Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

ELSEVIER

Stochastic Processes and their Applications 103 (2003) 257–276

stochastic processes and their applications

www.elsevier.com/locate/spa

Time fluctuations of the random average process with parabolic initial conditions

L.R.G. Fontes[a,*,1], D.P. Medeiros[b,2], M. Vachkovskaia[c,3]

[a] *Instituto de Matemática e Estatística, Universidade de São Paulo, R. do Matão 1010, Cidade Universitária, 05508-090 São Paulo SP, Brazil*

[b] *Instituto de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, Campus de Ondina, 40170-110 Salvador Ba, Brazil*

[c] *Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Caixa Postal 6065, 13083-970 Campinas SP, Brazil*

Received 29 November 2001; received in revised form 15 July 2002; accepted 29 July 2002

Abstract

The random average process is a randomly evolving d-dimensional surface whose heights are updated by random convex combinations of neighboring heights. The fluctuations of this process in case of linear initial conditions have been studied before. In this paper, we analyze the case of polynomial initial conditions of degree 2 and higher. Specifically, we prove that the time fluctuations of a initial parabolic surface are of order $n^{2-d/2}$ for $d=1,2,3$; $\log n$ in $d=4$; and are bounded in $d \geqslant 5$. We establish a central limit theorem in $d=1$. In the bounded case of $d \geqslant 5$, we exhibit an invariant measure for the process as seen from the average height at the origin and describe its asymptotic space fluctuations. We consider briefly the case of initial polynomial surfaces of higher degree to show that their time fluctuations are not bounded in high dimensions, in contrast with the linear and parabolic cases.

MSC: primary 60K35; 82C41

Keywords: Random average process; Random surfaces; Harness process; Linear process; Surface fluctuations; Central limit theorem

* Corresponding author. Fax: +55-11-3091-6130.

E-mail addresses: lrenato@ime.usp.br (L.R.G. Fontes), medeiros@ufba.br (D.P. Medeiros), marinav@ime.unicamp.br (M. Vachkovskaia).

[1] Partially supported by CNPq grants 300576/92-7 and 662177/96-7 (PRONEX) and FAPESP grant 99/11962-9.

[2] Partially supported by CAPES.

[3] Partially supported by FAPESP (00/11462-5).

0304-4149/03/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0304-4149(02)00210-7

1. Introduction

The random average process (RAP) was introduced by Ferrari and Fontes (1998) as a model of a randomly evolving d-dimensional surface in $d+1$ space. The evolution consists of the heights of the surface getting updated, at either discrete or continuous time, by random convex combinations of neighboring heights (see (1) below). In this way, starting out with a given surface, which can be deterministic or itself random, we get at all times evolved surfaces which are random (as functions of the random convex weights and possibly random initial condition).

Closely related processes are the *harness process* introduced by Hammersley (1965/1966) (see also Toom, 1997) and the *smoothing process* (Andjel, 1985; Liggett and Spitzer, 1981; Liggett, 1985), where height updates consist of *deterministic* convex combinations of neighboring heights plus an additive (for the former process) or multiplicative (for the latter one) random noise. The RAP (as well as the smoothing process) is a special case of Liggett's linear processes (Liggett, 1985, Chapter IX).

A much studied special case of the RAP (one which we discuss only briefly in this paper) is the *voter model* (Liggett, 1985; Durrett, 1996). This corresponds to having the random convex combination almost surely assign total mass to a neighbor chosen at random. As discussed by Ferrari and Fontes (1998), the behavior of the voter model is rather different from the more general case treated in that paper and also here.

In this paper, we study at length the discrete time RAP with a parabolic initial condition. One of our main results are upper and lower bounds of the same leading order to the time fluctuations of the evolving surface. Under suitable assumptions on the distribution of the convex weights, we obtain $n^{2-d/2}$ as the leading order of the variance (as a function of time n) of the height of the surface at a given site in dimensions $d=1,2,3$; $\log n$ in $d=4$; and constant for $d \geqslant 5$ (see Theorem 2.1). This compares firstly to the case of linear initial conditions (Ferrari and Fontes, 1998), where the analogous variance is of order $\sqrt{n}$ in $d=1$; $\log n$ in $d=2$; and constant for $d \geqslant 3$. The approach and techniques here are also comparable with those of Ferrari and Fontes (1998). Here, as there, we have a dual process with the same single time distribution as the RAP (see (3) and (2)) which, when centered, is a martingale. It is enough then to study this process (as far as single time distributions are concerned). Variances are then also shown to be related to moments of a space-inhomogeneous Markov chain, here in $2d$-space (see (8), (14), (23) and (25)), rather than in d-space. In view of the extra complication, we keep the analysis simple by making extra assumptions on the distribution of the convex weights vis-a-vis Ferrari and Fontes (1998) (see the first paragraph of the next section). We also take a convenient concrete form for the initial condition (see (5)). The problem is then further reduced to one involving the same d-dimensional Markov chain that enters the analysis of the linear case in Ferrari and Fontes (1998), but via a different, if related, quantity (see (31)). The analysis proceeds indirectly (as in the linear case) by taking generating functions. The argument for the parabolic case will actually involve also the derivative of the generating functions entering the analysis of the linear case. As in Ferrari and Fontes (1998), we compare with the analogous quantity for a d-dimensional space-homogeneous Markov chain (a random walk), and then get the upper bounds (see (33), (39), (41), (42) and (43)).

The argument for the lower bounds is similar, but simpler. It involves a d-dimensional random walk directly (see (46) and (51)).

We also prove, in one dimension, a central limit theorem for the time fluctuations of the surface (see Theorem 2.3). As in Ferrari and Fontes (1998), we verify the hypotheses of a martingale CLT in Hall and Heyde (1980).

The boundedness of the height variance of the RAP in high dimensions seems to be a distinguishing feature of linear and parabolic initial conditions, among polynomial ones. We show in Theorem 5.1 that, starting out with a cubic, the heights have variance of order of at least n^2 in all high enough dimensions. This divergence in time of the fluctuations can be argued for initial polynomials of higher degree as well.

One difference between the initial linear and parabolic cases is the following. Due to the martingale property of the dual processes, mentioned above, and the L_2-boundedness in high dimensions, the RAPs as seen from the average height at the origin with initial linear and parabolic surfaces converge weakly to invariant measures for the dynamics as seen from the average height at the origin, in those dimensions. The spatial fluctuations of these measures can be then studied and they are found to be bounded for linear initial conditions. This is *not* the case for the initial condition here. We show in Theorem 4.1 that the non-trivially scaled space fluctuations of the invariant measures converge weakly to a non-trivial limit. The variances also converge to the variance of the limit.

This paper grew out of the Ph.D. research of the second author, which consisted of the RAP with a parabolic initial condition of a different form from the one treated here. Medeiros (2001) shows essentially the same results we present here, except for the ones in Sections 5 and 6, obtained with the same approach and techniques, and more: sharp bounds were obtained in one dimension (see Theorem 2.2); fluctuations of the surface as seen from the height at the origin (of a slightly modified process) were shown to be bounded for $d \geqslant 3$, the scaled spatial fluctuations of the limiting invariant measures for the process as seen from the height at the origin arising in this context were proved to converge to non-trivial weak limits; and continuous time analogs of the discrete time results were established. The assumptions on the convex weights distribution made in Medeiros (2001) are less restrictive than the ones here. Some of the extra results will be the object of a future paper.

We close this introduction with a comparison to the harness process mentioned above. The underlying space is $\mathbb{Z}^d$. For the case of a deterministic initial surface (as here), translation invariant convex weights (in here, that is the case in the distributional sense) and i.i.d. L_2 mean zero noise, the surface (height vector) at time n can be written as

$$X_n = \mathscr{U}^n X_0 + \sum_{k=1}^{n} \mathscr{U}^{n-k} \mathscr{Z}_k,$$

where $\mathscr{Z}_k$, $k \geqslant 1$, are i.i.d. vectors of i.i.d. L_2 mean zero noise components and $\mathscr{U}$ is the convex weights matrix satisfying that $\mathscr{U}(i, i+j) \geqslant 0$ does not depend on i for all j, $\sum_j \mathscr{U}(i,j) = 1$ for all i and $\sum_j \|j\|^2 \mathscr{U}(0,j) < \infty$. It is clear then that, in this context, the height variances do not depend on X_0. With the same approach and techniques used by Ferrari and Fontes (1998) and here, it is possible to show that those variances have the same order of magnitude as the ones of the RAP with linear initial conditions in all dimensions (the argument is actually quite straightforward in this case). We conclude that the fluctuations of the RAP and the harness process behave rather differently for parabolic (and higher degree polynomial) initial conditions of the kind considered in this paper.

