

Improved Bounds for Testing Forbidden Order Patterns

Omri Ben-Eliezer*

Clément L. Canonne†

October 31, 2017

Abstract

A sequence $f: \{1, \dots, n\} \rightarrow \mathbb{R}$ contains a permutation π of length k if there exist $i_1 < \dots < i_k$ such that, for all x, y , $f(i_x) < f(i_y)$ if and only if $\pi(x) < \pi(y)$; otherwise, f is said to be π -free. In this work, we consider the problem of testing for π -freeness with one-sided error, continuing the investigation of [Newman et al., SODA'17].

We demonstrate a surprising behavior for non-adaptive tests with one-sided error: While a trivial sampling-based approach yields an ε -test for π -freeness making $\Theta(\varepsilon^{-1/k} n^{1-1/k})$ queries, our lower bounds imply that this is almost optimal for most permutations! Specifically, for most permutations π of length k , any non-adaptive one-sided ε -test requires $\varepsilon^{-1/(k-\Theta(1))} n^{1-1/(k-\Theta(1))}$ queries; furthermore, the permutations that are hardest to test require $\Theta(\varepsilon^{-1/(k-1)} n^{1-1/(k-1)})$ queries, which is tight in n and ε .

Additionally, we show two hierarchical behaviors here. First, for any k and $l \leq k-1$, there exists some π of length k that requires $\tilde{\Theta}_\varepsilon(n^{1-1/l})$ non-adaptive queries. Second, we show an adaptivity hierarchy for $\pi = (1, 3, 2)$ by proving upper and lower bounds for (one- and two-sided) testing of π -freeness with r rounds of adaptivity. The results answer open questions of Newman et al. and [Canonne and Gur, CCC'17].

*Blavatnik School of Computer Science, Tel-Aviv University. Email: omrib@mail.tau.ac.il.

†Columbia University. Email: ccanonne@cs.columbia.edu. Research supported by NSF grants CCF-1115703 and NSF CCF-1319788.

1 Introduction

1.1 Background and motivation

Introduced by Rubinfeld and Sudan [RS96] and Goldreich, Goldwasser, and Ron [GGR98], the field of property testing is concerned with obtaining query- and time-efficient randomized algorithms (usually running in sublinear time), which decide whether their input satisfies some prescribed property of interest, or differs significantly from any object with this property. Originating from the early work on Probabilistically Checkable Proofs, property testing quickly evolved to become an area of study of its own right, encompassing a variety of topics, including (but not limited to) testing of graphs, Boolean and real-valued functions, and probability distributions (see e.g. [Gol10, Gol17, BY17, Ron08, Ron09, Can15] for recent books and surveys).

In this work, we focus on testing properties of real-valued functions, of the form $f: [n] \rightarrow \mathbb{R}$, i.e., *real-valued sequences*. A significant body of work has been dedicated to such properties, with the examples of monotonicity [EKK⁺00, BGJ⁺12, DGL⁺99, Fis04, CS13, Ras14], the Lipschitz property [JR13, CS13], convexity [PRR03], and k -monotonicity [CGG⁺17, GKW17]. Very recently, Newman, Rabinovich, Rajendraprasad, and Sohler [NRRS17] initiated the study of a massively parameterized property,¹ *order pattern freeness*, which generalizes and subsumes some of the aforementioned properties as special cases. We refer the reader to [NRRS17] for a discussion of several motivations for testing order pattern freeness, stemming e.g. from combinatorics and time series analysis.

In this paper, we continue the investigation of testing order pattern freeness. The problem is formulated as follows. The *forbidden order pattern* is a permutation $\pi = (\pi_1, \dots, \pi_k)$ of $[k]$, viewed here as a sequence of length k . Two sequences $x = (x_1, \dots, x_k)$ and $y = (y_1, \dots, y_k)$ are *order-isomorphic* if for any $i \neq j$, $x_i < x_j$ holds if and only if $y_i < y_j$. That is, if the relative order of the elements in both sequences is the same. For the above permutation π , a sequence $f: [n] \rightarrow \mathbb{R}$ is said to *contain the pattern* π if π is order-isomorphic to a subsequence of f . In other words, f contains π if there exist k integers $i_1 < \dots < i_k \in [n]$ such that $f(i_a) < f(i_b)$ if and only if $\pi(a) < \pi(b)$. Accordingly, f is π -free if it does not contain the pattern π , and ε -far from being π -free if one needs to modify at least εn of its values to make it π -free.

An ε -test for order pattern freeness is an algorithm which, given query access to an unknown input sequence $f: [n] \rightarrow \mathbb{R}$, must distinguish (with good probability) between the case that the input is π -free and the case that it is ε -far from being π -free. As an example for $k = 2$, $(2, 1)$ -freeness is equivalent to being monotone non-decreasing, and testing $(2, 1)$ -freeness amounts to testing monotonicity.

1.2 Preliminaries and notation

We hereafter denote by $[n]$ the set of integers $\{1, \dots, n\}$, and by \log the logarithm in base two. As aforementioned, we consider the set \mathcal{F} of real-valued functions of the form $f: [n] \rightarrow \mathbb{R}$ (which we will equivalently refer to as *sequences*), equipped with the usual Hamming distance: $\text{dist}(f, g) = |\{i \in [n] : f(i) \neq g(i)\}|$. Given a property $\mathcal{P} \subseteq \mathcal{F}$, the distance of f to \mathcal{P} is then defined as the minimum distance of f to any function having the property, i.e. $\text{dist}(f, \mathcal{P}) := \inf_{g \in \mathcal{P}} \text{dist}(f, g)$

¹In the massively parameterized framework, the property to be tested depends on an underlying structure, typically of a significant size, which is considered “fixed” (i.e., not part of the input itself). See e.g. [New10] for a survey on testing massively parameterized properties.

(equivalently, this is the minimum number of values one needs to change in f so that it satisfies the property). We say that f is ε -far from \mathcal{P} if $\text{dist}(f, \mathcal{P}) > \varepsilon n$; otherwise, it is ε -close.

We work in the standard setting of property testing: Namely, an ε -test (for a fixed property \mathcal{P}) is a randomized algorithm which, given parameter $\varepsilon \in (0, 1]$ and black-box query access to an unknown function $f: [n] \rightarrow \mathbb{R}$, must, with probability $2/3$, accept if $f \in \mathcal{P}$ and reject if f is ε -far from \mathcal{P} . Moreover, for most of our results we shall be focusing specifically on *one-sided* tests, i.e. those which are required to accept with probability one when $f \in \mathcal{P}$, and are only allowed to err when f is ε -far from \mathcal{P} . The reason for this focus on one-sided tests instead of the weaker two-sided requirement is motivated by the connection to *finding* a violation to the property: Indeed, one-sided testing of π -freeness is essentially the algorithmic problem of finding a π -copy in a sequence that contains many pairwise-disjoint π -copies (instead of the simpler decision problem, which corresponds to the usual two-sided requirement).

With this in mind, we follow the standard use and define the *query complexity* of an ε -test \mathcal{T} for π -freeness as the maximum number of black-box queries of the unknown sequence $f: [n] \rightarrow \mathbb{R}$ that \mathcal{T} may make (as a function of the relevant parameters n , π , and ε). The (one-sided) query complexity of π -freeness testing is then the minimum query complexity among all (one-sided) tests for it; our goal is to design optimal tests for π -freeness, i.e. ones that achieve optimal query complexity.

Throughout the paper, the parameter n is always used to denote the length of the input sequence $f: [n] \rightarrow \mathbb{R}$, the parameter k denotes the length of the forbidden order pattern π , and the proximity parameter of the test is denoted by ε . We use the notations $\tilde{O}(h), \tilde{\Omega}(h)$ to hide polylogarithmic dependencies on n , i.e. for expressions of the form $O(h \log^c n)$ and $\Omega(h \log^c n)$ (for some absolute constant c). We also write O_h (or, analogously, \tilde{O}_h) to denote expressions where the hidden multiplicative constant is allowed to depend (usually polynomially) on the additional parameter h .

Running time of our algorithms All of our tests (except maybe that of [Section 7](#)) run in time *linear* in the number of queries they make. This is because they work by checking whether the queried subsequence contains the forbidden pattern. But this in turn can be performed efficiently, building on an algorithm of Guillemot and Marx [[GM14](#)] which determines whether a given sequence contains a fixed permutation in time that is linear in the size of the sequence.

Distance function All results are stated here for the Hamming distance function, but they also hold for the stronger *deletion distance*, defined as follows: $\text{dist}_{\text{del}}(f, g)$ is the minimal number of value modifications, deletions, and insertions needed to turn f into g . This follows from the fact that the Hamming distance and the deletion distance of a sequence f to π -freeness are always equal: Indeed, if S is a set of entries of a function $f: [n] \rightarrow \mathbb{R}$ whose deletion turns f into a π -free sequence, then it is possible to turn f into a π -free sequence using $|S|$ value modifications by initializing $T = S$ and iteratively applying the following until T is empty: Find an $x \in T$ with a neighboring entry $y \notin T$, set $f(x) = f(y)$, and remove x from T . This way, if f restricted to $[n] \setminus S$ is π -free, then so is f after these value modifications.

In particular, this implies that the distance of a sequence f to π -freeness is closely related to the maximum size of a set \mathcal{C} of pairwise-disjoint π -copies in f : On one hand, if f is ε -far from π -freeness then we cannot delete all π -copies in \mathcal{C} with less than εn entry deletions, so $|\mathcal{C}| \geq \varepsilon n / |\pi|$. On the other hand, if $|\mathcal{C}| \geq \varepsilon n$ then trivially f is ε -far from π -freeness.

Real-valued versus integer range As mentioned before, we are concerned here with real-valued functions, i.e. of the form $f: [n] \rightarrow \mathbb{R}$. However, this range is chosen merely for convenience and generality; all of our results should still hold when considering integer-valued functions, that is $f: [n] \rightarrow \mathbb{N}$.

On the respective parameters As mentioned earlier, π -freeness is a property massively parameterized by a given permutation π . We regard the length of π – denoted k – as a constant, and generally focus on obtaining query complexities that are optimal up to a multiplicative factor that depends on k . Similarly, the proximity parameter ε is to be thought of as either a small constant or a function of the main parameter n (the size of the domain), that slowly tends to zero.

1.3 Organization of the results

In [Section 2](#) we present the results of Newman et al. [[NRRS17](#)] on the problem of testing π -freeness. We then provide our results in [Section 3](#). Most of our results are in the non-adaptive case, and seem to yield a relatively good general understanding of this case. All results in [[NRRS17](#)] only consider one-sided testing, and we also mainly follow this paradigm.

In [Subsection 3.6](#), we turn to the following closely related question: How many queries are needed for a (one-sided or two-sided) test for π -freeness when we have multiple *rounds* of adaptivity (as defined in [[CG17](#)]; that is, each round is non-adaptive in itself, but after each round the test can make adaptive decisions)? We show that π -freeness for $\pi = (1, 3, 2)$ has an adaptivity hierarchy. That is, adding more rounds improves (asymptotically) the query complexity. This is the first known example of a natural property exhibiting an adaptivity hierarchy.

2 Previous work

We describe here the previous state of knowledge on testing pattern-freeness; all results here are established in [[NRRS17](#)], and focus on one-sided tests.

2.1 Sample-based non-adaptive upper bound

Any permutation of length k has a non-adaptive one-sided test making $O(\varepsilon^{-1/k} n^{1-1/k})$ queries. This is the sample-based test, that samples a uniformly random set of elements in the input sequence of the required size and accepts the input (i.e., indicates that it is π -free) if the queried subsequence is π -free. In what follows, this test is called the *sampler*.

2.2 Efficient non-adaptive testing of monotone permutations

π -freeness is efficiently testable non-adaptively when π is monotone: For any $k > 1$, if $\pi = (1, 2, \dots, k)$ or $\pi = (k, k-1, \dots, 1)$ then π -freeness has a one-sided non-adaptive ε -test making $(\varepsilon^{-1} \log n)^{O(k^2)}$ queries. The queries are chosen (non-adaptively) using a dedicated algorithm called the *dyadic sampler*, that iteratively tries to “guess” the typical structure of a π -copy and query according to this guess. This settles the special case of monotone permutations, up to a factor polynomial in $\varepsilon^{-1} \log n$.

2.3 Permutations of length 3, and an exponential gap

Due to symmetry considerations, to understand the behavior of non-monotone permutations of length 3 it is enough to consider the permutation $\pi = (1, 3, 2)$. For this choice of π , it is shown that:

- There is an *adaptive* one-sided ε -test for π -freeness making $(\varepsilon^{-1} \log n)^{O(1)}$ queries.
- Any *non-adaptive* $1/9$ -test for π -freeness has query complexity $\Omega(\sqrt{n})$ – an exponential separation from the adaptive case! It is interesting to note that while the lower bound in [NRRS17] was only obtained for one-sided tests, a similar lower bound for two-sided tests may be derived using similar (yet more technical) ideas.
- There is a *non-adaptive* one-sided ε -test for π -freeness making $\sqrt{n}(\varepsilon^{-1} \log n)^{O(1)}$ queries. Thus, the non-adaptive bounds for $\pi = (1, 3, 2)$ are tight up to an $(\varepsilon^{-1} \log n)^{O(1)}$ factor.

2.4 Non-adaptive lower bounds, and separations between permutations

The $\Omega(\sqrt{n})$ non-adaptive lower bound from [Subsection 2.3](#) actually applies to *any* non-monotone permutation. Moreover, this bound can be strengthened for certain permutations: For any odd k , any one-sided non-adaptive test for the permutation $\pi = (1, k, k - 1, 2, 3, k - 2, \dots, (k + 1)/2)$ requires $\Omega(n^{1-2/(k+1)})$ queries.

2.5 Discussion on previous results

The results in [NRRS17] essentially settle two special cases: The monotone permutations of any length, and the permutations of length 3. However, the general task of understanding the query complexity of optimal tests for π -freeness – for any π – both in the adaptive and the non-adaptive case, has remained wide open. The major open problems that Newman et al. proposed are the following.

Adaptive case Is it true that π -freeness is testable adaptively with query complexity polylogarithmic in n for *any* permutation π ?

Non-adaptive case How does the structure of a permutation π correlate with the query complexity of an optimal (one-sided) non-adaptive test for π -freeness? In particular, do there exist infinitely many permutations π for which π -freeness is testable with query complexity that is $O(n^{0.99})$?

3 Our contributions

In this paper, we mainly address the non-adaptive case, achieving good (though not yet complete) understanding of this case. Along the way, we discover many interesting and surprising phenomena. The details are presented in [Subsections 3.1 to 3.5](#).

Additionally, we explore how partial adaptivity helps in the problem of testing π -freeness, in particular for the permutation $\pi = (1, 3, 2)$. We observe that a hierarchy of adaptivity exists for this problem, making it the first known natural example of a single property with such an hierarchy. The main results concerning partial adaptivity are presented in [Subsection 3.6](#).

3.1 Tight non-adaptive upper bound

Our first main result is an improved upper bound for the non-adaptive one-sided case.

Theorem 3.1. *For any permutation π of length $k \geq 3$, π -freeness has a one-sided non-adaptive ε -test whose query complexity is $O(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}})$.*

This bound improves upon all previously known upper bounds for non-monotone permutations, as the query complexity it suggests is better than both the sample-based upper bound from [Subsection 2.1](#) and the upper bound for permutations of length 3, from [Subsection 2.3](#).

