PARAMETERIZED CONVEXITY TESTING

Abhiruk Lahiri

CHARLES UNIVERSITY

Ilan Newman

אוניברסיטת חיפה
University of Haifa
جامعة حيفا

Nithin Varma

CHENNAI MATHEMATICAL INSTITUTE

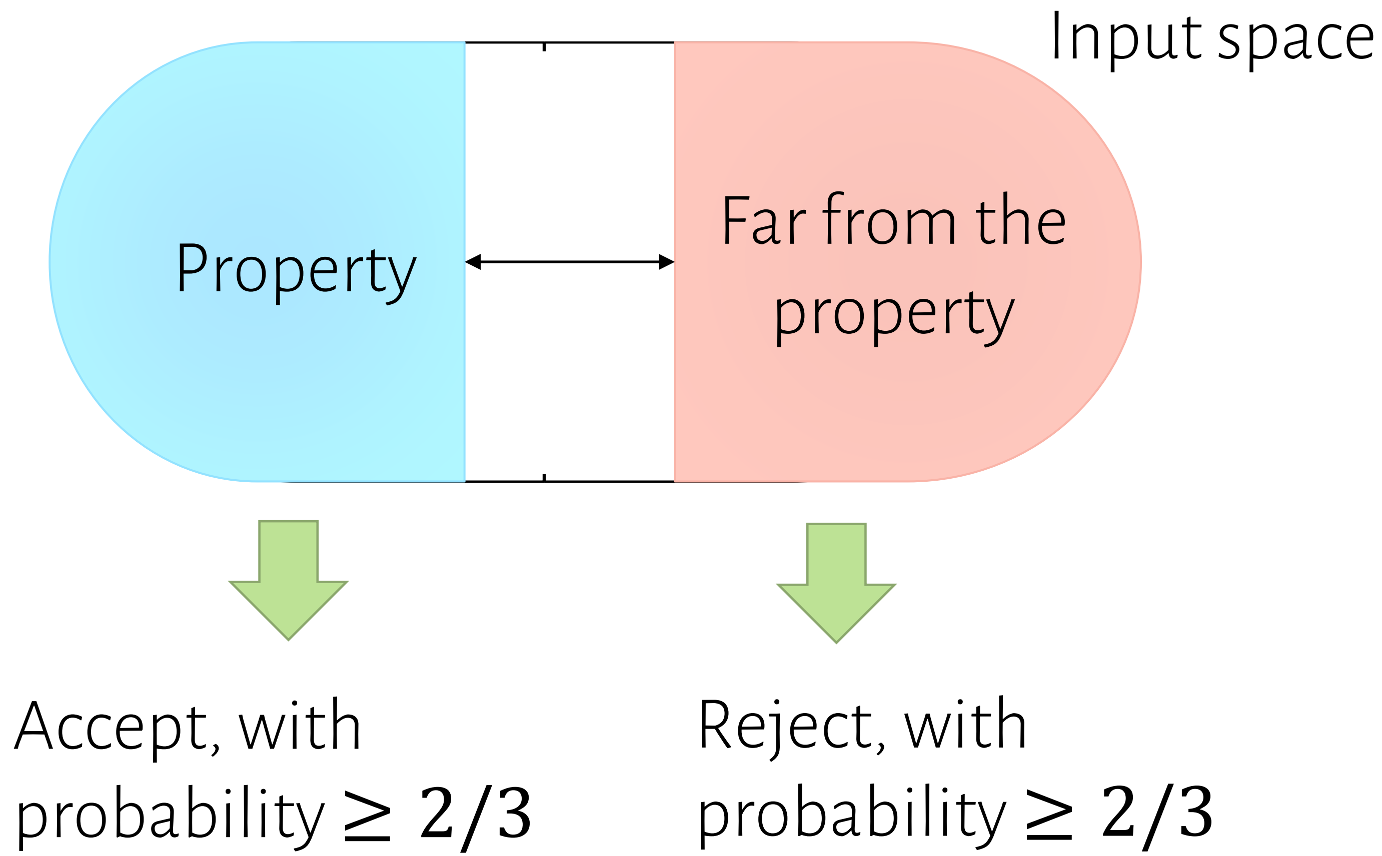
Broad Goal of our Work

"Identify parameters that can express the complexity of sublinear-time algorithms better!"

- Line of work initiated by [Pallavoor Raskhodnikova Varma '19]
- Complexity expressed in terms of input length n might be pessimistic
- Want to capture the fine-grained complexity of a problem
- Identifying the right parameters might enable circumventing worst-case lower bounds in terms of the input size

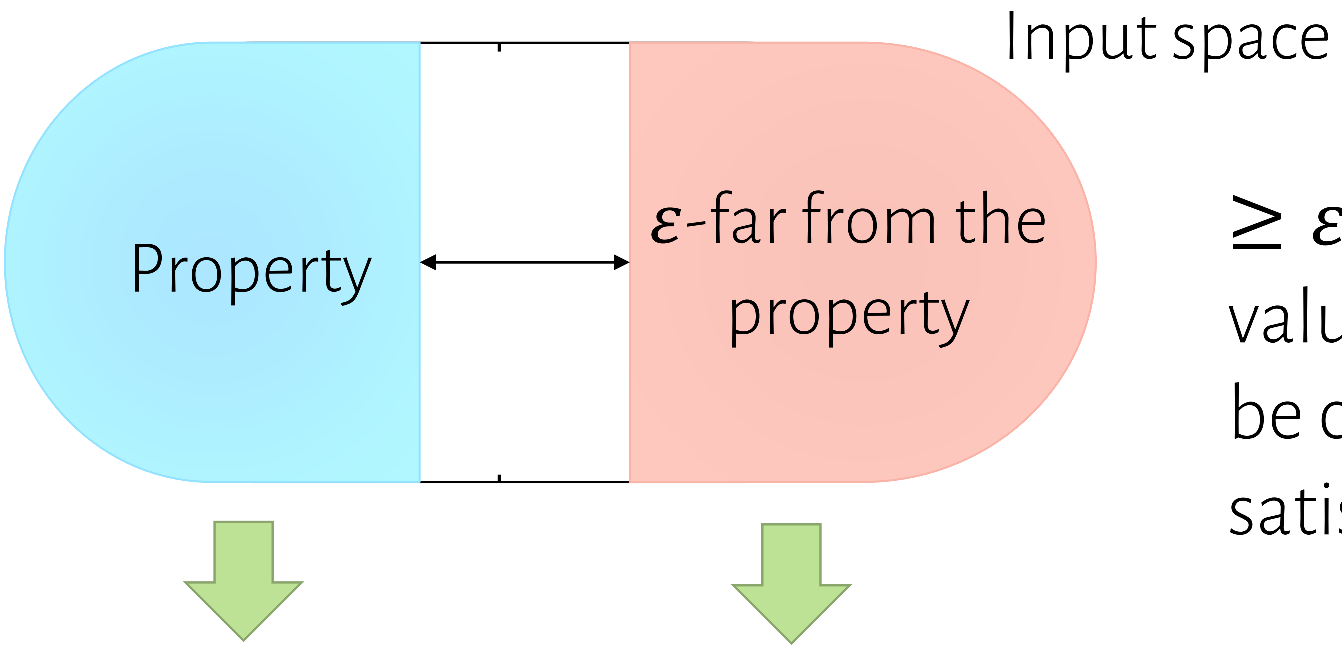
Computational Problem

- Property Testing [Rubinfeld Sudan '96, Goldreich Goldwasser Ron '98]



Property Testing

- Example properties: Sorted arrays, convex functions, bipartite graphs….