2. Definitions and main results

We briefly review the definition of the discrete time random average process. See Ferrari and Fontes (1998) for more details. Let $X_0 \in \mathbb{R}^{\mathbb{Z}^d}$ be a initial surface and define, for $n \geqslant 1$,

$$X_n(i) = \sum_{j \in \mathbb{Z}^d} u_n(i,j) X_{n-1}(j), \quad n \geqslant 1, \tag{1}$$

where $U = \{u_n(i, i+\cdot),\ n \geqslant 1, i \in \mathbb{Z}^d\}$ is a family of i.i.d. random probability vectors independent of the initial configuration X_0, all of which are defined in a suitable probability space $\{\Omega, \mathscr{G}, \mathbb{P}\}$.

Then, for all $n \geqslant 0$,

$$X_n := \{X_n(x);\ x \in \mathbb{Z}^d\} \stackrel{\mathrm{d}}{=} \{L_n(x);\ x \in \mathbb{Z}^d\} =: L_n, \tag{2}$$

where $\stackrel{\mathrm{d}}{=}$ means identity in distribution and, for $n \geqslant 0$ and $x \in \mathbb{Z}^d$,

$$L_n(x) = \mathbb{E}[X_0(Y_n^x)|\mathscr{F}] \tag{3}$$

and $(Y_k^x)_{k \geqslant 0}$ is a random walk defined in $\{\Omega, \mathscr{G}, \mathbb{P}\}$ with transition probabilities

$$\mathbb{P}(Y_k^x = j | Y_{k-1}^x = i, \mathscr{F}) = u_k(i,j). \tag{4}$$

The results of this paper will be stated and argued for L_n. By (2), they hold for X_n as well.

Let $\theta_n(k)$ be the mean of $u_n(k, k+\cdot)$ and e_j, $j = 1, \ldots, d$, the jth positive coordinate vector. Let $\mathscr{N} = \{\pm e_j,\ j = 1, \ldots, d\}$. We make the following assumptions on the distribution of $u_1(0, \cdot)$:

1. Nearest-neighbor range: $u_1(0,i) = 0$ almost surely for $i \notin \mathscr{N}$.
2. Symmetry: $\{u_1(0,i);\ i \in \mathscr{N}\} \stackrel{\mathrm{d}}{=} \{u_1(0,-i);\ i \in \mathscr{N}\}$.
3. Coordinate exchangeability: $\{u_1(0,\pi(i));\ i \in \mathscr{N}\} \stackrel{\mathrm{d}}{=} \{u_1(0,i);\ i \in \mathscr{N}\}$ for all permutations of coordinates π.
4. Non-voter model case: $\mathbb{P}(u_1(0,i) = 1 \text{ for some } i \in \mathscr{N}) < 1$.
5. Non-degeneracy: $\sigma^2 := \mathbb{E}\{[\mathbf{1}\theta_1(0)]^2\} > 0$,

where $\mathbf{1} \in \mathbb{R}^d$, $\mathbf{1} = (1, \ldots, 1)$. The first and third assumptions are for simplicity. The fourth one is to rule out a simpler (and qualitatively different) case, namely the voter model (but see Remark 4.1). The last one involves the particular initial condition we will consider (see (5)) and, for that case, rules out a trivial case, where there are no fluctuations.

In this paper we consider a parabolic initial condition of the form

$$X_0(x) = (\mathbf{1}x)^2 = \left(\sum_{k=1}^{d} x_k\right)^2 \tag{5}$$

for any $x \in \mathbb{Z}^d$. This satisfies (2.4) of Ferrari and Fontes (1998), so that the process is well defined.

Remark 2.1. Other parabolic forms, like $(\lambda x)^2 = (\sum_{k=1}^{d} \lambda_k x_k)^2$, where $\lambda \in \mathbb{R}^d$ is a fixed non-null vector, can be handled, with essentially the same results and techniques. The form $\|x\|^2 = \sum_{k=1}^{d} x_k^2$ was analyzed by Medeiros (2001).

With the choice (5), we have

$$L_n(x) = \mathbb{E}[(\mathbf{1}Y_n^x)^2 | \mathscr{F}]. \tag{6}$$

Since Y_n^x is a random walk starting from x and the u's are symmetric, we have $\mathbb{E}(L_n(x)) = \mathbb{E}((\mathbf{1}Y_n^x)^2) = \mathbb{E}((\mathbf{1}(Y_n^0 + x))^2) = (\mathbf{1}x)^2 + \mathbb{E}((\mathbf{1}Y_n^0)^2)$. Writing $Y_n^0 = \sum_{i=1}^{n} \Delta_i$, where $\Delta_1, \ldots, \Delta_n$ are i.i.d. random vectors such that $\mathbb{P}(\Delta_1 = \pm e_i) = 1/2d$, for $i = 1, \ldots, d$, we have $\mathbb{E}((\mathbf{1}Y_n^0)^2) = \sum_{i=1}^{n} \mathbb{E}((\mathbf{1}\Delta_i)^2) = n$, since $\mathbf{1}\Delta_1 = \pm 1$ almost surely. Thus $\mathbb{E}(L_n(x)) = (\mathbf{1}x)^2 + n$. Let $\bar{Y}_n^x = Y_n^x - \mathbb{E}(Y_n^x) = Y_n^x - x$. Then, $(\mathbf{1}Y_n^x)^2 = (\mathbf{1}\bar{Y}_n^x)^2 + 2(\mathbf{1}x)(\mathbf{1}\bar{Y}_n^x) + (\mathbf{1}x)^2$ and

$$L_n(x) = \bar{L}_n(x) + 2(\mathbf{1}x)\bar{Z}_n(x) + (\mathbf{1}x)^2, \tag{7}$$

where $\bar{L}_n(x) = \mathbb{E}((\mathbf{1}\bar{Y}_n^x)^2 | \mathscr{F}_n)$ and $\bar{Z}_n(x) = \mathbb{E}(\mathbf{1}\bar{Y}_n^x | \mathscr{F}_n)$. Note that, by translation invariance, the distributions of $\bar{L}_n(x)$ and $\bar{Z}_n(x)$ do not depend on x. Let $\bar{L}_n := L_n(0) = \bar{L}_n(0)$ and $\bar{Z}_n := Z_n(0) = \bar{Z}_n(0)$. So,

$$\mathbb{V}(L_n(x)) = \mathbb{V}(\bar{L}_n) + 4(\mathbf{1}x)^2 \mathbb{V}(\bar{Z}_n) + 4(\mathbf{1}x)\mathbb{C}\mathrm{ov}(\bar{L}_n, \bar{Z}_n). \tag{8}$$

Below, $c_1, c_2, \ldots$ will always denote positive real numbers which may depend only on d. One of our main results is the following:

Theorem 2.1. *For all $x \in \mathbb{Z}^d$, there exist c_1, c_2 such that*

$$c_1 \mathfrak{O}(n,d) \leqslant \mathbb{V}(L_n(x)) \leqslant c_2 \mathfrak{O}(n,d)$$

for all n, where

$$\mathfrak{O}(n,d) = \begin{cases} n^{2-d/2} & \text{if } d = 1,2,3, \\ \log n & \text{if } d = 4, \\ \text{constant} & \text{if } d \geqslant 5. \end{cases} \tag{9}$$

The proof of Theorem 2.1 will be presented in Section 4.

In dimension 1, it is possible to get a stronger result, for which we do not present a proof here, but rather refer to Medeiros (2001, Theorem 3.1.3):

Theorem 2.2. *If $d=1$, then there exists c_3 such that for all $x \in \mathbb{Z}^d$*

$$\frac{\mathbb{V}(L_n(x))}{n^{3/2}} \to c_3 \quad \textit{when } n \to \infty.$$

We establish also a Central Limit Theorem for $\bar{L}_n = L_n(0)$ in dimension 1 (the proof, sketched in Section 6, does not use Theorem 2.2). Let $\mathscr{V}_n := \mathbb{V}(\bar{L}_n)$.

Theorem 2.3. *In $d=1$, the distribution of $\mathscr{V}_n^{-1/2}(\bar{L}_n - n)$ converges to a standard Gaussian as $n \to \infty$.*

Our analysis yields that in dimensions 5 or more there exists an invariant measure for the dynamics of the surface as seen from the average height at the origin. This is related to the almost sure existence of the limits of $\bar{L}_n - n$ and $\bar{Z}_n$ as $n \to \infty$. In Section 4.2, we discuss this and prove a result about the asymptotic shape and magnitude of the space fluctuations of this measure.

In Section 3, we state and prove auxiliary results for the arguments of the proofs of our main results. In Section 5, we discuss the case of higher order polynomial initial conditions and prove a result that indicates a substantial difference with the linear and parabolic cases, namely the unboundedness of the time fluctuations at high dimensions.

3. Preliminaries

To prove Theorems 2.1 and 2.3 we will need some lemmas.

