ALEA, Lat. Am. J. Probab. Math. Stat. 9, 85–99 (2012)

Recognition of times a walker is close to the origin

Heinrich Matzinger and Angelica Pachon Pinzon

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332.
E-mail address: matzi@math.gatech.edu

Centre for Discrete Mathematics and its Applications DIMAP, University of Warwick, Coventry CV4 7AL, United Kingdom.
E-mail address: A.Y.Pachon-Pinzon@warwick.ac.uk

Abstract. In this article we consider a simple random walker moving on a random media. Whilst doing so, the random walker observes at each point of time the "color" of the location he is at. This process creates a sequence of observations. We consider the problem of determining when the walker is close to the origin. For this we are only given, the observations made by the walker as well as a small portion of the media close to the origin. With that information alone, we show that we can typically construct an exponential number of stopping times, which all occur whilst the walker is on the small piece of media available to us. The number is exponential in the size of that small piece of media. So far this problem could only be solved when the media contained 5 colors.

In the present article, we use a subtle entropy argument on the set of possible observations given the point where the walker starts and given the media in that neighborhood. This allows us to achieve our goal when the media contains 4 equiprobable colors. The random media is often called "scenery".

An important area of research is Scenery Reconstruction, in which one tries to retrieve the scenery based on the observations made by the random walker alone. Finding times when the random walker is close to the origin is the main hurdle for building a scenery reconstruction algorithm. Our present result, implies that the Scenery Reconstruction result in Hart et al. (2011) also applies with 4 colors as opposed to just 5 colors. The less colors the more difficult these problems become.

1. Introduction

Let S_t denote the position of a random walker at time t. We assume that S_t is a simple symmetric random walk starting at the origin. Let $\xi : \mathbb{Z} \to \{0, 1, 2, 3\}$ denote a coloring of the integers with 4-colors. We call the landscape ξ a scenery.

Received by the editors October 2, 2009; accepted January 10, 2012.
2000 *Mathematics Subject Classification.* 60K37, 60G50.
Key words and phrases. Random walk, Random media, Stopping times.

We assume that at every time $t \geq 0$, the random walker sees the color of the point he is at. This implies that at time t the random walker observes the color $\chi_t := \xi(S_t)$.

The problem we consider in this article is to figure out times when the random walker is in a close vicinity of the origin. For this we are given only the observations

$$\chi := \chi_0\chi_1\chi_2\ldots$$

and the restriction of ξ to $[0, n-1]$. The way we try to guess when the random walk is close to the origin, is by searching in the observations for the word

$$w^n := \xi_0\xi_1\xi_2\ldots\xi_{n-1}.$$

More precisely, consider the stopping times $\tau_1, \tau_2, \ldots$ defined as follows:

$$\tau_1 := \min\{t \geq n | w^n = \chi_{t-n+1}\chi_{t-n+2}\cdots\chi_t\}.$$

By induction on i, τ_{i+1} is the next time the pattern w^n appears in the observations:

$$\tau_{i+1} := \min\{t > \tau_i | w^n = \chi_{t-n+1}\chi_{t-n+2}\cdots\chi_t\}.$$

We take the scenery ξ to be random ergodic. The pattern w^n will appear infinitely often in the scenery ξ. So there is no hope that at all the times τ_i the walker is close to the origin, because he will also observe the pattern w^n in other locations. Instead, we will prove that an exponential number of the times τ_i tell us that the walker is close to the origin. For this B^n, is the event that the walker is close to 0 for all τ_i with $i \leq (v_2)^n$. Here v_2 is a constant not depending on n satisfying

$$\frac{2}{2^{H_2(0.25)}} > v_2 > 1, \tag{1.1}$$

where $H_2(x)$ is the entropy function:

$$H_2(x) := x\log_2(1/x) + (1-x)\log_2(1/(1-x)).$$

(Note that $2/2^{H_2(0.25)} > 1$, so that a constant v_2 satisfying 1.1 really exists!). In the event B^n, "close to the origin" is defined as in the interval $[0, n-1]$, so that

$$B^n := \{S_{\tau_i} \in [0, n-1], \forall i \leq v_2^n\}.$$

Our main theorem states that when the scenery ξ is taken i.i.d. with four equiprobable colors then the probability that B^n does not hold is exponentially small in n. (This is true for any constant v_2 satisfying 1.1 but not depending on n). Here comes our main theorem:

Theorem 1.1. *Assume that ξ_z with $z \in \mathbb{Z}$ is a collection of i.i.d. variables independent of the simple symmetric random walk S_t starting at the origin. We also assume that the variables ξ_z are equally likely to be equal to* 0,1, 2 *or* 3*:*

$$P(\xi_z = 0) = P(\xi = 1) = P(\xi = 2) = P(\xi = 3) = 1/4.$$

Then, all the stopping times τ_i up to $i = v_2^n$ stop the random walk S with high probability in $[0, n-1]$:

$$P(B^n) \geq 1 - e^{-c_B n}$$

for all $n \in \mathbb{N}$, where $c_B > 0$ is a constant not depending on n.

In Hart et al. (2011), it is proven that an exponential number of times τ_i stop the random walk close to the origin in the context of a 5-color scenery. However the proof in Hart et al. (2011) fails with less than 5 colors. We introduce a subtle entropy argument for the class of observations generated by a walker, which allows this improvement. The technique we develop here is important and we expect it to be useful in many other situations.

