^{(d)}\|}{\sqrt{n}}.$$

The right-hand side converges to zero in probability, since, for every fixed $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}\left[\sup_{i\in\{1,\dots,n\}} \frac{\|X_i^{(d)}\|}{\sqrt{n}} > \varepsilon\right] &\leq n\mathbb{P}[\|X^{(d)}\|^2 > \varepsilon^2 n] \leq \varepsilon^{-2}\mathbb{E}[\|X^{(d)}\|^2\,\mathbb{1}_{\{\|X^{(d)}\|^2>\varepsilon^2 n\}}] \\ &\leq \varepsilon^{-2}\sup_{\ell\in\mathbb{N}}\mathbb{E}[\|X^{(\ell)}\|^2\,\mathbb{1}_{\{\|X^{(\ell)}\|^2>\varepsilon^2 n\}}],\end{aligned}$$

and the latter converges to zero by (4). □

4.3. **Proof of Theorem 2.6.** The map $K \ni t \mapsto \mathbb{X}_d(t) \in \mathbb{K}_d$ is surjective. Similarly to the proof of Theorem 2.4 we use Corollary 7.3.28 on page 258 of [3] to infer that

$$d_{\mathrm{GH}}(\mathbb{K}_d, K) \leq 2\sup_{s,t\in K}\Big|\|\mathbb{X}_d(s) - \mathbb{X}_d(t)\| - \rho(s,t)\Big|.$$

To prove the theorem it suffices to show that the right-hand side converges to 0 a.s., that is

$$\sup_{s,t\in K}\Big|\|\mathbb{X}_d(s) - \mathbb{X}_d(t)\| - \sqrt{\mathrm{Var}\,(X(s)-X(t))}\Big| \xrightarrow[d\to\infty]{a.s.} 0. \tag{18}$$

The function $z \mapsto \sqrt{z}$ is uniformly continuous on every interval of the form $[0, A]$, with $A > 0$. Therefore, for non-negative bounded functions, $f_n \to f$ uniformly implies that $\sqrt{f}_n \to \sqrt{f}$ uniformly. Hence, to prove (18), it suffices to check that

$$\lim_{d\to\infty}\sup_{s,t\in K}\Big|\|\mathbb{X}_d(s) - \mathbb{X}_d(t)\|^2 - \mathrm{Var}\,(X(s)-X(t))\Big| = 0. \tag{19}$$

Define i.i.d. stochastic processes $(Y_k(s,t))_{(s,t)\in K\times K}$, $k\in\mathbb{N}$, by

$$Y_k(s,t) = (X_k(s)-X_k(t))^2 - \mathbb{E}(X_k(s)-X_k(t))^2, \quad (s,t)\in K\times K.$$

Note that Y_k has continuous sample paths on $K\times K$ (endowed with the product metric [3, p. 88]) and that

$$\mathbb{E}\sup_{(s,t)\in K\times K}|Y_k(s,t)| \le 2\mathbb{E}\sup_{(s,t)\in K\times K}(X_k(s)-X_k(t))^2 \le 4\mathbb{E}\sup_{(s,t)\in K\times K}(X_k^2(s)+X_k^2(t)) \le 8\mathbb{E}\sup_{s\in K}X_k^2(s)<\infty.$$

Then,

$$\|\mathbb{X}_d(s)-\mathbb{X}_d(t)\|^2 - \mathrm{Var}\,(X(s)-X(t)) = \frac{1}{d}\sum_{k=1}^{d}((X_k(s)-X_k(t))^2 - \mathbb{E}(X_k(s)-X_k(t))^2) = \frac{1}{d}\sum_{k=1}^{d}Y_k(s,t).$$

Note that $Y_1, Y_2, \ldots$ are i.i.d. random elements in the Banach space $C(K\times K)$ of continuous functions on the compact space $K\times K$. As we have shown, $\mathbb{E}\|Y_k\|_\infty<\infty$. By the strong law of large numbers in the Banach space $C(K\times K)$, see Theorem 1.1 on page 131 in [12], we have

$$\sup_{(s,t)\in K\times K}\left|\frac{1}{d}\sum_{k=1}^{d}Y_k(s,t)\right| \xrightarrow[d\to\infty]{a.s.} 0.$$

This proves (19) and completes the proof of Theorem 2.6. $\square$

5. Proofs of the distributional limit theorems for the norm

5.1. **General strategy.** To prove the results stated in Section 3, recall from (10), (11), (12) the decomposition

$$\|S_n^{(d)}\|^2 - n = \langle S_n^{(d)}, S_n^{(d)}\rangle - n = \langle X_1^{(d)}+\cdots+X_n^{(d)}, X_1^{(d)}+\cdots+X_n^{(d)}\rangle - n = T_n^{(d)} - n + Q_n^{(d)}.$$

Our aim is to derive distributional limit theorems for the "diagonal sum" $T_n^{(d)}$ and the "off-diagonal sum" $Q_n^{(d)}$. For the former quantity, this task is usually straightforward since $T_n^{(d)}$ is a sum of i.i.d. random variables. Suppose that

$$\frac{T_n^{(d)}-n}{\tau_n^{(d)}} \xrightarrow[d\to\infty]{w} T_\infty \tag{20}$$

for a suitable normalizing sequence $\tau_n^{(d)}>0$ and some stable random variable T_∞. For the off-diagonal sum, we shall prove, in all three models, a central limit theorem of the form

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1). \tag{21}$$

Having (20) and (21) at our disposal, we can determine the limit distribution of $\|S_n^{(d)}\|^2 - n$. Depending on which of the normalizing sequences, $\tau_n^{(d)}$ or $\sqrt{2n^2/d}$, is asymptotically larger, we distinguish the following cases.

Case 1: Off-diagonal fluctuations dominate meaning that $\tau_n^{(d)} = o(\sqrt{n^2/d})$. Then,

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} = \frac{T_n^{(d)} - n}{\tau_n^{(d)}} \cdot \frac{\tau_n^{(d)}}{\sqrt{2n^2/d}} + \frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

Case 2: Diagonal fluctuations dominate meaning that $\sqrt{n^2/d} = o(\tau_n^{(d)})$. Then,

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} = \frac{T_n^{(d)} - n}{\tau_n^{(d)}} + \frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \cdot \frac{\sqrt{2n^2/d}}{\tau_n^{(d)}} \xrightarrow[d\to\infty]{w} T_\infty.$$

Case 3: Both types of fluctuations are of the same order meaning that $\sqrt{2n^2/d}/\tau_n^{(d)} \to c \in (0,\infty)$. This case is somewhat more difficult and requires a separate analysis.

5.2. **Central limit theorem for the off-diagonal sum.** In all three models, the proof of the CLT for $Q_n^{(d)}$ is based on the representation (12).

To prove a central limit theorem for $Q_n^{(d)}$ we are going to apply the martingale central limit theorem (see Theorem 6.2 of the Appendix) to the martingale differences

$$\Delta_i^{(d)} := \frac{Y_i^{(d)}}{\sqrt{n^2/(2d)}}, \quad i = 1, \dots, n.$$

If the conditions of Theorem 6.2 are satisfied with $\sigma^2 = 1$, then

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} = \Delta_1^{(d)} + \cdots + \Delta_n^{(d)} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

In the following two lemmas we simultaneously verify condition (53) of Theorem 6.2 for all three models defined in Sections 3.1, 3.2, 3.3.

Lemma 5.1. *Consider a d-dimensional random walk $(S_i^{(d)})_{i=0}^\infty$ with i.i.d. zero-mean increments $X_1^{(d)}, X_2^{(d)}, \dots$ satisfying*

$$\mathbb{E}[X_{i,j}^{(d)} X_{i,k}^{(d)}] = 0, \quad \mathbb{E}(X_{i,j}^{(d)})^2 = 1/d, \quad j,k \in \{1,\dots,d\}, \quad j \neq k, \quad i \in \mathbb{N}. \tag{22}$$

Then, for all $n \in \mathbb{N}$,

$$D_n^{(d)} := \sum_{i=1}^n \mathbb{E}[(Y_i^{(d)})^2 | \mathcal{F}_{i-1}^{(d)}] = \frac{1}{d} \sum_{i=1}^{n-1} \|S_i^{(d)}\|^2, \tag{23}$$

and

$$\mathbb{E}D_n^{(d)} = \sum_{i=1}^{n} \mathbb{E}(Y_i^{(d)})^2 = \frac{n(n-1)}{2d}, \quad \operatorname{Var} Q_n^{(d)} = 4\mathbb{E}D_n^{(d)} = \frac{2n(n-1)}{d}. \tag{24}$$

