Parameterized Convexity Testing*

Abhiruk Lahiri[†] Ilan Newman[‡] Nithin Varma[§]

Abstract

In this work, we develop new insights into the fundamental problem of convexity testing of real-valued functions over the domain $[n]$. Specifically, we present a nonadaptive algorithm that, given inputs $\varepsilon \in (0,1), s \in \mathbb{N}$, and oracle access to a function, ε-tests convexity in $O(\log(s)/\varepsilon)$, where s is an upper bound on the number of distinct discrete derivatives of the function. We also show that this bound is tight. Since $s \leq n$, our query complexity bound is at least as good as that of the optimal convexity tester (Ben Eliezer; ITCS 2019) with complexity $O(\frac{\log \varepsilon n}{\varepsilon})$; our bound is strictly better when $s = o(n)$. The main contribution of our work is to appropriately parameterize the complexity of convexity testing to circumvent the worst-case lower bound (Belovs et al.; SODA 2020) of $\Omega(\frac{\log(\varepsilon n)}{\varepsilon})$ expressed in terms of the input size and obtain a more efficient algorithm.

1 Introduction

A function $f : [n] \to \mathbb{R}$ is convex if $f(x) - f(x-1) \leq f(x+1) - f(x)$ for all $x \in \{2, 3, \dots, n-1\}$. Convexity of functions is a natural and interesting property. Given oracle access to a function f, an ε-tester for convexity has to decide with high constant probability, whether f is a convex function or whether every convex function evaluates differently from f on at least εn domain points, where $\varepsilon \in (0,1)$. Parnas, Ron, and Rubinfeld [11] gave an ε-tester for convexity that has query complexity $O(\frac{\log n}{\varepsilon})$. Blais, Raskhodnikova, and Yaroslavtsev [4] showed that this bound is tight for constant ε for nonadaptive algorithms[1]. An improved upper bound of $O(\frac{\log(\varepsilon n)}{\varepsilon})$ was shown by Ben-Eliezer [3] in a work on the more general question of testing local properties. Recently, Belovs, Blais and Bommireddi [2] complemented this result by showing a tight lower bound of $\Omega(\frac{\log(\varepsilon n)}{\varepsilon})$.

In this work, we further investigate and develop new insights into this well-studied problem, thereby asserting that there is more way to go towards a full understanding of testing convexity of functions $f : [n] \to \mathbb{R}$. We show that the number of distinct discrete derivatives s, as opposed to the input size n, is the right input parameter to express the complexity of convexity testing, where a discrete derivative is a value of the form $f(x+1) - f(x)$ for $x \in [n-1]$. Specifically, we design a nonadaptive convexity tester with query complexity $O(\frac{\log s}{\varepsilon})$, and complement it with a nearly matching lower bound of $\Omega(\frac{\log(\varepsilon s)}{\varepsilon})$. Our work is motivated by the work of Pallavoor, Raskhodnikova and Varma [10] who introduced the notion of parameterization in the setting of sublinear algorithms.

Our results bring out the fine-grained complexity of the problem of convexity testing. In particular, $s \leq n$ always and therefore, our tester is at least as efficient as the state of the art convexity testers. Furthermore, the parameterization that we introduce, enables us to circumvent the worst case lower bounds expressed in terms of the input size n and obtain more efficient algorithms when $s << n$.

1.1 Our Results We begin our investigation with the simple and highly restricted case of testing convexity of functions $f : [n] \to \mathbb{R}$ having at most two distinct discrete derivatives. We design an adaptive algorithm that exactly decides convexity by making 5 queries and a nonadaptive algorithm that ε-tests convexity by making $O(1/\varepsilon)$ queries. The highlight is that both these algorithms are deterministic.

*The work of the first author was supported by the Israel Science Foundation, grant number 592/17 and 822/18. The work of the second author was supported by the Israel Science Foundation, grant number 379/21.

†Ariel University, Israel, abhiruk@ariel.ac.il

‡University of Haifa, Israel, ilan@cs.haifa.ac.il

§Chennai Mathematical Institute, India, nithinvarma@cmi.ac.in

[1] The queries of a nonadaptive algorithm does not depend on the answers to the previous queries. The algorithm is adaptive otherwise.

THEOREM 1.1. *There exists a* ***deterministic*** *algorithm that, given oracle access to a function* $f : [n] \to \mathbb{R}$ *having at most* 2 *distinct discrete derivatives, exactly decides convexity by making at most* 5 *adaptive queries.*

Theorem 1.1 is significant because one can construct simple examples of two distributions, both over functions having at most 2 distinct discrete derivatives, one over convex functions and the other over non-convex functions, such that no nonadaptive deterministic algorithm making $o(n)$ queries can distinguish the functions. Therefore, the above result shows the power of adaptivity even in this restricted setting.