At first glance, an upper bound of $O(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}})$ seems to only be a slight improvement over the $O(\varepsilon^{-\frac{1}{k}} n^{1-\frac{1}{k}})$ sample-based upper bound. However, quite surprisingly, this upper bound is tight in both n and ε for any $k \geq 3$. In other words, the optimal non-adaptive one-sided test for some permutations is only slightly more query-efficient than the sampler!

Theorem 3.2. *Let π be a permutation of length $k \geq 3$, and suppose that $|\pi^{-1}(1) - \pi^{-1}(k)| = 1$. Then the query complexity of any non-adaptive one-sided ε -test for π -freeness is $\Omega\left(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}}\right)$.*

This improves the non-adaptive lower bounds for any permutation of this type, whose length is at least four. For the non-monotone permutations of length 3, this result determines the correct dependence in ε .

The combination of [Theorem 3.2](#) with the results in [Subsection 2.2](#) demonstrates a surprising phenomenon: While the deletion distance between the permutations $\pi_1 = (1, 2, \dots, k)$ and $\pi_2 = (k, 1, 2, \dots, k-1)$ is only 2, the query complexity of non-adaptive one-sided testing for π_1 -freeness differs significantly from that of π_2 -freeness. For π_1 -freeness this query complexity is polylogarithmic in n , and so π_1 is the easiest to test among permutations of length k , while π_2 -freeness has a query complexity of $\Theta\left(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}}\right)$, making π_2 one of those permutations that are hardest to test with non-adaptive one-sided tests.

The proofs of [Theorems 3.1](#) and [3.2](#) appear in [Sections 5](#) and [6](#), respectively.

3.2 Almost tight non-adaptive bounds for random permutations

The next lower bound is perhaps even more surprising. It provides (along with [Theorem 3.1](#)) an almost tight bound on the query complexity of an optimal non-adaptive one-sided test for almost all permutations, implying that this query complexity is usually only marginally better than that of the sampler.

Theorem 3.3. *Let π be picked uniformly at random from all permutations of length k . The following holds with probability $1 - o(1)$ (where the $o(1)$ term tends to zero as $k \rightarrow \infty$): The query complexity of any non-adaptive one-sided ε -test for π -freeness is $\Omega\left(\varepsilon^{-\frac{1}{k-3}} n^{1-\frac{1}{k-3}}\right)$.*

3.3 General permutation-dependent non-adaptive lower bound

Both [Theorems 3.2](#) and [3.3](#) are actually special cases of a general permutation-dependent lower bound that we establish. This lower bound applies to any permutation, and depends heavily on the structure of the permutation. We believe that this lower bound is tight (up to polylogarithmic factors) for any permutation.

Interestingly, it is not clear how to describe the lower bound in a compact closed form, but given a permutation π of length k , the corresponding bound can be computed in constant time (that depends only on k). Later, as an important special case of this strong yet hard-to-digest bound, we provide a slightly weaker permutation-dependent lower bound that has a more natural

combinatorial characterization, and is therefore easier to analyze. See [Theorem 3.8](#) and the resulting Corollaries [3.9](#) and [3.10](#) for more details.

In order to describe our general lower bound, we shall first provide some definitions.

Definition 3.4. Let $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of length k . A subsequence σ of π is *consecutive* if $\sigma = (\pi_i, \dots, \pi_j)$ for some $1 \leq i \leq j \leq k$; in this case we write $\sigma = \pi[i, j]$.

A *partition* $\Lambda = (\sigma_1, \dots, \sigma_\ell)$ of the permutation π consists of consecutive subsequences $\sigma_1 = \pi[1, r_1], \sigma_2 = \pi[r_1 + 1, r_2], \dots, \sigma_\ell = \pi[r_{\ell-1} + 1, k]$, and its *size* is $|\Lambda| = \ell$.

A *signed partition* $P = (\Lambda, S)$ of the permutation π consists of a partition Λ as above, and a sign vector $S = (s_1, \dots, s_\ell) \in \{+, -\}^\ell$. For any σ_i of length bigger than one, the corresponding sign s_i must satisfy the following. If $\min \sigma_i$ appears before $\max \sigma_i$ in π , then the direction sign of σ_i is $-$, and otherwise, the direction sign is $+$. The *size* of P is $|\Lambda| = |S| = \ell$.

Let P be a signed partition as above. Define $r_0 = 0$, and for any $1 \leq i \leq \ell$, denote the length of σ_i by k_i (so $\sum_{i=1}^\ell k_i = k$). Consider the sequence $f_P: [k^2] \rightarrow \mathbb{R}$ defined as follows. For any $1 \leq j \leq k_i$ and $0 \leq m \leq k - 1$, we take $f_P(r_i k + m k_i + j) = m + \pi_{r_i+j}/2k$ for any $0 \leq i \leq \ell - 1$ where s_i is $+$, and $f_P(r_i k + m k_i + j) = (k - 1 - m) + \pi_{r_i+j}/2k$ for any $0 \leq i \leq \ell - 1$ where s_i is a $-$.

Note that for any $0 \leq m \leq k - 1$, the set of all entries $x \in [k^2]$ satisfying $m < f_P(x) < m + 1$ is a π -copy. We say that such a π -copy is *trivial*. We say that P is *unique* if f_P does not contain non-trivial π -copies, and denote by $\mathcal{U}(\pi)$ the set of all unique signed partitions of π . Finally, the *unique signed partition number* (USPN) of π is $u(\pi) = \max_{P \in \mathcal{U}(\pi)} |P|$.

Our lower bound for testing π -freeness is closely related to the USPN of π .

Theorem 3.5. *Let π be any permutation. Any non-adaptive one-sided ε -test for π -freeness has query complexity $\Omega\left(\varepsilon^{-1/u(\pi)} n^{1-1/u(\pi)}\right)$.*

The USPN of a permutation obviously depends only on the permutation (and not on the input sequence size), so it can be computed in constant time, that depends only on k . Thus, given a permutation π and parameters n, ε , one can compute the lower bound obtained from [Theorem 3.5](#) in constant time.

The proof of [Theorem 3.2](#) follows from [Theorem 3.5](#) by showing that for any permutation π of length k which satisfies $|\pi^{-1}(1) - \pi^{-1}(k)| = 1$, it holds that $u(\pi) = k - 1$; actually these are the only permutations of length k whose USPN is $k - 1$, and no permutation of length $k > 1$ has USPN that equals k , as can be derived from results that are discussed later.

We conjecture that the lower bound of [Theorem 3.5](#) is tight up to a multiplicative term that is polynomial in ε and $\log n$. That is, we conjecture that the USPN, $u(\pi)$, is the correct parameter of π that determines how hard it is to non-adaptively test π -freeness using one-sided tests.

Conjecture 3.6. *For any permutation π of any length, π -freeness has a non-adaptive one-sided ε -test making $\tilde{\Theta}_\varepsilon\left(n^{1-1/u(\pi)}\right)$ queries.*

A multiplicative term of $\log n$ is necessary to make [Conjecture 3.6](#) hold for monotone permutations π (for which $u(\pi) = 1$), since there is a lower bound of $\Omega(\log n)$ for testing monotonicity [[Fis04](#)], that can be generalized to testing π -freeness for any permutation π of length at least 2.

For non-monotone permutations, an even stronger conjecture can be given, namely that the number of queries required by a non-adaptive one-sided ε -test is $\Theta_\varepsilon(n^{1-1/u(\pi)})$ (without the polylogarithmic term in n).

3.4 Combinatorial characterizations related to the general lower bound

Motivated by [Theorem 3.5](#), it is desirable to find natural necessary and sufficient combinatorial conditions for uniqueness of a signed partition of a given permutation π . Our next main result provides a useful sufficient condition. For the result, we need some more definitions.

Definition 3.7. Let $\sigma = \pi[x, y]$ and $\sigma' = \pi[x', y']$ of π be disjoint consecutive subsequences of length at least two, and let $x' \leq m, M \leq y'$ be the indices satisfying $\pi_m = \min \pi[x', y']$ and $\pi_M = \max \pi[x', y']$. We say that σ' is *shadowed* with respect to σ if one of the following holds.

- $x' > y, m < M$, and $\pi_{x'-1} > \pi_M$.
- $x' > y, m > M$, and $\pi_{x'-1} < \pi_m$.
- $y' < x, m < M$, and $\pi_{y'+1} < \pi_m$.
- $y' < x, m > M$, and $\pi_{y'+1} > \pi_M$.

An *entangling* of π is a collection $E = (\sigma_1, \dots, \sigma_t)$ of pairwise disjoint consecutive subsequences of π , where $\sigma_i = \pi[a_i, b_i]$ for any $1 \leq i \leq t$, satisfying the following.

- For any $2 \leq j \leq t$, the following holds. Either $a_j > b_1$ and $\min_{i < j} \min \sigma_i < \pi_{a_j} < \max_{i < j} \max \sigma_i$, or $b_j < a_1$ and $\min_{i < j} \min \sigma_i < \pi_{b_j} < \max_{i < j} \max \sigma_i$.
- For any $2 \leq j \leq t$, σ_j is not shadowed with respect to σ_1 .
- For any $1 \leq \ell \leq k$, there exists $\sigma \in E$ such that $\min \sigma \leq \ell \leq \max \sigma$.

For the above entangling E of π , define $\Lambda(E)$ as the partition of π in which $\sigma_1, \dots, \sigma_t$ serve as parts, and any element of π not in $\bigcup_{i=1}^t \sigma_i$ has its own part. Denote $d(E) = |\Lambda(E)| = k - \sum_{\sigma \in E} (|\sigma| - 1)$. Finally, the *entangling number* of π is $d(\pi) = \max_E \{d(E)\}$ where E ranges over all valid entanglings of π .

Theorem 3.8. *For any permutation π and entangling E of π , there exists $S \in \{+, -\}^{|E|}$ for which the signed partition $P = (\Lambda(E), S)$ is unique. In particular, $d(\pi) \leq u(\pi)$ for any permutation π .*

The following is an immediate yet important corollary of [Theorems 3.5](#) and [3.8](#).

Corollary 3.9. *For any permutation π , any non-adaptive one-sided ε -test for π -freeness must make $\Omega\left(\varepsilon^{-1/d(\pi)} n^{1-1/d(\pi)}\right)$ queries.*

A useful simple special case of [Corollary 3.9](#) is the following.

Corollary 3.10. *For a permutation $\pi = (\pi_1, \dots, \pi_k)$, let $m(\pi) = \max_{1 \leq i \leq k-1} |\pi_{i+1} - \pi_i|$. Then $m(\pi) \leq d(\pi)$, and in particular, any non-adaptive one-sided ε -test for π -freeness must make $\Omega\left(\varepsilon^{-1/m(\pi)} n^{1-1/m(\pi)}\right)$ queries.*

Note that a permutation π of length k with $|\pi^{-1}(1) - \pi^{-1}(k)| = 1$ satisfies $m(\pi) = k - 1$, so [Theorem 3.2](#) is actually a special case of [Corollary 3.10](#).

[Theorem 3.3](#) follows from [Corollary 3.9](#) by observing that $d(\pi) \geq k - 3$ holds w.h.p. for a random permutation π of length k ; actually, both $d(\pi)$ and $u(\pi)$ are concentrated in the values $k - 2$ and $k - 3$, as $u(\pi) = k - 1$ holds with probability $O(1/k)$.

There exist permutations π for which $d(\pi) < u(\pi)$. In particular, partitions with a unique signed form are not necessarily entanglings, so the sufficient condition for uniqueness from [Theorem 3.8](#) is not a necessary one. For example, one can verify that $\pi = (4, 1, 2, 5, 6, 3)$ satisfies $d(\pi) = 3$

but $u(\pi) = 4$; a unique signed partition of size 4 for π is (Λ, S) where $\Lambda = ((4, 1), 2, 5, (6, 3))$ and $S = (+, -, -, +)$.

The following necessary condition for being a unique signed partition is easy to prove.

Observation 3.11. *Let π be a permutation of length k and let $P = (\Lambda, S)$ be a unique signed partition for π . Then Λ satisfies the following conditions.*

- For any $1 \leq \ell \leq k$ there exists $\sigma \in \Lambda$ of length bigger than one, such that $\min \sigma \leq \ell \leq \max \sigma$.
- Let $\sigma \in \Lambda$ with $|\sigma| > 1$. If $\max \sigma < k$ then there exists $\sigma' \in \Lambda$ satisfying $\min \sigma' < \max \sigma < \max \sigma'$, and similarly, if $\min \sigma > 1$ then there exists $\sigma' \in \Lambda$ satisfying $\min \sigma' < \min \sigma < \max \sigma'$.

The size $|\Lambda|$ of the largest partition Λ of π satisfying the conditions in [Observation 3.11](#) might be bigger than the USPN of π . For example, the partition $\Lambda = ((5, 1), 3, 2, 7, 6, (8, 4))$ of the permutation $\pi = (5, 1, 3, 2, 7, 6, 8, 4)$ satisfies these conditions, but one can verify that it is not a unique signed partition. By [Observation 3.11](#), none of the other partitions of π of size 6 have a unique signed form, so $u(\pi) < 6 = |\Lambda|$. In fact, $u(\pi) = 5$ in this case, as $((1, 3), (2, 7), (6, 8))$ is an entangling.

3.5 Permutation-dependent hierarchy in the non-adaptive case

The statement of [Conjecture 3.6](#) suggests that there is a permutation-dependent hierarchical behavior of the query complexity for one-sided non-adaptive testing of π -freeness as a function of π . The following result verifies that such an hierarchical structure indeed exists.

Theorem 3.12. *For any two positive integers $k \geq 2$ and $1 \leq \ell \leq k - 1$, there is a permutation π of length k with $m(\pi) = \ell$, for which the optimal non-adaptive ε -test makes $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$ queries, where the $\tilde{\Theta}_\varepsilon$ notation hides a term polynomial in $\log n$ and ε .*

In particular, we conclude that for any positive integer ℓ , there exist infinitely many permutations π for which the query complexity of one-sided non-adaptive testing of π -freeness is $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$. This answers and generalizes the open question of Newman et al. [[NRRS17](#)], who asked whether there exist infinitely many permutations π that have a non-adaptive one-sided test for π -freeness making at most $O(n^{0.99})$ queries (for a fixed ε).

3.6 Hierarchy of Adaptivity

Recent work of Canonne and Gur [[CG17](#)] introduced the notion of *amount of adaptivity* in property testing, which they define as follows.² An r -round test for some property \mathcal{P} is an algorithm which proceeds in r stages. At each stage, it produces and makes a batch of queries to the function, which cannot depend adaptively on each other (i.e., these queries are among themselves non-adaptive), and receives all answers to these queries. It then produces the queries for the next stage, which can depend adaptively on the answers just received. At the end of the r stages (the first one being fully non-adaptive, and each query in the subsequent rounds depending adaptively on those made in the previous rounds only), the algorithm must accept or reject the function; the query complexity is then the total number of queries made overall.

²Actually, for clarity of exposition we slightly depart from the notation of [[CG17](#)], and use *r-round test* for what they refer to as *(r - 1)-round-adaptive test*. The reason is that, with our convention, an r -round test is an algorithm which proceeds in r rounds; while in theirs, an r -round test is an algorithm which proceeds in $r + 1$ rounds, with r adaptive ones and one (the very first) being non-adaptive.

Note that in this formalization, non-adaptive tests correspond to 1-round tests, while adaptive ones are those with unbounded number of rounds. In [CG17], the authors establish a strong hierarchy theorem, showing separations between r -round tests and $(r + 1)$ -round tests, for any integer r – albeit for a somewhat contrived family of properties. They also establish another such theorem, slightly weaker, but this time for a *natural* property (of graphs). However, both results have the issue of producing a different property for every r : that is, “for every integer r there exists a property \mathcal{P}_r hard to test in r rounds, but easy in $r + 1$.” Determining whether there exists a *single* natural property which would witness such a hierarchy – “there exists a property \mathcal{P} which is, for infinitely many integers $r < r'$, harder to test in r rounds than in r' ” – is posed as an open question in [CG17].

