Sublinear Algorithms
Lecture 5

Sofya Raskhodnikova

Penn State University

Thanks to Madhav Jha (Penn State) for help with creating these slides.

Today

Lecture 5. Limitations of sublinear algorithms. Yao's Minimax Principle.

Query Complexity

- Query complexity of an algorithm is the maximum number of queries the algorithm makes.
  - Usually expressed as a function of input length (and other parameters)
  - Example: the test for sortedness (from Lecture 2) had query complexity O(log n) for constant ε.
  - running time $\geq$ query complexity
- Query complexity of a problem P, denoted $q(P)$, is the query complexity of the best algorithm for the problem.
  - What is $q(\text{testing sortedness})$? How do we know that there is no better algorithm?

Today: Techniques for proving lower bounds on $q(P)$.

Yao's Principle

A Method for Proving Lower Bounds

Yao's Minimax Principle

The following statements are equivalent.

Statement 1

For any **probabilistic** algorithm A of complexity q there exists an input x s.t.

$$\Pr_{coin\ tosses\ of\ A}[A(x) \text{ is wrong}] > 1/3.$$

Statement 2

There is a distribution *D* on the inputs,

s.t. for every **deterministic** algorithm of complexity q,

$$\Pr_{x \leftarrow D}[A(x) \text{ is wrong}] > 1/3.$$

- Need for lower bounds

Yao's Minimax Principle (easy direction): Statement 2 $\Rightarrow$ Statement 1.

 Prove it.

Yao's Minimax Principle as a game

Players: Evil algorithms designer Al and poor lower bound prover Lola.

Game1

Move 1. Al selects a q-query **randomized** algorithm A for the problem.

Move 2. Lola selects an input on which A errs with largest probability.

Game2

Move 1. Lola selects a distribution on inputs.

Move 2. Al selects a q-query **deterministic** algorithm with as large probability of success on Lola's distribution as possible.

*A Lower Bound for Testing 1**

Input: string of n bits

Question: Is the string contains only 1's or is it ε-far form the all-1 string?

Claim. Any algorithm needs $\Omega(1/\varepsilon)$ queries to answer this question w.p. $\geq \mathbf{2/3}$.

Proof: By Yao's Minimax Principle, enough to prove Statement 2.

Distribution on n-bit strings:

- Divide the input string into $1/\varepsilon$ blocks of size $\varepsilon \mathrm{n}$.
- Let y_i be the string where the ith block is 0's and remaining bits are 1.
- Distribution D gives the all-1 string w.p. 1/2 and y_i with w.p. 1/2, where i is chosen uniformly at random from 1, …, $1/\varepsilon$.

*A Lower Bound for Testing 1**

Claim. Any ε -test for 1* needs $\Omega(1/\varepsilon)$ queries.

Proof (continued): Now fix a deterministic tester A making q < $1/3\varepsilon$ queries.

1. A must accept if all answers are 1. Otherwise, it would be wrong on all-1 string, that is, with probability 1/2 with respect to D.
2. Let $i_1, \ldots, i_q$ be the positions A queries when it sees only 1s. The test can choose its queries based on previous answers. However, since all these answers are 1 and since A is deterministic, the query positions are fixed.

- At least $1/\varepsilon$ − q > $2/3\varepsilon$ of the blocks do not hold any queried indices.
- Therefore, A accepts > 2/3 of the inputs y_i. Thus, it is wrong with probability > $2/3\varepsilon \cdot \frac{\varepsilon}{2} = 1/3$

Context: [Alon Krivelevich Newman Szegedy 99]

Every regular language can be tested in $O(1/\varepsilon \text{ polylog } 1/\varepsilon)$ time

A Lower Bound for Testing Sortedness

Input: a list of *n* numbers $x_1, x_2, \ldots, x_n$

Question: Is the list sorted or ε-far from sorted?