Lemma 3.1. *The process $\bar{L}_n - n$ is a martingale with respect to $\{\mathscr{F}_n,\ n \geqslant 0\}$.*

Proof. Let $Y_n = Y_n^0$. So $\mathbb{E}(\bar{L}_n) = \mathbb{E}((\mathbf{1}Y_n)^2) = n$. We have

$$\begin{aligned}
\bar{L}_n &= \mathbb{E}[(\mathbf{1}Y_n)^2 | \mathscr{F}_n] \\
&= \sum_{k\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} (\mathbf{1}j)^2 \mathbb{P}(Y_n = j | Y_{n-1} = k, \mathscr{F}_n) \mathbb{P}(Y_{n-1} = k | \mathscr{F}_n) \\
&= \sum_{k\in\mathbb{Z}^d} \sum_{j\in\mathbb{Z}^d} \{\mathbf{1}[k + (j-k)]\}^2 u_n(k,j) \mathbb{P}(Y_{n-1} = k | \mathscr{F}_{n-1}) \\
&= \sum_{k\in\mathbb{Z}^d} (\mathbf{1}k)^2 \mathbb{P}(Y_{n-1} = k | \mathscr{F}_{n-1}) \\
&\quad + 2 \sum_{k\in\mathbb{Z}^d} (\mathbf{1}k) \left(\sum_{j\in\mathbb{Z}^d} [\mathbf{1}(j-k)] u_n(k,j) \right) \mathbb{P}(Y_{n-1} = k | \mathscr{F}_{n-1})
\end{aligned}$$

$$+\sum_{k\in\mathbb{Z}^d}\left(\sum_{j\in\mathbb{Z}^d}[\mathbf{1}(j-k)]^2u_n(k,j)\right)\mathbb{P}(Y_{n-1}=k|\mathscr{F}_{n-1})$$

$$=\mathbb{E}[(\mathbf{1}Y_{n-1})^2|\mathscr{F}_{n-1}]+2\sum_{k\in\mathbb{Z}^d}(\mathbf{1}k)(\mathbf{1}\theta_n(k))\mathbb{P}(Y_{n-1}=k|\mathscr{F}_{n-1})+1,$$

since $[\mathbf{1}(j-k)]^2=1$ every time $u_n(k,j)\neq 0$, due to the nearest-neighbor character of the u's.

Letting $W_n=\mathbb{E}[(\mathbf{1}Y_{n-1})(\mathbf{1}\theta_n(Y_{n-1}))|\mathscr{F}_n]$, we have $\bar{L}_n=\bar{L}_{n-1}+2W_n+1$, $n\geqslant 1$, with $\bar{L}_0=0$. Thus

$$\bar{L}_n-n=2\sum_{i=1}^{n}W_i. \tag{10}$$

Note that the distribution of $\theta_n(k)$ does not depend on n or k and $\mathbb{E}(\theta_1(0))=\mathbb{E}(\sum_{j\in\mathbb{Z}^d} ju_1(0,j))=0$. We have then

$$\mathbb{E}[W_n|\mathscr{F}_{n-1}]=\mathbb{E}\{\mathbb{E}[(\mathbf{1}Y_{n-1})(\mathbf{1}\theta_n(Y_{n-1}))|\mathscr{F}_n]|\mathscr{F}_{n-1}\} \tag{11}$$

$$=\mathbb{E}[(\mathbf{1}Y_{n-1})(\mathbf{1}\theta_n(Y_{n-1}))|\mathscr{F}_{n-1}]$$

$$=\sum_{k\in\mathbb{Z}^d}\mathbb{E}[(\mathbf{1}k)(\mathbf{1}\theta_n(k))|Y_{n-1}=k,\mathscr{F}_{n-1}]\mathbb{P}(Y_{n-1}=k|\mathscr{F}_{n-1})$$

$$=\sum_{k\in\mathbb{Z}^d}(\mathbf{1}k)(\mathbf{1}\mathbb{E}[\theta_n(k)])\mathbb{P}(Y_{n-1}=k|\mathscr{F}_{n-1})=0, \tag{12}$$

since $\theta_n(k)$ is independent of $\mathscr{F}_{n-1}$ for all n,k. Thus, $\bar{L}_n-n$ is a martingale with respect to $\{\mathscr{F}_n,\ n\geqslant 0\}$ and Lemma 3.1 is proved. □

Lemma 3.2. *Let $(\hat{Y}_n)_{n\geqslant 0}$ be an independent copy of $(Y_n)_{n\geqslant 0}$ given $\mathscr{F}$. Then $(Y_n,\hat{Y}_n)$ is a Markov chain in $\mathbb{Z}^d\times\mathbb{Z}^d$ with the following transition probabilities*:

$$\mathbb{P}(Y_n=k_n,\hat{Y}_n=l_n|Y_{n-1}=k_{n-1},\hat{Y}_{n-1}=l_{n-1})=v[u_1(k_{n-1},k_n)u_1(l_{n-1},l_n)], \tag{13}$$

where v denotes the marginal distribution of $u_1(\cdot,\cdot)$ and $v(X)$ means expectation of a random variable X with respect to v.

Proof. Straightforward. □

Corollary 3.1. *Let $D_n=Y_n-\hat{Y}_n$ and $S_n=Y_n+\hat{Y}_n$. Then $(D_n,S_n)_{n\geqslant 0}$ is a Markov chain in $\mathbb{Z}^d\times\mathbb{Z}^d$ with transition probabilities*

$$\mathbb{P}(D_n=d_n,S_n=s_n|D_{n-1}=d_{n-1},S_{n-1}=s_{n-1})$$

$$=v\left[u_1\left(\frac{s_{n-1}+d_{n-1}}{2},\frac{s_n+d_n}{2}\right)u_1\left(\frac{s_{n-1}-d_{n-1}}{2},\frac{s_n-d_n}{2}\right)\right]. \tag{14}$$

Proof. Straightforward from Lemma 3.2. □

Remark 3.1. Corollary 3.1 and the assumptions on the u's imply that

$$\mathbb{P}(D_n = d_{n-1} + d, S_n = s_{n-1} + s | D_{n-1} = d_{n-1}, S_{n-1} = s_{n-1})$$
$$= \begin{cases} v[u_1(0, \frac{s+d}{2}) u_1(0, \frac{s-d}{2})] & \text{if } d_{n-1} = 0, \\ v[u_1(0, \frac{s+d}{2})] v[u_1(0, \frac{s-d}{2})] & \text{if } d_{n-1} \neq 0 \end{cases} \tag{15}$$

and thus (D_n, S_n) is space homogeneous in $\{0\} \times \mathbb{Z}^d$ and $(\mathbb{Z}^d \backslash \{0\}) \times \mathbb{Z}^d$ separately, but not in $\mathbb{Z}^d \times \mathbb{Z}^d$.

Remark 3.2. It follows from (15) that D_n, $n \geqslant 0$, is a Markov chain with transition probabilities (see also Ferrari and Fontes, 1998, Lemma 2.5)

$$\gamma(l, k) = \begin{cases} \sum_{j \in \mathbb{Z}^d} v[u_1(0, j) u_1(0, j + k)] & \text{if } l = 0, \\ \sum_{j \in \mathbb{Z}^d} v[u_1(0, j)] v[u_1(l, j + k)] & \text{if } l \neq 0. \end{cases} \tag{16}$$

Our assumptions on the u's make the jumps of D_n have length only either 0 or 2, with $\gamma(0,0) < 1$. The jumps of length 2 can be (only) in any of the coordinate positive and negative directions. All of these possibilities have equal probabilities (which do not depend on the starting point, provided it is in $\mathbb{Z}^d \backslash \{0\}$), that is

$$\gamma(l, l \pm 2e_j) = (1 - \gamma(l, l))/(2d) \quad \text{for all } l \in \mathbb{Z}^d \text{ and } j = 1, \ldots, d,$$

$$\gamma(l, l) = \gamma(l', l') \quad \text{if } l, l' \neq 0.$$

Remark 3.3. Remark 3.1 allows us to construct (D_n, S_n) in the following way. Let $\{(\delta_n(i), \xi_n(i));\ i \in \mathbb{Z}^d \backslash \{0\}, n \geqslant 1\}$ and $\{(\delta_n(0), \xi_n(0));\ n \geqslant 1\}$ be two independent families of i.i.d. random vectors such that $(\delta_1(0), \xi_1(0))$ are distributed as the increments of (D_n, S_n) in $\{0\} \times \mathbb{Z}^d$, and $(\delta_1(i)), \xi_1(i))$ $(i \neq 0)$ as those in $(\mathbb{Z}^d \backslash \{0\}) \times \mathbb{Z}^d$, that is,

$$\mathbb{P}(\delta_1(0) = d, \xi_1(0) = s) = v\left[u_1\left(0, \frac{s+d}{2}\right) u_1\left(0, \frac{s-d}{2}\right)\right] \tag{17}$$

and

$$\mathbb{P}(\delta_1(i) = d, \xi_1(i) = s) = v\left[u_1\left(0, \frac{s+d}{2}\right)\right] v\left[u_1\left(0, \frac{s-d}{2}\right)\right]. \tag{18}$$

Then

$$(D_n, S_n) = \sum_{i=1}^{n} (\delta_i(D_{i-1}), \xi_i(D_{i-1})). \tag{19}$$

Remark 3.4. Remark 3.1 implies also that for all n, given D_{n-1} and $\delta_n(D_{n-1})$, $\xi_n(D_{n-1}) \stackrel{\mathrm{d}}{=} -\xi_n(D_{n-1})$.

