Cambridge University Press
978-1-107-02669-8 — Non-homogeneous Random Walks
Mikhail Menshikov , Serguei Popov , Andrew Wade
Excerpt
More Information

1

Introduction

1.1 Random Walks

Random walks are fundamental models in probability theory that exhibit deep mathematical properties and enjoy broad application across the sciences and beyond. Generally speaking, a random walk is a stochastic process modelling the random motion of a particle (or *random walker*) in space. The particle's trajectory is described by a series of random *increments* or *jumps* at discrete instants in time. Central questions for these models involve the long-time asymptotic behaviour of the walker.

Random walks have a rich history involving several disciplines. Classical one-dimensional random walks were first studied several hundred years ago as models for games of chance, such as the so-called gambler's ruin problem. Similar reasoning led to random walk models of stock prices described by Jules Regnault in his 1863 book [265] and Louis Bachelier in his 1900 thesis [14]. Many-dimensional random walks were first studied at around the same time, arising from the work of pioneers of science in diverse applications such as acoustics (Lord Rayleigh's theory of sound developed from about 1880 [264]), biology (Karl Pearson's 1906 [254] theory of random migration of species), and statistical physics (Einstein's theory of Brownian motion developed during 1905–8 [86]). The mathematical importance of the random walk problem became clear after Pólya's work in the 1920s, and over the last 60 years or so there have emerged beautiful connections linking random walk theory and other influential areas of mathematics, such as harmonic analysis, potential theory, combinatorics, and spectral theory. Random walk models have continued to find new and important applications in many highly active domains of modern science: see for example the wide range of articles in [287]. Specific recent developments include modelling of microbe locomotion in microbiology [23, 245], polymer conformation in molecular chemistry [15, 202], and financial systems in economics.

Spatially homogeneous random walks are the subject of a substantial literature, including [139, 195, 269, 293]. In many modelling applications, the classical assumption of spatial homogeneity is not realistic: the behaviour of the random walker may depend on the present location in space. Applications thus motivate the study of *non-homogeneous* random walks. These models are also motivated naturally from a mathematical perspective: non-homogeneous random walks are the natural setting in which to probe near-critical behaviour and obtain a finer understanding of phase transitions present in the classical random walk models.

The main theme of this book is the analysis of near-critical stochastic systems using the method of Lyapunov functions. The non-homogeneous random walk serves as a prototypical near-critical system; the Lyapunov function methodology is robust and powerful, and can be applied to many other near-critical models, including those with applications across modern probability and beyond, to areas such as queueing theory, interacting particle systems, and random media. In this chapter we give an informal introduction to non-homogeneous random walks, and how their behaviour differs from classical random walks; we also describe some fundamental ideas of the Lyapunov function technique. We state some theorems, but we often omit technical details and generally omit proofs. All of the results that we mention will be stated more precisely (and proved) later in the book, and also applied to a wide variety of near-critical stochastic systems: the non-homogeneous random walk serves as an expository bridge between well-known classical results and the near-critical behaviour that is the subject of this book.

1.2 Simple Random Walk

The most intensively studied random walk model is the symmetric *simple random walk*. Simple random walk is a discrete-time Markov process $(S_n, n \geq 0)$ on the d-dimensional integer lattice $\mathbb{Z}^d$: S_n can be thought of as the location (in the state space $\mathbb{Z}^d$) of the random walker at time n (or after n steps). The stochastic evolution of the process is as follows. Given S_n in $\mathbb{Z}^d$, the next point S_{n+1} is chosen uniformly at random from among the $2d$ lattice points adjacent to S_n, i.e., those points that differ from S_n by exactly ± 1 in a single coordinate. In other words, the transition probabilities of the Markov chain are given for $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^d$ by

$$\mathbb{P}[S_{n+1} = \mathbf{y} \mid S_n = \mathbf{x}] = \begin{cases} \frac{1}{2d} & \text{if } \|\mathbf{x} - \mathbf{y}\| = 1; \\ 0 & \text{otherwise;} \end{cases}$$

where $\| \cdot \|$ is the Euclidean norm on $\mathbb{R}^d$.

For example, when $d = 1$ the one-dimensional simple random walk jumps one unit to the left or right, each with probability $1/2$, while when $d = 2$ the two-dimensional simple random walk jumps to one of its four neighbours, with probability $1/4$ of each.

