Fast Property Testing and Metrics for Permutations

JACOB FOX[†] and FAN WEI

Department of Mathematics, Stanford University, Stanford, CA 94305, USA
(e-mail: `jacobfox@stanford.edu`, `fanwei@stanford.edu`)

The goal of property testing is to quickly distinguish between objects which satisfy a property and objects that are ϵ-far from satisfying the property. There are now several general results in this area which show that natural properties of combinatorial objects can be tested with 'constant' query complexity, depending only on ϵ and the property, and not on the size of the object being tested. The upper bound on the query complexity coming from the proof techniques is often enormous and impractical. It remains a major open problem if better bounds hold.

Maybe surprisingly, for testing with respect to the rectangular distance, we prove there is a universal (not depending on the property), polynomial in $1/\epsilon$ query complexity bound for two-sided testing hereditary properties of sufficiently large permutations. We further give a nearly linear bound with respect to a closely related metric which also depends on the smallest forbidden subpermutation for the property. Finally, we show that several different permutation metrics of interest are related to the rectangular distance, yielding similar results for testing with respect to these metrics.

2010 *Mathematics subject classification*: Primary 68W20
Secondary 68R05, 05D40, 05A05, 68R15

1. Introduction

Traditionally, algorithms that run in time polynomial in the input size were considered fast. However, as the desired input size has increased, this notion of fast is sometimes insufficient. Some examples include in algorithmic problems on networks like the internet or the brain, or in ranking websites for search algorithms, in which the structures being studied have billions of elements and are often not well understood. In order to handle such large structures, sublinear time algorithms are desired. One would not expect such algorithms to be able to determine properties of the structures with certainty. This is where property testing comes in.

The goal of property testing is to quickly distinguish between objects which satisfy a property and objects that are ϵ-far from satisfying the property. The study of this notion was initiated by Rubinfield and Sudan [28]. Subsequently, Goldreich, Goldwasser and Ron [17] began the investigation of property testers for combinatorial objects. There are now several quite general results in this area which show that properties can be tested with 'constant' query complexity, depending only on ϵ and the property, and not on the size of the object being tested. A property P is *one-sided testable* if there is a function $q(\epsilon)$ and a randomized algorithm with query complexity $q(\epsilon)$ which, on an input which has property P, correctly outputs that the object has property P, and on input that is ϵ-far from satisfying P, correctly outputs with probability at least 2/3 that the object is ϵ-far from satisfying P. Property P is *two-sided testable* if it correctly outputs in either case with probability at least 2/3. Note that if an input neither satisfies P nor is ϵ-far from P, it has no guarantee on the output. An exemplary result in this area, due to Alon and Shapira [4], states that every hereditary graph property is one-sided testable. However the query complexity bound it gives is at least of wowzer type in $1/\epsilon$, which is one level higher in the Ackermann hierarchy than the tower function, as it uses the strong regularity lemma. Enormous bounds like this on the query complexity are typical of many general results in this area and is a major drawback as the bounds are impractical. Besides regularity methods, which often give tower-type or worse estimates, some of the other results in this area are established using compactness arguments and thus yield no bound. There are some examples of progress on quantitative bounds for property testing. See, for example, [1, 3, 5, 6, 13]. However, it remains a major open problem if better bounds hold for the various property testing results.

In this paper, we address this problem for permutations. To properly understand our results, it is important to try to first determine what is a good notion (or notions) of distance between combinatorial objects. This is because we need to understand which metric we are using when we say that two objects are ϵ-far from each other in this metric. For graphs, the edit distance, which is the fraction of pairs which one needs to add or delete edges from in order to turn one graph into the other, is quite natural, and it is not surprising that it is the most studied with regard to graph property testing. For permutations, there are now several important metrics that naturally arise in ranking problems in statistics. See, for example, the book by Diaconis [9]. This makes it less clear for which metrics permutation property testing should be applied.

An early paper of Cooper [7] develops an analogue of Szemerédi's regularity lemma for permutations[†] and deduces a permutation removal lemma. Typically, removal lemmas are equivalent to saying that certain properties of combinatorial objects are one-sided testable with respect to some metric. However, Cooper's permutation removal lemma does not give such a metric and so does not actually translate to a result in property testing.

Since Cooper's work, there are now two different notions of distance which have been studied for permutation property testing, the rectangular distance (or cut distance)

[†] This regularity lemma was subsequently improved upon by Hoppen, Kohayakawa and Sampaio [19] and with much better quantitative estimates by Fox, Lovász and Zhao [14].

and Kendall's tau distance. The rectangular distance for permutations is an analogue of the Frieze–Kannan cut distance for graphs, which has played an important role in the development of the weak regularity lemma, graph limits and approximation algorithms for graphs. See, for example, the book by Lovász [25] and the paper [14].

A permutation π of length n is a bijection from $[n] := \{1, 2, \ldots, n\}$ to itself. We can represent π as n points in the plane with the coordinates of the ith point being $(i, \pi(i))$. The *rectangular (cut) distance* between two permutations π_1, π_2 of length n is defined to be

$$\rho_{\mathrm{r}}(\pi_1, \pi_2) = \frac{1}{n} \max_{S,T} ||\pi_1(S) \cap T| - |\pi_2(S) \cap T||,$$

where the maximum is over all sub-intervals S, T of $[n]$. Thus, the rectangular distance is the normalized maximum discrepancy in rectangles between the number of points of the form $(i, \pi_1(i))$ and the number of points of the form $(i, \pi_2(i))$. While the rectangular distance is defined globally, through a counting lemma, it can be shown that two permutations have small rectangular distance if and only if they have roughly the same densities of all small subpermutations. This is an analogue of similar results for graphs; see [25] and [20] for details on these results for graphs and permutations, respectively.

A copy of a permutation σ of length k in a permutation π of length n is a subsequence of π that has the same order type as σ. That is, a copy of σ in π is a sequence of integers $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$ such that $\pi(i_j) < \pi(i_\ell)$ if and only if $\sigma(j) < \sigma(\ell)$. If π contains a copy of σ, then we say that σ is a *subpermutation* of π. If π does not contain a copy of σ, then we say that π avoids σ or is σ-free. A permutation property is just a family of permutations. A permutation property $\mathcal{P}$ is *hereditary* if it is closed under subpermutations, that is, if every subpermutation of a permutation in $\mathcal{P}$ is also in $\mathcal{P}$. Hoppen, Kohayakawa, Moreira and Sampaio [21] proved that every hereditary permutation property is one-sided testable with respect to the rectangular distance. Their proof uses a compactness argument and does not give any bound on the query complexity. They also conjectured a stronger result that hereditary permutation properties are strongly testable, that is, they can be tested with respect to *Kendall's tau distance*: for two permutations π_1, π_2 of length n,

$$\rho_{\mathrm{KT}}(\pi_1, \pi_2) = \frac{1}{\binom{n}{2}} |\{(i, j) \text{ such that } \pi_1(i) < \pi_1(j), \pi_2(i) > \pi_2(j), i, j \in [n]\}|.$$

Alternatively, Kendall's tau distance between π_1, π_2 can also be defined as the minimum number of adjacent transpositions (*i.e.* swapping the y-coordinates of the points $(i, \pi(i))$ and $(i+1, \pi(i+1))$) required to turn π_1 into π_2, and normalized by dividing by $\binom{n}{2}$. This conjecture is stronger because the rectangular distance is small if Kendall's tau distance is small, but the converse is not true. For example, for two random permutations of length n almost surely have rectangular distance $o(1)$, but Kendall's tau distance $\Omega(1)$. The nice conjecture of Hoppen, Kohayakawa, Moreira and Sampaio [21] was verified by Klimošová and Král' [23]. However, even for the property of being σ-free for some fixed permutation σ, the bound on the query complexity is enormous, of Ackermann type in

$1/\epsilon$, and hence not primitive recursive.[‡] In another work [15], we prove that there is a polynomial in $1/\epsilon$ bound for one-sided testing σ-freeness, where the exponent depends on σ. The result generalizes to show that hereditary properties are one-sided testable with respect to Kendall's tau distance, and for typical properties, it gives a polynomial bound.

Another important permutation metric is Spearman's footrule distance. For two permutation $\pi_1, \pi_2 : [n] \to [n]$, their *Spearman's footrule distance* is

$$D(\pi_1, \pi_2) = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} |\pi_1(i) - \pi_2(i)|.$$

A fundamental result of Diaconis and Graham [10] states that

$$\rho_{\mathrm{KT}}(\pi_1, \pi_2) \leqslant D(\pi_1, \pi_2) \leqslant 2\rho_{\mathrm{KT}}(\pi_1, \pi_2).$$

Thus Kendall's tau distance and Spearman's footrule distance are within a factor of two, and so testing with respect to Kendall's tau distance is essentially equivalent to testing with respect to Spearman's footrule distance.

Maybe surprisingly, for testing with respect to the rectangular distance, we prove that there is a universal (not depending on the property), polynomial in $1/\epsilon$ query complexity bound for two-sided testing of hereditary properties of sufficiently large permutations. One drawback of the definition of the rectangular distance is that it is global, whereas with Kendall's tau distance, we see that we can make sequential local moves in order to get from one permutation to the other. We study a new distance for permutations, whose general form is called *earth mover's distance* in statistics. It turns out to be quite natural and defined based on local moves, yet we prove it is small if and only if the rectangular distance is small.

Definition 1.1 (planar tau distance). We say that σ is obtained from π by a *planar simple transposition* if there is an integer $1 \leqslant i < |\pi| - 1$ such that σ is the same as π except $\sigma(i) = \pi(i+1), \sigma(i+1) = \pi(i)$, or $\sigma^{-1}(i) = \pi^{-1}(i+1), \sigma^{-1}(i+1) = \pi^{-1}(i)$. The *planar tau distance* $\rho_{\mathrm{PT}}(\pi_1, \pi_2)$ between two permutations π_1, π_2 of length n is defined as $1/\binom{n}{2}$ times the minimum number of planar simple transpositions needed to transform π_1 into π_2. □

The $1/\binom{n}{2}$ factor in the definition is the proper normalization in order to guarantee that this distance is always at most one. We can also define the planar tau distance in terms of Kendall's tau distance, since the planar tau distance allows for adjacent transpositions in both the horizontal and vertical directions. Thus

$$\rho_{\mathrm{PT}}(\pi_1, \pi_2) = \min_{\text{bijection } \theta : [n] \to [n]} \rho_{\mathrm{KT}}(\mathrm{id}, \theta) + \rho_{\mathrm{KT}}(\pi_1, \pi_2 \circ \theta).$$

It is also useful and interesting to define a planar analogue of Spearman's footrule distance.

[‡] In the conference version of [23], it is incorrectly stated that the proof gives a double exponential bound for testing σ-freeness.

Definition 1.2 (Earth mover's distance (for permutations)). The *earth mover's distance* between two permutations π_1, π_2 of length n is defined as

$$\rho_{\mathrm{EMD}}(\pi_1, \pi_2) = \frac{1}{\binom{n}{2}} \min_{\text{bijection } \theta:[n]\to[n]} \left(\sum_{i=1}^{n} |i - \theta(i)| + |\pi_1(i) - \pi_2(\theta(i))| \right),$$

where the minimum is over all bijections $\theta : [n] \to [n]$. ☐

This is the sum of L_1 distances between a point in π_1 and the point in π_2 that it maps to under θ. We can treat θ as a permutation. Thus the earth mover's distance is equivalent to

$$\rho_{\mathrm{EMD}}(\pi_1, \pi_2) = \min_{\text{bijection } \theta:[n]\to[n]} D(\mathrm{id}, \theta) + D(\pi_1, \pi_2 \circ \theta).$$

When θ is the identity permutation id, we obtain $\rho_{\mathrm{PT}}(\pi_1, \pi_2) \leqslant \rho_{\mathrm{KT}}(\pi_1, \pi_2)$ and $\rho_{\mathrm{EMD}}(\pi_1, \pi_2) \leqslant D(\pi_1, \pi_2)$. Thus, the planar metrics are at most their classical analogues.

The earth mover's distance is a special case of more general metrics that have been extensively studied before in other contexts. It is called the earth mover's distance or the Monge–Kantorovich norm in computer science, and was first introduced by Monge [27] in 1781 as a central concept in transportation. It is a natural way of measuring the similarity between two digital images (see *e.g.* [29]). In the case of permutations, the digital image has a single one in each row and column. In analysis, it is known as the Wasserstein metric. It is also a special case of the minimum weighted matching problem (see *e.g.* [30]).

By the definitions of the planar tau distance through Kendall's tau distance and the earth mover's distance through Spearman's footrule distance, and by the Diaconis–Graham inequality, we therefore get the following planar analogue of the Diaconis–Graham inequality, which was pointed out by Diaconis [8].

Corollary 1.3.

$$\rho_{PT}(\pi_1, \pi_2) \leqslant \rho_{EMD}(\pi_1, \pi_2) \leqslant 2\rho_{PT}(\pi_1, \pi_2).$$ ☐

The following result shows that the planar tau distance is small if and only if the rectangular distance is small. Together with the previous result, it shows that testing with respect to any of these metrics is the same up to a quadratic change in the testing parameter ϵ.

Theorem 1.4. *For any two permutations π_1, π_2 of length n, we have*

$$\frac{1}{8}\rho_r(\pi_1, \pi_2)^2 \leqslant \rho_{EMD}(\pi_1, \pi_2) \leqslant 48\rho_r(\pi_1, \pi_2)^{1/2}.$$ ☐

Thus the new planar metrics share many of the advantages of both the rectangular distance and Kendall's tau distance. This result and other results relating permutation metrics are proved in the full version of this paper.
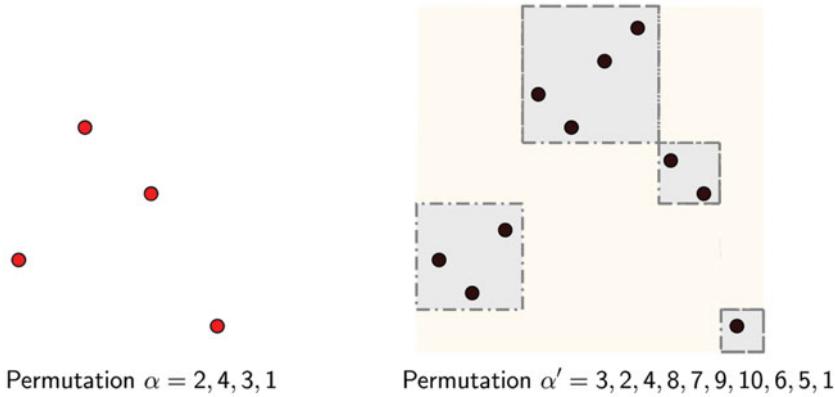
Permutation $\alpha = 2, 4, 3, 1$ Permutation $\alpha' = 3, 2, 4, 8, 7, 9, 10, 6, 5, 1$

Figure 1. Blow-up of a permutation. In this example, α' (the permutation in black) is a blow-up of α (the permutation in red), with the four blocks being four grey squares. The first, second, third and fourth points in α blow-up into blocks of sizes three, four, two and one respectively.

Definition 1.5 (blow-up of a permutation). A permutation α' is a *blow-up* of another permutation α if and only if we can find positive integers $1 = k_1 < k_2 < \cdots < k_{|\alpha|+1} = |\alpha'| + 1$ with the following property. If $i_1, i_2 \leqslant |\alpha'|$ satisfy that $k_{j_1} \leqslant i_1 < k_{j_1+1}$ and $k_{j_2} \leqslant i_2 < k_{j_2+1}$ with $j_1 \neq j_2$, then $\alpha'(i_1) > \alpha'(i_2)$ if and only if $\alpha(j_1) > \alpha(j_2)$. $\qquad\square$

Intuitively, it means that each point $(i, \alpha(i))$ blows up into a *block* $\{(i', \alpha'(i')), k_i \leqslant i' < k_{i+1}\}$. The ith block is of size $k_{i+1} - k_i$. Notice that we did not specify the permutation within each block. Figure 1 is an example of a blow-up α' of a permutation α.

A *k-blow-up* of α is a blow-up of α where each block has size k. We next define some important parameters for the property $\mathcal{P}$ related to blow-ups of permutations. Since we work with a single property $\mathcal{P}$, we leave it out of the notation to make the notation simpler.

Definition 1.6 (blow-up parameter for $\mathcal{P}$). Given a permutation α, let $k^*(\alpha)$ be the minimum positive integer k (if it exists) such that no k-blow-up of α is in $\mathcal{P}$. If no such integer k exists, that is, for every k there is a k-blow-up of α which is in $\mathcal{P}$, then we define $k^*(\alpha) = \infty$. Given a positive integer T, let $k^*(T)$ be the maximum of $k^*(\alpha)$ over all permutations α of length T for which $k^*(\alpha) < \infty$. If no such α exists, that is, $k^*(\alpha) = \infty$ for all permutations α of length T, then we define $k^*(T) = \infty$. $\qquad\square$

Note that if $\mathcal{P}$ is hereditary and has a forbidden subpermutation of length at most T, then it has a forbidden subpermutation α of length T, and α is a 1-blow-up of itself and is not in $\mathcal{P}$, which implies that $k^*(T)$ is finite in this case.