Proof. To prove (23), observe that

$$\begin{aligned} D_n^{(d)} &= \sum_{i=1}^{n} \mathbb{E}\left[\langle X_i^{(d)}, S_{i-1}^{(d)}\rangle^2 \middle| \mathcal{F}_{i-1}^{(d)}\right] = \sum_{i=1}^{n} \mathbb{E}\left[\left(\sum_{j=1}^{d} X_{i,j}^{(d)} S_{i-1,j}^{(d)}\right)^2 \middle| \mathcal{F}_{i-1}^{(d)}\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\sum_{k=1}^{d}\sum_{\ell=1}^{d} X_{i,k}^{(d)} X_{i,\ell}^{(d)} S_{i-1,k}^{(d)} S_{i-1,\ell}^{(d)} \middle| \mathcal{F}_{i-1}^{(d)}\right] = \sum_{i=1}^{n}\sum_{k=1}^{d}\sum_{\ell=1}^{d} S_{i-1,k}^{(d)} S_{i-1,\ell}^{(d)} \mathbb{E}\left[X_{i,k}^{(d)} X_{i,\ell}^{(d)}\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{d} (S_{i-1,j}^{(d)})^2 \mathbb{E}(X_{i,j}^{(d)})^2 = \frac{1}{d}\sum_{i=1}^{n}\sum_{j=1}^{d} (S_{i-1,j}^{(d)})^2 = \frac{1}{d}\sum_{i=1}^{n-1} \|S_{i,j}^{(d)}\|^2, \end{aligned}$$

where we used (22). To prove the first equation in (24), take expectation of (23) and observe that $\mathbb{E}\|S_i^{(d)}\|^2 = i$. To prove the second equation in (24), recall (12) and observe that $Y_1^{(d)}, \dots, Y_n^{(d)}$, being martingale differences, are uncorrelated. □

Lemma 5.2. *For a random walk satisfying conditions (a), (c), (d) of Section 2.1 and* (22) *we have*

$$\sum_{i=1}^{n} \mathbb{E}[(\Delta_i^{(d)})^2 | \mathcal{F}_{i-1}^{(d)}] = \frac{D_n^{(d)}}{n^2/(2d)} \underset{d\to\infty}{\overset{P}{\longrightarrow}} 1.$$

Proof. We know from Theorem 4.1 that

$$\max_{i\in\{1,\dots,n\}} \frac{1}{n}\left|\|S_i^{(d)}\|^2 - i\right| \underset{d\to\infty}{\overset{P}{\longrightarrow}} 0.$$

Taking some $\varepsilon > 0$ and denoting by $A_n^{(d)}$ the event that $\max_{i\in\{1,\dots,n\}} |\|S_i^{(d)}\|^2 - i| \le n\varepsilon$, we have that $\mathbb{P}[A_n^{(d)}] \to 1$ as $d \to \infty$. On the event $A_n^{(d)}$ we have the upper bound

$$D_n^{(d)} = \frac{1}{d}\sum_{i=1}^{n-1} \|S_i^{(d)}\|^2 \le \frac{1}{d}\sum_{i=1}^{n-1}(i + n\varepsilon) = \frac{n(n-1) + 2\varepsilon n(n-1)}{2d}$$

and the lower bound

$$D_n^{(d)} = \frac{1}{d}\sum_{i=1}^{n-1} \|S_i^{(d)}\|^2 \ge \frac{1}{d}\sum_{i=1}^{n-1}(i - n\varepsilon) = \frac{n(n-1) - 2\varepsilon n(n-1)}{2d}.$$

Taken together, these bounds imply the claim. □

5.3. **Model 1: Proofs of Theorems 3.1 and 3.2.** The main difficulty is to prove the following central limit theorem for $Q_n^{(d)}$.

Proposition 5.3. *In the setting of Section 3.1 suppose that $\mathbb{E}|\xi|^{2+\delta} < \infty$, for some $\delta > 0$. Then,*

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

Proof. Since condition (53) of the martingale central limit theorem (see Theorem 6.2) has already been verified in Lemma 5.2, it remains to verify Lyapunov's condition (56) which takes the form

$$\sum_{i=1}^{n} \mathbb{E}|Y_i^{(d)}|^{2+\delta} = o(n^{2+\delta}/d^{1+\frac{\delta}{2}}),$$

where $\delta > 0$ is such that $\mathbb{E}|\xi|^{2+\delta} < \infty$. To prove this estimate, it suffices to show that

$$\max_{i\in\{1,\dots,n\}} \mathbb{E}|Y_i^{(d)}|^{2+\delta} \le C(n/d)^{1+\frac{\delta}{2}}.$$

In the following, C denotes a sufficiently large constant that does not depend on d. Recall from Section 3.1 that

$$Y_i^{(d)} = \langle X_i^{(d)}, S_{i-1}^{(d)}\rangle = \frac{1}{d}\sum_{j=1}^{d} \xi_{i,j}(\xi_{1,j} + \dots + \xi_{i-1,j}) =: \frac{1}{d}\sum_{j=1}^{d} \xi_{i,j}\eta_{i-1,j},$$

where we defined $\eta_{i-1,j} := \xi_{1,j} + \dots + \xi_{i-1,j}$. The Rosenthal inequality, see Theorem 6.3, implies that

$$\mathbb{E}|Y_i^{(d)}|^{2+\delta} = d^{-2-\delta}\mathbb{E}\left|\sum_{j=1}^{d}\xi_{i,j}\eta_{i-1,j}\right|^{2+\delta} \le Cd^{-2-\delta}\max\left\{\sum_{j=1}^{d}\mathbb{E}|\xi_{i,j}\eta_{i-1,j}|^{2+\delta}, \left(\sum_{j=1}^{d}\mathbb{E}(\xi_{i,j}\eta_{i-1,j})^2\right)^{1+\frac{\delta}{2}}\right\},$$

for all $i \in \{1,\dots,n\}$. In the following, we estimate both terms appearing on the right-hand side. For the first term, we first recall that $\xi_{i,j}$ has finite moment of order $(2+\delta)$:

$$\sum_{j=1}^{d}\mathbb{E}|\xi_{i,j}\eta_{i-1,j}|^{2+\delta} \le C\sum_{j=1}^{d}\mathbb{E}|\eta_{i-1,j}|^{2+\delta}.$$

For each summand on the right-hand side we use the Rosenthal inequality to obtain

$$\mathbb{E}\left|\eta_{i-1,j}\right|^{2+\delta} = \mathbb{E}\left|\sum_{\ell=1}^{i-1}\xi_{\ell,j}\right|^{2+\delta} \le C\max\{i, i^{1+\frac{\delta}{2}}\} \le Cn^{1+\frac{\delta}{2}}.$$

Therefore,

$$\sum_{j=1}^{d}\mathbb{E}|\xi_{i,j}\eta_{i-1,j}|^{2+\delta} \le C\cdot d\cdot n^{1+\frac{\delta}{2}}.$$

To estimate the second term, we observe that $\mathbb{E}(\xi_{i,j}\eta_{i-1,j})^2 = \mathbb{E}(\eta_{i-1,j})^2 = i-1 < n$. It follows that

$$\left(\sum_{j=1}^{d} \mathbb{E}(\xi_{i,j}\eta_{i-1,j})^2\right)^{1+\frac{\delta}{2}} \le (dn)^{1+\frac{\delta}{2}}.$$

Altogether we arrive at

$$\mathbb{E}|Y_i^{(d)}|^{2+\delta} \le Cd^{-2-\delta}(dn)^{1+\frac{\delta}{2}} = C(n/d)^{1+\frac{\delta}{2}},$$

which proves the claim. □

Proof of Theorem 3.1. By Proposition 5.3, the off-diagonal sum satisfies

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

To derive a distributional limit theorem for the diagonal sum, we observe that

$$T_n^{(d)} - n = \sum_{i=1}^{n}(\|X_i^{(d)}\|^2 - 1) = \frac{1}{d}\sum_{i=1}^{n}\sum_{j=1}^{d}(\xi_{i,j}^2 - 1) \tag{25}$$

Recall the assumption $\mathbb{E}\xi^4 < \infty$. Applying the classical CLT to the right-hand side of (25) yields

$$\frac{T_n^{(d)} - n}{\sqrt{n/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, \mathrm{Var}\,[\xi^2]).$$

Since $\sqrt{n/d} = o(\sqrt{2n^2/d})$, the fluctuations of the off-diagonal sum $Q_n^{(d)}$ dominate. □

Proof of Theorem 3.2. By (25) and (7), we have

$$\frac{T_n^{(d)} - n}{d^{-1}(nd)^{1/\alpha}L(nd)} = \frac{1}{(nd)^{1/\alpha}L(nd)}\sum_{i=1}^{n}\sum_{j=1}^{d}(\xi_{i,j}^2 - 1) \xrightarrow[d\to\infty]{w} \zeta_\alpha.$$

The normalizing sequence for $T_n^{(d)} - n$ is thus $\tau_n^{(d)} = d^{-1}(nd)^{1/\alpha}L(nd)$. If, for some $\delta > 0$ and all sufficiently large d, $n > d^{\frac{2-\alpha}{2\alpha-2}+\delta}$, respectively, $n < d^{\frac{2-\alpha}{2\alpha-2}-\delta}$, then $\sqrt{n^2/d} = o(\tau_n^{(d)})$ (meaning that the fluctuations of $T_n^{(d)}$ dominate), respectively, $\tau_n^{(d)} = o(\sqrt{2n^2/d})$ (meaning that the fluctuations of $Q_n^{(d)}$ dominate).