Let H_n be a space homogeneous Markov chain (random walk) with transition probabilities

$$\gamma_H(l,k) = \sum_{j \in \mathbb{Z}^d} v(u_1(0,j))v(u_1(l,j+k)) \quad \text{for all } l,k \in \mathbb{Z}^d. \tag{20}$$

Remark 3.5. The transition probabilities of H_n and D_n differ only at the origin. We have also that $0 < \gamma_H(0,0) < \gamma(0,0) < 1$.

For $0 < s < 1$, let $f(s)$ and $g(s)$ be the power series of $\mathbb{P}(D_n = 0|D_0 = 0)$ and $\mathbb{P}(H_n = 0|H_0 = 0)$, respectively; that is, for $0 \leqslant s < 1$

$$f(s) = \sum_{n \geqslant 0} \mathbb{P}(D_n = 0|D_0 = 0)s^n \quad \text{and} \quad g(s) = \sum_{n \geqslant 0} \mathbb{P}(H_n = 0|H_0 = 0)s^n.$$

Let $\gamma := \gamma(0,0)$, $\bar{\gamma} := \gamma_H(0,0)$ and, for $0 \leqslant p < 1$,

$$h(s,p) := \{1 - \phi(s,p)\}^{-1}, \quad \bar{h}(s,p) := \{1 - \bar{\phi}(s,p)\}^{-1}$$

with

$$\phi(s,p) := ps + (1-p)\psi(s), \quad \bar{\phi}(s,p) := ps + (1-p)\bar{\psi}(s)$$

and

$$\psi(s) := \sum_{n \geqslant 0} \mathbb{P}(T = n)s^{n+1}, \quad \bar{\psi}(s) := \sum_{n \geqslant 0} \mathbb{P}(T' = n)s^{n+1},$$

where T is the time of the first return to the origin of D_n after first leaving it, and T' is the analogous return time for H_n. In Ferrari and Fontes (1998), (3.7)–(3.9), it was shown (with a different notation) that

$$f(s) = h(s,\gamma) = \frac{1}{1 - \gamma s - (1-\gamma)\psi(s)}$$

and

$$g(s) = \bar{h}(s,\bar{\gamma}) = \frac{1}{1 - \bar{\gamma}s - (1-\bar{\gamma})\bar{\psi}(s)}.$$

Note that, from the nearest-neighbor and exchangeability assumptions we made, T has the same distribution as the return time starting from $2e_1$, say, since, after leaving the origin, D_n necessarily jumps to a site of the form $\pm 2e_i$ for some $i = 1,\dots,d$, and the distribution of this latter return time does not depend on which of these sites D_n starts from, by the coordinate exchangeability assumption on the u's. By the same considerations and Remark 3.5, we conclude that $T \stackrel{\mathrm{d}}{=} T'$ and, thence, $\bar{h}(s,p) \equiv h(s,p)$.

Lemma 3.3. *Let $f^{(k)}$ and $g^{(k)}$ denote the kth derivatives of f and g, respectively. Then, for $k = 0,1,2,3,4$*

$$\lim_{s \uparrow 1} \frac{f^{(k)}(s)}{g^{(k)}(s)} < \infty. \tag{21}$$

Remark 3.6. (21) holds for all k. We give proof of the cases we explicitly use in this paper. The general argument is similar.

Proof. The case $k=0$ was proved by Ferrari and Fontes (1998, Lemma 3.2). The argument in our case is much simpler, so we give it in the next paragraph, for completeness.

For $0 \leqslant s < 1$

$$\frac{f^{(0)}(s)}{g^{(0)}(s)} = \frac{f(s)}{g(s)} = \frac{\bar{\gamma}(1-s)+(1-\bar{\gamma})[1-\psi(s)]}{\gamma(1-s)+(1-\gamma)[1-\psi(s)]} = \frac{\bar{\gamma}p_s+(1-\bar{\gamma})q_s}{\gamma p_s+(1-\gamma)q_s}, \tag{22}$$

where $p_s=(1-s)/\{(1-s)+[1-\psi(s)]\}$ and $q_s=1-p_s$. Since $p_s\in[0,1]$, the result for $k=0$ follows from Remark 3.5.

We thus have a positive constant M such that $f(s)\leqslant Mg(s)$ for all $0\leqslant s<1$. We can also have M which satisfies $1-\gamma\leqslant M(1-\bar{\gamma})$, $\gamma\leqslant M\bar{\gamma}$.

For the case $k=1$, notice first that

$$f^{(1)}(s)=[f(s)]^2\,\mathrm{d}\phi(s,\gamma)/\mathrm{d}s=[f(s)]^2[\gamma+(1-\gamma)\psi^{(1)}(s)]$$

and, analogously,

$$g^{(1)}(s)=[g(s)]^2\,\mathrm{d}\phi(s,\bar{\gamma})/\mathrm{d}s=[g(s)]^2[\bar{\gamma}+(1-\bar{\gamma})\psi^{(1)}(s)].$$

It follows that $\lim_{s\uparrow 1} f^{(1)}(s)/g^{(1)}(s)\leqslant M^3$.

For the next case, notice that

$$\begin{aligned} f^{(2)}(s) &= [f(s)]^2\frac{\mathrm{d}^2\phi(s,\gamma)}{\mathrm{d}s^2}+2[f(s)]^3\left(\frac{\mathrm{d}\phi(s,\gamma)}{\mathrm{d}s}\right)^2 \\ &= [f(s)]^2(1-\gamma)\psi^{(2)}(s)+2[f(s)]^3[\gamma+(1-\gamma)\psi^{(1)}(s)]^2 \end{aligned}$$

and a similar expression holds for $g^{(2)}(s)$, with $\bar{\gamma}$ replacing γ and g replacing f. From the above considerations, it follows that $\lim_{s\uparrow 1} f^{(2)}(s)/g^{(2)}(s)\leqslant 3M^5$.

Similarly, we find that

$$\lim_{s\uparrow 1} f^{(3)}(s)/g^{(3)}(s)\leqslant 13M^7 \quad\text{and}\quad \lim_{s\uparrow 1} f^{(4)}(s)/g^{(4)}(s)\leqslant 75M^9. \qquad \square$$

4. Fluctuations of L_n

4.1. Proof of Theorem 2.1

By Lemma 3.1 and (10),

$$\mathbb{V}(\bar{L}_n)=4\sum_{i=1}^{n}\mathbb{V}(W_i). \tag{23}$$

Now

$$\mathbb{V}(W_i)=\mathbb{E}(W_i)^2=\mathbb{E}\left[\left(\sum_{k\in\mathbb{Z}^d}(\mathbf{1}k)(\mathbf{1}\theta_i(k))\mathbb{P}(Y_{i-1}=k|\mathscr{F}_{i-1})\right)^2\right]$$

$$= \mathbb{E} \sum_{k,r \in \mathbb{Z}^d} (\mathbf{1}k)(\mathbf{1}r)(\mathbf{1}\theta_i(k))(\mathbf{1}\theta_i(r))\mathbb{P}(Y_{i-1} = k|\mathscr{F}_{i-1})\mathbb{P}(Y_{i-1} = r|\mathscr{F}_{i-1})$$

$$= \sum_{k \in \mathbb{Z}^d} (\mathbf{1}k)^2 \mathbb{E}\{(\mathbf{1}\theta_i(k))^2 \mathbb{P}^2(Y_{i-1} = k|\mathscr{F}_{i-1})\}$$

$$= \sum_{k \in \mathbb{Z}^d} (\mathbf{1}k)^2 v[(\mathbf{1}\theta_i(k))^2]\mathbb{E}[\mathbb{P}^2(Y_{i-1} = k|\mathscr{F}_{i-1})]$$

$$= \sigma^2 \sum_{k \in \mathbb{Z}^d} (\mathbf{1}k)^2 \mathbb{E}[\mathbb{P}^2(Y_{i-1} = k|\mathscr{F}_{i-1})], \tag{24}$$

by the independence between $\theta_i(\cdot)$ and $\mathscr{F}_{i-1}$, between $\theta_i(k)$ and $\theta_i(r)$ when $k \neq r$, and the zero mean of the latter.