The stopping time problem considered here is an essential step for scenery reconstruction. Once many stopping times are constructed, it is relatively easy to reconstruct a large portion of the scenery around the origin. Once the stopping times are available, the scenery reconstruction can be performed exactly as in Hart et al. (2011). The present result implies that the scenery reconstruction result proven in Hart et al. (2011) for 5 color sceneries also holds with 4-color sceneries. We explain more details on this in subsection 1.1.

Let us at this stage explain the scenery reconstruction problem: This problem goes back to questions from Kolmogorov, Kesten, Keane, Benjamini, Perez, Den Hollander and others. The scenery reconstruction problem considers the same setting as in the present article. This means that a recurrent random walk makes observations of a random media, called scenery. One tries to reconstruct the random media based on the observations alone. In general it is only possible to reconstruct the scenery up to translation and reflection around the origin. There exist sceneries which can not be reconstructed. (Lindenstrauss in [12] exhibited sceneries which can not be reconstructed.) To overcome this obstacle, we take the scenery to be random and prove that almost every scenery can a.s. be reconstructed up to translation and reflection.

The first positive result Matzinger (1999) on Scenery Reconstruction was Matzinger's PhD-thesis in which he showed that one can reconstruct up to equivalence almost every 2-color i.i.d scenery seen along the path of a simple symmetric random walk with holding. Many other results followed: for non skip-free walker Löwe et al. (2004); Lember and Matzinger (2008, 2003) in two dimensions Löwe and Matzinger (2002) a very different approach to Matzinger (2005) is needed. The reconstruction of a finite piece in polynomial time is treated in Matzinger and Rolles (2003a, 2006b,a), whilst reconstruction with errors in the observations is the subject of Matzinger and Rolles (2003b); Hart and Matzinger (2006). In Matzinger and Popov (2007), a continuous analogon is considered. Scenery reconstruction methods differ greatly depending on the distribution of the scenery and the random walk.

1.1. *Implication of present result for scenery reconstruction.* The research in this area started with people investigating the ergodic properties of the observations made by a random walker on a random media. Kesten Kesten (1996) proved that with five colors, if one knows the scenery in every point except in one, then, it becomes possible to reconstruct the missing color in that one location. For this purpose, the observations ξ are "observable". This problem is called "distinguishing a single defect in a scenery". But, at that time, specialists believed that it might not be possible to distinguish single defects with less than 5 colors in the scenery. Hence, the result in Matzinger (1999) came as a surprise. Nontheless, the general question remains open: when does it become impossible to reconstruct a scenery? When does reducing the entropy in the scenery whilst increasing it for the walker lead to a critical phenomena where the scenery becomes unreconstructable?

The article Hart et al. (2011) is the first and only, where scenery reconstruction is shown to be possible despite the increment of the random walk having a non-bounded support. In Hart et al. (2011), a symmetric random walk has its distribution close to a symmetric random walk, but with small probability $\delta > 0$ the steps can be larger than 1 unit. The conditions in Hart et al. (2011) for the random walk can be written as

$$P(|S_{t+1} - S_t| \neq 1) \leq \delta \tag{1.2}$$

and

$$P(|S_{t+1} - S_t| = m \mid |S_{t+1} - S_t| \neq 1) \leq e^{-cm}, \ \forall m \in \mathbb{N} \tag{1.3}$$

for a constant $c > 0$ not depending on m. Hence, the step length has an exponentially decaying tail, but non-bounded support. Even when $\delta > 0$ is taken very small and $c > 0$ very large, all other reconstruction methods fail. The algorithm in Hart et al. (2011) achieves reconstruction for $\delta > 0$ small enough and $c > 0$ large enough. It reconstructs the scenery on larger and larger intervals. Once it has reconstructed the restriction of ξ to $[-n, n]$, it proceeds in determining ξ_{n+1} and ξ_{-n-1}. For this, it first obtains an exponential number of stopping times. It uses the observations χ and the already reconstructed piece $\xi_{-n}\xi_{-n+1} \dots \xi_{n-1}\xi_n$. These stopping times are shown to typically all occur whilst the walker is in $[-n, n]$. With the availability of these stopping times, the reconstruction of ξ_{n+1} and ξ_{-n-1} is relatively easy. We can use the stopping times constructed in this paper in the same way as in Hart et al. (2011) to obtain ξ_{n+1} and ξ_{-n-1}. The stopping times are even defined in the same way, both here and in Hart et al. (2011). The only difference is that here we prove them to work with only 4 colors instead of 5. The fact that in Hart et al. (2011), we do not only consider a simple random walk but also a slightly disturbed version of a simple random walk does not matter: the proof and methods provided here does also carry over to that case. This implies that scenery reconstruction is possible with a slightly disturbed random walk (i.e. taking $\delta > 0$ small enough and $c > 0$ large and assuming that 1.2 and 1.3 hold for a symmetric random walk S), even if there are only 4 equiprobable colors in the scenery.