For definiteness, suppose that the walk starts at the origin of $\mathbb{Z}^d$: $S_0 = \mathbf{0}$. By the Markov property and spatial homogeneity, the increments (or jumps) $S_{n+1} - S_n$ of the walk are independent and identically distributed (i.i.d.) random vectors. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_d\}$ is the standard orthonormal basis on $\mathbb{R}^d$, let

$$U_d := \{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_d\}$$

for the possible values of the increments of the walk. Then we can write $S_n = \sum_{k=1}^{n} Z_k$, where $Z_1, Z_2, \ldots$ are i.i.d. with

$$\mathbb{P}[Z_1 = \mathbf{e}] = \frac{1}{2d} \quad \text{for } \mathbf{e} \in U_d.$$

(With the usual convention that an empty sum is zero, $S_0 = \mathbf{0}$.) Thus we may represent the random walk S_n via a sequence of *partial sums* of the i.i.d. random increments Z_n.

A fundamental question, addressed by Pólya [259], concerns the *recurrence* or *transience* of the random walk: what is the probability that the walk eventually returns to $\mathbf{0}$? If we write $\tau_d := \min\{n \geq 1 : S_n = \mathbf{0}\}$ for the time of the first return to $\mathbf{0}$ (with the usual convention that $\min \emptyset := \infty$), the recurrence question concerns

$$p_d := \mathbb{P}[\tau_d < \infty].$$

The random walk is recurrent if $p_d = 1$, in which case with probability one the random walk will visit $\mathbf{0}$ infinitely often. On the other hand, if $p_d < 1$ the random walk is transient, and will, with probability one, visit $\mathbf{0}$ only finitely many times, before eventually leaving, never to return.

The following fundamental result is due to Pólya [259].

Theorem 1.2.1 *Simple random walk is recurrent in 1 or 2 dimensions, but transient in 3 or more; i.e.,* $p_1 = p_2 = 1$ *but* $p_d < 1$ *for all* $d \geq 3$. □

The content of the theorem is nicely captured by an aphorism attributed to Shizuo Kakutani: 'A drunk man will eventually find his way home, but a drunk bird may get lost forever' (see [83, p. 191]).

Pólya's theorem (Theorem 1.2.1) tells us that the walk returns to $\mathbf{0}$ eventually when $d = 1$ or $d = 2$. But how long might we have to wait? The answer is, potentially, a very long time, since τ_d has very *heavy tails*:

$$\mathbb{P}[\tau_1 > n] \sim \frac{1}{\sqrt{\pi n}}, \text{ and } \mathbb{P}[\tau_2 > n] \sim \frac{\pi}{\log n}, \quad \text{as } n \to \infty;$$

see e.g. [259, p. 159] for the first expression and [84, pp. 356–7] for the second. So in $d = 1$, $\mathbb{E}[\tau_1^{1/2}] = \infty$, while in $d = 2$, τ_2 has no moments at all. According to Hughes [139, p. 42],

> the failure of certain moments of distributions or densities to converge … is pregnant with physical meaning, and indicative of connections to scaling laws, renormalization group methods, and fractals.

The recurrence exhibited by the simple random walk for $d \in \{1, 2\}$ is *null recurrence*, meaning that $\mathbb{E}\,\tau_d = \infty$; more stable processes may exhibit *positive recurrence*, meaning that the analogue of τ_d is integrable.

1.3 Lamperti's Problem

There are several proofs of Theorem 1.2.1 in the literature, the most popular being those that are largely combinatorial (such as Pólya's original argument [259]) and those based on potential theory and electrical networks (see e.g. [81]). A drawback of each of these approaches is that they rapidly break down when one tries to generalize Pólya's theorem to other random walks. In this section we describe a robust approach to proving Pólya's theorem, due to Lamperti, which enables very broad generalization. This approach is based on the methodology of *Lyapunov functions*.

Again let S_n be the symmetric simple random walk on $\mathbb{Z}^d$, starting at $\mathbf{0}$. In the context of Pólya's recurrence theorem, we are interested in the events $\{S_n = \mathbf{0}\}$. We can reduce this d-dimensional problem to a one-dimensional problem by considering a transformation of the process (a Lyapunov function) given by

$$X_n := \|S_n\|, \tag{1.1}$$

i.e., X_n is the distance from the origin of the walker at time n. The stochastic process $(X_n, n \geq 0)$ takes values in the countable set $\mathcal{S} = \{\|\mathbf{x}\| : \mathbf{x} \in \mathbb{Z}^d\}$, a subset of the half-line $\mathbb{R}_+$, and $X_n = 0$ if and only if $S_n = \mathbf{0}$. So we can study the recurrence or transience of S_n via the recurrence or transience of X_n, a one-dimensional process.