In this work, we give a positive answer for this question by analyzing tests for the specific pattern $\pi = (1, 3, 2)$. Newman et al. established in [NRRS17] an exponential gap between the query complexity of *adaptive* and *non-adaptive* tests; focusing on r -round tests, and building on their algorithm, we establish a finer separation for testing this particular pattern, leading to our adaptivity hierarchy result.

Theorem 3.13. *For every $1 \leq r \leq (\log n)^{O(1)}$, there exists an r -round (one-sided) testing algorithm for $(1, 3, 2)$ -freeness with query complexity $\tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right)$.*

We then complement the upper bound part of our adaptivity separation by the following lower bound counterpart:

Theorem 3.14. *For every $0 \leq r \leq (\log \log n)^{O(1)}$, any r -round (two-sided) testing algorithm for $(1, 3, 2)$ -freeness must make $\Omega\left(n^{1/2^{r+3}}\right)$ queries.*

Note that the type of tests involved in Theorems 3.13 and 3.14 is best possible. Indeed, it implies that there is a test \mathcal{T} using 2^{r+3} rounds whose number of queries is polynomially better than that of any one-sided or two-sided r -round test, and in addition that this \mathcal{T} belongs to the more restricted class of one-sided tests. We remark that one could hope for an even stronger theorem, which would separate r rounds from $r + 1$ (as opposed to r vs. $\exp(r)$ as above). We conjecture that the best r -round-adaptive test has query complexity $\tilde{\Theta}_\varepsilon\left(n^{1/(2r+O(1))}\right)$.

4 Discussion and open problems

The problem of (one-sided) testing of π -freeness demonstrates a wide array of interesting phenomena: An exponential separation between the adaptive and the non-adaptive case, surprising hardness results and permutation-dependent hierarchical behaviors in the non-adaptive case, and a hierarchy of adaptivity that is the first of its kind. We believe that these results serve as a strong motivation to try to achieve a complete understanding of the problem. Below we suggest several possible directions for future research.

The adaptive case Testing π -freeness in the adaptive case is still far from being understood. In particular, the question whether all permutations are testable adaptively with number of queries that is polylogarithmic in n is still wide open, even if we allow for two-sided tests. At this point, this seems to be the most intriguing open question regarding testing π -freeness.

Improving bounds in the non-adaptive case While our understanding of the non-adaptive case is far better than that of the adaptive case, there are still gaps in it. The main goal here is to obtain good permutation-dependent upper bounds: [Conjecture 3.6](#) states that our lower bound is actually tight, and it will be obviously interesting to understand if it holds.

Understanding the USPN Another interesting direction would be to obtain a simple complete combinatorial characterization of the USPN of any given permutation. Currently we have lower and upper bounds for the USPN of a permutation ([Theorem 3.8](#) and [Observation 3.11](#), respectively), that are usually tight for small permutations, and we know that the USPN of a permutation is computable in constant time.

Two-sided testing All known results so far are for one-sided testing, aside from our two-sided lower bound in the partially adaptive setting. It is worth to note that the $\Omega_\epsilon(n^{1/2})$ lower bound for one-sided testing of all non-monotone permutations can be (carefully) translated into the same bound for two-sided tests. However, the proofs of other one-sided non-adaptive lower bounds do not seem to translate well to the two-sided setting.

Therefore, it will be interesting to understand what is the query complexity of optimal *two-sided* tests, both in the adaptive and the non-adaptive case. Specific questions of interest include (but are not limited to) the following: When do the non-adaptive two-sided lower bounds match the one-sided ones? Can one obtain a general two-sided upper bound that beats the tight one-sided upper bound of [Theorem 3.1](#) for permutations of size bigger than three? Does two-sidedness help testing in the adaptive case?

Families of forbidden order patterns It will be interesting to investigate the case where more than one order pattern is forbidden (note that there are families for which the question does not make sense; for example, the famous theorem of Erdős-Szekeres [[ES35](#)] implies that any sequence of length at least $k^2 - 2k + 2$ must contain one of the monotone permutations of length k). As mentioned in [[NRRS17](#)], all one-sided upper bounds from the single-pattern case carry over to the multiple-pattern case, but the lower bounds do not; for example, there exists a family of two non-monotone permutations of size 3 that has a one-sided non-adaptive test whose query complexity is polylogarithmic in n . Some specific open questions of interest: Is the upper bound from [Theorem 3.1](#) tight in this case? How does the non-adaptive family-dependent hierarchy look like?

Forbidden order patterns in multi-dimensional structures How does π -freeness behave in structures of higher dimensions, such as the hypergrid or the Boolean hypercube? The sample-based upper bound from [Subsection 2.1](#) still holds in these cases, provided that the input contains many pairwise-disjoint copies of the forbidden structure π . However, in contrast to the one-dimensional case, it is then no longer clear whether being far from π -freeness implies that the input indeed has many pairwise-disjoint π -copies. Interestingly, recent work of Grigorescu, Kumar, and Wimmer [[GKW17](#)] gives strong evidence that testing order pattern freeness on the hypercube is hard.

5 Upper bound

In this section we provide the proof of [Theorem 3.1](#). The test that is used to prove the upper bound is one-sided, and indicates that the input sequence $f: [n] \rightarrow \mathbb{R}$ has a π -copy only if it finds one. Thus, the testing problem reduces to the following search problem: Given query access to an unknown sequence f that is ε -far from π -freeness, the goal is to find a π -copy in f . Here and henceforth, we omit floor and ceiling signs, as they do not make an essential difference in the arguments.

The proof of [Theorem 3.1](#) follows immediately from the next lemma, which provides a sublinear algorithm to find a π -copy in a sequence f , assuming that f is far enough from π -freeness.

Lemma 5.1. *Let π be a permutation of length $k \geq 3$, and suppose that $f: [n] \rightarrow \mathbb{R}$ is ε -far from π -freeness for some $\varepsilon \geq c_k n^{-1/9}$, where c_k depends only on k , and n is large enough (as a function of k). Then there is an algorithm that finds, with probability $2/3$, a copy of π in f by querying $O(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}})$ entries in f .*

Remark 5.2. [Lemma 5.1](#) is stronger than what is needed to obtain a one-sided test, in the sense that ε is allowed to scale with n ; for the proof of [Theorem 3.1](#) a lemma that applies to a constant ε would have been sufficient. However, the added flexibility of the lemma reflects that the statement of [Theorem 3.1](#) would still be true should we take ε^{-1} as a slowly-growing function of n .

Proof. The proof idea is as follows. Given an input sequence $f: [n] \rightarrow \mathbb{R}$, we partition $[n]$ into a collection \mathcal{I} of n/m intervals of size m each, for a suitable choice of m ; we may assume, for convenience, that m divides n . Suppose that f is ε -far from π -freeness. Then f contains a set \mathcal{A} of $\varepsilon n/k$ pairwise disjoint π -copies. We consider two cases, where for each of the cases the queries made are different. Our actual algorithm is a combination of the algorithms for each of the cases.

In the first case, most π -copies in \mathcal{A} have at least two entries in the same interval; the algorithm for this case queries a set of whole intervals, chosen uniformly at random, and a set of single elements, also chosen uniformly at random. The analysis of this case does not use the fact that π is a permutation. In the second case, most π -copies in \mathcal{A} do not have two entries in the same interval, and it can be shown that the sampler (which samples entries of f uniformly at random) suffices for this case. Here we do use the fact that π is a permutation, and the analysis actually shows that the required number of queries is much smaller (for constant ε) than in the first case.

We now give the full details. Pick the interval size to be $m = (\varepsilon n)^{1-1/(k-1)}$, and write $\pi = (\pi_1, \dots, \pi_k)$. The π -copies are represented in \mathcal{A} as k -tuples $t = (t_1, \dots, t_k)$ where t_i is the location of the element corresponding to π_i in the copy. Write $\mathcal{A} = \mathcal{B} \cup \mathcal{C}$, where \mathcal{B} contains all π -copies from \mathcal{A} that have at least two entries in the same interval, and \mathcal{C} contains all π -copies that have at most one entry in each interval. Then either $|\mathcal{B}| \geq \varepsilon n/2k$ or $|\mathcal{C}| \geq \varepsilon n/2k$.

Case 1: $|\mathcal{B}| \geq \varepsilon n/2k$ Our algorithm for this case is described as follows. We first pick a set $Q_1 \subseteq \mathcal{I}$ of intervals, where every $I \in \mathcal{I}$ is included in Q_1 with probability $p = cm/\varepsilon n = c(\varepsilon n)^{-1/(k-1)}$, independently of other intervals. Here $c = 100k^2$ is a constant that depends (polynomially) on k . Next, we pick a set Q_2 of elements from $[n]$, where each element is added to Q_2 with probability p , uniformly and independently of other elements. Up to this point, the algorithm does not make any queries.

An independent sampling trick Variants of the following simple idea are used several times along the paper. Let E_{found} be the event that the subsequence of f induced by Q_1 and Q_2 contains a π -copy. Let E_{big} be the event that $|Q_1| > 100c/\varepsilon$ or $|Q_2| > 100cm/\varepsilon$. By Markov's inequality, $\Pr(E_{\text{big}}) \leq 1/50$. If E_{big} occurs, then the algorithm stops without making any queries (and hence it does not find a π -copy in f). If E_{big} has not occurred, then the algorithm now queries all elements induced by Q_1 and Q_2 . Thus, the algorithm finds a π -copy if and only if E_{found} occurs and E_{big} does not occur. The number of queries made is at most $200cm/\varepsilon = O(\varepsilon^{-\frac{1}{k-1}} n^{1-\frac{1}{k-1}})$, as desired. The probability that the algorithm finds a π -copy is at least $\Pr(E_{\text{found}}) - \Pr(E_{\text{big}}) \geq \Pr(E_{\text{found}}) - 1/50$. Thus, it remains to show that $\Pr(E_{\text{found}}) \geq 2/3 + 1/50$ (note that we consider here the unconditional probability of E_{found} , and in particular, we do not condition on E_{big} not happening).

Analyzing the event E_{found} For each $I \in \mathcal{I}$, let t_I denote the number of π -copies from \mathcal{B} that have at least two entries in I , and let $t = \sum_{I \in \mathcal{I}} t_I$, so $\varepsilon n/2k \leq t \leq \varepsilon n$. Let X be the random variable that counts the number of π -copies from \mathcal{B} that have at least two entries in some $I \in \mathcal{Q}_1$. The expectation of X is $\mathbb{E}[X] = tp \geq cm/2k = 50km$, and the variance of X is bounded by

$$\mathbb{E}[X^2] \leq p \sum_{I \in \mathcal{I}} t_I^2 \leq pm^2 \varepsilon n/m = cm^2$$

where the second inequality follows from convexity arguments, using the facts that $0 \leq t_I \leq m$ for every I and $\sum t_I = t \leq \varepsilon n$. Thus, the standard deviation of X is bounded by $m\sqrt{c} = 10km$. By Chebyshev's inequality, we get that $X \geq m$ with probability at least $9/10$.

Assume now that $X \geq m$, that is, there exists a set $\mathcal{B}' \subseteq \mathcal{B}$ of m π -copies that have at least two of their entries in intervals from \mathcal{Q}_1 . For each such copy, the event that all other $k-2$ (or less) entries of it are in \mathcal{Q}_2 has probability at least $p^{k-2} = c^{k-2}/m$, and is independent of the corresponding events of the other copies in \mathcal{B}' . Thus, the probability that none of these events occurs is bounded by $(1 - c^{k-2}/m)^m \leq e^{-c^{k-2}} < 1/100$. This finishes the proof.

Case 2: $|\mathcal{C}| \geq \varepsilon n/2k$ We start with some notation. For a copy $t = (t_1, \dots, t_k) \in \mathcal{C}$, we define $I_i(t)$ as the interval in \mathcal{I} containing t_i .

Non-extremal π -copies For any interval $I \in \mathcal{I}$, let $y_1 \leq \dots \leq y_m$ be the elements of $f(I) = \{f(x) : x \in I\}$, and let $y_I^- = y_{\lceil \varepsilon m/6k \rceil}$ and $y_I^+ = y_{\lfloor m - \varepsilon m/6k \rfloor}$. We say that a π -copy $t = (t_1, \dots, t_k)$ is *top-high* if $f(t_{\pi^{-1}(k)}) > y_{I_{\pi^{-1}(k)}(t)}^+$, and *bottom-low* if $f(t_{\pi^{-1}(1)}) < y_{I_{\pi^{-1}(1)}(t)}^-$. A copy that is neither top-high nor bottom-low is said to be *non-extremal*. In other words, a π -copy is non-extremal if its highest point is not too high with respect to the interval it lies in, and similarly, its lowest point is not too low with respect to its interval. Note that the number of top-high π -copies in \mathcal{C} is bounded by $\varepsilon n/6k$ (as each interval contributes no more than $\varepsilon m/6k$ such copies), and similarly for the number of bottom-low π -copies. Thus, there exists a set $\mathcal{C}' \subseteq \mathcal{C}$ of $\varepsilon n/6k$ non-extremal π -copies.

The main idea is that with sufficiently many queries, the sampler – a sample-based algorithm to find a π -copy – will be able to capture all entries of a non-extremal π -copy $t = (t_1, \dots, t_k) \in \mathcal{C}'$ besides the lowest entry $t_{\pi^{-1}(1)}$ and the highest entry $t_{\pi^{-1}(k)}$, which will be replaced by a small enough entry from $I_{\pi^{-1}(1)}(t)$ and a large enough entry from $I_{\pi^{-1}(k)}(t)$, respectively. Note that this is a valid π -copy.

Analyzing the sampler Let $p = cm/\varepsilon n = c(\varepsilon n)^{-1/(k-1)}$ as above. Let E_{found} be the event that a subsequence g of f , constructed by putting each entry of f in it with probability p , contains a π -copy. Using the sampling trick from the first case, it is enough to show that $\Pr(E_{\text{found}}) \geq 2/3 + 1/50$. Before we continue, we define the events A_t, B_t, E_t for any π -copy $t = (t_1, \dots, t_k) \in \mathcal{C}'$ as follows. A_t is the event that all entries $\{t_{\pi^{-1}(j)}\}_{j=2}^{k-1}$ of t are included in g , so $\Pr(A_t) = p^{k-2} = c^{k-2}/m$. B_t is the event that g contains $x \in I_{\pi^{-1}(1)}(t)$ and $x' \in I_{\pi^{-1}(k)}(t)$ such that $f(x) \leq f(t_{\pi^{-1}(1)})$ and $f(x') \geq f(t_{\pi^{-1}(k)})$, so $\Pr(B_t) \geq 1 - 2(1-p)^{\varepsilon m/6k}$. Finally, $E_t = A_t \cap B_t \subseteq E_{\text{found}}$ indicates that g contains a π -copy. Note that any event A_t is independent of all other events, and B_t is only dependent on events $B_{t'}$ where $I_{\pi^{-1}(1)}(t) = I_{\pi^{-1}(1)}(t')$ or $I_{\pi^{-1}(k)}(t) = I_{\pi^{-1}(k)}(t')$; there are at most $2m$ such events for each B_t .

The analysis of permutations of length 3 differs from that of longer ones.