Let us explain a little more on how the present proof in this article can be adapted for the situation where the random walk is not exactly simple but very close to it. At the beginning of the next section we define a constant $r > 0$ which does not depend on n. We then go on defining T to be the first time that the random walk S visits the set $\{-r^n, r^n\}$. Our main theorem 1.1 asserts that with high probability the first v_2^n stopping times τ_i all stop S in the interval $[0, n-1]$. But, in the proof of theorem 1.1 we actually prove slightly more: we also show that with high probability those first v_2^n stopping times all occur before time T. (See lemma 2.1.) Now, the proof of the main theorem in this paper, which is valid for S being a simple random walk, can be summarized as follows:

let $\mathcal{P}_x^n$ to be the set of all nearest neighbor paths of length n starting at x and such that the percentage of back-forth steps is less than $q_0 > 1/4$. (Defined at the beginning of the next section). Let x be a non-random point so that $|x| > 2n$. Then for any (non-random) nearest neighbor path R in $\mathcal{P}_x^n$, we have that the probability that the word w is generated by R is equal to $(1/4)^n$. That is:

$$P(w^n = \xi \circ R) = \frac{1}{4^n}.$$

The above probability follows from the fact that in the scenery ξ there are 4 equally likely colors and the scenery is i.i.d.. Also, since the scenery ξ is i.i.d., we get that $\xi \circ R$ is independent of the word $w^n = \xi_0\xi_1 \dots \xi_{n-1}$. (This is because, R can not enter the interval $[0, n-1]$, since it starts outside $[-2n, 2n]$ and moves by one unit at a time). Hence, the probability that there exists a path R in $\mathcal{P}_x^n$ generating w can be bounded as follows:

$$P(\exists R \in \mathcal{P}_x^n, \text{so that } \xi \circ R = w) \leq \frac{|\mathcal{P}_x^n|}{4^n}. \tag{1.4}$$

In other words, if the cardinality $|\mathcal{P}_x^n|$, is of a much lesser order than 4^n, then with high probability there will be no nearest neighbor path starting in x and generating w. Now, our main proof works by showing that the right side of 1.4 is even of much lesser order than r^n. From this it follows then immediately that with high probability:

within the interval $[-r^n, r^n]$, there is no place where a simple nearest neighbor walk can generate w except very close to the origin. The only thing needed to show the above is that

$$|\mathcal{P}_x^n| < r^n \cdot 4^n, \tag{1.5}$$

where $<$ stands for "lesser exponential order than". The inequality 1.5 is what truly makes the present paper "work". By itself it guaranties that with high probability, until the random walk hits $\{r^n, r^n\}$, all the stopping times τ_i stop the random walk S in $[-2n, 2n]$. Inequality 1.5 is shown in the proof of lemma 2.3. For proving the main theorem in the case of a random walk which is not simple, but only close to simple, we only need to prove inequality 1.5 for that case. So, instead of the set of nearest neighbor walks $\mathcal{P}_x^n$, we need the set $\mathcal{P}_x^n(\lambda, s)$ of λ, s-walks: a map $R : [0, n-1] \to \mathbb{Z}$ is a called *λ, s-walk* if the number of i's in $[0, n-2]$ for which $|R(i+1) - R(i)| \neq 1$ are less than λn. We also require that the sum of the absolute value of the steps which are not $+-1$ does not exceed sn:

$$\sum_{i \in [0,n-2], |R(i+1)-R(i)| \neq 1} |R(i+1) - R(i)| \leq sn.$$

Note that by taking the constants $s, \delta > 0$ small enough (not depending on n), the exponential rate of

$$|\mathcal{P}_x^n(\lambda, s)|$$

gets as close as we want to $|\mathcal{P}_x^n|$. In this manner, since equation 1.5 holds, we get by choosing $s, \lambda > 0$ small enough (but not depending on n), that

$$|\mathcal{P}_x^n(\lambda, s)| < r^n \cdot 4^n, \tag{1.6}$$

This would then make the stopping time work for a "close to simple" random walk S, provided we can show that with high probability, up to time T, S only follows λ, s-paths. More precisely, let G^n be the event that for all $t \leq T$, we have that

$$S_t, S_{t+1}, \dots, S_{t+n-1}$$

is a λ, s-path. By taking the parameters $\delta > 0$ and $c > 0$ in 1.2 and 1.3 small enough (but not depending on n), we get that the event G^n has probability close to 1 up to an exponentially small quantity in n. Together with inequality 1.6, this does the trick of proving the present result for a non-simple random walk given that the parameters $\lambda, s > 0$ are small enough

The new idea used here gives us hope for sceneries with lower entropy. When entropy is low, scenery reconstruction becomes way more difficult. However, the present technique offers a new approach: If the string

$$w^n := \xi_0 \xi_1 \dots \xi_{n-1}$$

has low entropy, then we should also be able to very much restrict the collection of strings generated by a walker starting in a given point x on the scenery ξ and which might lead to w^n. (See below the argument restricting path which might generate w^n).

2. Proof of main theorem

2.1. *Definition of events and combinatorics.* Recall that w^n designates the word obtained by restricting the scenery ξ to $[0, n-1]$:

$$w^n := \xi_0\xi_1\xi_2 \dots \xi_{n-1}.$$

Note that

$$\frac{4}{2^{H_2(0.25)}} > 2.$$

In the introductory section we defined the constant $v_2 > 1$ to be any constant not depending on n and satisfying 1.1. We will also need the constants r and v_1 which shall not depend on n, but satisfy the equation

$$\frac{4}{2^{H_2(0.25)}} > r > v_1 > 2v_2 > 2. \tag{2.1}$$

Let $q_0 > 0.25$ denote a constant not depending on n such that

$$\frac{4}{2^{H_2(0.25)}} > \frac{4}{2^{H_2(q_0)}} > r. \tag{2.2}$$

Note that we can always find such a constant q_0 because the entropy function $H_2(.)$ is strictly increasing in the interval $[0, 0.5]$.

Let us quickly give a sneak preview of where the constants r, v_1 and v_2 make their appearance:

-With high likelihood, within a radius r^n of the origin there is no place where the word w^n can be read except at the origin.

-Typically at least $(v_1)^n$ visits to the origin occur before the random walk S leaves the interval $[-r^n, r^n]$.

