Matching random colored points with rectangles

Josué Corujo[1], David Flores-Peñaloza[2⋆], Clemens Huemer[3⋆⋆], Pablo Pérez-Lantero[4⋆⋆⋆], and Carlos Seara[5†]

[1] Université Paris-Dauphine, Paris, France. Facultad de Matemática y Computación, Universidad de La Habana, Cuba, corujo@ceremade.dauphine.fr

[2] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, México, dflorespenaloza@gmail.com

[3] Departament de Matemàtiques, Universitat Politècnica de Catalunya, Spain, clemens.huemer@upc.edu

[4] Departamento de Matemática y Ciencia de la Computación, USACH, Chile, pablo.perez.l@usach.cl

[5] Departament de Matemàtiques, Universitat Politècnica de Catalunya, Spain, carlos.seara@upc.edu

Abstract. Let $S \subset [0,1]^2$ be a set of n points, randomly and uniformly selected. Let $R \cup B$ be a random partition, or coloring, of S in which each point of S is included in R uniformly at random with probability $1/2$. We study the random variable $M(n)$ equal to the number of points of S that are covered by the rectangles of a maximum strong matching of S with axis-aligned rectangles. The matching consists of closed rectangles that cover exactly two points of S of the same color. A matching is strong if all its rectangles are pairwise disjoint. We prove that almost surely $M(n) \geq 0.83\,n$ for n large enough. Our approach is based on modeling a deterministic greedy matching algorithm, that runs over the random point set, as a Markov chain.

1 Introduction

Given a point set $S \subset \mathbb{R}^2$ of n points, and a class $\mathcal{C}$ of geometric objects, a *geometric matching* of S is a set $M \subseteq \mathcal{C}$ such that each element of M contains exactly two points of S and every point of S lies in at most one element of M. A geometric matching is *strong* if the geometric objects are pairwise disjoint,

⋆ Research supported by project PAPIIT IN117317 (UNAM, Mexico).

⋆⋆ Research supported by projects MTM2015-63791-R (MINECO/FEDER) and Gen. Cat. DGR 2017SGR1336.

⋆⋆⋆ Partially supported by projects CONICYT FONDECYT/Regular 1160543 (Chile), DICYT 041933PL Vicerrectoría de Investigación, Desarrollo e Innovación USACH (Chile), and Programa Regional STICAMSUD 19-STIC-02.

† Research supported by projects MTM2015-63791-R MINECO/FEDER and Gen. Cat. DGR 2017SGR1640.

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 734922.

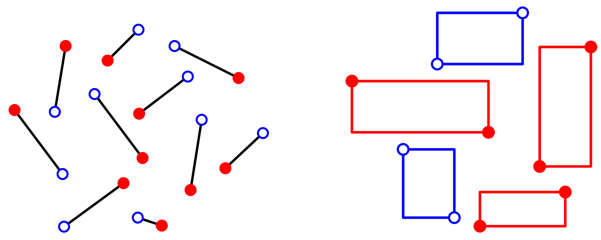


Fig. 1: (a) A perfect, strong bichromatic matching of 10 red points and 10 blue points with segments. (b) A perfect, strong monochromatic matching of 6 red points and 4 blue points with axis-aligned rectangles.

and *perfect* if every point of S belongs to (or is covered by) some element of M. This type of geometric matching problems was considered by Ábrego et al. [1], who studied the existence and properties of matchings for point sets in the plane when $\mathcal{C}$ is the class of axis-aligned squares, or the class of disks.

Let $S = R \cup B \subset \mathbb{R}^2$ be a set of n colored points in the plane, each point colored red or blue, where R and B are the sets of red and blue points, respectively. A geometric matching of S is called *monochromatic* if all matching objects cover points of the same color, and *bichromatic* if all matching objects cover points of different colors. For example, monochromatic matchings of two-colored point sets in the plane with straight segments have been studied [4,5]. In the case of bichromatic matchings with straight segments, a classical result in discrete geometry asserts that for any planar point set S consisting of n red points and n blue points in general position (i.e., no three points of S are collinear) there exists a perfect, strong bichromatic matching of S with straight segments [6] (see Figure 1a).