We also design a constant-query deterministic nonadaptive testing algorithm for convexity of functions $f : [n] \to \mathbb{R}$ having at most 2 distinct discrete derivatives.

THEOREM 1.2. *Let* $\varepsilon \in (0, 1)$. *There exists a* ***deterministic nonadaptive*** 1*-sided error* ε*-tester for convexity of functions* $f : [n] \to \mathbb{R}$ *having at most* 2 *distinct discrete derivatives with query complexity* $O(1/\varepsilon)$.

Next, we consider the general case of functions $f : [n] \to \mathbb{R}$ having at most s distinct discrete derivatives and design the following nonadaptive tester.

THEOREM 1.3. *Let* $\varepsilon \in (0, 1)$. *There exists a nonadaptive* 1*-sided error* ε*-tester with query complexity* $O(\frac{\log s}{\varepsilon})$ *for convexity of real-valued functions* $f : [n] \to \mathbb{R}$ *having at most* s *distinct discrete derivatives.*

We complement Theorem 1.3 with the following lower bound that is tight for constant $\varepsilon \in (0, 1)$. The bound holds even for adaptive testers, thereby showing that one cannot hope for a separation between adaptive and nonadaptive testers for this general setting.

THEOREM 1.4. *For every sufficiently large* $s \in \mathbb{N}$, *every* $\varepsilon \in [1/s, 1/9]$, *and for every sufficiently large* $n \geq s$, *every* ε*-tester for convexity of functions* $f : [n] \to \mathbb{R}$ *having at most* s *distinct derivatives has query complexity* $\Omega\left(\frac{\log(\varepsilon s)}{\varepsilon}\right)$.

1.2 Related Work The study of property testing was initiated by Rubinfeld and Sudan [12] and Goldreich, Goldwasser and Ron [6]. The first example where parameterization has helped in the design of efficient testers is the work of Jha and Raskhodnikova [7] on testing the Lipschitz property. A systematic study of parameterization in sublinear-time algorithms was initiated by Pallavoor, Raskhodnikova and Varma [10] and studied further by [1, 5, 13, 9].

In this work, we are concerned only with convexity of real-valued functions over a $1D$ domain. We would like to note that not much is known about testing convexity of functions over higher dimensional domains. One possible reason behind this could be the following: there is no single definition of discrete convexity for real-valued functions of multiple variables. For a good overview of this topic, we refer interested readers to the textbook by Murota on discrete convex analysis [8]. Ben Eliezer [3], in his work on local properties, studied the problem of testing convexity of functions of the form $f : [n]^2 \to \mathbb{R}$ and designed a nonadaptive tester with query complexity $O(n)$. Later, Belovs, Blais and Bommireddi [2] showed a nonadaptive query lower bound of $\Omega(\frac{n}{d})^{\frac{d}{2}}$ for testing convexity of real-valued functions over $[n]^d$. For functions of the form $f : [3] \times [n] \to \mathbb{R}$, they design an adaptive tester with query complexity $O(\log^2 n)$ and show that the complexity of nonadaptive testing is $O(\sqrt{n})$.

2 Preliminaries

For a natural number $n \in \mathbb{N}$, we denote by $[n]$, the set $\{1, 2, \ldots, n\}$. Let $B \subseteq [n]$. For $x \in B$, the $\mathsf{order}(x)$ is the number of points $y \in B$ such that $y \leq x$. Points $x, y \in B$ are *consecutive* if $|\mathsf{order}(x) - \mathsf{order}(y)| = 1$. A function $f : B \to \mathbb{R}$ is *convex* if and only if

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y} \tag{2.1}$$

for all $x, y, z \in B$ such that $x < y < z$. A set of points $V \subseteq B$ is said to *violate* convexity if $f|_V$ is not convex.

For a function $f : [n] \to \mathbb{R}$, and $i \in [n]$, we use the term 'discrete derivative' at i for $\Delta(f, i) = f(i+1) - f(i)$. We denote by $\Delta_f : [n-1] \to \mathbb{R}$, the derivative function $\Delta_f(i) = \Delta(f, i),\ i = 1, \ldots, n-1$. The cardinality of the range of Δ_f is referred to as the *number of distinct discrete derivatives* of f. A function $f : [n] \to \mathbb{R}$ is convex if and only if Δ_f is monotone non-decreasing.

FACT 2.1. *If $f : B \to \mathbb{R}$ is not convex, then there exists three consecutive points $x, y, z \in B$ that violate eq.* (2.1).

We note that although convexity of f is equivalent to monotonicity of Δ_f, it is not true that if f is ε-far from being convex then Δ_f is ν-far from monotonicity for some positive constant ν. E.g., consider f that is defined by $f(i) = i,\ i \in [k],\ f(k+1) = k$ and $f(j) = j-1,\ j = k+2, \ldots n$. Then f if nearly $\frac{1}{2}$-far from being convex for $k = n/2$, while Δ_f is almost the 1-constant function.