$\geq \varepsilon$ fraction of values of input to be changed to satisfy property

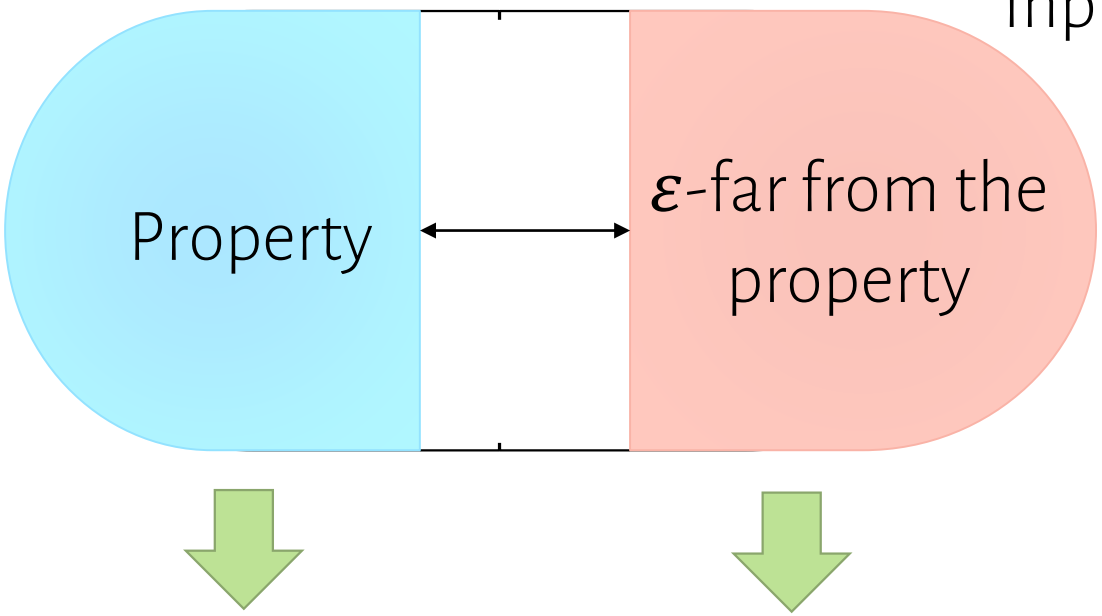
Accept, with probability $\geq 2/3$

Reject, with probability $\geq 2/3$

Property Testing

- Example properties: Sorted arrays, convex functions, bipartite graphs….

Input space



Accept, with probability $\geq 2/3$

Reject, with probability $\geq 2/3$

$\geq \varepsilon$ fraction of values of input to be changed to satisfy property

Algorithms have query access to inputs represented as functions

Parameterized Sublinear-time Algorithms

- For testing monotonicity of real-valued functions
  - Cardinality of the image of the function [Pallavoor Raskhodnikova Varma '19], [Black Kalemaj Raskhodnikova '21]
- For testing monotonicity of Boolean functions over $\{0,1\}^d$
  - Influence of the function [Chakrabarty Seshadhri '19]
- For testing Lipschitz property of functions $f: [n] \to \mathbb{R}$
  - Image diameter of function [Jha Raskhodnikova '13]
- Estimating the length of the longest increasing subsequence in an array
  - Number of distinct values in the array [Newman Varma '21]

We Study: Testing Convexity

- Function $f : [n] \to \mathbb{R}$ is convex:
  - if for all $i \in \{2, \ldots, n-1\}$

$$f(i) - f(i-1) \leq f(i+1) - f(i)$$

    i.e., adjacent discrete derivatives are monotone

  - If for $x, y, z \in [n]$,

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

Given query access to $f : [n] \to \mathbb{R}$ and input $\epsilon \in (0,1)$, decide with probability at least $2/3$, if f is convex or ϵ-far from convex

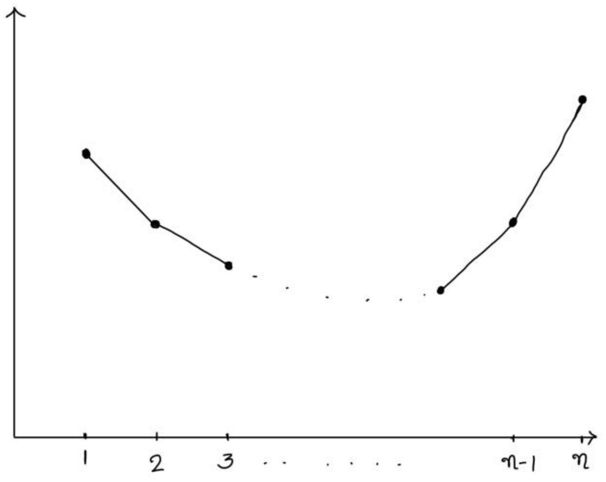
Results on Convexity Testing

- $O\left(\frac{\log n}{\epsilon}\right)$-query nonadaptive tester [Parnas Ron Rubinfeld '06]
- $\Omega(\log n)$-query lower bound for nonadaptive algorithms [Blais Rakshodnikova Yaroslavtsev '14]
- $O\left(\frac{\log \epsilon n}{\epsilon}\right)$-query nonadaptive tester [Ben-Eliezer '19, Belovs Blais Bommireddi '20]
- $\Omega\left(\frac{\log \epsilon n}{\epsilon}\right)$-query lower bound [Belovs Blais Bommireddi '20]

Optimal complexity of convexity testing is $\Theta(\log n)$!