This reduction in dimensionality of the problem comes at a price: X_n is not in general a Markov process. For instance, when $d = 2$, given one of the two

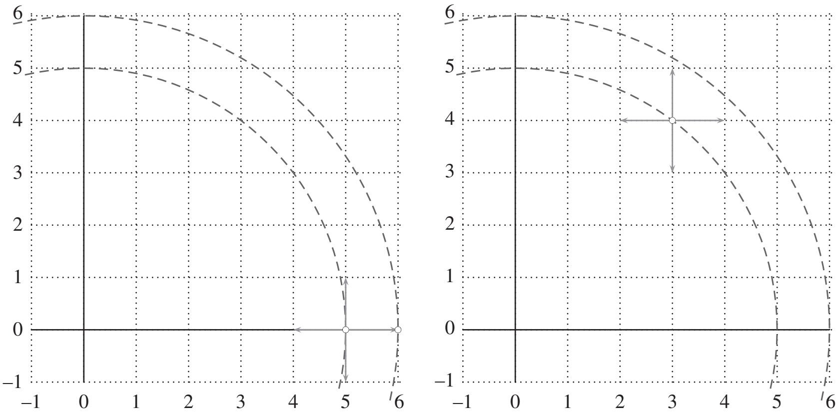

Figure 1.1 Illustration of the non-Markovian nature of X_n in $d = 2$.

events $\{S_n = (5,0)\}$ and $\{S_n = (3,4)\}$ we have $X_n = 5$ in each case but X_{n+1} has two different distributions; X_{n+1} can take the value 6 (with probability 1/4) in the first case, but this is impossible in the second case. See Figure 1.1. Thus the tools that we use to study X_n must not rely too heavily on the Markov property.

First we compute the expected increment of X_n given $S_n = \mathbf{x}$, namely

$$\mathbb{E}[X_{n+1} - X_n \mid S_n = \mathbf{x}] = \frac{1}{2d} \sum_{i=1}^{d} (\|\mathbf{x} + \mathbf{e}_i\| + \|\mathbf{x} - \mathbf{e}_i\| - 2\|\mathbf{x}\|) . \tag{1.2}$$

To proceed we apply Taylor's theorem in an elementary way. Using the Taylor expansion

$$(1+y)^{1/2} = 1 + \frac{1}{2}y - \frac{1}{8}y^2 + O(y^3), \text{ as } y \to 0,$$

we obtain that, for any $\mathbf{e} \in \mathbb{S}^{d-1}$,

$$\begin{aligned} \|\mathbf{x} + \mathbf{e}\| - \|\mathbf{x}\| &= \sqrt{(\mathbf{x}+\mathbf{e}) \cdot (\mathbf{x}+\mathbf{e})} - \|\mathbf{x}\| \\ &= \|\mathbf{x}\| \left[\left(1 + \frac{2\mathbf{e} \cdot \mathbf{x} + 1}{\|\mathbf{x}\|^2} \right)^{1/2} - 1 \right] \\ &= \|\mathbf{x}\| \left(\frac{2\mathbf{e} \cdot \mathbf{x} + 1}{2\|\mathbf{x}\|^2} - \frac{(\mathbf{e} \cdot \mathbf{x})^2}{2\|\mathbf{x}\|^4} + O(\|\mathbf{x}\|^{-3}) \right) . \end{aligned} \tag{1.3}$$

It follows that

$$\|\mathbf{x} + \mathbf{e}\| + \|\mathbf{x} - \mathbf{e}\| - 2\|\mathbf{x}\| = \frac{1}{\|\mathbf{x}\|} - \frac{(\mathbf{e} \cdot \mathbf{x})^2}{\|\mathbf{x}\|^3} + O(\|\mathbf{x}\|^{-2}).$$

Thus from (1.2), using the fact that $\sum_{i=1}^{d} (\mathbf{e}_i \cdot \mathbf{x})^2 = \|\mathbf{x}\|^2$, we obtain

$$\mathbb{E}[X_{n+1} - X_n \mid S_n = \mathbf{x}] = \frac{1}{2d} \sum_{i=1}^{d} (\|\mathbf{x} + \mathbf{e}_i\| + \|\mathbf{x} - \mathbf{e}_i\| - 2\|\mathbf{x}\|)$$
$$= \left(\frac{d-1}{2d}\right) \frac{1}{\|\mathbf{x}\|} + O(\|\mathbf{x}\|^{-2}). \tag{1.4}$$

