Markers for error-corrupted observations

Andrew Hart[a], Heinrich Matzinger[b,*]

[a] *Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, UMR 2071 CNRS-UCHILE, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 170-3, Correo 3, Santiago, Chile*
[b] *Fakultät für Mathematik, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany*

Abstract

A *scenery* is a coloring ξ of the integers. Let $(S(t))_{t\geq 0}$ be a recurrent random walk on the integers. Observing the scenery ξ along the path of this random walk, one sees the color $\xi(S(t))$ at time t. The *scenery reconstruction problem* is concerned with trying to retrieve the scenery ξ, given only the sequence of observations $\chi := (\xi(S(t)))_{t\geq 0}$. Russel Lyons and Yuval Peres have both posed the question of whether two-color sceneries can be reconstructed when the observations are corrupted by random errors. The random errors happening at different times are independent conditional on χ. It has been proved that it is possible to do reconstruction in the case where the observations are contaminated with errors and the scenery has several colors, provided the error probability is small enough. However, the reconstruction problem is more difficult with fewer colors. Although the scenery reconstruction problem for two-color sceneries from error-free observations has been solved, the reconstruction of two-color sceneries from error-corrupted observations remains an open problem. In this paper, we solve one of the two remaining problems needed in order to reconstruct two-color sceneries when the observations are corrupted with random errors. We prove that given only the corrupted observations, we are able to determine a large amount of times, when the random walk is back at the same place (*marker*) in the scenery.

Keywords: Scenery reconstruction; Scenery distinguishing; Large deviations

* Corresponding author. Tel.: +49 5 21 1 06 47.
E-mail address: matzing@mathematik.uni-bielefeld.de (H. Matzinger).

1. Introduction

We call ξ a scenery if it is a coloring of the integers $\xi : \mathbb{Z} \mapsto \{1, 2, \dots, C\}$ where C is the number of colors. Let $\{S_t\}_{t\in\mathbb{N}}$ be a recurrent random walk on the integers. We call $\chi_t := \xi(S_t)$ the observation made of the scenery by the random walk at time $t \in \mathbb{N}$. A realization of the process $\chi := \{\chi_t\}_{t\in\mathbb{N}}$ is called the "observation".

The **scenery reconstruction problem** can be formulated as follows: If we do not know the scenery ξ but are only given one path realization of χ, can we almost surely recover ξ? In other words, does one path realization of the process $\{\chi_t\}_{t\in\mathbb{N}}$ determine ξ a.s.? We should point out that it is only possible to reconstruct sceneries up to shift and reflection in general. Thus the scenery reconstruction problem is the problem of trying to reconstruct ξ up to shift and reflection given only one realization of χ. A result of Lindenstrauss [15] implies that there exist an uncountable number of sceneries which cannot be reconstructed. Fortunately, these unreconstructable sceneries are in a certain sense "untypical". So, in general we take the scenery to be generated by a random process which is independent of the random walk and then show that almost every scenery can be reconstructed a.s. up to shift and reflection from a single realization of χ.

Let $\{\nu_t\}_{t\in\mathbb{N}}$ be an i.i.d. sequence of Bernoulli random variables which is independent of ξ and S. The variables ν_t are used to indicate at which times there are errors in the observations. More precisely, when $\nu_t = 1$ then there is an error in the observation at time t. Let $\tilde{\chi}$ denote the observations χ corrupted by the errors $\{\nu_t\}_{t\in\mathbb{N}}$. Thus,

$$\tilde{\chi}_t = \chi_t$$

when $\nu_t = 0$ and $\tilde{\chi}_t \neq \chi_t$ otherwise. (The exact value of $\tilde{\chi}_t$ when $\tilde{\chi}_t \neq \chi_t$ is not very important.)

The scenery reconstruction problem with errors can now be formulated as follows: Try to reconstruct ξ a.s. up to shift and reflection when you are only given one realization of $\tilde{\chi}$. For the case where the scenery has many colors and the error probability is small, the problem was solved by Rolles and Matzinger in [26]. In this article, we show how we can construct markers and stopping times telling us when the random walk is back at the markers, despite the errors. The method we use is different from that used to deal with the non-error-corrupted case. This solves one of the two remaining problems for scenery reconstruction with two colors and errors in the observations.

Let us at this stage explain what the remaining open problem is for achieving scenery reconstruction with error-corrupted observations and two-color sceneries. For this consider at first a simplified example of reconstruction without errors in the case of a simple random walk. Assume the scenery ξ is binary except at a point x and a point y where we observe the colors 2 and 3. Hence $\xi(x) = 2$ and $\xi(y) = 3$. Since the colors 2 and 3 appear only at one point in the scenery, we know every time the random walk visits x or y. We can reconstruct the portion of the scenery between x and y. For this, note that since the random walk is simple, the shortest way to go from $|x|$ to $|y|$ is in $|x - y|$ steps. When the random walk goes in shortest way from x to y it makes steps only in one direction. Hence, during that time we see in the observations χ a copy of the piece of ξ lying between x and y. In [18], Lember and Matzinger use this idea to prove that it is possible to reconstruct a two-color scenery. But in a two-color scenery the two extra colors 2 and 3 are not available. So instead one uses markers to find out when one is back at the same point. Actually the markers are not yet powerful enough. So, one collects information about the surrounding of each marker. The information about the surroundings of several nearby markers jointly gives us a powerful way of determining when the random walk is back in the same place.

There is one problem left: for the method of Lember and Matzinger to work, one needs to be able to determine that the random walk is to the right of or at x and to the left of or at y, at the times when it is in the surrounding of x or y. (We assume that $x < y$.) For this problem, the Lember–Matzinger method fails in the presence of errors.

Scenery reconstruction is closely related to the scenery distinguishing problem. We give a brief account. Let ξ^a and ξ^b be two non-equivalent sceneries which are known to us. Assume that the scenery ξ is equal either to ξ^a or to ξ^b but we do not know which. If we are only given one realization of the observation process χ of the scenery ξ by the random walk S, can we almost surely determine whether ξ is equal to ξ^a or whether it is equal to ξ^b? If so, we say the sceneries ξ^a and ξ^b are distinguishable. Kesten and Benjamini [1] showed that almost every pair of sceneries is distinguishable, even in the two-dimensional case and with only two colors. To do this, they took ξ^a to be any non-random scenery and ξ^b to be an i.i.d. scenery with two colors. Earlier, Howard [9] showed that any pair of periodic, non-equivalent sceneries are distinguishable, as are periodic sceneries with a single defect [8].

The problem of distinguishing two sceneries which differ at only one point is called detecting a single defect in a scenery. Kesten [12] was able to show that one can a.s. detect single defects in the case of four-color sceneries. A question Kesten raised concerning the detection of defects in sceneries led Matzinger [24,25,23] to investigate the scenery reconstruction problem.

As with scenery reconstruction, there is a version of the scenery distinguishing problem with observations that are corrupted. Once again, the scenery ξ is equal either to ξ^a or to ξ^b, both of which are known to us. However, the observations are now corrupted through an error process $\{\nu_t\}_{t\geq 0}$, which is assumed to be a sequence of i.i.d. Bernoulli random variables with parameter strictly smaller than $1/2$ and independent of ξ and S. The variables ν_t are used to indicate at which times there are errors in the observations. More precisely, when $\nu_t = 1$ then there is an error in the observation at time t. Let $\tilde{\chi}$ denote the observations χ corrupted by the errors $\{\nu_t\}_{t\in\mathbb{N}}$. Thus,

$$\tilde{\chi}_t = \chi_t$$

when $\nu_t = 0$ and $\tilde{\chi}_t \neq \chi_t$ otherwise. Knowing ξ^a and ξ^b, can we decide a.s. if $\xi = \xi^a$ or if $\xi = \xi^b$ based on one path realization of the process $\tilde{\chi}$? This constitutes the scenery distinguishing problem in the case of error-corrupted observations. The subject of this article is closely related to a random coin tossing problem which was first investigated by Harris and Keane in [6] and later by Levin, Pemantle and Peres in [22]. They take the error probability to be equal to $1/2$. The **coin tossing problem of Harris and Keane** can be described as follows:

Let $X_1, X_2, \ldots$ denote a sequence of Bernoulli variables where X_k is the result of the kth coin toss. We consider two ways of doing this.

- The first method is to toss an unbiased coin independently each time. In this case the variables X_k are a sequence of i.i.d. Bernoulli random variables with parameter $1/2$.
- Let $\tau_1, \tau_2, \ldots$ denote a sequence of return times of a random walk to the origin. We toss fair coins independently at all times except at the times τ_k, at which we toss a biased coin with fixed bias ω instead.

The problem investigated by Harris and Keane in [6] and later by Levin, Pemantle and Peres in [22] can now be described as follows: If we are only given one realization of the process $\{X_k\}_{k\geq 0}$, but do not know whether it was generated by mechanism 1 or mechanism 2, can we determine a.s. from which of the two processes the observation comes? Harris and Keane were able to show that, depending on the finiteness of the moments of the stopping times, we may or may not be able to deduce the method used to generate the observed sequence. Later, Levin, Pemantle and Peres were able to show that there is a phase transition depending on the size of the bias. Furthermore, they were also able to solve the problem in the case where the stopping times halt a random walk at a finite number of points instead of just at the origin.