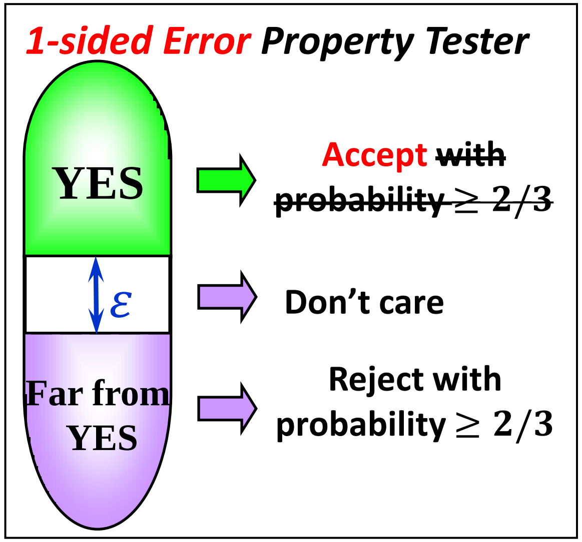
Already saw: two different $O((\log n)/\varepsilon)$ time testers.

Known [Ergün Kannan Kumar Rubinfeld Viswanathan 98, Fischer 01]:

$\Omega(\log n)$ queries are required for all constant $\varepsilon \leq 1/2$

Today: $\Omega(\log n)$ queries are required for all constant $\varepsilon \leq 1/2$ for every 1-sided error nonadaptive test.

- A test has 1-sided error if it always accepts all YES instances.
- A test is nonadaptive if its queries do not depend on answers to previous queries.



1-Sided Error Tests Must Catch "Mistakes"

- A pair (x_i, x_j) is **violated** if $x_i < x_j$

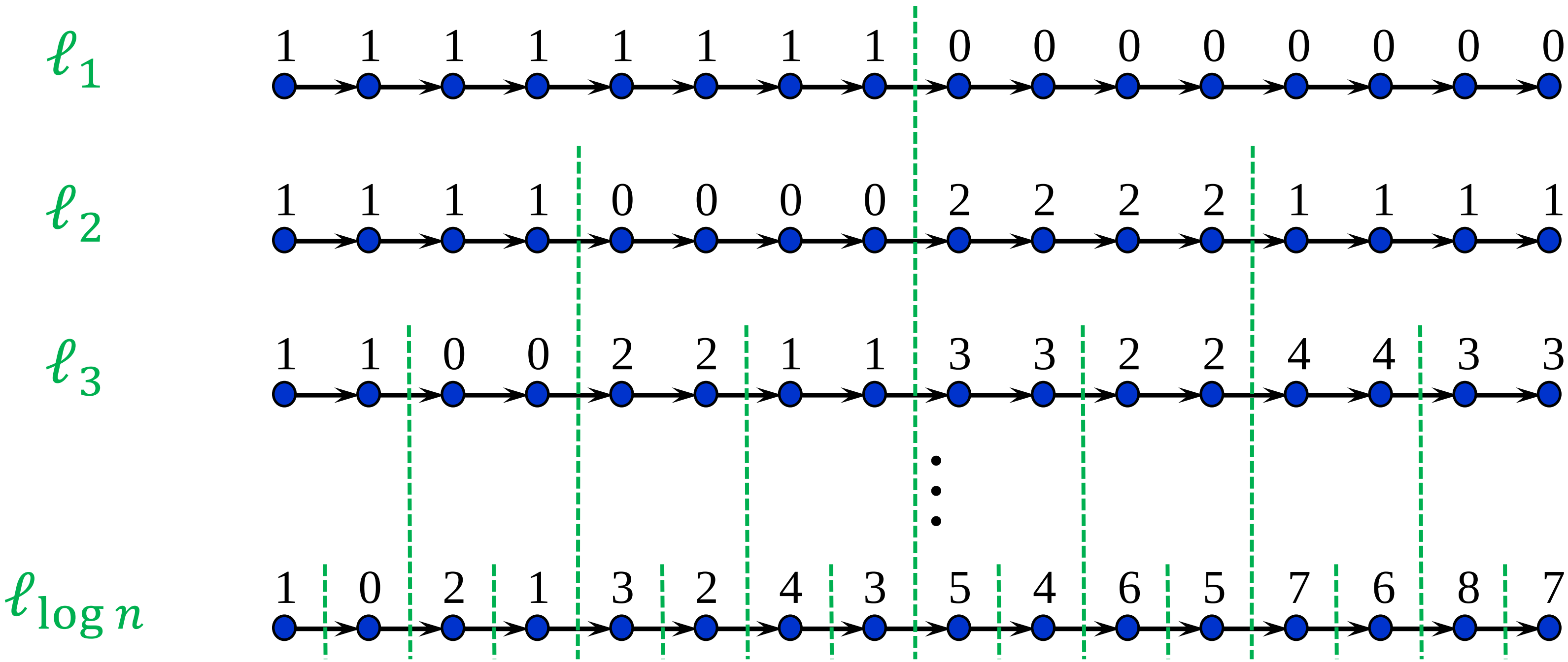
Claim. A 1-sided error test can reject only if it finds a violated pair.