We have that

$$\begin{aligned}\mathbb{E}[\mathbb{P}^2(Y_{i-1} = k|\mathscr{F}_{i-1})] &= \mathbb{E}[\mathbb{P}(Y_{i-1} = k|\mathscr{F}_{i-1})\mathbb{P}(\hat{Y}_{i-1} = k|\mathscr{F}_{i-1})] \\ &= \mathbb{E}[\mathbb{P}(Y_{i-1} = k, \hat{Y}_{i-1} = k|\mathscr{F}_{i-1})] \\ &= \mathbb{P}(Y_{i-1} = \hat{Y}_{i-1} = k) = \mathbb{P}\left(\frac{S_{i-1}}{2} = k, D_{i-1} = 0\right),\end{aligned}$$

by the definition of S_n and D_n (see Lemma 3.2 and Corollary 3.1 above). Thus,

$$\mathbb{V}(W_i) = \sigma^2 \sum_{k \in \mathbb{Z}^d} (\mathbf{1}k)^2 \mathbb{P}\left(\frac{S_{i-1}}{2} = k, D_{i-1} = 0\right) = \frac{\sigma^2}{4} \mathbb{E}[(\mathbf{1}S_{i-1})^2; D_{i-1} = 0]. \tag{25}$$

Thus from (25) and (23), we get

$$\mathbb{V}(\bar{L}_n) = \sigma^2 \sum_{i=0}^{n-1} \mathbb{E}[(\mathbf{1}S_i)^2; D_i = 0]. \tag{26}$$

4.1.1. Upper bound for $\mathbb{V}(\bar{L}_n)$

Writing S_i as a sum of its increments, we get

$$\begin{aligned}\mathbb{E}[(\mathbf{1}S_i)^2; D_i = 0] &= \mathbb{E}\left[\left(\sum_{j=1}^{i} \mathbf{1}\xi_j(D_{j-1})\right)^2 ; D_i = 0\right] \\ &= \sum_{j=1}^{i} \mathbb{E}\{[\mathbf{1}\xi_j(D_{j-1})]^2; D_i = 0\} \\ &\quad + 2\sum_{1=k<j}^{i} \mathbb{E}\{[\mathbf{1}\xi_k(D_{k-1})][\mathbf{1}\xi_j(D_{j-1})]; D_i = 0\}.\end{aligned} \tag{27}$$

Since $\xi_j(D_{j-1}) = (Y_j - Y_{j-1}) + (\hat{Y}_j - \hat{Y}_{j-1})$, we have $|\mathbf{1}\xi_j(D_{j-1})| \leqslant |\mathbf{1}(Y_j - Y_{j-1})| + |\mathbf{1}(\hat{Y}_j - \hat{Y}_{j-1})| \leqslant 2$. So, $(\mathbf{1}\xi_j(D_{j-1}))^2 \leqslant 4$ and $\mathbb{E}[(\mathbf{1}\xi_j(D_{j-1}))^2; D_i = 0] \leqslant 4\mathbb{P}(D_i = 0)$.

Thus, for the first term on the right-hand side of (27), we obtain

$$\sum_{j=1}^{i} \mathbb{E}[(\mathbf{1}\xi_j(D_{j-1}))^2; D_i = 0] \leqslant \sum_{j=1}^{i} 4\mathbb{P}(D_i = 0) = 4i\,\mathbb{P}(D_i = 0). \tag{28}$$

The expectation in the second term on the right-hand side of (27) can be written as

$$\begin{aligned}
&\mathbb{E}[(\mathbf{1}\xi_k(D_{k-1}))(\mathbf{1}\xi_j(D_{j-1})); D_i = 0] \\
&\quad = \sum_{l\in\mathbb{Z}^d}\sum_{r\in\mathbb{Z}^d} \mathbb{E}[(\mathbf{1}\xi_k(r))(\mathbf{1}\xi_j(l)); D_{k-1} = r, D_{j-1} = l, D_i = 0] \\
&\quad = \sum_{l,r\in\mathbb{Z}^d}\sum_{w_1,w_2\in\mathbb{Z}^d} \mathbb{E}[(\mathbf{1}\xi_k(r))(\mathbf{1}\xi_j(l)); D_{k-1} = r, D_k = r + w_2, \\
&\qquad D_{j-1} = l, D_j = l + w_1, D_i = 0] \\
&\quad = \sum_{l,r\in\mathbb{Z}^d}\sum_{w_1,w_2\in\mathbb{Z}^d}\sum_{z_1,z_2\in\mathbb{Z}^d} (\mathbf{1}z_1)(\mathbf{1}z_2)\mathbb{P}(\xi_k(r) = z_2, \xi_j(l) = z_1, \\
&\qquad D_{k-1} = r, D_{j-1} = l, D_k = r + w_2, D_j = l + w_1, D_i = 0).
\end{aligned} \tag{29}$$

By the Markov property of D_n and the identities $\{D_j = l + w_1, D_{j-1} = l\} = \{D_{j-1} = l, \delta_j(l) = w_1\}$, $\{D_k = r + w_2, D_{k-1} = r\} = \{D_{k-1} = r, \delta_k(r) = w_2\}$, we get that, for all $k < j$, the latter probability equals $\mathbb{P}(D_i = 0|D_j = l + w_1)$ times

$$\begin{aligned}
&\mathbb{P}(\xi_j(l) = z_1, \delta_j(l) = w_1, \xi_k(r) = z_2, \delta_k(r) = w_2, D_{j-1} = l, D_{k-1} = r) \\
&\quad = \mathbb{P}(\xi_j(l) = z_1, \delta_j(l) = w_1)\mathbb{P}(\xi_k(r) = z_2, \delta_k(r) = w_2, D_{j-1} = l, D_{k-1} = r)
\end{aligned}$$

and thus the right-hand side of (29) can be written as

$$\begin{aligned}
&\sum_{l,r\in\mathbb{Z}^d}\sum_{w_1,w_2\in\mathbb{Z}^d}\sum_{z_1,z_2\in\mathbb{Z}^d} (\mathbf{1}z_1)(\mathbf{1}z_2)\mathbb{P}(\xi_j(l) = z_1, \delta_j(l) = w_1) \\
&\quad \times\mathbb{P}(\xi_k(r) = z_2, \delta_k(r) = w_2, D_{j-1} = l, D_{k-1} = r)\mathbb{P}(D_i = 0|D_j = l + w_1) \\
&\qquad = \sum_{l,r\in\mathbb{Z}^d}\sum_{w_1,w_2\in\mathbb{Z}^d}\sum_{z_2\in\mathbb{Z}^d} \mathbf{1}\left\{\sum_{z_1\in\mathbb{Z}^d} z_1\mathbb{P}(\xi_j(l) = z_1, \delta_j(l) = w_1)\right\}(\mathbf{1}z_2) \\
&\qquad\quad \times\mathbb{P}(\xi_k(r) = z_2, \delta_k(r) = w_2, D_{j-1} = l, D_{k-1} = r)\mathbb{P}(D_i = 0|D_j = l + w_1).
\end{aligned}$$

Now, Remark 3.4 says that

$$\mathbb{P}(\xi_j(l) = z_1, \delta_j(l) = w_1) = \mathbb{P}(\xi_j(l) = -z_1, \delta_j(l) = w_1)$$

and thus the expression within braces above vanishes and consequently so does the second term on the right-hand side of (27). Using this and (28) in (27), we get that

$$\mathbb{V}(W_i) \leqslant \sigma^2(i-1)\,\mathbb{P}(D_{i-1} = 0). \tag{30}$$

Now, from (23),

$$\mathbb{V}(\bar{L}_n) \leqslant 4\sigma^2 \sum_{i=1}^{n-1} i\, \mathbb{P}(D_i = 0). \tag{31}$$

(a) This is already enough to obtain an upper bound of the form (9) in $d = 1$, if we use the following bound. For some c_4

$$\sum_{i=0}^{n-1} \mathbb{P}(D_i = 0) \leqslant c_4 n^{1/2}. \tag{32}$$

This was established in Ferrari and Fontes (1998, Lemma 3.3). From this and (31),

$$\mathbb{V}(\bar{L}_n) \leqslant 4\sigma^2 \sum_{i=1}^{n-1} i\, \mathbb{P}(D_i = 0) \leqslant 4\sigma^2 n \sum_{i=1}^{n-1} \mathbb{P}(D_i = 0) \leqslant 4c_4\sigma^2 n^{3/2}. \tag{33}$$

In $d \geqslant 2$, this argument does not give the correct order. We also do not know the asymptotic behavior of $\mathbb{P}(D_i = 0)$ as $i \to \infty$ (which we do in $d = 1$ (see Lemma 6.1); this would give another argument for (a) above). We thus make a more circuitous argument, via Lemma 3.3. By the latter result, there exists c_5 such that for all $0 \leqslant s < 1$

$$\sum_{i\geqslant 1} i\mathbb{P}(D_i = 0)s^i \leqslant c_5 \sum_{i\geqslant 1} i\mathbb{P}(H_i = 0)s^i. \tag{34}$$

