Scenery Reconstruction

Angélica Pachón
joint with Serguei Popov and Heinrich Matzinger

University of Torino, Italy
Campinas University, Brazil and Georgia Institute of Technology, USA

Young Women in Probability
Bonn 2014

Scenery Reconstruction Problem

The basic **scenery reconstruction problem** can be described as follows:

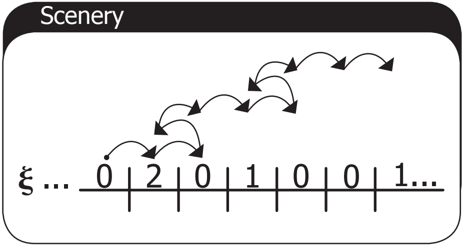
1. A **scenery** is a coloring of the integers with a color of the set $C = \{0, 1, \dots, c-1\}$, i.e,

$$\xi : \mathbb{Z} \to \{0, 1, \dots, c-1\}$$

2. A simple random walk $\{S_t\}_{t \in \mathbb{N}}$ starts to move on these colored integers registering the color it sees at each time t producing a new sequence of colors $\{\chi_t\}_{t \geq 0}$, i.e,

$$\chi_t = \xi(S_t).$$

Example:



so, following the path above on ξ we obtain the sequence of observations χ

0	2	0	2	0	1	0	1	0	0	1

Scenery Reconstruction Problem

The interesting question is:

Can the original coloring of the integer numbers be reconstructed from the sequence produced by the process?

The answer in general is NO. However, under appropriate restrictions, the answer will become YES.

Let us explain these restrictions:

Not every scenery can be reconstructed

1. E. Lindenstrauss (1999), shows that there exist **sceneries** which can not be reconstructed.
2. Then, the **scenery** is taken to be the outcome of a random process and one tries to show that almost all sceneries can be reconstructed.

The reconstruction works up to equivalence

1. Two **sceneries** are called equivalent if one of them is obtained from the other by a translation, reflection or rotation. We denoted that by $\varepsilon \sim \varepsilon'$
2. The reconstruction works in the best case only almost surely up to shift and/or reflection.

The evolution of scenery reconstruction

- H. Matzinger (1999) shows that typical 2-color **sceneries** can be reconstructed almost surely up to equivalence.
- H. Kesten (1999) noticed that Matzinger's result depends heavily on the one dimensionality of the problem and the skip-freeness of the random walk.
- The development of the theory of scenery reconstruction took place in three phases.

  1. Combinatorial case: Sceneries with two or more colors along a simple random walk path
  2. Semi-combinatorial case: 2-color scenery along a simple random walk with holding
  3. Purely statistical case: 2-color scenery along a random walk with jumps

Results in $\mathbb{Z}^d$, $d \geq 2$

- M. Löwe and H. Matzinger (2002) shows the reconstruction of a 2-dimensional scenery seen along a random walk path with many colors.

- S. Popov and A. P. (2010) study the scenery reconstruction problem in d-dimension using observations registered in boxes of size k (for k fixed), seen along a branching random walk. In other words, they suppose that the particles see at each time a window around their current position.

Definition of the model in $\mathbb{Z}^d$ with a branching random walk

- Let $\{\varepsilon_z\}_{z\in\mathbb{Z}^d}$, be a sequence of i.i.d.r.v, taking values in $\{1, 2 \ldots, C\}$.
- Let ε be a realization of $\{\varepsilon_z\}_{z\in\mathbb{Z}^d}$ obtained with uniform probability, it is thus a (random) coloring of $\mathbb{Z}^d$.
- Consider a branching random walk in $\mathbb{Z}^d$ (BRW), described in the following way: Let it starts with only one particle at the origin at time 0, then at each step, any particle is substituted by two particles with probability b and is left intact with probability $1 - b$, for some fixed $b \in (0, 1)$. Furthermore, each particle follows the path of a simple random walk.
- Let χ be the observed sequence of colors $\{\chi_t\}_{t\geq 0}$, i.e, $\chi_t = \xi(S(v_i^t))$, where $S(v_i^t)$ is the position of the i-*th* particle at time t, so $v_i^t \in \mathbb{Z}^d$.

Using a branching random walk with boxes: S.Popov, A. P.