In this paper, we consider strong monochromatic matchings with axis-aligned rectangles. Refer to Figure 1b for an example of a perfect matching of this type. Throughout the paper, every rectangle will be considered axis-aligned and a closed subset of the plane.

Caraballo et al. [2] studied both monochromatic and bichromatic strong matchings of S with rectangles from the algorithmic point of view. That is, they studied two combinatorial optimization problems for given $S = R \cup B$: find a monochromatic strong matching of S with the maximum number of rectangles, and find a bichromatic strong matching of S with the maximum number of rectangles; proving that both problems are NP-hard and giving a polynomial-time 4-approximation algorithm in each case. As noted by Caraballo et al., these two problems are special cases of the Maximum Independent Set of Rectangles problem (MISR): Given a finite set $\mathcal{R}$ of rectangles in the plane, find a subset $\mathcal{R}' \subseteq \mathcal{R}$ of maximum cardinality, denoted $\alpha(\mathcal{R})$, such that every pair of rectangles in $\mathcal{R}'$ are disjoint.

Indeed, suppose that we want to find a monochromatic matching of S with the maximum number of rectangles. For every distinct $p, q \in \mathbb{R}^2$, let $D(p,q)$ denote the minimum axis-aligned rectangle (i.e., the rectangle such that both dimensions are minimum) that encloses p and q. Let $\mathcal{R}(S)$ be the set of all rectangles $D(p,q)$ such that $p, q \in S$, p and q have the same color, and $D(p,q)$ contains no points of S different from p and q. Finding a monochromatic strong matching of S with the maximum number of rectangles is equivalent to finding in $\mathcal{R}(S)$ a maximum subset of pairwise disjoint rectangles, whose size is $\alpha(\mathcal{R}(S))$, that is, solving the MISR problem in $\mathcal{R}(S)$.

In this paper, we study monochromatic strong matchings of S with rectangles from the combinatorial point of view. From this point forward, every rectangle will cover precisely two points of S. Point sets $S = R \cup B$ exist in which no matching rectangle is possible (e.g., S is a color-alternating sequence of points on the line $y = x$), and point sets in which a perfect strong matching with rectangles exists (e.g., an even number of red points in the negative part of the line $y = x$, and an even number of blue points in the positive part). These two extreme cases show that it is not worth studying the number $\alpha(\mathcal{R}(S))$ for fixed, or given, colored point sets S. This does not happen, for example, for monochromatic strong matchings with segments for fixed sets of red and blue points: Dumitrescu and Kaye [4] proved that every two-colored point set $S = R \cup B$ of n points in general position admits a strong matching with segments that covers at least $\frac{6}{7}n - O(1)$ of the points; furthermore, there exist n-point sets such that every strong matching with segments covers at most $\frac{94}{95}n + O(1)$ points. Instead, we want to study $\alpha(\mathcal{R}(S))$ when S is a random point set in the square $[0,1]^2$, in which the positions of the n points of S are random and the color of each point of S is also random. Formally:

Let $n > 0$ be an integer, and let $S \subset [0,1]^2$ be a set of n points, randomly and uniformly selected. Let $R \cup B$ be a random partition (i.e., coloring) of S in which each point of S is included in R uniformly at random with probability $1/2$. We study the random variable $M(n) = 2 \cdot \alpha(\mathcal{R}(S))$ equal to the number of points of S that are covered by the rectangles of a maximum monochromatic strong matching of S with rectangles.

Given a set S of n points, randomly and uniformly selected in the square $[0,1]^2$, Chen et al. [3] studied a similar variable: the random variable $\alpha(D(S))$, where $D(S)$ is the random graph with vertex set S and two points $p, q \in S$ define an edge if and only if $D(p,q) \cap S = \{p, q\}$. Here, $\alpha(D(S))$ denotes the size of a maximum independent set of $D(S)$.