π of length $k > 3$. The probability that none of the events A_t for $t \in \mathcal{C}'$ holds is at most $(1 - p^{k-2})^{|\mathcal{C}'|} \leq \exp(-p^{k-2}\varepsilon n/6k) = \exp(-c^{k-2}(\varepsilon n)^{1/(k-1)}/6k) < 1/10$. Suppose then that A_t holds for some $t \in \mathcal{C}'$. The probability that B_t does not occur is bounded by $2(1-p)^{\varepsilon m/6k} \leq 2\exp(-p\varepsilon m/6k) = 2\exp(-cm^2/6kn) = 2\exp(-\varepsilon^{2-2/(k-1)}n^{1-2/(k-1)}c/6k) < 1/10$ for large enough n . Thus, $\Pr(E_{\text{found}}) = \Pr(\exists t: A_t \wedge B_t) > 8/10 > 2/3 + 1/50$ in this case, as desired.

π of length $k = 3$. Let X_t denote the indicator random variable of E_t and let $X = \sum_{t \in \mathcal{C}'} X_t$. For every $t \in \mathcal{C}'$, we have $\Pr(A_t) = p = c/m$ and $\Pr(B_t) \geq (1 - (1-p)^{\varepsilon m/18})^2 \geq (1 - \exp(-p\varepsilon m/18))^2 = (1 - e^{-c\varepsilon/18})^2 \geq c^2\varepsilon^2/400$, where the last inequality holds when $\varepsilon \leq \alpha k^{-2}$ for a small enough α , as $e^{-x} \leq 1 - 9x/10$ for small enough x . Thus, $\mathbb{E}[X] = \sum_{t \in \mathcal{C}'} \Pr(A_t) \Pr(B_t) \geq \frac{\varepsilon n}{18} \frac{c}{m} \frac{c^2 \varepsilon^2}{400} = \frac{c^3}{7200} \varepsilon^{5/2} n^{1/2}$. On the other hand,

$$\text{Var}(X) = \sum_{t \in \mathcal{C}'} \text{Var}(X_t) + \sum_{t \neq t' \in \mathcal{C}'} \text{Cov}(X_t, X_{t'}) \leq \mathbb{E}[X] + 2m|\mathcal{C}'| \max_{t, t'} \Pr(A_t) \Pr(A_{t'}) \Pr(B_t), \quad (1)$$

where the inequality builds on the following two facts. The first is that $\text{Cov}(X_t, X_{t'}) \leq \mathbb{E}(X_t X_{t'}) \leq \Pr(A_t) \Pr(A_{t'}) \Pr(B_t)$, as the events $A_t, A_{t'}, B_t$ are mutually independent. The second fact is that, for any $t \in \mathcal{C}'$, $\text{Cov}(X_t, X_{t'}) = 0$ for all but $2m$ of the tuples t' , as discussed above.

The second term in (1) is bounded by $2c^2\sqrt{\varepsilon n}$. Therefore, the standard deviation of X is bounded by $\sqrt{\mathbb{E}[X]} + \sqrt{2}c\varepsilon^{1/4}n^{1/4} \leq \mathbb{E}[X]/10$, where the bound on ε in the statement of the lemma is chosen so that the last inequality holds (note that $\varepsilon^{1/4}n^{1/4} = \varepsilon^{5/2}n^{1/2}$ for $\varepsilon = n^{-1/9}$, and thus the smallest possible value of ε for which this proof works is $\Theta_k(n^{-1/9})$). Using Chebyshev's inequality, $\Pr(E_{\text{found}}) = \Pr(X > 0) \geq 9/10 > 2/3 + 1/50$, concluding the proof. \square

6 Lower bounds

In this section we provide proofs for our lower bounds in the non adaptive case. These are Theorems 3.2, 3.5 and 3.8. We start with the proof of Theorem 3.5. Then, we use it to finish the proof of Theorem 3.2, which requires us to prove a relatively simple special case of Theorem 3.8. Finally, we prove Theorem 3.8 in its full generality. We chose to present the proofs in this order for the sake of readability, as the proof of Theorem 3.8 will be easier to understand after tackling the special case considered in Theorem 3.2.

Proof of Theorem 3.5. Fix a permutation π of length k , and let $P = (\Lambda, S)$ be a unique signed partition of π of size u , where $\Lambda = (\sigma_1, \dots, \sigma_u)$ and $S = (s_1, \dots, s_u)$. Let $0 < \varepsilon < \varepsilon_0(k)$ and let $n > n_0(k)$ be an integer, where $\varepsilon_0(k) \leq 1/2k$ is small enough as a function of k and $n_0(k)$ is large enough as a function of k . We may assume, for convenience, that $m = n/k$ is an integer and that εm is an integer bigger than k (translating the result to any n and ε comes at a “price” of a multiplicative constant that depends only on k).

Recall that a one sided ε -test for π -freeness must always accept π -free sequences, and reject sequences that are ε -far from π -freeness with probability at least $2/3$. Thus, any one sided test T for π -freeness must always accept its input if the subsequence it queries is π -free. This follows from the fact that for any π -free sequence $g: [q] \rightarrow \mathbb{R}$, any $n > q$ and any $1 \leq t_1 < \dots < t_q \leq n$, there exists a π -free sequence $f: [n] \rightarrow \mathbb{R}$ such that $f(t_j) = g(j)$ for any $j = 1, \dots, q$. That is, any π -free queried subsequence might possibly be contained in a π -free sequence, and hence must be accepted by any one-sided test. Therefore, any one-sided test for π -freeness can be seen as a non-adaptive *search algorithm for π in f* , whose goal is to find a π -copy in an unknown input sequence f that is guaranteed to be ε -far from π -freeness, with success probability at least $2/3$.

We use Yao’s principle, constructing a family \mathcal{F} of sequences $f: [n] \rightarrow \mathbb{R}$ that are ε -far from π -freeness, which satisfies the following property for some constant $c_k > 0$. For any $1 \leq t_1 < \dots < t_q \leq n$ where $q < c_k \varepsilon^{-1/u} n^{1-1/u}$, it holds that

$$\Pr_{f \in \mathcal{F}} (\text{subsequence of } f \text{ in indices } t_1, \dots, t_q \text{ contains a } \pi\text{-copy}) < 2/3. \quad (2)$$

Combining (2) with a standard Yao-type argument completes the proof, as it implies that any (possibly probabilistic) search algorithm for π in f , where f is chosen uniformly at random from \mathcal{F} , must make $c_k \varepsilon^{-1/u} n^{1-1/u}$ queries to have success probability $2/3$.

Constructing \mathcal{F} In the rest of the proof, we present a family $\mathcal{F} = \mathcal{F}(P)$ for which (2) holds. Let us describe the structure of the sequences $f \in \mathcal{F}$ before diving into the technical details. A sequence $f \in \mathcal{F}$ looks like a blowup of f_P , where each blown up part is planted, starting at a random location, inside a longer interval (making it hard for a non-adaptive test to “guess” where each part is located inside its interval). More specifically, each part σ_i of the partition Λ corresponds to an interval I_i in f whose length is $|\sigma_i|m$. In this interval, there are εn copies of σ_i ordered in an increasing manner if s_i is a $+$, and a decreasing manner if σ_i is a $-$, where each σ_i -copy is a consecutive subsequence of f . The value of f on these σ_i -copies (for each i) is “aligned” with other intervals, so that f contains a set of εn pairwise disjoint π -copies, without containing any other π -copy (here we use the fact that $P = (\Lambda, S)$ is unique). The rest of the elements in each interval are chosen in a manner that does not create any other π -copy. To make \mathcal{F} “random enough,” the points where the consecutive copies begin in each interval are chosen uniformly at random. This assures that the probability for each k -tuple of entries in f to induce a π -copy is sufficiently small, proving (2).

We now provide the technical details. For every $1 \leq i \leq u$, write $\sigma_i = \pi[j_{i-1} + 1, j_i]$, where $j_0 = 0$ and let $\delta_i = j_i - j_{i-1}$ denote the length of σ_i . For any $1 \leq i \leq u$, let $I_i = \{mj_{i-1} + 1, \dots, mj_i\}$. A sequence $f: [n] \rightarrow \mathbb{R}$ is in \mathcal{F} if for any $1 \leq i \leq u$ there exists $0 \leq n_i \leq (1 - k\varepsilon)\delta_i m$, such that the following conditions hold.

- For every $1 \leq i \leq u$ where s_i is a $+$, and every $1 \leq l \leq \delta_i$ and $0 \leq r \leq \varepsilon n - 1$, we take $f(mj_{i-1} + n_i + r\delta_i + l) = r + \pi_{j_{i-1}+l}/2k$. We also take $f(x) = -1$ for any $mj_{i-1} + 1 \leq x \leq mj_{i-1} + n_i$, and $f(x) = n$ for any $mj_{i-1} + n_i + \varepsilon n \delta_i + 1 \leq x \leq mj_i$.

- For every $1 \leq i \leq u$ where s_i is a $-$, and every $1 \leq l \leq \delta_i$ and $0 \leq r \leq \varepsilon n - 1$, we take $f(mj_{i-1} + n_i + r\delta_i + l) = (\varepsilon n - 1 - r) + \pi_{j_{i-1}+l}/2k$. We also take $f(x) = n$ for any $mj_{i-1} + 1 \leq x \leq mj_{i-1} + n_i$, and $f(x) = -1$ for any $mj_{i-1} + n_i + \varepsilon n\delta_i + 1 \leq x \leq mj_i$.

Any $f \in \mathcal{F}$ is ε -far from π -freeness Any such f is ε -far from π -freeness. Indeed, for any $0 \leq r \leq \varepsilon n - 1$, the subsequence of f consisting of all k entries $x \in [n]$ for which $r < f(x) < r + 1$ is a π -copy, so there is a set \mathcal{D}_f of εn pairwise-disjoint π -copies in f .

Any $f \in \mathcal{F}$ does not contain non-trivial π -copies On the other hand, f does not contain any other (i.e., non-trivial) π -copy. To show this we use the fact that $P = (\Lambda, S)$ is a unique signed partition.

Claim 6.1. *Let $f \in \mathcal{F}$. If f contains a non-trivial π -copy, then it contains a non-trivial copy without the values -1 and m .*

Proof sketch. Recall that $k < \varepsilon m$. The proof follows by applying iteratively the following fact, and its symmetric counterpart. If $t = (t_1, \dots, t_k)$ is a π -copy in f , and if there exist $0 \leq r < \varepsilon n - 1$ and $i \in [k]$ such that $r - 1 \leq f(t_i) < r$, but there is no $j \in [k]$ for which $r < f(t_j) < r + 1$, then f also contains a π -copy created by the following “lifting process,” that replaces all entries with values between $r - 1$ (inclusive) and r (exclusive) with entries whose values are bigger than r and smaller than $r + 1$.

If $r > 0$, we replace any t_i satisfying $r - 1 < f(t_i) < r$ with the unique entry t' satisfying $f(t') = f(t_i) + 1$. If $r = 0$ we replace t_i with the closest entry t' among those satisfying $0 < f(t') < 1$. \square

Suppose now to the contrary that f contains a non-trivial π -copy in the entries $x_1 < \dots < x_k \in n$, without the values -1 and m , and let $R = \{ \lfloor f(x_i) \rfloor : 1 \leq i \leq k \} \subseteq \{0, 1, \dots, \varepsilon n - 1\}$, so $2 \leq |R| \leq k$. We now arbitrarily add elements from $\{0, 1, \dots, \varepsilon m - 1\}$ to R to obtain a set R' of size *exactly* k .

Let g be the subsequence of f over the set of entries $W(R') = \{w \in [n] : \lfloor f(w) \rfloor \in R'\}$. In particular $x_1, \dots, x_k \in W(R')$, so g contains a non-trivial π -copy. But this is a contradiction – the nature of our construction (and in particular, the choice of signs) implies that g is order-isomorphic to the sequence f_P given in [Definition 3.4](#), which does not contain non-trivial π -copies (as P is unique). Thus, the only π -copies in f are the trivial copies that come from \mathcal{D}_f .

Analysis: \mathcal{F} satisfies desired conditions Finally, we show that the probability for a k -tuple $1 \leq t_1 < \dots < t_k \leq n$ to induce a π -copy in a sequence $f \in \mathcal{F}$ chosen uniformly at random is sufficiently small. We may restrict ourselves to tuples containing exactly δ_i entries in I_i for any $1 \leq i \leq u$, as these are the only tuples with positive probability to induce a π -copy. Suppose that $f \in \mathcal{F}$ contains a π -copy in entries $t_1 < \dots < t_k$. This copy must come from \mathcal{D}_f , and so there exists some $0 \leq r \leq \varepsilon n - 1$ such that $r < f(t_l) < r + 1$ for any $1 \leq l \leq k$. The values of r and $t_{j_1}, t_{j_2}, \dots, t_{j_u}$ determine n_1, \dots, n_u uniquely. In other words, f is the only sequence, among all $|\mathcal{F}| > (n/2k)^u \geq (2k)^{-k} n^u$ sequences from \mathcal{F} , that has a π -copy of height between r and $r + 1$ whose j_i -th entry lies in t_{j_i} , for any $1 \leq i \leq u$. In total, only at most εn such possible choices $f \in \mathcal{F}$ have a π -copy whose j_i -th entry lies in t_{j_i} , for any $1 \leq i \leq u$. The following inequality summarizes the discussion.

$$\Pr_{f \in \mathcal{F}} (\text{subsequence of } f \text{ in indices } t_{j_1}, \dots, t_{j_u} \text{ is contained in a } \pi\text{-copy}) \leq \frac{\varepsilon n}{|\mathcal{F}|} < \frac{(2k)^{k\varepsilon}}{n^{u-1}} \quad (3)$$

We are now ready to finish the proof of (2). Pick $c_k = (3k)^{-k/u}$, and let $t = (t_1, \dots, t_q)$ with $1 \leq t_1 < \dots < t_q \leq n$ be a q -tuple, where $q < c_k \varepsilon^{-1/u} n^{1-1/u}$. t contains $\binom{q}{u} \leq q^u$ u -tuples, so by a union bound, the expected number of u -tuples contained in a π -copy (over a uniform choice $f \in \mathcal{F}$) is less than $(2k)^k \varepsilon q^u n^{1-u} < 2/3$. Thus, the probability that the subsequence of f on t contains a π -copy is less than $2/3$, as desired. \square

Proof of Theorem 3.2. Using Theorem 3.5, it is enough to show that $u(\pi) = k-1$ for any permutation π of length k satisfying $|\pi^{-1}(1) - \pi^{-1}(k)| = 1$. Let $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of length k , and assume, without loss of generality, that $\pi_\ell = 1$ and $\pi_{\ell+1} = k$ for some $1 \leq \ell \leq k-1$. We take the following signed partition $P = (\Lambda, S)$ of size $k-1$. $\Lambda = (\sigma_1, \dots, \sigma_{k-1})$ where σ_i consists of the single element π_i for any $i < \ell$, $\sigma_\ell = (1, k)$, and σ_i is the single element π_{i+1} for any $i > \ell$. The sign vector $S = (s_1, \dots, s_{k-1})$ is defined as follows. s_ℓ is a $-$, and for any $i \neq \ell$, s_i is a $+$ if and only if $\pi_i > \pi_{i+1}$.

We now show that P is unique, implying that $u(\pi) \geq |P| = k-1$, as needed. Consider the sequence $f = f_P$, as defined in Definition 3.4. We partition the entries of f_P into intervals I_1, \dots, I_{k-1} , where I_i contains all entries that participate in the σ_i -part of some π -copy in f_P . In other words, $I_i = \{(i-1)k+1, \dots, ik\}$ for any $1 \leq i < \ell$, $I_\ell = \{(\ell-1)k+1, \dots, (\ell+1)k\}$ and $I_i = \{ik+1, \dots, (i+1)k\}$ for any $\ell < i \leq k-1$.