We give a nearly linear bound for the query complexity for testing $\mathcal{P}$ which depends on the smallest forbidden subpermutation for the property. We first need an important definition. A 0–1 matrix A *contains* another 0–1 matrix B if there is a submatrix A' of A of the same size as B such that for every one entry of B, the corresponding entry of A' is a one. For a permutation σ, the extremal number $\mathrm{ex}(n, \sigma)$ is the maximum number of

one entries in an $n \times n$ matrix with entries 0 or 1 which does not contain the permutation matrix of σ. Füredi and Hajnal [16] conjectured that for each permutation σ, the limit

$$c(\sigma) := \lim_{n \to \infty} \frac{\text{ex}(n, \sigma)}{n}$$

exists. Klazar [22] proved that the Füredi–Hajnal conjecture implies the well-known Stanley–Wilf conjecture. A celebrated result of Marcus and Tardos [26] verifies the Füredi–Hajnal conjecture, and hence the Stanley–Wilf conjecture. It shows that $c(\sigma) = 2^{O(|\sigma|\log|\sigma|)}$. The first author [12] improved the bound to $c(\sigma) = 2^{O(|\sigma|)}$, and showed that $c(\sigma) = 2^{|\sigma|^{\Omega(1)}}$ for almost all permutations σ of a given order. The constant $c(\sigma)$ is known as the *Füredi–Hajnal constant* of σ. The fact that $\text{ex}(n, \sigma)$ is superadditive in n implies that $\text{ex}(n, \sigma) \leqslant c(\sigma)n$ for all n.

Theorem 1.7. *For each proper hereditary permutation property $\mathcal{P}$ and $\epsilon > 0$, let $C = C(\mathcal{P}) = 1000c(\sigma)$, where $c(\sigma)$ is the Füredi–Hajnal constant of a smallest forbidden subpermutation σ for $\mathcal{P}$. Let*

$$M = M(\epsilon) = \left(\frac{2000C}{\epsilon}\right)\log^2(\epsilon/2), \quad n_0 = n_0(M, \epsilon) = 32M^5k^*(M)/\epsilon^3.$$

There is a two-sided tester for $\mathcal{P}$ with respect to the planar tau distance of query complexity M for permutations of size at least n_0. □

The tester works as follows. Let M be as specified in Theorem 1.7. For a permutation π, we pick a subpermutation π' of π of size M uniformly at random; call it an *M-sample*. That is, we pick a subset $S \subset [n]$ of size M uniformly at random, and π' is the subpermutation of π induced on S. If, for all integers k, there exists a k-blow-up of $\pi' \in \mathcal{P}$, our algorithm outputs 'π is in $\mathcal{P}$'. If there exists an integer k such that no k-blow-up of π' is in $\mathcal{P}$, our algorithm outputs 'π is not in $\mathcal{P}$'. We remark that the constant dependence can sometimes be improved by using extremal properties of the family of forbidden subpermutations rather than just the smallest forbidden subpermutation.

The next theorem gives a universal quadratic bound (not depending on the property) on the query complexity for testing sufficiently large permutations.

Theorem 1.8. *For any hereditary permutation property $\mathcal{P}$ and $\epsilon > 0$, let $M = M(\epsilon) = 20000/\epsilon^2$ and $n_0 = n_0(M, \epsilon) = 32M^5k^*(M)/\epsilon^3$. There is a two-sided tester for $\mathcal{P}$ with respect to the planar tau distance of query complexity M for permutations of size at least n_0.* □

Theorems 1.7 and 1.8 are both with respect to two-sided testing. For one-sided testing, we can still get reasonably good bounds, as stated in Theorem 1.9, by showing that very likely a permutation has the property that it is close to a blow-up of a random subpermutation α and a somewhat larger random subpermutation π' very likely contains a $k^*(|\alpha|)$-blow-up of α. These bounds are polynomial in $1/\epsilon$ as long as the blow-up parameter $k^*(M)$ for $\mathcal{P}$ is bounded above by a polynomial in M. In particular, for almost all permutations σ of length s, we get a universal bound for testing σ-freeness. This is because almost all permutations σ of length s have no sub-interval of length three whose
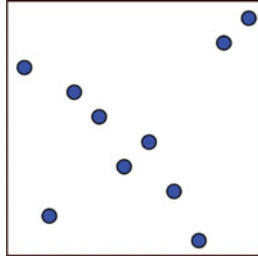
J. Fox and F. Wei

Figure 2. Permutation representation in $[0,1]^2$. This is an example of the permutation $8, 2, 7, 6, 4, 5, 3, 1, 9, 10$.

image is an interval of length three, and hence k^* is at most three for the property of being σ-free.

Theorem 1.9. *Let $\mathcal{P}$ be a hereditary permutation property and $\epsilon > 0$. Let $M_1(\epsilon), M_2(\epsilon)$ be the value M in terms of ϵ in Theorems 1.7 and 1.8, respectively. Let $n_0(M, \epsilon) = 32M^5 k^*(M)/\epsilon^3$. Let*

$$M' = \min(n_0(M_1(\epsilon/2), \epsilon/2), n_0(M_2(\epsilon/2), \epsilon/2)).$$

There is a one-sided tester for $\mathcal{P}$ with respect to the planar tau distance of query complexity M' for permutations of size at least M'. □

Finally, we show that several different permutation metrics of interest are closely related to the cut metric, yielding similar results for testing with respect to these metrics.

We often consider the input permutation π of length n as a collection of points $(i/n, \pi(i)/n)$ in the unit square $[0,1]^2 := [0,1] \times [0,1]$. See Figure 2 for an example.

If we are testing for a hereditary permutation property $\mathcal{P}$, and we know that the following hold:

- there is a permutation $\sigma \notin \mathcal{P}$ (or even if large blow-ups of σ are not in $\mathcal{P}$),
- there are $|\sigma|$ rectangles whose horizontal intervals are disjoint and whose vertical intervals are disjoint,
- each of the rectangles contains a significant fraction of the points corresponding to π, and
- if we pick one point from each of the rectangles, then we form a copy of σ,

then a large sample of points from π will likely have many points in each rectangle and thus contain a large blow-up of σ and certify that $\pi \notin \mathcal{P}$. This simple idea is very important for our various property testing algorithms, and is demonstrated in Figure 3.

2. Equivalence between different metrics on permutations

We next define several different metrics between permutations of the same length. Each of the metrics we use here is normalized such that the maximum distance between two permutations cannot exceed 1. We then study properties of these metrics and the relationships between them. Each of these metrics has the property that if two
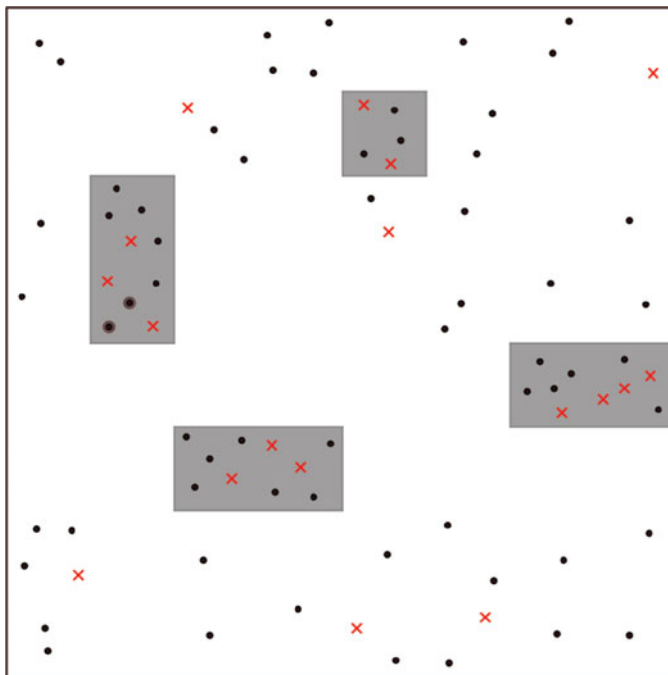
Figure 3. Illustration of a permutation containing a large blow-up. The black dots denote the points in π; while the red crosses are the points being picked by the M-sample.

permutations have small distance in one metric, they also have small distance in the other metrics. Furthermore, for two permutations to have small distance in any of these metrics is equivalent to having, for each small permutation μ, roughly the same density of μ as a subpermutation.

Recall that we sometimes view the permutation π of length n as the collection of points $(i, \pi(i))$ with $i \in [n]$ or, normalized, as the collection of points $(i/n, \pi(i)/n) \in [0,1]^2$. It should be clear from the context which is used. Viewing a permutation as a collection of points in $[0,1]^2$, given any rectangle $S \subset [0,1]^2$, let $S(\pi)$ be the number of points of π inside S (including the boundary).

Definition 2.1. Let π_1, π_2 be permutations of length n.

(1) *Rectangular distance*, or *cut distance*. The rectangular (cut) distance between π_1, π_2 is defined as

$$\rho_r(\pi_1, \pi_2) = \frac{1}{n} \max_{I, J \text{ intervals} \subset [0,1]} |(I \times J)(\pi_1) - (I \times J)(\pi_2)|,$$

where the maximum is over all closed intervals $I \subset [0,1]$ and $J \subset [0,1]$.

(2) *Dyadic distance*. A closed interval $I \subset [0,1]$ is called a *dyadic interval* if there exist positive integers i, k such that $I = [i/2^k, (i+1)/2^k]$. The dyadic distance between π_1, π_2

is defined as

$$\rho_{\mathrm{d}}(\pi_1, \pi_2) = \frac{1}{n} \max_{I,J \text{ dyadic intervals}} |(I \times J)(\pi_1) - (I \times J)(\pi_2)|,$$

where the maximum is over all intervals $I \subset [0,1]$ and $J \subset [0,1]$ and I, J are dyadic intervals.

(3) *Square distance.* The square distance between π_1, π_2 is defined as

$$\rho_{\mathrm{s}}(\pi_1, \pi_2) = \frac{1}{n} \max_{I,J} |(I \times J)(\pi_1) - (I \times J)(\pi_2)|,$$

where the maximum is over all intervals $I \subset [0,1]$ and $J \subset [0,1]$ and $|I| = |J|$. Thus $I \times J$ is a square in $[0,1]^2$.

(4) *Dyadic square distance.* The dyadic square distance between π_1, π_2 is defined as

$$\rho_{\mathrm{ds}}(\pi_1, \pi_2) = \frac{1}{n} \max_{I,J \text{ dyadic intervals}, |I|=|J|} |(I \times J)(\pi_1) - (I \times J)(\pi_2)|,$$

where the maximum is over all *dyadic squares* $I \times J$, which means I, J are dyadic intervals and $|I| = |J|$.

(5) *Earth mover's distance.* The earth mover's distance between π_1, π_2 is defined as

$$\rho_{\mathrm{EMD}}(\pi_1, \pi_2) = \frac{1}{\binom{n}{2}} \min_{n\text{-permutation } \theta} \left(\sum_{i=1}^{n} |i - \theta(i)| + |\pi_1(i) - \pi_2(\theta(i))| \right),$$

where the minimum is over all permutations $\theta : [n] \to [n]$ of length n. This is the sum of L_1 distances between a point in π_1 and the point in π_2 that it maps to under θ.

(6) *Planar tau distance.* A permutation σ can be obtained from a permutation π of the same length by a *planar simple transposition* if there exists an integer $1 \leqslant i < |\pi| - 1$ such that σ is the same as π except either $\sigma(i) = \pi(i+1)$, $\sigma(i+1) = \pi(i)$, or $\sigma^{-1}(i) = \pi^{-1}(i+1)$, $\sigma^{-1}(i+1) = \pi^{-1}(i)$. The planar tau distance between π_1, π_2, denoted as $\rho_{\mathrm{PT}}(\pi_1, \pi_2)$, is the minimum number of planar simple transpositions required to transform π_1 into π_2, and then normalized by $1/\binom{n}{2}$. Restated, the planar tau distance between π_1 and π_2 is the normalized (so divided by $1/\binom{n}{2}$) minimum number of consecutive row or column swaps needed to obtain the permutation matrix of π_2 from the permutation matrix of π_1. Recall that the permutation matrix of a permutation π of length n is an $n \times n$ matrix M_π with $M_{i\pi(i)} = 1$ for each $i \in [n]$, and the remaining entries are 0. □

We first make several remarks about these metrics. For two permutations of length n, their planar tau distance is at most their Kendall's tau distance, as Kendall's tau distance is defined in the same way but is more restrictive on the allowed moves (only allowing consecutive column swaps). Similarly, their earth mover's distance is at most their Spearman's tau distance, as taking θ to be the identity permutation of length n, we obtain Spearman's footrule distance. Summarizing, the planar distances are at most their classical variants. We also recall Corollary 1.3, which shows that $\rho_{\mathrm{PT}}(\pi_1, \pi_2) \leqslant \rho_{\mathrm{EMD}}(\pi_1, \pi_2) \leqslant 2\rho_{\mathrm{PT}}(\pi_1, \pi_2)$. Thus the planar tau distance and earth mover's distance are within a factor of two of each other.

Also, it is worth discussing the complexity of computing these metrics for two permutations of length n. The rectangular distance is defined as the minimum over $O(n^4)$ choices of pairs of intervals, while the square distance is only over $O(n^3)$ choices of pairs of intervals, and the dyadic distance and the dyadic square distance is defined only over $O(n^2)$ choices of pairs of intervals. Hence, the dyadic distances appear to be considerably faster to determine exactly.

If we only want to approximate these distances, we can do a much faster computation. Two squares whose horizontal and vertical intervals differ in endpoints by at most $\epsilon/8$ in each coordinate differ by at most $\epsilon/2$ in the fraction of points of the permutation in the rectangle for a given permutation. Hence, by considering only multiples of $\epsilon/8$ as possible endpoints, we can approximate the rectangular distance within ϵ using only $O(\epsilon^{-4})$ rectangles. Similarly, we can approximate within ϵ the square distance by using at most $O(\epsilon^{-3})$ squares, and the dyadic distance or the dyadic square distance within ϵ using only $O(\epsilon^{-2})$ dyadic rectangles. Thus, these distances can be determined or approximated rather quickly, with the dyadic distances being the fastest to approximate.

On the other hand, while very natural, the earth mover's distance is defined as the minimum over $n!$ permutations, which requires a huge computation. Similarly, it is unclear if there is an efficient algorithm for computing the planar tau distance efficiently. However, it is possible to efficiently approximate these planar distances. By partitioning into boxes of side length about $\epsilon/4$, just using the information about the fraction of points in each box, by considering roughly the fraction of points in each box that match up to points in other boxes, it is possible to show that one can compute the planar tau distance and the earth mover's distance each within ϵ in time which is a function only of ϵ.

We next prove Theorem 1.4, which states that for any two permutations π_1, π_2 of length n, we have

$$\frac{1}{8}\rho_{\mathrm{r}}(\pi_1, \pi_2)^2 \leqslant \rho_{\mathrm{EMD}}(\pi_1, \pi_2) \leqslant 48\rho_{\mathrm{r}}(\pi_1, \pi_2)^{1/2}.$$

Proof of Theorem 1.4. The earth mover's distance between two permutations π_1, π_2 is the normalized minimum, over all permutations $\theta : [n] \to [n]$, which is the same as the normalized minimum over all matchings between points in π_1 and π_2, of the sum over all matched pairs of the taxicab (L_1) distance between the two points that are matched.

We first show that if $\rho_{\mathrm{r}}(\pi_1, \pi_2) \geqslant \epsilon$, then $\rho_{\mathrm{EMD}}(\pi_1, \pi_2) \geqslant \epsilon^2/8$. Since $\rho_{\mathrm{r}}(\pi_1, \pi_2) \geqslant \epsilon$, we can find a rectangle $I \times J \subset [0, 1]^2$ such that there are at least ϵn more points in this rectangle in π_1 than π_2, or *vice versa*. Without loss of generality, we assume there are ϵn more points in π_1 than in π_2. Therefore for any bijection θ that maps points in π_1 to the ones in π_2, at least ϵn points in π_1 have to map to the points of π_2 that are outside the rectangle $I \times J$.

Assume $I = [x_1, x_2], J = [y_1, y_2]$. Let

$$I' = [\max(x_1 - \epsilon/8, 0), \min(x_2 + \epsilon/8, 1)], \quad J' = [\max(y_1 - \epsilon/8, 0), \min(y_2 + \epsilon/8, 1)].$$

Thus $I \times J \subset I' \times J'$, and the difference between the two rectangles has margin at most $\epsilon/8$. Figure 4 illustrates these two rectangles. Since π_2 is a permutation, there are at most
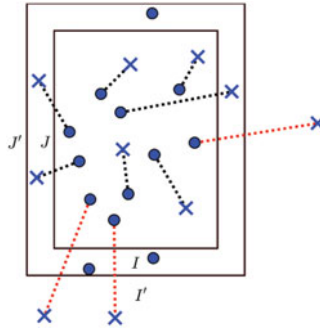
Figure 4. Example of the matching process. The inner rectangle is $I \times J$ while the outer rectangle is $I' \times J'$. The round dots are the points of π_1 and the crosses are the points of π_2. If a point in π_1 is connected to a point of π_2 by a dashed line, it means they are matched. Since there are many more points of π_1 than of π_2 in $I \times J$ and not many points of π_2 in $(I' \times J') \setminus (I \times J)$, many left-over points of π_1 in $I \times J$ have to map to points of π_2 outside $I' \times J'$; each such match has a distance at least $\epsilon/8$ between its points.