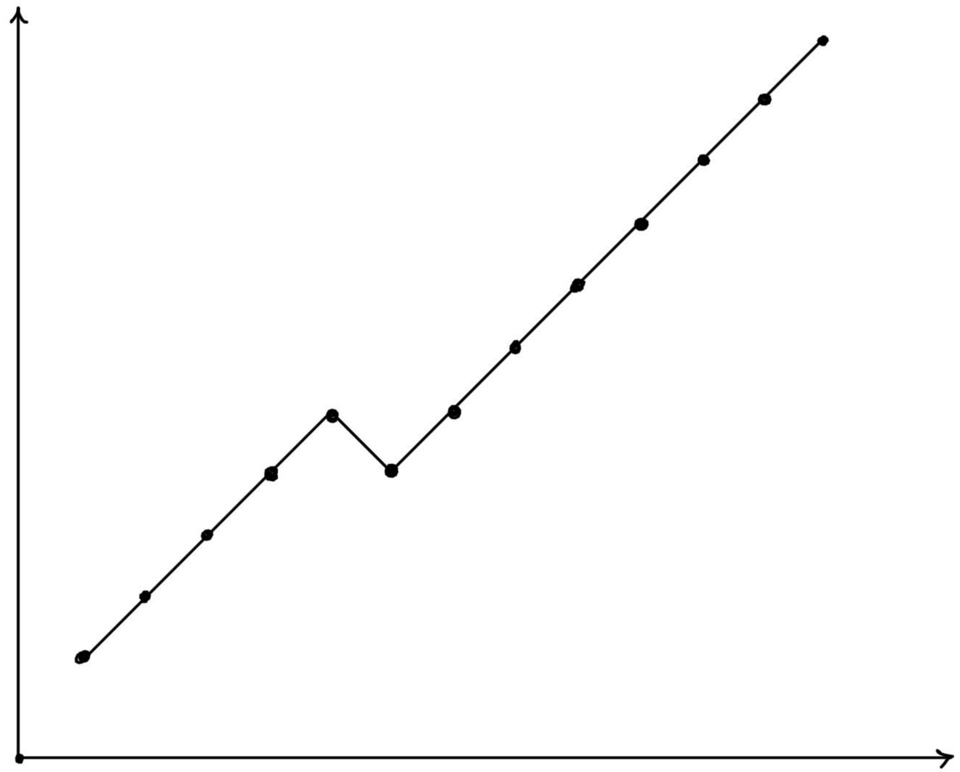
Convex Functions Over $[n]$

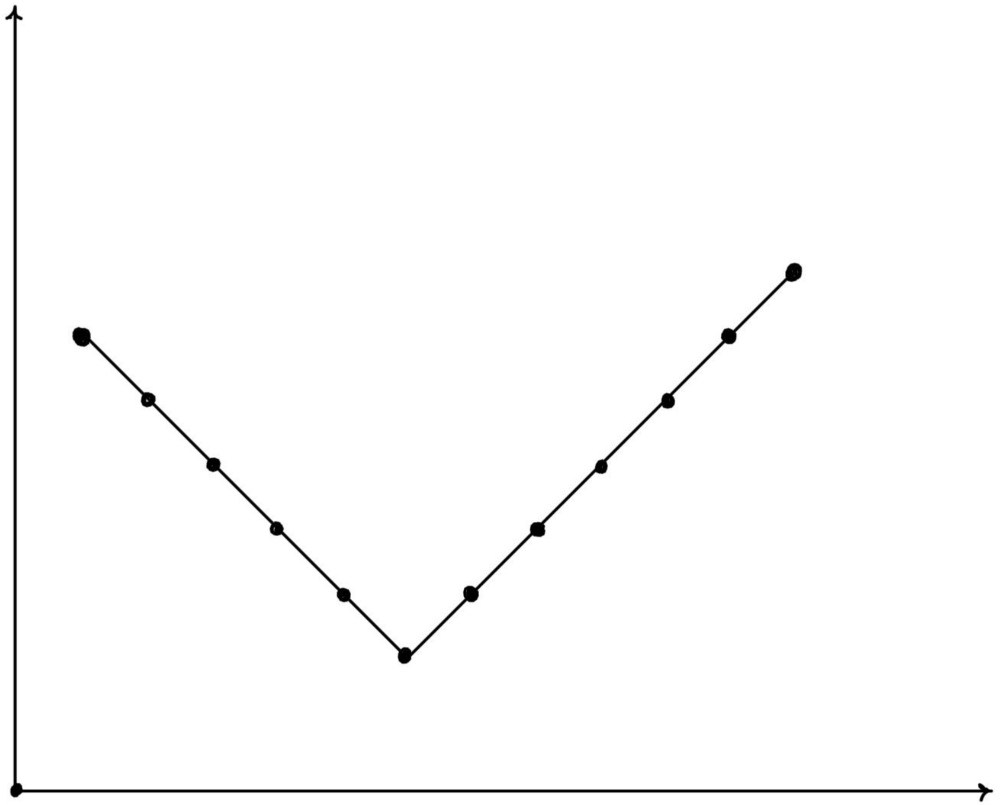
- Piecewise linear function
  - Number of pieces is the number of distinct adjacent discrete derivatives



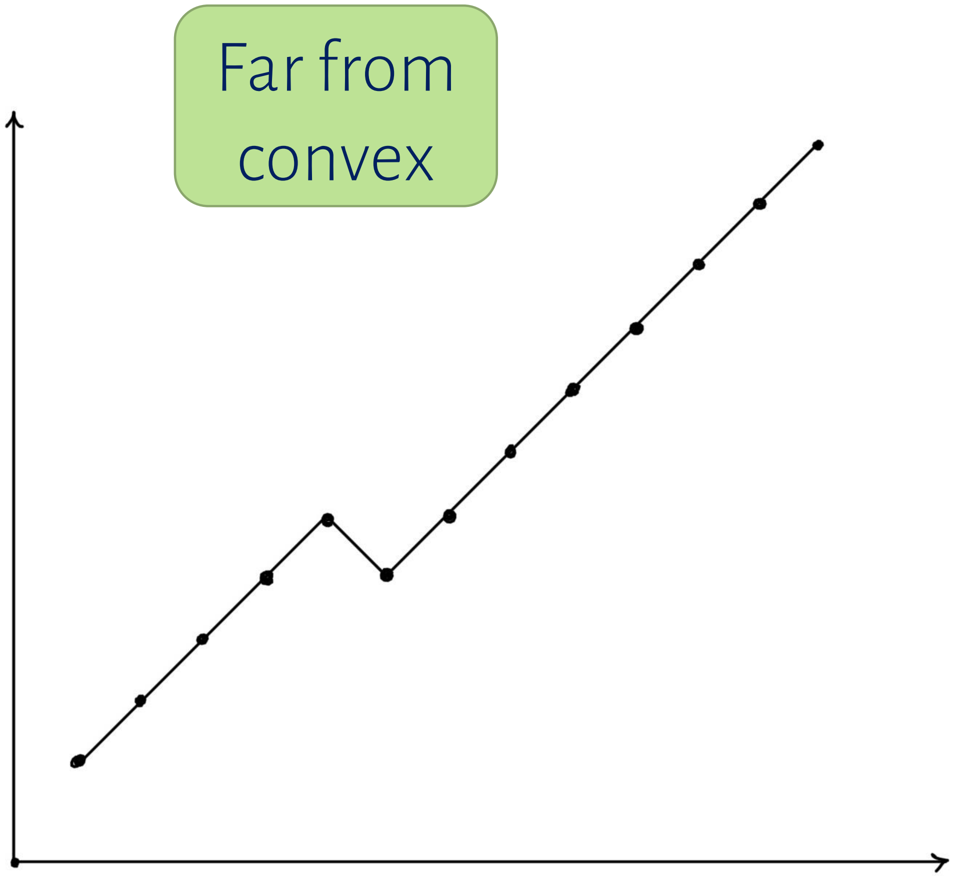
What is the complexity of testing convexity if the number of distinct derivatives in input function is small?