It is also clear that, in fact, a more precise result has been deduced. Namely, if

$$\lim_{d\to\infty} n^{1/\alpha-1}d^{1/\alpha-1/2}L(nd) = 0,$$

then the convergence in Part (a) holds true, whereas if the above limit is equal to $+\infty$, the convergence in Part (b) holds true. □

5.4. **Model 2: Proofs of Theorems 3.3 and 3.5.** The main difficulty is again to prove the CLT for the off-diagonal sum.

Proposition 5.4. *In addition to the setting of Section 3.2 suppose that* $\mathbb{E}R^{2+\delta} < \infty$ *for some* $\delta > 0$. *Then,*

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0,1).$$

Proof. We again apply the martingale central limit theorem. Condition (53) of Theorem 6.2 has been verified in Lemma 5.2. We shall verify the Lyapunov condition (55) which takes the form

$$\frac{1}{(n/\sqrt{d})^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|Y_i^{(d)}|^{2+\delta} | \mathcal{F}_{i-1}^{(d)}\right] \xrightarrow[d\to\infty]{P} 0. \tag{26}$$

Recall from Section 3.2 that $X_i = R_i U_i^{(d)}$, where $R_i \geq 0$, $U_i^{(d)}$ and $S_{i-1}^{(d)}$ are independent. It follows that

$$\begin{aligned}
\mathbb{E}[|Y_i^{(d)}|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] &= \mathbb{E}[|\langle X_i, S_{i-1}^{(d)}\rangle|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] = \mathbb{E}[|\langle R_i U_i^{(d)}, S_{i-1}^{(d)}\rangle|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] \\
&= \mathbb{E}|R|^{2+\delta} \cdot \mathbb{E}[|\langle U_i^{(d)}, S_{i-1}^{(d)}\rangle|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] = \mathbb{E}|R|^{2+\delta} \cdot |S_{i-1}^{(d)}|^{2+\delta} \cdot \mathbb{E}|\langle U_i^{(d)}, e_1\rangle|^{2+\delta} \\
&\leq C \cdot d^{-1-\frac{\delta}{2}} \cdot |S_{i-1}^{(d)}|^{2+\delta},
\end{aligned}$$

where in the penultimate step we used the isotropy of $U_i^{(d)}$. In the last step, we used that $\sqrt{d}\langle U_i^{(d)}, e_1\rangle$ converges to the standard normal distribution together with all moments [18, Theorem 1] and, consequently, $\mathbb{E}|\langle U_i^{(d)}, e_1\rangle|^{2+\delta} \leq C d^{-1-\frac{\delta}{2}}$. For the Lyapunov sum we obtain the estimate

$$\sum_{i=1}^{n} \mathbb{E}[|Y_i^{(d)}|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] \leq C \cdot d^{-1-\frac{\delta}{2}} \cdot \sum_{i=1}^{n} |S_i^{(d)}|^{2+\delta}.$$

We know from Theorem 4.1 that the event

$$A_n^{(d)} := \left\{ \max_{i=1,\dots,n} \left| \|S_i^{(d)}\|^2 - i \right| \leq n \right\}$$

satisfies $\lim_{d\to\infty} \mathbb{P}[A_n^{(d)}] = 1$. So, on the event $A_n^{(d)}$ we have $\|S_i^{(d)}\|^2 \leq 2n$ for all $i = 1, \dots, n$ and $\sum_{i=1}^{n-1} |S_i^{(d)}|^{2+\delta} \leq n^{2+(\delta/2)}$. It follows that, on $A_n^{(d)}$,

$$\sum_{i=1}^{n} \mathbb{E}[|Y_i^{(d)}|^{2+\delta}|\mathcal{F}_{i-1}^{(d)}] \leq C \cdot d^{-1-\frac{\delta}{2}} \cdot \sum_{i=1}^{n} |S_i^{(d)}|^{2+\delta} \leq C \cdot d^{-1-\frac{\delta}{2}} \cdot n^{2+\frac{\delta}{2}} = o((n/\sqrt{d})^{2+\delta}). \tag{27}$$

This proves (26). $\square$

Proof of Theorem 3.3. By the classical CLT and Proposition 5.4,

$$\frac{T_n^{(d)}-n}{\sqrt{n}}=\frac{\sum_{i=1}^n(\|X_i^{(d)}\|^2-1)}{\sqrt{n}}=\frac{\sum_{i=1}^n(R_i^2-1)}{\sqrt{n}}\overset{w}{\underset{d\to\infty}{\longrightarrow}}\mathrm{N}(0,\mathrm{Var}\,[R^2]),\qquad\frac{Q_n^{(d)}}{\sqrt{2n^2/d}}\overset{w}{\underset{d\to\infty}{\longrightarrow}}\mathrm{N}(0,1).\tag{28}$$

Proof of (a): If $n/d\to 0$, then $\sqrt{2n^2/d}=o(\sqrt{n})$ and (since $\mathrm{Var}\,[R^2]\neq 0$) the diagonal sum $T_n^{(d)}$ dominates.

Proof of (b): If $n/d\to\infty$, then $\sqrt{n}=o(\sqrt{2n^2/d})$ and the off-diagonal sum $Q_n^{(d)}$ dominates. The same conclusion applies if $R=1$ is deterministic since then the diagonal sum equals n.

Proof of (c). The proof in the "critical case" when $n\sim\gamma d$ follows essentially the same idea as described in Sections 5.1 and 5.2, but requires more refined estimates. We start with the decomposition

$$\frac{\|S_n^{(d)}\|^2-n}{\sqrt{n}}=\sum_{i=1}^n\Delta_i^{(d)},\qquad\Delta_i^{(d)}:=\frac{Y_i^{(d)}}{\sqrt{n}},\qquad Y_i^{(d)}:=R_i^2-1+2R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle.$$

The sequence $\Delta_1^{(d)},\dots,\Delta_n^{(d)}$ forms a martingale difference since

$$\mathbb{E}[Y_i^{(d)}|\mathcal{F}_{i-1}]=\mathbb{E}[(R_i^2-1)|\mathcal{F}_{i-1}]+\mathbb{E}R_i\cdot\mathbb{E}[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\mathcal{F}_{i-1}]=0,$$

where we used that $\mathbb{E}\langle U_i^{(d)},x\rangle=0$ for every fixed vector $x\in\mathbb{R}^d$. The latter relation and $\mathbb{E}[\langle U_i^{(d)},x\rangle^2]=\|x\|^2/d$ imply

$$\begin{aligned}\mathbb{E}[(Y_i^{(d)})^2|\mathcal{F}_{i-1}]&=\mathbb{E}[(R_i^2-1)^2]+4\mathbb{E}[R_i(R_i^2-1)]\mathbb{E}[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\mathcal{F}_{i-1}]+4\mathbb{E}[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2|\mathcal{F}_{i-1}]\\&=\mathrm{Var}\,(R^2)+\frac{4}{d}\|S_{i-1}^{(d)}\|^2.\end{aligned}$$

Thus, it follows that

$$\sum_{i=1}^n\mathbb{E}\left[\left(\Delta_i^{(d)}\right)^2|\mathcal{F}_{i-1}^{(d)}\right]=\mathrm{Var}\,(R^2)+\frac{4}{dn}\sum_{i=1}^n\|S_{i-1}^{(d)}\|^2\overset{P}{\underset{d\to\infty}{\longrightarrow}}\mathrm{Var}\,(R^2)+2\gamma,$$

where we utilized that $\frac{2}{n^2}\sum_{i=1}^n\|S_{i-1}^{(d)}\|^2$ converges in probability to 1, which can be verified in the same way as in the proof of Lemma 5.2. It remains to check the Lindeberg condition (54) which takes the following form. For every $\varepsilon>0$,

$$\frac{1}{n}\sum_{i=1}^n\mathbb{E}\left[\left(R_i^2-1+2R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle\right)^2\mathbb{1}_{\left\{|R_i^2-1+2R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\geq\varepsilon\sqrt{n}\right\}}\Big|\mathcal{F}_{i-1}^{(d)}\right]\overset{P}{\underset{d\to\infty}{\longrightarrow}}0.\tag{29}$$