Similarly

$$\mathbb{E}[X_{n+1}^2 - X_n^2 \mid S_n = \mathbf{x}] = \frac{1}{2d} \sum_{i=1}^{d} \left(\|\mathbf{x} + \mathbf{e}_i\|^2 + \|\mathbf{x} - \mathbf{e}_i\|^2 - 2\|\mathbf{x}\|^2\right) = 1.$$

Then, since $(X_{n+1} - X_n)^2 = X_{n+1}^2 - X_n^2 - 2X_n(X_{n+1} - X_n)$ we obtain

$$\mathbb{E}[(X_{n+1} - X_n)^2 \mid S_n = \mathbf{x}] = \frac{1}{d} + O(\|\mathbf{x}\|^{-1}). \tag{1.5}$$

Informally speaking, (1.4) says that the mean increment of X_n at $x \in \mathcal{S}$ is $\frac{1}{2x}(1 - \frac{1}{d}) + O(x^{-2})$, and similarly (1.5) says that the second moment of the increment at x is $\frac{1}{d} + O(x^{-1})$; however, to formalize these statements we need to be careful since X_n is not a Markov process, and we have to clarify what we mean by the increment moments 'at x'. We deal with these technicalities later, since they complicate the notation (although they will not complicate the proofs, which are based on martingale arguments). For now, for the purposes of exposition, we switch to the case where X_n *is* a Markov process.

Suppose now that $(X_n, n \geq 0)$ is a time-homogeneous Markov process on an unbounded subset $\mathcal{S}$ of $\mathbb{R}_+$. Consider the *increment moment functions*

$$\mu_k(x) := \mathbb{E}[(X_{n+1} - X_n)^k \mid X_n = x].$$

For simplicity, suppose that X_n has uniformly bounded increments, so that

$$\mathbb{P}[|X_{n+1} - X_n| \leq B] = 1 \tag{1.6}$$

for some $B \in \mathbb{R}_+$; under condition (1.6), the μ_k are well-defined functions of $x \in \mathcal{S}$. The first moment function, $\mu_1(x)$, is the one-step *mean drift* of X_n at x.

Lamperti [190, 191, 192] investigated the extent to which the asymptotic behaviour of such a process is determined by the μ_k; essentially, μ_1 and μ_2 turn out to govern the asymptotic behaviour. For example, the following result is a version Lamperti's fundamental recurrence classification.

Theorem 1.3.1 *Suppose that X_n is a Markov process on $\mathcal{S}$ satisfying* (1.6). *Under mild conditions, the following recurrence classification holds. Let $\varepsilon > 0$.*

- *If* $2x\mu_1(x) + \mu_2(x) < -\varepsilon$, *then* X_n *is positive recurrent;*
- *If* $2x|\mu_1(x)| \le \mu_2(x) + O(x^{-\varepsilon})$, *then* X_n *is null recurrent;*
- *If* $2x\mu_1(x) - \mu_2(x) > \varepsilon$, *then* X_n *is transient.* □

The mild conditions mentioned in the statement of Theorem 1.3.1 are related to issues of irreducibility; the exact nature of the conditions required depends on the state space $\mathcal{S}$ and the notion of recurrence and transience desired. We leave the technical details until later in the book.

A version of Theorem 1.3.1 applies to non-Markov processes X_n when formulated correctly, using appropriate versions of the μ_k. In particular, we can use such results to study the process X_n defined by (1.1): in this case, the analogue of $2x\mu_1(x)$ is, by (1.4),

$$1 - \frac{1}{d} + O(x^{-1})$$

and the analogue of $\mu_2(x)$ is, by (1.5),

$$\frac{1}{d} + O(x^{-1}).$$

An application of the generalized version of Theorem 1.3.1 shows that X_n is transient if and only if

$$1 - \frac{1}{d} > \frac{1}{d},$$

or, in other words, $d > 2$. This gives a very robust strategy for proving Pólya's theorem (Theorem 1.2.1) and its generalizations, based on computations of increment moments for X_n defined by (1.1). These computations use elementary Taylor's theorem ideas, and do not rely at all on special structure of the original process S_n.