-Typically, a number $(v_2)^n$ of stopping times all stop the random walk in $[0, n-1]$.

Let R be a map from the integer interval $[0, n-1]$ into $\mathbb{Z}$, i.e, $R : [0, n-1] \to \mathbb{Z}$, and such that

$$|R(i+1) - R(i)| = 1$$

for all $i \in [0, n-2]$. We call R a *nearest neighbor walk path* of length n and say R starts in x if $R(0) = x$.

Let $\mathcal{P}_x^n$ denote the set of all nearest neighbor paths of length n starting at x and such that the percentage of back-forth steps is less than $q_0 > 1/4$. Hence, $R : [0, n-1] \to \mathbb{Z}$ is in $\mathcal{P}_x^n$ iff both of the following conditions hold

(1) $R(0) = x$, and
(2) $|\{\, i \in [1, n-2] \mid (R(i) - R(i-1))(R(i+1) - R(i)) = -1\}| \leq q_0(n-2)$.

Let T denote the first visit by the random walk S to the set of two points $\{-r^n, r^n\}$:

$$T := \min\{\, t \mid |S_t| = r^n \,\}.$$

Next we define the events which we will use:

- Let ν_i denote the i-th visit after 0 by S to the origin:

$$\nu_{i+1} := \{t > \nu_i | S_t = 0\},$$

whilst $\nu_0 = 0$. (The random walk starts at the origin). Let C^n be the event that the random walk visits the origin at least $(v_1)^n$ times before time T:

$$C^n := \left\{\nu_i \leq T, \forall i \leq (v_1)^n \ \right\}.$$

- Let D_1^n be the event that there is no path R starting in $[-r^n, r^n] - [-2n, 2n]$ with less than q_0-percentage of back-forth steps and generating the word w^n. In other words, the event D_1^n means that if

$$x \in [-r^n, r^n] \text{ and } x \notin [-2n, 2n]$$

and $R \in \mathcal{P}_x^n$ then

$$\xi(R_0)\xi(R_1)\xi(R_2)\ldots\xi(R_{(n-1)}) \neq w^n.$$

- Let D_2^n be the event that no path starting in $[-2n, 2n]$ and ending outside $[0, n-1]$ whilst having less than q_0-percentage of back-forth steps can generate the word w^n. More precisely, the event D_2^n means that $\forall x \in [-2n, 2n]$ and all $R \in \mathcal{P}_x^n$, we have that

$$\xi(R_0)\xi(R_1)\xi(R_2)\ldots\xi(R_{(n-1)}) \neq w^n,$$

if $R(n-1) \notin [0, n-1]$.
- Let E^n be the event that the random walk crosses the interval $[0, n-1]$ in a straight way at least $(v_2)^n$ times among the first $(v_1)^n$ visits to the origin. More precisely, let E^n be the event the (random) set

$$\left\{\nu_i | i \leq (v_1)^n \ ; \ S_{j+1} - S_j = +1, \forall j \in [\nu_i, \nu_i + n - 1]\right\}$$

contains more than $(v_2)^n$ elements.
- Finally let F^n denote the event that the word w^n has a proportion less or equal to q_0 of letters w_i such that $w_i = w_{i+2}$. Hence, F^n is the event that

$$\text{Cardinality}\{i \in [1, n-1] | w_{i+1} = w_{i-1}\} \ \leq \ q_0(n-2).$$

Recall that B^n stands for the event that the first $(v_2)^n$ stopping times τ_i all occur whilst the random walk S is in the interval $[0, n-1]$.

Lemma 2.1. *We have that*

$$C^n \cap D_1^n \cap D_2^n \cap E^n \cap F^n \subset B^n.$$

Proof: With the event C^n we know that before time T, there are at least $(v_1)^n$ visits to the origin before time T. The event E^n guaranties that among the first $(v_1)^n$ visits to the origin, there are at least $(v_2)^n$ followed by a direct crossing of the interval $[0, n-1]$. When, the random walk S crosses the interval $[0, n-1]$ in a straight way, then during that time we see the pattern $w^n = \xi_0\xi_1\ldots\xi_{n-1}$ appearing in the observations. Thus, when C^n and E^n both hold, we see the pattern w^n appear at least $(v_2)^n$ times in the observations χ before time T. So, there will be at least $(v_2)^n$ stopping times τ_i before time T:

$$\tau_i \leq T \ , \forall i \leq (v_2)^n.$$

The next question is if those stopping times really stop the random walk in the interval $[0, n-1]$. With the event F^n, in the word w^n there are less than $q_0(n-2)$, letters w_i such that $w_i = w_{i+1}$. So, any nearest neighbor walk path with more than

$q_0(n-2)$ "back and forth" steps can not generate w^n on the scenery ξ. In other words, any nearest neighbor walk path $R:[0,n-1]\to\mathbb{Z}$ starting in x but not in $\mathcal{P}^n_x$ can not generate w^n:

$$\xi(R_0)\xi(R_1)\dots\xi(R_{n-1})\neq w^n.$$

So, when F^n holds, for a nearest neighbor walk path $R:[0,n-1]\to\mathbb{Z}$ to generate w^n, we need to have $R\in\mathcal{P}^n_x$ where $x:=R(0)$. By the event $D^n_1\cap D^n_2$, for all $x\in[-r^n,r^n]$, and all $R\in\mathcal{P}^n_x$, R can generate w^n only if it ends in $[0,n-1]$. That means that with the event $D^n_1\cap D^n_2$, for all $x\in[-r^n,r^n]$ and all $R\in\mathcal{P}^n_x$, we have