Let $\varepsilon \in (0,1)$. A function $f : [n] \to \mathbb{R}$ is ε-far from convex if every convex function evaluates differently from f on at least εn points. A *basic ε-tester* for convexity gets oracle access to a function f, a parameter ε, and is such that, it **accepts** if f is convex, and **rejects**, with probability at least ε, if f is ε-far from convex.

3 Deterministic Convexity Testers for Functions having at most 2 Distinct Discrete Derivatives

We start with the very simple case in which f has only 2 distinct derivatives (if f has only 1 distinct derivative then f is a degree 1 function and, in particular, convex) and prove Theorem 1.1 and Theorem 1.2.

Let the range of Δ_f be $\{r_1 < r_2\}$. We do not assume that the algorithm knows the values r_1, r_2 but we will refer to them in our proofs and reasoning below. As it turns out, in this case there is a deterministic adaptive algorithm that can precisely decide if f is convex, making only 5 queries. This is based on the fact that the class of convex functions having at most two distinct derivatives is very restricted.

OBSERVATION 1. *If f is convex and Δ_f takes at most two values $\{r_1 < r_2\}$ then f is of the following form: there is $j \in [n]$ such that $f(i) = a + r_1(i-1), i = 1, \ldots, j$ and $f(i) = a + r_1(j-1) + r_2(i-j),\ i = j+1, \ldots n$ for some $r_1 < r_2$ and some a. We denote such f as f_{a,r_1,j,r_2}.*

Observation 1 suggests Algorithm 3 as a test for convexity.

Algorithm 1

Require: oracle access to function $f : [n] \to \mathbb{R}$ having at most 2 distinct derivatives

1: Query $f(1), f(2), f(n-1), f(n)$.
2: Define $f_1(x) := f(1) + (x-1)(f(2) - f(1))$ and $f_2(x) := f(n) - (n-x)(f(n) - f(n-1))$.
3: Query $f(j)$ such that $f_1(j) = f_2(j)$.
4: **Reject** if the function restricted to $1, 2, j, n-1, n$ is not convex or $f(j) \neq f_1(j)$, and **accept** otherwise.

REMARK 1. *Algorithm 3 is a* deterministic, adaptive *algorithm; it can make the last query only after knowing the values of f at the first* 4 *points. Moreover, if f is a function having at most* 2 *distinct discrete derivatives, then there is always an integer point $j \in [n]$ such that $f_1(j) = f_2(j)$.*

LEMMA 3.1. *Algorithm 3 accepts every convex function having at most* 2 *distinct derivatives, and rejects every function having at most* 2 *distinct derivatives that is not convex.*

We note that the lemma asserts that Algorithm 3 decides convexity correctly on every function that has at most 2 distinct derivatives, regardless of the distance to convexity.

Proof. If f is convex, then the restriction of f to every subset of $[n]$ is also convex and Algorithm 3 accepts.

Suppose f is not convex. This immediately implies that f has two distinct discrete derivatives, which we denote by $r_1 < r_2$. Now, it is necessary that $r_1 = \Delta_f(1)$ and $r_2 = \Delta_f(n-1)$ for the restriction of f to $\{1, 2, n-1, n\}$ to be convex, for otherwise Algorithm 3 immediately rejects.

Assuming that the restriction of f to the set $S = \{1, 2, n-1, n\}$ is convex, Observation 1 implies that the only convex function with 2 distinct derivatives that is consistent with f on the points in S is the function f^* (from Observation 1), where the value of $j \in [n]$ is unique and is as determined by Algorithm 3. If $f(j) \neq f^*(j)$, then the restriction of f to $S \cup \{j\}$ is not convex and Algorithm 3 rejects. In the rest of the proof, we argue that if $f(j) = f^*(j)$, then f and f^* has to evaluate to the same value on every point in $[n]$ and that f is convex. Specifically, each one of the discrete derivatives of f upto the j must be r_1, for otherwise, $f(j)$ will be larger than $f^*(j)$. Moreover, each one of the discrete derivatives of f from j upto n must be r_2, for otherwise, $f(j)$ will be smaller than $f^*(j)$. That is, the functions are identical on every point in $[n]$. ∎

3.1 A Nonadaptive Deterministic Convexity Tester for Functions having at most 2 Distinct Discrete Derivatives

As remarked above, Algorithm 3 is deterministic and decides convexity exactly under the promise that f has at most 2 distinct derivatives. However, it is *adaptive*. What can be said about nonadaptive algorithms for the same problem? It is easy to see that for any deterministic algorithm that makes $q < n-1$ *nonadaptive* queries $Q \subset [n]$, there are two functions g, h, both having at most 2 distinct derivatives, for which $g|_Q = h|_Q$ but g is convex while h is not convex. Hence there is no deterministic nonadaptive algorithm that decides convexity exactly, while making at most $n-2$ queries. This line of reasoning immediately extends to a $\Omega(n)$ lower bound on randomized nonadaptive algorithms that exactly decide convexity.