$4 \cdot \epsilon n/8 = \epsilon n/2$ points of π_2 that are inside the region $(I' \times J') \setminus (I \times J)$. Therefore there are at least $\epsilon n - \epsilon n/2 = \epsilon n/2$ points of π_1 inside $I \times J$ that have to map to points of π_2 outside $I' \times J'$. However for these points of π_1, the distance between it and the point in π_2 it maps to is at least $\epsilon/8$. Therefore

$$\rho_{\mathrm{EMD}}(\pi_1, \pi_2) \geqslant \epsilon/8 \cdot \epsilon n/2 \cdot n / \binom{n}{2} \geqslant \epsilon^2/8.$$

Thus we have proved $\frac{1}{8}\rho_{\mathrm{r}}(\pi_1, \pi_2)^2 \leqslant \rho_{\mathrm{EMD}}(\pi_1, \pi_2)$.

We now show that if $\rho_{\mathrm{r}}(\pi_1, \pi_2) \leqslant \epsilon$, then $\rho_{\mathrm{EMD}}(\pi_1, \pi_2) \leqslant 48\sqrt{\epsilon}$. We may assume that $n \geqslant 2$ as otherwise these distances are all 0. We will find a bijection θ such that

$$\frac{1}{\binom{n}{2}} \sum_{i=1}^{n} |i - \theta(i)| + |\pi_1(i) - \pi_2(\theta(i))| \leqslant 48\sqrt{\epsilon}.$$

To see this, we partition $[0,1]^2$ into d^2 squares, each of side length $1/d$, where $d = 2^{-h}$ with $h = \lfloor \log_2(1/2\sqrt{\epsilon}) \rfloor$. We will define θ which matches points of π_1 to points in π_2 recursively in rounds. In round 0, we match up as many points of π_1 to points in π_2 as possible that lie in the same dyadic square of side length d. In round ℓ, we match up as many not yet matched points of π_1 to not yet matched points in π_2 as possible that lie in the same dyadic square of side length $2^\ell d$. For each pair of points matched in level ℓ, their L_1 distance is at most $2^{\ell+1}d$. In the last round $\ell = h$, the remaining unmatched points in π_1 necessarily get matched to the unmatched points in π_2 as they all lie in the square of side length $1 = 2^h d$.

As the discrepancy in the number of points in π_1 and in π_2 in any square is at most ϵn, after round ℓ, the number of unmatched points is at most $\epsilon n (2^{-\ell}/d)^2$. Thus round $\ell + 1$ matches at most $\epsilon n (2^{-\ell}/d)^2$ pairs of points, each such pair of points has L_1 distance at most $2^{\ell+2}d$. Hence, the sum of the L_1 distances of the pairs matched at level $\ell + 1$ is at most $\epsilon n (2^{-\ell}/d)^2 \cdot 2^{\ell+2}d = (2^{2-\ell}/d)\epsilon n$. Summing over all $\ell \geqslant 0$ gives a total sum of L_1 distances in $[0,1]^2$ less than $8\epsilon d^{-1}n$. Also, there are at most n points in π_1 that get matched

Figure 5. Illustration of partitioning most of an interval I into dyadic intervals. Interval I is the segment EG. It is mostly covered by four dyadic intervals $I_1 = AB, I_2 = DA, I_3 = BC, I_4 = CF$. The only parts in I not covered by dyadic intervals are I_0, I_0'.

to a point in π_2 in the same dyadic square of side length d so that the L_1 distance is at most $2d$. The sum of these L_1 distances is at most $2dn$. Thus, the earth mover's distance between π_1 and π_2 is at most

$$\frac{n}{\binom{n}{2}}(2dn + 8\epsilon d^{-1}n) \leqslant 8(d + 4\epsilon d^{-1}) \leqslant 48\sqrt{\epsilon},$$

where the last inequality uses $\epsilon^{1/2} \leqslant d \leqslant 2\epsilon^{1/2}$. Thus we have proved

$$\rho_{\mathrm{EMD}}(\pi_1, \pi_2) \leqslant 48\rho_{\mathrm{r}}(\pi_1, \pi_2)^{1/2}. \qquad \square$$

The next lemma shows that, up to two logarithmic factors, the rectangular distance is the same as the dyadic distance, which is much faster to compute or approximate. It is easy to give a construction showing that the upper bound is sometimes tight, and reverse-engineering the proof gives a construction showing that the lower bound is sometimes tight up to an absolute constant factor.

Lemma 2.2.

$$\frac{\rho_{\mathrm{r}}(\pi_1, \pi_2)}{8\log_2^2(8/\rho_{\mathrm{r}}(\pi_1, \pi_2))} \leqslant \rho_{\mathrm{d}}(\pi_1, \pi_2) \leqslant \rho_{\mathrm{r}}(\pi_1, \pi_2).$$

Proof. It is clear from the definition that $\rho_{\mathrm{r}}(\pi_1, \pi_2) \geqslant \rho_{\mathrm{d}}(\pi_1, \pi_2)$. We thus need to show the other inequality. Given $\rho_{\mathrm{r}}(\pi_1, \pi_2) \geqslant \epsilon$, we will show

$$\rho_{\mathrm{d}}(\pi_1, \pi_2) \geqslant \frac{\epsilon}{2\log^2(1/\epsilon)}.$$

Since $\rho_{\mathrm{r}}(\pi_1, \pi_2) \geqslant \epsilon$, by definition, there exist intervals $I, J \subset [0, 1]$ such that

$$\frac{1}{n}\max_{I,J}|(I \times J)(\pi_1) - (I \times J)(\pi_2)| \geqslant \epsilon.$$

We want to tile most of $I \times J$ by dyadic rectangles, and thus one of them has large difference between the number of points in π_1 and π_2. We do this by first covering most of the interval I by dyadic intervals, shown in Figure 5.

We know I must contain a dyadic interval I_1 such that $I_1 \subset I$ and $|I_1| \geqslant \frac{1}{2}|I|$. Removing I_1 from I, we are left with at most two other intervals I_2' and I_3' with I_2' coming before I_3' and each of length at most $|I|/2$. In the next level, notice that since I_1 is a dyadic interval, I_2', I_3' each have one endpoint being dyadic (*i.e.* of the form $i/2^k$ for some integers i, k). Thus again we can find a dyadic interval $I_2 \subset I_2'$ with the same right endpoint as I_2' and $|I_2| \geqslant |I_2'|/2$; again $I_2' \setminus I_2$ is another interval with one endpoint dyadic. We can similarly
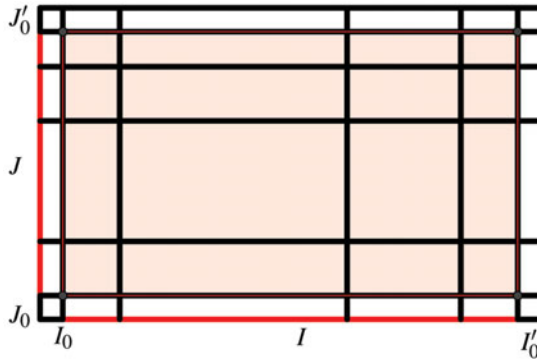
Figure 6. Covering most of the rectangle $I \times J$ by dyadic rectangles by the dyadic partitioning of most of I, J. $I \times J$ is the largest rectangle. The sub-interval $\tilde{I}$ of I covered by dyadic intervals is shown in red. $\tilde{I} = I \setminus (I_0 \cup I'_0)$. Similar for the sub-interval $\tilde{J}$ of J covered by dyadic intervals. The dyadic rectangles are defined by the dyadic intervals covering $\tilde{I}, \tilde{J}$. Thus the shaded rectangle $\tilde{I} \times \tilde{J}$ is partitioned by dyadic rectangles.

find a dyadic interval $I_3 \subset I'_3$ with the same left endpoint as I'_3 with $|I_3| \geqslant |I'_3|/2$ and such that $I'_3 \setminus I_3$ is an interval with one endpoint having a dyadic coordinate. We know

$$|I \setminus (I_1 \cup I_2 \cup I_3)| \leqslant \left(|I \setminus I_1| + \frac{|I_1|}{2} \right) \leqslant \frac{|I|}{4}.$$

Therefore removing I_1, I_2, I_3 from I leaves us with at most two remaining intervals with each having a dyadic endpoint and their total length is at most $|I|/4$. We repeat this process. In each step, we find at most two new dyadic intervals and removing them further from I leaves us with two remaining intervals each with one dyadic endpoint and each of these remaining intervals has length at most half of the length of the intervals they came from in the previous step. Thus, we can partition I into at most $1 + 2\log(4/\epsilon) = 5 + 2\log(1/\epsilon)$ dyadic intervals $I_1, I_2, \ldots$ and at most two intervals I_0, I'_0 such that $|I_0| + |I'_0| \leqslant \epsilon/4$. This is because in the first step we used one dyadic interval I_1, and in each further step, we picked out two dyadic intervals, and the total number of steps used is at most

$$1 + \log_2\left(\frac{|I|/2}{\epsilon/8} \right) \leqslant 1 + \log(4/\epsilon).$$

Similarly, we can partition J into at most

$$1 + 2\log(4/\epsilon) = 5 + 2\log(1/\epsilon)$$

dyadic intervals $J_1, J_2, \ldots$ and at most two intervals J_0, J'_0 such that $|J_0| + |J'_0| \leqslant \epsilon/4$.

Therefore we can cover most of the rectangle $I \times J$ by at most $(5 + 2 \cdot \log_2(1/\epsilon))^2$ dyadic rectangles $I_i \times J_j$ except for four rectangles $S_1 = I_0 \times J$, $S_2 = I'_0 \times J$, $S_3 = I \times J_0$, $S_4 = I \times J'_0$. This is illustrated in Figure 6.

However, notice that for any rectangle $I' \times J'$, we have $0 \leqslant (I' \times J')(\pi_1) \leqslant n \cdot \min(|I'|, |J'|)$ simply because π_1 is a permutation, and the same applies to π_2. Therefore

$$0 \leqslant \left(\bigcup_{i=1}^{4} S_i \right)(\pi_1) \leqslant |I_0| + |I_0'| + |J_0| + |J_0'| \leqslant n\epsilon/2.$$

Similarly,

$$0 \leqslant \left(\bigcup_{i=1}^{4} S_i \right)(\pi_2) \leqslant n\epsilon/2.$$

Thus

$$\left| \left(\bigcup_{i=1}^{4} S_i \right)(\pi_1) - \left(\bigcup_{i=1}^{4} S_i \right)(\pi_2) \right| \leqslant n\epsilon/2.$$

Let $\tilde{I} = I \setminus (I_0 \cup I_0')$, and $\tilde{J} = J \setminus (J_0 \cup J_0')$. Since $|(I \times J)(\pi_1) - (I \times J)(\pi_2)| \geqslant n\epsilon$ and

$$\bigcup_{i=1}^{4} S_i = (I \times J) \setminus (\tilde{I} \times \tilde{J}),$$

we have that

$$|(\tilde{I} \times \tilde{J})(\pi_1) - (\tilde{I} \times \tilde{J})(\pi_2)| \geqslant n\epsilon - n\epsilon/2 \geqslant n\epsilon/2.$$

However, we have shown that $\tilde{I} \times \tilde{J}$ is covered by at most $(5 + 2 \cdot \log_2(1/\epsilon))^2$ dyadic rectangles; therefore there must be a dyadic rectangle S such that

$$|S(\pi_1) - S(\pi_2)| \geqslant \frac{n\epsilon}{2(5 + 2 \cdot \log_2(1/\epsilon))^2} \geqslant \frac{n\epsilon}{8(\log_2(8/\epsilon))^2}.$$

This implies

$$\rho_{\mathrm{d}}(\pi_1, \pi_2) \geqslant \frac{\epsilon}{8(\log_2(8/\epsilon))^2}. \qquad \square$$

The next lemma relates the rectangular distance and the square distance. It is easy to see that the lower bound is sometimes tight, and reverse-engineering the proof shows that the upper bound is sometimes tight up to an absolute constant factor.

Lemma 2.3.

$$\rho_s(\pi_1, \pi_2) \leqslant \rho_r(\pi_1, \pi_2) \leqslant \sqrt{7} \rho_s(\pi_1, \pi_2)^{1/2}.$$

Proof. It is clear from the definition that $\rho_r(\pi_1, \pi_2) \geqslant \rho_s(\pi_1, \pi_2)$. We thus need to show the other inequality. Given $\rho_r(\pi_1, \pi_2) \geqslant \epsilon$, we will show $\rho_s(\pi_1, \pi_2) \geqslant \epsilon^2/7$.

As $\rho_r(\pi_1, \pi_2) \geqslant \epsilon$, there exist intervals $I, J \subset [0, 1]$ such that

$$\frac{1}{n} \max_{I,J} |(I \times J)(\pi_1) - (I \times J)(\pi_2)| \geqslant \epsilon.$$
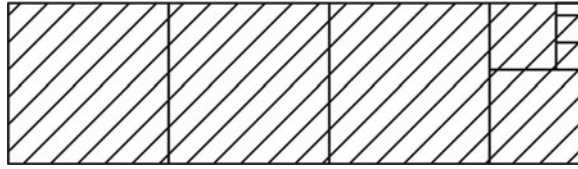
Figure 7. Illustration of covering most of a rectangle by squares. In this example, the rectangle $I \times J$ is covered by squares of four different sizes (the shaded squares) and a remaining rectangle R (the unshaded rectangle in the top right corner) with one side length smaller than $\epsilon/2$.

We want to find a square in $I \times J$ in which there are many more points from π_1 than from π_2, or *vice versa*.

Assuming $|I| \geqslant |J|$, then we partition I into $\lfloor |I|/|J| \rfloor$ intervals of length $|J|$ and a remaining interval with length $|I| - |J| \lfloor |I|/|J| \rfloor$. Thus the rectangle $J \times I$ is partitioned into $\lfloor |I|/|J| \rfloor$ squares each of side length $|J|$ and a remaining rectangle of size $|J| \times (|I| - |J| \lfloor |I|/|J| \rfloor)$. An example of this step can be seen as the three largest squares in Figure 7. And we do the same partitioning procedure for the remaining rectangle by covering the longer side (now it is the interval of length $|J|$) by as many intervals of length $(|I| - |J| \lfloor |I|/|J| \rfloor)$ as possible. We repeat this process until the remaining interval has length at most $\epsilon/2$. In Figure 7, we repeat this step for four rounds (corresponding to squares of four different sizes), and stop at the smallest white rectangle since its shorter side has length at most $\epsilon/2$.

We want to bound the number of squares we obtained. Notice that for each rectangle of size $a_1 \times a_2$, assuming $a_1 \geqslant a_2$, after cutting it into as many squares of size $a_2 \times a_2$ as possible, the remaining rectangle has size $a_3 \times a_2$, where $0 \leqslant a_3 < a_2$ satisfies $a_1 = k_1 a_2 + a_3$ for some positive integer k. Since $k_1 \geqslant 1$ and $a_3 < a_2$, clearly we have $a_3 \leqslant a_1/2$. We then partition the rectangle $a_3 \times a_2$ into $\lfloor a_2/a_3 \rfloor$ squares of size $a_3 \times a_3$, and a remaining rectangle is of size $a_3 \times a_4$ where $0 \leqslant a_4 < a_3$. By the same argument, again we have $a_4 \leqslant a_2/2$. We repeat the process, obtaining squares of side lengths $a_1 > a_2 > \cdots$ until the side length of some rectangle is no more than $\epsilon/2$. Say it stops after subdividing a rectangle of size $a_{N-1} \times a_N$ into many squares, and the remaining rectangle is of size $a_N \times a_{N+1}$ where $0 \leqslant a_{N+1} \leqslant \epsilon/2$. Since $a_i \leqslant a_{i-2}/2$ and $a_N \geqslant \epsilon/2$, we have that

$$N \leqslant 2 \log_2(2/\epsilon). \tag{2.1}$$

The total number of squares is bounded above by

$$a_1/a_2 + a_2/a_3 + \cdots + a_{N-1}/a_N. \tag{2.2}$$

Now, to upper-bound the sum in (2.2), notice that the product of the terms is $a_1/a_N < 1/(\epsilon/2) = 2/\epsilon$ since $a_N > \epsilon/2$. Furthermore, for each $i \leqslant N - 1$, $1 < a_i/a_{i+1} \leqslant 1/(\epsilon/2) = 2/\epsilon$ as $a_i \leqslant 1, a_{i+1} \geqslant \epsilon/2$ for $i \leqslant N - 1$. Given an upper bound of the product of non-negative terms and the range for each term, we know that the sum of the terms is maximized when the terms are at their extreme values; in our case, it would be that all the terms are either $2/\epsilon$ or 1. Therefore (2.2) is bounded above by $(N - 2) + 2/\epsilon$. Plugging in (2.1), the number of squares is bounded above by $2 \log_2(2/\epsilon) + 2/\epsilon$.