Let $q = (q_1, \dots, q_\ell)$ be a π -copy in f_P . The following claim sheds light on the structure of q with respect to the intervals I_1, \dots, I_{k-1} .

Claim 6.2. *For any $i = 1, \dots, k$ let $\text{ind}(i)$ denote the index of the interval containing q_i , that is, $q_i \in I_{\text{ind}(i)}$. Then $\text{ind}(i) \geq i$ for any $i \leq \ell$ and $\text{ind}(i) \leq i-1$ for any $i \geq \ell+1$.*

Proof. Suppose to the contrary that $\text{ind}(i) < i$ for some $i \leq \ell$, and consider the smallest i satisfying this. Then $\text{ind}(i-1) = \text{ind}(i) = i-1$, that is, $q_{i-1}, q_i \in I_{i-1}$. This is a contradiction: If $\pi_i > \pi_{i-1}$ then s_{i-1} is a $-$, so the subsequence of f restricted to I_{i-1} is decreasing, contradicting the fact that q is a π -copy, that must satisfy $f(q_i) > f(q_{i-1})$ since $\pi_i > \pi_{i-1}$. If $\pi_i < \pi_{i-1}$ then s_{i-1} is a $+$ and, symmetrically, we have a contradiction. Thus, $\text{ind}(i) \geq i$ for any $i \leq \ell$. The proof that $\text{ind}(i) \leq i-1$ for any $i \geq \ell+1$ is symmetric. \square

As a special case of Claim 6.2, we conclude that $q_\ell, q_{\ell+1} \in I_\ell$ for any π -copy $q = (q_1, \dots, q_k)$. This implies that $f_P(q_\ell) = r + 1/2k$ and $f_P(q_{\ell+1}) = r + 1/2$ for some integer r (since the only length-2 subsequences of f inside I_ℓ that are increasing are $(r + 1/2k, r + 1/2)$, for any integer $0 \leq r \leq k-1$). Hence, for any $i \neq \ell, \ell+1$, q_i must be the unique entry satisfying $f(q_i) = r + \pi_i/2k$. We conclude that f_P does not contain non-trivial π -copies, so P is unique. \square

The proof of Theorem 3.8 is based on ideas that are similar to those of the proof of Theorem 3.2, and in particular, a generalized form of Claim 6.2 serves as an important tool in the proof.

Proof of Theorem 3.8. Let π be a permutation of length k , and let $E = \{\tau_1, \dots, \tau_t\}$ be an entangling of π whose resulting partition $\Lambda = \Lambda(E) = (\sigma_1, \dots, \sigma_d)$ is of size $d = d(\pi)$. For any $1 \leq \ell \leq t$, denote by $\lambda(\ell)$ the unique index satisfying $\tau_\ell = \sigma_{\lambda(\ell)}$. For any $1 \leq i \leq d$, write $\sigma_i = \pi[j_{i-1} + 1, j_i]$ where $j_0 = 0$ and $j_d = k$. We choose the sign vector $S = (s_1, \dots, s_d)$ as follows.

- For any $1 \leq i \leq d$ where σ_i contains more than one element, s_i is a $+$ if $\min \sigma_i$ lies after $\max \sigma_i$ in π , and otherwise s_i is a $-$.
- For any $i < \lambda(1)$ where σ_i is a single element, s_i is a $+$ if and only if $\pi_{j_i} > \pi_{j_i+1}$.

- For any $i > \lambda(1)$ where σ_i is a single element, s_i is a + if and only if $\pi_{j_i} < \pi_{j_{i-1}}$.

To finish the proof, we shall show that the signed partition $P = (\Lambda, S)$ is unique, implying that $u(\pi) \geq d = d(\pi)$. For this, we need to show that $f = f_P$ does not contain non-trivial π -copies. As in the proof of [Theorem 3.2](#), for any $1 \leq i \leq d$ let $I_i = \{kj_{i-1} + 1, \dots, kj_i\}$. Let $q = (q_1, \dots, q_k)$ be a π -copy in f_P . The following claim is the equivalent of [Claim 6.2](#) for our more general case.

Claim 6.3. *For any $i = 1, \dots, k$ let $\text{ind}(i)$ denote the index for which $q_i \in I_{\text{ind}(i)}$. Then $\text{ind}(j_{i-1} + 1) \geq i$ for any $i \leq \lambda(1)$ and $\text{ind}(j_i) \leq i$ for any $i \geq \lambda(1)$.*

Proof. We shall prove the claim for $i \leq \lambda(1)$, as the proof for $i \geq \lambda(1)$ is symmetric. Suppose to the contrary that there exists $i \leq \lambda(1)$ for which $\text{ind}(j_{i-1} + 1) < i$, and consider the smallest such i . Then $\text{ind}(j_{i-2} + 1) \geq i - 1$, and so $\text{ind}(j) = i - 1$ for any $j_{i-2} + 1 \leq j \leq j_{i-1} + 1$.

We show that I_{i-1} does not contain a copy of $\pi[j_{i-2} + 1, j_{i-1} + 1]$, leading to a contradiction. If $|\sigma_{i-1}| = 1$ then the choice of the sign s_{i-1} is a + if $\pi_{j_{i-1}+1} < \pi_{j_{i-1}}$, and a - otherwise; in the first case, the entries in I_{i-1} are increasing and so it cannot contain $\pi[j_{i-2}, j_{i-1} + 1]$, which is a decreasing sequence, a contradiction. In the other case we also get a contradiction, symmetrically. Thus, from here onwards we may assume that σ_{i-1} contains more than one element.

The choice of the sign s_{i-1} implies that the only σ_{i-1} -copies in the subsequence of f_P on the interval I_{i-1} are the trivial ones, i.e., those that contain all $|\sigma_{i-1}|$ elements between r and $r + 1$ for some integer $0 \leq r \leq k - 1$. Thus, we may assume that $r < f_P(q_j) < r + 1$ for any $j_{i-2} + 1 \leq j \leq j_{i-1}$.

Without loss of generality, assume that s_{i-1} is a +; this corresponds to the case where $\max \sigma_{i-1}$ lies before $\min \sigma_{i-1}$ in π . Since E is an entangling, we know that σ_{i-1} is not shadowed with respect to $\sigma_{\lambda(1)}$. This means that $\pi_{j_{i-1}+1} < \max \sigma_{i-1}$, and so $f_P(q_{j_{i-1}+1}) < r + 1$. But this contradicts the fact that $\text{ind}(j_{i-1} + 1) = i - 1$: All $|\sigma_{i-1}|$ entries in I_{i-1} whose value is between r and $r + 1$ are assigned to $q_{j_{i-2}+1}, \dots, q_{j_{i-1}}$, and all entries x of I_{i-1} that come after these entries satisfy $f_P(x) > r + 1$. In particular, $q_{j_{i-1}+1} \in I_{i-1}$ so $f_P(q_{j_{i-1}+1}) > r + 1$, a contradiction. \square

To show that q is a trivial π -copy, we prove the following claim by induction.

Claim 6.4. *There exists an integer $r = r(q)$ where $0 \leq r \leq k - 1$, satisfying the following. For any $1 \leq l \leq t$, and any $j_{\lambda(\ell)-1} + 1 \leq j \leq j_{\lambda(\ell)}$, it holds that $q_j \in I_{\lambda(\ell)}$, and more specifically, $f_P(q_j) = r + \pi_j/2k$.*

Proof. The proof is by induction on ℓ . By [Claim 6.3](#), $q_{j_{\lambda(1)-1}+1}, \dots, q_{j_{\lambda(1)}} \in I_{\lambda(1)}$. By our choice of the sign $s_{\lambda(1)}$, there must be some integer $0 \leq r \leq k - 1$, such that for any $j_{\lambda(1)-1} + 1 \leq j \leq j_{\lambda(1)}$, q_j is the unique entry of f_P satisfying $f_P(q_j) = r + \pi_j/2k$. This settles the case $\ell = 1$.

Now let $\ell > 1$, and assume that $f_P(q_j) = r + \pi_j/2k$ for any $j_{\lambda(\ell')-1} + 1 \leq j \leq j_{\lambda(\ell')}$ where $1 \leq \ell' < \ell$. We need to show that $f_P(q_j) = r + \pi_j/2k$ for any $j_{\lambda(\ell)-1} + 1 \leq j \leq j_{\lambda(\ell)}$.

For any $j', j'' \in [k]$ for which we already know that $f_P(q_{j'}) = r + \pi_{j'}/2k$, $f_P(q_{j''}) = r + \pi_{j''}/2k$, and $\pi_{j'} < \pi_{j''}$, it must be true that $f_P(q_j) = r + \pi_j/2k$ for any j satisfying $\pi_{j'} < \pi_j < \pi_{j''}$. To see this, note that the number of entries of f_P with value between $f_P(q_{j'})$ and $f_P(q_{j''})$ (not including $f_P(q_{j'}), f_P(q_{j''})$ themselves) is exactly $\pi_{j''} - \pi_{j'} - 1$. Since q is a π -copy, it also contains exactly $\pi_{j''} - \pi_{j'} - 1$ entries with value between $f_P(q_{j'})$ and $f_P(q_{j''})$, so these entries of q must be precisely all entries of f_P whose value lies in this range.

Without loss of generality, assume that $\lambda(\ell) < \lambda(1)$ (that is, $\tau_\ell = \sigma_{\lambda(\ell)}$ lies before $\tau_1 = \sigma_{\lambda(1)}$ in π). Since E is an entangling, we know that $\pi_{j'} < \pi_{j_{\lambda(\ell)}} < \pi_{j''}$ for some $\pi_{j'}, \pi_{j''} \in \bigcup_{\ell' < \ell} \tau_{\ell'}$. By the previous paragraph, $f_P(q_{j_{\lambda(\ell)}}) = r + \pi_{j_{\lambda(\ell)}}/2k$, also implying that $\text{ind}(j_{\lambda(\ell)}) = \lambda(\ell)$. By [Claim 6.3](#),

$\text{ind}(j_{\lambda(\ell)-1} + 1) \geq \lambda(\ell)$, so we get that $\text{ind}(j) = \lambda(\ell)$ for any $j_{\lambda(\ell)-1} + 1 \leq j \leq j_{\lambda(\ell)}$. Considering our choice of the sign s_ℓ , it follows that $f_P(q_j) = r + \pi_j/2k$ must hold for any $j_{\lambda(\ell)-1} + 1 \leq j \leq j_{\lambda(\ell)}$. This concludes the inductive proof. \square

With [Claim 6.4](#) it is easy to finish the proof. Since E is an entangling, there exist $1 \leq \ell, \ell' \leq d$ such that $1 \in \tau_\ell = \sigma_{\lambda(\ell)}$ and $k \in \tau_{\ell'} = \sigma_{\lambda(\ell')}$, implying that $f_P(q_{\pi^{-1}(1)}) = r+1/2k$ and $f_P(q_{\pi^{-1}(k)}) = r+1/2$ for some $0 \leq r \leq r+1$. Thus, $r < f_P(q_j) < r+1$ for any $1 \leq j \leq k$. Since there are exactly k entries $x \in [k^2]$ for which $r < f_P(x) < r+1$, q must be a trivial π -copy. Therefore, P is unique. \square

We finish with an (easy) proof of [Theorem 3.3](#) that builds on [Corollary 3.9](#).

Proof of [Theorem 3.3](#). Let $\pi = (\pi_1, \dots, \pi_k)$ be a permutation of length k chosen uniformly at random. Without loss of generality assume that $\pi_i = 1, \pi_j = k$ for some $i < j$. The probability that $\pi_{i+1} \leq k^{3/4}$ or $\pi_{j-1} \geq k - k^{3/4}$ or $|i - j| < k^{3/4}$ is $O(k^{-1/4})$. Conditioning on the event that none of the above happens, the probability that there exists no $i+1 < x < j-1$ for which $\pi_x < k^{3/4}$ and $\pi_{x+1} > k - k^{3/4}$ is also bounded by $O(k^{-1/4})$ (it is actually exponentially smaller than that). If none of these events happens, then $d(\pi) \geq k - 3$, as there exists some $i < x < y$ for which $((\pi_x, \pi_{x+1}), (1, \pi_{i+1}), (\pi_{j-1}, k))$ is an entangling. Indeed, $(1, \pi_{i+1})$ and (π_{j-1}, k) cannot be shadowed with respect to (π_x, π_{x+1}) , and the two other conditions of an entanglement hold since $\pi_x < \pi_{i+1}, \pi_{j-1} < \pi_{x+1}$. Thus $d(\pi) \geq k - 3$ with probability at least $1 - O(k^{-1/4})$, as desired.

As an added bonus, note that $\Pr(d(\pi) \geq k - 2) \geq 19/24 - O(1/k)$: Suppose that $i > 1, j < n$, and $j \geq i + 2$ (all of these hold with probability $1 - O(1/k)$). Consider the event where either $\max\{\pi_{i-1}, \pi_{i+1}\} \geq \pi_{j-1}$ or $\min\{\pi_{j-1}, \pi_{j+1}\} \leq \pi_{i+1}$. This event has probability $19/24$, and if it occurs, one can verify that $d(\pi) \geq k - 2$. \square

7 Hierarchy of adaptivity

In this section, we establish an adaptivity hierarchy theorem for testing $(1, 3, 2)$ -freeness, by proving [Theorem 3.13](#) and [Theorem 3.14](#). The algorithm behind the former is obtained by a rather natural modification of the adaptive test of [\[NRRS17\]](#): We replace the only adaptive component of this test – a variant of a binary search – by a less query-efficient, but adaptivity-limited subroutine performing the corresponding search by a recursive partitioning of the search space. The lower bound in our case is shown by a reduction to a clean problem, `TEMPLATE-SEARCH` (a variant of a similar problem introduced by Newman et al., `INTERSECTION-SEARCH` in the context of *non-adaptive* algorithms). Showing a lower bound on `TEMPLATE-SEARCH` for adaptive tests with r rounds, however, turns out to be far from straightforward. Our inductive proof relies on a connection with anti-concentration of Binomial distributions to argue that the “uncertainty” left to the algorithm does not decay too fast with every stage, but rather at most by a square root (or, put differently, that the algorithm cannot restrict its search space by more than a square root at each step).

7.1 The upper bound part of the hierarchy

We hereby prove [Theorem 3.13](#), restated below:

Theorem 3.13. *For every $1 \leq r \leq (\log n)^{O(1)}$, there exists an r -round (one-sided) testing algorithm for $(1, 3, 2)$ -freeness with query complexity $\tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right)$.*

Proof. As aforementioned, our upper bound relies on a modification of the adaptive algorithm of Newman et al. ([NRRS17, Theorem 5.1]). Before explaining our modification of this algorithm, we briefly sketch how the original works, and introduce the necessary notation.

The (1, 3, 2) adaptive test of Newman et al. The adaptive test in [NRRS17] runs (in parallel) two different tests, call them TEST1 and TEST2, each trying to catch a violation of a specific kind; and rejects if any of them does find a violation. (Hence, the test is clearly one-sided.) We hereafter assume that f is ε -far from (1, 3, 2)-free. This implies that there exists a matching T of (1, 3, 2)-tuples of size at least $\varepsilon n/3$; which can be partitioned as $T = T_1 \cup T_2$, where T_1 contains the tuples (i, j, k) such that $j - i \geq k - j$ and T_2 those for which $j - i < k - j$.