One result of Chen et al. [3, Theorem 1] states that if n tends to infinity, then $\alpha(D(S)) = O(n(\log^2 \log n)/\log n)$ with probability tending to 1. This result implies that if $C(n)$ denotes the number of points of S that are covered by a maximum monochromatic matching of S with rectangles, where the rectangles may overlap (i.e., the matching is not necessarily strong), then $C(n) = n - o(n)$ with probability tending to 1. In fact, let M' be a maximum monochromatic matching of S with rectangles, where M' is not necessarily strong, and let $S' \subset S$ be the points not covered by M'. Note that at least $|S'|/2$ points of S' have the same color, and they form an independent set in the graph $D(S)$. Then, with probability tending to 1, we have that M' covers at least $n - |S'| = n - O(n(\log^2 \log n)/\log n) = n - o(n)$ points.

2 Preliminaries

Since for matching S with rectangles, only the left-to-right and bottom-to-top orders of S are relevant, and since the probability that two points of S are in the same vertical or horizontal line is zero, we consider S equal to the point set $S_\pi = \{(i, \pi(i)) \mid i = 1, 2, \ldots, n\}$, where $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ is a randomly and uniformly selected permutation. This assumption was also done by Chen et al. [3].

We have implemented a `Python` program that, given n, generates a uniform random permutation π, and selects the color of each $p \in S_\pi$ (red or blue) randomly and uniformly with probability $1/2$. The program then runs a deterministic algorithm on $S_\pi = R \cup B$ that greedily finds a maximum independent subset of rectangles in $\mathcal{R}(S_\pi)$. The greedy algorithm iterates the points of S_π from left to right, and for each point p in the iteration, it performs the following action: If p is not matched with any point prior to p in the iteration, it finds (if it exists) the first point q to the right of p such that $D(p, q) \in \mathcal{R}(S_\pi)$ and $D(p, q)$ has empty intersection with all matching rectangles already reported. If q exists, then the algorithm reports $D(p, q)$ as a matching rectangle. In any case, regardless of whether q exists, the algorithm continues the iteration to the next unmatched point p.

For large n, say $n = 10000$, the implemented algorithm reports a matching covering approximately $\frac{97}{100}n$ of the points. In fact, we run the algorithm $N = 100$ times for $n = 10000$, and average the outputs (i.e., the percentage of matched points) and computed the standard deviation. The average output is 0.978 and the standard deviation 0.0022. See in Table 1 the row $k = \infty$, where k is a parameter that will be explained later.

Then, it seems that $M(n) \geq \frac{97}{100}n$ for n large enough and probability close to 1. More formally, using the Central Limit theorem, we have that

$$\left(0.9780 - \frac{0.0022}{\sqrt{N}} z_{0.99}, 1\right] \;\subset\; (0.97, 1]$$

is a 99% confidence interval for the expected value of $M(n)/n$. We denote by $z_{0.99} \approx 2.33$ the real value which satisfies $\text{Prob}(Z \leq z_{0.99}) = 0.99$, for Z a normal random variable with mean 0 and variance 1.

Analyzing the algorithm, when run over the random S_π, seems to be a good approach for obtaining a high lower bound for $M(n)$. One way to analyze the algorithm is to consider a parameterized version of it, with a parameter k, such that each unmatched point p finds its match point q among only the next k points of S_π to the right of p. Let $\mathcal{A}_k$ denote this parameterized algorithm. For further experimental results, see Table 1.

	$n=1000$		$n=10000$	
k	mean	sdev	mean	sdev
1	0.6653	0.0175	0.6673	0.0052
2	0.7948	0.0104	0.7934	0.0036
3	0.8301	0.0097	0.8304	0.0034
4	0.8555	0.0094	0.8562	0.0028
5	0.8727	0.0090	0.8736	0.0026
6	0.8860	0.0087	0.8864	0.0026
7	0.8953	0.0084	0.8962	0.0026
8	0.9031	0.0079	0.9041	0.0025
∞	0.9724	0.0062	0.9780	0.0022

Table 1: The table shows the experimental results obtained when running the greedy matching algorithm for $n \in \{1000, 10000\}$ points, parameterized with $k \in \{1,2,3,\ldots,8\}$, or not parameterized ($k=\infty$). For each combination of n,k, we run the algorithm 100 times, and measured the mean and standard deviation of the ratio between the total number of matched points and n.