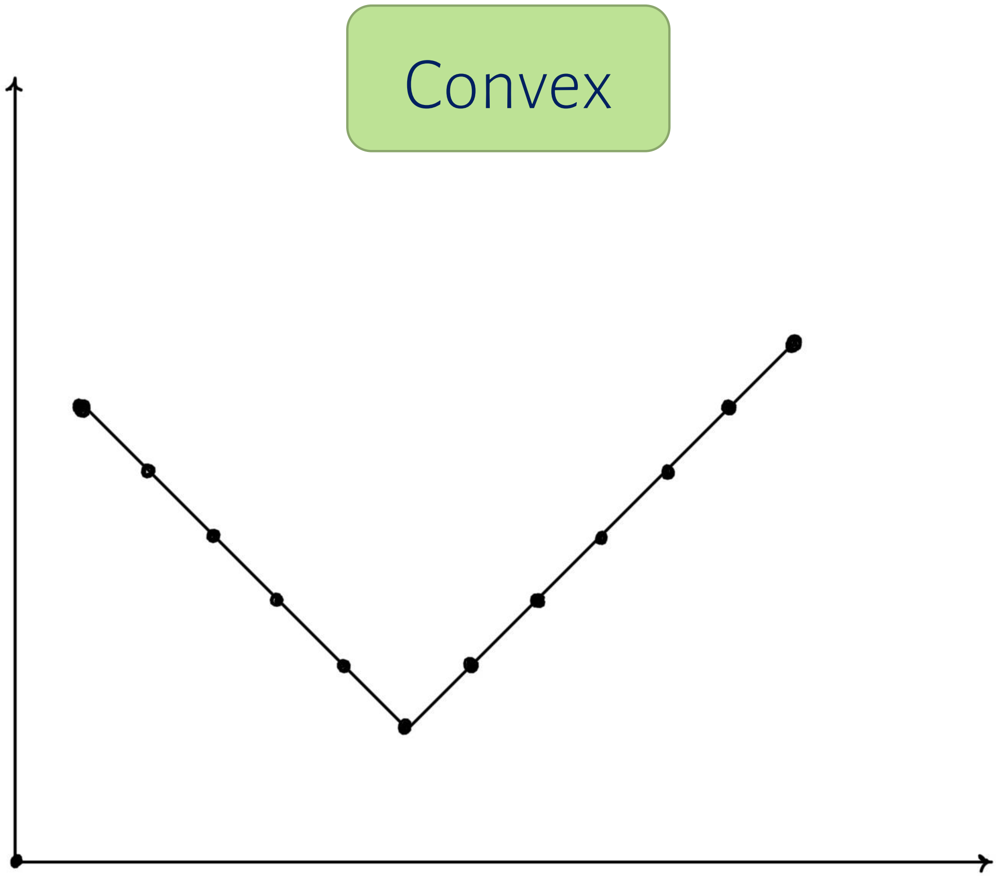
Functions with 2 Distinct Derivatives



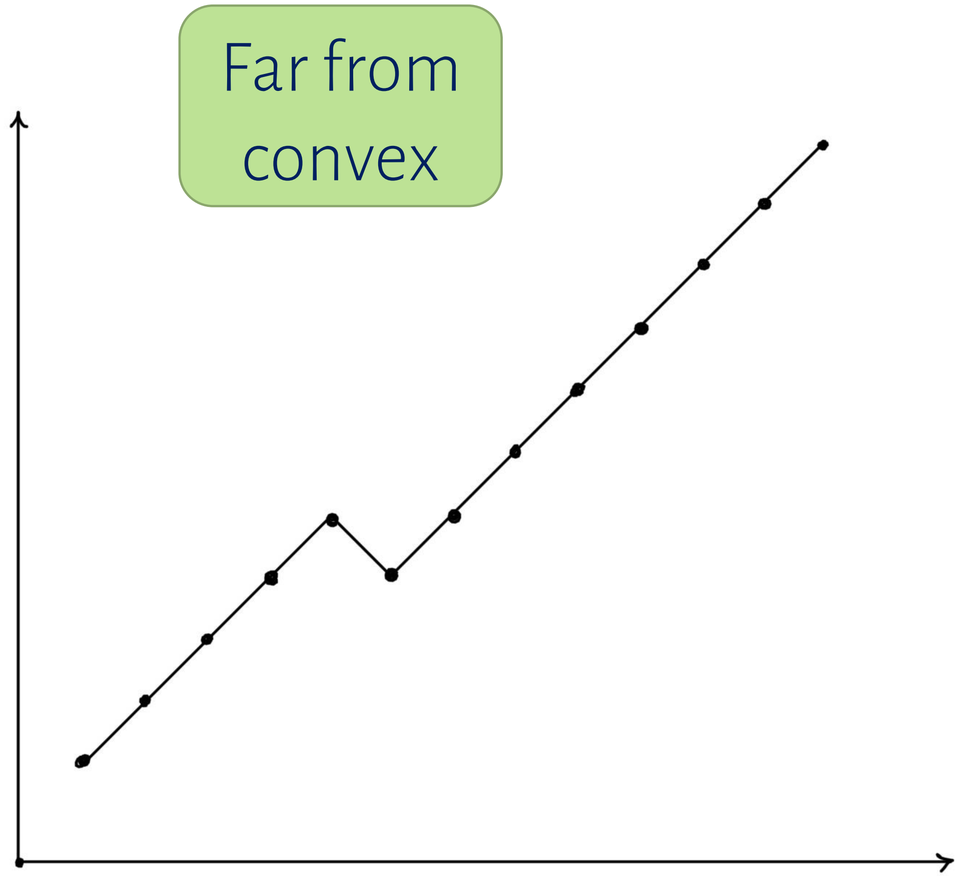


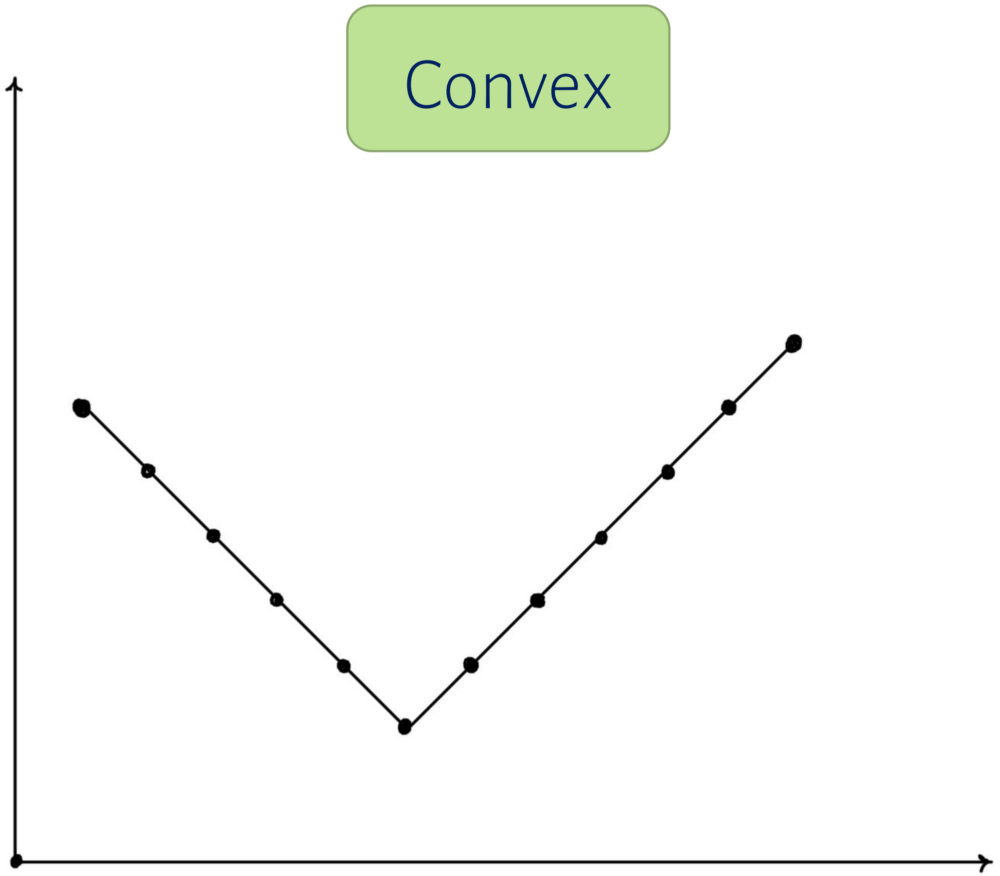
Functions with 2 Distinct Derivatives





Functions with 2 Distinct Derivatives

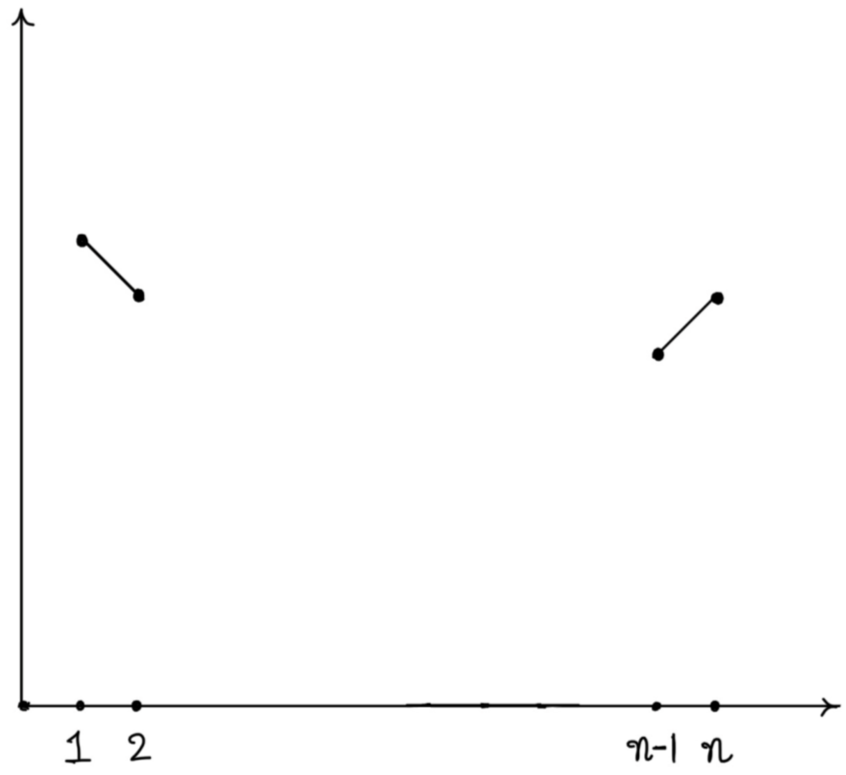




Is it easier to distinguish between these two cases?

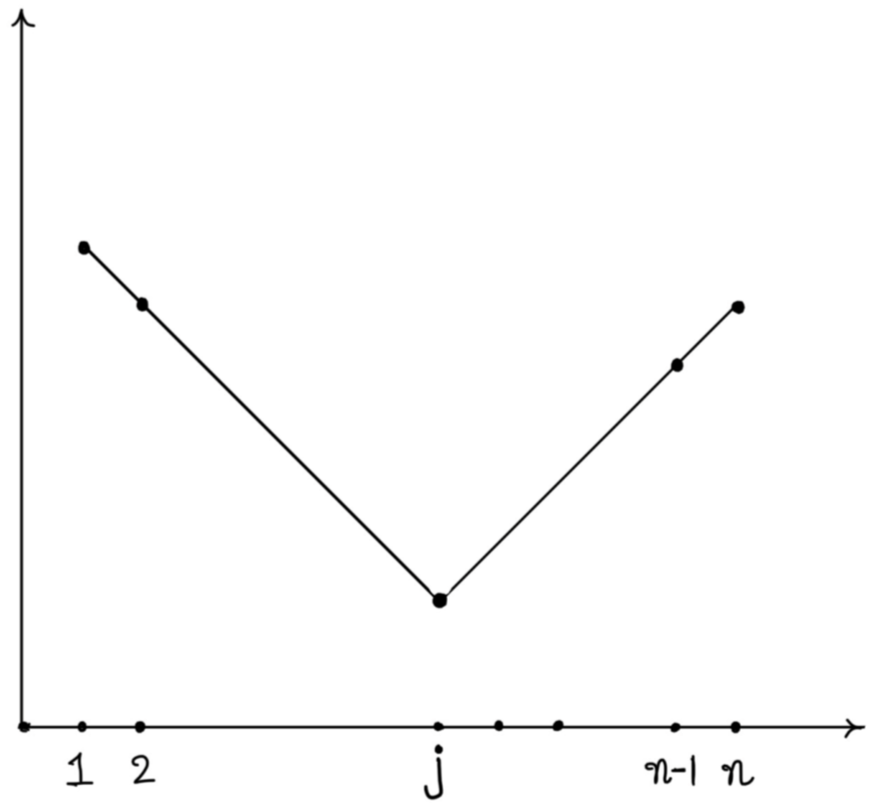
Functions with 2 Distinct Derivatives

- **Observation**: Can decide convexity exactly in **5** queries!!