Proof: Every sorted partial list can be extended to a sorted list.

1	?	?	4	…	7	?	?	9

Yao's Principle Game [Jha]

Lola's distribution is uniform over the following $\log n$ lists:

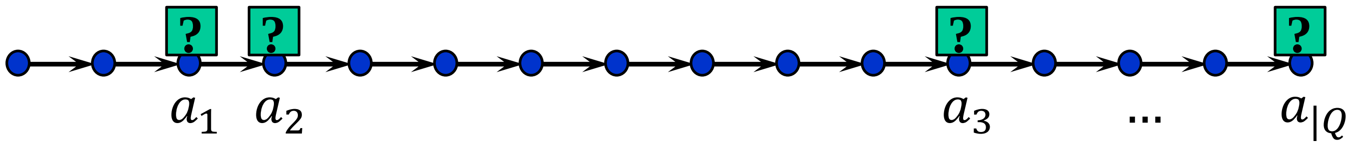


Claim 1. All lists above are 1/2-far from sorted.

Claim 2. Every pair (x_i, x_j) is violated in exactly one list above.

Yao's Principle Game: Al's Move

Al picks a set $Q = \{a_1, a_2, \ldots, a_{|Q|}\}$ of positions to query.



- His test must be correct, i.e., must find a violated pair with probability $\geq 2/3$ when input is picked according to Lola's distribution.
- Q contains a violated pair $\Leftrightarrow$ (a_i, a_{i+1}) is violated for some i

$$\Pr_{\ell \leftarrow \text{Lola's distribution}}[(a_i, a_{i+1}) \text{ for some } i \text{ is vilolated in list } \ell] \leq \frac{|Q|-1}{\log n}$$

By the Union Bound

- If $|Q| \leq \frac{2}{3} \log n$ then this probability is $< \frac{2}{3}$
- So, $|Q| = \Omega(\log n)$
- By Yao's Minimax Principle, every randomized 1-sided error nonadaptive test for sortedness must make $\Omega(\log n)$ queries. ✓

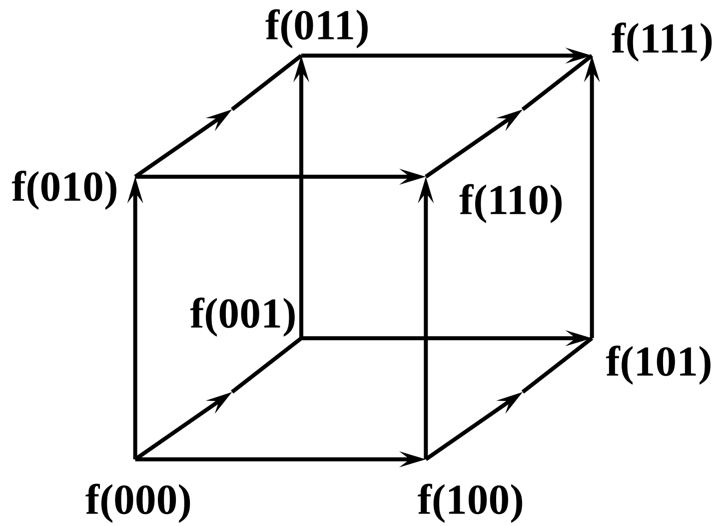
Testing Monotonicity of functions on Hypercube

Non-adaptive 1-sided error Lower Bound

Boolean Functions $f: \{0,1\}^n \to \{0,1\}$

Graph representation:

n-dimensional hypercube



- vertices: bit strings of length n
- edges: (x, y) is an edge if y can be obtained from x by increasing one bit from 0 to 1