It is evident that the Harris–Keane coin tossing problem can be viewed as a scenery distinguishing problem with errors. In particular, take ξ^a as the scenery which is everywhere equal to zero, and ξ^b as the scenery which is zero everywhere except at the origin. In the case studied by Levin, Pemantle and Peres [22], set the scenery $\xi^a \equiv 0$ and ξ^b to be zero everywhere except at a finite number of points. They take the error probability to be $1/2$, except when a "one" is observed. Hence, in their case, $P(\tilde{\chi}_t = 0|\chi_t = 0) = 1/2$, but $P(\tilde{\chi}_t = 0|\chi_t = 1) \neq 1/2$.

There is an excellent overview of scenery reconstruction and scenery distinguishing by Kesten [13]. Scenery distinguishing and reconstruction belongs to the general area of probability theory which deals with the ergodic properties of observations made by a random process in a random media. An important related problem is the T, T^{-1} problem studied by Kalikow [10]. Several important contributions about the properties of the observations were made later. These include those of Keane and den Hollander [11], den Hollander [2], den Hollander and Steif [3], Heicklen et al. [5], and Levin and Peres [21]. Interest in the scenery distinguishing problem was sparked when Keane and den Hollander, as well as Benjamini, asked whether all non-equivalent sceneries could be distinguished. Lindentstrauss was able to prove that there exist pairs of sceneries which cannot be distinguished [15]. After Matzinger showed the validity of scenery reconstruction in the simple case of error-free observations made by a one-dimensional random walk without jumps (see [25,24]), Kesten noticed that Matzinger's method was inadequate for solving the reconstruction problem in the two-dimensional case, as well as in the case when the random walk is allowed to jump. Subsequently, Löwe and Matzinger [17] were able to prove that scenery reconstruction is also possible on two-dimensional sceneries with many colors. Later, Matzinger, Merkl and Löwe [19] proved that with enough colors in one dimension one can do reconstruction even if the random walk is allowed to jump and thus is not a simple random walk. In general, scenery reconstruction becomes more difficult as the number of colors decreases (except in the trivial case when there is only one color). The most difficult case of reconstruction from observations made by a random walk with jumps on two-color sceneries was solved by Lember and Matzinger [18]. Den Hollander asked whether it would be possible to do reconstruction if the jumps made by the random walk are not bounded. Lenstra and Matzinger [16] were able to answer this question. Finally, following a question of den Hollander, Löwe and Matzinger [20] investigated the possibility of reconstructing sceneries that are not i.i.d. but have some correlation. The possibility of reconstructing finite pieces of sceneries in polynomial time following a question of Benjamini was investigated by Rolles and Matzinger [27].

In this article, we study one of the crucial techniques for finding markers used in scenery reconstruction and show that one can still construct and use markers when the observations are error-corrupted.

The paper is organized as follows. In Section 2, we consider a simplified example without errors. We show how in this simplified case, markers can be constructed and used for scenery reconstruction. Since many scenery reconstruction methods are very complicated, it seems worthwhile to present this simple case. In addition, it also serves as motivation, demonstrating the usefulness of markers. The following sections are concerned with how to define markers in the context of error-corrupted observations and construct stopping times that tell us when the random walk has returned to such a marker. In Section 3, we consider the likelihood of a marker being present in the scenery, given that some tail-event has occurred in the error-infested observation process. In Section 4, we show how to construct a multitude of stopping times which tell us when the random walk has returned to the location of a marker. It is assumed that there is a marker close to the random walk's starting state. Finally, in Section 5, we show how to find a marker for the first time and then construct a series of stopping times which tell us when the random walk returns to that marker.

2. An example of scenery reconstruction using a single marker

In this section, we shall illustrate the use of markers in scenery reconstruction. Let us make some special assumptions which will only apply within this section:

- The scenery $\xi : \mathbb{Z} \to \{0, 1, 2\}$ is a three-color scenery, with colors from the set $\{0, 1, 2\}$.
- The origin is colored with color 2: $\xi(0) = 2$.
- $\xi_1, \xi_{-1}, \xi_2, \xi_{-2}, \xi_3, \xi_{-3}, \dots$ is a sequence of i.i.d. Bernoulli variables with parameter $1/2$. This means that, excepting the origin, the scenery ξ is a two-color scenery.
- The random walk $\{S_t\}_{t\in\mathbb{N}}$ is a simple random walk starting at the origin.

The only place where there is a 2 in the scenery is the origin. We can use this "2" as a "marker": Every time we see a 2 in the observations, we are at the origin. This implies:

$$\chi_t = 2 \Longrightarrow S_t = 0.$$

Let τ_k be the time of the kth visit of S to the origin. Note that τ_k is observable since it is also the kth time we observe a 2 in χ:

$$\tau_k := \min\{t \in \mathbb{N} \mid \chi_t = 2, t > \tau_{k-1}\}, \quad k \geq 1.$$

By convention, we set $\tau_0 := 0$. Consider the following sequence of binary words:

$$w_1 = 001100, \quad w_2 = 0011001100, \quad w_3 = 00110011001100, \dots.$$

Since the scenery ξ is i.i.d., every finite pattern will occur in ξ infinitely often. Hence all the strings w_k will occur in ξ infinitely often. Let x_k denote the closest place to the origin where w_k occurs in the scenery. (If there are two such places, choose the one to the right of the origin.) Hence x_k is a point z minimizing $|z|$ under the following constraints:

1. If $z > 0$, then

$$\xi_z\xi_{z+1}\dots\xi_{z+4k+1} = w_k.$$

2. If $z < 0$, then

$$\xi_z\xi_{z-1}\dots\xi_{z-4k-1} = w_k.$$

It is easy to see that the only way the string w_k can appear in the observations χ is by walking in a straight line over a portion of the scenery where w_k appears. In other words, we observe the word w_k at time t, that is,

$$\chi_t\chi_{(t+1)}\cdots\chi_{(t+4k+1)} = w_k,$$

if and only if, for all $i = 0, \dots, i + 4k + 1$, we have

$$S_{(t+i)} = S_t + iu$$

and

$$\xi_{S_t}\xi_{S_t+u}\ldots\xi_{S_t+4ku+u}=w_k,$$

where $u=\pm 1$.

Let us give a numerical example. Assume that at time $t\in\mathbb{N}$, we observe:

$$\chi_t\chi_{t+1}\ldots\chi_{t+5}=001100.$$

Then we have either $S_{t+5}=S_t+5$ or $S_{t+5}=S_t-5$. In the first case, we have that

$$\xi_{S_t}\xi_{(S_t+1)}\xi_{(S_t+2)}\ldots\xi_{(S_t+5)}=001100$$

whilst in the second case, we would have:

$$\xi_{S_t}\xi_{(S_t-1)}\xi_{(S_t-2)}\ldots\xi_{(S_t-5)}=001100.$$

Almost surely, we have that

$$\lim_{k\to\infty}|x_k|=\infty$$

and on both sides of the origin there are infinitely many points from the sequence x_k, $k\in\mathbb{N}$.

The shortest time after a 2 at which we can observe the word w_k is x_k. It takes the random walk $|x_k|$ steps to go from the origin to x_k in minimal time. When doing so, the random walk must walk in a straight line only taking steps towards x_k. Whenever the random walk travels in a straight line, it produces a copy of the portion of the scenery which it has traversed. This copy is manifest and plain to see in the observations. The random walk goes from the origin to x_k infinitely often in the shortest possible time. This implies that when we observe 2 at time t followed by w_k at time $t+|x_k|$, then we have a copy of $\xi_0\xi_u\xi_{2u}\ldots\xi_{x_k}$ in the observations χ. Here, we take $u=x_k/|x_k|$. Hence we can reconstruct

$$\xi_0\xi_u\xi_{2u}\ldots\xi_{x_k} \tag{1}$$

using the following algorithm:

Algorithm 2.1. 1. Let μ_s denote the first time we observe the word w_k after time τ_s:

$$\mu_s:=\min\{t>\tau_s\mid w_k=\chi_t\chi_{t+1}\ldots\chi_{t+4k+1}\}.$$

2. Let d_k denote the minimum time at which we can observe the finite string w_k after a 2:

$$d_k:=\min\{\mu_s-\tau_s\mid s\in\mathbb{N}\}.$$

3. Let s^* denote any s minimizing $\mu_s-\tau_s$. In other words, s^* is such that

$$\mu_{s^*}-\tau_{s^*}=d_k.$$

4. The output of our algorithm is

$$\chi_{\tau_{s^*}}\chi_{\tau_{s^*}+1}\cdots\chi_{\mu_{s^*}}. \tag{2}$$

For the reasons explained above, the output of the above algorithm is equal to the piece of the scenery located between the origin and x_k inclusive with probability one. This should demonstrate the usefulness of markers in scenery reconstruction.

3. Existence of a marker

In this section, we take the scenery $\xi : \mathbb{Z} \to \{0, 1\}$ to be a two-color i.i.d. scenery. Thus, it is a realization of the process $\{\xi_z\}_{z\in\mathbb{Z}}$ where the ξ_z's are i.i.d. Bernoulli random variables with parameter $1/2$.