In the next two sections, we show how to model (an adaptation of) $\mathcal{A}_k$ as a Markov chain, for any fixed $k \in \{1,2,3,\ldots\}$. Then, we show that the algorithm $\mathcal{A}_3$ almost surely guarantees $M(n) \geq \frac{83}{100}n$, for n large enough, by computing the stationary distribution of the Markov chain and applying the Ergodic theorem. For the theory on Markov chains, refer to Norris [7].

3 The Markov chains

From this point forward, we also consider $S = S_\pi$, and whenever we say point i, for $i \in \{1,2,\ldots,n\}$, or just i when it is clear from the context, we are referring to the point $p_i := (i, \pi(i)) \in S$. Let $\text{color}(i) \in \{R, B\}$ be the color of point i.

Let $k \in \{1,2,3,\ldots\}$ be a constant, and let $\tilde{\mathcal{A}}_k$ be the following adaptation of algorithm $\mathcal{A}_k$, consisting in the next idea:

Suppose that $\mathcal{A}_k$ matches points i and j, with $i < j \leq i+k$, when the iteration of S_π is on point i. Algorithm $\tilde{\mathcal{A}}_k$ iterates S_π from left to right, and will also match i and j but, in contrast with $\mathcal{A}_k$, when the iteration is on j, or on a point to the right of j. Using $\tilde{\mathcal{A}}_k$ instead of $\mathcal{A}_k$, allows us to describe in a more compact way the states of the memory of the algorithm during the iteration of the elements of S_π.

Let $E(j)$ be the data structure associated with point $j \in \{1,2,\ldots,n\}$, that is maintained by $\tilde{\mathcal{A}}_k$ during the iteration of S_π. For any j, let $i = i(j)$ be the smallest element in the set $\{\max(1, j-(k-1)), \ldots, j\}$ such that the point i is not matched, and each point in $\{i+1, \ldots, j\}$ is matched with a point to the left of i or is not yet matched. If i exists, then $E(j)$ consists of the following elements:

– The set $U(j) \subseteq \{i, i+1, \ldots, j\}$ of the points that are not matched, with $i \in U(j)$.
– The set $\mathrm{Rect}(j)$ of the (pairwise disjoint) rectangles that match the points in $\{i+1, \ldots, j\} \setminus U(j)$ with points to the left of i.
– The number $f(j)$ of points of S_π that are matched while the iteration is at point j.

Otherwise, if i does not exist, then $E(j)$ consists of the same three above elements with $U(j) = \emptyset$ and $\mathrm{Rect}(j) = \emptyset$.

For $j = 1$, we have $U(1) = \{1\}$, $\mathrm{Rect}(1) = \emptyset$, and $f(1) = 0$. We show now how to obtain $E(j+1)$ from $E(j)$, for any $j \in \{1, \ldots, n-1\}$.

First, we match points i and $j+1$ if and only if $j+1 \leq i+k$, $\mathrm{color}(i) = \mathrm{color}(j+1)$, and the rectangle $D(p_i, p_{j+1})$ does not overlap any rectangle in $\mathrm{Rect}(j)$.

After that, we match other points in $(U(j) \setminus \{i\}) \cup \{j+1\}$ if and only if i was matched in the previous step, or we have finished with point i. We say that we have *finished* with point i if there do not exist more chances for point i to be matched, which is equivalent to $i + k \leq j + 1$. This final matching procedure consists in running the original algorithm $\mathcal{A}_k$ with input the points $\{i+1, \ldots, j, j+1\}$, but with the extra condition that the algorithm terminates if the current point t on the iteration of $\{i+1, \ldots, j, j+1\}$ from left to right, cannot be matched with any other one to its right (i.e., the matching on points in $(U(j) \setminus \{i\}) \cup \{j+1\}$ is performed by running the original algorithm $\mathcal{A}_k$). This is because t must find its match among the points in $\{j+2, \ldots, t+k\}$, before any matching between points in $\{t+1, \ldots, j+1\}$ occurs.

We set $f(j+1)$ equal to the total number of points matched at iteration j. Obtaining $U(j+1)$ and $\mathrm{Rect}(j+1)$ is straightforward.