- If T_1 has size at least $|T_1| \geq \frac{\varepsilon n}{6}$, then TEST1 outputs reject with probability at least $2/3$, and makes $O(\log^5 n / \varepsilon^3)$ non-adaptive queries.
- If T_2 has size at least $|T_2| \geq \frac{\varepsilon n}{6}$, then TEST2 outputs reject with probability at least $\text{poly}(\varepsilon, 1/\log(n))$, which can be amplified by running it independently $\text{poly}(1/\varepsilon, \log(n))$ times. Here too, there are two cases to consider, where in what follows

1. $I_L, I_R \subseteq [n]$ are disjoint intervals considered by (and known to) the algorithm, where all elements of the “left interval” I_L lie before those of the “right interval” I_R .
2. $I'_R \subseteq I_R$ is a subset implicitly defined as $I'_R = \{k \in I_R : f(k) > \alpha\}$ (where α is a value obtained and known by the algorithm in a previous step), and unknown to the algorithm.

I_R, I_L, α are all induced by the behavior of the algorithm TEST2 on steps 1 to 4 of the algorithm described in [NRRS17]; these steps perform a total of $q := O(\log^{20} n / \varepsilon^2)$ queries to f . We condition on these steps to be successful, which happens with probability $\Omega(\varepsilon / \log^2 n)$ by the analysis of [NRRS17]), so that

1. If $f|_{I'_R}$ is $\frac{1}{\log^5 n}$ -far from monotone (i.e., does not contain any monotone subsequence of length at least $(1 - \frac{1}{\log^5 n})$), then Step 5 (which also makes q queries to f) will output reject with probability $1 - o(1)$.
2. If $f|_{I'_R}$ is $\frac{1}{\log^5 n}$ -close to monotone, then Step 6 (which makes $O(1)$ queries to f) will return with probability $\Omega(\varepsilon^2 / \log^4 n)$ a pair $(i, j) \in I_L^2$ such that
 - (a) $i < j$ and $\alpha < f(i) < f(j)$.
 - (b) There exists $k \in I'_R$ such that $(i, j, k) \in T$.

Then, Step 7 leverages (a), (b), and the near-monotonicity of $f|_{I'_R}$, to find some $k' \in I_R$ such that $f(k') \in (f(i), f(j))$.

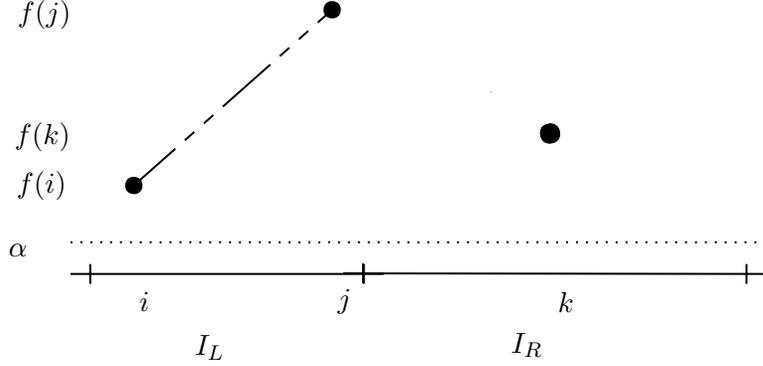


Figure 1: Our starting assumptions for Step 7.

The only adaptive part of the test is Step 7 of the second case (TEST2), and this is the one which is amenable to improvements with bounded adaptivity. Indeed, this step is implemented in [NRRS17] in two different ways, to obtain respectively an adaptive and a non-adaptive test:

- (a) An *adaptive* method making a “filtered” binary search, with query complexity $\text{poly}(\log n, 1/\varepsilon)$.
- (b) A *non-adaptive* sampling method, with query complexity $O((\sqrt{n} \log^2 n)/\varepsilon)$.

To prove our theorem, it is thus sufficient to explain how to replace the adaptive method with one that uses r rounds of queries, which we do next.

Our goal, and high-level idea We have to solve the following problem. We are given values $a := f(i)$, $b := f(j)$ with $a < b$, and a lower bound $\alpha := f(i_0)$ with $\alpha < a$; and granted query access to f over an interval $I := I_R$ (known), with the guarantee that there exists an (unknown) subset $J := I'|_R \subseteq I$ such that:

- $i \in I$ is in J iff $f(i) > \alpha$;
- J has size at least $\delta |I|$ for $\delta := \frac{\varepsilon}{8 \log n}$;
- J contains at least ℓ indices k such that $a < f(k) < b$ (for ℓ to be determined later);
- $f|_J$ is η -close to monotone non-decreasing, for $\eta := \frac{1}{\log^5 n}$.

The goal is to output, with non-trivial probability, either an index $k \in I$ such that $a < f(k) < b$ or two indices $k < k' \in I$ such that $\alpha < f(k') < f(k)$.

At a high-level, our algorithm acts recursively as follows. In any given round (other than the last one), our algorithm divides the interval I_R into s randomly chosen intervals, that partition I'_R into roughly equal pieces with high probability, and queries the endpoints of these intervals. Then, using the near-monotonicity of $f|_{I'_R}$, the algorithm focuses on a single one of these intervals, which will be the I_R of choice for the next round of queries. The last round just consists of sampling enough points uniformly at random, which should hit (with constant probability) at least one of the k witnesses which are now concentrated in our last, much smaller interval. The issues here are that, of course: (i) we do not know I'_R , only I_R (though testing membership in I'_R is easy); (ii) that f is only *close* to monotone, not exactly monotone; and (iii) that f is close to monotone on I'_R , *not* on I_R . In spite of this, we can make the above approach work, losing only some logarithmic factors in the query complexity.

We need the following well-known fact.

Proposition 7.1 (See [DN04, Section 6.4]). *Suppose $N - 1$ points are drawn uniformly and independently at random from $[0, 1]$, and let X_1, \dots, X_N be the lengths of the N segments they partition $[0, 1]$ into. Letting $\Delta_N := \max_{1 \leq k \leq N} X_k$, we have $\mathbb{E}[\Delta_N] = \frac{H_N}{N}$ where H_N is the N -th Harmonic number.*

Combining **Proposition 7.1** with Markov inequality, it follows that by taking $N = O(\frac{s}{r} \log \frac{s}{r})$ uniformly random and independent points from any ordered set S (with $s = o(|S|)$), with probability at least $1 - \frac{1}{10r}$ the partition of S into N “intervals” induced by these points is such that no interval contains more than an $\frac{1}{s}$ fraction of S .

The last thing we need, for our recursion to go through, is to ensure that J (which has density δ in I) still has density roughly δ in all sub-intervals of I . This is not necessarily true *per se*, but we can rely on [NRRS17, Lemma 5.3] to get the following. There exists a subset $J' \subset J$ of size at least $|J'| \geq (1 - \frac{3}{\log^2 n})|J|$ such that every $i \in J'$ satisfies the property below. For every interval $K \ni i$, we have $|K \cap J'| \geq \gamma|K|$, for $\gamma := \varepsilon/\log^2 n$. (In the terminology of [NRRS17], no element of J' is γ -deserted.)

Then, we have the following:

- $J' \subseteq J$;
- J' has size at least $\delta'|I|$ for $\delta' := \frac{\varepsilon}{8 \log n} (1 - \frac{3}{\log^2 n}) \geq \gamma$, and contains no γ -deserted elements;
- J contains at least ℓ indices k such that $a < f(k) < b$ (for ℓ to be determined later);
- $f|_{J'}$ is $(\eta + \gamma)$ -close to monotone non-decreasing.

(The only non-immediate point is the third, which will follow from the way we guarantee the existence of these ℓ elements.) With all this in place, we can describe the algorithm.

- 1: Set $s := n^{\frac{1}{r+1}}$
- 2: **for** $1 \leq t \leq r - 1$, given a current interval I_t **do**
- 3: Query f on $N := O(\frac{\log r}{\gamma} \cdot \frac{s}{r} \log \frac{s}{r})$ uniformly random and independent points from I_t
- 4: Let S_t be those of these points which belong to J , i.e. whose f -values is at least a
- 5: **if** $S_t = \emptyset$ **then** ▷ Happens with negligible probability if f when far from $(1, 3, 2)$ -free
- 6: **return accept**
- 7: **end if**
- 8: **if** any of these values is in (a, b) **then**
- 9: **return reject** ▷ We found some k' with $(f(i), f(j), f(k'))$ being an $(1, 3, 2)$ -pattern
- 10: **else if** the sequence corresponding to the f -values of S_t is not non-increasing **then**
- 11: **return reject.** ▷ We found some k, k' with $(f(i_0), f(k'), f(k))$ being an $(1, 3, 2)$ -pattern
- 12: **else**
- 13: Let I_{t+1} be the (only) interval induced by the points in S_t such that $I_{t+1} \subseteq (f(i), f(j))$;
- 14: Recurse on I_{t+1} .
- 15: **end if**
- 16: **end for**

We condition on not outputting **accept** during any of the $r - 1$ rounds, and then on every of the rounds being such that $|I_{t+1}| \leq \frac{1}{s}|I_t|$. By a union bound, this happens with probability at least $9/10$; the first, by a union bound and the fact that $|J' \cap I_t| \geq \gamma|I_t|$ (from the non-desertion property) and the $O((\log r)/\gamma)$ factor in N . The second, again by a union bound over the $r - 1$ rounds, and **Proposition 7.1**.

Thus, at the end, we either have found already an $(1, 3, 2)$ -pattern and rejected, or are left with

an interval I_r of size at most $\frac{n}{s^{r-1}} = \frac{n}{r-1} = n^{\frac{2}{r+1}}$ which contains at least ℓ indices $k \in I'_R \subseteq I_R$ such that $(i, j, k) \in T$ (This last point by monotonicity: all such indices must be in the remaining interval, or we would have found one already). At this stage, by taking (in the r th (last) round of queries) $O\left(n^{\frac{2}{r+1}}/\ell\right)$ independent and uniformly distributed queries in I_r , we find such a k with high constant probability. The total number of queries made is upper bounded by

$$(r-1) \cdot \left(n^{\frac{1}{r+1}} + 1\right) + O\left(n^{\frac{2}{r+1}}/\ell\right)$$

which for $\ell := \tilde{O}\left(\varepsilon n^{\frac{1}{r+1}}\right)$ (recalling that $r \leq \text{poly log}(n)$) is $\tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right)$.

Last step: the promise of these ℓ witnesses It only remains to describe how to achieve the guarantee of having at least ℓ “witnesses” in J' (with high constant probability). This is done by calling the `DyadicSampler` (Algorithm 3.1) of [NRRS17] $O(\ell \varepsilon^{-2} \log^2 n)$ times; doing so, one gets with high constant probability a set of ℓ many $(1, 2)$ pairs, each dominating a different $(1, 2)$ -pair in T_2 (similarly as in [NRRS17, Section 5.2]). By considering (i, j) among these $(1, 2)$ pairs such that $(f(i), f(j))$, we then have (i, j) for which there exist at least ℓ different indices $k \in I'_R$ such that (i, j, k) is a $(1, 3, 2)$ -pattern.

Since $|J'| = (1 - o(1)) |I'_R|$, we can additionally guarantee at the cost of a small multiplicative factor that we get ℓ points in J' (not only in $J = I'_R$). Further, these calls to the `DyadicSampler` can all be done in parallel, and in parallel to the first round of queries, thus preserving the number of rounds r .

Query and round complexity Thus, overall the query complexity of this step is

$$O(\ell \varepsilon^{-2} \log^2 n) + \tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right) = \tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right) + \tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right) = \tilde{O}\left(\varepsilon^{-1} n^{\frac{1}{r+1}}\right)$$

as claimed, using r rounds of adaptivity; and this algorithm, under the assumptions laid out when describing “our goal,” rejects with constant probability. Integrating it into the (non-adaptive) rest of the algorithm behind [NRRS17, Theorem 5.1] yields the theorem. \square

7.2 The lower bound part of the hierarchy

We now establish the lower bound part of our adaptivity hierarchy for testing $(1, 3, 2)$ -freeness, namely [Theorem 3.14](#):

Theorem 3.14. *For every $0 \leq r \leq (\log \log n)^{O(1)}$, any r -round (two-sided) testing algorithm for $(1, 3, 2)$ -freeness must make $\Omega\left(n^{1/2^{r+3}}\right)$ queries.*

In order to do so, we first introduce a related problem, that we refer to as `TEMPLATE-SEARCH` – a variant of the `INTERSECTION-SEARCH` problem defined in [NRRS17]. We then show our lower bound on the query complexity of any r -round algorithm solving this problem in [Section 7.2.1](#), before showing in [Section 7.2.2](#) how to reduce the `TEMPLATE-SEARCH` problem to testing $(1, 3, 2)$ -freeness, while preserving both query complexity and number of rounds of adaptivity.

7.2.1 Lower bound on Template-Search

We start by defining the problem we consider in this subsection:

Definition 7.2 (TEMPLATE-SEARCH). An instance of this problem is a tuple (S, T) , where $|T| = m$, $|S| = 3m$, and S, T are two non-decreasing arrays of elements from \mathbb{R} . Furthermore, T is a consecutive subarray of S , that is there exists an integer Δ such that $S_{i+\Delta} = T_i$ for all $1 \leq i \leq m$. The goal is, granted query access to both S and T (i.e., one can ask for the i -th element of either array), to determine the value Δ .

Our goal is to prove that this problem is “hard” for r -round algorithms, as formalized in the theorem below:

Theorem 7.3. *Let $r \leq \log^{O(1)} m$. Any (possibly randomized) r -round algorithm that correctly solves TEMPLATE-SEARCH with probability at least $2/3$ must make $\Omega(m^{1/2^{r+3}})$ queries.*

Proof. By (the easy direction of) Yao’s principle, it is sufficient to present a particular distribution over TEMPLATE-SEARCH instances such that any *deterministic* algorithm which succeeds with probability at least $2/3$ over a random choice of $(S, T) \sim \mathcal{D}$ must make $\Omega(m^{1/2^{r+3}})$.

Our distribution \mathcal{D} is as follows: We generate two tuples S, T of elements from \mathbb{R} and of size $|S| = 3m$ and $|T| = m$, as follows.

- T is obtained by drawing m numbers independently at random from $(0, 1)$, and sorting them;
- S is obtained by choosing an offset Δ uniformly at random in $\{0, \dots, 2m\}$, and setting

$$S := \underbrace{(-1, \dots, -1)}_{\Delta \text{ times}} \sqcup T \sqcup \underbrace{(2, \dots, 2)}_{2m-\Delta \text{ times}}$$

Note that with probability 1 all elements of T are distinct; we will assume this is the case in the rest of the proof.

Fix an arbitrary r -round algorithm \mathcal{A} for TEMPLATE-SEARCH with query complexity q . For an index $i \in [m]$ (corresponding to an element T_i in T), we write $\delta(i) = \Delta + i$ for its position in S , i.e. $S_{\delta(i)} = T_i$ for all $i \in [m]$ (and accordingly write $\delta(T)$ for the position of the “template” of T inside S). Moreover, for simplicity (up to losing some poly $\log(n)$ factors in the lower bound), we can assume that the algorithm makes exactly $2q$ queries in every of the r rounds, namely q to S and q to T ; and will proceed by induction.

We let $d_0 := m$, and $d_{\ell+1} = \alpha \frac{\sqrt{d_\ell}}{r q^2}$ for $0 \leq \ell \leq r-1$ (where $\alpha > 0$ is an absolute constant determined in the course of the analysis). First, we apply the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [DKW56, Mas90] to the m i.i.d. samples defining T to argue that, with probability at least $5/6$, the empirical distribution defined by these m samples has Kolmogorov distance (i.e., maximum pointwise distance between the cumulative distribution functions) at most $O(1/\sqrt{m})$ from the uniform distribution; or, equivalently, that simultaneously every interval $[a, b] \subseteq [m]$ (recalling that T is sorted) is such that $T_b - T_a$ is within an additive $O(1/\sqrt{m})$ of $\frac{1}{m}(b-a)$. We hereafter condition on this.

