LECTURE 2

Last time

- Introduction
- Basic models for sublinear-time computation
- Simple examples of sublinear algorithms

Today

- Properties of lists and functions.
- Testing if a list is sorted/Lipschitz and if a function is monotone.
- Revisiting half-plane testing

Sofya Raskhodnikova;Boston University

Reminders

HW1 is due next Thursday at 11am
It is posted on the course webpage:
https://cs-people.bu.edu/sofya/sublinear-course/

Use Piazza for questions and discussions

Office hours:
Wednesdays, 1:30PM-3:00PM

Testing if a List is Sorted

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

- Question: Is the list sorted?

  Requires reading entire list: $\Omega(n)$ time

- Approximate version: Is the list sorted or ε-far from sorted?

  (An ε fraction of x_i's must be changed to make it sorted.)

  [Ergün Kannan Kumar Rubinfeld Viswanathan 98, Fischer 01]: $O((\log n)/\varepsilon)$ time

  $\Omega(\log n)$ queries

- Best known bounds:

  $\Theta(\log(\varepsilon n)/\varepsilon)$ time

  [Belovs, Chakrabarty Dixit Jha Seshadhri 15]

Testing Sortedness: Attempts

1. **Test**: Pick a uniformly random $i \in \{1, \dots, n-1\}$ and reject if $x_i > x_{i+1}$.

Fails on:

1	1	1	1	0	0	0	0

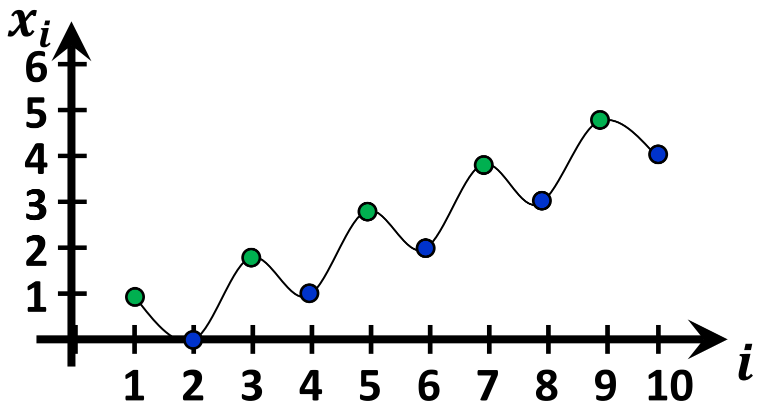
Ã 1/2-far from sorted

2. **Test**: Pick uniformly random $i < j$ in $\{1, \dots, n\}$ and reject if $x_i > x_j$.

Fails on:

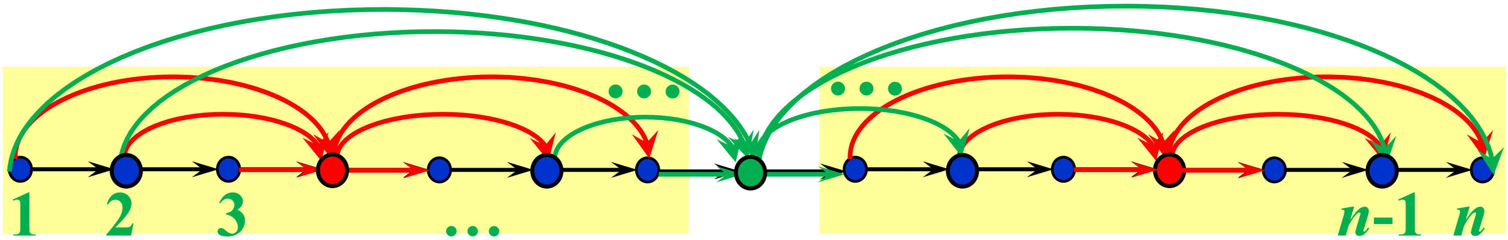
1	0	2	1	3	2	4	3	5	4

Ã 1/2-far from sorted



Is a List Sorted or ε-far from Sorted?

Idea: Associate positions in the list with vertices of the directed line.