Here we come back to the property testing scenario. We show that there is a *deterministic nonadaptive* tester, Algorithm 2, that accepts every convex function f and rejects every function f that is ε-far from convex, under the promise that f has at most 2 distinct derivatives. Algorithm 2 makes only $O(1/\varepsilon)$ nonadaptive deterministic queries.

Algorithm 2

Require: $\varepsilon \in (0, 1)$; oracle access to function $f : [n] \to \mathbb{R}$ having at most 2 discrete derivatives
1: $S \leftarrow \{x_i(= i\varepsilon n)|\ i = 1, \ldots, 1/\varepsilon\}$
2: Query $f(1), f(2), f(n-1), f(n)$ and $f(x_i), f(x_i + 1)$ for all $i = 1, \ldots, 1/\varepsilon$
3: Set $r_1 \leftarrow f(2) - f(1)$ and $r_2 \leftarrow f(n) - f(n-1)$
4: **Reject** if $r_1 > r_2$
5: Let $j \in [1/\varepsilon]$ be the largest integer such that $f(x_j) = f(1) + r_1(x_j - 1)$
6: **Reject** if for some $i \geq j+1$, $f(x_i) \neq f(n) - r_2(n - x_i)$ and **accept** otherwise

---

Claim 1. *Algorithm 2 accepts every convex function f and rejects every function f that is ε-far from convex, provided that f has at most 2 distinct derivatives. Further, Algorithm 2 is a deterministic nonadaptive tester making $O(1/\varepsilon)$ queries.*

Proof. The claim about the query complexity is clear. Further, by Observation 1, if f is convex having at most 2 distinct derivatives, then for some $j^* \in [n]$, the function f is of the form given in the observation. Let $x_j \leq j^* \leq x_{j+1}$. Then f is consistent with j in the acceptance criterion of Algorithm 2, and hence will be accepted.

Next, consider a function f that is accepted by Algorithm 2. That is, there exists $j \in [1/\varepsilon]$ such that $f(x_i) = f(1) + r_1(x_i - 1)$ for every $i \leq j$, and that $f(x_i) = f(n) - r_2(n - x_i)$ for every $i \geq j+1$, where $r_1 \leq r_2$ are the two distinct discrete derivatives of f. Since $f(x_j) = f(1) + r_1(x_j - 1)$, the function f is a linear function when restricted to the set $[x_j]$. Similarly, when restricted to the set $[n] \setminus [x_{j+1} - 1]$, the function f is linear with slope r_2. Further, it can be seen that f can be corrected to be convex by changing the values for $x_j + 1 \leq i \leq x_{j+1} - 1$ to be consistent with $f_{f(1), r_1, j^*, r_2}$. As this changes at most εn points, it implies that f is ε-close to convex. ◻

4 Convexity Tester for Functions having at most s Distinct Discrete Derivatives

In this section, we describe our convexity tester for the case that the function $f : [n] \to \mathbb{R}$ has at most s distinct discrete derivatives and prove Theorem 1.3. A basic tester is presented in Algorithm 3. For simplicity, we assume throughout this section, that s/ε is an integer that divides n.

The top level idea is the following: suppose that f is convex with at most s distinct discrete derivatives, and let B be a set of $\ell = 1 + \frac{2s}{\varepsilon}$ nearly equally spaced consecutive pairs of points in $[n]$ starting with 1, 2, namely, $B = \{1, 2\} \cup \{i \cdot \frac{\varepsilon n}{2s} - 1, i \cdot \frac{\varepsilon n}{2s} :\ i = 1, \ldots \ell - 1\}$. Let $x_i = i \cdot \frac{\varepsilon n}{2s}$ for $i \in [\ell - 1]$. By the assumption on f, the function $\Delta_f|_I$ is the constant function on at least $\ell - s$ of the intervals $I = [x_i, x_{i+1} - 1]$. Further, if $\Delta_f|_I$ is constant on $I = [x_i, x_{i+1} - 1]$, then obviously $f(j) = f(x_i) + (j - x_i) \cdot \Delta_f(x_i)$ for $j \in I$. Thus in order to check that f is convex, we first check that $f|_B$ is convex using the nonadaptive, 1-sided error basic ε-tester of Belovs et al. [2] by making $O(\log(\varepsilon|B|))$ queries. Afterwards, we test that f is close to being a linear function on most intervals I. To test "linearity" of $f|_I$ on most such I, it is enough to pick a random such interval and test the distance to the appropriate linear function, which will result in a large enough success probability. The details follow.

Algorithm 3 invokes Algorithm 4 as a subroutine, where Algorithm 4 is a basic tester for convexity of functions defined over subdomains of $[n]$. The following theorem can be proven by modifying the analysis of the convexity tester by Belovs et al. [2] in a fairly straightforward manner. We have included its proof in the Appendix.