Consider now the ℓ -th round, for $0 \leq \ell \leq r-1$. Letting $S_\ell = \{s_i\}_{1 \leq i \leq \ell q}$ and $T_\ell = \{t_i\}_{1 \leq i \leq \ell q}$ be the set of queries made, in the previous rounds, to S and T respectively. Our induction hypothesis is that

$$(\dagger) \text{ for all } 1 \leq i, j \leq \ell q, |s_i - \delta(t_j)| \geq d_\ell;$$

In other words, this asks that no query in S lies “too close” to the location in S of a query made in T . (Note that as $S_0, T_0 = \emptyset$, the initial case of our induction trivially holds.) Our goal then is to show that, with high probability over the choice of (S, T) , we will have that **item (†)** still hold at round $\ell + 1$, for $d_{\ell+1} \approx \sqrt{d_\ell}$ as defined above.

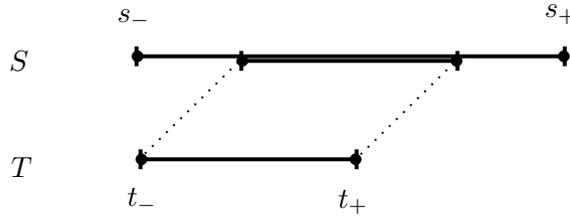
Fix any pair $(s, t) \in [n] \times [m]$ that \mathcal{A} queries (in respectively S and T) in the ℓ -th round; and let $s_-, s_+ \in S_\ell, t_-, t_+ \in T_\ell$ be the indices previously queried with

$$\begin{aligned} s_- &:= \max \{ s' \in S_\ell : s' < s \} & s_+ &:= \max \{ s' \in S_\ell : s' > s \} \\ t_- &:= \max \{ t' \in T_\ell : t' < t \} & t_+ &:= \max \{ t' \in T_\ell : t' > t \} \end{aligned}$$

so that (s_-, s_+) and (t_-, t_+) are the minimal intervals delimited by previously queried points which contain s and t . Note that if $s_- > \Delta$ or $s_+ < \Delta$, then s will fall at least as far from $\delta(T)$ as s_- and s_+ , revealing no new information; thus, we assume afterwards that $[s, s_-] \cap \delta(T) \neq \emptyset$.

We divide the analysis in 3 cases, depending on the relative position of these points:

Case $s_- < \delta(t_-) < \delta(t_+) < s_+$: by **item (†)**, we must have $|\delta(t_-) - s_-|, |\delta(t_+) - s_+| \geq d_\ell$. Then, no matter where the query $t \in T$ is made, we will have $\min_{s' \in S_\ell} |t - s'| = \min(|t - s_-|, |t - s_+|) \geq d_\ell$, so our bound on $d_{\ell+1}$ is determined by $|t - s|$ (i.e., by the position of the new query s).



Over the choice of S, T , the number of elements that fall in $(s_-, \delta(t_-))$ is distributed as a Binomial random variable X with parameters

$$N := (s_+ - \delta(t_+)) + (\delta(t_-) - s_-) \geq 2d_\ell, \quad p := \frac{T_{t_-} - S_{s_-}}{(T_{t_-} - S_{s_-}) + (S_{s_+} - T_{t_+})}$$

which has standard deviation $\sqrt{Np(1-p)} \geq \sqrt{Np/2}$.³

³Without loss of generality, we here assume $p \leq 1/2$; otherwise, we proceed with the same argument, but considering the number of elements in $(s_+, \delta(t_+))$ instead of $(s_-, \delta(t_-))$, and $1 - p$ instead of p .

We lower bound this variance as follows.

$$\begin{aligned}
Np &= (s_- - \delta(t_-)) \cdot \left(1 + \frac{\delta(t_+) - s_+}{s_- - \delta(t_-)}\right) \cdot \frac{1}{1 + \frac{S_{s_+} - T_{t_+}}{T_{t_-} - S_{s_-}}} \\
&\geq d_\ell \cdot \left(1 + \frac{\delta(t_+) - s_+}{s_- - \delta(t_-)}\right) \cdot \frac{1}{1 + \frac{S_{s_+} - T_{t_+}}{T_{t_-} - S_{s_-}}} \\
&\geq d_\ell \cdot \left(1 + \frac{\delta(t_+) - s_+}{s_- - \delta(t_-)}\right) \cdot \frac{1}{1 + \frac{\delta(t_+) - s_+ + O(\sqrt{m})}{s_- - \delta(t_-) - O(\sqrt{m})}} \quad (\text{DKW}) \\
&\geq d_\ell \cdot \left(1 + \frac{\delta(t_+) - s_+}{s_- - \delta(t_-)}\right) \cdot \frac{1}{1 + \frac{2(\delta(t_+) - s_+)}{(s_- - \delta(t_-))/2}} \quad (d_\ell = \Omega(\sqrt{m})) \\
&\geq \frac{d_\ell}{4} \quad (\dagger)
\end{aligned}$$

which holds as long as $d_\ell \geq C\sqrt{m}$ for some absolute constant $C > 0$.

Since this number of elements X fully characterizes, in this first case, the distance from s (whose position in (s_-, s_+) is known to the algorithm) to $\delta(t_-)$, $\delta(t)$, and $\delta(t_+)$ (whose relative position is the same as that of t_-, t, t_+ and thus also known to the algorithm), the distance of s to $\delta(T)$ is equivalent to the realization of X . Thus, it is sufficient to show that with high probability any algorithm would have error at least $d_{\ell+1} \approx \sqrt{d_\ell}$ when guessing X . This follows from anticoncentration of Binomial distributions: namely,

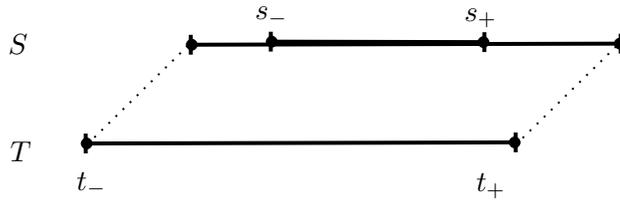
$$\inf_{x \in \mathbb{R}} \Pr_X[|X - x| \leq d_{\ell+1}] = \Pr_X[|X - \mathbb{E}[X]| \leq d_{\ell+1}] = O\left(\frac{d_{\ell+1}}{\sqrt{pN}}\right) \leq \frac{1}{30rq^2}$$

the last equality by our choice of $d_{\ell+1} = O\left(\frac{\sqrt{d_\ell}}{rq^2}\right)$, and the fact that $Np = \Omega(d_\ell)$.

By the above discussion, and a union bound over the 3 events, we get that

$$\Pr\left[\min_{u \in \{t', t, t''\}} |s - \delta(u)| \leq d_{\ell+1}\right] \leq \frac{1}{10rq^2}.$$

Case $\delta(t_-) < s_- < s_+ < \delta(t_+)$: by **item (\dagger)** , we must also have $|\delta(t_-) - s_-|, |\delta(t_+) - s_+| \geq d_\ell$.



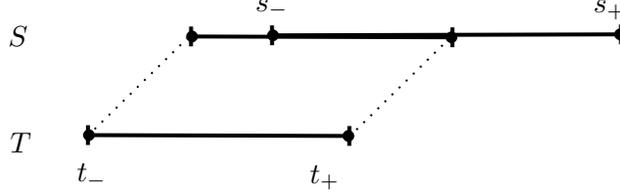
The situation is the same as in the first case, with the roles of (t_-, t, t_+) and (s_-, s, s_+) swapped. Namely, the distance of $\delta(t)$ to s_-, s_+ is now fully captured by the number of elements in $(\delta(t_-), s_-)$, which is (over the choice of S, T) distributed as a random variable X following a Binomial distribution with parameters

$$N := (\delta(t_+) - s_+) + (s_- - \delta(t_-)) \geq 2d_\ell, \quad p := \frac{S_{s_-} - T_{t_-}}{(S_{s_-} - T_{t_-}) + (S_{s_+} - T_{t_+})}.$$

The same argument then shows that

$$\Pr \left[\min_{u \in \{s', s, s''\}} |\delta(t) - u| \leq d_{\ell+1} \right] \leq \frac{1}{10rq^2}.$$

Case $\delta(t_-) < s_- < \delta(t_+) < s_+$ (equivalent by symmetry to $s_- < \delta(t_-) < s_+ < \delta(t_+)$): by **item (†)**, we must have $|\delta(t_-) - s_-|, |\delta(t_+) - s_-|, |\delta(t_+) - s_+| \geq d_\ell$.



In this case, the distance of $\delta(t)$ to s_-, s_+ and s to $\delta(s_-), \delta(s_+)$ are entirely characterized (from the point of view of \mathcal{A}) by the unknown number of elements in $(s_-, \delta(t_+))$, which is this time (over the choice of S, T) distributed as a random variable X following a Binomial distribution with parameters

$$N := s_+ - s_- = (s_+ - \delta(t_+)) + (\delta(t_+) - s_-) \geq 2d_\ell, \quad p := \frac{T_{t_+} - S_{s_-}}{S_{s_+} - S_{s_-}} = \frac{T_{t_+} - S_{s_-}}{(S_{s_+} - T_{t_+}) + (T_{t_+} - S_{s_-})}.$$

The same argument as before then shows that

$$\Pr \left[\min_{\substack{u \in \{s', s, s''\} \\ v \in \{t', t, t''\}}} |\delta(v) - u| \leq d_{\ell+1} \right] \leq \frac{1}{6rq^2}$$

(where we did a union bound over the 5 events “ s close to $\delta(t')$,” “ s close to $\delta(t)$,” “ s close to $\delta(t'')$,” “ $\delta(t)$ close to s' ,” and “ $\delta(t)$ close to s'' ”).

A union bound over all q^2 pairs of queries (s, t) then guarantees that, with probability at least $\frac{5}{6r}$, **item (†)** still holds at round $\ell + 1$.

This concludes the induction part of the argument; to finish the proof, observe that by a union bound over all r rounds and the application of the DKW inequality, we get that **item (†)** holds all through the execution with probability at least $1 - (r \cdot 1/(6r) + 1/6) = 2/3$ (over the choice of (S, T)). But since the algorithm is only successful when it finds the value of Δ (i.e., when $S_r \cap \delta(T_r) \neq \emptyset$), we must have $d_r \leq 1$. In particular, there exists some stage $1 \leq L \leq r$ such that $d_L \leq C\sqrt{m} < d_{L-1}$,⁴ which leads to

$$C\sqrt{m} \geq d_L = \frac{\alpha}{rq^2} \sqrt{d_{L-1}} = \dots = \left(\frac{\alpha}{rq^2} \right)^{\sum_{a=0}^{L-1} 1/2^a} (d_0)^{1/2^{L+1}} = \left(\frac{\alpha}{rq^2} \right)^{2(1-1/2^{L+1})} m^{1/2^{L+1}}$$

from which

$$q^2 \geq \frac{\alpha}{r} m^{\frac{1}{2^{L+2}}} \geq \frac{\alpha}{r} m^{\frac{1}{2^{r+2}}}$$

which yields the lower bound $q = \tilde{\Omega}(m^{1/2^{r+3}})$. □

⁴Recall that (†) is only valid for $d_\ell \geq C\sqrt{m}$.

7.2.2 Reduction from Template-Search to $(1, 3, 2)$ -testing

It remains to describe and analyze a reduction from `TEMPLATE-SEARCH` to (two-sided) testing of $(1, 3, 2)$ -freeness. We first describe the reduction, before analyzing it and establishing the required properties (i.e., that it preserves both the query complexity and the number of rounds of adaptivity).

As a first step, we note that we can without loss of generality assume any test for $(1, 3, 2)$ -freeness to be *order-based*, i.e. bases its decisions only on the relative order of the values of f on its queries (and not on the values themselves). This is possible by invoking a result of Fischer [Fis04], along with the fact that π -freeness is a strongly order-based property (in the terminology of [Fis04]).⁵

Construction and simulation Given an instance (S, T) of the `TEMPLATE-SEARCH` problem with $|T| = m$, we build two (random) instances $f_{\text{yes}}, f_{\text{no}}$ of $(1, 3, 2)$ -freeness testing of size $n := 5m$. To obtain them, we start by describing the deterministic construction of a related function $f: [n] \rightarrow \mathbb{R}$ that both $f_{\text{yes}}, f_{\text{no}}$ will be based on.

1. set $\delta_i := \frac{1}{4} \min(T_{i+1} - T_i, T_i - T_{i-1})$, for all $1 \leq i \leq m - 1$;
2. define f on the first $2m$ elements by $f(2i - 1) = f(2i) := T_i$ for all $i \in [m]$;
3. define f on the remaining $3m$ elements by $f(i + 2m) := S_i$ for all $i \in [3m]$.

In other terms, f corresponds to duplicating each element of T into two adjacent identical elements, and concatenating the $2m$ resulting sequence T' with S .

Now, we define f_{yes} and f_{no} based on this f , which correspond to specific element-wise perturbations of the above f :

1. set $f_{\text{yes}}(2i - 1) = f_{\text{no}}(2i - 1) := f(2i - 1) - \delta_i$ and $f_{\text{yes}}(2i) = f_{\text{no}}(2i) := f(2i) + \delta_i$ for all $i \in [m]$ (that is, we decrease the first copy of T_i by δ_i and increase the second by δ_i);
2. set $f_{\text{yes}}(i + 2m) := f(i + 2m) + 2\delta_i$ and $f_{\text{no}}(i + 2m) := f(i + 2m)$ for all $i \in [m]$.

It is easy to see that f_{yes} is always $(1, 3, 2)$ -free. On the other hand, our function f_{no} will be $\Omega(1)$ -far from $(1, 3, 2)$ -free: indeed, one can check that every f_{no} has exactly m $(1, 3, 2)$ -tuples (one for each element of T , including the two first adjacent (perturbed) copies and the counterpart in the last $3m$ elements). Thus, since “fixing” one such triple requires modifying one of its elements, we have that f_{no} is $\frac{1}{5}$ -far from $(1, 3, 2)$ -free.

Furthermore, it is straightforward to simulate query access to either f_{yes} or f_{no} , provided query access to (S, T) , while only blowing up the number of queries by a factor 4 (and preserving the number of rounds of adaptivity). Indeed, when an algorithm queries $f_{\text{yes}}(i)$ (f_{no} is similar), it suffices to

- query the corresponding three adjacent elements of T to compute the relevant δ_j ;
- if $i > 2m$, also query S_{i-2m}

which can be done in parallel for all queries in a given round.

Reduction and simulation Assume now we have a q -query two-sided (order-based) test \mathcal{T} for $(1, 3, 2)$ -freeness with r rounds of adaptivity, which succeeds with probability $5/6$. Given an instance

⁵We actually need to be relatively careful in applying the result of [Fis04], as (i) we also require that the equivalence between tests and order-based tests preserve the number of rounds of adaptivity, and (ii) [Fis04] deals with integer-valued functions, while in our case they take values in \mathbb{R} . However, it is not hard to see by inspection of the proof of [Fis04, Theorem 3.2] that the argument does preserve the number of rounds of adaptivity; as for (ii), Section 5 of Fischer’s paper mentions the extension of his results to real-valued functions.