Algorithm 1 Two-sided tester

Let M be as specified in Theorem 3.1. Given a permutation π of length n, we pick a subpermutation π' of π of size M uniformly at random. That is, we pick a subset $S \subset [n]$ of size M uniformly at random, and π_0 is the subpermutation of π induced on S.

1: If for all integers k, there exists a k-blow-up of $\pi' \in \mathcal{P}$, our algorithm outputs 'π is in $\mathcal{P}$'. Equivalently, if $k^*(M) = \infty$, then output 'π is in $\mathcal{P}$', and otherwise $k^*(M)$ is finite (recall the definition of $k^*(M)$ in Definition 1.6), and we output 'π is in $\mathcal{P}$' if and only if there is a $k^*(M)$-blow-up of $\pi' \in \mathcal{P}$.

2: If no $k^*(M)$-blow-up of π' is in $\mathcal{P}$, our algorithm outputs 'π is not in $\mathcal{P}$'.

Thus we have partitioned the rectangle $I \times J$ into at most $2 \log_2(2/\epsilon) + 2/\epsilon$ squares and a rectangle R with one side length at most $\epsilon/2$. Thus $|R(\pi_1) - R(\pi_2)| \leqslant \epsilon/2$. Therefore

$$\frac{1}{n}|((I \times J) \setminus R)(\pi_1) - ((I \times J) \setminus R)(\pi_2)| \geqslant \epsilon/2,$$

where $(I \times J) \setminus R$ is partitioned into at most $2 \log_2(2/\epsilon) + 2/\epsilon$ squares. Thus one of the squares, call it S, satisfies

$$\frac{1}{n}|S(\pi_1) - S(\pi_2)| \geqslant \frac{\epsilon}{2(2 \log_2(2/\epsilon) + 2/\epsilon)} \geqslant \epsilon^2/7.$$

The last inequality holds for all $0 < \epsilon \leqslant 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

3. Two-sided property testing under the planar metrics

Recall from Corollary 1.3 that the planar tau distance and the earth mover's distance are within a factor of two. Thus, to test with respect to either metric is essentially the same thing. Hence, we can pick the metric which is the easiest to analyse, and, for our purposes, it is the earth mover's distance.

For a hereditary property $\mathcal{P}$, we propose Algorithm 1, a two-sided tester for property $\mathcal{P}$. It is closely related to the blow-up parameter for $\mathcal{P}$, which is defined in Definition 1.6. We will analyse this algorithm and show its effectiveness in the next theorem. Let $\mathrm{dist}_{\mathrm{PF}}(\pi, \mathcal{P})$ be the minimum of $\rho_{\mathrm{EMD}}(\pi, \pi_0)$ over all $\pi_0 \in \mathcal{P}$ of the same size as π. If $\mathcal{P}$ has no permutation of size $|\pi|$, then $\mathrm{dist}_{\mathrm{PF}}(\pi, \mathcal{P})$ is defined to be 1.

We restate Theorem 1.7 below for convenience.

Theorem 3.1. *Let $\mathcal{P}$ be a proper hereditary family of permutations, let σ be a smallest permutation not in $\mathcal{P}$, and $s = |\sigma|$. Let $c(\sigma)$ be the Füredi–Hajnal constant of σ.*

For any fixed $0 < \epsilon \leqslant 1$, let

$$M = \frac{2000 c(\sigma) \log^2(2/\epsilon)}{\epsilon},$$

so $M \geqslant |\sigma|$ and hence $k^*(M)$ is finite. Then the following holds for any permutation π of length at least $32 M^5 k^*(M)/\epsilon^3$.

(1) *If $\pi \in \mathcal{P}$, then with probability at least $1 - \epsilon$, the algorithm outputs 'π is in $\mathcal{P}$'.*
(2) *If $dist_{PF}(\pi, \mathcal{P}) > \epsilon$, then with probability at least $1 - \epsilon$, the algorithm outputs 'π is not in $\mathcal{P}$'.* $\square$

We justify these two assertions in the next two subsections, respectively.

3.1. When $\pi \in \mathcal{P}$.

We first show that when $\pi \in \mathcal{P}$, with large probability the algorithm outputs $\pi \in \mathcal{P}$. For this part, we need the following key lemma.

Lemma 3.2. *Suppose $|\pi| = n$. Let T, k be positive integers and ϵ' a positive number. Suppose $n \geqslant 32T^5k/\epsilon'^3$. Let π' be a subpermutation of π of size T picked uniformly at random. Then, with probability at least $1 - \epsilon'$, some k-blow-up of π' is in π.*

Proof. We consider the permutation as a collection of points represented in the unit square $[0, 1]^2$. Let $t = (4T^2)/\epsilon'$. We equally partition $[0, 1]^2$ into t columns and t rows, thus each of the $t \times t$ squares is of size $1/t \times 1/t$. We say a square is dense if it contains at least δn points of π, where $\delta = \epsilon'^3/(32T^5)$. Otherwise we say the square is sparse.

For the samples of size T, the probability that those samples are in different columns is $t(t-1) \cdots (t - T + 1)/t^T$, where the nominator is the number of ways of assigning the T samples to columns if they are in different columns, and the denominator is an upper bound of the total number of ways which columns these T samples can be in. Therefore, the probability that at least two samples are in the same column is at most

$$1 - \frac{t(t-1) \cdots (t - T + 1)}{t^T} \leqslant 1 - \left(\frac{t - T + 1}{t}\right)^T \leqslant \frac{T(T-1)}{t}.$$

Similarly, the probability that at least two samples are in the same row is at most $(T(T-1))/t$. Thus, by the union bound, the probability that there are two samples in the same column or row is at most $(2T(T-1))/t$.

For a sample point, the probability that it lies in a sparse square is at most δt^2 since each sparse square has at most δn points in π and there are at most t^2 sparse regions. Thus by the union bound, the probability that at least one of the samples is in the sparse region is at most $\delta t^2 T$.

Combining the results above, the probability that the T samples are in different columns and different rows and all the samples are in dense squares is at least

$$1 - \frac{2T(T-1)}{t} - \delta T t^2 \geqslant 1 - \epsilon'.$$

The inequality holds by our choice of δ and t. We thus now assume the T sample points are in different dense squares and they are in different rows and columns. However, in each of the dense squares that the samples are in, there are at least $\delta n \geqslant k$ points of π in that square. By picking k points of π from each of these squares give us a k-blow-up of the T-sample π'. Therefore a k-blow-up of π' is in π. $\square$

With Lemma 3.2 we can prove the first assertion of Theorem 3.1.

Proof of the first assertion in Theorem 3.1. Recall the definition of the blow-up parameter $k^*(\cdot)$ discussed in Definition 1.6.

Let $q = k^*(M)$, which is finite. Let π' be the M-subpermutation of π picked uniformly at random. Applying Lemma 3.2 with $k = q, \epsilon' = \epsilon, T = M$, we know as long as $n \geqslant 32M^5q/\epsilon^3$, with probability at least $1 - \epsilon$, the M-sample π' has a q-blow-up in π. Since $\pi \in \mathcal{P}$ and $\mathcal{P}$ is hereditary, this q-blow-up of π', which is a subpermutation of π, is in $\mathcal{P}$. However, if our algorithm outputs 'π is not in $\mathcal{P}$', it means there exists a least positive integer ℓ such that no ℓ-blow-up of π' is in $\mathcal{P}$. By our definition of $k^*(M)$, we have $l \leqslant k^*(M) = q$. But we have just found a q-blow-up of π' in $\mathcal{P}$, which contradicts $\mathcal{P}$ is hereditary. $\qquad\square$

3.2. When π is ϵ-far from satisfying $\mathcal{P}$

We now prove the second part of Theorem 3.1. We want to show that if $\rho_{\mathrm{EMD}}(\pi, \pi_0) \geqslant \epsilon$ for all $\pi_0 \in \mathcal{P}$, then with large probability, our algorithm will output 'π is not in $\mathcal{P}$'.

Proof outline. The proof idea is as follows. We view the permutation π as a collection of points in $[0, 1]^2$. We carry out a 'partition procedure', later described as Algorithm 2, to partition $[0, 1]^2$ into dyadic squares (of possibly different sizes), where some squares are dense (*i.e.* informally containing many points from π) and some squares are sparse (*i.e.* informally containing only few points from π). With large probability, the M-sample will hit each of the dense squares (Lemma 3.4).

Roughly speaking, if there are many squares that are dense, then π contains many copies of σ, the smallest permutation not in $\mathcal{P}$. Thus, with large probability, the M-sample will contain a copy of σ. This is proved in Lemma 3.5. Hence, our algorithm will output 'π is not in $\mathcal{P}$'.

Therefore, we can restrict our attention to the case where not many squares are dense. In this case, we show that with large probability, the M-sample π' contains a subpermutation α which has a blow-up $\tilde{\alpha}$ that is close to π. In fact, $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}) < \epsilon/2$ (Lemma 3.9). If our algorithm outputs 'π is in $\mathcal{P}$', then for any k, there is some k-blow-up of π' in $\mathcal{P}$. Since $\mathcal{P}$ is hereditary, there is some blow-up of α, call it $\tilde{\alpha}'$, which has the same block sizes as $\tilde{\alpha}$, and $\tilde{\alpha}' \in \mathcal{P}$. But $\rho_{\mathrm{EMD}}(\tilde{\alpha}, \tilde{\alpha}') < \epsilon/2$ (Lemma 3.10). Thus π is within distance ϵ from $\tilde{\alpha}' \in \mathcal{P}$ (Lemma 3.12). This contradicts the fact that π is ϵ-far from $\mathcal{P}$ under ρ_{EMD}.

We first partition $[0, 1]^2$ into squares of different types depending on the number of points of the permutation in the square.

Dyadic partition procedure. This is done through a dynamic procedure in different levels, starting from level 0. Throughout the procedure, we maintain and update three sets of squares, called ACTIVE, MATURE and FROZEN, which roughly correspond to squares which are dense, modestly dense, and sparse, respectively. From now on, let

$$\delta = \frac{\epsilon}{512c(\sigma)\log(32/\epsilon)} \tag{3.1}$$

be the density parameter for a square. Here density means the fraction of the points of π that is in this square. (Notice here we do not normalize the number of points of π by the area of the square.) A square in ACTIVE means it has density at least δ (thus dense); a

square in FROZEN means it has density less than δ (thus sparse). We will specify MATURE soon. Figure 8 illustrates this procedure and should help the reader to understand it.

To initialize the procedure, which begins at level 0, we put $[0,1]^2$ into the set ACTIVE since the density is $1 > \delta$. Initially, the sets MATURE and FROZEN are empty.

In level i, for $i \geqslant 1$, we consider squares at level i, working only with squares in ACTIVE which are of side length $1/2^{i-1}$. This procedure will maintain that at the beginning of level i, all the squares in ACTIVE are of side length $1/2^{i-1}$, and at the end of level i, all squares in ACTIVE are of side length $1/2^i$. We remove a square R of side length $1/2^{i-1}$ from ACTIVE, and subdivide it into four *children squares* R_1, R_2, R_3, R_4, each having side length $1/2^i$. We say R is the *parent square* of $R_i, i = 1, 2, 3, 4$. For any of those four squares, it is added to ACTIVE if it has density at least δ, and is added to FROZEN if otherwise. Furthermore, if all those four squares are added to FROZEN, then we add R itself to MATURE. This is roughly saying that R is dense, since it was once in ACTIVE, but it is not very dense, since it has density at most 4δ as all of its four children squares are added to FROZEN. After processing all the squares in ACTIVE of side length $1/2^{i-1}$, we move to the next level.

This procedure stops when ACTIVE is empty, or after level $\lceil \log(64/\epsilon) \rceil$. This means the smallest possible squares in the three sets have side length at least

$$d = 2^{-\lceil \log(64/\epsilon) \rceil}. \tag{3.2}$$

We have $\epsilon/128 < d \leqslant \epsilon/64$.

In summary, the procedure is to subdivide parent squares which are dense into children squares, and subdivide its children squares further only if some children squares are still dense. We never subdivide squares in FROZEN. Eventually we obtain a partition of $[0,1]^2$ by squares in ACTIVE and FROZEN. Notice that if a square is labelled as frozen, then its parent square is either mature or once active.

This procedure is also presented in the pseudocode in Algorithm 2. Based on this procedure, we obtain the following simple observations.

Lemma 3.3.

(a) *The squares in* ACTIVE, FROZEN *form a partition of* $[0,1]^2$.

(b) *Each square in* ACTIVE *is disjoint from each square in* MATURE.

(c) *At the end of each level, all squares in* ACTIVE *have the same side length. If* ACTIVE $\neq \emptyset$ *when the procedure stops, then all squares in* ACTIVE *have side length* d.

(d) *If* $R \in$ FROZEN, *then the parent square of* R *has density at least* δ.

(e) *If* $R \in$ MATURE, *then its density is at most* 4δ; *if* $R \in$ ACTIVE, *then its density is at most* d.

Proof.   Items (a), (b), (d) are direct consequences of the procedure.

To see (c), notice that if R is active and has side length greater than d, then the procedure has not stopped yet; thus it should have been subdivided.

To see (e), notice that R is in MATURE only if none of its four children squares have density at least δ. Therefore R has density at most 4δ. If $R \in$ ACTIVE when the procedure stops, then by (c) it has side length d, but since π is a permutation, it means R contains at most dn points.   □

Algorithm 2 Dyadic partition procedure

1: **Input:** The $[0,1]^2$ representation of π.
2: **Output:** Three sets of squares ACTIVE, FROZEN, MATURE.
3: ACTIVE $= \{[0,1]^2\}$.
4: MATURE $= \emptyset$.
5: FROZEN $= \emptyset$.
6: **procedure** PARTITION $[0,1]^2$
7:     **for** Levels $i = 1,2,\ldots,\log(1/d)$ **do**
8:         **if** ACTIVE $= \emptyset$ **then**
9:             **Stop**
10:         **else**
11:             **for** all squares R in ACTIVE with side length $1/2^{i-1}$ **do**
12:                 ACTIVE $=$ ACTIVE $\setminus \{R\}$.
13:                 Split R into four equal sized squares $R_1,\ldots,R_4$, each of side length $1/2^i$.
14:                 **if** None of R_1,R_2,R_3,R_4 has density at least δ **then**
15:                     FROZEN $=$ FROZEN $\cup \{R_1,R_2,R_3,R_4\}$.
16:                     MATURE $=$ MATURE $\cup \{R\}$.
17:                 **else**
18:                     ACTIVE $=$ ACTIVE $\cup \{R_i, i = 1,2,3,4$ which has density at least $\delta\}$;
19:                     FROZEN $=$ FROZEN $\cup \{R_i, i = 1,2,3,4$ which has density less than $\delta\}$;
        return ACTIVE, FROZEN, MATURE.

Let RICH be the set of squares at the end of the procedure containing many points, defined as RICH $=$ ACTIVE $\cup$ MATURE. By definition, any square in RICH has density at least δ. In this way we have identified regions in $[0,1]^2$ which contain many points in π (the squares in the set RICH). We claim that with large probability, the M-sample of π hits each of these dense regions.