THEOREM 4.1. (BELOVS ET AL. [2]) *Let $B \subseteq [n]$. There exists a basic ε-tester for convexity of functions of the form $f : B \to \mathbb{R}$ that works for all $\varepsilon \in (0,1)$ with query complexity $O(\log(\varepsilon|B|))$.*

Algorithm 3 Convexity Tester

Require: parameter $\varepsilon \in (0,1)$; oracle access to function $f : [n] \to \mathbb{R}$; upper bound s on the number of distinct discrete derivatives in f

1: Let $B = \{1, 2\} \cup \{i \cdot \frac{\varepsilon n}{2s} - 1, i \cdot \frac{\varepsilon n}{2s} : i = 1, \ldots \frac{2s}{\varepsilon}\}$.
2: Test convexity of $f|_B$ with parameter $\varepsilon/32$ using Algorithm 4 and **reject** if that execution rejects.
3: Sample a point $x \in_R [n]$ u.a.r.
4: Let $y \leftarrow \lfloor \frac{2sx}{\varepsilon n} \rfloor$.
5: **Reject** if the points $y-1, y, x, y-1+\frac{\varepsilon n}{2s}, y+\frac{\varepsilon n}{2s}$ violate convexity.

Lemma 4.1 shows that Algorithm 3 is indeed a basic ε-tester for convexity. Our convexity tester with query complexity $O(\log(s)/\varepsilon)$ is obtained by $O(1/\varepsilon)$ repetitions of Algorithm 3. This completes the proof of Theorem 1.3.

LEMMA 4.1. *Algorithm 3, rejects with probability at least $\varepsilon/32$, every function f having at most s distinct discrete derivatives that is ε-far from convex.*

Proof. It is enough to argue that Steps 2-5 of Algorithm 3 rejects with probability at least $\varepsilon/32$.

If $f|_B$ is $\varepsilon/32$-far from being convex, by Theorem 4.1, one iteration of Algorithm 4 rejects with probability at least $\varepsilon/32$.

In the rest of the proof, we assume that $f|_B$ is $\varepsilon/32$-close to convex. In other words, it is possible to modify $f|_B$ in at most $\frac{\varepsilon|B|}{32}$ points in B in order to make $f|_B$ convex.

Let b_k for $k \in [2s/\varepsilon]$ be shorthand for the index $k \cdot \frac{\varepsilon n}{2s} - 1$, and let b_0 stand for the index 1. Let I_k denote the interval of indices $\{b_k + 1, \ldots, b_{k+1}\}$ for $k \in [(2s/\varepsilon) - 1]$. Let I_0 denote the interval of indices $\{2, \ldots, b_1\}$. For $k \in [(2s/\varepsilon) - 1]$, the interval I_k is *nearly linear* if

$$f(b_k + 1) - f(b_k) = \frac{f(b_{k+1}) - f(b_k + 1)}{b_{k+1} - (b_k + 1)} = f(b_{k+1} + 1) - f(b_{k+1}). \tag{4.2}$$

The interval I_0 is nearly linear if

$$f(2) - f(1) = \frac{f(b_1) - f(1)}{b_1 - 1} = f(b_1 + 1) - f(b_1). \tag{4.3}$$

We first prove a lower bound on the number of nearly linear intervals. Recall that there is a way to modify $f|_B$ by changing its values on at most $\varepsilon \cdot |B|/32$ *bad* points. For $k \in [(2s/\varepsilon) - 1]$, the interval I_k is bad if one among $b_k, b_k + 1, b_{k+1}, b_{k+1} + 1$ is a bad point, and is good otherwise. Likewise, I_0 is bad if one among $1, 2, b_1, b_1 + 1$ is a bad point, and is good otherwise. The number of bad intervals is, therefore, at most $\varepsilon|B|/16$, since a bad point can make at most two intervals bad. Now, $\varepsilon|B|/16$ is at most $15s/16$, by substituting the value of $|B|$.

For $k \in [(2s/\varepsilon) - 1]$, if the interval I_k is good, then none of the points in $\{b_k, b_k + 1, b_{k+1}, b_{k+1} + 1\}$ are bad, and hence we have

$$f(b_k + 1) - f(b_k) \le \frac{f(b_{k+1}) - f(b_k + 1)}{b_{k+1} - (b_k + 1)} \le f(b_{k+1} + 1) - f(b_{k+1}). \tag{4.4}$$

Similarly, if I_0 is good, then

$$f(2) - f(1) \le \frac{f(b_1) - f(1)}{b_1 - 1} \le f(b_1 + 1) - f(b_1). \tag{4.5}$$