By the estimate $(a+b)^2 \le 2a^2 + 2b^2$ and Markov's inequality it suffices to verify the following claims:

$$\lim_{d\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[(R_i^2-1)^2\,\mathbb{1}_{\{|R_i^2-1|\ge\frac{1}{2}\varepsilon\sqrt{n}\}}\right] = 0, \tag{30}$$

$$\lim_{d\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[(R_i^2-1)^2\,\mathbb{1}_{\left\{|R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge\frac{1}{4}\varepsilon\sqrt{n}\right\}}\right] = 0, \tag{31}$$

$$\lim_{d\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\{|R_i^2-1|\ge\frac{1}{2}\varepsilon\sqrt{n}\}}\right] = 0, \tag{32}$$

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\left\{|R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge\frac{1}{4}\varepsilon\sqrt{n}\right\}}\,\middle|\,\mathcal{F}_{i-1}^{(d)}\right] \xrightarrow[d\to\infty]{P} 0. \tag{33}$$

Condition (30) is fulfilled by the monotone convergence theorem since the R_i's are independent copies of R and $\mathbb{E}R^4 < \infty$. To prove the remaining conditions we first observe that

$$\mathbb{E}\left[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\right] = \mathbb{E}\left[\mathbb{E}\left[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2|S_{i-1}^{(d)}\right]\right] = \frac{1}{d}\mathbb{E}\|S_{i-1}^{(d)}\|^2 = \frac{i-1}{d} \le \frac{n}{d} \le C, \tag{34}$$

for all $i\in\{1,\dots,n\}$. Recall also that R_i, $U_i^{(d)}$ and $S_{i-1}^{(d)}$ are independent. To prove (31), note that

$$\begin{aligned}\mathbb{E}\left[(R_i^2-1)^2\,\mathbb{1}_{\left\{|R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge\frac{1}{4}\varepsilon\sqrt{n}\right\}}\right] &\le \mathbb{E}\left[(R_i^2-1)^2\,\mathbb{1}_{\{R_i\ge\frac{1}{4}\varepsilon n^{1/4}\}}\right] + \mathbb{E}\left[(R_i^2-1)^2\,\mathbb{1}_{\left\{|\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge n^{1/4}\right\}}\right]\\ &\le \mathbb{E}\left[(R^2-1)^2\,\mathbb{1}_{\{|R|\ge\frac{1}{4}\varepsilon n^{1/4}\}}\right] + \mathbb{E}\left[(R^2-1)^2\right]\mathbb{P}\left[|\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge n^{1/4}\right].\end{aligned}$$

Both summands on the right-hand side go to 0 (for the second summand this follows from (34) and Markov's inequality). To prove condition (32), we note that

$$\begin{aligned}&\mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\{|R_i^2-1|\ge\frac{1}{2}\varepsilon\sqrt{n}\}}\right]\\ &\qquad= \mathbb{E}\left[R_i^2\,\mathbb{1}_{\{|R_i^2-1|\ge\frac{1}{2}\varepsilon\sqrt{n}\}}\right]\cdot\mathbb{E}\left[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\right] \le \mathbb{E}\left[R^2\,\mathbb{1}_{\{|R^2-1|\ge\frac{1}{2}\varepsilon\sqrt{n}\}}\right]\cdot C\end{aligned}$$

and observe that the right-hand side goes to 0 as $n\to\infty$ by the monotone convergence theorem. To prove (33), we argue as follows:

$$\begin{aligned}&\mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\left\{|R_i\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge\frac{1}{4}\varepsilon\sqrt{n}\right\}}\,\middle|\,\mathcal{F}_{i-1}^{(d)}\right]\\ &\quad\le \mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\{|R_i|\ge\frac{1}{4}\varepsilon n^{1/4}\}}\,\middle|\,\mathcal{F}_{i-1}^{(d)}\right] + \mathbb{E}\left[R_i^2\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\left\{|\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge n^{1/4}\right\}}\,\middle|\,\mathcal{F}_{i-1}^{(d)}\right]\\ &\quad\le \frac{1}{d}\|S_{i-1}^{(d)}\|^2\cdot\mathbb{E}\left[R^2\,\mathbb{1}_{\{|R|\ge\frac{1}{4}\varepsilon n^{1/4}\}}\right] + \mathbb{E}\left[\langle U_i^{(d)},S_{i-1}^{(d)}\rangle^2\,\mathbb{1}_{\left\{|\langle U_i^{(d)},S_{i-1}^{(d)}\rangle|\ge n^{1/4}\right\}}\,\middle|\,\mathcal{F}_{i-1}^{(d)}\right].\end{aligned}$$

The expectation of the first summand can be bounded above by $C\mathbb{E}[R^2 \mathbb{1}_{\{|R|\geq \frac{1}{4}\varepsilon n^{1/4}\}}]$ uniformly over $i \in \{1,\dots,n\}$, which goes to 0 by the monotone convergence theorem. To bound the second summand, we observe that, for every $\delta > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\langle U_i^{(d)}, S_{i-1}^{(d)}\rangle^2 \mathbb{1}_{\left\{|\langle U_i^{(d)}, S_{i-1}^{(d)}\rangle|\geq n^{1/4}\right\}}\middle|\mathcal{F}_{i-1}^{(d)}\right] \leq n^{-1-\frac{\delta}{4}}\sum_{i=1}^{n}\mathbb{E}\left[|\langle U_i^{(d)}, S_{i-1}^{(d)}\rangle|^{2+\delta}\middle|\mathcal{F}_{i-1}^{(d)}\right]$$
$$\leq C\cdot n^{-1-\frac{\delta}{4}}\cdot d^{-1-\frac{\delta}{2}}\cdot n^{2+\frac{\delta}{2}},$$

where the last inequality holds on the event $A_n^{(d)}$, as we have shown in the proof of Proposition 5.4, see formula (27). The right-hand side goes to 0 if $n \sim \gamma d$, and we have $\mathbb{P}[A_n^{(d)}] \to 1$, which completes the verification of (33) and the Lindeberg condition (29). An appeal to the martingale CLT stated in Theorem 6.2 completes the proof of Part (c). □

Proof of Theorem 3.5. It follows from (8) and Proposition 5.4 that

$$\frac{T_n^{(d)} - n}{n^{1/\alpha}L(n)} = \frac{R_1^2 + \dots + R_n^2 - n}{n^{1/\alpha}L(n)} \underset{n\to\infty}{\overset{w}{\longrightarrow}} \zeta_\alpha \quad \text{and} \quad \frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \underset{d\to\infty}{\overset{w}{\longrightarrow}} \mathrm{N}(0,1). \tag{35}$$

If, for some $\delta > 0$ and all sufficiently large d, we have $n > d^{\frac{\alpha}{2\alpha-2}+\delta}$, respectively, $n < d^{\frac{\alpha}{2\alpha-2}-\delta}$, then $n^{1/\alpha}L(n) = o(\sqrt{2n^2/d})$, respectively, $\sqrt{2n^2/d} = o(n^{1/\alpha}L(n))$, and the claims of (a) and (b) follow. A similar observation as at the end of the proof of Theorem 3.2 applies here. Namely, if

$$\lim_{d\to\infty} d^{1/2}n^{1/\alpha-1}L(n) = 0,$$

then the convergence in Part (a) holds true, whereas if the above limit is equal to $+\infty$, then the convergence in Part (b) holds true. □

Remark 5.5. In the missing critical case of Theorem 3.5, i.e. when (8) holds and $\sqrt{2n^2/d} \sim \gamma n^{1/\alpha}L(n)$ for some constant $\gamma \in (0,\infty)$, we conjecture that

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} \underset{d\to\infty}{\overset{w}{\longrightarrow}} N + \gamma^{-1}\zeta_\alpha, \tag{36}$$

where N has the standard normal law, N and ζ_α are independent. Let us explain the intuition behind this conjecture (in fact, similar arguments apply to all cases of Theorems 3.3 and 3.5). It is known, see Theorem 4 in [18], that for every fixed $m \in \mathbb{N}$, the collection of $m(m-1)/2$ random variables $\sqrt{d}\cdot(\langle U_i^{(d)}, U_j^{(d)}\rangle)_{1\leq i<j\leq m}$ converges in distribution to a collection of i.i.d. standard normal variables $(N_{i,j})_{1\leq i<j\leq m}$. This suggests the approximation

$$\|S_n^{(d)}\|^2 - n = \sum_{i=1}^{n}(R_i^2 - 1) + 2\sum_{1\leq i<j\leq n} R_iR_j\langle U_i^{(d)}, U_j^{(d)}\rangle \approx \sum_{i=1}^{n}(R_i^2 - 1) + \frac{2}{\sqrt{d}}\sum_{1\leq i<j\leq n} R_iR_jN_{i,j}. \tag{37}$$

Conditionally on $R_1, R_2, \dots$, the distribution of the second term term is centered normal with the variance

$$\frac{2}{d}\left(\sum_{i=1}^{n} R_i^2\right)^2 - \frac{2}{d}\sum_{i=1}^{n} R_i^4.$$

By the law of large numbers, $(\sum_{i=1}^n R_i^2)^2 \sim n^2$ a.s., while $\sum_{i=1}^n R_i^4 = o(n^2)$ since R_i^4 is in the domain of attraction of an $\alpha/2$-stable distribution with $\alpha/2 > 1/2$. Hence, the variance of the normal distribution is asymptotic to $2n^2/d$. We see that the fluctuations of the first term on the right-hand side of (37) are determined by the R_i's, while the fluctuations of the second term are determined by the $N_{i,j}$'s only. Hence, these fluctuations are asymptotically independent. Recalling (8) for the first term, we arrive at (36).