Let $j \in \{1, 2, \ldots, n\}$. The *state* of $E(j)$ is a 2-tuple formed by:

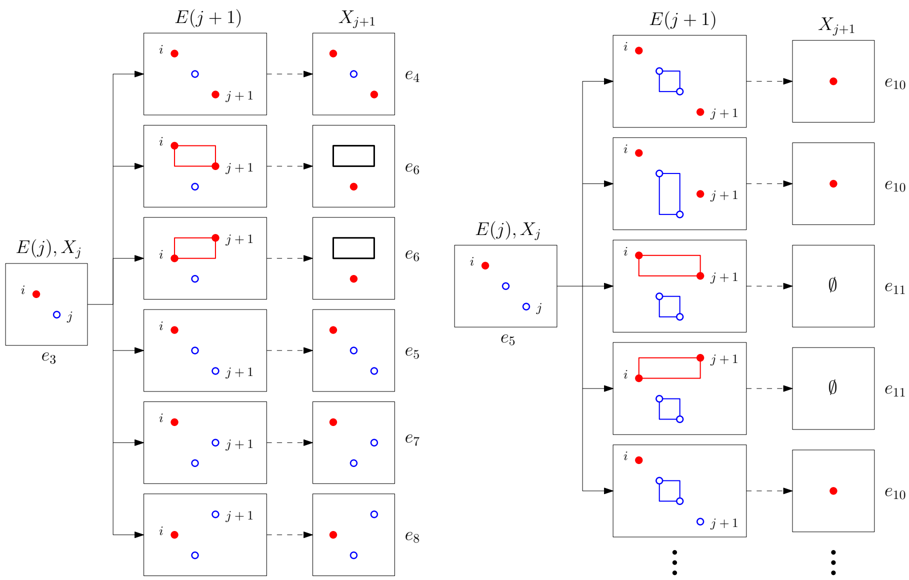
As first component, (a certificate of) the relative positions between the points of $U(j)$ and the rectangles of $\mathrm{Rect}(j)$, together with the color of each point of $U(j)$. If the leftmost point is colored blue, then we switch the color of every point such that the leftmost one is always red.

As second component, the number $f(j)$ of matched points. We say that two states e and e' are *equal* (i.e., $e = e'$) if: (i) the first components are equal, or one first component is symmetric to the other in the vertical direction, and (ii) the second components are equal.

Let $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$ denote the set of all possible states of $E(j)$, which is a finite set, and let $X_j \in \mathcal{E}$ be the random variable equal to the state of $E(j)$. Let $e \in \mathcal{E}$ be a state, and assume that e is the state of $E(j)$ for some j. Let $f(e) = f(j)$ (with abuse of notation), and let $N(e)$ be the *neighborhood* of e, which is the multiset consisting of the state of $E(j+1)$ for every color and every different relative position, with respect to the elements of both $U(j)$ and $\mathrm{Rect}(j)$, of point $j+1$. See for example Figure 2.

Lemma 1. *Let $e, e' \in \mathcal{E}$ be two states. For every $j \geq 2$, we have:*

$$\mathrm{Prob}(X_{j+1} = e' \mid X_j = e) \;=\; \frac{m}{2\left(|U(j)| + 2|\mathrm{Rect}(j)| + 1\right)},$$



(a) (b)

Fig. 2: (a) Example of the data structure $E(j)$, its state $X_j = e_3$, and the states in the neighborhood $N(e_3) = \{e_4, e_5, e_6, e_6, e_7, e_8\}$ corresponding to $E(j+1)$, for each position and color of point $j+1$. Note that $f(e_6) = 2$, and $f(e_i) = 0$ for all $e_i \in \{e_4, e_5, e_7, e_8\}$. (b) Example of $E(j)$ and its state $X_j = e_5$, with $N(e_5) = \{e_{10}, e_{10}, e_{10}, e_{10}, e_{10}, e_{10}, e_{11}, e_{11}\}$. Note that $f(e_{10}) = 2$ and $f(e_{11}) = 4$. For every of the four positions of the blue point $j+1$, the resulting state is e_{10}.

where m is the multiplicity of e' in $N(e)$.