(S, T) of `TEMPLATE-SEARCH`, we draw and simulate access to two function $f_{\text{yes}}, f_{\text{no}}$ as above, and run \mathcal{T} in parallel on both (which costs in total, per the above, at most $2 \cdot 4q$ queries to (S, T)) *on the same randomness* ω . By a union bound, both instances are correct with probability at least $2/3$ over the choice of ω , i.e. \mathcal{T} rejects f_{no} while accepting f_{yes} . Fix any ω such that this holds.

We now use the fact that \mathcal{T} is order-based. Since for every $i < j$ such that (i, j) is not of the form $(2\ell, \ell + \Delta + 2m)$ with $\ell \in [m]$ (that is, where i is the index of the second perturbed copy of an element T_ℓ , and j the index in f of the element of S corresponding to T_ℓ), the order relation is the same under f_{yes} and f_{no} , all the answers given to \mathcal{T} for such queries will be the same under both functions. (This is by construction of f_{yes} and f_{no} , and the choice of the δ_i 's.)

In addition, as the two instances of the test are run on the same coin tosses, in order for them to give different answers (respectively `accept` and `reject`) they must have received a different answer to the same query at some point. That is, there exist a query pair $(i, j) \in [n]^2$ queried by both instances, such that $f_{\text{yes}}(i) < f_{\text{yes}}(j)$ yet $f_{\text{no}}(i) > f_{\text{no}}(j)$. But, by the above discussion, this only happens for (i, j) of the form $(2\ell, \ell + \Delta + 2m)$ (or $(\ell + \Delta + 2m)$ if $i > j$), from which the offset Δ can be immediately computed.

Thus, the above simulation allows one to solve the `TEMPLATE-SEARCH` problem on any instance (S, T) with probability $2/3$, while preserving the number of rounds of adaptivity r , and with at most $6q$ queries to (S, T) . By [Theorem 7.3](#), this implies $6q = \Omega(m^{1/2^{r+3}})$, establishing [Theorem 3.14](#).

8 A permutation dependent non-adaptive hierarchy

This section contains the proof of [Theorem 3.12](#). Note that the corresponding lower bound follows from [Corollary 3.10](#), so we only need to prove the upper bound.

Following the notation of [\[NRRS17\]](#), for a set \mathcal{A} of disjoint π -copies in $f: [n] \rightarrow \mathbb{R}$ we define $T_i = T_i(\mathcal{A}) = \{t_i : (t_1, \dots, t_k) \in \mathcal{A}\}$ for any $1 \leq i \leq k$. We also define

$$\begin{aligned} T^*(\mathcal{A}) &= \{(t_1, \dots, t_k) : \forall i \ t_i \in T_i(\mathcal{A}), \ \forall i \neq j \ f(t_i) < f(t_j) \iff \pi(i) < \pi(j)\} \\ T_{i,j}^*(\mathcal{A}) &= \{(t_i, \dots, t_j) : (t_1, \dots, t_k) \in T^*(\mathcal{A})\} \end{aligned}$$

That is, \mathcal{T}^* is the set of all π -copies induced by copies from \mathcal{A} , where the entry t_i from a copy $(t_1, \dots, t_k) \in \mathcal{A}$ is only allowed to play the role of an i -th entry of a copy. $T_{i,j}^*$ is the projection of \mathcal{T}^* onto coordinates i to j .

The proof of [Theorem 3.12](#) uses the `DyadicSampler` (Algorithm 3.1 of [\[NRRS17\]](#)) to efficiently find a large set of monotone subsequences of a desired form, and combines it with uniform sampling of the entries of the input sequence, to show that a π -copy can be obtained with good probability.

Proof of [Theorem 3.12](#). Let $2 \leq \ell \leq k - 1$. (the case $\ell = 1$ corresponds to a monotone permutation, and is settled by Newman et al.) Take $\pi = (\pi_1, \dots, \pi_k)$ to be any permutation in which $\pi_\ell = 1$ and $\pi_{\ell+i} = \ell + i$ for any $1 \leq i \leq k - \ell$.

The lower bound follows from [Theorem 3.5](#) and it remains to obtain the corresponding upper bound. Let $f: [n] \rightarrow \mathbb{R}$ be ε -far from π -freeness, so f contains a set \mathcal{A} of $\varepsilon n/k$ pairwise-disjoint π -copies. Our algorithm for finding a π -copy in f with probability $2/3$ is described below. To simplify the presentation, we do not try to optimize the (polynomial) dependence of the number of queries in ε and $\log n$.

- We run the `DyadicSampler` sufficiently many times, where each run is independent of all other runs. Our goal here is to obtain a set $T' \subseteq T_{\ell,k}^*$ of $n^{1-1/\ell}$ monotone increasing subsequences of

f , that are ℓ -dominating in the following sense. A sequence $s = (s_1, \dots, s_{k-\ell+1}) \in T_{\ell,k}^*$ is ℓ -dominating if the unique $t(s) = (t_1, \dots, t_k) \in \mathcal{A}$ for which $s_1 = t_\ell$ also satisfies $f(t_{\ell+1}) \leq f(s_2)$. We also require that the subsequences of T' have disjoint first entries. That is, for any two tuples $s = (s_1, \dots, s_{k-\ell+1}), s' = (s'_1, \dots, s'_{k-\ell+1}) \in T'$, it holds that $s_1 \neq s'_1$. Later, we show that after (independently) running the dyadic sampler $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$ times (making $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$ queries in total, since each run makes k queries), a set T' of the required size is obtained with probability $9/10$, provided that the constants hidden in the $\tilde{\Theta}_\varepsilon$ term are large enough.

- Suppose now that a set T' with the desired size was obtained in the first step. The second step is to sample single entries of our input sequence f uniformly at random, where the probability for each entry to be sampled is $10n^{-1/\ell}$, independently of other samples. The crucial idea here is that the set of entries of a sequence $s \in T'$, and the first $\ell - 1$ entries of the ℓ -dominated π -copy $t = t(s) \in \mathcal{A}$, can be combined together to obtain a π -copy. For a single subsequence $s \in T'$, Let E_s be the event that all of the first $\ell - 1$ entries of $t(s)$ are sampled. Then $\Pr(E_s) = 10^{\ell-1}n^{-1+1/\ell}$, and E_s is independent of all other events $E_{s'}$ for $s' \in T'$. Thus, it is not hard to see that with probability at least $9/10$, there exists $s \in T'$ for which E_s holds.
- The number of entries sampled in the second step is bounded by $100n^{1-1/\ell}$ with probability at least $9/10$ (so if the number of proposed samples exceeds this, we may stop and return an arbitrary answer, similarly to what was done in [Section 5](#)). Thus, the probability that a π -copy is found using our test, which makes $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$ queries, is at least $7/10$.

It remains to show that a set T' of $n^{1-1/\ell}$ monotone increasing subsequences as above may indeed be produced with probability $9/10$ using $\tilde{\Theta}_\varepsilon(n^{1-1/\ell})$ runs of the dyadic sampling algorithm. The fact that the dyadic algorithm essentially generates ℓ -dominating monotone subsequences at no additional cost is a direct consequence of the proof of [Theorem 3.2](#) in [\[NRRS17\]](#) (as was first observed by Newman et al., see the beginning of [Section 5](#) in their paper).

It was shown in [\[NRRS17\]](#) that the success probability of the dyadic sampler (i.e., the probability to generate an ℓ -dominating monotone subsequence as defined above) is at least $1/\beta_k(n, \varepsilon)$, where β_k is polynomial in $\log n$ and ε .

First, we make $g_{k,\ell}(n, \varepsilon) = 20\alpha_k\beta_k(n, \varepsilon)n^{1-1/\ell} \log n$ mutually independent runs of the dyadic sampler (where α_k is determined in [Lemma 8.1](#)), and denote by S the set of output tuples of all successful runs. By Chernoff's bound, $\Pr(|S| \geq 10\alpha_k n^{1-1/\ell} \log n) > 99/100$ for large enough n . The next lemma can be used to show that T' is large enough with good probability, assuming that S is large enough.

Lemma 8.1. *Let $s = (s_1, \dots, s_r)$ be the output of a single run of the `DyadicSampler` of Newman et al. ([Algorithm 3.1](#) in [\[NRRS17\]](#)) with parameters I and r , where I is a sequence of length $m \geq r$. Then for any element $x \in I$, the probability that s_1 equals x is bounded by α_r/m , where α_r depends only on r .*

Proof. For $r = 1$ this is obvious, and $\alpha_1 = 1$ suffices. The rest of the proof is by induction. If the “split point” is chosen to be $\ell \in [r - 1]$ in step 2 of the sampler, and the “slice-width” is chosen to be $W = 2^w$ in step 3 of the algorithm, then the `DyadicSampler` recursively takes (in step 5) its first ℓ elements from an interval I_L of size at most $2\ell W < 2^{w+1}r$, chosen uniformly (in step 4) among a set of $\Theta(m/2^w)$ possible intervals, such that each interval has a shift of 2^w from its predecessor. Therefore, each element is contained in at most $2r$ such intervals. Note that an interval chosen to be I_L in `DyadicSampler` must not be the last one – otherwise I_R would be empty – so $|I_L| \geq 2^w$.

Combining the inductive assumption with all of the above considerations, the probability to choose a certain entry as s_1 is bounded by

$$\frac{1}{\log m} \sum_{w=0}^{\lfloor \log m-1 \rfloor} \sum_{\ell=1}^{r-1} \left(O \left(\frac{2^w r}{m} \right) \frac{\alpha_\ell}{|I_L|} \right) \leq \frac{r}{m \log m} \sum_{w=0}^{\lfloor \log m-1 \rfloor} \sum_{\ell=1}^{r-1} \alpha_\ell \leq \frac{1}{m} r \sum_{\ell=1}^{r-1} \alpha_\ell$$

So picking $\alpha_r = r \sum_{\ell=1}^{r-1} \alpha_\ell$ suffices for our purposes. \square

With [Lemma 8.1](#) in hand, the rest of the proof is quite straightforward. We may assume that n is large enough, so that $g_{k,l}(n, \varepsilon) < n/10\alpha_k$. For a specific $x \in [n]$ the probability that among the $g_{k,l}(n, \varepsilon)$ runs of the dyadic sampler, at least $\log n$ outputted a tuple in which x is the first element is bounded by

$$\binom{n/10\alpha_k}{\log n} \left(\frac{\alpha_k}{n} \right)^{\log n} \leq \left(\frac{n}{10\alpha_k} \right)^{\log n} \left(\frac{\alpha_k}{n} \right)^{\log n} \leq \frac{1}{10}$$

Thus, with probability at least $89/100$, both $|S| \geq 10\alpha_k n^{1-1/\ell} \log n$ and no entry from $[n]$ appears as the first entry of more than $\log n$ tuples from S . In this case, $|T'| \geq n^{1-1/\ell}$ as required. \square

9 Acknowledgments

The authors wish to thank Noga Alon for fruitful discussions and invaluable feedback, and the anonymous reviewers for helpful comments and especially for pointing out an inaccuracy in one of the definitions in an early version of this paper.

References

- [BGJ⁺12] Arnab Bhattacharyya, Elena Grigorescu, Kyomin Jung, Sofya Raskhodnikova, and David P. Woodruff. Transitive-closure spanners. *SIAM Journal on Computing*, 41(6):1380–1425, 2012. [1.1](#)
- [BY17] Arnab Bhattacharyya and Yuichi Yoshida. *Property Testing*. Forthcoming, 2017. [1.1](#)
- [Can15] Clément L. Canonne. A Survey on Distribution Testing: your Data is Big. But is it Blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, April 2015. [1.1](#)
- [CG17] Clément L. Canonne and Tom Gur. An adaptivity hierarchy theorem for property testing. In *Computational Complexity Conference (CCC)*, 2017. [1.3](#), [3.6](#), [2](#)
- [CGG⁺17] Clément L. Canonne, Elena Grigorescu, Siyao Guo, Akash Kumar, and Karl Wimmer. Testing k -monotonicity. In *ITCS*, 2017. [1.1](#)
- [CS13] Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing, STOC '13*, pages 419–428, New York, NY, USA, 2013. ACM. [1.1](#)
- [DGL⁺99] Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. In Dorit S. Hochbaum, Klaus Jansen, José D. P. Rolim, and Alistair Sinclair, editors, *RANDOM-APPROX*, pages 97–108, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg. [1.1](#)
- [DKW56] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 09 1956. [7.2.1](#)
- [DN04] H.A. David and H.N. Nagaraja. *Order Statistics*. Wiley Series in Probability and Statistics. Wiley, 2004. [7.1](#)
- [EKK⁺00] Funda Ergün, Sampath Kannan, S.Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. *Journal of Computer and System Sciences*, 60(3):717–751, 2000. [1.1](#)
- [ES35] Paul Erdős and George Szekeres. A combinatorial problem in geometry. *Compos. Math.*, 2:463–470, 1935. [4](#)
- [Fis04] Eldar Fischer. On the strength of comparisons in property testing. *Inform. and Comput.*, 189:107–116, 2004. [1.1](#), [3.3](#), [7.2.2](#), [5](#)
- [GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998. [1.1](#)
- [GKW17] Elena Grigorescu, Akash Kumar, and Karl Wimmer. K -Monotonicity is Not Testable on the Hypercube. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:88, 2017. [1.1](#), [4](#)

- [GM14] Sylvain Guillemot and Daniel Márx. Finding small patterns in permutations in linear time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 82–101, 2014. [1.2](#)
- [Gol10] Oded Goldreich, editor. *Property Testing - Current Research and Surveys [outgrow of a workshop at the Institute for Computer Science (ITCS) at Tsinghua University, January 2010]*, volume 6390 of *Lecture Notes in Computer Science*. Springer, 2010. [1.1](#)
- [Gol17] Oded Goldreich. *Introduction to Property Testing*. Forthcoming, 2017. [1.1](#)
- [JR13] Madhav Jha and Sofya Raskhodnikova. Testing and reconstruction of Lipschitz functions with applications to data privacy. *SIAM J. Comput.*, 42(2):700–731, 2013. [1.1](#)
- [Mas90] Pascal Massart. The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 07 1990. [7.2.1](#)
- [New10] Ilan Newman. Property testing of massively parametrized problems – A survey. In Oded Goldreich, editor, *Property Testing - Current Research and Surveys [outgrow of a workshop at the Institute for Computer Science (ITCS) at Tsinghua University, January 2010]*, volume 6390 of *Lecture Notes in Computer Science*, pages 142–157. Springer, 2010. [1](#)
- [NRRS17] Ilan Newman, Yuri Rabinovich, Deepak Rajendraprasad, and Christian Sohler. Testing for forbidden order patterns in an array. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1582–1597, 2017. [1.1](#), [1.3](#), [2](#), [2.3](#), [2.5](#), [3.5](#), [3.6](#), [4](#), [7](#), [7.1](#), [7.1](#), [7.1](#), [7.1](#), [7.1](#), [7.1](#), [7.1](#), [7.1](#), [7.2](#), [8](#), [8.1](#)
- [PRR03] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. On testing convexity and submodularity. *SIAM Journal on Computing*, 32(5):1158–1184, 2003. [1.1](#)
- [Ras14] Sofya Raskhodnikova. Testing if an array is sorted. In Ming-Yang Kao, editor, *Encyclopedia of Algorithms*, pages 1–5. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. [1.1](#)
- [Ron08] Dana Ron. Property testing: A learning theory perspective. *Foundations and Trends in Machine Learning*, 1(3):307–402, 2008. [1.1](#)
- [Ron09] Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5(2):73–205, 2009. [1.1](#)
- [RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996. [1.1](#)