1.4 General Random Walk

Simple random walk is an attractive model and can be studied using combinatorial methods based on counting sample paths, for example; it is, however, a very specific model. Naturally, it is of interest to study a much broader class of random walks. In particular, for what class of models does a result similar to Theorem 1.2.1 hold? To put the question in another way, what are the essential properties possessed by simple random walk that imply Theorem 1.2.1? To answer such questions, we start by describing a much more general model of a random walk.

By a *random walk* we mean a discrete-time Markov process $(\xi_n, n \ge 0)$ on an unbounded state space $\Sigma \subseteq \mathbb{R}^d$. We assume that the random walk is time homogeneous. This means that the distribution of ξ_{n+1} given $(\xi_0, \xi_1, \ldots, \xi_n)$ depends only on ξ_n (and not on n). We also need some form of irreducibility to ensure that the random walk cannot get 'trapped' in some part of the state space: it is simplest to take Σ to be a locally finite set (such as $\mathbb{Z}^d$) to avoid technical issues at this point.

We are thus using the term random walk in a rather general sense, requiring that the Markov process inhabit Euclidean space. We also want to impose some regularity assumptions on the increments of the process, to rule out very long jumps for the random walk. For this chapter, for convenience we often suppose that the increments of the walk are uniformly bounded, so that there exists $B \in \mathbb{R}_+$ for which, almost surely (a.s.),

$$\mathbb{P}[\|\xi_{n+1} - \xi_n\| \leq B \mid \xi_n = \mathbf{x}] = 1, \quad \text{for all } \mathbf{x} \in \Sigma. \tag{1.7}$$

In many cases this assumption can be replaced by a weaker assumption on the existence of higher moments for the increments, without producing fundamentally new behaviour. On the other hand, the case of genuinely heavy-tailed increments leads to different phenomena, as discussed in Chapter 5.

Under the assumption (1.7), the random vectors $\xi_{n+1} - \xi_n$ have well-defined moments, which may depend on ξ_n. In particular, an important quantity is the one-step *mean drift* vector

$$\mu(\mathbf{x}) := \mathbb{E}[\xi_{n+1} - \xi_n \mid \xi_n = \mathbf{x}],$$

which is the average change in position in a single step starting from $\mathbf{x} \in \Sigma$. (Note that the definition via the conditional expectation on $\{\xi_n = \mathbf{x}\}$ is clear in the case of a countable state space Σ, and makes sense when correctly interpreted for uncountable Σ as well.)

An important and well-studied subclass of random walks are *spatially homogeneous*, for which the distribution of the increment $\xi_{n+1} - \xi_n$ does not depend on the current location ξ_n. Writing $\theta_n := \xi_{n+1} - \xi_n$, spatial homogeneity implies that $\theta_0, \theta_1, \ldots$ are i.i.d. random vectors. Then the representation

$$\xi_n = \sum_{k=0}^{n-1} \theta_k \tag{1.8}$$

as a sum of i.i.d. random vectors enables classical tools of probability theory, such as Fourier methods, to be brought to bear in the analysis of the random walk in the spatially homogeneous case.

Example 1.4.1 (Simple random walk) Using $\mathbf{e}_1, \ldots, \mathbf{e}_d$ to denote the standard orthonormal basis vectors of $\mathbb{R}^d$, simple random walk is spa-

Cambridge University Press
978-1-107-02669-8 — Non-homogeneous Random Walks
Mikhail Menshikov , Serguei Popov , Andrew Wade
Excerpt
More Information

tially homogeneous with θ_n uniformly distributed on the set $U_d = \{\pm \mathbf{e}_1, \ldots, \pm \mathbf{e}_d\}$. △

Example 1.4.2 (Pearson–Rayleigh random walk) Take θ_n to be uniformly distributed on the unit-radius sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. The corresponding d-dimensional random walk, which proceeds via a sequence of unit-length steps, each in an independent and uniformly random direction, is sometimes called a *Pearson–Rayleigh* random walk: see the bibliographical notes at the end of this chapter. △

Spatially homogeneous random walks have been extensively studied in classical probability theory. Non-homogeneous random walks, on the other hand, require new techniques and, just as importantly, new intuitions.

1.5 Recurrence and Transience

For our general random walks, we need a more general definition of recurrence and transience. In the absence of any structural assumptions, our basic use of the terminology is as follows. Note that, in general, the two behaviours are not *a priori* exhaustive.