Proof. Through each point of $U(j)$ draw a horizontal line, and for each rectangle of $\mathrm{Rect}(j)$ draw a horizontal line through the top side and a horizontal line through the bottom side. Each of these $K = |U(j)|+2|\mathrm{Rect}(j)|$ lines goes through a different element of S_π, and they subdivide the plane into $K+1$ strips. Since the point $j+1$ is to the right of both every point of $U(j)$ and every rectangle of $\mathrm{Rect}(j)$, the relative position of point $j+1$ with respect to the elements of $U(j)$ and $\mathrm{Rect}(j)$ is to be in one of these strips, and this happens with probability $1/(K+1)$. Furthermore, the color of point $j+1$ is given with probability $\frac{1}{2}$. The lemma follows. □

Note that the value of X_{j+1} depends on the value of X_j, and does not depend in any of the values of $X_1, X_2, \ldots, X_{j-1}$. Formally,

$$\mathrm{Prob}(X_{j+1} = x_{j+1} \mid X_j = x_j, \ldots, X_1 = x_1) = \mathrm{Prob}(X_{j+1} = x_{j+1} \mid X_j = x_j)$$

for all $x_1, \ldots, x_{j+1} \in \mathcal{E}$ such that $\mathrm{Prob}(X_j = x_j, \ldots, X_1 = x_1) > 0$.

Thus, the sequence $(X_n)_{n\geq 1}$ is a Markov chain, denoted $\mathcal{C}_k$, over the set $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$ of states. Let P denote the transition matrix, of dimensions $N \times N$, such that $P_{i,j} = \text{Prob}(X_{\ell+1} = e_j \mid X_\ell = e_i)$. The key observation is that the total number of points matched by the algorithm $\tilde{\mathcal{A}}_k$, denoted $M_k(n)$, is precisely

$$M_k(n) = \sum_{j=1}^{n} f(X_j).$$

A Markov chain is *irreducible* if with positive probability any state can be reached from any other state [7]. We need that this property holds in $\mathcal{C}_k$, as stated in the next lemma.

Lemma 2. *The Markov chain $\mathcal{C}_k$ is irreducible.*

Proof. Assume without loss of generality that e_1 is the state of $E(1)$, which consists of a single red point and ensures $f(e_1) = 0$. Let $e \in \mathcal{E} \setminus \{e_1\}$ be any other state, which by definition is the state of $E(j)$ for some j. Then, in $\mathcal{C}_k$ the state e can be reached from e_1 with positive probability (Lemma 1).

We prove now that also with positive probability, the state e_1 can be reached from e, which implies that $\mathcal{C}_k$ is irreducible. Note that with positive probability, the point $j+1$ may be matched with point $i(j)$ in $E(j+1)$. Then, for some point $t \geq j+1$ we have with positive probability that for $\ell = j+1, \ldots, t$ the points $i(\ell-1)$ and ℓ are matched in $E(\ell)$, and $U(t) = \emptyset$ and $\text{Rect}(t) = \emptyset$.

Let e' denote the state of $E(t)$, and we have $\text{Prob}(X_{t+1} = e_1 \mid X_t = e') = 1$. Hence, the state e_1 can be reached from e with positive probability, and $\mathcal{C}_k$ is thus irreducible. □

Since $\mathcal{C}_k$ is irreducible (Lemma 2) and has a finite set of states, it has a unique stationary distribution $s = (s_1, s_2, \ldots, s_N)$, which is the solution of the system

$$s \ = \ s \cdot P, \quad s_1 + s_2 + \cdots + s_N \ = \ 1$$

of linear equations [7]. Furthermore, since $f(e) \in \left\{0, 2, 4, \ldots, 2\lceil \frac{k+1}{2} \rceil\right\}$ for all $e \in \mathcal{E}$, the function f is bounded and then the Ergodic theorem ensures

$$\lim_{n\to\infty} \frac{M_k(n)}{n} \ = \ \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(X_j) \ = \ \sum_{i=1}^{N} s_i f(e_i),$$

almost surely [7]. Let $\alpha_k = \sum_{i=1}^{N} s_i f(e_i)$. We then arrive to the main result of this paper:

Theorem 1. *Let $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ be a uniform random permutation. Let*

$$S_\pi = \{(i, \pi(i)) \mid i = 1, 2, \ldots, n\}$$

be a random point set, where the color (red or blue) of each point of S_π is selected randomly and uniformly with probability $1/2$. Let $k \in \{1, 2, 3, \ldots\}$ be a constant. For all constant $\varepsilon > 0$ and n large enough, almost surely the number $M_k(n)$ of points of S_π that are matched by the algorithm $\tilde{\mathcal{A}}_k$ satisfies $M_k(n) \geq (\alpha_k - \varepsilon)n$.