We also use the well-known result that for every $d \geqslant 1$, there exists c_6 such that

$$\mathbb{P}(H_i = 0) \sim c_6 i^{-d/2}, \tag{35}$$

where, as usually, $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = 1$ (see, for example, Spitzer (1976, P7.9)). So, we get that for some c_7

$$\sum_{i\geqslant 1} i\mathbb{P}(D_i = 0)s^i \leqslant c_5 c_7 \sum_{i\geqslant 1} i^{1-d/2} s^i =: c_5 c_7 \varphi_d(s). \tag{36}$$

Now notice that

$$\sum_{i=1}^{n} i\mathbb{P}(D_i = 0) \leqslant 4 \sum_{i\geqslant 1} i\mathbb{P}(D_i = 0)(1 - 1/n)^i \tag{37}$$

for all $n \geqslant 2$. From this, (36) and (31), we get

$$\begin{aligned}\mathbb{V}(\bar{L}_n) &\leqslant 4\sigma^2 \sum_{i=1}^{n} i\mathbb{P}(D_i = 0)\\ &\leqslant 16c_5c_7\sigma^2 \sum_{i\geqslant 1} i^{1-d/2}(1 - 1/n)^i = 16c_5c_7\sigma^2\varphi_d(1 - 1/n).\end{aligned} \tag{38}$$

(b) $\varphi_2(s) = \sum_{i\geqslant 1} s^i = s/(1 - s)$. Thus, $\varphi_2(1 - 1/n) = n - 1$ and from this and (38) we have

$$\mathbb{V}(\bar{L}_n) \leqslant 16c_5c_7\sigma^2 n \quad \text{if } d = 2. \tag{39}$$

(c)

$$(\varphi_3(s))^2 = \left(\sum_{n\geqslant 1} \frac{1}{\sqrt{n}} s^n\right)^2 = \sum_{n\geqslant 1} \left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{i}\sqrt{n-i}}\right) s^n$$

$$= \sum_{n\geqslant 1} \left(\sum_{i=1}^{n-1} \frac{1}{\sqrt{i/n}\sqrt{1-i/n}} \frac{1}{n}\right) s^n. \tag{40}$$

Notice that the expression within parentheses on the right-hand side of (40) is a Riemann sum for a definite integral which equals π. Thus, there exists c_8 such that

$$(\varphi_3(s))^2 \leqslant \sum_{n\geqslant 1} c_8 s^n = c_8\, s/(1-s).$$

Thus $\varphi_3(1-1/n) \leqslant c_8^{1/2} n^{1/2}$ and from this and (38) we have

$$\mathbb{V}(\bar{L}_n) \leqslant 16 c_5 c_7 c_8^{1/2} \sigma^2 n^{1/2} \quad \text{if } d = 3. \tag{41}$$

(d) $\varphi_4(s) = \sum_{i\geqslant 1} s^i/i = \log[1/(1-s)]$. Thus $\varphi_4(1-1/n) = \log n$ and from this and (38) we have

$$\mathbb{V}(\bar{L}_n) \leqslant 16 c_5 c_7 \sigma^2 \log n \quad \text{if } d = 4. \tag{42}$$

(e) Finally, if $d \geqslant 5$ and $0 \leqslant s < 1$, then

$$\varphi_d(s) = \sum_{i\geqslant 1} i^{-(d/2-1)} s^i \leqslant \sum_{i\geqslant 1} i^{-3/2} < \infty.$$

We get from this and (38) that

$$\mathbb{V}(\bar{L}_n) \text{ is bounded in } n \text{ if } d \geqslant 5. \tag{43}$$

Gathering (33), (39) and (41)–(43), we have that there exists c_2 such that

$$\mathbb{V}(\bar{L}_n) \leqslant c_2 \mathfrak{D}(n,d). \tag{44}$$

4.1.2. Lower bound for $\mathbb{V}(\bar{L}_n)$

We only need to consider $d \leqslant 4$. By Jensen's inequality, the expectation in (24) is bounded from below by

$$\mathbb{E}^2[\mathbb{P}(Y_{i-1} = k | \mathscr{F}_{i-1})] = \mathbb{P}^2(Y_{i-1} = k) = \mathbb{P}(Y_{i-1} = Y'_{i-1} = k), \tag{45}$$

where Y' is an independent copy of Y. It follows that

$$\mathbb{V}(\bar{L}_n) \geqslant 4\sigma^2 \sum_{i=1}^{n-1} \sum_{k\in\mathbb{Z}^d} (\mathbf{1}k)^2 \mathbb{P}^2(Y_i = k). \tag{46}$$

Thus, if $\varepsilon > 0$

$$\mathbb{V}(\bar{L}_n) \geqslant 4\sigma^2 \varepsilon \sum_{i=1}^{n-1} i \sum_{k\in\mathbb{Z}^d:(\mathbf{1}k)^2 > \varepsilon i} \mathbb{P}^2(Y_i = k)$$

$$= 4\sigma^2\, \varepsilon \sum_{i=1}^{n-1} i \left[\sum_{k\in\mathbb{Z}^d} \mathbb{P}^2(Y_i = k) - \sum_{k\in\mathbb{Z}^d:(\mathbf{1}k)^2 \leqslant \varepsilon i} \mathbb{P}^2(Y_i = k)\right]. \tag{47}$$

Notice that $Y_n - Y_n'$ is a random walk distributed as H_n from the previous subsection. Thus,

$$\sum_{k\in\mathbb{Z}^d} \mathbb{P}^2(Y_i = k) = \sum_{k\in\mathbb{Z}^d} \mathbb{P}(Y_i = Y_i' = k) = \mathbb{P}(Y_i = Y_i') = \mathbb{P}(H_i = 0). \tag{48}$$

Using (48) in (47), we obtain

$$\mathbb{V}(\bar{L}_n) \geqslant 4\sigma^2 \varepsilon \sum_{i=1}^{n-1} i \left\{ \mathbb{P}(H_i = 0) - \sup_k \mathbb{P}(Y_i = k)\mathbb{P}((\mathbf{1}Y_i)^2 \leqslant \varepsilon i) \right\}. \tag{49}$$

Using (35) again, there exists c_9 such that $\mathbb{P}(H_i = 0) \geqslant c_9 i^{-d/2}$ for all i. It is also well known that there exists c_{10} such that $\sup_k \mathbb{P}(Y_i = k) \leqslant c_{10} i^{-d/2}$. Thus

$$\mathbb{V}(\bar{L}_n) \geqslant 4\sigma^2 \varepsilon \sum_{i=1}^{n-1} i^{1-d/2} \left[c_9 - c_{10}\, \mathbb{P}((\mathbf{1}Y_i/\sqrt{i})^2 \leqslant \varepsilon) \right]. \tag{50}$$

Now, by Berry–Esseen, applied to $\mathbf{1}Y_i$, which is a one-dimensional simple symmetric random walk, the above probability is bounded above by $O(\varepsilon) + O(1/\sqrt{i})$. We conclude that there exists $\varepsilon > 0$ for which the expression in brackets on the right-hand side of (50) is bounded below by a positive constant c_{11} for all large enough i. Thus, for some c_{12}, c_1, all n and $d = 1, 2, 3, 4$,

$$\mathbb{V}(\bar{L}_n) \geqslant 4c_{12}\varepsilon \sum_{i=1}^{n-1} i^{1-d/2} \geqslant c_1 \mathfrak{O}(n, d). \tag{51}$$

Now, by Proposition 2.3 of Ferrari and Fontes (1998),

$$\mathbb{V}(\bar{Z}_n) = \sigma^2 \sum_{j=0}^{n-1} \mathbb{P}(D_j = 0). \tag{52}$$

Corollary 3.4 of the same paper states that

$$\mathbb{V}(\bar{Z}_n) \text{ is of order } \begin{cases} n^{1/2} & \text{if } d = 1, \\ \log n & \text{if } d = 2, \\ \text{constant} & \text{if } d \geqslant 3. \end{cases} \tag{53}$$

Thus, from (51), for all $d \geqslant 1$, $\mathbb{V}(\bar{L}_n)$ dominates $\mathbb{V}(\bar{Z}_n)$. Since the order of $\mathbb{C}\mathrm{ov}(\bar{L}_n, \bar{Z}_n)$ is intermediate to the orders of $\mathbb{V}(\bar{Z}_n)$ and $\mathbb{V}(\bar{L}_n)$, it is also dominated by the latter. Thus, by (8), the order of $\mathbb{V}(L_n(x))$ is the same as the order of $\mathbb{V}(\bar{L}_n)$ and, from (51) and (44), the proof of Theorem 2.1 is complete. □

Note that by (31) and (52), if the non-degeneracy assumption on the u's does not hold, then $\mathbb{V}(\bar{L}_n(x))$ and $\mathbb{V}(\bar{Z}_n)$ vanish identically and thus so does $\mathbb{V}(L_n(x))$.