5.5. **Model 3: Proofs of Theorems 3.6, 3.7, 3.8.** The random walk described in Model 3 can be coupled with the classical allocation scheme [15] in which n balls are independently dropped into d equiprobable boxes. Each time a random walk makes a jump along the line spanned by the basis vector e_j, we drop a ball into the box with the number j. Let

$$k_j(\ell) = \sum_{i=1}^{\ell} \mathbb{1}_{\{V_i^{(d)} = e_j\}}$$

be the number of balls in box $j \in \{1, \dots, d\}$ after $\ell \in \mathbb{N}$ balls have been placed into boxes. Let $(R_{i;j})_{i,j\in\mathbb{N}}$ be independent copies of the random variable R and consider independent random walks $(Z_{k;j})_{k\in\mathbb{N}_0}$, $j \in \mathbb{N}$, defined by

$$Z_{k;j} := R_{1;j} + \dots + R_{k;j}, \quad Z_{0;j} := 0, \quad k \in \mathbb{N}, \quad j \in \mathbb{N}.$$

Then, it follows from the definition of Model 3 given in Section 3.3 that

$$S_n^{(d)} = (S_{n,1}^{(d)}, \dots, S_{n,d}^{(d)}) \stackrel{d}{=} \left(Z_{k_1(n);1}, \dots, Z_{k_d(n);d}\right). \tag{38}$$

In particular, this shows that $\|S_n^{(d)}\|^2$ is a particular case of the so-called randomized decomposable statistics whose limit behaviour has been extensively studied. A survey on this topic with pointers to the original literature including [5] and the thesis of S. I. Bykov [4] can be found in [17]. It would be possible to prove most of Theorems 3.6 and 3.8 by verifying the (quite technical) conditions of Theorems 1.2.1 and 1.3.1 in [17] (which are due to S. I. Bykov), but we prefer to give independent proofs since these are quite simple. We begin with a CLT for the off-diagonal sum.

Proposition 5.6. *In addition to the setting of Section 3.3 suppose that $\mathbb{E}R^{2+\delta} < \infty$ and $n/\sqrt{d} \to \infty$. Then,*

$$\frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \underset{d\to\infty}{\overset{w}{\longrightarrow}} \mathrm{N}(0,1).$$

Proof. We again apply the martingale central limit theorem; see Theorem 6.2. Its condition (53) has been verified in Lemma 5.2. It suffices to verify the Lyapunov condition (56) which takes the form

$$\sum_{i=1}^{n} \mathbb{E}|Y_i^{(d)}|^{2+\delta} = o(n^{2+\delta}/d^{1+\frac{\delta}{2}}), \tag{39}$$

where $\delta > 0$ is such that $\mathbb{E}R^{2+\delta} < \infty$. Without loss of generality we assume that $\delta < 2$.

Let $i \in \{1, \dots, n\}$ be fixed. By definition of our model, see Section 3.3,

$$\begin{aligned}\mathbb{E}|Y_i^{(d)}|^{2+\delta} &= \mathbb{E}|\langle R_i V_i^{(d)}, S_{i-1}^{(d)}\rangle|^{2+\delta} = \mathbb{E}|R|^{2+\delta} \cdot \mathbb{E}|\langle V_i^{(d)}, S_{i-1}^{(d)}\rangle|^{2+\delta} \\ &= \mathbb{E}|R|^{2+\delta} \cdot \frac{1}{d} \cdot \sum_{j=1}^{d} \mathbb{E}|S_{i-1,j}^{(d)}|^{2+\delta} = \mathbb{E}|R|^{2+\delta} \cdot \mathbb{E}|S_{i-1,1}^{(d)}|^{2+\delta} = C \cdot \mathbb{E}|S_{i-1,1}^{(d)}|^{2+\delta},\end{aligned}$$

where we recall the notation $S_{i-1}^{(d)} = (S_{i-1,1}^{(d)}, \dots, S_{i-1,d}^{(d)})$ for the components of $S_{i-1}^{(d)}$. In view of (38)

$$\mathbb{E}|S_{i-1,1}^{(d)}|^{2+\delta} = \mathbb{E}|Z_{k_1(i-1);1}|^{2+\delta} \leq C\mathbb{E}(k_1(i-1))^{1+\delta/2},$$

where we used Rosenthal's inequality in the last estimate. Taking everything together, we arrive at

$$\sum_{i=1}^{n} \mathbb{E}|Y_i^{(d)}|^{2+\delta} \leq C \cdot \sum_{i=0}^{n-1} \mathbb{E}|S_{i,1}^{(d)}|^{2+\delta} \leq C \cdot \sum_{i=1}^{n-1} \mathbb{E}(k_1(i))^{1+\frac{\delta}{2}}. \tag{40}$$

Note that $k_1(i)$ has a binomial distribution $\mathrm{Bin}(i, 1/d)$. We claim that, for all $i, d \in \mathbb{N}$ and $\delta \in (0, 2]$,

$$\mathbb{E}(k_1(i))^{1+\frac{\delta}{2}} \leq C \cdot (i/d) + C \cdot (i/d)^{1+\frac{\delta}{2}}. \tag{41}$$

Observe that $\mathbb{E}k_1(i) = i/d$. Using the inequality $|a+b|^{1+\frac{\delta}{2}} \leq 2^{\delta/2}(|a|^{1+\frac{\delta}{2}} + |b|^{1+\frac{\delta}{2}})$, we obtain

$$\mathbb{E}(k_1(i))^{1+\frac{\delta}{2}} = \mathbb{E}|k_1(i) - i/d + i/d|^{1+\frac{\delta}{2}} \leq 2^{\delta/2}\mathbb{E}|k_1(i) - i/d|^{1+\frac{\delta}{2}} + 2^{\delta/2}(i/d)^{1+\frac{\delta}{2}}.$$

We can write $k_1(i) - i/d = \varepsilon_1 + \dots + \varepsilon_i$, where $\varepsilon_1, \dots, \varepsilon_d$ are zero-mean i.i.d. with $\mathbb{P}[\varepsilon_i = 1 - 1/d] = 1/d$, $\mathbb{P}[\varepsilon_i = -1/d] = 1 - 1/d$. By Corollary 8.2 on p. 151 of [10], we have

$$\mathbb{E}|k_1(i) - i/d|^{1+\frac{\delta}{2}} = \mathbb{E}|\varepsilon_1 + \dots + \varepsilon_i|^{1+\frac{\delta}{2}} \leq Ci\mathbb{E}|\varepsilon_1|^{1+\frac{\delta}{2}} \leq Ci/d.$$

This proves (41). Now we can complete the proof of (39) as follows. By (40) and (41),

$$\sum_{i=1}^{n} \mathbb{E}|Y_i^{(d)}|^{2+\delta} \le C \cdot \sum_{i=1}^{n-1} \mathbb{E}(k_1(i))^{1+\frac{\delta}{2}} \le C \cdot n \cdot \Big((n/d) + (n/d)^{1+\frac{\delta}{2}}\Big) = C \cdot \left(\frac{n^2}{d} + \frac{n^{2+\frac{\delta}{2}}}{d^{1+\frac{\delta}{2}}}\right) = o\left(\frac{n^{2+\delta}}{d^{1+\frac{\delta}{2}}}\right)$$

by the assumption $n/\sqrt{d} \to \infty$. The proof of (39) is complete. □

Proof of Theorem 3.6. By the classical CLT,

$$\frac{T_n^{(d)} - n}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\|X_i^{(d)}\|^2 - 1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (R_i^2 - 1) \xrightarrow[d\to\infty]{w} \mathrm{N}(0, \mathrm{Var}\,[R^2]). \tag{42}$$

Proof of (a). If $n/d \to 0$, then Lemma 5.1 yields $\mathrm{Var}\, Q_n^{(d)} = 2n(n-1)/d = o(n)$ and hence

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} = \frac{T_n^{(d)} - n}{\sqrt{n}} + \frac{Q_n^{(d)}}{\sqrt{2n(n-1)/d}} \cdot \frac{\sqrt{2n(n-1)/d}}{\sqrt{n}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, \mathrm{Var}\,[R^2]).$$

Note that we did not use Proposition 5.6. In particular, we do not need assumptions imposed therein.