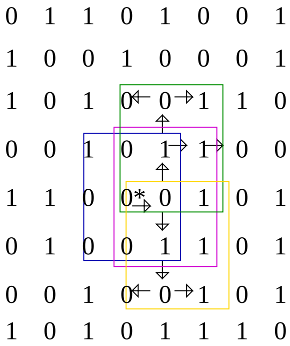


Figure: A branching random walk on a given scenery (0* is the origin). At each step, any particle is substituted by two particles with probability b and is left intact with probability $1 - b$, for some fixed $b \in (0, 1)$.

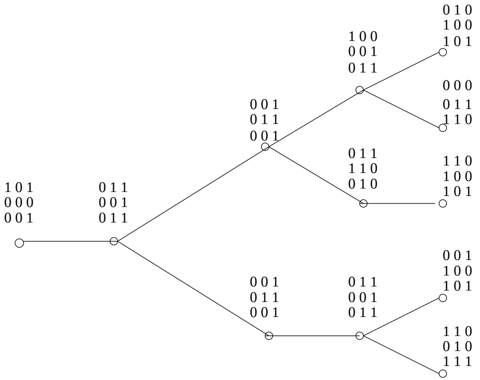


Figure: χ the observed sequence of colors is a tree with colored boxes observed at each instant of time.

Without windows

- $\Omega_1 = \{(\eta_t)_{t\in\mathbb{N}}\}$ be the space of all possible "trajectories" of the branching random walk, where $\eta_n(z)$ be the number of particles in z at time t, $z \in \mathbb{Z}^d$.
- $\Omega_2 = \{1, \ldots, C\}^{\mathbb{Z}^d}$ be the space of all possibles sceneries ξ, and
- $\Omega_3 = \{(\chi_t)_{t\in\mathbb{N}}\}$ be the space of all possibles sceneries looked χ.

Theorem (H.Matzinger, S.Popov and A.P.)

For any $b \in (0,1)$ and $C > 2d$ there exists a measurable function $\Lambda : \Omega_3 \to \Omega_2$ such that

$$P(\Lambda(\chi) \sim \xi) = 1,$$

where P is the product measure defined on the space $\Omega_1 \otimes \Omega_2$

Algorithm Λ^n: reconstructing a finite piece of scenery

- **First phase:** Construct the set of $SHORTWORDS^n$ (words of length $(\ln n)^2$) such that

$$W(\xi_{\mathcal{R}(4n)}) \subseteq SHORTWORDS^n \subseteq W(\xi_{\mathcal{R}(n^2)}).$$

- **Second phase:** Construct the set of $LONGTWORDS^n$ (words of length $4n$), i.e., assemble the words of $SHORTWORDS^n$ into longer words of size $4n$. **Rule:** The words of $SHORTWORDS^n$ to get assembled must coincide on $(\ln n)^2 - 1$ consecutive letters.
- **Third phase:** Select a seed word close to the origin, i.e., select from the previous long words one which is close to the origin. For that we apply the algorithm as described before, but with the parameter being equal to $n^{1/4}$ instead of n.
- **Forth phase:** Determining which long words are neighbors of each other.

Main ideas of the proof: the DNA sequence trick

We use the same trick as is used in modern method for DNA-sequencing.

- First determine small pieces of order at least $(\ln n)^2$ and then assembled them to obtain the whole sequence.
  Example

ξ	...	4	1	4	2	1	3	4	...

- Assume that the set of all pieces of length 5 (and their reverses) is given in one bag without extra information, i.e.,

$$41421, 14213, 42134 \quad \text{and} \quad 12414, 31241, 43124$$

- then, we could reconstruct $\ldots\xi\ldots$ by assembling these pieces using the rule that two pieces must coincide on a sub-piece of length 4.

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, 1241, \mathbf{3124}, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.

     Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad 12414, \mathbf{31241}, 43124$$

$$\begin{array}{cccccc} & 3 & 1 & 2 & 4 & 1 \\ \hline 4 & 3 & 1 & 2 & 4 & 1 \end{array}$$

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, 1241, 3124, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.
     Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad 12414, \mathbf{31241}, 43124$$

$$\underline{4 \quad 3 \quad 1 \quad 2 \quad 4 \quad 1 \qquad\quad}$$

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, \mathbf{1241}, 3124, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.
  Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad \mathbf{12414}, 31241, 43124$$

4	3	1	2	4	1	
		1	2	4	1	4

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, 1241, 3124, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.
     Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad 12414, 31241, 43124$$

$$\underline{\;4 \quad 3 \quad 1 \quad 2 \quad 4 \quad 1 \quad 4\;\;\;\;}$$

- We will need the bag of all strings of length $(\ln n)^2$ restricted to a portion of the scenery ξ, $\xi_{\mathcal{K}(4n)}$,
- and we will need that all the substrings of length $(\ln n)^2 - 1$ appear only once in $\xi_{\mathcal{K}(4n)}$.