As before, the observation of the scenery ξ by the random walk S at time t is denoted by $\chi_t := \xi(S_t)$. We assume that the errors are i.i.d. with the probability $P(\nu_t = 1) = \epsilon$ of an error at time t being strictly smaller than $1/2$. Then, the error-corrupted observation $\tilde{\chi}_t$ at time t is given by

$$\tilde{\chi}_t := (\chi_t + \nu_t) \bmod 2.$$

We write $\chi = (\chi_0, \chi_1, \ldots)$ for the error-free observations and $\tilde{\chi} := (\tilde{\chi}_0, \tilde{\chi}_1, \ldots)$ for the error-corrupted observations. The scenery ξ, the random walk S and the error process $\{\nu_t\}_{t\in\mathbb{N}}$ are all assumed to be independent of each other.

For completeness, the following list details all the assumptions we make in this section.

- Let $S = \{S_t\}_{t\geq 0}$ be a recurrent random walk on $\mathbb{Z}$ starting at the origin which can visit any point $z \in \mathbb{Z}$ with positive probability.
- The distribution of the increments of the random walk S has bounded support.
- The process $\xi = \{\xi_z\}_{z\in\mathbb{Z}}$ is such that the ξ_k's are i.i.d. Bernoulli variables with parameter $1/2$.
- The errors ν_t, for $t \geq 0$, are i.i.d. Bernoulli variables with parameter $\epsilon = P(\nu_t = 1)$ strictly smaller than $1/2$. Here ϵ denotes the probability of an error.
- The three processes ξ, S and $\nu = \{\nu_t\}_{t\geq 0}$ are independent of each other.

Next we need to define a few events. Firstly, define

$$A^n := \left\{ \sum_{t=0}^{n^2} \tilde{\chi}_t \leq \epsilon n^2 \right\}.$$

Let B^n be the event that there exists a contiguous block of zeros in ξ of length greater than $n^{0.1}$ in the interval $[-Ln^2, Ln^2]$. More precisely,

$$B^n := \left\{ \begin{array}{l} \exists z \in [-Ln^2, Ln^2 - n^{0.1}] \text{ such that} \\ \xi_z = \xi_{z+1} = \ldots = \xi_{(z+n^{0.1})} = 0 \end{array} \right\}.$$

Remark. The reader might object that $n^{0.1}$ is not necessarily an integer. That reader is right. But for scenery reconstruction to work we do not need to show that for every n things work. It is enough to show that for an increasing sequence of n's (which does not go to infinity faster than polynomially) things work. Hence the reader can imagine that n in the first place is the tenth power of an integer so that $n^{0.1}$ becomes an integer. Throughout this article, whenever a fractional power of n is encountered, the reader should view that fractional power as an integer. One could also take the integer part and every n, but then notationally things could become unpleasant.

Let C^n be the event that the error-free observation process reveals more than $n^{1.7}$ 1's in the first n^2 observations:

$$C^n := \left\{ \sum_{t=0}^{n^2} \chi_t \geq n^{1.7} \right\}.$$

We shall denote the complement of an event E by E^c. Next, let us present the main result of this section.

Theorem 3.1. *For n large enough,*

$$P(B^{nc} \mid A^n) \leq \exp\left(-(1-2\epsilon)^2 n^{1.4}/3\right).$$

Hence, $P(B^n \mid A^n) \to 1$ as $n \to \infty$.

Theorem 3.1 says that if in the first n^2 error-corrupted observations we observe a significantly low number of 1's, then with very high probability there is a contiguous block of zeros of length $n^{0.1}$ very close to the origin in the scenery ξ. This unbroken block of zeros will be used in the next section as a marker to tell us when the random walk is back near the origin.

To prove Theorem 3.1, we will need a number of lemmas. The first two of these will be used numerous times throughout this and the following section. Let us start with a large deviation result.

Lemma 3.1. *Let $\Delta > 0$. Let $X_1, X_2, \ldots$ be a sequence of zero-mean random variables such that $\{S_t\}_{t\geq 0}$ is a martingale, where $S_0 = 0$ and $S_t = \sum_{i=1}^t X_i$ for $t \geq 1$. Assume furthermore that the random variables all have bounded range, that is, for some $a > 0$, $|X_i| \leq a$ for all $i = 1, 2, \ldots$. Then, for all $k \geq 1$,*

$$P\left(\frac{\sum_{i=1}^{k} X_i}{k} \geq \Delta\right) \leq \exp\left(-\frac{k\Delta^2}{2a^2}\right). \tag{3}$$

Proof. From Chernov's inequality, we have

$$\begin{aligned}
P\left(\frac{\sum_{i=1}^{k} X_i}{k} \geq \Delta\right) &= \mathbb{E}[\mathbf{1}_{S_k \geq k\Delta}] \\
&\leq \mathbb{E}[\exp(\beta S_k - \beta k\Delta)], \quad \text{for all } \beta > 0, \\
&= \exp(-\beta k\Delta)\mathbb{E}[\exp(\beta S_k)].
\end{aligned}$$

Then, an application of the Azuma–Hoeffding lemma (see [7], p. 252) yields

$$P\left(\frac{\sum_{i=1}^{k} X_i}{k} \geq \Delta\right) \leq \exp(-\beta k\Delta)\exp(k\beta^2 a^2/2) = \exp(-\beta k\Delta + k\beta^2 a^2/2).$$

The right-hand side of this final expression obtains its optimal (minimum) value at $\beta = \Delta/a^2$. Substituting $\beta = \Delta/a^2$ into the equation yields the desired result. $\square$

Lemma 3.2. *Let $S := \{S_t\}_{t\geq 0}$ be a random walk with bounded jumps which starts at the origin. Then:*

1. *There exists a constant* $C' > 0$ *such that*

$$P\left(\max_{0 \le t \le n^2} |S_t| \le n\right) \ge C'$$

for all $n \ge 0$.

2. *As* $n \to \infty$,

$$P\left(\max_{0 \le t \le n^{2-\gamma}} |S_t| \le n\right) \to 1,$$

for any $0 < \gamma < 2$.

Proof. 1. Define $Z_k := \{Z_k(s)\}_{s \in \mathbb{N}}$, where $Z_k(s) := \frac{1}{k} S_{sk^2}$ and let $W := \{W_t\}_{t \ge 0}$ denote a Brownian motion. Then, by the invariance principle, $Z_n \xrightarrow{\mathcal{D}} W$ as $n \to \infty$. In particular, $Z_n(s) \xrightarrow{\mathcal{D}} W_s$ and so

$$\max_{0 \le t \le n^2} \left| \frac{S_t}{n} \right| = \max_{s=0,1/n^2,2/n^2,\ldots,1} |Z_n(s)| \xrightarrow{\mathcal{D}} \max_{0 \le s \le 1} |W_s|$$

as $n \to \infty$. Thus,

$$P\left(\max_{0 \le t \le n^2} |S_t| \le n\right) \to P\left(\max_{0 \le s \le 1} |W_s| \le 1\right) = P(\psi_{[-1,1]} > 1) > 0,$$

where $\psi_{[-1,1]}$ is the first time of exit of the Brownian motion W from the interval $[-1, 1]$. The positivity of $P(\psi_{[-1,1]} > 1)$ may be deduced from the analytic expression

$$P(\psi_{[-1,1]} \in dt) = \frac{2}{\sqrt{2\pi t^3}} \sum_{n=-\infty}^{\infty} (4n+1) \mathrm{e}^{-\frac{(4n+1)^2}{2t}} \, \mathrm{d}t,$$

which is a special case of an expression derived in [14].

Thus, since $P(\max_{0 \le t \le n^2} |S_t| \le n) > 0$ for all n, it follows that there exists $C' > 0$ such that $P(\max_{0 \le t \le n^2} |S_t| \le n) \ge C'$ for all $n \ge 0$.

2. As we are assuming that S has i.i.d. increments, let σ^2 denote the variance of an increment. Then, applying the Kolmogorov inequality (for example, see Chapter 14.6 of [28]), we have

$$\begin{aligned} P\left(\max_{0 \le t \le n^{2-\gamma}} |S_t| \le n\right) &= 1 - P\left(\max_{0 \le t \le n^{2-\gamma}} |S_t| \ge n+1\right) \\ &\ge 1 - \frac{n^{2-\gamma}\sigma^2}{(n+1)^2} \ge 1 - \sigma^2 n^{-\gamma} \to 1, \end{aligned}$$

as $n \to \infty$. □

Lemma 3.3. *There exists a constant* $c > 0$ *not depending on* n *such that, for all* $n \ge 0$,

$$P(A^n) \ge c \left(\frac{1}{4}\right)^n . \tag{4}$$

Proof. Let D^n and E^n be events defined as follows:

$$D^n := \{\forall z \in [-n, n],\ \xi_z = 0\} \text{ and}$$
$$E^n := \{\forall t \in [0, n^2],\ S_t \in [-n, n]\}.$$

By Part 1 of Lemma 3.2, we know that there exists a constant $C' > 0$, not depending on n, such that $P(E^n) \geq C'$. Furthermore, $P(D^n) = (1/2)^{2n+1}$. Since, conditional on $D^n \cap E^n$, $\sum_{t=0}^{n^2} \tilde{\chi}_t = \sum_{t=0}^{n^2} \nu_t \sim \mathrm{Bin}(n^2+1, \epsilon)$, we see that $P(A^n \mid D^n \cap E^n) > 0$ for all $n \geq 0$. (Here $\mathrm{Bin}(n^2+1, \epsilon)$ denotes a binomial distribution with parameter n^2 and ϵ.) Furthermore, by the central limit theorem, $P(A^n \mid D^n \cap E^n) \longrightarrow \frac{1}{2}$ as $n \to \infty$. Thus, there exists $c'' > 0$ such that $P(A^n \mid D^n \cap E^n) \geq c''$ for all n. Consequently,

$$P(A^n) \geq c'' P(D^n \cap E^n) = c'' P(D^n) P(E^n) \geq c \left(\frac{1}{2}\right)^{2n},$$

where $c = C'c''/2$. $\square$

Lemma 3.4.

$$P(A^n \mid C^n) \leq \exp\left(-\frac{(1-2\epsilon)^2 n^{1.4}}{2}\right) \tag{5}$$

for all $n \geq 1$.