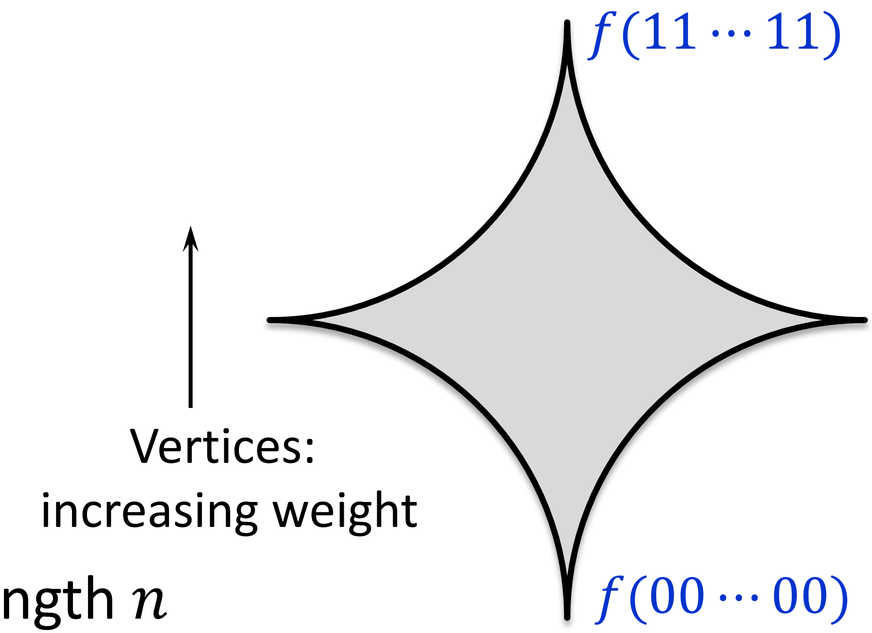
x	001001
y	011001

- each vertex x is labeled with $f(x)$

Boolean Functions $\boldsymbol{f : \{0,1\}^n \to \{0,1\}}$

Graph representation:

n-dimensional hypercube



- 2^n vertices: bit strings of length n
- $2^{n-1}n$ edges: (x, y) is an edge if y can be obtained from x by increasing one bit from 0 to 1

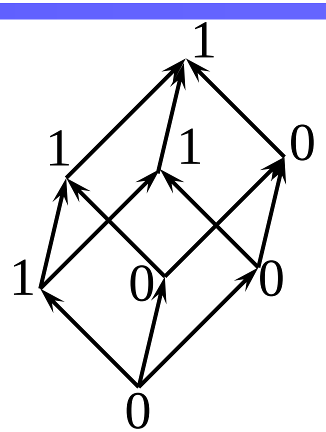
x	001001
y	011001

- each vertex x is labeled with $f(x)$

Monotonicity of Functions

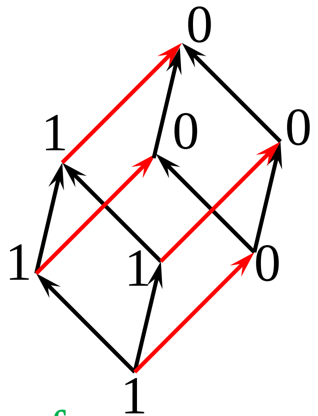
[Goldreich Goldwasser Lehman Ron Samorodnitsky,
Dodis Goldreich Lehman Raskhodnikova Ron Samorodnitsky]

- A function $f : \{0,1\}^n \to \{0,1\}$ is monotone
  if increasing a bit of x does not decrease $f(x)$.



monotone

- Is f monotone or ε-far from monotone?
  - Edge $x \rightarrow y$ is violated by f if $f(x) > f(y)$.



$\frac{1}{2}$-far from monotone

Time:

- $O(n/\varepsilon)$, logarithmic in the size of the input, 2^n
- $\Omega(\sqrt{n}/\varepsilon)$ for restricted class of tests

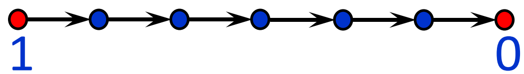
Hypercube 1-sided Error Lower Bound

Lemma [Fischer Lehman Newman Raskhodnikova Rubinfeld Samorodnitsky]

Every 1-sided error non-adaptive test for monotonicity of functions $f : \{0,1\}^n \to \{0,1\}$ requires $\Omega(\sqrt{n})$ queries.