For a good interval I_k (or I_0) for $k \in [(2s/\varepsilon) - 1]$ that is not nearly linear, one of the inequalities in Equation 4.4 (Equation 4.5, respectively) must be a strict inequality. Since the number of distinct discrete derivatives in f is at most s, the number of distinct discrete derivatives among restricted to the good points is also at most s. Since the function f restricted to the good points is convex, the number of good intervals with strict inequalities (in

Algorithm 4 Basic Tester for Convexity over Subdomains of $[n]$

Require: oracle access to a function $g : B \to \mathbb{R}$; parameter ε

```
loop ⌈24 log(2ε|B|)⌉ times:
    Draw a point a ∈ B uniformly at random.
    Let a′ ∈ B be such that order(a′) = order(a) + 1.
    Pick a number k uniformly at random from {0, 1, ..., ⌈log 2ε|B|⌉}.
    Let h ∈ B be a point such that order(h) is a multiple of 2^k, where, with probability 1/2, the point h is the
  smallest such point larger than a, and with probability 1/2, it is the largest such point smaller than a.
    Let h′ ∈ B be such that order(h′) = order(h) + 1.
    Reject if the set {a, a′, h, h′} violates convexity.
end loop
```

Equation 4.4 or Equation 4.5) is at most $s-1$. Hence, the number of good intervals that are not nearly linear is at most $s-1$.

Since f is ε-far from being convex and each interval has at most $\frac{\varepsilon n}{2s}$ indices, the restriction of f to the set of indices belonging to good nearly linear intervals, has distance at least $\varepsilon n - \frac{15s}{16} \cdot \frac{\varepsilon n}{2s} - s \cdot \frac{\varepsilon n}{2s} = \varepsilon n - \frac{\varepsilon n}{2} - \frac{15\varepsilon n}{32}$ from convexity.

For $k \in [(2s/\varepsilon) - 1]$, for a good nearly linear interval I_k, we use D_k to denote the Hamming distance of $f|_{I_k}$ to the linear function $g_k : I_k \to \mathbb{R}$ defined as $g_k(x) = f(b_k + 1) + (x - b_k - 1) \cdot t$ for $x \in I_k$, where $t = f(b_{k+1} + 1) - f(b_{k+1})$. Similarly, if I_0 is good, we use D_0 to denote the Hamming distance of $f|_{I_0}$ to the linear function $g_0 : I_0 \to \mathbb{R}$ defined as $g_0(x) = f(1) + (x-1) \cdot t$, where $t = f(2) - f(1)$.

Consider the restriction of f to the set of indices that belong to the good nearly linear intervals. We can make this restriction convex by replacing $f|_{I_k}$ with g_k for each k such that I_k is a good nearly linear interval. Hence,

$$\sum_{\substack{k: I_k \\ \text{nearly linear} \\ \text{and good}}} D_k \geq \varepsilon n - \frac{\varepsilon n}{2} - \frac{15\varepsilon n}{32} \geq \frac{\varepsilon n}{32}.$$

The proof will be completed by arguing that there are at least D_k points x in a good nearly linear interval I_k such that Algorithm 3 rejects by sampling x in Step 5.

Consider a good nearly linear interval I_k such that $f(b_k + 1) - f(b_k) = t$. Consider the (favorable) set F consisting of all points $z \in I_k$ such that $f(z) - f(b_k + 1) = t \cdot (z - b_k - 1)$ and $f(b_{k+1}) - f(z) = t \cdot (b_{k+1} - z)$. Clearly, both $b_k + 1$ and b_{k+1} are in F. If Algorithm 3 samples, in Step 5, a point $z \notin F$, it rejects. We now show that we can repair the function values at points not in F and make $f|_{I_k}$ be equal to g_k. Consider an interval of points $\{x, x+1, \ldots, y\}$ such that none of them are in F, and where, both $x-1$ and $y+1$ are in F. Since $x-1$ and $y+1$ are both in F, we have that $f(y+1) - f(x-1) = t \cdot (y - x + 2)$. We repair the function on the interval $\{x, x+1, \ldots, y\}$ by assigning the value $f(x-1) + (x' - x + 1) \cdot t$ for all $x' \in \{x, x+1, \ldots, y\}$. We can repair the function on the whole interval and make $f|_{I_k}$ be equal to g_k by applying the same modification on every such maximal subinterval, where the maximality is in the sense of not belonging to F.

Hence, the probability that the tester rejects in Step 5 is at least $\varepsilon/32$. This completes the proof. □

5 Lower Bound

In this section, we prove Theorem 1.4.