Remark 4.1. In the case of the voter model, $\gamma = \gamma(0,0) = 1$ and thus $D \equiv 0$. From (26), we get that $\mathbb{V}(\bar{L}_n) = \sigma^2 \sum_{i=0}^{n-1} \mathbb{E}[(\mathbf{1}S_i)^2]$, which one then easily estimates to be of the order of n^2 in all dimensions. From (52), we have that $\mathbb{V}(\bar{Z}_n)$ is of the order of n in all dimensions. Thus, by (8), $\mathbb{V}(L_n(x))$ is of the order of n^2 in all dimensions for all x.

4.2. Invariant measure in high dimensions and its fluctuations

Since in 5 and higher dimensions $\bar{L}_n - n$ is an L_2-bounded martingale, it converges almost surely as $n \to \infty$. The same holds for $\bar{Z}_n$ (see Ferrari and Fontes (1998, Lemma 2.2), and (53) above). We thus have from (7) that $L_n - n$ converges almost surely as $n \to \infty$, say to $\tilde{L}$. We thus have that the distribution of $\tilde{L}$ is invariant for the surface dynamics corresponding to applying the random averaging (1) and subtracting 1 at each site. A similar argument as that for Proposition 5.2 in Ferrari and Fontes (1998) can be made for that. We omit it here. From these facts and (7), we are able to argue the following:

Theorem 4.1. *The distribution of* $(\tilde{L}(x) - (\mathbf{1}x)^2)/|x|$ *converges weakly to the distribution of* $2\mu\tilde{Z}$ *as* $|x| \to \infty$, *where* $\tilde{Z}$ *is a non-trivial random variable and* μ *is a real number, provided* $\mathbf{1}x/|x| \to \mu$ *as* $|x| \to \infty$. *The variance of* $(\tilde{L}(x)-(\mathbf{1}x)^2)/|x|$ *converges to the variance of the weak limit.*

Proof. Let $L' = \lim_{n\to\infty} \bar{L}_n - n$ and $Z' = \lim_{n\to\infty} \bar{Z}_n$. Then, from (7), $\tilde{L}(x) = L'(x) + 2(\mathbf{1}x)Z'(x) + (\mathbf{1}x)^2$ and thus $(\tilde{L}(x) - (\mathbf{1}x)^2)/|x| = (L'(x)/|x|) + 2[(\mathbf{1}x)/|x|]Z'(x)$.

The distributions of both $L'(x)$ and $Z'(x)$ do not depend on x, since they are the limits of distributions that do not depend on x, and $L'(x)$ and $Z'(x)$ have finite second moments, by the L_2-Martingale Convergence Theorem. We conclude that $L'(x)/|x| \to 0$ as $|x| \to \infty$ in L_2. We take $\tilde{Z}=Z'(0)$. The non-triviality of $\tilde{Z}$ follows from the positivity of the variance of $Z'(0)$ (which equals $\sigma^2 \sum_{i=0}^{\infty} \mathbb{P}(D_i = 0)$; see (52) above). The result follows. □

5. Remark on higher degrees

The boundedness of the fluctuations in high dimensions, characteristic of the cases of parabolic (studied above) and linear initial conditions (Ferrari and Fontes, 1998), does not necessarily occur for a polynomial initial condition of higher degree. We illustrate this with the case of a cubic.

Proposition 5.1. *Let* $\hat{L}_n(x) = \mathbb{E}[(\mathbf{1}Y_n^x)^3|\mathscr{F}_n]$, $x \in \mathbb{Z}^d$. *For all high enough dimensions, there exists* c_{13} *such that for all* $n \geqslant 1$

$$\mathbb{V}(\hat{L}_n(0)) \geqslant c_{13}n^2. \tag{54}$$

Proof.

$$\begin{aligned}\hat{L}_n(0) = \mathbb{E}[(\mathbf{1}Y_n)^3|\mathscr{F}_n] &= \sum_{k\in\mathbb{Z}^d}\sum_{j=1}^{d}\sum_{\alpha=\pm1}[(\mathbf{1}k)^3 + 3\alpha(\mathbf{1}k)^2 + 3(\mathbf{1}k) + \alpha] \\ &\quad \times u_n(k, k+\alpha e_j)\mathbb{P}(Y_{n-1} = k|\mathscr{F}_{n-1}) \\ &= \hat{L}_{n-1} + 3\bar{Z}_{n-1} + 3\mathbb{E}[(\mathbf{1}Y_{n-1})^2(\mathbf{1}\theta_n(Y_{n-1}))|\mathscr{F}_n] \\ &\quad + \mathbb{E}[\mathbf{1}\theta_n(Y_{n-1})|\mathscr{F}_n],\end{aligned}$$

where $\bar{Z}_n$ is as in the previous sections. Since the distribution of the u_n's are symmetric and independent of $\mathscr{F}_{n-1}$, $\mathbb{E}(\hat{L}_n|\mathscr{F}_{n-1}) = \hat{L}_{n-1} + 3\bar{Z}_{n-1}$. From this and the fact that $\bar{Z}_n$ is a martingale we get that $\hat{L}_n - 3n\bar{Z}_n$ is a martingale. So $\hat{L}_n - 3n\bar{Z}_n = 3\sum_{i=1}^n \mathbb{E}[(\mathbf{1}Y_{i-1})^2(\mathbf{1}\theta_i(Y_{i-1}))|\mathscr{F}_i] + \sum_{i=1}^n \mathbb{E}[\mathbf{1}\theta_i(Y_{i-1})|\mathscr{F}_i]$. The latter sum equals $\bar{Z}_n$ (see Ferrari and Fontes, 1998, (2.24)). We thus get that $\hat{L}_n - 3(n+1)\bar{Z}_n$ is also a martingale with $\hat{L}_n - 3(n+1)\bar{Z}_n = 3\sum_{i=1}^n \hat{W}_i$, where $\hat{W}_i := \mathbb{E}[(\mathbf{1}Y_{i-1})^2(\mathbf{1}\theta_i(Y_{i-1}))|\mathscr{F}_i]$.

The variance of $\hat{W}_i$ can be (roughly) estimated as follows:

$$\mathbb{V}(\hat{W}_i) = \sigma^2 \mathbb{E} \sum_{k\in\mathbb{Z}^d} (\mathbf{1}k)^4 \mathbb{P}^2(Y_{i-1} = k|\mathscr{F}_{i-1})$$

$$= \sigma^2 \mathbb{E}[(\mathbf{1}S_{i-1})^4; D_{i-1} = 0] \leqslant 16\sigma^2(i-1)^4\, \mathbb{P}(D_{i-1} = 0),$$

where in the above inequality we used the fact that S has jumps of length at most 2. Thus $\mathbb{V}(\hat{L}_n)$ is bounded above by constant times $\sum_{i=1}^\infty i^4 \mathbb{P}(D_i = 0)$. From Lemma 3.3, we can obtain an upper bound for the latter sum by replacing $\mathbb{P}(D_i = 0)$ in it by constant times $\mathbb{P}(H_i = 0)$. We conclude that the sum is bounded by constant times $\sum_{i=1}^\infty i^{4-d/2}$, which is finite if $d \geqslant 11$. On the other hand, (52) and (53) above tell us that the fluctuations of $\bar{Z}_n$ are positive and bounded for $d \geqslant 3$ and thus $\mathbb{V}(\hat{L}_n)$ is of order n^2, if $d \geqslant 11$. □

Remark 5.1. It is possible, with a similar estimation as the one done in Section 4.1.1, to get a sharper estimate of the variance of $\sum_{i=1}^n \hat{W}_i$ and then obtain an upper bound of constant times $\sum_{i=1}^n i^{2-d/2}$ for $\mathbb{V}(\hat{L}_n)$. This and (53) above would imply that the term $3(n+1)\bar{Z}_n$ gives the dominating contribution for the variance of $\hat{L}_n$ in $d \geqslant 2$ and thus the conclusion of Proposition 5.1 would hold for these dimensions. (The case of $d=1$ would demand further analysis, since the contributions of $3(n+1)\bar{Z}_n$ and $\sum_{i=1}^n \hat{W}_i$ would be of the same order and one would have to rule out cancellations.) We chose not to make an exhaustive analysis here, but rather to indicate the phenomenon of unboundedness of the fluctuations in all high dimensions as simply as possible.

Remark 5.2. Unboundedness of the fluctuations in high dimensions should also occur for higher degree initial polynomials such as $X_0(x) = (\mathbf{1}x)^k$, $k \geqslant 4$. Results similar to Proposition 5.1 should hold for those cases, with similar arguments.