- How?
  - Query f at $1,2,n-1,n$
  - If convexity is violated on these points, reject immediately



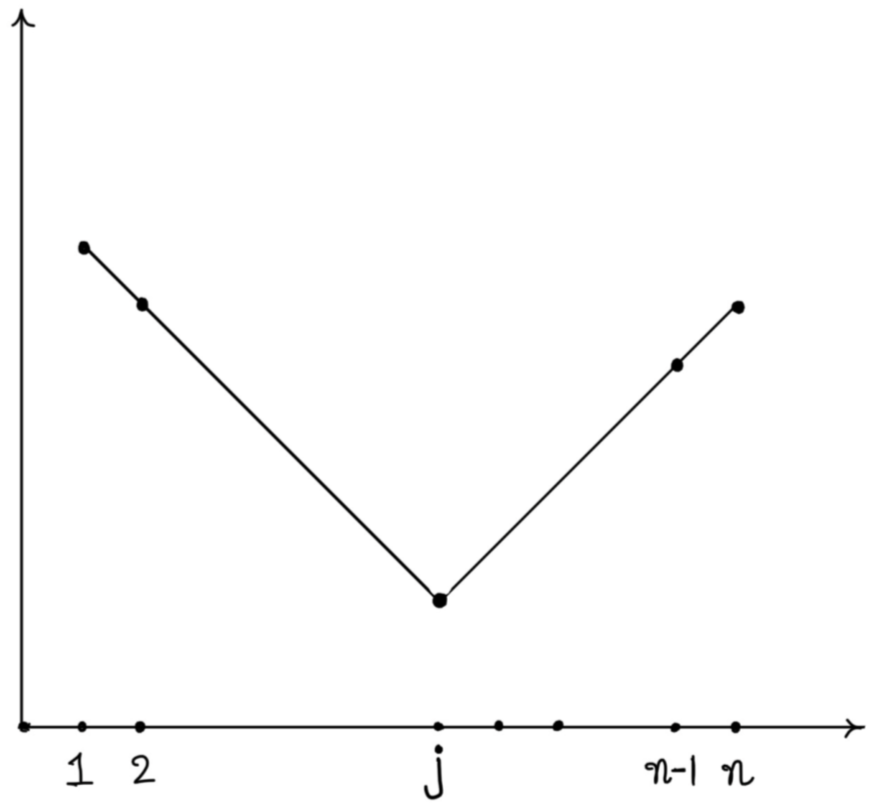
Functions with 2 Distinct Derivatives

- **Observation**: Can decide convexity exactly in **5** queries!!
- How?
  - Query f at $1,2,n-1,n$
  - If convexity is violated on these points, **reject** immediately
  - Else, find $j \in [n]$ where the linear extensions intersect
  - If $f(j)$ is not the same as the predicted value, **reject**