LEMMA 5.1. *For every sufficiently large $s \in \mathbb{N}$, every $\varepsilon \in [1/s, 1/9]$, and for every sufficiently large $n \geq s$, every ε-tester for convexity of functions $f : [n] \to \mathbb{R}$ having at most s distinct derivatives has query complexity $\Omega\left(\frac{\log(\varepsilon s)}{\varepsilon}\right)$.*

Proof. We use Yao's principle. Let $s \in \mathbb{N}$ and $\varepsilon \in [1/s, 1/9]$. Consider the distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ from Belovs et al. [2] (proof of Theorem 1.3) of functions $f : [s] \to \mathbb{R}$. Every function sampled from these distributions have at most s distinct derivatives. Moreover, every function sampled from $\mathcal{D}_0$ is convex, and every function sampled from $\mathcal{D}_1$ is ε-far from convex. They show that every tester distinguishing these distributions, with probability at least 2/3, has to make at least $\Omega\left(\frac{\log(\varepsilon s)}{\varepsilon}\right)$ queries.

Consider an integer $n \geq s$ that is an integer multiple of s. Let k denote n/s. We define distributions $\mathcal{D}'_0$ and $\mathcal{D}'_1$ of functions $f : [n] \to \mathbb{R}$ as follows.

For $b \in \{0, 1\}$, to sample a function f from $\mathcal{D}'_b$, first sample a function g from $\mathcal{D}_b$. For $i \in [s]$, let $f((i-1)k+1) = g(i)$. For $i \in [s-1]$, let $\mathsf{slope}_i = \frac{g(i+1)-g(i)}{k}$. Now, for all $j \in [k-1]$ and for all $i \in [s-1]$, set $f((i-1)k+1+j) = f((i-1)k+1) + j \cdot \mathsf{slope}_i$.

By construction, every function sampled from $\mathcal{D}'_0$ is convex. Additionally, every function sampled from $\mathcal{D}'_1$ is ε-far from convex. To see this, consider a function f sampled from $\mathcal{D}'_1$ that is ε-close to being convex. Let g denote the function sampled from $\mathcal{D}_1$ from which we constructed f. Let $B \subseteq [n]$ denote the set of *bad* points such that changing the values of f on points in B makes it convex. Since f is piecewise linear, it is clear that for each point in B of the form $(i-1)k+1$ for $i \in [s]$, either the set of $k-1$ points $\{(i-1)k+1+j : j \in [k-1]\}$, or the set of $k-1$ points $\{(i-2)k+1+j : j \in [k-1]\}$ has to belong to B. Thus, the number of points in B of the form $(i-1)k+1$ for $i \in [s]$ is at most $|B|/k = \varepsilon s$. By construction of f, we know that for all $i \in [s]$, it holds that $f((i-1)k+1) = g(i)$. Thus, the distance of g to convexity is at most εs.

Consider a deterministic algorithm A that distinguishes these distributions by making $o\left(\frac{\log(\varepsilon s)}{\varepsilon}\right)$ queries. One can use A to distinguish, with the same success probability, the distributions $\mathcal{D}_0$ and $\mathcal{D}_1$ by making at most twice the number of queries as A, which leads to a contradiction. $\square$

References

[1] Aleksandrs Belovs. Adaptive lower bound for testing monotonicity on the line. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, volume 116 of *LIPIcs*, pages 31:1–31:10, 2018.

[2] Aleksandrs Belovs, Eric Blais, and Abhinav Bommireddi. Testing convexity of functions over finite domains. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 2030–2045, 2020.

[3] Omri Ben-Eliezer. Testing local properties of arrays. In *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 11:1–11:20, 2019.

[4] Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 309–320. IEEE Computer Society, 2014.

[5] Deeparnab Chakrabarty and C. Seshadhri. Adaptive boolean monotonicity testing in total influence time. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 20:1–20:7, 2019.

[6] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[7] Madhav Jha and Sofya Raskhodnikova. Testing and reconstruction of lipschitz functions with applications to data privacy. *SIAM J. Comput.*, 42(2):700–731, 2013.

[8] Kazuo Murota. *Discrete convex analysis*, volume 10 of *SIAM monographs on discrete mathematics and applications*. SIAM, 2003.

[9] Ilan Newman and Nithin Varma. New sublinear algorithms and lower bounds for LIS estimation. In *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 100:1–100:20, 2021.

[10] Ramesh Krishnan S. Pallavoor, Sofya Raskhodnikova, and Nithin Varma. Parameterized property testing of functions. *ACM Trans. Comput. Theory*, 9(4):17:1–17:19, 2018.

[11] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *J. Comput. Syst. Sci.*, 72(6):1012–1042, 2006.

[12] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996.

[13] Ramesh Krishnan Pallavoor Suresh. *Improved algorithms and new models in property testing*. PhD thesis, Boston University, 2020.

A Basic Tester for Convexity over Subdomains of $[n]$

In this section, we prove Theorem 4.1. All definitions and lemmas in this section are straightforward generalizations of those by Belovs et al. [2].

DEFINITION 1. (TEST-SET, AND ITS ROOT, HUB, AND SCALE) *Consider a point $a \in B$ and an integer k. Let $h \in B$ be one among the two points closest to a such that $\mathsf{order}(h)$ is a multiple of 2^k. Let $a', h' \in B$ be such that $\mathsf{order}(a') = \mathsf{order}(a) + 1$, and $\mathsf{order}(h') = \mathsf{order}(h) + 1$. We refer to $\{a, a', h, h'\}$ as the* test-set *with* root a, hub h, *and* scale 2^k.