6. Central limit theorem

In $d = 1$, the following holds:

Lemma 6.1. *Let D be as in the previous sections. Then*

$$\mathbb{P}(D_n = 0|D_0 = 0) \simeq n^{-1/2}. \tag{55}$$

From here on, $a_n \simeq b_n$ means that $\lim_{n\to\infty} a_n/b_n$ exists and is positive.

Proof. This is loosely argued in Remark 3.3 of Ferrari and Fontes (1998). We repeat that argument. The result follows from P20.2 of Spitzer (1976), Lemma 3.3 and the fact that (1) of P20.2 of Spitzer (1976) holds for H. We leave more details to the reader. □

Proof of Theorem 2.3. To prove Theorem 2.3, it is enough that we verify the two conditions of the corollary to Theorem 3.2 of Hall and Heyde (1980). In the notation of that reference, $X_{ni}=\mathscr{V}_n^{-1/2}(2W_i+1)$ and $\mathscr{F}_{ni}=\mathscr{F}_i$ for all i,n. We will be sketchy, since most of the arguments are either standard or have been employed (in essence) by Ferrari and Fontes (1998). We recall that $W_i = \mathbb{E}[Y_{i-1}\theta_i(Y_{i-1})|\mathscr{F}_i] = \sum_k k\theta_i(k)\mathbb{P}(Y_{i-1} = k|\mathscr{F}_{i-1})$.

6.1. First condition

Since $\mathscr{V}_n$ is of the order of $n^{3/2}$, it is enough to show that

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}(W_i^2;|W_i|>\varepsilon n^{3/4}|\mathscr{F}_{i-1})\to 0 \tag{56}$$

in probability as $n\to\infty$ for any $\varepsilon>0$. For that, it suffices to prove that

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}(W_i^2;|W_i|>\varepsilon n^{3/4})\to 0 \tag{57}$$

as $n\to\infty$, and this can be achieved by repeated use of the Cauchy–Schwarz inequality and the Chebyshev inequality, together with standard bounds on moments of the simple symmetric random walk and (55). We leave details to the reader.

6.2. Second condition

In the notation of Hall and Heyde (1980), it can be written as

$$V_n^2 := \mathscr{V}_n^{-1}\sum_{i=1}^{n}(\mathbb{E}[(2W_i+1)^2|\mathscr{F}_{i-1}]-\mathbb{E}[(2W_i+1)^2])\to 0 \tag{58}$$

in probability as $n\to\infty$. Since $\mathscr{V}_n$ is of the order of $n^{3/2}$, and also noting that $\mathbb{E}(W_i|\mathscr{F}_{i-1})=0$ for all $i\geqslant 1$, it is enough to show that

$$n^{-3/2}\sum_{i=1}^{n}[\mathbb{E}(W_i^2|\mathscr{F}_{i-1})-\mathbb{E}(W_i^2)]\to 0 \tag{59}$$

in probability as $n\to\infty$. The above expression can be written as

$$\sigma^2 n^{-3/2}\sum_{i=0}^{n-1}(\mathbb{E}(S_i^2;D_i=0|\mathscr{F}_i)-\mathbb{E}(S_i^2;D_i=0)). \tag{60}$$

We will argue that the variance of the latter expression tends to 0 as $n \to \infty$. We write the variance of the sum as

$$\sum_{j=1}^{n-1} \mathbb{E}\left(\sum_{i=j}^{n-1}[\mathbb{E}(S_i^2; D_i = 0|\mathscr{F}_j) - \mathbb{E}(S_i^2; D_i = 0|\mathscr{F}_{j-1})]\right)^2.$$

After some calculation, the inner sums on the right-hand side of (61) can be rewritten as

$$\sum_{l,l'}[A^n_{j,l,l'}\overline{u_{j,l'}(1-u_{j,l})} - A^n_{j,l,l'-2}\overline{u_{j,l'}} + A^n_{j,l+2,l'}\overline{u_{j,l'}u_{j,l}}]$$
$$\times\mathbb{P}(Y_{j-1} = l, \hat{Y}_{j-1} = l'|\mathscr{F}_{j-1}), \tag{61}$$

where

$$A^n_{j,l,l'} := \sum_{i=j}^{n-1}(\mathbb{E}(S_i^2; D_i = 0|S_j = l + l', D_j = l - l' - 2)$$
$$- \mathbb{E}(S_i^2; D_i = 0|S_j = l + l' - 2, D_j = l - l')) \tag{62}$$

and bar means centering (that is, $\bar{X} = X - \mathbb{E}(X)$).

We present next an estimation of $A^n_{j,l,l'}$. After that, the rest of the argument of the proof of the second condition, which relies on standard use of the Cauchy–Schwarz inequality, bounds on moments of the simple symmetric random walk and (55), will be left to the reader.

We begin by observing that, by Remark 3.4, we can represent S_n in one dimension as $\sum_{i=1}^n \hat{\xi}_i\eta_i(D_{i-1})$, where $\hat{\xi}_1, \hat{\xi}_2, \ldots$ are i.i.d. with $\mathbb{P}(\hat{\xi}_1 = 2) = \mathbb{P}(\hat{\xi}_1 = -2) = 1/2$ and $\eta_i(l) \stackrel{d}{=} \mathbb{1}(\xi_i(l) \neq 0)$ for all i, l, where $\mathbb{1}(\cdot)$ is the indicator function. Clearly, $\eta_i = \mathbb{1}(S_i \neq S_{i-1})$. Now, from the nearest-neighbor character of the jumps of $Y, \hat{Y}$, we have that $\mathbb{1}(S_i \neq S_{i-1}) = \mathbb{1}(D_i = D_{i-1})$. We can thus rewrite the summands of (62) as

$$\mathbb{E}[(l + l' + S_i - S_j)^2; D_i = 0|D_j = l - l' - 2]$$
$$- \mathbb{E}[(l + l' - 2 + S_i - S_j)^2; D_i = 0|D_j = l - l']$$
$$= (l + l')^2[\mathbb{P}(D_i = 0|D_j = l - l' - 2) - \mathbb{P}(D_i = 0|D_j = l - l')]$$
$$+ 4(l + l' - 1)\mathbb{P}(D_i = 0|D_j = l - l')$$
$$+ 4\sum_{k=j+1}^{i}[\mathbb{P}(D_i = 0, D_k = D_{k-1}|D_j = l - 2)$$
$$- \mathbb{P}(D_i = 0, D_k = D_{k-1}|D_j = l)].$$

From now on, we leave it to the reader to verify that the absolute value of the sum in i from j to $n - 1$ of the above terms is bounded above by constant times $(l + l')^2$, for the first term; $|l + l' - 1|\sqrt{n}$, for the second one; and n, for the third one. Two ingredients one can use for this is the uniform boundedness of $\sum_{i=j}^n a_{i,j,l,l'}$ in l, l', j

and n, where $a_{i,j,l,l'} := \mathbb{P}(D_i = 0|D_j = l - l' - 2) - \mathbb{P}(D_i = 0|D_j = l - l')$ and the fact that $\mathbb{P}(D_n = 0|D_0 = m) \leqslant \mathbb{P}(D_n = 0|D_0 = 0)$ for all m (as shown in Ferrari and Fontes, 1998).

We conclude that $|A^n_{j,l,l'}|$ is bounded above by constant times

$$(l + l')^2 + |l + l' - 1|\sqrt{n} + n \tag{63}$$

and the estimation of $A^n_{j,l,l'}$ is complete.

Acknowledgements

We thank P. Ferrari for many discussions on these and related questions and models. We also thank an anonymous referee, whose suggestions led to improvements in this presentation.

References

Andjel, E., 1985. Invariant measures and long time behaviour of the smoothing process. Ann. Probab. 13, 62–71.

Durrett, R., 1996. Stochastic Spatial Models, PCMI Lecture Notes. IAS, Princeton.

Ferrari, P.A., Fontes, L.R.G., 1998. Fluctuations of a surface submitted to a random average process. Electron. J. Probab. 3 (6), 1–34.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.

Hammersley, J.M., 1965/1966. Harnesses, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, California, Vol. III: Physical Sciences. University of California Press, Berkeley, CA, pp. 89–117.

Liggett, T.M., 1985. Interacting Particle Systems. Springer, Berlin.

Liggett, T.M., Spitzer, F., 1981. Ergodic theorems for coupled random walks and other systems with locally interacting components. Z. Wahrsch. Verw. Gebiete 56, 443–468.

Medeiros, D.P., 2001. Processo de médias aleatórias com configuração inicial parabólica. Ph.D. Thesis, University of São Paulo, 151pp (in Portuguese). Version in http://www.ime.usp.br/∼lrenato/tese.ps, 128pp (in Portuguese).

Spitzer, F., 1976. Principles of Random Walk. Springer, New York.

Toom, A., 1997. Tails in harnesses. J. Statist. Phys. 88, 347–364.