$$\xi(R_0)\xi(R_1)\dots\xi(R_{n-1})=w^n$$

implies $R(n-1)\in[0,n-1]$. Summarizing: when F^n and $D^n_1\cap D^n_2$ both hold, then the only way a nearest neighbor walk $R:[0,n-1]\to\mathbb{Z}$ can start in $[-r^n,r^n]$ and generate w^n on ξ, is when it ends in $[0,n-1]$, i.e. when $R(n-1)\in[0,n-1]$. Recall that by definition, up to time T the random walk S remains in $[-r^n,r^n]$. So, up to time T, when F^n and $D^n_1\cap D^n_2$ both hold, we can "only observe w^n when the random walk S follows a nearest neighbor walk path of length $n-1$ ending in $[0,n-1]$". In other words, for all $\tau_i\leq T$, we have that

$$S_{\tau_i}\in[0.n-1].$$

We have seen in the beginning of this proof, that when C^n and E^n both hold, then before time T we see the pattern w^n appear at least $(v_2)^n$ times in the observations χ. So, there are at least $(v_2)^n$ stopping times τ_i before time T. With F^n and $D^n_1\cap D^n_2$ holding all these stopping times occur when S_{τ_i} is in $[0,n-1]$. Hence, all this together implies that the first $(v_2)^n$ stopping times τ_i happen whilst S_{τ_i} is in $[0,n-1]$. Formally, we have proven that when all the events

$$C^n,E^n,F^n,D^n_1,D^n_2$$

hold then $S_{\tau_i}\in[0,n-1]$ for all $i\leq(v_2)^n$. This is the definition of the event B^n, so we have that

$$C^n\cap E^n\cap F^n\cap D^n_1\cap D^n_2\subset B^n.$$

□

2.2. *High probability of events.*

Lemma 2.2. *We have that*

$$P(C^n)\geq 1-\left(\frac{v_1}{r}\right)^n.$$

Proof: Let S^1_t be a simple random walk such that $S^1_0=1$. Define the stopping time

$$\tau^1=\min_t\Big\{S^1_t=0\text{ or }S^1_t=r^n\Big\}.$$

We know that for a stopping time thus defined

$$E(S^1_{\tau_1})=E(S^1_0),$$

then

$$r^nP\Big(S^1_{\tau_1}=r^n\Big)=1$$

and

$$P\Big(S^1_{\tau_1}=r^n\Big)=\frac{1}{r^n},$$

but it is just the probability of S^1_t visits r^n before visits the origin.

For the case of a simple random walk starting at minus one, S_t^{-1}, the probability that it visits $-r^n$ before visiting the origin is also $\frac{1}{r^n}$. So the probability of a simple random walk hitting r^n or $-r^n$ before hitting the origin is $p = \frac{1}{r^n}$.

Let C_i^n be the event that after the i-visit to the origin, the random walk first gets back to the origin before visiting the set $\{-r^n, r^n\}$.

(Recall that ν_i denotes the i-th visit after 0 by S to the origin:

$$\nu_{i+1} := \min\{t > \nu_i | S_t = 0\},$$

whilst $\nu_0 = 0$.) So, C_i^n is the event that

$$\min\{t > \nu_i || S_t| = r^n\} > \min\{t > \nu_i | S_t = 0\}.$$

by the strong Markov property of S, we have that

$$P(C_i^{nc}) = \frac{1}{r^n}. \tag{2.3}$$

But we have that

$$C^n = \bigcap_{i=0}^{(v_1)^n} C_i^n$$

and hence

$$P(C^{nc}) \leq \sum_{i=0}^{(v_1)^n} P(C_i^{nc})$$

The last inequality together with 2.3, yields

$$P(C^{nc}) \leq \sum_{i=0}^{(v_1)^n} \frac{1}{r^n} = \left(\frac{v_1}{r}\right)^n.$$

Note that the constants r and v_1 where defined so that $(v_1/r) < 1$, which implies that the last bound above is exponentially small in n. □

Lemma 2.3. *We have that*

$$P(D_1^n) \geq 1 - 4\left(\frac{r \cdot 2^{H_2(q_0)}}{4}\right)^n.$$

Proof: Let D_{1x}^n denote the event that there is no nearest neighbor walk path R in $\mathcal{P}_x^n$ and generating w^n on ξ. In other words, D_{1x}^n is the event that there is no nearest neighbor walk path $R : [0, n-1] \to \mathbb{Z}$, starting in x with less than q_0-percentage of back-and-forth steps and such that

$$\xi(R(0))\xi(R(1))\xi(R(2)) \ldots \xi(R(n-1)) = w^n.$$

We have that

$$D_1^n = \bigcap_{x \in [-r^n, r^n] - [-2n, 2n]} D_{1x}^n$$

so that

$$P(D_1^{nc}) \leq \sum_{x \in [-r^n, r^n] - [-2n, 2n]} P(D_{1x}^{nc}). \tag{2.4}$$