LEMMA A.1. *Consider points $x, y \in B$ such that $\mathsf{order}(x) < \mathsf{order}(y) - 1$. Then there exists test-sets with roots x and y respectively, having a common hub and scales at most $2(\mathsf{order}(y) - \mathsf{order}(x))$.*

Proof. Consider the smallest integer k such that there exists a unique multiple m of 2^k satisfying $\mathsf{order}(x) \leq m \leq \mathsf{order}(y)$. There are at least 2 multiples of 2^{k-1} in the range $[\mathsf{order}(x), \mathsf{order}(y)]$. If there were only one multiple of 2^{k-1} in that range, then it contradicts our assumption that k is the smallest integer such that there is a unique multiple of 2^k in $[\mathsf{order}(x), \mathsf{order}(y)]$. Now, since there are at least 2 multiples of 2^{k-1} in the range $[\mathsf{order}(x), \mathsf{order}(y)]$, we have that $2^{k-1} \leq (\mathsf{order}(y) - \mathsf{order}(x))$.

The lemma follows by setting $h \in B$ such that $\mathsf{order}(h) = m$ as the common hub of the test-sets with roots x and y, and 2^k as their scale. ∎

LEMMA A.2. *Consider a set $\{x, y, z\} \subseteq B$ that violates convexity such that $\mathsf{order}(x) < \mathsf{order}(y) < \mathsf{order}(z)$. At least one of the test-sets involving x, y, or z violate convexity. Moreover, scale is not exceeding $2 \cdot \max\{\mathsf{order}(z) - \mathsf{order}(y), \mathsf{order}(y) - \mathsf{order}(x)\}$.*

Proof. Let $x', y', z' \in B$ be such that $\mathsf{order}(x') = \mathsf{order}(x)+1$, $\mathsf{order}(y') = \mathsf{order}(y)+1$, and $\mathsf{order}(z') = \mathsf{order}(z)+1$.

Let h_1 be the common hub for test-sets involving x and y, as guaranteed by Lemma A.1 and let $h_1' \in B$ be such that $\mathsf{order}(h_1') = \mathsf{order}(h_1) + 1$. By Lemma A.1, the test-sets $\{x, x', h_1, h_1'\}$, and $\{h_1, h_1', y, y'\}$ both have scale at most $2(\mathsf{order}(y) - \mathsf{order}(x))$.

Similarly, let $h_2 \in B$ be the common hub for test-sets of y and z with scale at most $2(\mathsf{order}(z) - \mathsf{order}(y))$ and let $h_2' \in B$ be such that $\mathsf{order}(h_2') = \mathsf{order}(h_2) + 1$. Note that $\{y, y', h_2, h_2'\}$, and $\{h_2, h_2', z, z'\}$ are the test-sets being alluded to in this case.

If none of the aforementioned four test-sets violate convexity, then, we immediately have the following inequalities:

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(y) - f(h_1')}{y - h_1'} \leq f(y') - f(y) \leq \frac{f(z) - f(y)}{z - y}.$$

This contradicts our assumption that $\{x, y, z\}$ violates convexity. ∎

We are now ready to prove Theorem 4.1. The query complexity of Algorithm 4 is clear from its description. Additionally, it always accepts convex functions. Consider a function $f : B \to \mathbb{R}$ that is ε-far from convex. The total number of possible scales for test-sets (see Algorithm 4) is at most $1 + \lceil \log_2(2\varepsilon|B|) \rceil$. Hence, the total number of test-sets is at most $2|B| \cdot (1 + \lceil \log_2(2\varepsilon|B|) \rceil)$.

We will construct a set $C \subseteq B$ such that $|C| \geq \varepsilon \cdot |B|$ and for every $a \in C$, one of the test-sets rooted at a of scale at most $1 + \lceil \log_2(2\varepsilon|B|) \rceil$ violate convexity. Initialize C to be $\emptyset$. If $|C| < \varepsilon \cdot |B|$, then $f|_{B \setminus C}$ is not convex. We can find consecutive points $x < y < z \in B \setminus C$, such that f violates convexity on these three points. Since C contains fewer than $\varepsilon|B|$ many points, we have that $\max\{\mathsf{order}(y) - \mathsf{order}(x), \mathsf{order}(z) - \mathsf{order}(y)\}$ is at most $\varepsilon|B|$. Hence, by Lemma A.2, we know that there exists a violating test-set involving either x, y, or z of scale at most $2\varepsilon|B|$. We add such a point to set C. Hence, we have at least εn violating test-sets. Therefore, the tester rejects with probability at least $\frac{2\varepsilon}{\log(2\varepsilon n)}$ in a single iteration.