Functions with 2 Distinct Derivatives

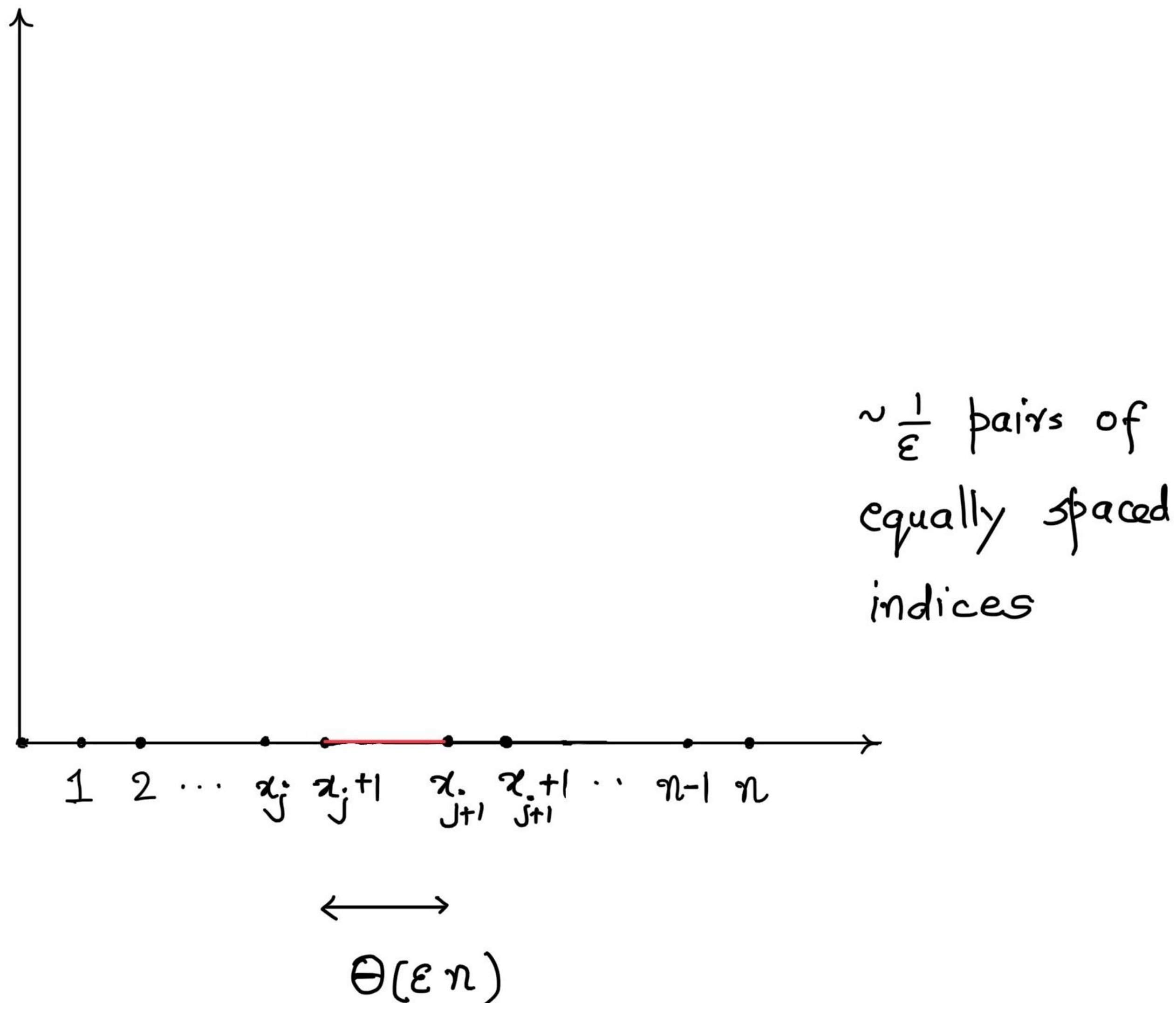
- **Observation**: Can decide convexity exactly in **5** queries!!
- How?
  - Query f at $1, 2, n-1, n$
  - If convexity is violated on these points, **reject** immediately
  - Else, find $j \in [n]$ where the linear extensions intersect
  - If $f(j)$ is not the same as the predicted value, **reject**



Adaptive algorithm! Last query depends on answers to previous ones.

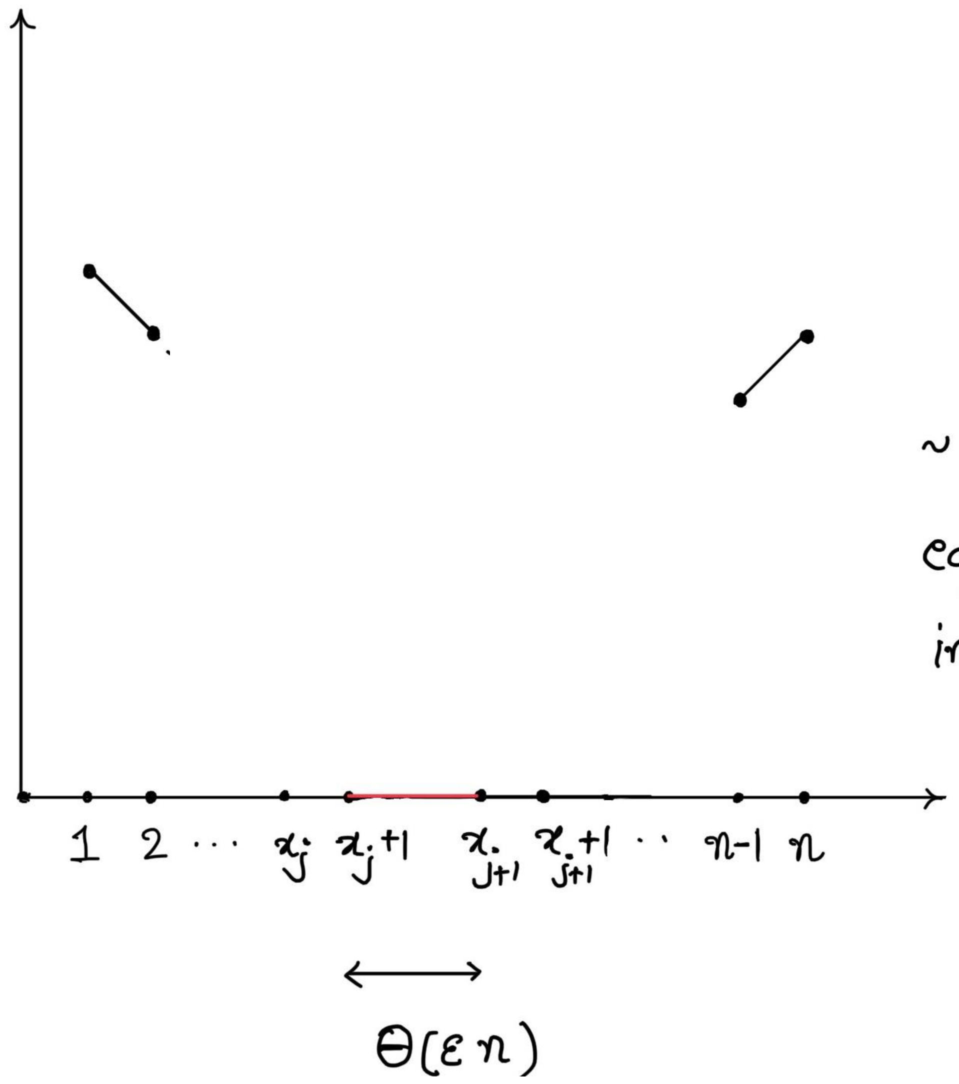
Functions with 2 Distinct Derivatives

- **Observation**: Can ϵ-test convexity in $O\left(\frac{1}{\epsilon}\right)$ nonadaptive queries.
- **How?**
  - Query f at $\Theta\left(\frac{1}{\epsilon}\right)$ equally spaced adjacent pairs of points