Lemma 3.4. *With probability at least $1 - \epsilon$, every square $R \in$ RICH contains a sample point.*

Proof. There are at most $(1/d)^2 = d^{-2}$ squares in RICH since each square has side length at least d and the squares in RICH are disjoint. For each $R \in$ RICH, it contains at least δn points of π. The probability that none of the M samples are in R is bounded above by $(1 - \delta)^M$. Thus, by the union bound, the probability that there is a dense square in RICH which is not hit by any of the M samples is bounded above by

$$d^{-2}(1-\delta)^M \leqslant d^{-2}e^{-\delta M}.$$

However, recall that

$$M = \frac{2000 c(\sigma)\log^2(2/\epsilon)}{\epsilon},$$

and the values of δ, d are determined in (3.1) and (3.2); by a simple computation, $d^{-2}e^{-\delta M} < \epsilon$, since $M > -\ln(\epsilon d^2)/\delta$ by our choices of M, d, δ. $\quad\square$

Level 0   Level 1

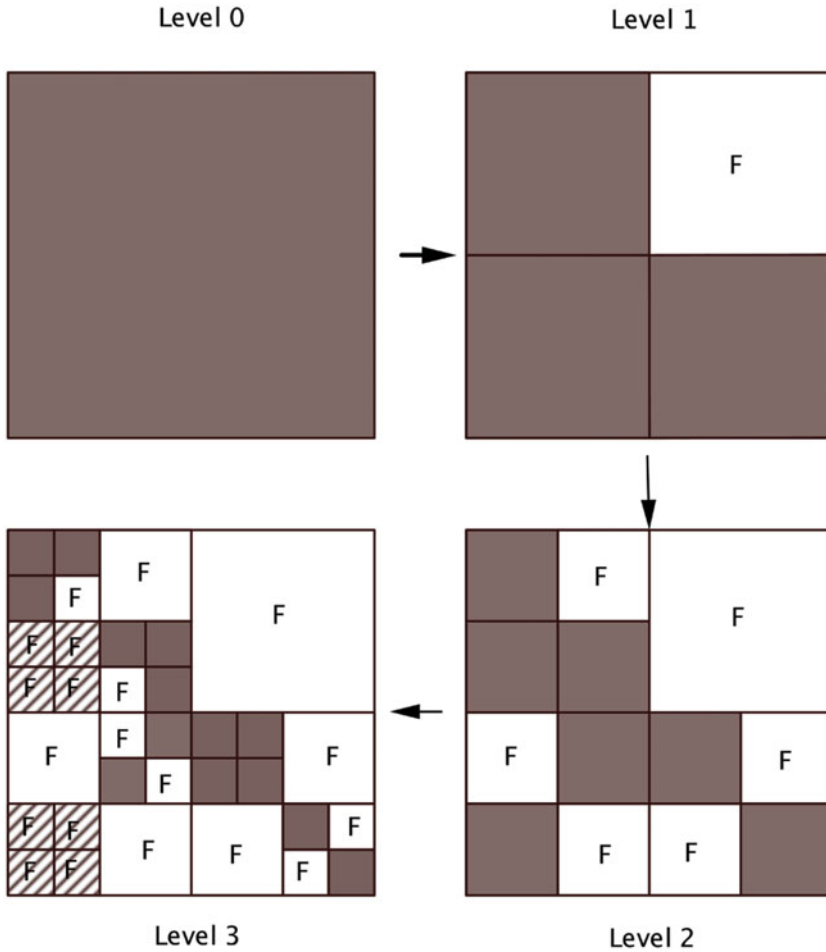Level 3   Level 2

Figure 8. Illustration of the dyadic partition procedure. We provide an example of the first four levels of the partition procedure. The squares in ACTIVE that resulted in each level are denoted by the black squares. At each level, the active squares in ACTIVE splits. The squares in FROZEN at each level are denoted by the white squares with label 'F' (stands for 'Frozen'). The squares in MATURE are given by the squares with the parallel lines filled in and each such square consists of four frozen squares. In this example, MATURE is empty at levels 0,1,2. After finishing the algorithm at level 3, there are two squares in MATURE.

Thus we know with probability at least $1 - \epsilon$ that each square in RICH contains a sample point. We pick one sample point from each of the squares in RICH (if there are multiple samples in one square, we arbitrarily pick one). These samples of π induce a permutation α of length $|\text{RICH}|$. For any $1 \leqslant i \leqslant |\alpha|$, define j_i to be such that the ith element (from left to right) of α corresponds to the sample $(j_i, \pi(j_i))$ of π. Thus $1 \leqslant j_1 < \cdots < j_{|\alpha|} \leqslant n$. Furthermore, $\alpha(i) > \alpha(i')$ if and only if $\pi(j_i) > \pi(j_{i'})$. Figure 9(a) shows an example of the M-sample in π, and Figure 9(b) shows α; in this example, $\alpha = 4213$.

By Lemma 3.3, the parent of each frozen square was once active or mature. If there are many frozen squares of the same size, then we get many parent squares of the same size which are dense. The next lemma shows that if this is the case, our algorithm will
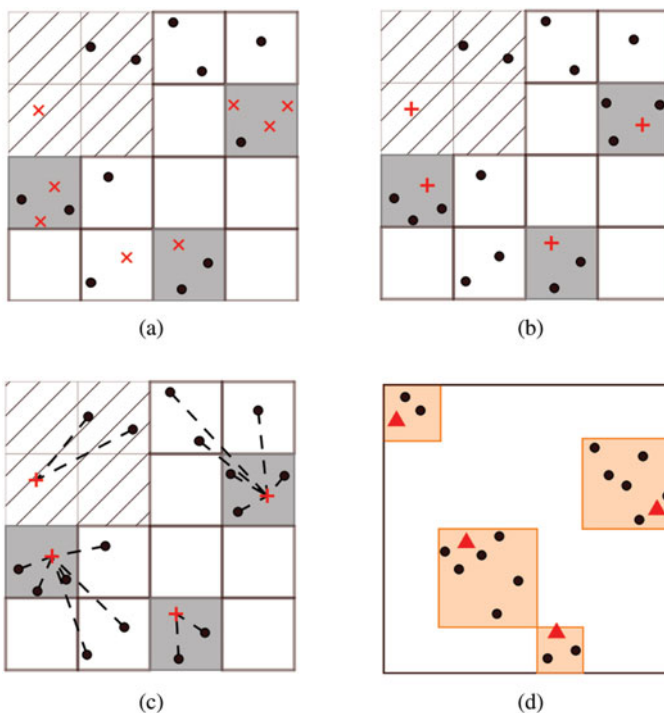
Figure 9. Obtaining a blow-up of a subpermutation of the *M*-sample that is close to π. The black and red points are points in π. In this example $|\pi| = 20$. In (a), (b) and (c), the squares are the results of the partition procedure. The squares in grey are the squares in ACTIVE. The squares in white and the four smaller squares with stripes are in FROZEN. The larger striped square, which consists of four smaller striped squares, is in MATURE. In this example $|\text{ACTIVE}| = 3$, $|\text{FROZEN}| = 13$, $|\text{MATURE}| = 1$. (a) The red crosses are the *M* sample points; each region in RICH contains a sample point. (b) The permutation α consists of one sample point from each square in RICH, given by the red '+'; if one square in RICH contains multiple sample points, arbitrarily pick one. We ignore the sample points not contained in squares in RICH. In this example, the permutation α induced by the subset of sample points is 4213. (c) Map ϕ maps each point in π to a point in α. The dashed line connects the point of α that a point of π is mapped to. A point is mapped to a point of α sharing the same parent square. (d) The permutation $\tilde{\alpha}$. Notice that in the *t*th blow-up block, the subpermutation is the one induced by the points in π which are mapped to the *t*th element of α.

output 'π is not in $\mathcal{P}$' with large probability. This lemma implicitly uses the Marcus–Tardos theorem [26] discussed in the Introduction, which verifies the existence of the Füredi–Hajnal constant.

Lemma 3.5. *Let $C = 2c(\sigma)$ where $c(\sigma)$ is the Füredi–Hajnal constant. If there is a positive integer j such that there are at least $4C \cdot 2^{j-1}$ squares in* FROZEN *with side length $1/2^j$, then with probability at least $1 - \epsilon$, our algorithm will output 'π is not in $\mathcal{P}$'.*

Proof. Let the set of squares in FROZEN with side length $1/2^j$ be $\mathcal{F}_j$. For each square in $\mathcal{F}_j$, its parent square is either in MATURE or it was once in ACTIVE. Consider the set of all these parent squares; call the set $\mathcal{S}$. Thus $|\mathcal{F}_j| \leqslant 4|\mathcal{S}|$ since each parent square has four children squares. Thus $|\mathcal{S}| \geqslant C \cdot 2^{j-1}$.

Now we consider $[0, 1]^2$ as being partitioned into $(2^{j-1})^2$ regions where each region is a square with side length $1/2^{j-1}$. The parent squares in $\mathcal{S}$ are some of those regions. Each region corresponding to a parent square in $\mathcal{S}$ has density at least δ. Recall $s = |\sigma|$ and $C = 2c(\sigma)$. Since $|\mathcal{S}| \geqslant C \cdot 2^{j-1}$, by the Marcus–Tardos theorem, we can find s regions corresponding to s squares in $\mathcal{S}$ that are in different columns and rows. Furthermore, the points of π in these s regions form a blow-up of σ.

Furthermore, if a square in $\mathcal{S}$ is not in MATURE, then there must be a square in MATURE $\cup$ ACTIVE which is a subset of this square. Thus by Lemma 3.4, with probability at least $1 - \epsilon$, these s regions each contain a sample point. This means, with probability at least $1 - \epsilon$, σ is a subpermutation of the sampled π'. However, since $\sigma \notin \mathcal{P}$, the 1-blow-up of π' is not in $\mathcal{P}$. Since each permutation has exactly one 1-blow-up, which is itself, Algorithm 1 will output 'π is not in $\mathcal{P}$'.   □

From the previous lemma, it suffices to only further consider the case where there are not many frozen squares in any level. Formally, for each j, FROZEN has at most $4C \cdot 2^{j-1}$ squares of side length 2^{-j}. We have already obtained α, a subpermutation of π'. The definition of α is right after Lemma 3.4, and recall α is induced by the points $(j_t, \pi(j_t)), t = 1, 2, \ldots, |\alpha|$. We now want to construct a blow-up of α – call it $\tilde{\alpha}$ – such that $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$ is small.

To define a blow-up of a given permutation, it suffices to define the length and shape of the subpermutation within each of the blow-up blocks. In the case of $\tilde{\alpha}$, there are in total $|\alpha|$ blow-up blocks. The high-level idea for creating an $\tilde{\alpha}$ with small distance $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$ is as follows. Recall that to bound $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$, we find a matching $\psi : [n] \to [n]$ telling us that the ith point of π is matched with the $\psi(i)$th point in $\tilde{\alpha}$.

We first define a map $\phi : [n] \to \{j_t : 1 \leqslant t \leqslant |\alpha|\}$ which associates with each point of π one of the sampled points in π that make up α. This is shown in Figure 9(c), in which the dashed line indicates where a point of π is mapped to. We will use the map ϕ to determine ψ and $\tilde{\alpha}$. The map ϕ will have the property that the sum of the (taxicab) distance between points in π and the sampled points they are mapped to is small. Also, if the ith point of π is associated with the tth point in α under ϕ, then $\psi(i)$ is in the tth blow-up block in $\tilde{\alpha}$. Thus ϕ determines the size of each blow-up block, which for the tth blow-up block is $|\phi^{-1}(j_t)|$. To finish the construction of $\tilde{\alpha}$, simply let the subpermutation in the tth blow-up block be the permutation induced by the subpermutation of π restricted to the positions in $\phi^{-1}(j_t)$. This is shown in Figure 9(d). To complete the construction of ψ, if $\psi(i), \psi(i')$ are mapped to the same blow-up block, let $\psi(i) < \psi(i')$ if and only if $i < i'$. We will prove that the matching ψ shows that the earth mover's distance $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$ is small.

We now go through the details of this argument. First we define ϕ. Let

$$\phi : [n] \to \{j_i, i = 1, 2, \ldots, |\alpha|\}$$

be such that if $\phi(i) = j_t$, we associate $(i, \pi(i))$ by ϕ to the tth point in α. The association rule is as follows.

(1) If the ith point of π is in a square in ACTIVE, and $(j_t, \pi(j_t)) \in \pi$ is the sample point in that square chosen to contribute to α, we define $\phi(i) = j_t$.

(2) If the ith point of π is in some sparse square in FROZEN, it means that its parent square was once active. If the parent square is in MATURE, then let $(j_t, \pi(j_t)) \in \pi$ be the unique point in that square chosen to be an element of α and define $\phi(i) = j_t$. If the parent square is not in MATURE, then after subdividing the parent square into four squares, one of them, R, is active. R will further subdivide, and at the end of the partition procedure, either R is mature, or it contains a square that is mature or active. Arbitrarily pick one such square that is mature or active, and let $(j_t, \pi(j_t)) \in \pi$ be the unique element of α in that square and define $\phi(i) = j_t$.

This association rule is illustrated in Figure 9(c). We thus find for the ith point in π a target point $(\phi(t), \pi(\phi(t)))$ of α.

For the point $(j_t, \pi(j_t)) \in \pi$ which gives the tth element in α, let $P_t = \phi^{-1}(j_t)$. Thus P_t tells us which points in π are mapped to the tth point in α. We create a blow-up $\tilde{\alpha}$ of α in the following way: the tth point of α blows up into a block B_t, where $|B_t| = |\phi^{-1}(j_t)| = |P_t|$ and the permutation restricted to B_t in $\tilde{\alpha}$ is the subpermutation of π restricted to the positions in P_t, that is, $\{(v, \pi(v)), v \in P_t\}$. By construction, $\tilde{\alpha}$ is of length n. Figures 9(c) and 9(d) illustrate how $\tilde{\alpha}$ is obtained.

Summarizing, we found a mapping of the points of π to the points of a subpermutation α. We naturally formed a blow-up $\tilde{\alpha}$ of α. We want to show $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$ is small. Recall that by the definition of the earth mover's distance, it suffices to find a bijection between points in π and points in $\tilde{\alpha}$ for which the sum of the L_1 distances between the point in π and the points in $\tilde{\alpha}$ that they map to is small.

We choose the bijection ψ as follows. For any $i \neq j \leqslant n$, $\psi(i) < \psi(j)$, if and only if one of the following situation holds: (1) $\phi(i) < \phi(j)$, or (2) $\phi(i) = \phi(j)$ but $i < j$. Equivalently, for each i (corresponding to the ith point in π), let $(\psi(i), \tilde{\alpha}(\psi(i)))$ lie in the block $B_{\phi(i)}$. And if among all the points in $\phi^{-1}(\phi(i))$, i is the tth smallest, then $\psi(i)$ is the tth point (from left to right) in $B_{\phi(i)}$. In other words, $\psi(i) = |B_1| + \cdots + |B_{\phi(i)-1}| + t$.

To show $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha})$ is small, it suffices to show that

$$\sum_{i=1}^n d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n}\right), \left(\frac{\psi(i)}{n}, \frac{\tilde{\alpha}(\psi(i))}{n}\right)\right)$$

is small, where $d((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$ is the L_1 distance. The next lemma shows that this is within a factor of two of the sums of the L_1 distance between the points in π and the points in α that they map to. This in turn is easier to bound as the L_1 distance between two points in the same axis-parallel square is at most twice the side length of the square.

Lemma 3.6.

$$\sum_{i=1}^n d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n}\right), \left(\frac{\psi(i)}{n}, \frac{\tilde{\alpha}(\psi(i))}{n}\right)\right) \leqslant 2 \sum_{i=1}^n d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n}\right), \left(\frac{\phi(i)}{n}, \frac{\pi(\phi(i))}{n}\right)\right). \qquad \square$$

We first prove the following useful claim.

Claim 3.7 (generalized Spearman's footrule). *Let $\gamma(i) : [n] \to [n]$ be any map. Define a bijection $\beta : [n] \to [n]$ as follows. We have $\beta(i) < \beta(j)$ if and only if (1) $\gamma(i) < \gamma(j)$ or (2) $\gamma(i) = \gamma(j)$ and $i < j$.*

Let $D(\beta) = \sum_{i=1}^{n} |i - \beta(i)|$, which is Spearman's footrule distance between the identity permutation and the permutation β. Let $D(\gamma) = \sum_{i=1}^{n} |i - \gamma(i)|$, which is a footrule distance between the identity permutation and the mapping γ. Then

$$D(\beta) \leqslant 2D(\gamma).$$

Proof of Claim 3.7. For each integer $1 \leqslant j \leqslant n$, let γ_j^+ be the number of integers i such that $i \leqslant j < \gamma(i)$. Similarly, let γ_j^- be the number of i such that $\gamma(i) \leqslant j < i$. Define β_j^+, β_j^- similarly.