4 The Markov chain for $k = 3$

In this section, we consider the algorithm $\tilde{\mathcal{A}}_3$ and give a precise value for α_3. In Table 2, we describe the states, and the transitions between the states, of the Markov chain $\mathcal{C}_3$. The transition matrix P is in Figure 3.

Since $f(e) = 2$ for all $e \in \{e_2, e_6, e_9, e_{10}, e_{16}, e_{17}, e_{18}\}$, $f(e_{11}) = 4$, $f(e) = 0$ for all other state e, and the stationary distribution $s = (s_1, \ldots, s_{18})$ satisfies

$$s_2 = \frac{167959}{816233},\ s_6 = \frac{69640}{816233},\ s_9 = \frac{6800}{816233},\ s_{10} = \frac{58650}{816233},$$
$$s_{11} = \frac{13600}{816233},\ s_{16} = \frac{5950}{816233},\ s_{17} = \frac{1360}{816233},\ s_{18} = \frac{1190}{816233},$$

we obtain

$$\alpha_3 = 2(s_2 + s_6 + s_9 + s_{10} + s_{16} + s_{17} + s_{18}) + 4s_{11} = \frac{677498}{816233} \approx 0.830030151.$$

By Theorem 1, taking $\varepsilon = \alpha_3 - 0.83 > 0$, for n large enough we have almost surely that

$$M(n) \geq M_3(n) \geq 0.83\, n.$$

It can be noted in Table 1 that in practice this lower bound is satisfied. Then, we obtain our second result:

Theorem 2. *Let $n > 0$ be an integer, and let $S \subset [0,1]^2$ be a set of n points, randomly and uniformly selected. Let $R \cup B$ be a random partition (i.e., coloring) of S in which each point of S is included in R uniformly at random with probability $1/2$. For n large enough, almost surely we have that the maximum number $M(n)$ of points of S that are covered by a monochromatic strong matching of S with rectangles satisfies $M(n) \geq 0.83\, n$.*

5 Discussion and open problems

Using the theory of Markov chains, by modeling a deterministic greedy matching algorithm that runs over the random colored n-point set, we have proved that almost surely at least $0.83\, n$ points can be matched with pairwise disjoint axis-aligned rectangles, that is, $M(n) \geq 0.83\, n$. This lower bound was obtained from the stationary distribution of a Markov chain for the parameter $k = 3$. Building the Markov chain for any $k \geq 4$ and computing its stationary distribution will give, by Theorem 1, higher lower bounds, as the results of Table 1 suggest. Experimental results suggest that we may be able to prove the bound $M(n) \geq 0.97\, n$.

The trivial upper bound for $M(n)$ is $M(n) \leq n$. Obtaining tighter lower and upper bounds for $M(n)$ seems to be more challenging. There are cases in which we must consider matching strategies more general than that of the greedy algorithm. Hence, the main open question here is whether $\lim_{n\to\infty} M(n)/n = 1$ or $\lim_{n\to\infty} M(n)/n < 1$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	1/2	1/2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	1/6	1/6	1/3	1/6	1/6	0	0	0	0	0	0	0	0	0	0
4	0	0	0	1/8	1/8	3/8	1/8	0	1/4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	3/4	1/4	0	0	0	0	0	0	0
6	0	1/4	0	0	0	0	0	0	0	0	0	1/8	1/8	1/4	1/4	0	0	0
7	0	0	0	0	0	0	0	0	0	3/4	1/4	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	3/4	0	0	0	0	0	1/4	0	0
9	0	3/10	0	0	0	0	0	0	0	0	0	1/10	0	1/5	1/5	0	1/5	0
10	0	1/2	1/2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1/1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	1/2	3/10	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0
13	0	1/2	3/10	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0
14	0	3/10	1/2	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1/2	3/10	0	0	1/5	0	0	0	0	0	0	0	0	0	0	0	0
16	0	1/5	0	0	0	0	0	0	0	0	0	1/10	1/10	1/10	3/10	0	0	1/5
17	5/6	1/6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	5/6	1/6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 3: The transition matrix P of the Markov chain for $k = 3$.