Functions with 2 Distinct Derivatives

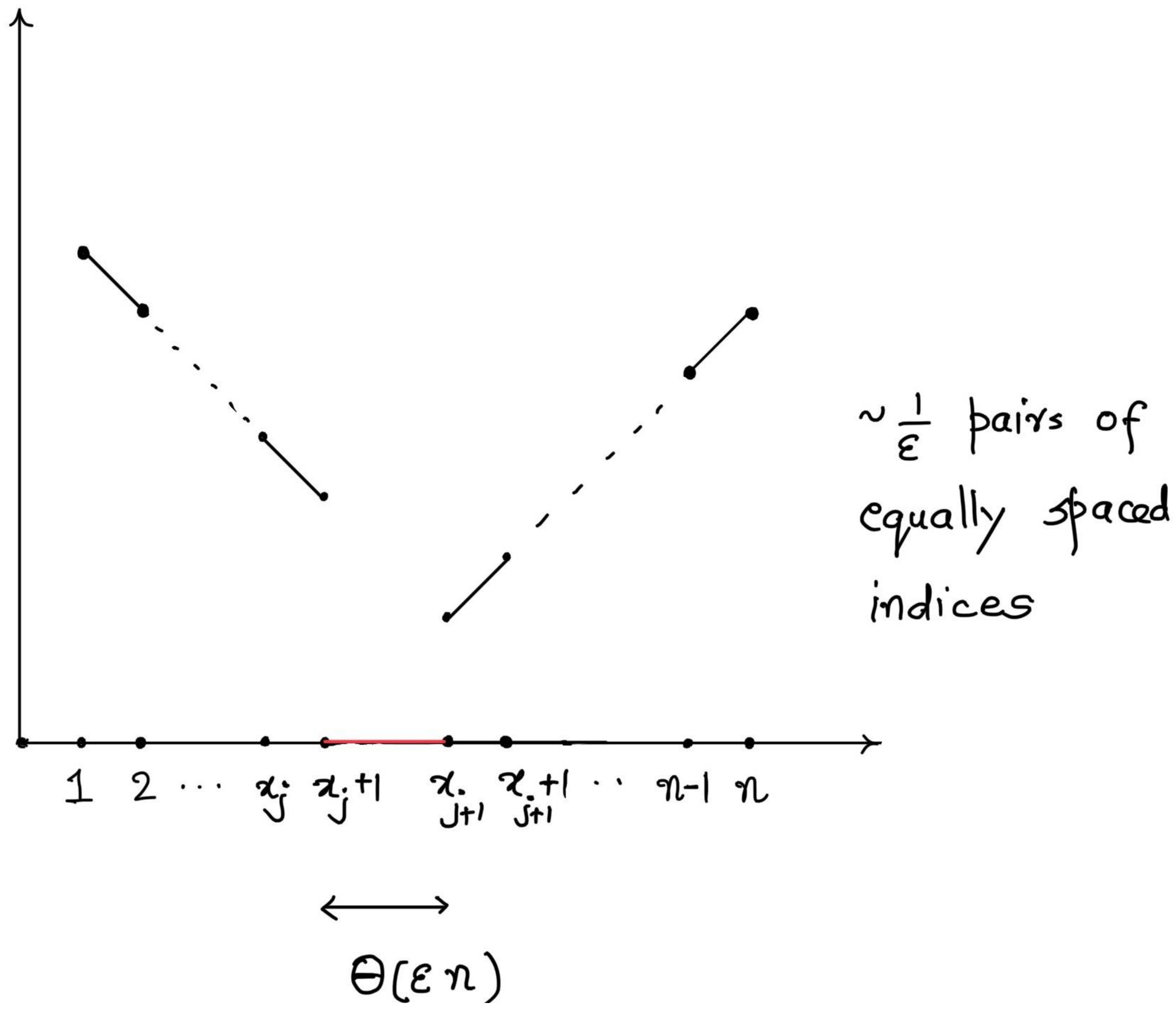
- **Observation**: Can ϵ-test convexity in $O\left(\frac{1}{\epsilon}\right)$ nonadaptive queries.
- How?
  - Query f at $\Theta\left(\frac{1}{\epsilon}\right)$ equally spaced adjacent pairs of points



$\sim \frac{1}{\varepsilon}$ pairs of equally spaced indices

Functions with 2 Distinct Derivatives

- **Observation**: Can ϵ-test convexity in $O\left(\frac{1}{\epsilon}\right)$ nonadaptive queries.
- **How?**
  - Query f at $\Theta\left(\frac{1}{\epsilon}\right)$ equally spaced adjacent pairs of points
  - **Reject** if the values do not correspond to the linear extensions



Our Results

Functions with $\leq s$ Distinct Derivatives

- Nonadaptive convexity ϵ-tester with query complexity $O\left(\frac{\log s}{\epsilon}\right)$
  - Circumvents the lower bound of $\Omega(\log n)$ when $s \ll n^{o(1)}$
  - Bridge between the case of 2 distinct derivatives and the general case
  - Brings out the fine-grained complexity of convexity testing
- Every adaptive tester must make $\Omega\left(\frac{\log \epsilon s}{\epsilon}\right)$ queries
  - Generalization of the lower bound by Belovs Blais Bommireddi '20

Our Results

Functions with $\leq s$ Distinct Derivatives

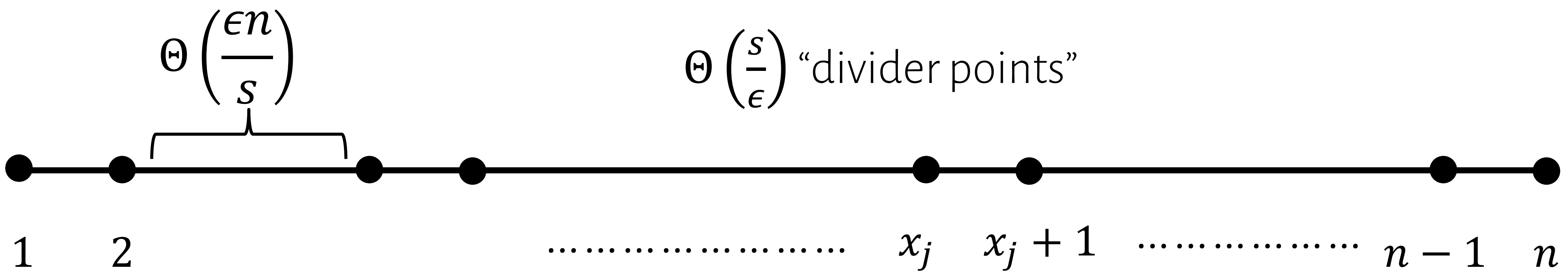
Rest of the talk

- Nonadaptive convexity ϵ-tester with query complexity $O\left(\frac{\log s}{\epsilon}\right)$
  - Circumvents the lower bound of $\Omega(\log n)$ when $s \ll n^{o(1)}$
  - Bridge between the case of 2 distinct derivatives and the general case
  - Brings out the fine-grained complexity of convexity testing
- Every adaptive tester must make $\Omega\left(\frac{\log \epsilon s}{\epsilon}\right)$ queries
  - Generalization of the lower bound by Belovs Blais Bommireddi '20