- 1-sided error test must accept if no violated pair is uncovered.

Violated pair:

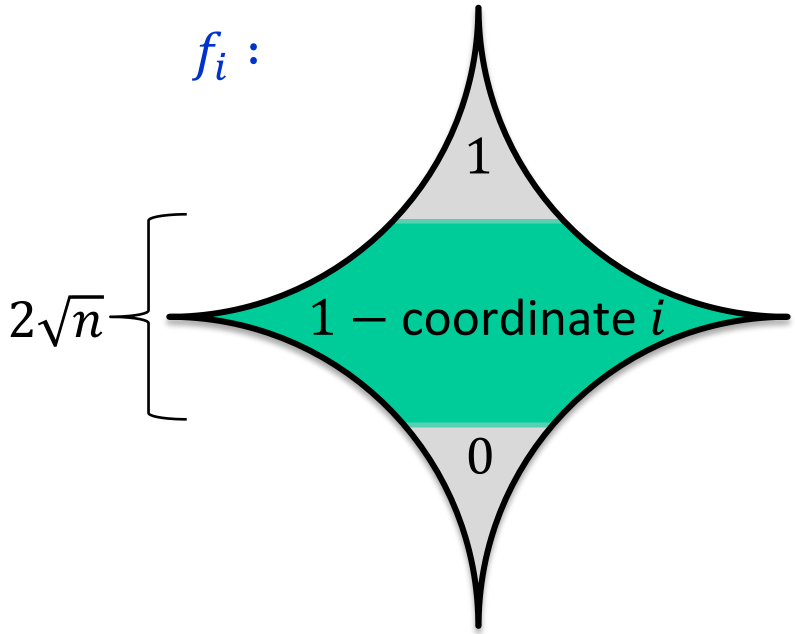


  - Only a distribution on far from monotone values suffices.

Hypercube 1-sided Error Lower Bound

- Hard distribution: pick coordinate i at random and output f_i.

f_i :

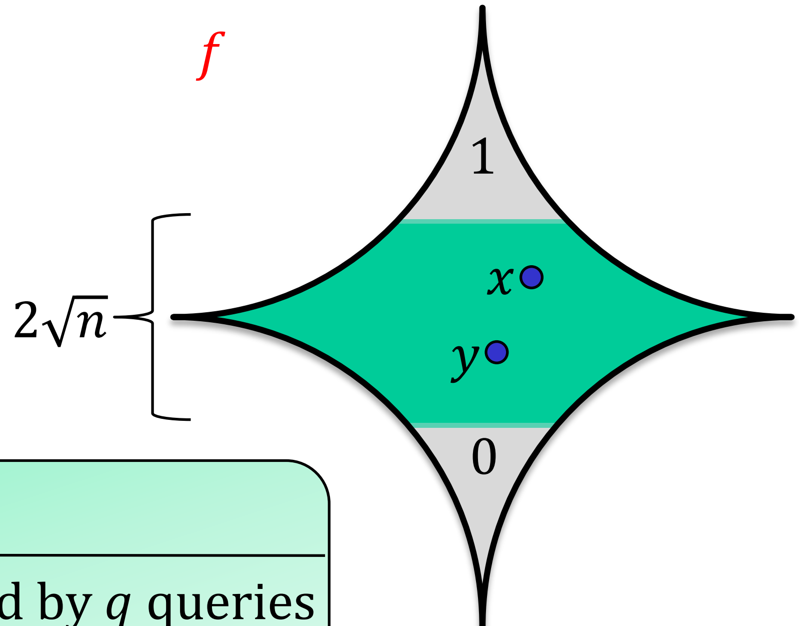


Analysis

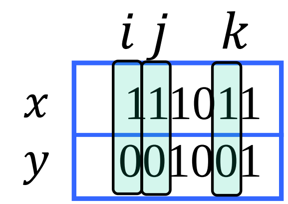
- Edges from $(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$ to $(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)$ are violated if both endpoints are in the middle.
- The middle contains a constant fraction of vertices.
- All n functions are ε-far from monotone for some constant ε.