References

1. B. M. Ábrego, E. M. Arkin, S. Fernández-Merchant, F. Hurtado, M. Kano, J. S. Mitchell, and J. Urrutia. Matching points with squares. *Discrete & Computational Geometry*, 41(1):77–95, 2009.
2. L. E. Caraballo, C. Ochoa, P. Pérez-Lantero, and J. Rojas-Ledesma. Matching colored points with rectangles. *Journal of Combinatorial Optimization*, 33(2):403–421, 2017.
3. X. Chen, J. Pach, M. Szegedy, and G. Tardos. Delaunay graphs of point sets in the plane with respect to axis-parallel rectangles. *Random Structures and Algorithms*, 34(1):11–23, 2009.
4. A. Dumitrescu and R. Kaye. Matching colored points in the plane: Some new results. *Computational Geometry*, 19(1):69–85, 2001.
5. A. Dumitrescu and W. L. Steiger. On a matching problem in the plane. *Discrete Mathematics*, 211:183–195, 2000.
6. L. C. Larson. *Problem-solving through problems*. Springer Science & Business Media, 1983.
7. J. R. Norris. *Markov chains*. Number 2. Cambridge University Press, 1998.

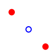
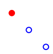
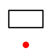
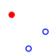
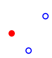
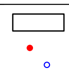
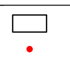
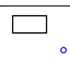
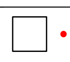
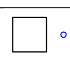
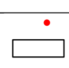
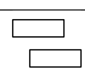
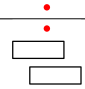
e_i	elem. of e_i	$f(e_i)$	neighbors of e_i
e_1		0	$(e_2, 1/2), (e_3, 1/2)$
e_2	$\emptyset$	2	$(e_1, 1)$
e_3		0	$(e_4, 1/6), (e_5, 1/6), (e_6, 1/3), (e_7, 1/6), (e_8, 1/6)$
e_4		0	$(e_4, 1/8), (e_5, 1/8), (e_6, 3/8), (e_7, 1/8), (e_9, 1/4)$
e_5		0	$(e_{10}, 3/4), (e_{11}, 1/4)$
e_6		2	$(e_2, 1/4), (e_{12}, 1/8), (e_{13}, 1/8), (e_{14}, 1/4), (e_{15}, 1/4)$
e_7		0	$(e_{10}, 3/4), (e_{11}, 1/4)$
e_8		0	$(e_{10}, 3/4), (e_{16}, 1/4)$
e_9		2	$(e_2, 3/10), (e_{12}, 1/10), (e_{14}, 1/5), (e_{15}, 1/5), (e_{17}, 1/5)$
e_{10}		2	$(e_2, 1/2), (e_3, 1/2)$
e_{11}	$\emptyset$	4	$(e_1, 1)$
e_{12}		0	$(e_2, 1/2), (e_3, 3/10), (e_6, 1/5)$
e_{13}		0	$(e_2, 1/2), (e_3, 3/10), (e_6, 1/5)$
e_{14}		0	$(e_2, 3/10), (e_3, 1/2), (e_6, 1/5)$
e_{15}		0	$(e_2, 1/2), (e_3, 3/10), (e_6, 1/5)$
e_{16}		2	$(e_2, 1/5), (e_{12}, 1/10), (e_{13}, 1/10),$ $(e_{14}, 1/10), (e_{15}, 3/10), (e_{18}, 1/5)$
e_{17}		2	$(e_1, 5/6), (e_2, 1/6)$
e_{18}		2	$(e_1, 5/6), (e_2, 1/6)$

Table 2: The table shows the 18 states of the Markov chain for $k = 3$. In the second column we show the first component of e_i, and in the third column we show the second component $f(e_i)$. In the last column we show the neighbor states of e_i as a list of tuples of the form $(e_j, P_{i,j})$, where $P_{i,j} = \text{Prob}(X_{\ell+1} = e_j \mid X_\ell = e_i) > 0$ is the transition probability from e_i to e_j.