Definition 1.5.1 A stochastic process $(\xi_n, n \geq 0)$ taking values in $\Sigma \subseteq \mathbb{R}^d$ is *transient* if $\lim_{n\to\infty} \|\xi_n\| = \infty$, a.s. The process is *recurrent* if, for some constant $r_0 \in \mathbb{R}_+$, $\liminf_{n\to\infty} \|\xi_n\| \leq r_0$, a.s.

If ξ_n is an irreducible time-homogeneous Markov chain whose state space is a locally finite subset of $\mathbb{R}^d$ (such as $\mathbb{Z}^d$), then recurrence and transience in the sense of Definition 1.5.1 coincide with the usual Markov chain definition in terms of returns to any given state. Definition 1.5.1 allows more general processes, and is the most convenient definition in the context of Lamperti's problem outlined in Section 1.3.

First we return to the classical spatially homogeneous random walk described in Section 1.4, in which case the increments θ_n in (1.8) are i.i.d. In this case, when defined, $\mathbb{E}[\xi_{n+1} - \xi_n \mid \xi_n = \mathbf{x}] = \mathbb{E}\,\theta_0 = \mu$ does not depend on $\mathbf{x}$.

If $\mu \neq \mathbf{0}$, then the strong law of large numbers shows that the walk is transient, and $\lim_{n\to\infty} n^{-1}\xi_n = \mu$, a.s., so the walk escapes to infinity at positive speed. The most subtle case is that of *zero drift* when $\mu = \mathbf{0}$. Here, under mild conditions, Pólya's theorem (Theorem 1.2.1) for simple symmetric random walk extends to the case of general spatially homogeneous random walks with zero drift. Recall that we view θ_0 and other d-dimensional vectors as column vectors throughout.

Theorem 1.5.2 *For a spatially homogeneous random walk ξ_n on $\mathbb{R}^d$, suppose that $\mathbb{E}[\|\theta_0\|^2] < \infty$, $\mathbb{E}\,\theta_0 = \mathbf{0}$, and $\mathbb{E}[\theta_0\theta_0^\top]$ is positive definite. Then ξ_n is recurrent for $d \in \{1, 2\}$ and transient for $d \geq 3$.* □

The positive-definite covariance condition ensures that the increments are not supported on a lower-dimensional subspace.

How does the situation change if the walk is allowed to be *spatially non-homogeneous*? In the general, non-homogeneous case, $\mu(\mathbf{x}) = \mathbb{E}[\theta_n \mid \xi_n = \mathbf{x}]$ will depend on $\mathbf{x}$. *Even if* $\mu(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x}$, however, the non-homogeneous zero-drift walk can behave completely differently to the homogeneous zero-drift walk.

Theorem 1.5.3 *Let ξ_n be a spatially non-homogeneous random walk on $\mathbb{R}^d$ with zero drift, so that $\mu(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x}$.*

- *If $d = 2$, then we can exhibit such a walk that is* transient.
- *If $d \geq 3$, then we can exhibit such a walk that is* recurrent.

In all cases, we may take these examples to have uniformly bounded increments, as at (1.7). □

We emphasize that, for example, in two dimensions, zero drift does *not* imply recurrence for a *non-homogeneous* random walk with bounded jumps. This fact is contrary to intuition built from homogeneous random walks, but should not be surprising to readers who have encountered random walks in random environments: for example, Zeitouni [319, pp. 90–91] discusses an example of a transient walk in $d = 2$ with symmetric increments; see also the examples in Chapter 4.

So non-homogeneous random walks can show *anomalous* (non-classical) recurrence behaviour. We can reassert some control by imposing additional regularity structure on the second moments of the increments $\theta_n = \xi_{n+1} - \xi_n$.

Assuming (1.7), then the matrix function

$$M(\mathbf{x}) = \mathbb{E}[\theta_n\theta_n^\top \mid \xi_n = \mathbf{x}] \tag{1.9}$$

is well defined, since for any $\mathbf{e} \in \mathbb{S}^{d-1}$, $|\mathbf{e}^\top\theta_n\theta_n^\top\mathbf{e}| = (\mathbf{e}\cdot\theta_n)^2 \leq \|\theta_n^2\|$; for each $\mathbf{x}$, $M(\mathbf{x})$ is a symmetric, non-negative-definite matrix. We refer to $M(\mathbf{x})$ defined at (1.9) as the *increment covariance matrix*; this is a slight abuse of