Proof of (b). If $n/d \to \infty$, then $n/\sqrt{d} \to \infty$ and we can apply Proposition 5.6, resulting in

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{2n^2/d}} = \frac{T_n^{(d)} - n}{\sqrt{n}} \cdot \frac{\sqrt{n}}{\sqrt{2n^2/d}} + \frac{Q_n^{(d)}}{\sqrt{2n^2/d}} \xrightarrow[d\to\infty]{w} \mathrm{N}(0, 1).$$

Proof of (c). The starting point is the decomposition

$$\frac{\|S_n^{(d)}\|^2 - n}{\sqrt{n}} = \sum_{i=1}^{n} \Delta_i^{(d)}, \qquad \Delta_i^{(d)} := \frac{Y_i^{(d)}}{\sqrt{n}}, \quad Y_i^{(d)} := R_i^2 - 1 + 2R_i \langle V_i^{(d)}, S_{i-1}^{(d)} \rangle.$$

The sequence $\Delta_1^{(d)}, \dots, \Delta_n^{(d)}$ forms a martingale difference since

$$\mathbb{E}[Y_i^{(d)} | \mathcal{F}_{i-1}] = \mathbb{E}[(R_i^2 - 1)|\mathcal{F}_{i-1}] + 2\mathbb{E}R_i \cdot \mathbb{E}[\langle V_i^{(d)}, S_{i-1}^{(d)} \rangle | \mathcal{F}_{i-1}] = 0,$$

where we used that $\mathbb{E}R_i = 0$. Next we observe that

$$\begin{aligned}\mathbb{E}[(Y_i^{(d)})^2 | \mathcal{F}_{i-1}] &= \mathbb{E}[(R_i^2 - 1)^2] + 4\mathbb{E}[R_i(R_i^2 - 1)]\mathbb{E}[\langle V_i^{(d)}, S_{i-1}^{(d)} \rangle | \mathcal{F}_{i-1}] + 4\mathbb{E}[\langle V_i^{(d)}, S_{i-1}^{(d)} \rangle^2 | \mathcal{F}_{i-1}] \\ &= \mathrm{Var}\,(R^2) + 4\mathbb{E}R^3 \cdot \frac{1}{d}(S_{i-1,1}^{(d)} + \dots + S_{i-1,d}^{(d)}) + \frac{4}{d}\|S_{i-1}^{(d)}\|^2,\end{aligned}$$

where we used that $\mathbb{E}\langle V_i^{(d)}, x \rangle = \frac{1}{d}(x_1 + \dots + x_d)$ and $\mathbb{E}\langle V_i^{(d)}, x \rangle^2 = \|x\|^2/d$ for each fixed vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Let us check that

$$\frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d} S^{(d)}_{i-1,j} \overset{P}{\underset{d\to\infty}{\longrightarrow}} 0. \tag{43}$$

Indeed,

$$\frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d} S^{(d)}_{i-1,j} = \frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d}\sum_{k=1}^{i-1} R_k \mathbb{1}_{\{V^{(d)}_k=e_j\}} = \frac{1}{dn}\sum_{i=1}^{n}\sum_{k=1}^{i-1} R_k = \frac{1}{dn}\sum_{k=1}^{n-1}(n-k)R_k,$$

and the right-hand side converges to zero in probability by Chebyshev's inequality, since the variance of the right-hand side is $O(n^3/(nd)^2) = O(1/d)$.

Formula (43) together with the fact that $\frac{2}{n^2}\sum_{i=1}^n \|S^{(d)}_{i-1}\|^2$ converges in probability to 1, which can be verified in the same way as in the proof of Lemma 5.2, yield

$$\sum_{i=1}^{n}\mathbb{E}\left[\left(\Delta^{(d)}_i\right)^2 |\mathcal{F}^{(d)}_{i-1}\right] = \operatorname{Var}(R^2) + 4\mathbb{E}R^3\cdot\frac{1}{dn}\sum_{i=1}^{n}\sum_{j=1}^{d} S^{(d)}_{i-1,j} + \frac{4}{dn}\sum_{i=1}^{n}\|S^{(d)}_{i-1}\|^2 \overset{P}{\underset{d\to\infty}{\longrightarrow}} \operatorname{Var}(R^2)+2\gamma.$$

It remains to verify the following Lindeberg condition that implies (54). For every $\varepsilon>0$,

$$\lim_{d\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\left(R_i^2-1+2R_i\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle\right)^2 \mathbb{1}_{\left\{|R_i^2-1+2R_i\langle V^{(d)}_i,S^{(d)}_{i-1}\rangle|\geq\varepsilon\sqrt{n}\right\}}\right]=0. \tag{44}$$

The proof proceeds by the same method as in the proof of Theorem 3.6 (c) with the following modifications. The analogues of (30), (31), (32) (with $U^{(d)}_i$ replaced by $V^{(d)}_i$) can be established in the same way as above upon replacing (34) by

$$\mathbb{E}\left[\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle^2\right] = \mathbb{E}\left[\mathbb{E}\left[\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle^2 | S^{(d)}_{i-1}\right]\right] = \frac{1}{d}\mathbb{E}\|S^{(d)}_{i-1}\|^2 = \frac{i-1}{d}\leq\frac{n}{d}\leq C. \tag{45}$$

Instead of (33) we verify the following condition:

$$\lim_{d\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[R_i^2\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle^2 \mathbb{1}_{\left\{|R_i\langle V^{(d)}_i,S^{(d)}_{i-1}\rangle|\geq\frac14\varepsilon\sqrt{n}\right\}}\right]=0. \tag{46}$$

Take some $\delta>0$. Then,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[R_i^2\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle^2 \mathbb{1}_{\left\{|R_i\langle V^{(d)}_i,S^{(d)}_{i-1}\rangle|\geq\frac14\varepsilon\sqrt{n}\right\}}\right] &\leq \frac{1}{n}\left(\frac14\varepsilon\sqrt{n}\right)^{-\delta}\sum_{i=1}^{n}\mathbb{E}\left[|R_i|^{2+\delta}|\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle|^{2+\delta}\right]\\ &\leq C\cdot n^{-1-\frac{\delta}{2}}\sum_{i=1}^{n}\mathbb{E}\left[|\langle V^{(d)}_i, S^{(d)}_{i-1}\rangle|^{2+\delta}\right]\\ &\leq C\cdot n^{-1-\frac{\delta}{2}}\cdot\left(\frac{n^2}{d}+\frac{n^{2+\frac{\delta}{2}}}{d^{1+\frac{\delta}{2}}}\right),\end{aligned}$$

where the last inequality was established in the proof of Proposition 5.6. The right-hand side converges to 0 in the regime when $n \sim \gamma d$, which proves (46) and completes the verification of the Lindeberg condition (44). Thus, Part (c) follow by another appeal to Theorem 6.2. □

Proof of Theorem 3.7. By (8), we have

$$\frac{T_n^{(d)} - n}{n^{1/\alpha} L(n)} \xrightarrow[d\to\infty]{w} \zeta_\alpha. \tag{47}$$

Proof of (a). If $n > d^{\frac{\alpha}{2\alpha-2}+\delta}$ for some $\delta > 0$ and all sufficiently large d, then we also have $n/d \to \infty$ since $\alpha \in (1,2)$. Hence, Proposition 5.6 applies and $Q_n^{(d)}$ satisfies a CLT with normalization $\sqrt{2n^2/d}$. We have $n^{1/\alpha}L(n) = o(\sqrt{2n^2/d})$, meaning that the off-diagonal fluctuations dominate, and the claim follows.

Proof of (b). If $n < d^{\frac{\alpha}{2\alpha-2}-\delta}$ for all sufficiently large d, then $\operatorname{Var} Q_n^{(d)} = 2n(n-1)/d$ by (24) and $\sqrt{2n(n-1)/d} = o(n^{1/\alpha}L(n))$. It follows that

$$\frac{\|S_n^{(d)}\|^2 - n}{n^{1/\alpha} L(n)} = \frac{T_n^{(d)} - n}{n^{1/\alpha} L(n)} + \frac{Q_n^{(d)}}{\sqrt{2n(n-1)/d}} \frac{\sqrt{2n(n-1)/d}}{n^{1/\alpha} L(n)} \xrightarrow[d\to\infty]{w} \zeta_\alpha,$$

which proves the claim. □

Proof of Theorem 3.8. Proof of (a): If $n = o(\sqrt{d})$, then by a well-known result on the birthday problem (see, e.g., Example 3.2.5 in [7] or p. 42 in [15]), the probability that no box contains ≥ 2 balls (equivalently, that the vectors $V_1^{(d)}, \dots, V_n^{(d)}$ are pairwise different) converges to 1. On this event, we evidently have $\|S_n^{(d)}\|^2 = n$.