Proof. Let Z and $\tilde{Z}$ denote the sums

$$Z := \sum_{t=0}^{n^2} \chi_t$$

and

$$\tilde{Z} := \sum_{t=0}^{n^2} \tilde{\chi}_t.$$

Conditional on $\chi_t = 1$, $\tilde{\chi}_t$ has expectation $1-\epsilon$ whilst conditional on $\chi_t = 0$, $\tilde{\chi}_t$ has expectation ϵ. Thus, $\tilde{Z}$ conditional on Z has the same distribution as the sum of n^2+1 independent Bernoulli variables where Z of them have expectation $1-\epsilon$ and the other n^2+1-Z have expectation ϵ. It follows that the conditional expectation of $\tilde{Z}$ given Z is $\mathbb{E}[\tilde{Z} \mid Z] = (n^2+1)\epsilon + (1-2\epsilon)Z$. Now,

$$P(A^n \mid Z) = P(\tilde{Z} \leq \epsilon n^2 \mid Z) = P\left(\frac{\tilde{Z} - (\epsilon n^2 + (1-2\epsilon)Z)}{n^2} \leq -\frac{(1-2\epsilon)Z}{n^2} \,\middle|\, Z\right). \tag{6}$$

Since, conditional on Z, $\tilde{Z}$ is distributed like a sum of n^2+1 independent Bernoulli variables, it follows that we can apply Lemma 3.1. Taking $k = n^2$, $a = 1$ and $\Delta = (1-2\epsilon)Z/n^2$, we find that the expression on the right-hand side of (6) is bounded by

$$\exp\left(-\frac{(1-2\epsilon)^2 Z^2}{2n^2}\right).$$

Hence, when $Z \geq n^{1.7}$ is assumed given, we obtain

$$P(A^n \mid C^n) = P(A^n \mid Z \geq n^{1.7}) \leq \exp\left(-\frac{(1-2\epsilon)^2 n^{1.4}}{2}\right)$$

and the proof is complete. $\square$

Next, we define $q_{x,y}^n$ to be the probability that the random walk S visits the point x or y before time $n^{0.21}$:

$$q_{x,y}^n := P\left(\exists t \leq n^{0.21}, S_t \in \{x, y\}\right).$$

Let q^n denote the minimum

$$q^n := \min_{(x,y)\in G_n} q_{x,y}^n,$$

where $G_n := \left\{(x, y) \in [-n^{0.1}, n^{0.1}]^2 | x < 0 < y\right\}$. The following lemma will be needed to prove Lemma 3.6.

Lemma 3.5. $\lim_{n\to\infty} q^n = 1$.

Proof. Let n be large and choose two points $x, y \in [-n^{0.1}, n^{0.1}]$ such that $x < 0 < y$. Also, let I_x and I_y denote the intervals $I_x := [x - L, x + L]$ and $I_y := [y - L, y + L]$ respectively. Then, we define τ_{xy} to be the time of the first visit by S to $I_x \cup I_y$ and use E_{xy}^n to denote the event that S visits x or y before time $n^{0.21}$:

$$E_{xy}^n := \left\{\exists t \leq n^{0.21}, S_t \in \{x, y\}\right\}.$$

Further, let $E_{a,xy}^n$ denote the event that, within time $n^{0.2}$ of the stopping time τ_{xy}, the random walk visits all the points in a neighborhood of radius L of the point $S_{\tau_{xy}}$. Hence, $E_{a,xy}^n$ denotes the event that for all z satisfying $\left|z - S_{\tau_{xy}}\right| \leq L$, there exists $t \in [\tau_{xy}, \tau_{xy} + n^{0.2}]$ such that $S_t = z$.

Lastly, define E_b^n to be the event that the random walk S is outside the interval $[-n^{0.1}, n^{0.1}]$ at time $t = n^{0.205}$:

$$E_b^n := \left\{S_{n^{0.205}} \notin [-n^{0.1}, n^{0.1}]\right\}.$$

Since the random walk S starts at the origin, it must cross (but not necessarily hit) either x or y before leaving the interval $[-n^{0.1}, n^{0.1}]$. Since the step lengths of S are bounded by L, the random walk must visit either I_x or I_y in order to exit the interval $[-n^{0.1}, n^{0.1}]$. Hence, when E_b^n holds, we have

$$\tau_{xy} \leq n^{0.205}. \tag{7}$$

Now, whenever (7) and $E_{a,xy}^n$ hold, the set $\{x, y\}$ will be visited before time $n^{0.205} + n^{0.2}$. For n large enough, $n^{0.205} + n^{0.2} < n^{0.21}$. Hence,

$$E_{a,xy}^n \cap E_b^n \subset E_{xy}^n,$$

for any $(x, y) \in G_n$. This implies that

$$P(E_{xy}^{nc}) \leq P(E_{a,xy}^{nc}) + P(E_b^{nc}).$$

Next, let E_a^n denote the event that the random walk visits all the points in $[-L, L]$ before time $n^{0.2}$. By the strong Markov property of S, we see that $P(E_{a,xy}^{nc}) = P(E_a^{nc})$ and hence we obtain

$$P(E_{xy}^{nc}) \leq P(E_a^{nc}) + P(E_b^{nc}). \tag{8}$$

Note that the bound on the right side does not depend on either x or y and that (8) holds for all $(x, y) \in G_n$. Therefore,

$$q^n = \min_{x,y} P(E_{xy}^n) \geq 1 - P(E_a^{nc}) - P(E_b^{nc}). \tag{9}$$

Now, by the central limit theorem, we have

$$\lim_{n\to\infty} P(E_b^{nc}) = 0. \tag{10}$$

Also, by the assumption that S is recurrent and hence has positive probability of visiting all points in $\mathbb{Z}$, we find that $P(E_a^\infty) = 1$. (Here the event E_a^∞ is defined to be equal to $\cup_{n>1} E_a^n$.) Then, by continuity of probability,

$$\lim_{n\to\infty} P(E_a^n) = P(E_a^\infty) = 1. \tag{11}$$

Then, by applying (10) and (11) to (9), we conclude that

$$\lim_{n\to\infty} q^n = 1. \quad \square$$

Lemma 3.6. *For sufficiently large n,*

$$P(C^{nc} \mid B^{nc}) \leq \exp(-n^{1.79}/8). \tag{12}$$

Proof. We begin by defining Bernoulli variables $\{Y_k\}_{k\geq 1}$ in the following way:

$$Y_k = \mathbf{1}_{\sum_{t=(k-1)n^{0.21}}^{kn^{0.21}} \chi_t \geq 1} = \mathbf{1}_{\exists t \in](k-1)n^{0.21}, kn^{0.21}] \quad \text{such that } \chi_t = 1}.$$

Clearly, $Y_k \leq \sum_{t=(k-1)n^{0.21}-1}^{kn^{0.21}} \chi_t$ and

$$\sum_{k=1}^{n^{1.79}} Y_k \leq \sum_{t=0}^{n^2} \chi_t.$$

Thus,

$$C^{nc} = \left\{ \sum_{t=0}^{n^2} \chi_t < n^{1.7} \right\} \subseteq \left\{ \sum_{k=1}^{n^{1.79}} Y_k < n^{1.7} \right\}$$

and

$$P(C^{nc} \mid B^{nc}) \leq P\left(\sum_{k=1}^{n^{1.79}} Y_k < n^{1.7} \mid B^{nc} \right). \tag{13}$$

Let $\mathcal{F} := \bigcup_{k=1}^\infty \mathcal{F}_k$ be the σ-algebra defined by the filtration $\{\mathcal{F}_k\}_{k\geq 1}$, where

$$\mathcal{F}_k := \sigma(S_t, \xi_z \mid t \leq kn^{0.21}, z \in \mathbb{Z}).$$

The sequence $\{Y_k\}_{k\in\mathbb{N}}$ is $\mathcal{F}$-adapted. Furthermore, $M_k = \sum_{i=1}^k (Y_i - \mathbb{E}[Y_i \mid \mathcal{F}_{i-1}])$ is a martingale with respect to $\{\mathcal{F}_k\}_{k\geq 1}$.