Then, if R starts in x (that is $R(0) = x$), with $x \notin [-2n, 2n]$ and since the nearest neighbor walk path moving at most one unit by step, it follows that R can not enter the interval $[0, n-1]$. Assuming that R is non-random, we then obtain that the observation generated by R, that is the string

$$\xi(R_0)\xi(R_1) \ldots \xi(R_{n-1})$$

is independent of ξ restricted to $[0, n-1]$. This is because the scenery ξ is i.i.d. In other words, we obtain that w^n is independent of $\xi(R_0)\dots\xi(R_{n-1})$. Since, we have 4 equiprobable colors in w^n, this leads to

$$P\left(w^n \;=\; \xi(R_0)\xi(R_1)\dots\xi(R_{n-1})\;\right) = \left(\frac{1}{4}\right)^n, \tag{2.5}$$

for any non-random $R \in \mathcal{P}_x^n$ as soon as $x \notin [-2n, 2n]$. Now,

$$D_{1x}^n = \cap_{R\in\mathcal{P}_x^n}\{w^n \neq \xi(R_0)\xi(R_1)\dots\xi(R_{n-1})\}$$

so that

$$P(D_{1x}^{nc}) \leq \sum_{R\in\mathcal{P}_x^n} P\left(\; w^n = \xi(R_0)\dots\xi(R_{n-1})\;\right).$$

Applying now 2.5 to the last inequality above yields

$$P(D_{1x}^{nc}) \leq \sum_{R\in\mathcal{P}_x^n}\left(\frac{1}{4}\right)^n, \tag{2.6}$$

when $x \notin [-2n, 2n]$.

Since the number of different sequences with length n and proportion q_0 of back-and-forth steps is

$$\begin{pmatrix} n-2 \\ q_0(n-2) \end{pmatrix},$$

Using inequality 2.18 from the appendix, we get that there are less than $2^{H_2(q_0)(n-2)}$ elements in the set $\mathcal{P}_x^n$. Thus 2.6 can be written as:

$$P(D_{1x}^{nc}) \leq \frac{2^{H_2(q_0)(n-2)}}{4^n}$$

for all $x \in [r^n, r^n] - [-2n, 2n]$. Applying the last equation above to inequality 2.4, we obtain

$$P(D_1^{nc}) \leq \sum_{x\in[-r^n,r^n]-[-2n,2n]} \frac{2^{H_2(q_0)(n-2)}}{4^n}.$$

Since in the set $[-r^n, r^n] - [-2n, 2n]$ there are less than $2r^n$ elements, we find

$$P(D_1^{nc}) \leq 2\left(\frac{r}{4}\right)^n 2^{H_2(q_0)(n-2)}.$$

The expression on the right side of the last equation above is an exponential negative bound, since by inequality 2.1, we have

$$\frac{r\cdot 2^{H_2(q_0)}}{4} \;<\; 1.$$

□

Lemma 2.4. *We have that*

$$P(D_2^n) \geq 1 - \frac{n}{4}\left(\frac{2^{H_2(q_0)}}{4}\right)^{(n-2)}.$$

Proof: Let $R : [0, n-1] \mapsto \mathbb{Z}$ be a (non-random) nearest neighbor path ending outside $[0, n-1]$. Assume first that $R(n-1) > n-1$. Then, since each step R travels no more than one unit, we get that $n-1-i < R(n-1-i)$ for all $i \in [0, n-1]$. Hence, since the scenery ξ is i.i.d., we find that $\xi_{n-1-i} = w_{n-1-i}$ is independent of

$$\xi(R_{n-1-i})\xi(R_{n-1-i+1})\dots\xi(R_{n-1}) \tag{2.7}$$

Let Z_i be the Bernoulli variable which is equal to 1 if $\xi(R_{n-1-i}) = w_{n-1-i}$ and $Z_i = 0$ otherwise. Because of the independence of expression 2.7, we get

$$P(Z_i = 1|Z_{i-1}Z_{i-2}\dots Z_0) = 1/4.$$

It follows that the variables $Z_0Z_1\dots Z_n$ are i.i.d. so that

$$P(Z_0 = 1, Z_1 = 1, \dots, Z_n = 1) \;=\; \left(\frac{1}{4}\right)^n$$

But having all the Z_i's equal to 1 for $i = 1, 2, \dots, n$ is the same as saying that R generates the word w^n on the scenery ξ. Hence,

$$P(w^n = \xi(R_0)\xi(R_1)\dots\xi(R_n)) = \left(\frac{1}{4}\right)^n. \tag{2.8}$$

The last inequality was obtained assuming $R(n-1) > n-1$. The same inequality can be obtained for when $R(n-1) < 0$ and so inequality 2.8 holds for all nearest neighbor paths not ending in $[0, n-1]$. Now, the event B_2^n is the event that there exists no nearest neighbor walk path $R \in \mathcal{P}_x^n$, with $x \in [-2n, 2n]$ and generating w^n on ξ whilst ending outside $[0, n-1]$. Hence

$$B_2^n = \bigcap_R \{\, w^n \neq \xi(R_0)\xi(R_1)\dots\xi(R_n)\,\}, \tag{2.9}$$

where the intersection is taken over all R in

$$\bigcup_{x\in[-2n,2n]} \mathcal{P}_x^n \tag{2.10}$$

ending outside $[0, n-1]$, i.e. such that $R(n-1) \notin [0, n-1]$. For those paths ending outside $[0, n-1]$, equation 2.8 applies. We can use this in conjunction with equation 2.9, (since in equation 2.9 all paths considered end outside $[0, n-1]$). We obtain:

$$P(B_2^{nc}) \le \sum_R P(w^n = \xi(R_0)\xi(R_1)\dots\xi(R_n)) = \sum_R \left(\frac{1}{4}\right)^n. \tag{2.11}$$

where in the last sums above, R is taken over the set 2.10 and such that $R(n-1) \notin [0, n-1]$. The set $\mathcal{P}_x^n$ for given x contains less than $2^{(n-2)H_2(q_0)}$ elements. So the set 2.10 contains less than $4n2^{(n-2)H_2(q_0)}$ elements. Applying this to inequality 2.11, we get:

$$P(B_2^{nc}) \le 4n\frac{2^{(n-2)H_2(q_0)}}{4^n}.$$

Note that the bound on the last inequality above is exponentially small in n since by 2.1, we have $(2^{H_2(q_0)}/4) < 0.5$ □

Lemma 2.5. *We have that*

$$P(E^{nc}) \le \frac{4n}{(v_1/2)^n},$$

for all n large enough.