We first show that $D(\gamma) = \sum_{j=1}^{n} \gamma_j^+ + \gamma_j^-$. We use a double-counting argument. Note that if $i \leqslant \gamma(i)$, then the number of j for which i contributes one to γ_j^+ is precisely $|i - \gamma(i)|$, and the number of j for which i contributes to γ_j^- is zero. Similarly, if $i > \gamma(i)$, then the number of j for which i contributes one to γ_j^- is $|i - \gamma(i)|$, and the number of j for which i contributes to γ_j^+ is zero. Therefore,

$$D(\gamma) = \sum_{i:i<\gamma(i)} |i - \gamma(i)| + \sum_{i:i>\gamma(i)} |i - \gamma(i)| = \sum_{j=1}^{n} \gamma_j^+ + \sum_{j=1}^{n} \gamma_j^-.$$

Similarly, we have

$$D(\beta) = \sum_{j=1}^{n} \beta_j^+ + \beta_j^-.$$

To prove $D(\beta) \leqslant 2D(\gamma)$, it suffices to show for any j, $\beta_j^+ \leqslant \gamma_j^+ + \gamma_j^-$ and $\beta_j^- \leqslant \gamma_j^+ + \gamma_j^-$. The proofs of both inequalities are essentially the same argument, so we only show the first. If i contributes one to β_j^+, then $i \leqslant j < \beta(i)$. If $\gamma(i) > j$, then i also contributes one to γ_j^+. Otherwise, $i, \gamma(i) \leqslant j < \beta(i)$. Since β is a permutation, for the set of such i, since $i \leqslant j < \beta(i)$, we can match each such i with a unique i' such that $\beta(i') \leqslant j < i'$. However, since $\beta(i') < \beta(i)$ and $i' > i$, we must have $\gamma(i') < \gamma(i)$, thus $\gamma(i') < j < i'$, which implies that i' contributes one to γ_j^-. We thus get $\beta_j^+ \leqslant \gamma_j^+ + \gamma_j^-$. Similarly, $\beta_j^- \leqslant \gamma_j^+ + \gamma_j^-$. Summing over all j, we get $D(\beta) \leqslant 2D(\gamma)$. $\qquad\square$

Proof of Lemma 3.6. It suffices to prove the following two inequalities:

$$\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\psi(i)}{n} \right| \leqslant 2 \sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi(i)}{n} \right|,$$

$$\sum_{i=1}^{n} \left| \frac{\pi(i)}{n} - \frac{\tilde{\alpha}(\psi(i))}{n} \right| \leqslant 2 \sum_{i=1}^{n} \left| \frac{\pi(i)}{n} - \frac{\pi(\phi(i))}{n} \right|.$$

To prove the first inequality, we, apply the claim with $\gamma(i) = \phi(i)$ and $\beta(i) = \psi(i)$. By the definition of ψ which is $\psi(i) < \psi(j)$ if and only if (1) $\phi(i) < \phi(j)$, or (2) $\phi(i) = \phi(j)$ but $i < j$, the condition of the claim holds. For the second inequality, we apply the claim

with $\gamma(i) = \pi \circ \phi \circ \pi^{-1}(i)$ and $\beta(i) = \tilde{\alpha} \circ \psi \circ \pi^{-1}(i)$. Similarly, by the definition of $\tilde{\alpha}$, the conditions for the claim holds. $\square$

Lemma 3.8. *Let $\mathcal{P}$ be a proper hereditary family of permutations, let σ be a smallest permutation not in $\mathcal{P}$, and $s = |\sigma|$. Let $C = 2c(\sigma)$, where $c(\sigma)$ is the Füredi–Hajnal constant for σ. Assume, for permutation π, that for each j, there are at most $4C \cdot 2^{j-1}$ frozen squares of side length 2^{-j}. Further suppose that the random subpermutation α of π contains at least one point in each rich square. Then*

$$\sum_{i=1}^{n} d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n}\right), \left(\frac{\phi(i)}{n}, \frac{\pi(\phi(i))}{n}\right)\right) \leqslant 2dn + 8\log(1/d)C\delta n.$$

Proof. For any point in an active square, the L_1 distance between it and the point of α it maps to is at most $2d$ since they are both in the same active square, which has side length d by Lemma 3.3. There are at most n points of π which are in active squares. Thus, the sum of the distances coming from points in active squares is at most $2dn$.

For a point $(i/n, \pi(i)/n)$ in a frozen square $R \in \text{FROZEN}$, suppose the side length of the square is $1/2^j$. The L_1 distance from it to the point in α it associates with is at most $2 \cdot 1/2^{j-1}$ since these two points are in the same parent square whose side length is $1/2^{j-1}$. However, by assumption there are at most $4C \cdot 2^{j-1}$ frozen squares of side length 2^{-j}. Therefore, for each j, the sum of L_1 distances coming from points in frozen squares of side length $1/2^j$ is at most $(4C \cdot 2^{j-1}) \cdot \delta n \cdot 2/2^{j-1}$, where $(4C \cdot 2^{j-1}) \cdot \delta n$ gives an upper bound for the number of points in frozen squares of side length $1/2^j$ and $2/2^{j-1}$ is an upper bound for the L_1 distances between a point in π and its image in α. Therefore the sum of the distances coming from points in frozen squares of different side lengths is at most

$$\sum_{j=1}^{\log(1/d)} 8C\delta n = 8\log(1/d)C\delta n.$$

Combining the distances from points in active squares and points in frozen squares, we have that

$$\sum_{i=1}^{n} d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n}\right), (a_{\phi(i)}, b_{\phi(i)})\right) \leqslant 2dn + 8\log(1/d)C\delta n. \qquad \square$$

Lemmas 3.6 and 3.8, together with the values of δ, d as given in (3.1) and (3.2), give the following lemma.

Lemma 3.9. *Assume that for each j, there are at most $4C \cdot 2^{j-1}$ frozen squares of side length 2^{-j}, and the sample contains at least one point in each rich square. Then*

$$\rho_{EMD}(\pi, \tilde{\alpha}) \leqslant 2(2d + 8\log(1/d)C\delta) < \epsilon/2. \qquad \square$$

We have shown $\tilde{\alpha}$ and π are close in the earth mover's distance. We now show that $\tilde{\alpha}$ is close in the earth mover's distance to any other blow-up of α, where the jth point in α blows up into a block of size $|\phi^{-1}(j)| = |B_j|$. Thus π and any of these blow-ups are close.

Lemma 3.10. *Let $\tilde{\alpha}'$ be any blow-up of α where the jth point in α blows up into a block of size $|\phi^{-1}(j)| = |B_j|$. Then*

$$\rho_{EMD}(\tilde{\alpha}', \tilde{\alpha}) \leqslant \epsilon/2. \qquad \square$$

Before proving Lemma 3.10, we first obtain the following useful claim.

Claim 3.11.

$$\delta n \leqslant |B_j| \leqslant 3\delta n \log(1/d) + \max(dn, 4\delta n).$$

Proof. By definition, $B_j = \phi^{-1}(j)$. Since $(t_j, \pi(t_j))$ is in an active or mature square, and all the other points in this square are mapped to it by ϕ, we have $|B_j| \geqslant \delta n$.

Given any mature or active square R of side length l, by the partition procedure, we know for each possible length $l' > l$, there are at most three frozen squares of side length l' whose points are mapped to a sample in R. Since there are at most $\log(1/l)$ possible side lengths of squares greater than l, the number of points coming from the frozen squares mapped to the sample point in R is at most $\delta n \cdot (3 \log(1/l)) \leqslant 3\delta n \log(1/d)$.

Furthermore, by Lemma 3.3, the number of points in R is at most $\max(dn, 4\delta n)$. The points in R and points in a frozen square mapped to R are all the possible points mapping to R. Thus

$$|\phi^{-1}(j)| \leqslant 3\delta n \log(1/d) + \max(dn, 4\delta n). \qquad \square$$

Proof of Lemma 3.10. By the assumption, we know $\tilde{\alpha}, \tilde{\alpha}'$ have the same corresponding blow-up block sizes. To get an upper bound for $\rho_{EMD}(\tilde{\alpha}', \tilde{\alpha})$, we simply pick the identity permutation θ given by $\theta(i) = i$ for $i \in [n]$. Thus we would like to upper-bound

$$\sum_{i=1}^{n} d((i/n, \tilde{\alpha}(i)/n), (i/n, \tilde{\alpha}'(i)/n)) = \sum_{i=1}^{n} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n|.$$

Notice that if the ith point in $\tilde{\alpha}$ is in the jth blow-up block, then the ith point in $\tilde{\alpha}'$ is also in the jth block. Thus $|\tilde{\alpha}(i) - \tilde{\alpha}'(i)| \leqslant |B_j|$. This gives us

$$\sum_{i=1}^{n} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n| = \sum_{j=1}^{|\alpha|} \left(\sum_{i:(i,\tilde{\alpha}(i)) \in B_j} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n| \right) \leqslant \sum_{j=1}^{|\alpha|} |B_j|^2/n. \qquad (3.3)$$

We have

$$\sum_{j=1}^{|\alpha|} |B_j| = n \quad \text{and} \quad 0 \leqslant |B_j| \leqslant 3\delta n \log(1/d) + \max(dn, 4\delta n).$$

To maximize $\sum_{j=1}^{|\alpha|} |B_j|^2/n$ with these constraints, we want the $|B_j|$ as close to the extreme values as possible. This gives

$$\sum_{j=1}^{|\alpha|} |B_j|^2/n \leqslant (3\delta \log(1/d) + \max(d, 4\delta))n,$$

where we assign $n/(3\delta n \log(1/d) + \max(dn, 4\delta n))$ number of $|B_j|$ to achieve the maximum possible value $3\delta n \log(1/d) + \max(dn, 4\delta n)$, and the rest are set to be 0.

Therefore, we have

$$
\begin{aligned}
\rho_{\text{EMD}}(\tilde{\alpha}', \tilde{\alpha}) &\leqslant \frac{1}{(n-1)/2} \sum_{i=1}^{n} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n| \\
&\leqslant \frac{n}{(n-1)/2} (3\delta \log(1/d) + \max(d, 4\delta)) \leqslant \epsilon/2.
\end{aligned}
$$

The first inequality is by the definition of ρ_{EMD}, and the last inequality is by our choices of δ, d as in (3.1) and (3.2). $\qquad\square$

Combining the previous two lemmas, we have the following lemma.

Lemma 3.12. *Suppose the subpermutation α contains at least one point in each rich square. Let $\tilde{\alpha}'$ be any blow-up of α such that the jth point in α blows up into a block of size $|\phi^{-1}(j)| = |B_j|$. Assume that for each j, there are at most $4C \cdot 2^{j-1}$ frozen squares of side length 2^{-j}. Then*

$$\rho_{EMD}(\pi, \tilde{\alpha}') < \epsilon.$$

Proof. This is a direct consequence of Lemmas 3.10 and 3.9 and the triangle inequality:

$$\rho_{\text{EMD}}(\pi, \tilde{\alpha}') \leqslant \rho_{\text{EMD}}(\pi, \tilde{\alpha}) + \rho_{\text{EMD}}(\tilde{\alpha}, \tilde{\alpha}') < \epsilon. \qquad\square$$

Assuming Lemma 3.12, we show that the second assertion of Theorem 3.1 holds.

Proof. We have obtained that with probability at least $1 - \epsilon$, we can obtain a permutation α being a subpermutation of the sample permutation π', and α hits each of the dense squares in RICH once. If our algorithm outputs '$\pi \in \mathcal{P}$', it implies that there is an n-blow-up of α in $\mathcal{P}$. Therefore we can find a subpermutation $\tilde{\alpha}'$ of the n-blow-up, where $\tilde{\alpha}'$ is of length n, and it is a blow-up of α where the jth point of α blows up into a block of size $|B_j|$. (Notice that this $\tilde{\alpha}'$ may be different from the $\tilde{\alpha}$ defined right before Lemma 3.6.) Since $\mathcal{P}$ is hereditary, $\tilde{\alpha}' \in \mathcal{P}$. However, by Lemma 3.12, we know $\rho_{\text{EMD}}(\pi, \tilde{\alpha}') < \epsilon$. This contradicts the assumption that π is at distance at least ϵ from $\mathcal{P}$ and completes the proof. $\qquad\square$

4. Two-sided property testing under the planar metrics with universal bound

In this section we prove Theorem 1.8. The tester we use is the same as the two-sided tester in Algorithm 1, but with a different choice of M. The proof is similar to that of Theorem 1.7, but with a somewhat different analysis. When $\pi \in \mathcal{P}$, by Lemma 3.2 and the same argument as in the proof of Theorem 1.7, it can be seen that when n is sufficiently large (depending on $\mathcal{P}$), the algorithm outputs $\pi \in \mathcal{P}$ with probability at least $1 - \epsilon$.

We now handle the case when π is ϵ-far from $\mathcal{P}$. Let

$$M = 20000/\epsilon^2 \tag{4.1}$$

be the sample size. We will first prove the following lemma, which states that any permutation is very close to some blow-up of a sufficiently large sample with high probability.

Lemma 4.1. *Let π be a permutation of length n and M be as in (4.1). An M-sample chosen uniformly at random in π, given by the sample points $(j_t, \pi(j_t))$ for $t = 1, 2, \ldots, M$, induces a permutation α. Then, with probability at least $1 - \epsilon/2$, there exists a blow-up of α which is within distance $\epsilon/2$ (under ρ_{EMD}) with π.* ☐

Remark. The sample size $M = O(1/\epsilon^2)$ is tight. The following is the main idea of the proof. For a permutation of length n picked uniformly at random with $n \gg \epsilon^{-2} \log(1/\epsilon)$, if $M < 2^{-7}\epsilon^{-2}$, then a square of side length 8ϵ around a point of π with probability at least $2/3$ does not contain any sample point. Thus, we expect that most points have to move distance more than 2ϵ in any blow-up of α. ☐

To prove Lemma 4.1, we proceed similar to the proof of Theorem 1.7. Define a map $\phi : [n] \to [n]$ which maps each point of π to the closest (under the Euclidean distance[§] in $[0, 1]^2$) sample point. In other words, $\phi(i) = j$ is equivalent to the point $(i, \pi(i))$ of π being closest to the sample point $(j, \pi(j))$. As in the proof of Theorem 1.7, from ϕ we also define a permutation $\tilde{\alpha} : [n] \to [n]$ which is a blow-up of α. For each $t = 1, 2, \ldots, M$, let $P_t = \phi^{-1}(j_t)$; then the permutation restricted to the tth blow-up block B_t of $\tilde{\alpha}$ is the subpermutation of π restricted to the positions indicated in P_t. Clearly $|B_t| = |P_t|$. We also define a matching $\psi : [n] \to [n]$, such that for any $i \neq j \leqslant n, \psi(i) < \psi(j)$ if and only if one of the following holds: (1) $\phi(i) < \phi(j)$, or (2) $\phi(i) = \phi(j)$ and $i < j$. Thus ψ gives a matching between points in π and points in ψ. As before, we will show that with probability at least $1 - \epsilon/2$,

$$\rho_{\text{EMD}}(\pi, \tilde{\alpha}) < \epsilon/2, \tag{4.2}$$

which is a stronger statement of Lemma 4.1. Further, we will prove that with probability at least $1 - \epsilon/2$, if $\tilde{\alpha}'$ is another blow-up of α where the jth point of α blows up into a

---

[§] Note that one can also easily use the L_∞ or L_1 distance as they are Lipschitz-equivalent; the proofs are basically the same but it is easier to explain it with figures using the Euclidean distance.

block of size $|B_j|$ as well, then

$$\rho_{\mathrm{EMD}}(\tilde{\alpha}, \tilde{\alpha}') < \epsilon/2. \tag{4.3}$$

Inequalities (4.2) and (4.3) and the triangle inequality imply that $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}') < \epsilon$. Assuming we can prove that $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}') < \epsilon$, if the algorithm outputs '$\pi \in \mathcal{P}$', then there is an n-blow-up of α in $\mathcal{P}$. Therefore we can find a length-n subpermutation $\tilde{\alpha}'$ of this n-blow-up where the jth point of α blows up into a block of size $|B_t|$. Clearly $\tilde{\alpha}' \in \mathcal{P}$ by the hereditary property. Since we have assumed $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}') < \epsilon$, we reach a contradiction since we also assume $\rho_{\mathrm{EMD}}(\pi, \mathcal{P}) \geqslant \epsilon$.

Therefore we just need to prove two things: Lemma 4.1 or equivalently inequality (4.2), and (4.3). Note that as a consequence inequalities (4.2) and (4.3) will simultaneously hold with probability at least $1 - \epsilon$. The proof consists of two parts. The first part is to show inequality (4.2), and the second part is to show inequality (4.3). The first part is done through the dyadic partition procedure again.

4.1. $\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}) < \epsilon/2$

As in the proof of Theorem 1.7, we again conduct a dynamic procedure in different levels, starting from level 0, where in level i we look at squares of side lengths 2^{-i}. Recall that the density of a square is the fraction of points in the permutation in the square (so it is not the relative density with respect to the area of the square). Let

$$d = 2^{-\lceil \log(48/\epsilon) \rceil} \tag{4.4}$$

be the smallest side length of squares. Let $l_0 = \log(1/d)$ be the number of different levels. Let

$$\delta_1 = \sqrt{2}/96 \cdot \epsilon^2 \log(1/\epsilon), \quad \delta_2 = \epsilon^2/5200, \quad K = \lceil -\log(12 \cdot 16 \cdot \delta_1/\epsilon) \rceil, \tag{4.5}$$

where δ_1, δ_2 are two density thresholds: if $i \leqslant K$, then a square of side length 2^{-i} is in ACTIVE if it has density at least δ_1; if $i > K$, then a square of side length 2^{-i} is in ACTIVE if it has density at least δ_2. We go through the same procedure as in Algorithm 2, with the only difference being that the density threshold for a square to be in ACTIVE depends on its side length. In summary, throughout the procedure, repeatedly some parent square in ACTIVE splits into four children squares, and a child square splits if is dense with respect to the corresponding density threshold.

For each level $0 \leqslant i \leqslant \log(1/d)$, let FROZEN$_i$ be the set of squares in FROZEN which are of side length 2^{-i}. Let MATURE$_i$ be the set of squares in MATURE which are of side length 2^{-i}. Note that each square in MATURE$_i$ is a union of four squares in FROZEN$_{i+1}$. Let ACTIVE$_i$ be the squares of side length 2^{-i} that are once active. Clearly ACTIVE$_{l_0}$ = ACTIVE.

We will define a map $\phi' : [n] \to [n]$ which maps points of π to some sample point. First notice that ϕ is the map which maps each point of π to its closest sample point; thus

$$\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi(i))}{n} \right| \leqslant \sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi'(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi'(i))}{n} \right|. \tag{4.6}$$

We now define ϕ' such that we can bound the right-hand side of (4.6), and thus upper-bounds the left-hand side of (4.6) which is about ϕ.