Starting from the origin, the random walk S takes steps with lengths bounded by L. This implies that S stays in the set $[-Ln^{0.21}, Ln^{0.21}]$ during the time interval $[0, n^{0.21}]$. When the event B^{nc} holds, there exists, for every point $z \in [-Ln^{0.21}, Ln^{0.21}]$, two random points x^* and y^* such that $z-n^{0.1} < x^* < z < y^* < z+n^{0.1}$ with $\xi_{x^*} = \xi_{y^*} = 1$. By the strong Markov property, given that the random walk is at z at time t, the probability of visiting x^* or y^* during the time interval $(t, t+n^{0.21}]$ is equal to $q^n_{x^*-z,y^*-z}$. Hence this probability is larger than q^n. In this case the conditional probability that we observe at least one 1 in χ during the time interval $[t, t+n^{0.21}]$ is larger than or equal to q^n. (Conditional on B^{nc} and S_t, where $S_t \in [-Ln^{0.21}, Ln^{0.21}]$.) This means that, when the event B^{nc} holds, then

$$P(Y_k = 1 \mid \mathcal{F}_{k-1}) = \mathbb{E}[Y_k \mid \mathcal{F}_{k-1}] \geq q^n,$$

for all $1 \leq k \leq n^{1.79}$. Since

$$\lim_{n\to\infty} q^n = 1$$

by Lemma 3.5, we can assume that n is large enough so that $q^n > 3/4$. Thus,

$$\mathbb{E}[Y_k \mid \mathcal{F}_{k-1}] \geq \frac{3}{4} \tag{14}$$

for n large enough when B^{nc} holds and $k \leq n^{1.79}$. Because of (14) and since B^{nc} is $\mathcal{F}_0$-measurable, we find

$$P\left(\sum_{k=1}^{n^{1.79}} Y_k < n^{1.7} \mid B^{nc}\right) \leq P\left(\frac{\sum_{k=1}^{n^{1.79}}(Y_k - \mathbb{E}[Y_k \mid \mathcal{F}_{k-1}])}{n^{1.79}} < \frac{n^{1.7}}{n^{1.79}} - \frac{3}{4} \,\middle|\, B^{nc}\right) \tag{15}$$

for large n. Since $\{M_k\}_{k\geq 0}$ constitutes a martingale with respect to the filtration $\{\mathcal{F}_k\}_{k\geq 0}$ and since B^{nc} is $\mathcal{F}_0$-measurable, M_k remains a martingale when we condition on B^{nc}. Therefore, we can apply Lemma 3.1 to bound the probability on the right side of (15). For this we take $a = 1$ and $k = n^{1.79}$. For n large enough we have that $n^{-0.09} - \frac{3}{4} < -\frac{1}{2}$, which allows us to take the value $\frac{1}{2}$ for the Δ of Lemma 3.1. In this way we find that the right side of (15) is smaller than $\exp(-n^{1.79}/8)$. Combining inequalities (13) and (15) with this bound completes the proof. $\square$

Lemma 3.7. *For n sufficiently large,*

$$P(A^n \mid B^{nc}) \leq 3\exp\left(-\frac{(1-2\epsilon)^2 n^{1.4}}{2}\right). \tag{16}$$

Proof. We have

$$\begin{aligned} P(A^n \mid B^{nc}) &= P\left((A^n \cap C^n) \cup (A^n \cap C^{nc}) \mid B^{nc}\right) \\ &= P(A^n \cap C^n \mid B^{nc}) + P(A^n \cap C^{nc} \mid B^{nc}) \\ &\leq P(A^n \mid C^n) P(B^{nc})^{-1} + P(C^{nc} \mid B^{nc}). \end{aligned}$$

Note that for n large, $P(B^{nc})$ is close to one. Thus, let us assume that $P(B^{nc}) > 1/2$. With this assumption we obtain

$$P(A^n \mid B^{nc}) \leq 2P(A^n \mid C^n) + P(C^{nc} \mid B^{nc}).$$

We can now apply the bounds from inequalities (5) and (12) to the right-hand side of this last inequality. Note that the first of these two bounds is much larger than the second. We therefore find that $P(A^n \mid B^{nc})$ is smaller than 3 times the larger bound, provided that n is large enough. In other words,

$$P(A^n \mid B^{nc}) \leq 3 \exp\left(-\frac{(1-2\epsilon)^2 n^{1.4}}{2}\right)$$

for large n. This completes the proof. □

We can now prove Theorem 3.1.

Proof of Theorem 3.1. We have

$$P(B^{nc} \mid A^n) = P(A^n \mid B^{nc}) \cdot \frac{P(B^{nc})}{P(A^n)} \leq \frac{P(A^n \mid B^{nc})}{P(A^n)}. \tag{17}$$

Applying inequalities (4) and (16) to this expression, we obtain

$$P(B^{nc} \mid A^n) \leq \frac{3\exp\left(-(1-2\epsilon)^2 n^{1.4}/2\right)}{c(1/4)^n} = \exp\left(-(1-2\epsilon)^2 n^{1.4}/2 + n\ln 4 + \ln(3/c)\right)$$

for n sufficiently large. In the expression $-(1-2\epsilon)^2 n^{1.4}/2 + n\ln 4 + \ln(3/c)$, the dominating term is the first. This implies that for n large enough, $-(1-2\epsilon)^2 n^{1.4}/2 + n\ln 4 + \ln(3/c)$ is smaller than $-(1-2\epsilon)^2 n^{1.4}/3$. This in turn implies that

$$P(B^{nc} \mid A^n) \leq \exp\left(-(1-2\epsilon)^2 n^{1.4}/3\right)$$

for large n and this yields the desired result. □

We conclude this section with a lemma that will be useful in the next section.

Lemma 3.8. *For n large,*

$$P(A^n) \leq 2Ln^2 \left(\frac{1}{2}\right)^{n^{0.1}}.$$

Proof. Let B_z^n denote the event that there is a contiguous block of zeros in the scenery between z and $z + n^{0.1}$ inclusive:

$$B_z^n := \left\{\xi_z = \xi_{z+1} = \cdots = \xi_{z+n^{0.1}} = 0\right\}.$$

Since the scenery is generated by i.i.d. Bernoulli random variables, $P(B_z^n) = (1/2)^{n^{0.1}+1}$. Furthermore, with this definition, $B^n = \bigcup_z B_z^n$, where the union is taken over z in $[-Ln^2, Ln^2 - n^{0.1}]$. The length of this interval is smaller than $2Ln^2$. Thus we see that

$$P(B^n) \leq \sum_z P(B_z^n) \leq 2Ln^2 \left(\frac{1}{2}\right)^{n^{0.1}+1}. \tag{18}$$

Now,

$$P(A^n) = P(A^n \mid B^{nc})P(B^{nc}) + P(A^n \mid B^n)P(B^n) \leq P(A^n \mid B^{nc}) + P(B^n). \tag{19}$$

We can bound $P(A^n|B^{nc})$ using the inequality (16) and $P(B^n)$ with the aid of (18). The bound given on the right-hand side of (18) is asymptotically much larger than that given in (16). Thus, for large enough n, we can bound (19) by twice the larger of the two bounds and obtain

$$P(A^n) \le 2Ln^2 \left(\frac{1}{2}\right)^{n^{0.1}}. \quad \square$$

4. Returning to a marker

The main result of the last section states that, given that we observe a significantly low number of 1's in the first n^2 error-corrupted observations (the event A^n), there is a high probability that the scenery ξ has a contiguous block of $n^{0.1}$ or more zeros in the interval $[-Ln^2, Ln^2]$. In the context of sceneries observed with errors, we shall call such a block a marker.

In this section we shall prove that, by just looking at the observations $\tilde{\chi}$, we can tell $\exp(n^{0.001})$ times with high probability when the random walk is back at the marker. More precisely, we shall show that we can construct $\exp(n^{0.001})$ stopping times, which are observable, that is, $\sigma(\tilde{\chi})$ measurable, and will stop the random walk close to the marker in the interval $[-2Ln^2, 2Ln^2]$. Of course, we need to make the assumption that there is such a marker in the interval $[-Ln^2, Ln^2]$. In order to do this, we will assume that the probability distribution governing our whole world of scenery, random walk and errors has properties similar to the measure we obtain by taking the distribution used in the previous section conditional on the event B^n. To simplify notation, we do not use $P(\cdot \mid B^n)$, but a measure $P_2(\cdot)$ having very similar properties to $P(\cdot \mid B^n)$ instead. Throughout this section, $P_2(\cdot)$ denotes a measure which satisfies the following conditions:

- The random walk S and the scenery ξ are independent of each other. (See also the assumptions at the beginning of section 3.)
- The random walk S has the same distribution under $P_2(\cdot)$ as it had in the previous section. Moreover, it starts at the origin.
- The scenery outside the interval $[-Ln^2, Ln^2]$ is i.i.d. Bernoulli with parameter $1/2$.
- The portion of the scenery inside the interval $[-Ln^2, Ln^2]$ is independent of the remainder outside the interval.
- The scenery ξ P_2-almost surely contains a contiguous block of zeros longer than $n^{0.1}$ in $[-Ln^2, Ln^2]$. We require that

$$P_2\left(\exists z \in [-Ln^2, Ln^2 - n^{0.1}] \text{ such that } \xi_z = \xi_{z+1} = \cdots = \xi_{z+n^{0.1}} = 0\right) = 1.$$

- The errors under $P_2(\cdot)$ are distributed as before and are independent of the random walk and the scenery. In other words, the process $\{\nu_t\}_{t\ge0}$ is P_2-independent of $\{S_t\}_{t\ge0}$ and $\{\xi_z\}_{z\in\mathbb{Z}}$. Also, $P_2(\nu_t = 1) = \epsilon$. Once again, for all $t \in \mathbb{N}$,

$$\chi_t := \xi(S_t) \quad \text{and} \quad \tilde{\chi}_t := \chi_t + \nu_t \bmod 2.$$

Next, we define an increasing set of stopping times that are supposed to tell us when the random walk S is back close to the origin.