Hypercube 1-sided Error Lower Bound

- How many functions does a set of q queries expose?



● queries



Pair (x, y)
can expose only
functions f_i, f_j and f_k

Naïve Analysis

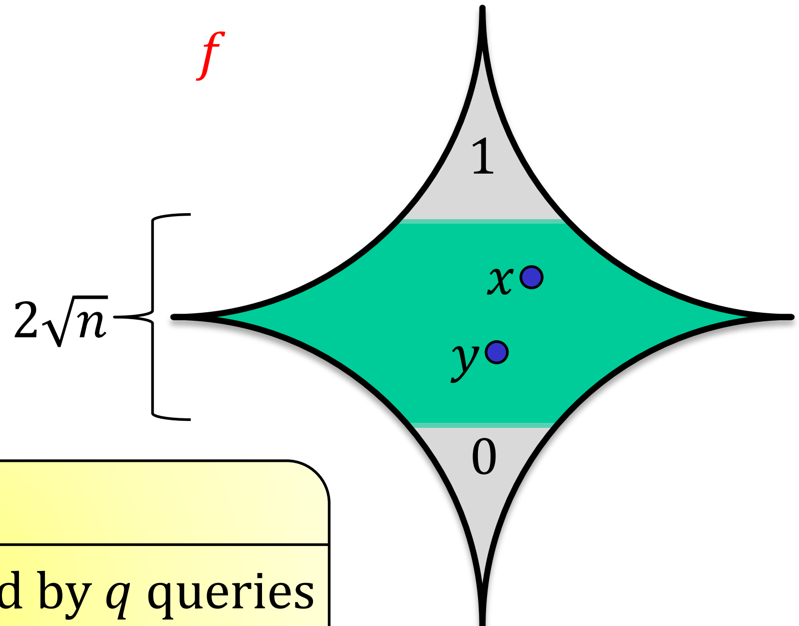
\# functions exposed by q queries
$\leq q^2 \cdot 2\sqrt{n}$

\# functions that a query pair (x, y) exposes
$\leq$ # coordinates on which x and y differ
$\leq 2\sqrt{n}$

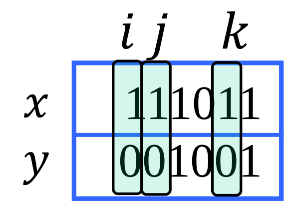
Only queries in the Green Band can be violated $\Rightarrow$ disagreements $\leq 2\sqrt{n}$

Hypercube 1-sided Error Lower Bound

- How many functions does a set of q queries expose?



● queries



Pair (x, y)
can expose only
functions f_i, f_j and f_k

Claim

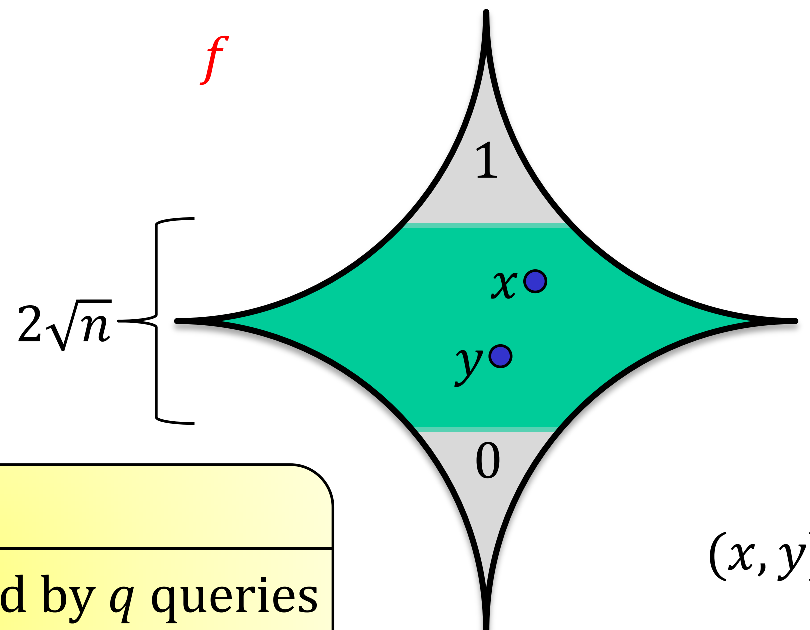
functions exposed by q queries
$\leq (q-1) \cdot 2\sqrt{n}$

functions that a query pair exposes
$\leq$ # disagreements between vertices of the pair
$\leq 2\sqrt{n}$

Only queries in the Green Band can be violated $\Rightarrow$ disagreements $\leq 2\sqrt{n}$

Hypercube 1-sided Error Lower Bound

- How many functions does a set of q queries expose?