For a point $(i, \pi(i))$ in π, let $1 \leqslant j \leqslant l_0$ be the largest integer such that there exists a square in ACTIVE_j containing both the point $(i, \pi(i))$ of π and some sample points. Let $\phi'(i) \in \{j_t, t = 1, 2, \ldots, M\}$ be the position of a closest such sample point to $(i, \pi(i))$. In other words, we find the smallest square which was once in ACTIVE containing this point and some sample points and ϕ' maps $(i, \pi(i))$ to a closest such sample point.

To analyse the right-hand side of (4.6), we first partition points in π into different groups. For each $1 \leqslant i \leqslant \log(1/d)$, we define a set of squares CANDIDATE_i as follows.

(1) If $0 \leqslant i < \log(1/d)$, CANDIDATE_i consists of squares in $\text{MATURE}_i \cup \text{FROZEN}_{i+1}$. (Notice $[0, 1]^2$ cannot be in MATURE.)

(2) If $i = \log(1/d)$, CANDIDATE_i consists of squares in ACTIVE, which by definition are of side length d. Denote the number of points of π in CANDIDATE_i as n_i. Thus $\sum_{i=0}^{\log(1/d)} n_i = n$.

The following lemma is easy to check from the definitions.

Lemma 4.2.

(a) *The squares in the sets* CANDIDATE_i *partition* $[0, 1]^2$.

(b) *Given* $1 \leqslant i \leqslant \log(1/d)$*, the points in squares in* $\bigcup_{i \leqslant j \leqslant \log(1/d)} \text{CANDIDATE}_j$ *are the same as the points in squares in* ACTIVE_i*. The points of* π *in squares in* CANDIDATE_i *are the points in squares in* ACTIVE_i *but not in squares in* $\bigcup_{j>i} \text{ACTIVE}_j$*.*

(c) *For each* $K + 1 \leqslant i < \log(1/d)$*,* $n_i \leqslant 4\delta_2 \cdot 2^{2i} n$*. For each* $0 \leqslant i \leqslant K$*,* $n_i \leqslant 4\delta_1 \cdot 2^{2i} n$*.*

Proof.   The first two assertions are clear from the definition of CANDIDATE_i. To show (c) holds, notice that when $i < \log(1/d)$, CANDIDATE_i consists of only squares in FROZEN_{i+1} or MATURE_i. If $i \leqslant K$, each dyadic square of side length 2^{-i} has at most $4\delta_1 n$ points of π coming from squares in MATURE_i or FROZEN_{i+1}, and there are 2^{2i} dyadic squares of side length 2^{-i}. Thus the bound is achieved. A similar argument applies to the case when $i > K$. $\square$

For a point in a square in CANDIDATE_i, we say that it is *resolved in level j* if j is the largest integer such that there is a square in ACTIVE_j containing this point and a sample point. Clearly by the definition of CANDIDATE_i, $j \leqslant i$. And by the definition of ϕ' defined above, if a point $(s, \pi(s))$ of π is resolved in level j, then

$$|s - \phi(s)| + |\pi(s) - \pi(\phi(s))| \leqslant |s - \phi'(s)| + |\pi(s) - \pi(\phi'(s))| \leqslant 2 \cdot 2^{-j}. \quad (4.7)$$

Let the random variables $N_{i,j}$ be the set of points in squares in CANDIDATE_i which are resolved in level j. Let $n_{i,j} = |N_{i,j}|$. Thus $\sum_{j: 0 \leqslant j \leqslant i} n_{i,j} = n_i$, the number of points of π in squares in CANDIDATE_i. Let NRESOLVE_j be the set of points in squares in $\bigcup_{l_0 \geqslant i \geqslant j} \text{CANDIDATE}_i$ which are not resolved in levels $l_0, l_0 - 1, \ldots, j$. Thus by the definition of NRESOLVE_j,

$$|\text{NRESOLVE}_j| = \sum_{j \leqslant s \leqslant l_0} \sum_{0 \leqslant t \leqslant j-1} n_{s,t}. \quad (4.8)$$

We first show that under some conditions, the quantity on the right-hand side of (4.6) is indeed small; then we show that with high probability, these conditions hold.

Lemma 4.3. *Let $c = 1/2$. Suppose the following conditions holds simultaneously.*

(i) *For each $i \leqslant K$, points in squares in CANDIDATE$_i$ are resolved in level i.*

(ii) *For each $i > K$, a point in a square in CANDIDATE$_i$ is resolved in level j with $j \geqslant K$.*

(iii) *For each $i > K$, $|\text{NRESOLVE}_j| \leqslant c\delta_2 2^{2j}n$.*

Then, under these three conditions,

$$\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi(i))}{n} \right| \leqslant \left(16\delta_1 2^K + \frac{8\delta_2}{d} + 2 \cdot d + \frac{2c\delta_2}{d} \right)n.$$

Proof. By (4.7) and the definition of $n_{i,j}$, we have the bound

$$\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi'(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi'(i))}{n} \right|$$

$$\leqslant \sum_{\log(1/d) \geqslant i \geqslant j \geqslant 0} 2 \cdot 2^{-j} n_{i,j}$$

$$= \sum_{i=0}^{K} \sum_{j=0}^{i} 2 \cdot 2^{-j} n_{i,j} + \sum_{i=K+1}^{l_0} \sum_{j=0}^{i} 2 \cdot 2^{-j} n_{i,j}. \qquad (4.9)$$

The first inequality is by (4.7) and the definition of $n_{i,j}$, and the equality is simply regrouping the terms into two groups. By conditions (i) and (ii), we have for $0 \leqslant i \leqslant K$, $n_{i,i} = n_i$ and $n_{i,t} = 0$ for $t < j$. For $i \geqslant K + 1$, $n_{i,t} = 0$ for $t < K$. Thus combining with Lemma 4.2(c) for the bound on n_i, the first summand in (4.9) can be bounded by

$$\sum_{i=0}^{K} \left(\sum_{j=0}^{i} 2 \cdot 2^{-j} n_{i,j} \right) = \sum_{i=0}^{K} 2 \cdot 2^{-i} n_i \leqslant \sum_{i=0}^{K} 2 \cdot 2^{-i} 4\delta_1 2^{2i} n = 8\delta_1 \sum_{i=0}^{K} 2^i n \leqslant 16\delta_1 2^K n. \quad (4.10)$$

We now bound the second summand in (4.9). Note that $|\text{NRESOLVE}_K| = 0$ by conditions (i) and (ii). Furthermore, by (4.8),

$$|\text{NRESOLVE}_j| - |\text{NRESOLVE}_{j-1}| = n_{j,j-1} + \cdots + n_{l_0,j-1} - (n_{j-1,j-2} + \cdots + n_{j-1,K}). \quad (4.11)$$

We now rewrite the second summand in (4.9):

$$\sum_{i=K+1}^{l_0} \sum_{0 \leqslant j \leqslant i} 2 \cdot 2^{-j} n_{i,j}$$

$$= \sum_{j=K+1}^{l_0} 2 \cdot 2^{-j} n_{j,j} + \sum_{t=K}^{l_0-1} \sum_{j=t+1}^{l_0} 2 \cdot 2^{-t} n_{j,t}$$

$$= \sum_{j=K+1}^{l_0} 2 \cdot 2^{-j} n_{j,j} + \sum_{t=K}^{l_0-1} 2 \cdot 2^{-t} (|\text{NRESOLVE}_{t+1}| - |\text{NRESOLVE}_t| + n_{t,t-1} + \cdots + n_{t,K})$$

$$= \sum_{j=K+1}^{l_0} 2 \cdot 2^{-j} (n_{j,j} + n_{j,j-1} + \cdots + n_{j,K}) + \sum_{t=K}^{l_0-1} 2 \cdot 2^{-t} (|\text{NRESOLVE}_{t+1}| - |\text{NRESOLVE}_t|)$$

$$
= \sum_{j=K+1}^{l_0} 2 \cdot 2^{-j} n_j + \sum_{t=K}^{l_0-1} 2 \cdot (2^{-(t-1)} - 2^{-t}) |\mathrm{NRESOLVE}_t|
$$

$$
= \sum_{j=K+1}^{l_0} 2 \cdot 2^{-j} n_j + \sum_{t=K}^{l_0-1} 2 \cdot 2^{-t} |\mathrm{NRESOLVE}_t|
$$

$$
\leqslant \sum_{j=K+1}^{l_0-1} 2 \cdot 2^{-j} (4\delta_2 2^{2j} n) + 2 \cdot dn + \sum_{t=K}^{l_0-1} 2 \cdot 2^{-t} (c\delta_2 2^{2t} n), \tag{4.12}
$$

where the second equality is by (4.11), the first inequality is by Lemma 4.2(c) and $n_{l_0} \leqslant n$, and condition (iii) in the lemma statement. Using a simple upper bound on a geometric series, (4.12) is bounded above by

$$
8\delta_2 \cdot 2^{l_0} n + 2dn + 2c\delta_2 \cdot 2^{l_0} n = \frac{8\delta_2}{d} n + 2dn + \frac{2c\delta_2 n}{d}. \tag{4.13}
$$

Therefore, combining (4.10) and (4.13), we have

$$
\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi'(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi'(i))}{n} \right| \leqslant \left(16\delta_1 2^K + \frac{8\delta_2}{d} + 2 \cdot d + \frac{2c\delta_2}{d} \right) n. \tag{4.14}
$$

Thus, by (4.6), the claim holds. $\square$

Now we show that with probability at least $1 - \epsilon/2$, the conditions in Lemma 4.3 hold.

Lemma 4.4. *Let $c = 1/2$. With probability at least $1 - \epsilon/2$, the following conditions hold simultaneously.*

(a) *For each $i \leqslant K$, points in squares in $\mathrm{CANDIDATE}_i$ are resolved in level i.*
(b) *For each $i > K$, a point in a square in $\mathrm{CANDIDATE}_i$ is resolved in level j with $j \geqslant K$.*
(c) *For each $i > K$, $|\mathrm{NRESOLVE}_j| \leqslant c\delta_2 2^{2j} n$.*

Proof.   For each $0 \leqslant i \leqslant K$, and for each square in ACTIVE_i, we know it has at least $\delta_1 n$ points by the definition of ACTIVE_i. The probability that this square does not contain a sample point is bounded above by $(1 - \delta_1)^M \leqslant e^{-\delta_1 M}$. There are at most 2^{2i} squares in ACTIVE_i, so by the union bound, the probability that the sample misses some square in ACTIVE_i is at most $2^{2i} e^{-\delta_1 M}$. By the union bound again, the probability that there exists $0 \leqslant i \leqslant K$ such that the sample misses some square in ACTIVE_i is bounded above by

$$
\sum_{i=0}^{K} 2^{2i} e^{-\delta_1 M} < 2^{2K+1} e^{-\delta_1 M} \leqslant \epsilon/4, \tag{4.15}
$$

where the last inequality is by our choices of δ_1, K as in (4.5).

Now we consider the case $K + 1 \leqslant i \leqslant l_0$. Recall $\mathrm{NRESOLVE}_i$ are the points in squares in ACTIVE_i which do not contain a sample point. Each square of side length 2^{-i} in ACTIVE_i has at least $\delta_2 n$ points of π by definition. Suppose $|\mathrm{ACTIVE}_i| = T_i$. We know $|T_i| \leqslant 2^{2i}$ since there are at most 2^{2i} squares of side length 2^{-i}. Given any set of squares in ACTIVE_i which consist of at least $c\delta_2 2^{2i} n$ points of π, the probability that these squares do not contain

a sample point is at most $(1 - c\delta_2 2^{2i})^M \leqslant e^{-c\delta_2 2^{2i}M}$. Since there are at most $2^{|T_i|} \leqslant 2^{2^{2i}}$ choices for these squares, by the union bound, the probability that some squares in ACTIVE$_i$ containing at least $c\delta_2 2^{2i}n$ points and do not contain a sample point is bounded above by $2^{2^{2i}}e^{-c\delta_2 2^{2i}M}$. Therefore

$$\Pr(|\mathrm{NRESOLVE}_i| \geqslant c\delta_2 2^{2i}n) \leqslant 2^{2^{2i}}e^{-c\delta_2 2^{2i}M} = e^{c2^{2i}(c' - \delta_2 M)} \tag{4.16}$$

where $c' = \ln(4e)$. By the union bound, the probability that $|\mathrm{NRESOLVE}_i| \geqslant c\delta_2 2^{2i}n$ for some $K + 1 \leqslant i \leqslant l_0$ is bounded above by

$$\sum_{i=K+1}^{l_0} e^{c2^{2i}(c' - \delta_2 M)}.$$

Since

$$c' - \delta_2 M = \ln(4e) - 20000/5200 = -1.46$$

by the values of M and δ_2 as in (4.1) and (4.5), we have

$$\Pr(|\mathrm{NRESOLVE}_i| \geqslant c\delta_2 2^{2i}n \text{ for some } K + 1 \leqslant i \leqslant l_0) \tag{4.17}$$

$$\leqslant \sum_{i=K+1}^{l_0} e^{c2^{2i}(c' - \delta_2 M)} < \sum_{j=2^{2(K+1)}}^{\infty} e^{cj(c' - \delta_2 M)}$$

$$= e^{c(c' - \delta_2 M)(2^{2(K+1)} - 1)}/(e^{-c(c' - \delta_2 M)} - 1) < \epsilon/8. \tag{4.18}$$

Thus the claim in the lemma holds by (4.15) and (4.18). $\square$

Combining Lemma 4.3 and Lemma 4.4, we know that with probability at least $1 - \epsilon/2$,

$$\sum_{i=1}^{n} \left| \frac{i}{n} - \frac{\phi(i)}{n} \right| + \left| \frac{\pi(i)}{n} - \frac{\pi(\phi(i))}{n} \right| \leqslant \left(16\delta_1 2^K + \frac{8\delta_2}{d} + 2 \cdot d + \frac{2c\delta_2}{d} \right) < \epsilon n/4.$$

Recall the definitions from earlier: $\tilde{\alpha}$ is the blow-up of the subpermutation α induced by the sample in π where the tth blow-up block is the subpermutation induced by the locations $\phi^{-1}(j_t)$ in π, and ψ is the bijection between locations of π and $\tilde{\alpha}$ with $\psi(i) < \psi(j)$ if and only if $\phi(i) < \phi(j)$ or $\phi(i) = \phi(j)$ but $i < j$. Then for the bijection ψ and the blow-up permutation $\tilde{\alpha}$, similar to the proof above, we have by Lemma 3.6 that

$$\rho_{\mathrm{EMD}}(\pi, \tilde{\alpha}) \leqslant \frac{1}{(n-1)/2} \sum_{i=1}^{n} d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n} \right), \left(\frac{\psi(i)}{n}, \frac{\tilde{\alpha}(\psi(i))}{n} \right) \right)$$

$$\leqslant \frac{2}{(n-1)/2} \sum_{i=1}^{n} d\left(\left(\frac{i}{n}, \frac{\pi(i)}{n} \right), \left(\frac{\phi(i)}{n}, \frac{\pi(\phi(i))}{n} \right) \right) < \epsilon/2.$$

4.2. $\rho_{\mathrm{EMD}}(\tilde{\alpha}, \tilde{\alpha}') < \epsilon/2$
In this section we show the following lemma.

Lemma 4.5. *With probability at least $1 - \epsilon/2$, any blow-up $\tilde{\alpha}'$ of α with the jth point of α blows up into a block of size $|\phi^{-1}(j)| = |B_j|$, which satisfies*

$$\rho_{EMD}(\tilde{\alpha}', \tilde{\alpha}) < \epsilon/2.$$

Proof.   As in the proof of Lemma 3.10, we again obtain (3.3), which states

$$\sum_{i=1}^{n} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n| = \sum_{j=1}^{|\alpha|} \left(\sum_{i:(i,\tilde{\alpha}(i))\in B_j} |(\tilde{\alpha}(i) - \tilde{\alpha}'(i))/n| \right) \leqslant \sum_{j=1}^{|\alpha|} |B_j|^2/n. \qquad (4.19)$$

To analyse $|B_j|$, we conduct another dyadic partition similar to that in Algorithm 2 with the following parameters. Let

$$\delta_3 := \epsilon^2 \log(2/\epsilon)/448 \qquad (4.20)$$

be the density threshold for a square to be in ACTIVE. The smallest side length of the squares is $d' = 2^{-\lceil 2\log(1/\epsilon)\rceil}$. Thus $\epsilon^2/2 < d' \leqslant \epsilon^2$. Again, let ACTIVE$_i$ be the set of squares of side length 2^{-i} that were once in ACTIVE; let FROZEN$_i \subset$ FROZEN be the set of squares of side length 2^{-i}. We have the following claim.

Claim 4.6.   *With probability at least $1 - \epsilon/2$, for any $1 \leqslant j \leqslant |\alpha| = M$,*

$$|B_j| \leqslant 56 \log(2/d')\delta_3 n.$$

We delay the proof of this claim to later. Assuming this claim holds, we complete the proof of Lemma 4.5.