Definition 4.1. Let T denote the random integer set

$$T := \left\{ t \ge 0 : \sum_{s=t}^{t+n^{0.1}} \tilde{\chi}_s \le \epsilon n^{0.1} \right\}.$$

For $k > 0$, let τ_k denote the kth element (under the usual ordering on $\mathbb{N}$) of the set T.

We can now state the principal result of this section. It says that, with high P_2-probability, all of the first $\exp(n^{0.001})$ stopping times τ_k stop the random walk in the interval $[-2Ln^2, 2Ln^2]$ near a contiguous block of more than $n^{0.1}$ zeros. Furthermore, it also says that these stopping times all occur prior to time $\exp(n^{0.003})$ with high P_2-probability.

Theorem 4.1. *For large n,*

$$P_2\left(\forall k \le \exp(n^{0.001}),\, S_{\tau_k} \in [-2Ln^2, 2Ln^2] \text{ and } \tau_{\exp(n^{0.001})} \le \exp(n^{0.003})\right)$$
$$\ge 1 - 3\exp(-n^{0.003}/4).$$

Before continuing, we shall define a few useful intervals and a number of events that we shall need in the following.

$$I_1^n := [-Ln^2, Ln^2], \quad I_2^n := [-1.5Ln^2, 1.5Ln^2],$$
$$I_3^n := [-2Ln^2, 2Ln^2], \quad I_4^n := [-Ln^2, Ln^2 - n^{0.1}],$$
$$I_5^n := [-Ln^2 - n^{0.005}, Ln^2], \quad I_6^n := [-Ln^2 + n^{0.1}/2, Ln^2 - n^{0.1}/2],$$
$$I_7^n := [-L\exp(n^{0.003}), L\exp(n^{0.003})].$$

The first event $E^n_{\text{no-error}}$ says that we never see a significantly low average of 1's in the observations up to time $t = \exp(n^{0.003})$ when we are outside I_2^n.

$$E^n_{\text{no-error}} := \left\{ \sum_{s=t}^{t+n^{0.1}} \tilde{\chi}_s > \epsilon n^{0.1},\ \forall t \le \exp(n^{0.003}) \text{ such that } S_t \notin I_2^n \right\}.$$

We know that under $P_2(\cdot)$ there is a block of color zero having length $n^{0.1}$ in I_1^n with probability one. Let z_c denote the center of such a block. Thus, $z_c \in I_6^n$ P_2-almost surely and

$$P_2(\xi_z = 0,\ \forall z \in [z_c - n^{0.1}/2, z_c + n^{0.1}/2]) = 1.$$

Note that, by assumption, z_c is P_2-independent of $\{S_t\}_{t\ge 0}$ and $\{\nu_t\}_{t\ge 0}$. Let κ_l^* denote the lth visit by S to the point z_c. Let κ_k denote the $l = kn^{0.1}$th stopping time κ_l^*. More precisely,

$$\kappa_k := \kappa^*_{kn^{0.1}}, \quad k \in \mathbb{N}.$$

We define the stopping times κ_k in this way to ensure they are separated by time periods of length at least $n^{0.1}$.

Let E^n_{visits} denote the event that there are more than $\exp(n^{0.002})$ visits to z_c before time $\exp(n^{0.003})$:

$$E^n_{\text{visits}} := \left\{\kappa_{\exp(n^{0.002})} \le \exp(n^{0.003})\right\}.$$

Let Y_k denote the Bernoulli variable which is equal to one if and only if

$$\sum_{s=\kappa_k}^{\kappa_k+n^{0.1}} \tilde{\chi}_s \le \epsilon n^{0.1}.$$

Let $E^n_{\text{marker-works}}$ denote the event that we observe a significantly low number of ones more than $1/3$ of the time after a stopping time κ_k, $k \le \exp(n^{0.002})$:

$$E^n_{\text{marker-works}} := \left\{ \sum_{k=1}^{\exp(n^{0.002})} Y_k \ge \frac{\exp(n^{0.002})}{3} \right\}.$$

The final event we shall need is E^n_{OK} which is the event that our stopping times work the way we want, that is,

$$E^n_{\text{OK}} := \left\{ \forall k \le \exp(n^{0.001}),\ S_{\tau_k} \in I^n_3 \text{ and } \tau_{\exp(n^{0.001})} \le \exp(n^{0.003}) \right\}.$$

With these definitions, we are ready to formulate the four intermediate results which we will need in order to prove Theorem 4.1. The first lemma is of a combinatorial nature.

Lemma 4.1. *For n sufficiently large,* $E^n_{\text{no-error}} \cap E^n_{\text{visits}} \cap E^n_{\text{marker-works}} \subset E^n_{\text{OK}}$.

Proof. When it occurs, the event $E^n_{\text{no-error}}$ guarantees that all the stopping times in T up to time $\exp(n^{0.003})$ stop the random walk inside the interval I^n_2. Since $I^n_2 \subset I^n_3$, $E^n_{\text{no-error}}$ implies that

$$S_{\tau_k} \in I^n_3 \quad \text{for all } \tau_k \le \exp(n^{0.003}).$$

Next, if E^n_{visits} and $E^n_{\text{marker-works}}$ both hold, then there are at least $\exp(n^{0.002})/3$ stopping times in T which occur prior to time $\exp(n^{0.003})$. In other words,

$$\tau_{\exp(n^{0.002})/3} \le \exp(n^{0.003}).$$

Now, when n is sufficiently large, $n^{0.001} \le n^{0.002}/3$ and so

$$\tau_{\exp(n^{0.001})} \le \exp(n^{0.003}).$$

Consequently, the simultaneous occurrence of both E^n_{visits} and $E^n_{\text{marker-works}}$ implies that $\tau_k \le \exp(n^{0.003})$ for all $k \le \exp(n^{0.001})$ when n is large.

Finally, if $E^n_{\text{no-error}}$ holds in addition to E^n_{visits} and $E^n_{\text{marker-works}}$, then we also see that $S_{\tau_k} \in I^n_3$ for all $k \le \exp(n^{0.001})$. Thus, when all three events

$$E^n_{\text{no-error}}, \quad E^n_{\text{visits}} \quad \text{and} \quad E^n_{\text{marker-works}}$$

occur simultaneously, then E^n_{OK} must also occur. □

The next three results yield lower bounds on the quantities $P_2(E^n_{\text{no-error}})$, $P_2(E^n_{\text{visits}})$ and $P_2(E^n_{\text{marker-works}})$.

Lemma 4.2. *For n large,*

$$P_2(E^n_{\text{no-error}}) \ge 1 - (0.6)^{n^{0.005}}. \tag{20}$$

Proof. Let $\kappa_{z,l}$ denote the time of the lth visit by the random walk S to the point z. Let $E^n_{\text{no-error},z,l}$ denote the event that there is no significantly low number of ones immediately following the stopping time $\kappa_{z,l}$, that is,

$$E^n_{\text{no-error},z,l} := \left\{ \sum_{s=\kappa_{z,l}}^{\kappa_{z,l}+n^{0.1}} \tilde{\chi}_s > \epsilon n^{0.1} \right\}.$$

Up to time $t = \exp(n^{0.003})$, the random walk cannot visit points z further away from the origin than $L\exp(n^{0.003})$ nor can it visit a point more than $\exp(n^{0.003})$ times. Thus, all the times which appear in the definition of the event $E^n_{\text{no-error}}$, that is all the times t for which $t \leq \exp(n^{0.003})$ and $S_t \notin I^n_2$, include the set of times $\kappa_{z,l}$ for which $z \in (I^n_7 \setminus I^n_2)$ and $l \leq \exp(n^{0.003})$. This implies that

$$\bigcap_{z,l} E^n_{\text{no-error},z,l} \subset E^n_{\text{no-error}}, \tag{21}$$

where the intersection is taken over all $z \in (I^n_7 - I^n_2)$ and $l \leq \exp(n^{0.003})$.