Proof: As before, let ν_i denote the i-th visit by the random walk to the origin and define the following sequence. Let

$$k_1 = \nu_1$$

and, for $i \ge 1$, let

$$k_{i+1} := \min\{\nu_j \ge k_i + n : j \in \mathbb{N}\},$$

The sequence of k_i's denotes a set of visits by S to the origin, such that two consecutive visits are separated by at least n steps.

Let Y_i be a Bernoulli variable, where $Y_i = 1$ if after time k_i, S takes $n-1$ steps to the right and $Y_i = 0$ otherwise. Hence $Y_i = 1$ if and only if

$$S_{j+1} - S_j = 1 \ , \ \forall j \in [k_i, k_i + n - 1].$$

The variables $Y_1, Y_2, \dots$ are $i.i.d$ with

$$p = P(Y_i = 1) = \left(\frac{1}{2}\right)^{n-1}.$$

Note that among the first $(v_1)^n$ visits ν_i to the origin, there are at least $(v_1)^n/n$ visits k_i, and hence:

$$\left\{ \sum_{i=1}^{v_1^n/n} Y_i \geq (v_2)^n \right\} \subseteq E^n.$$

From the last inequality above it follows that

$$P(E^{nc}) \leq P\Big(\sum_{i=1}^{v_1^n/n} Y_i < (v_2)^n \Big) \tag{2.12}$$

At this point we simply use the Chebycheff inequality. Set

$$Z := \sum_{i=1}^{v_1^n/n} Y_i$$

so that

$$E[Z] = \frac{(v_1)^n}{n} E[Y_1] = \frac{(v_1)^n}{n} \left(\frac{1}{2}\right)^n = (1/n) \left(\frac{v_1}{2}\right)^n$$

and

$$VAR[Z] = \frac{(v_1)^n}{n} VAR[Y_1] = \frac{v_1^n}{n} \left(\frac{1}{2}\right)^n \left(1 - \frac{1}{2^n}\right) \leq \frac{(v_1)^n}{n} \left(\frac{1}{2}\right)^n. \tag{2.13}$$

The constants v_1 and v_2 do not depend on n and satisfy inequality 2.1, so that $v_1/2 > v_2$. It follows that for n large enough, $E[Z] = (v_1/2)^n/n$ is much larger than $(v_2)^n$. So for n large enough, we have

$$\left| (1/n) \left(\frac{v_1}{2}\right)^n - (v_2)^n \right| \geq (1/2n) \left(\frac{v_1}{2}\right)^n$$

and hence

$$|E[Z] - (v_2)^n| \geq (1/2n) \left(\frac{v_1}{2}\right)^n. \tag{2.14}$$

Applying now Chebycheff's inequality to 2.12, we obtain

$$P(E^{nc}) \leq \frac{VAR[Z]}{(E[Z] - v_2^n)^2}.$$

Applying equation 2.13 and inequality 2.14 to the last inequality above we find

$$P(E^{nc}) \leq \frac{(v_1/2)^n}{(v_1/2)^{2n}} \frac{4n^2}{n} = \frac{4n}{(v_1/2)^n}. \tag{2.15}$$

Since by inequality 2.1 we have $v_1/2 > 1$ it follows that the bound on the right side of inequality 2.15, is an exponentially small quantity in n. □

Lemma 2.6. *We have that*

$$P(F^{nc}) \leq (c_F)^n,$$

where $0 < c_F < 1$ *does not depend on* n*.*

Proof: For any integer $z \in [1, n-1]$, define the event

$$A_z = \{\xi(z+1) = \xi(z-1)\}.$$

Since the scenery-process $\{\xi\}_{z\in\mathbb{Z}}$ is a sequence of $i.i.d$ random variables with uniform probability on a set of 4 colors, we get: $P(A_z) = 4\left(\frac{1}{16}\right) = \frac{1}{4}$.

Let X_z be the Bernoulli variable, such that $X_z = 1$ iff A_z holds. Note that any sequence of colors with size n has exactly $(n-2)$ possible pairs of positions for A_z to occur, so that

$$\begin{aligned} P(F^{nc}) &= P(X_1 + X_2 + \cdots + X_{n-2} > (n-2)q_0) \\ &\leq P((X_1 - q_0) + \cdots + (X_{n-2} - q_0) \geq 0) \\ &\leq E(e^{Y_1 t})^{n-2}, \end{aligned}$$

where $Y_1 = (X_1 - q_0)$. Here we use the same argument as in the proof of lemma 2.5. That is we use that any random variable Z and for any $t > 0$, $P(Z \geq 0) \leq E[e^{Zt}]$.