Since $\sum_{j=1}^{M} |B_j| = n$ and by the claim, $|B_j| \leqslant B^* := 56 \log(2/d')\delta_3 n$, by convexity we have

$$\sum_{j=1}^{M} |B_j|^2/n \leqslant (B^*)^2 \cdot 1/n \cdot (n/B^*) = B^* = 56 \log(2/d')\delta_3 n. \qquad (4.21)$$

Therefore by (4.19) and applying the same argument as in Lemma 3.10, we have for any $\epsilon > 0$,

$$\rho_{\text{EMD}}(\tilde{\alpha}, \tilde{\alpha}') \leqslant 2 \cdot 56 \log(2/d')\delta_3 \leqslant 112 \log(4/\epsilon^2)\delta_3 \leqslant \epsilon/2.$$

The last inequality holds by our choice of δ_3 as in (4.20). $\square$

We now finish this section by completing the proof of Claim 4.6.

Proof of Claim 4.6.   Notice that a smallest square is of side length d' and can contain at most $d'n < \delta_3 n$ points of π. Hence, the smallest squares cannot be active, so ACTIVE $= \emptyset$. We next show that with probability at least $1 - \epsilon/2$, for each $0 \leqslant i \leqslant \log(1/d')$, each square in ACTIVE$_i$ for some i contains a sample point. The reasoning is the same as in Lemma 3.4. For each square in ACTIVE$_i$, the probability that this square does not contain
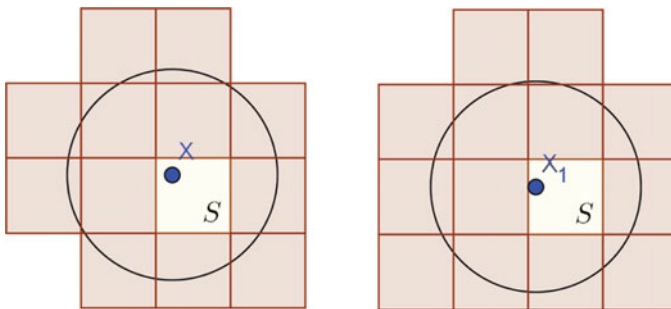
Figure 10. Case 1. The yellow squares are the square S containing the sample points X and X_1 respectively. The circles with centres X and X_1 and radius $\sqrt{2} \cdot 2^{-(i-1)}$, respectively, each intersect at most 13 other squares in ACTIVE_s.

a sample point is at most $(1 - \delta_3)^M$. There are at most 2^{2i} squares in ACTIVE_i; thus there are at most

$$\sum_{i=0}^{\log(1/d')} 2^{2i} \leqslant 2 \cdot d'^{-2}$$

squares in $\bigcup_{i=0}^{\log(1/d')} \text{ACTIVE}_i$. Thus, by the union bound, the probability that for every i, every square in ACTIVE_i contains a sample point is at least

$$1 - 2d'^{-2}(1 - \delta_3)^M \geqslant 1 - 2d'^{-2}e^{-\delta_3 M} \geqslant 1 - \epsilon/2.$$

We assume now that each square once in ACTIVE contains a sample point. We know $[0,1]^2$ is now partitioned into squares in FROZEN. Let $X = (j_t, \pi(j_t))$ be a sample point. Suppose X is in FROZEN_s for some s, which means it is in a square S of side length 2^{-s} that is in FROZEN. We show that $|\phi^{-1}(j)|$ is small by counting the number of squares in FROZEN_i containing points of π that are mapped to this sample point X for each i. Note that a square in FROZEN_i must have a parent square in ACTIVE_{i-1}; on the other hand, each square in ACTIVE_{i-1} contains at most four squares in FROZEN_i. It thus suffices to upper-bound the number of squares in ACTIVE_{i-1} that contain a point of π which maps to X.

Case 1. $i - 1 = s$. Since each square in ACTIVE_{i-1} contains a sample point, the Euclidean distance between any point of π in this square and the sample point in ACTIVE_{i-1} is at most $\sqrt{2} \cdot 2^{-(i-1)}$. Thus for any square S' other than S in ACTIVE_{i-1}, if S' contains a point of π mapped to the sample point X in S, the distance between X and S' is at most $\sqrt{2} \cdot 2^{-(i-1)}$. Since the disc with centre X and radius $\sqrt{2} \cdot 2^{-(i-1)}$ can intersect at most 14 squares of side length 2^{-i} (see Figure 10), including S itself, we know there are at most 14 squares in ACTIVE_{i-1} containing points of π mapped to X. Therefore there are at most $4 \cdot 14$ squares in FROZEN_i containing points of π mapped to X; this implies at most $4 \cdot 14\delta_3 n$ points of π contained in squares in FROZEN_i mapped to X.

Case 2. $i - 1 > s$. Similarly, if a square $S' \in \text{ACTIVE}_{i-1}$ contains a point mapped to the sample point X, then the distance from X to S' is at most $\sqrt{2} \cdot 2^{-(i-1)}$, since the disc with
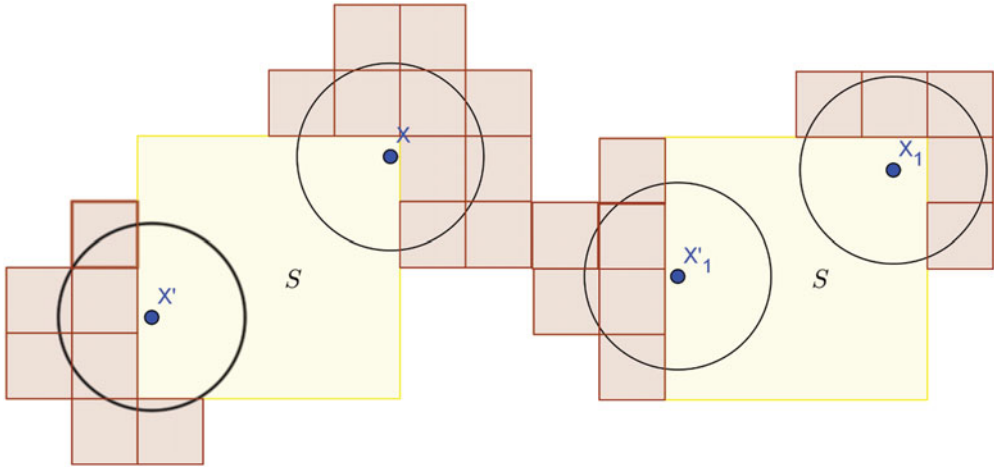
Figure 11. Case 2. The yellow squares are the squares containing the sample points X, X', X_1, X'_1. The four circles with the four sample points as centres and radius $\sqrt{2} \cdot 2^{-(i-1)}$ each intersect at most 12 squares in ACTIVE$_s$ respectively.

centre X and radius $\sqrt{2} \cdot 2^{-(i-1)}$ can intersect at most 12 squares of side length 2^{-i} (see Figure 11). Thus, by a similar argument, there are at most $4 \cdot 12\delta_3 n$ points of π contained in squares in FROZEN$_i$ mapped to X.

Case 3. $i - 1 < s$. By a similar argument, there are at most $4 \cdot 13\delta_3 n$ points of π contained in squares in FROZEN$_i$ mapped to X, since the disc with centre X and radius $\sqrt{2} \cdot 2^{-(i-1)}$ can intersect at most 13 squares of side length 2^{-i} (see Figure 12).

Combining the three cases above, we know that for each $0 \leqslant i \leqslant \log(1/d')$ and $i \neq s$, there are at most $56\delta_3 n$ points of π contained in squares in FROZEN$_i$ mapped to the sample point X. And there are at most $56\delta_3 n$ points of π in FROZEN$_s$ mapped to the sample point X. Therefore

$$|B_j| \leqslant \sum_{i=0}^{\log(1/d')} 56\delta_3 n = 56\log(2/d')\delta_3 n. \tag{4.22}$$

$\square$

5. One-sided property testing under the planar metrics

We prove Theorem 1.9 in this section. The referee pointed out that an approach similar to that of Goldreich and Trevisan given as Proposition D.2 in [18] can be used, which shows that a (non-adaptive) two-sided tester for a hereditary property $\mathcal{P}$ can be used as a black box to obtain a (canonical) one-sided tester for $\mathcal{P}$. This argument simplifies our original proof and also provides a better bound.

We first describe the one-sided tester, which is a modification of the two-sided tester as described in Algorithm 1. With the parameters $M_1(\epsilon/2)$, $M_2(\epsilon/2)$, and M' as described in Theorem 1.9, the tester works as shown in Algorithm 3.
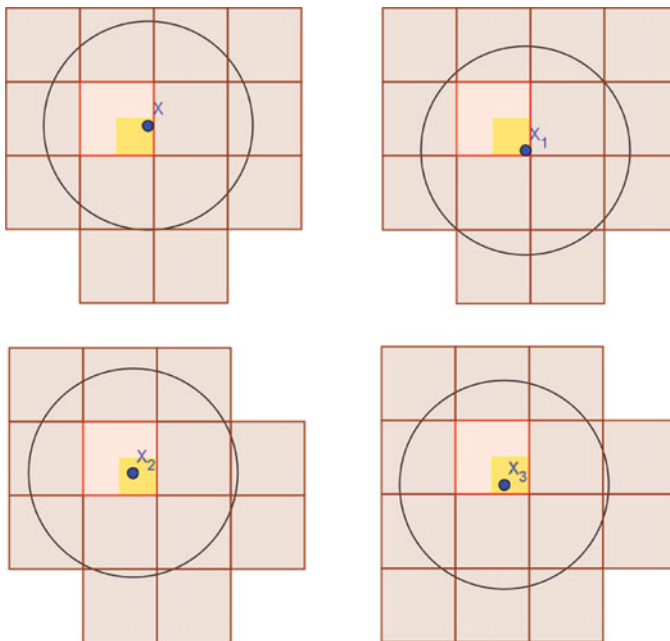
Figure 12. Case 3. The yellow squares are the squares S containing the sample points X, X_1, X_2, X_3. The four circles with the four sample points as centres and radius $\sqrt{2} \cdot 2^{-(i-1)}$ each intersect at most 13 squares in ACTIVE$_s$ respectively.

Algorithm 3 One-sided tester

Let M' be as specified in Theorem 3.1. Given a permutation π, we pick a subpermutation π' of π of size M' uniformly at random. Output 'π is in $\mathcal{P}$' if π' is in $\mathcal{P}$, and otherwise output 'π is not in $\mathcal{P}$'.

Proof of Theorem 1.9. We first assume

$$32 M_1^5 k^*(M_1)/\epsilon^3 \leqslant 32 M_2^5 k^*(M_2)/\epsilon^3.$$

The other case is similar. If $\pi \in \mathcal{P}$ then clearly the result holds since the property is hereditary. We can assume now that π is ϵ-far from $\mathcal{P}$. It suffices to prove that $\Pr(\pi' \in \mathcal{P}) < \epsilon$.

Let π'' be a subpermutation of π' of size $M_1(\epsilon/2)$ picked uniformly at random. Recall we have a two-sided tester in Algorithm 1 in Section 3; we call it Tester 1. Given any permutation π', applying Tester 1 with parameter $\epsilon/2$, we pick a subpermutation π'' of size $M_1(\epsilon/2)$ of π' uniformly at random. For simplicity, we say 'π' is accepted by Tester 1' if Tester 1 outputs '$\pi' \in \mathcal{P}$'. Therefore, for π' a subpermutation of π of size M' picked uniform at random from π, we have

$$\Pr(\pi' \text{ is accepted by Tester 1}) \geqslant \Pr(\pi' \text{ is accepted by Tester 1}|\pi' \in \mathcal{P})\Pr(\pi' \in \mathcal{P})$$
$$\geqslant (1 - \epsilon/2)\Pr(\pi' \in \mathcal{P}).$$

The second inequality holds by a direct consequence of Theorem 1.7. Since the size of π' is at least the value of $n_0(M_1(\epsilon/2), \epsilon/2)$ in Theorem 1.7 and $\pi' \in \mathcal{P}$, a subpermutation π'' of size $M_1(\epsilon/2)$ of π' will be accepted with probability at least $1 - \epsilon/2$.

Furthermore, since π is ϵ-far from $\mathcal{P}$ and is of size at least the value $n_0(M_1(\epsilon/2), \epsilon/2)$ in Theorem 1.7, and π'' is a subpermutation of size $M_1(\epsilon/2)$ of π picked uniformly at random, Theorem 1.7 tells us

$$\Pr(\pi' \text{ is accepted by Tester 1}) \leqslant \epsilon/2.$$

Therefore we have

$$\Pr(\pi' \in \mathcal{P}) \leqslant (\epsilon/2)/(1 - \epsilon/2) < \epsilon. \qquad \square$$

Acknowledgement

References

[1] Alon, N. (2002) Testing subgraphs in large graphs. *Random Struct. Alg.* **21** 359–370.

[2] Alon, N. and Fox, J. (2015) Easily testable graph properties. *Combin. Probab. Comput.* **24** 646–657.

[3] Alon, N. and Shapira, A. (2006) A characterization of easily testable induced subgraphs. *Combin. Probab. Comput.* **15** 791–805.

[4] Alon, N. and Shapira, A. (2008) A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.* **37** 1703–1727.

[5] Conlon, D. and Fox, J. (2012) Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.* **22** 1191–1256.

[6] Conlon, D. and Fox, J. (2013) Graph removal lemmas. In *Surveys in Combinatorics 2013*, Vol. 409 of London Mathematical Society Lecture Note Series, Cambridge University Press, pp. 1–49.

[7] Cooper, J. N. (2006) A permutation regularity lemma. *Electron. J. Combin.* **13** R22.

[8] Diaconis, P. personal communication.

[9] Diaconis, P. (1988) *Group Representations in Probability and Statistics*, Vol. 11 of Lecture Notes Monograph Series, Institute of Mathematical Statistics.

[10] Diaconis, P. and Graham, R. L. (1977) Spearman's footrule as a measure of disarray. *J. Roy. Statist. Soc. Ser. B* **39** 262–268.

[11] Erdős, P. and Szekeres, G. (1935) A combinatorial problem in geometry. *Compositio Math.* **2** 463–470.

[12] Fox, J. Stanley–Wilf limits are typically exponential. *Adv. Math.*, to appear.

[13] Fox, J. and Lovász, L. M. (2016) A tight bound for Green's arithmetic triangle removal lemma in vector spaces. *Adv. Math.* **321** 287–297.

[14] Fox, J., Lovász, L. M. and Zhao, Y. (2016) On regularity lemmas and their algorithmic applications. *Combin. Probab. Comput.* **26** 481–505.

[15] Fox, J. and Wei, F. Strongly testing hereditary permutation properties with polynomial query complexity. In preparation.

[16] Füredi, Z. and Hajnal, P. (1992) Davenport–Schinzel theory of matrices. *Discrete Math.* **103** 233–251.

[17] Goldreich, O., Goldwasser, S. and Ron, D. (1998) Property testing and its applications to learning and approximation. *J. Assoc. Comput. Mach.* **45** 653–750.

[18] Goldreich, O. and Trevisan, L. (2003) Three theorems regarding testing graph properties. *Random Struct. Alg.* **23** 23–57.

[19] Hoppen, C., Kohayakawa, Y. and Sampaio, R. M. (2012) A note on permutation regularity. *Discrete Appl. Math.* **160** 2716–2727.

[20] Hoppen, C., Kohayakawa, Y., Moreira, C. G., Ráth, B. and Sampaio, R. M. (2013) Limits of permutation sequences. *J. Combin. Theory Ser. B* **103** 93–113.

[21] Hoppen, C., Kohayakawa, Y., Moreira, C. G. and Sampaio, R. M. (2011) Testing permutation properties through subpermutations. *Theoret. Comput. Sci.* **412** 3555–3567.

[22] Klazar, M. (2000) The Füredi–Hajnal conjecture implies the Stanley–Wilf conjecture. In *Formal Power Series and Algebraic Combinatorics* (D. Krob, A. A. Mikhalev and A. V. Mikhalev, eds), Springer, pp. 250–255.

[23] Klimošová, T. and Král', D. (2012) Hereditary properties of permutations are strongly testable. arXiv:1208.2624 An early version appeared in *SODA '14: 25th Annual ACM–SIAM Symposium on Discrete Algorithms*, ACM (2014), pp. 1164–1173.

[24] Kruskal, J. B. Jr (1953) Monotonic subsequences. *Proc. Amer. Math. Soc.* **4** 264–273.

[25] Lovász, L. (2012) *Large Networks and Graph Limits*, Vol. 60 of American Mathematical Society Colloquium Publications, AMS.

[26] Marcus, A. and Tardos, G. (2004) Excluded permutation matrices and the Stanley–Wilf conjecture. *J. Combin. Theory Ser. A* **107** 153–160.

[27] Monge, G. (1781), Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, pp. 666–704.

[28] Rubinfield, R. and Sudan, M. (1996) Robust characterization of polynomials with applications to program testing. *SIAM J. Comput.* **25** 252–271.

[29] Rubner, Y., Tomasi, C. and Guibas, L. J. (2000) The earth mover's distance as a metric for image retrieval. *Int. J. Comput. Vision.* **40** 99–121.

[30] Vaidya, P. (1989) Geometry helps in matching. *SIAM J. Comput.* **18** 1201–1225.