If the random walk S is outside the interval I^n_2 at time t, then it is impossible for the random walk to reach the interval I^n_1 within time $n^{0.1}$. Thus if $S_t \notin I^n_2$ then S_s cannot be in I^n_1 for all times $s \in [t, t+n^{0.1}]$. However, outside the interval I^n_1, the scenery ξ has the same distribution under $P(\cdot)$ as it does under $P_2(\cdot)$. Thus, for $z \notin I^n_2$,

$$P_2(E^n_{\text{no-error},z,l}) = P(E^n_{\text{no-error},z,l}).$$

Furthermore, since the distribution of the scenery under $P(\cdot)$ is both time and spatially homogeneous, an application of the strong Markov property yields

$$P(E^n_{\text{no-error},z,l}) = P(E^n_{\text{no-error},0,0}) = P\left(\sum_{s=0}^{n^{0.1}} \tilde{\chi}_s > \epsilon n^{0.1}\right),$$

for all $z \notin I^n_2$. However the event $\left\{\sum_{s=0}^{n^{0.1}} \tilde{\chi}_s > \epsilon n^{0.1}\right\}$ is just the event A^{mc} from Section 3 with $m = n^{0.05}$. Hence, from Lemma 3.8 we obtain

$$P(E^{nc}_{\text{no-error},z,l}) = P(A^m) \leq 2Lm^2 \left(\frac{1}{2}\right)^{m^{0.1}} = 2Ln^{0.1}\left(\frac{1}{2}\right)^{n^{0.005}} \tag{22}$$

for all $z \notin I^n_2$. By combining this with (21), we arrive at

$$P(E^{nc}_{\text{no-error}}) \leq \sum_{z \in I^n_7 \setminus I^n_2, 0 \leq l \leq \exp(n^{0.003})} P(E^{nc}_{\text{no-error},z,l}) \leq 2L\exp(2n^{0.003}) \cdot 2Ln^{0.1}\left(\frac{1}{2}\right)^{n^{0.005}}.$$

The final inequality comes about by recognizing that there are fewer than $2L\exp(2n^{0.003})$ pairs (z,l) with $z \in I^n_7 \setminus I^n_2$ and $0 \leq l \leq \exp(n^{0.003})$.

Now, the dominating term in the bound on the right-hand side of this inequality is $(1/2)^{n^{0.005}}$. Thus, for n big enough, the expression on the right-hand side of the last inequality is smaller than $(0.6)^{n^{0.005}}$. The result follows by applying this bound to $E^n_{\text{no-error}}$. $\square$

Lemma 4.3. *For large n,*

$$P_2(E^n_{\text{marker-works}}) \geq 1 - \exp(-0.225\exp(n^{0.002})). \tag{23}$$

Proof. Let R be a random walk with increments identical to those of the random walk S but starting at the random point z_c. Thus, $R_t := S_t + z_c$. Let χ^R_t denote the observation made by the random walk R at time t of the scenery ξ, that is, $\chi^R_t := \xi(R_t)$. We shall use $\tilde{\chi}^R_t$ to denote that same observation made with an error:

$$\tilde{\chi}^R_t := \chi^R_t + \nu_t \bmod 2.$$

Let E_R^n denote the event that R does not stray from z_c by a distance greater than $n^{0.1}/2$ before time $n^{0.1}$:

$$E_R^n := \left\{ \forall t \leq n^{0.1}, |R_t - z_c| \leq n^{0.1}/2 \right\}.$$

Note that when E_R^n occurs, the random walk R stays within the contiguous block of zeros in ξ having z_c at its center during its first $n^{0.1}$ steps. Consequently, if E_R^n holds, we have

$$\sum_{t=0}^{n^{0.1}} \chi_t^R = 0.$$

It follows, conditional on E_R^n, that $\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \sim \mathrm{Bin}(n^{0.1}, \epsilon)$. Then, by the central limit theorem, as n tends to infinity,

$$P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \mid E_R^n \right)$$

converges to $1/2$. Now,

$$P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \right) = P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \mid E_R^n \right) P_2(E_R^n)$$
$$+ P_2\left(\left\{ \sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \right\} \cap E_R^{nc} \right). \tag{24}$$

By Part 2 of Lemma 3.2, $P_2(E_R^n)$ converges to one as n converges to infinity. It also follows that $P_2(\{\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1}\} \cap E_R^{nc})$ converges to zero as n tends to infinity. Hence,

$$P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \right) \longrightarrow \frac{1}{2}$$

as $n \to \infty$.

Next, let us assume that n is large enough so that

$$P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1} \right) \geq 0.49. \tag{25}$$

Define $\mathcal{G}_k$ to be the σ-algebra

$$\mathcal{G}_k := \sigma(z_c, \xi_z; S_0, S_1, \ldots, S_{\kappa_k} + n^{0.1} \mid z \in \mathbb{Z})$$

and let $\mathcal{G}$ denote the filtration $\mathcal{G} := \bigcup_k \mathcal{G}_k$. It can be seen that the sequence of random variables $Y_1, Y_2, \ldots$ is $\mathcal{G}$-adapted. Furthermore, by definition, the stopping times κ_k are at least $n^{0.1}$ time steps apart from each other. It follows that κ_{k+1} happens no earlier than time $\kappa_k + n^{0.1}$. By the strong Markov property of the random walk S, when we stop the process at a point, it then continues on as though it were a new random walk which was started at that point, independent of what happened beforehand. Putting it another way, conditional on $\mathcal{G}_k$, S is distributed after time κ_{k+1} like R. So,

$$P_2(Y_{k+1} = 1 \mid \mathcal{G}_k) = P_2\left(\sum_{s=\kappa_{k+1}}^{\kappa_{k+1}+n^{0.1}} \tilde{\chi}_s \leq \epsilon n^{0.1} \,\middle|\, \mathcal{G}_k\right) = P_2\left(\sum_{t=0}^{n^{0.1}} \tilde{\chi}_t^R \leq \epsilon n^{0.1}\right) P_2\text{-a.s.}$$

According to (25), the final expression in the equality above is greater than 0.49 for n sufficiently large and, hence, $\mathbb{E}[Y_{k+1}] \geq 0.49$. We can therefore use Lemma 3.1. Setting $k = \exp(n^{0.002})$, $a = 1/\sqrt{2}$ and $\Delta = 0.15$, we obtain

$$\begin{aligned}
P_2(E^{nc}_{\text{marker-works}}) &= P_2\left(\sum_{k=1}^{\exp(n^{0.002})} Y_k < \exp(n^{0.002})/3\right) \\
&\leq P_2\left(\frac{\sum_{k=1}^{\exp(n^{0.002})} (Y_k - \mathbb{E}[Y_k])}{\exp(n^{0.002})} < 1/3 - 0.49\right) \\
&\leq P_2\left(\frac{\sum_{k=1}^{\exp(n^{0.002})} (Y_k - \mathbb{E}[Y_k])}{\exp(n^{0.002})} \leq -0.15\right) \\
&\leq \exp(-0.225 \exp(n^{0.002})).
\end{aligned}$$

Thus, $P_2(E^n_{\text{marker-works}}) \geq 1 - \exp(-0.225 \exp(n^{0.002}))$ asymptotically. $\square$

Lemma 4.4. *For large n,*

$$P_2(E^n_{\text{visits}}) \geq 1 - \exp(-n^{0.003}/4). \tag{26}$$

Proof. Let $s := n^{0.1} \exp(n^{0.002})$ and observe that

$$E^n_{\text{visits}} = \left\{\kappa_{\exp(n^{0.002})} \leq \exp(n^{0.003})\right\} = \left\{\kappa^*_s \leq \exp(n^{0.003})\right\}.$$

Without loss of generality, assume that $z_c = 0$. If z_c is not zero, the proof is virtually the same since z_c is at most a distance polynomial in n away from the origin, which has negligible influence on the event, since we are considering exponentially long times in n. When $z_c = 0$, the event E^n_{visits} is simply the event that the random walk S visits the origin no less than s times before time $\exp(n^{0.003})$. Let Z_k denote the kth interarrival time between consecutive visits by S to the origin. Hence, $\sum_{l=1}^{k} Z_l$ is the time of the kth visit by S to the origin. Note that the random variables Z_k, $k \in \mathbb{N}$, are i.i.d. Define n_3 to be the number $n_3 := \exp(n^{0.003})$. Under the assumption that $z_c = 0$ (which changes the ultimate bound we shall find in only a minor way), we have that

$$P_2(E^{nc}_{\text{visits}}) = P_2\left(\sum_{k=1}^{s} Z_k > n_3\right).$$

Now,

$$P_2\left(\sum_{k=1}^{s} Z_k > n_3\right) = P_2\left(\left(\sum_{k=1}^{s} Z_k\right)^{1/3} > n_4\right),$$

where $n_4 := (n_3)^{1/3}$. For any set of positive numbers $\{a_l\}_1^j$, it is always true that $(\sum_{l=1}^{j} a_l)^3 \geq \sum_{l=1}^{j}(a_l)^3$. Hence, $\sum_{k=1}^{s}(Z_k)^{1/3} \geq (\sum_{k=1}^{s} Z_k)^{1/3}$ and so

$$P_2(E^{nc}_{\text{visits}}) \leq P_2\left(\sum_{k=1}^{s}(Z_k)^{1/3} > n_4\right).$$

By the Markov inequality,

$$P_2(E^{nc}_{\text{visits}}) \leq \frac{s\mathbb{E}_2[(Z_1)^{1/3}]}{n_4} = \frac{n^{0.1}\exp(n^{0.002})\mathbb{E}_2[(Z_1)^{1/3}]}{\exp(n^{0.003}/3)}. \tag{27}$$