Proof of (b): Using the notation introduced at the beginning of the present Section 5.5, we can write

$$\|S_n^{(d)}\|^2 = \sum_{j=1}^{d} Z^2_{k_j(n);j} = \sum_{j=1}^{d} Z^2_{k_j(n);j} \, \mathbb{1}_{\{k_j(n)\geq 3\}} + \sum_{j=1}^{d} Z^2_{k_j(n);j} \, \mathbb{1}_{\{k_j(n)\leq 2\}}. \tag{48}$$

Let us show that the first sum (which is the total contribution of the boxes containing at least 3 balls) converges in distribution to 0. The expectation of this term is given by

$$\mathbb{E}\left[\sum_{j=1}^{d} Z^2_{k_j(n);j} \, \mathbb{1}_{\{k_j(n)\geq 3\}}\right] = d\,\mathbb{E}\left[Z^2_{k_1(n);1} \, \mathbb{1}_{\{k_1(n)\geq 3\}}\right] = d \sum_{k=3}^{\infty} \mathbb{E}\left[Z^2_{k;1} \, \mathbb{1}_{\{k_1(n)=k\}}\right] = d \sum_{k=3}^{\infty} k \mathbb{P}[k_1(n) = k].$$

Using that $k_1(n)$ has a binomial distribution $\mathrm{Bin}(n,1/d)$ and that $n \le 2c\sqrt{d}$ for sufficiently large d, we can write

$$\mathbb{E}\left[\sum_{j=1}^{d} Z^2_{k_j(n);j}\,\mathbb{1}_{\{k_j(n)\ge 3\}}\right] = d\sum_{k=3}^{\infty} k\binom{n}{k}\frac{1}{d^k}\left(1-\frac{1}{d}\right)^{n-k} \le d\sum_{k=3}^{\infty} k\frac{n^k}{k!}\frac{1}{d^k} \le d\sum_{k=3}^{\infty}\frac{(2c\sqrt{d})^k}{(k-1)!}\frac{1}{d^k} \le d^{-1/2}\sum_{k=3}^{\infty}\frac{(2c)^k}{(k-1)!} \le Cd^{-1/2}, \tag{49}$$

which converges to 0, as $d \to \infty$. Let us now analyze the second sum in (48). For each $k \in \mathbb{N}_0$ let $\mu_k(n) := \sum_{j=1}^{d} \mathbb{1}_{\{k_j(n)=k\}}$ be the number of boxes containing k balls. It is known, see Example 2 on pp. 14–15 in [1] or Theorems 3, 5 on pp. 67–68 in [15], that in the regime when $n \sim c\sqrt{d}$ with $c \in (0,\infty)$,

$$\mu_2(n) \underset{d\to\infty}{\overset{w}{\longrightarrow}} \mathrm{Poi}(c^2/2). \tag{50}$$

The number of boxes containing at least 3 balls is denoted by

$$\mu_{\ge 3}(n) := \sum_{k=3}^{\infty} \mu_k(n) = \sum_{j=1}^{d} \mathbb{1}_{\{k_j(n)\ge 3\}}.$$

Almost the same estimate as in (49) shows that $\mathbb{E}\mu_{\ge 3}(n) \to 0$ and hence $\mu_{\ge 3}(n) \to 0$ in probability. If some box j contains 1 ball, then the corresponding contribution $Z^2_{1;j}$ is 1. If some box j contains 2 balls, then $Z^2_{2;j}$ is either $(1+1)^2 = (-1-1)^2 = 4$ or $(+1-1)^2 = (-1+1)^2 = 0$, both possibilities having probability 1/2. Denoting by $\eta_1, \eta_2, \ldots$ i.i.d. random variables with $\mathbb{P}[\eta_\ell = 4] = \mathbb{P}[\eta_\ell = 0] = 1/2$, we can write

$$\sum_{j=1}^{d} Z^2_{k_j(n);j}\,\mathbb{1}_{\{k_j(n)\le 2\}} - n \overset{d}{=} \mu_1(n) + \sum_{\ell=1}^{\mu_2(n)} \eta_\ell - n = \sum_{\ell=1}^{\mu_2(n)} \eta_\ell - \mu_2(n) - \mu_{\ge 3}(n) = \sum_{\ell=1}^{\mu_2(n)} (\eta_\ell - 1) - \mu_{\ge 3}(n). \tag{51}$$

Since the random variables $\eta_\ell - 1$ take values 3 and -1 with probability 1/2 each and since $\mu_2(n)$ converges in distribution to $\mathrm{Poi}(c^2/2)$ by (50), it follows that the right-hand side of (51) converges in distribution to $3P' - P''$, where P', P'' are independent and both have a Poisson distribution with parameter $c^2/4$.

Proof of (c): For the simple random walk, the diagonal sum is deterministic: $T_n^{(d)} = n$. If $n/\sqrt{d} \to \infty$, then the off-diagonal sum $Q_n^{(d)}$ satisfies a CLT by Proposition 5.6, and the claim follows. $\square$

6. Appendix

In the present section we collect some facts that have been used in our proofs.

6.1. **A law of large numbers.** In Section 4 we used the following version of the weak law of large numbers for triangular arrays.

Lemma 6.1. *Assume that $(\theta_{d,i})_{i\in\mathbb{N}}$, for every $d \in \mathbb{N}$, is a sequence of independent copies of a random variable θ_d. Suppose that $\mathbb{E}\theta_d = 1$ for all $d \in \mathbb{N}$, and the family $(\theta_d)_{d\in\mathbb{N}}$ is uniformly integrable. Then, for every integer sequence $n = n(d)$ such that $n(d) \to \infty$, as $d \to \infty$,*

$$\frac{1}{n}\sum_{i=1}^{n}\theta_{d,i} \xrightarrow[d\to\infty]{P} 1.$$

Proof. The proof is standard and goes along the same lines as the proof of Theorem on p. 105 in [9]. Put $\theta_{d,i}(n) := \theta_{d,i}\,\mathbb{1}_{\{|\theta_{d,i}|\le n\}}$ and note that

$$\mathbb{P}\left\{\sum_{i=1}^{n}\theta_{d,i} \neq \sum_{i=1}^{n}\theta_{d,i}(n)\right\} \le n\mathbb{P}\{|\theta_d| \ge n\} \le \mathbb{E}\Big(|\theta_d|\,\mathbb{1}_{\{|\theta_d|\ge n\}}\Big) \le \sup_{d\in\mathbb{N}}\mathbb{E}\Big(|\theta_d|\,\mathbb{1}_{\{|\theta_d|\ge n\}}\Big).$$

The left-hand side converges to zero, as $n \to \infty$, by the definition of the uniform integrability of the family $(\theta_d)_{d\in\mathbb{N}}$. By the same reasoning,

$$\lim_{d\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\theta_{d,i}(n) = 1.$$

Thus, by Chebyshev's inequality, it remains to show that

$$\operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\theta_{d,i}(n)\right) \to 0, \quad d \to \infty.$$

Clearly, it suffices to check

$$\lim_{d\to\infty}\frac{\mathbb{E}\theta_d^2(n)}{n} = 0. \tag{52}$$

Observe that by Fubini's theorem

$$\begin{aligned}\mathbb{E}\theta_d^2(n) = 2\int_{[0,n]}\left(\int_0^y s\mathrm{d}s\right)\mathbb{P}\{|\theta_d| \in \mathrm{d}y\} &= 2\int_0^n s\left(\int_{[s,n]}\mathbb{P}\{|\theta_d| \in \mathrm{d}y\}\right)\mathrm{d}s\\ &\le 2\int_0^n s\mathbb{P}\{|\theta_d| \ge s\}\mathrm{d}s \le 2\int_0^n \mathbb{E}\Big(|\theta_d|\,\mathbb{1}_{\{|\theta_d|\ge s\}}\Big)\mathrm{d}s \le 2\int_0^n \sup_{d\in\mathbb{N}}\mathbb{E}\Big(|\theta_d|\,\mathbb{1}_{\{|\theta_d|\ge s\}}\Big)\mathrm{d}s.\end{aligned}$$

Thus, (52) follows by an appeal to L'Hôpital's rule. □

6.2. **Martingale central limit theorem.** In Section 5 we used a central limit theorem for martingale triangular arrays which can be found in [8, Theorem 2] and, in a slightly less general setting, in [2, Theorem 2]. To state it, we need some notation. For every $d \in \mathbb{N}$, let $\mathcal{G}_0^{(d)} \subset \mathcal{G}_1^{(d)} \subset \cdots \subset \mathcal{G}_n^{(d)}$ be a filtration on a probability space $(\Omega, \mathcal{G}, \mathbb{P})$, where $n = n(d)$ is a sequence of positive integers such that $n(d) \to \infty$, as $d \to \infty$. The random variables $\Upsilon_1^{(d)}, \dots, \Upsilon_n^{(d)}$ are said to form an *array of martingale differences* if $\Upsilon_i^{(d)}$ is $\mathcal{G}_i^{(d)}$-measurable and $\mathbb{E}[\Upsilon_i^{(d)} | \mathcal{G}_{i-1}^{(d)}] = 0$ for all $i = 1, \dots, n$. The following result [8, Theorem 2] provides sufficient conditions under which the CLT holds for the random variables $\Upsilon_1^{(d)} + \cdots + \Upsilon_n^{(d)}$.