● queries

(x, y) a violation pair

$\Downarrow$

Some adjacent pair of vertices in a minimum spanning forest on the query set is also violated

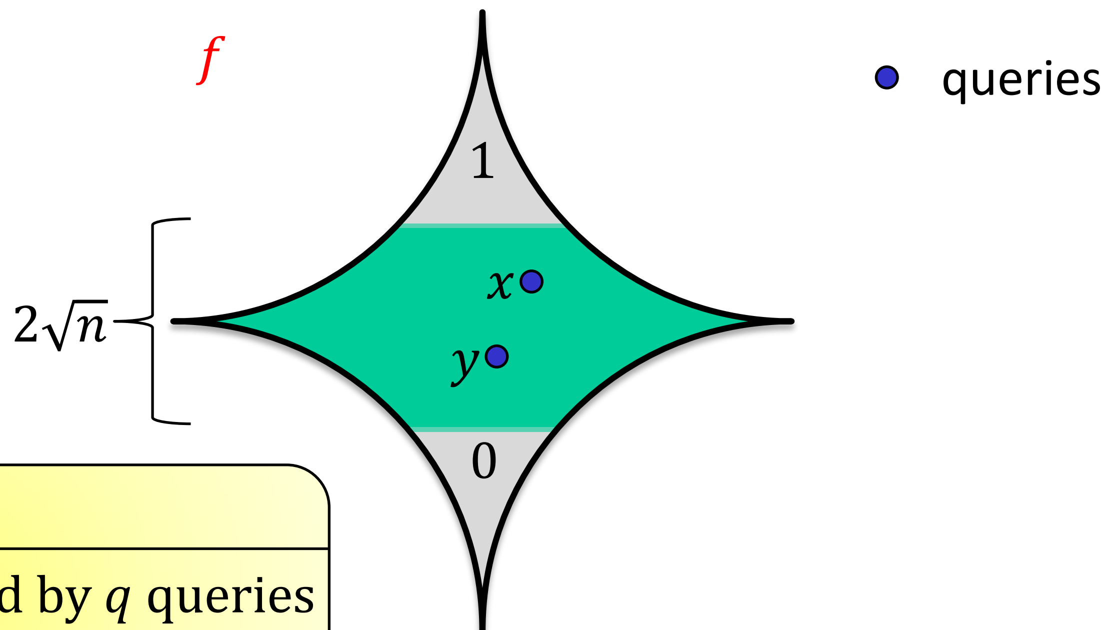
Claim

\# functions exposed by q queries
$\leq (q-1) \cdot 2\sqrt{n}$

sufficient to consider adjacent vertices in a minimum spanning forest on the query set

Hypercube 1-sided Error Lower Bound

- How many functions does a set of q queries expose?



Claim

\# functions exposed by q queries
$\leq (q-1) \cdot 2\sqrt{n}$

$\Downarrow$

Claim

Every deterministic test that makes a set Q of q queries (in the middle) succeeds with probability $O\left(\frac{q}{\sqrt{n}}\right)$ on our distribution.

Testing Monotonicity of functions on Hypercube

Non-adaptive 2-sided error Lower Bound

Hypercube 2-sided Error Lower Bound

Lemma [Fischer Lehman Newman Raskhodnikova Rubinfeld Samorodnitsky]

Every test for monotonicity of functions $f : \{0,1\}^n \to \{0,1\}$ requires $\Omega(\log n)$queries.

Hard distribution: randomly pick a subset B of coordinates from $[n]$ by independently choosing each coordinate to lie in B with probability $\frac{1}{10\sqrt{n}}$.

Uniformly choose good_B or bad_B.