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, 1241, 3124, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.
     Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad 12414, 31241, 43124$$

$$\underline{4 \quad 3 \quad 1 \quad 2 \quad 4 \quad 1 \quad 4}$$

- We will need the bag of all strings of length $(\ln n)^2$ restricted to a portion of the scenery ξ, $\xi_{\mathcal{K}(4n)}$,
- and we will need that all the substrings of length $(\ln n)^2 - 1$ appear only once in $\xi_{\mathcal{K}(4n)}$.

How?

- Consider the set of all substrings of length 4 (and their reverses), i.e,

$$4142, 1421, 4213, 2134 \quad \text{and} \quad 2414, 1241, 3124, 4312$$

- Given the bag of pieces of length 5 we assemble them one after other:
  1. We start by picking any word.
  2. Then, we put down the next word from our bag which coincides with the previous one on at least 4 colors.
     Example: Given

$$41421, 14213, 42134 \quad \text{and} \quad 12414, 31241, 43124$$

$$\underline{\;4 \quad 3 \quad 1 \quad 2 \quad 4 \quad 1 \quad 4\;}$$

- We will need the bag of all strings of length $(\ln n)^2$ restricted to a portion of the scenery ξ, $\xi_{\mathcal{K}(4n)}$,
- and we will need that all the substrings of length $(\ln n)^2 - 1$ appear only once in $\xi_{\mathcal{K}(4n)}$.

Reconstruction of short sequences

How does one manage to reconstruct short sequences (of length $(\ln n)^2$) given only observations?

- Assume for a moment that the scenery ξ is nonrandom, and is a four-color scenery $\{1, 2, 3, 4\}$.
- Let us imagine furthermore, that there are two integers x, y such that $\xi(x) = 2$ and $\xi(y) = 3$, but outside from x and y the scenery has everywhere color 1 or 4.

Example: Let a portion of the scenery ξ be such that

z	...	-2	-1	0	1	2	3	4	...
ξ		1	**2**	1	4	1	**3**	4	

where, $x = -1$ and $y = 3$.

- Take the shortest time of the random walk between x and y.

The diamond trick to reconstruct all the words of $\xi_{(4n)}$

The basic idea is to use the diamond associated with a word appearing in the scenery

- Take for example the following piece of two dimensional scenery which would be the restriction of ξ to the $[0,6]*[0,4]$:

2	1	9	4	3	7	4
7	5	0	7	6	1	1
7	4	3	9	1	2	1
8	6	4	4	0	4	3
2	2	7	8	0	3	9

  that is the diamond associated with the word $w :=$43912, let it noted by D_w.

- We will use that with high probability for a given non-random point z **outside** the diamond associated with w, **there is no** a nearest neighbor walk path starting (or ending) at z and generating w.

How can we use this to reconstruct short words?

- For this let the restriction of the scenery ξ to $[0,16]*[0,4]$: be equal

2	1	9	4	3	7	4	1	2	5	2	2	7	8	0	6	9
7	5	0	7	6	1	1	8	2	5	8	6	7	4	0	4	2
7	4	3	9	1	2	1	7	8	4	7	6	1	7	7	7	4
8	6	4	4	0	4	3	5	3	6	7	5	1	9	9	9	1
2	2	7	8	0	3	9	4	3	7	2	1	9	4	5	7	0

- Let $w_1 = 43912$, $w_3 = 61777$ and $w_2 = 17847$.
- Note that there is only one shortest nearest neighbor path to go from D_1 to D_3 (walking straight). When doing so a walker will see the word w_2.

Conclusions and Open Problems

Conclusions

Not only the techniques for solving the reconstruction problem change with the number of colors, when the random walk is not skip-free and when the 2-dimension space is taking. These also change in d dimensions when windows of observations or just one observation are taken.

Open Problems

- Which are the fundamental properties of the distribution of the scenery for a plausible reconstruction.
- Is it possible to find an universal reconstruction algorithm for solving the scenery reconstruction problem?