Theorem 6.2. *In the setting just described, assume that the following conditions hold:*

(a) *The variables $\Upsilon_1^{(d)}, \dots, \Upsilon_n^{(d)}$ have finite second moments and*

$$\sum_{i=1}^{n} \mathbb{E}\left[\left(\Upsilon_i^{(d)}\right)^2 \middle| \mathcal{G}_{i-1}^{(d)}\right] \underset{d\to\infty}{\overset{P}{\longrightarrow}} \sigma^2 \in (0, \infty). \tag{53}$$

(b) *For every $\varepsilon > 0$,*

$$\sum_{i=1}^{n} \mathbb{E}\left[\left(\Upsilon_i^{(d)}\right)^2 \mathbb{1}_{\left\{|\Upsilon_i^{(d)}| \geq \varepsilon\right\}} \middle| \mathcal{G}_{i-1}^{(d)}\right] \underset{d\to\infty}{\overset{P}{\longrightarrow}} 0. \tag{54}$$

Then, $\Upsilon_1^{(d)} + \cdots + \Upsilon_n^{(d)}$ converges weakly to the normal distribution $\mathrm{N}(0, \sigma^2)$, *as $d \to \infty$.*

The Lindeberg-type condition stated in (54) follows from the following Lyapunov-type condition: for some $\delta > 0$,

$$\sum_{i=1}^{n} \mathbb{E}\left[\left(\Upsilon_i^{(d)}\right)^{2+\delta} \middle| \mathcal{G}_{i-1}^{(d)}\right] \underset{d\to\infty}{\overset{P}{\longrightarrow}} 0. \tag{55}$$

To prove this implication, observe that $(\Upsilon_i^{(d)})^2 \mathbb{1}_{\{|\Upsilon_i^{(d)}| \geq \varepsilon\}} \leq \varepsilon^{-\delta} (\Upsilon_i^{(d)})^{2+\delta}$. Note also that, by the Markov inequality, condition (55) follows from

$$\lim_{d\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[\left(\Upsilon_i^{(d)}\right)^{2+\delta}\right] = 0. \tag{56}$$

6.3. **Rosenthal inequality.** In Section 5 we frequently used the following Rosenthal inequality; see Theorem 9.1 on p. 152 in [10].

Theorem 6.3. *For every $\delta > 0$ there is a universal constant $B = B(\delta)$ such that the following holds: If $Z_1, \dots, Z_d$ are independent random variables with $\mathbb{E}[Z_j] = 0$ and $\mathbb{E}[|Z_j|^{2+\delta}] < \infty$ for all $j = 1, \dots, d$, then*

$$\mathbb{E}\left|\sum_{j=1}^{d} Z_j\right|^{2+\delta} \leq B \max\left\{\sum_{j=1}^{d} \mathbb{E}|Z_j|^{2+\delta}, \left(\sum_{j=1}^{d} \mathbb{E}[Z_j^2]\right)^{(2+\delta)/2}\right\}.$$

If the random variables $Z_1, \dots, Z_d$ are identically distributed, then the Rosenthal inequality yields

$$\mathbb{E}\left|\sum_{j=1}^{d} Z_j\right|^{2+\delta} \le B \max\left\{d \cdot \mathbb{E}|Z_1|^{2+\delta}, \left(d\mathbb{E}[Z_1^2]\right)^{(2+\delta)/2}\right\} \le C d^{1+\frac{\delta}{2}}, \tag{57}$$

for all $d \in \mathbb{N}$, with a constant C depending only on the distribution of Z_1 and δ. Alternatively, this inequality follows from Corollary 8.2 on page 151 in [10].

Acknowledgements

ZK was supported by the German Research Foundation under Germany's Excellence Strategy EXC 2044 – 390685587, Mathematics Münster: Dynamics - Geometry - Structure and by the DFG priority program SPP 2265 Random Geometric Systems. AM was supported by the Alexander von Humboldt Foundation.

References

[1] R. Arratia, L. Goldstein, and L. Gordon. Two moments suffice for Poisson approximations: the Chen-Stein method. *Ann. Probab.*, 17(1):9–25, 1989. URL `https://doi.org/10.1214/aop/1176991491`.

[2] B. M. Brown. Martingale central limit theorems. *Ann. Math. Statist.*, 42:59–66, 1971. URL `https://doi.org/10.1214/aoms/1177693494`.

[3] D. Burago, Y. Burago, and S. Ivanov. *A course in metric geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. URL `https://doi.org/10.1090/gsm/033`.

[4] S. I. Bykov. Limit theorems for randomized decomposable statistics and their applications. PhD Thesis. Institute of Automatics, Moscow, 1991.

[5] S. I. Bykov and V. A. Ivanov. Conditions for the asymptotic normality of multidimensional randomized decomposable statistics. *Diskret. Mat.*, 1(2):57–64, 1989. URL `https://doi.org/10.1515/dma.1991.1.2.219`.

[6] Victor H. de la Peña. Decoupling and Khintchine's inequalities for U-statistics. *Ann. Probab.*, 20(4):1877–1892, 1992. URL `https://doi.org/10.1214/aop/1176989533`.

[7] R. Durrett. *Probability: Theory and Examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition, 2010. URL `https://doi.org/10.1017/CBO9780511779398`.

[8] P. Gaenssler, J. Strobel, and W. Stute. On central limit theorems for martingale triangular arrays. *Acta Math. Acad. Sci. Hungar.*, 31(3-4):205–216, 1978. URL https://doi.org/10.1007/BF01901971.

[9] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1968.

[10] A. Gut. *Probability: A Graduate Course*. Springer Texts in Statistics. Springer, New York, 2005.

[11] P. R. Halmos. *A Hilbert space problem book*, volume 17 of *Encyclopedia of Mathematics and its Applications*. Springer-Verlag, New York-Berlin, second edition, 1982.

[12] J. Hoffmann-Jørgensen. Probability in Banach space. In *École d'Été de Probabilités de Saint-Flour, VI-1976*, Lecture Notes in Math., Vol. 598, pages 1–186. Springer-Verlag, Berlin, 1977.

[13] G. G. Johnson. A crinkled arc. *Proc. Amer. Math. Soc.*, 25:375–376, 1970. URL https://doi.org/10.2307/2037223.

[14] Z. Kabluchko and A. Marynych. Random Walks in the High-Dimensional Limit II. 2023+. In preparation.

[15] V. F. Kolchin, B. A. Sevastyanov, and V. P. Chistyakov. *Random allocations*. V. H. Winston & Sons, Washington, D.C.; Halsted Press [John Wiley & Sons], New York-Toronto, Ont.-London, 1978.

[16] A. N. Kolmogoroff. Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 26:115–118, 1940.

[17] V. G. Mikhailov. Polynomial and polynomial-like allocations: recent developments. In *Probabilistic methods in discrete mathematics (Petrozavodsk, 1992)*, volume 1 of *Progr. Pure Appl. Discrete Math.*, pages 40–59. VSP, Utrecht, 1993.

[18] A. J. Stam. Limit theorems for uniform distributions on spheres in high-dimensional Euclidean spaces. *J. Appl. Probab.*, 19(1):221–228, 1982. URL https://doi.org/10.2307/3213932.

[19] R. Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. URL https://doi.org/10.1017/9781108231596.

[20] R. A. Vitale. Representation of a crinkled arc. *Proc. Amer. Math. Soc.*, 52:303–304, 1975. URL https://doi.org/10.2307/2040149.

Zakhar Kabluchko: Institut für Mathematische Stochastik, Westfälische Wilhelms-Universität Münster
Email address: zakhar.kabluchko@uni-muenster.de

Alexander Marynych, Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, 01601 Kyiv, Ukraine
Email address: marynych@unicyb.kiev.ua