One Iteration of Our Tester

$\Theta\left(\frac{1}{\epsilon}\right)$ iterations

- **Inputs**: $\epsilon \in (0,1)$, upper bound s on distinct derivatives, oracle access to function $f: [n] \to \mathbb{R}$
- **Phase 1**: Run a convexity ϵ'-tester on f restricted to $\Theta\left(\frac{s}{\epsilon}\right)$ equally spaced pairs of adjacent points, where $\epsilon' = \Omega(\epsilon)$ and **reject** if it **rejects.**

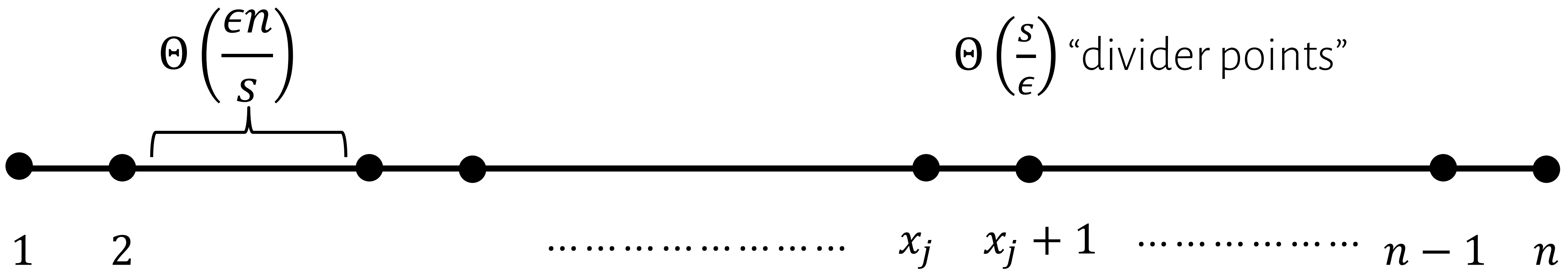


- **Phase 2**: Sample a random interval $I_j = [x_j, x_{j+1}]$ and a random index $y \in I_j$. **Reject** if convexity violated on $x_j, x_j + 1, y, x_{j+1}, x_{j+1} + 1$.

Analysis Idea

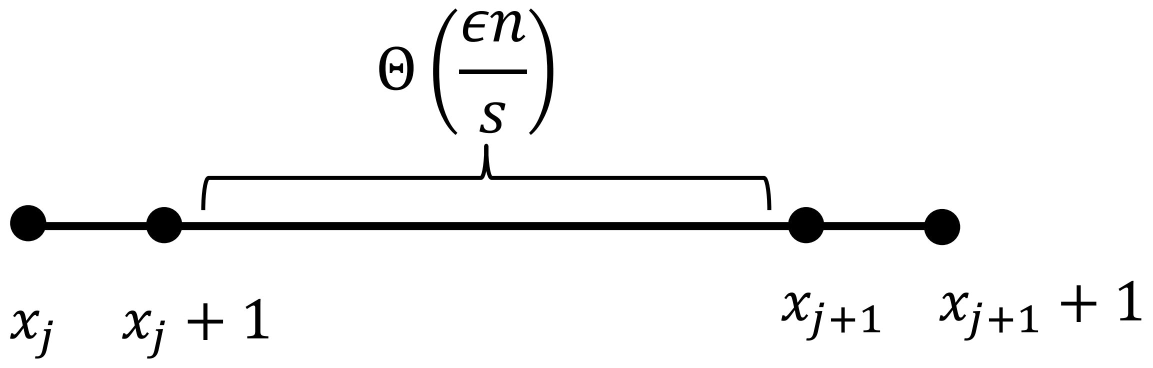
- Convex functions are not rejected by the tester
  - Easy direction
- If f is ϵ-far from convex, an iteration rejects with probability $\Omega(\epsilon)$
  - Phase 1 uses a variant of convexity tester by [Belovs Blais Bommireddi '20] generalized to work for arbitrary subdomains of $[n]$
  - We will analyze the rejection probability of Phase 2, in case Phase 1 does not reject
- Query complexity per iteration is $O(\log s)$
  - Phase 2 makes constant number of queries
  - Phase 1 makes $O(\log s)$ queries

If Phase 1 does not reject…



- If Phase 1 does not reject, f restricted to the "divider" points can be modified on $O(\epsilon) \cdot \Theta\left(\frac{s}{\epsilon}\right) = O(s)$ "**bad** divider points" and made convex
- Interval $[x_j, x_{j+1}]$ is **bad** if either $x_j, x_j + 1$, x_{j+1} or $x_{j+1} + 1$ is a bad divider point
- Number of **bad intervals** is $O(s)$

Good intervals



- For every good interval,
$$f(x_j + 1) - f(x_j) \leq \frac{f(x_{j+1}) - f(x_j + 1)}{x_{j+1} - x_j - 1} \leq f(x_{j+1} + 1) - f(x_{j+1})$$
- Number of good intervals with some inequality strict is $O(s)$
  - Number of distinct adjacent derivatives in f restricted to the good divider points is $\leq s$

Phase 2 rejection event

- $O(s)$ intervals that are either **Bad** or **Good with strict inequalities**
- Number of indices in such intervals is $O(s) \cdot \Theta\left(\frac{\epsilon n}{s}\right) = O(\epsilon n)$
- Since f is ϵ-far from convex, there are at least ϵn indices with "bad values"
- At least $\Omega(\epsilon n)$ indices with "bad values" belong to good intervals satisfying
  $$f(x_j + 1) - f(x_j) = \frac{f(x_{j+1}) - f(x_j + 1)}{x_{j+1} - x_j - 1} = f(x_{j+1} + 1) - f(x_{j+1})$$
- Phase 2 will capture such an index with probability $\Omega(\epsilon)$ and the tester will reject

Summary of our results

For functions with $\leq s$ distinct derivatives

- Nonadaptive convexity ϵ-tester with query complexity $O\left(\frac{\log s}{\epsilon}\right)$
  - Circumvents the lower bound of $\Omega(\log n)$ when $s \ll n^{o(1)}$
  - Bridge between the case of 2 distinct derivatives and the general case
  - Brings out the fine-grained complexity of convexity testing
- Every ϵ-tester must make $\Omega\left(\frac{\log \epsilon s}{\epsilon}\right)$ queries

Open questions and directions

- What about convexity of functions over higher dimensional domains?
  - Several definitions of convexity exist – separate convexity, M-convexity, L-convexity, …
  - Not much is known for the general case of functions $f: [n]^d \to \mathbb{R}$
    - Nonadaptive tester with complexity $O(n^{d-1})$ [Ben-Eliezer '19]
    - Lower bound of $\Omega\left(\left(\frac{n}{d}\right)^{\frac{d}{2}}\right)$ [Belovs Blais Bommireddi '20]
  - Restricting attention to functions with bounded number of distinct derivatives might be useful to improve understanding
- Investigate the right parameters for other problems having sublinear algorithms

Thank you!