Since $q_0 > 0.25$, it follows that

$$E(Y_1) = E[X_1] - q_0 = 0.25 - q_0 < 0.$$

Hence, there exists a $t_0 \geq 0$ such that $E(e^{Y_1 t_0}) < 1$. Call this value c_F:

$$c_F := E(e^{Y_1 t_0}) < 1.$$

Thus we have an upper bound for $P(F^{nc})$ which decreases exponentially fast to zero:

$$P(F^{nc}) \leq c_F^n.$$

□

2.3. *Proof of main theorem.* In lemma 2.1, we prove that

$$C^n \cap D_1^n \cap D_2^n \cap E^n \cap F^n \subset B^n$$

It follows that

$$P(B^{nc}) \leq P(C^{nc}) + P(D_1^{nc}) + P(D_2^{nc}) + P(E^{nc}) + P(F^{nc}). \tag{2.16}$$

In Subsection 2.2, we get upper bounds for each of the probabilities $P(C^{nc})$, $P(D_1^{nc})$, $P(D_2^{nc})$, $P(E^{nc})$ and $P(F^{nc})$ that are negative exponentially small in n. Hence, together with inequality 2.16, this implies that $P(B^{nc})$ is also exponentially small in n. Hence, there exists a constant $c_B > 0$ not depending on n such that for all n, we have

$$P(B^{nc}) \leq e^{-c_B n}.$$

2.4. *Appendix.* This appendix contains a proof of an inequality that is used for bounding the number of paths in $\mathcal{P}_x^n$. This bound is required in Lemma 2.3.

Let q be a value in the open interval $(0, 1)$. Assume that n is an integer such that nq is an integer. Let Z be a binomial variable with parameter n and q. The probability that Z is equal to nq is less than 1. Hence:

$$P(Z = nq) = \binom{n}{nq} q^{qn}(1-q)^{(1-q)n} \leq 1.$$

Dividing on both sides of the last inequality above by $q^{qn}(1-q)^{(1-q)n}$, we obtain:

$$\binom{n}{nq} \leq \left(\left(\frac{1}{q}\right)^q \left(\frac{1}{1-q}\right)^{1-q}\right)^n = 2^{n \cdot H_2(q)}. \tag{2.17}$$

Next assume that $q < 0.5$. We are going to show an improved version of inequality 2.17. That is we are going to prove that

$$\sum_{i=1}^{nq} \binom{n}{i} \leq 2^{n \cdot H_2(q)}. \tag{2.18}$$

For this note that

$$1 \geq P(Z \leq qn) = \sum_{i=1}^{nq} \binom{n}{i} q^i (1-q)^{n-i}. \tag{2.19}$$

But

$$\sum_{i=1}^{nq} \binom{n}{i} q^i (1-q)^{n-i} = \sum_{i=1}^{nq} \binom{n}{i} 2^{-H_2(i/n)n}. \tag{2.20}$$

Now note that from $q = 0$ up to $q = 0.5$ the entropy function $q \mapsto H_2(q)$ is increasing. Hence $-H_2(i/n) \geq -H_2(q)$ for all $i \leq nq$. Hence,

$$\sum_{i=1}^{nq} \binom{n}{i} 2^{-H_2(i/n)n} \geq 2^{-H_2(q)n} \sum_{i=1}^{nq} \binom{n}{i}.$$

Combining the last inequality above with inequalities 2.19 and 2.20, we obtain

$$1 \geq 2^{-H_2(q)n} \sum_{i=1}^{nq} \binom{n}{i},$$

wish completes the proof of 2.18.

References

A. Hart, F. Machado and H. Matzinger. Information recovery from observations by a random walk having jump distribution with exponential tails. *ArXiv Mathematics e-prints* (2011). http://arxiv.org/abs/1110.6853v1.

A. Hart and H. Matzinger. Markers for error-corrupted observations. *Stochastic Process. Appl.* **116** (5), 807–829 (2006). MR2218337.

H. Kesten. Detecting a single defect in a scenery by observing the scenery along a random walk path. In *Itô's stochastic calculus and probability theory*, pages 171–183. Springer, Tokyo (1996). MR1439524.

J. Lember and H. Matzinger. Reconstructing a 2-color scenery by observing it along a recurrent random walk path with bounded jumps (2003). Eurandom. In preparation.

J. Lember and H. Matzinger. Information recovery from a randomly mixed up message-text. *Electron. J. Probab.* **13**, no. 15, 396–466 (2008). MR2386738.

M. Löwe and H. Matzinger. Scenery reconstruction in two dimensions with many colors. *Ann. Appl. Probab.* **12** (4), 1322–1347 (2002). MR1936595.

M. Löwe, H. Matzinger and F. Merkl. Reconstructing a multicolor random scenery seen along a random walk path with bounded jumps. *Electron. J. Probab.* **9**, no. 15, 436–507 (electronic) (2004). MR2080606.

H. Matzinger. *Reconstructing a 2-color scenery by observing it along a simple random walk path with holding.* Ph.D. thesis, Cornell University (1999).

H. Matzinger. Reconstructing a two-color scenery by observing it along a simple random walk path. *Ann. Appl. Probab.* **15** (1B), 778–819 (2005). MR2114990.

H. Matzinger and S. Popov. Detecting a local perturbation in a continuous scenery. *Electron. J. Probab.* **12**, no. 22, 637–660 (2007). MR2318405.

H. Matzinger and S. W. W. Rolles. Reconstructing a piece of scenery with polynomially many observations. *Stochastic Process. Appl.* **107** (2), 289–300 (2003a). MR1999792.

H. Matzinger and S. W. W. Rolles. Reconstructing a random scenery observed with random errors along a random walk path. *Probab. Theory Related Fields* **125** (4), 539–577 (2003b). MR1974414.

H. Matzinger and S. W. W. Rolles. Finding blocks and other patterns in a random coloring of $\mathbb{Z}$. *Random Structures Algorithms* **28** (1), 37–75 (2006a). MR2187482.

H. Matzinger and S. W. W. Rolles. Retrieving random media. *Probab. Theory Related Fields* **136** (3), 469–507 (2006b). MR2257132.