It is known that $\mathbb{E}_2[(Z_k)^{1/3}]$ is finite (see for example Durrett [4]) and thus is a constant not depending on n. Furthermore, the dominating factor in the bound given in (27) is $\exp(-n^{0.003}/3)$. It follows that, for n large enough, the right-hand side of (27) is smaller than $\exp(-n^{0.003}/4)$. □

Proof of Theorem 4.1. Lemma 4.1 yields

$$P_2(E^{nc}_{\text{OK}}) \leq P_2(E^{nc}_{\text{no-error}}) + P_2(E^{nc}_{\text{visits}}) + P_2(E^{nc}_{\text{marker-works}}). \tag{28}$$

For the three quantities $P_2(E^{nc}_{\text{no-error}})$, $P_2(E^{nc}_{\text{visits}})$ and $P_2(E^{nc}_{\text{marker-works}})$, we have the bounds (20), (26) and (23) respectively. The largest of these bounds is given by (26). Since $P_2(E^{nc}_{\text{OK}})$ is asymptotically smaller than 3 times this bound, we can write $P_2(E^{nc}_{\text{OK}}) \leq 3\exp(-n^{0.003}/4)$ for n large. □

5. Recognizing markers in error-corrupted observations

In the preceding section, we investigated the case where we condition on the event B^n. Unfortunately, B^n is not an observable event. So instead, we need to condition on an event we are able to observe. We shall therefore choose to condition on A^n, which is observable. From Theorem 3.1, we know that, whenever A^n is observed, there is a block of zeros of length greater than $n^{0.1}$ close to the origin with high probability. (Here, close to the origin means belonging to $[-Ln^2, Ln^2]$.) We can then use this abnormally long block of zeros as a marker. This enables us to construct a total of $\exp(n^{0.001})$ stopping times τ_k and, with high probability, these stopping times all stop the random walk S in the interval $[-2Ln^2, 2Ln^2]$. This is a situation similar to the one described in Section 2, where we had a 2 at the origin. When we previously conditioned on the event B^n, we "forced" the scenery to have a marker close to the origin. We did this in order to simplify notation in the preceding argument. In reality, we have to search for a marker first. We shall now show how this can be done. (For simplification the following explanation is for $P_2(\cdot)$ rather than for $P(\cdot|B^n)$.)

Let τ^* denote the first time t at which we see a string of length n^2 with less than ϵn^2 ones in the error-corrupted observations:

$$\tau^* := \min\left\{t > 0 : \sum_{s=0}^{n^2} \tilde{\chi}_{t+s} \leq \epsilon n^2\right\}.$$

Since ξ, S and ν are mutually independent and S is a recurrent random walk, the stopping time τ^* must be almost surely finite, that is, $P(\tau^* < \infty) = 1$. The neighborhood of S_{τ^*} is very similar to the origin under the conditional probability measure $P_2(\cdot)$. Due to the spatial homogeneity of the scenery, the theory which we developed in the last section holds for the point $z = S_{\tau^*}$ instead of the origin. Hence, with high probability, there is a block of more than $n^{0.1}$ contiguous zeros in the interval

$$I_{\tau^*} := [S_{\tau^*} - 2Ln^2, S_{\tau^*} + 2Ln^2].$$

Using this block of zeros as a marker, we can then construct a total of $\exp(n^{0.001})$ stopping times which, with high probability, all stop the random walk S in I_{τ^*}. We shall denote this sequence of stopping times by $\{\bar{\tau}_k\}_{k>0}$. They are defined as follows:

Definition 5.1. For $k > 0$, let $\bar{\tau}_k$ denote the kth element (under the usual ordering on $\mathbb{N}$) of the set $T \cap [\tau^*, \infty)$. Note that $\bar{\tau}_1 = \tau^*$.

The result is that with high probability the first $\exp(n^{0.001})$ stopping times $\bar{\tau}_k$ stop S in I_ν.

Theorem 5.1. *The probability*

$$P\left(\forall k \le \exp(n^{0.001}),\ S_{\bar{\tau}_k} \in I_{\tau^*} \text{ and } (\bar{\tau}_{\exp(n^{0.001})} - \tau^*) \le \exp(n^{0.003})\right)$$

tends to one as $n \to \infty$.

Proof. The proof is analogous to that of Theorem 4.1. $\square$

These stopping times can be used to reconstruct a little piece of the scenery ξ in the neighborhood of the point S_{τ^*}. The methods which can be used for this are similar to what was described in Section 2.

In [18], Lember and Matzinger show how being able to reconstruct a small amount of information contained in the neighborhood of markers implies that the whole scenery ξ can be reconstructed almost surely. Their proof, however, only pertains to the case of observations made without errors. The question as to whether or not it is possible to perform scenery reconstruction from error-corrupted observations of a two-color scenery remains open.

Acknowledgement

I would like to thank the referee for correcting this paper very carefully and giving useful suggestions for improvement.

References

[1] I. Benjamini, H. Kesten, Distinguishing sceneries by observing the scenery along a random walk path, J. Anal. Math. 69 (1996) 97–135.
[2] W.Th.F. den Hollander, Mixing properties for random walk in random scenery, Ann. Probab. 16 (1988) 1788–1802.
[3] F. den Hollander, J.E. Steif, Mixing properties of the generalized T, T^{-1}-process, J. Anal. Math. 72 (1997) 165–202.
[4] R. Durrett, Probability: Theory and Examples, second ed., Duxbury Press, Belmont, CA, 1996.
[5] D. Heicklen, C. Hoffman, D.J. Rudolph, Entropy and dyadic equivalence of random walks on a random scenery, Adv. Math. 156 (2000) 157–179.
[6] M. Harris, M. Keane, Random coin tossing, Probab. Theory Related Fields 109 (1997) 27–37.

[7] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Assoc. 58 (1963) 215–247.
[8] C.D. Howard, Orthogonality of measures induced by random walks with scenery, Combin. Probab. Comput. 5 (1996) 247–256.
[9] C.D. Howard, Distinguishing certain random sceneries on $\mathbb{Z}$ via random walks, Statist. Probab. Lett. 34 (1997) 123–132.
[10] S.A. Kalikow, T, T^{-1} transformation is not loosely Bernoulli, Ann. Math. (2) 115 (1982) 393–409.
[11] M. Keane, W.Th.F. den Hollander, Ergodic properties of color records, Physica A 138 (1986) 183–193.
[12] H. Kesten, Detecting a single defect in a scenery by observing the scenery al a random walk path, in: Itô's Stochastic Calculus and Probability Theory, Springer, Tokyo, 1996, pp. 171–183.
[13] H. Kesten, Distinguishing and reconstructing sceneries from observations along random walk paths, in: Microsurveys in Discrete Probability (Princeton, NJ, 1997), Amer. Math. Soc., Providence, RI, 1998, pp. 75–83.
[14] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, Springer-Verlag, New York, 1991, p. 113.
[15] E. Lindenstrauss, Indistinguishable sceneries, Random Structures Algorithms 14 (1999) 71–86.
[16] A. Lenstra, H. Matzinger, Reconstructing a 4-color scenery by observing it along a recurrent random walk path with unbounded jumps, Eurandom (2001) (in preparation).
[17] M. Löwe, H. Matzinger, Scenery reconstruction in two dimensions with many colors, Ann. Appl. Probab. 12 (2002) 1322–1347.
[18] J. Lember, H. Matzinger, Reconstructing a 2-color scenery by observing it along a recurrent random walk path with bounded jumps, Eurandom (2003) (in preparation).
[19] M. Löwe, H. Matzinger, F. Merkl, Reconstructing a multicolor random scenery seen along a random walk path with bounded jumps, Electronic Journal of Probability 15 (2004) 436–507.
[20] M. Löwe, H. Matzinger, Reconstruction of sceneries with correlated colors, Stochastic Process. Appl. 105 (2003) 175–210.
[21] D. Levin, Y. Peres, Identifying several biased coins encountered by a hidden random walk, Random Structures Algorithms 25 (2004) 91–114.
[22] D.A. Levin, R. Pemantle, Y. Peres, A phase transition in random coin tossing, Ann. Probab. 29 (2001) 1637–1669.
[23] H. Matzinger, Reconstructing a 2-color scenery by observing it along a simple random walk path with holding, Ph.D. Thesis, Cornell University, 1999.
[24] H. Matzinger, Reconstructing a three-color scenery by observing it along a simple random walk path, Random Structures Algorithms 15 (1999) 196–207.
[25] H. Matzinger, Reconstructing a 2-color scenery seen along a simple random walk path, Ann. Appl. Probab. 1B (2005) 778–819.
[26] H. Matzinger, S.W.W. Rolles, Reconstructing a random scenery observed with random errors along a random walk path, Probability Theory and Related Fields 125 (2003) 539–577.
[27] H. Matzinger, S.W.W. Rolles, Reconstructing a piece of scenery with polynomially many observations, Stochastic Processes and their Applications 107 (2002) 289–300.
[28] D. Williams, Probability with Martingales, Cambridge University Press, Cambridge, 1991.