Construct a graph (2-spanner)

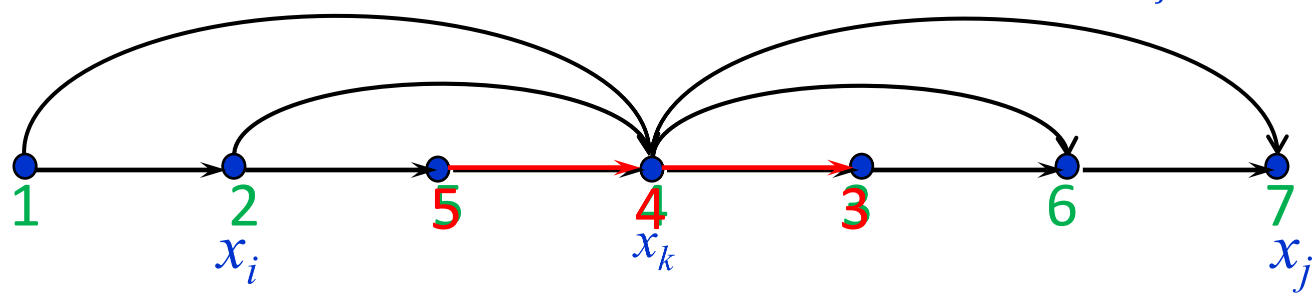
$\leq n \log n$ edges

- by adding a few “shortcut” edges (i, j) for $i < j$
- where each pair of vertices is connected by a path of length at most 2 ✓

Is a List Sorted or ε-far from Sorted?

Test [Dodis Goldreich Lehman Raskhodnikova Ron Samorodnitsky 99]

Pick a random edge (i, j) from the 2-spanner and **reject** if $x_i > x_j$.



Analysis:

- Call an edge (i, j) **violated** if $x_i > x_j$, and **satisfied** otherwise.
- If i is an endpoint of a **violated** edge, call x_i **bad**. Otherwise, call it **good**.

Claim 1. All **good** numbers x_i are sorted.

Proof: Consider any two good numbers, x_i and x_j.

They are connected by a path of (at most) two **satisfied** edges $(i, k), (k, j)$

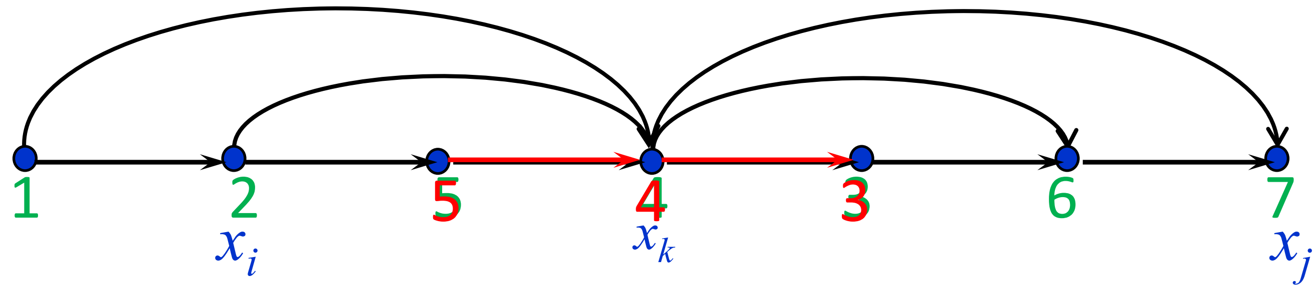
$\Rightarrow x_i \leq x_k$ and $x_k \leq x_j$

$$\Rightarrow x_i \leq x_j$$

Is a List Sorted or ε-far from Sorted?

Test [Dodis Goldreich Lehman Raskhodnikova Ron Samorodnitsky 99]

Pick a random edge (i, j) from the 2-spanner and **reject** if $x_i > x_j$.



Analysis:

- Call an edge (i, j) **violated** if $x_i > x_j$, and **satisfied** otherwise.
- If i is an endpoint of a **violated** edge, call x_i **bad**. Otherwise, call it **good**.

Claim 1. All **good** numbers x_i are sorted.

Claim 2. An ε-far list **violates** $\geq \varepsilon/(2 \log n)$ fraction of edges in 2-spanner.

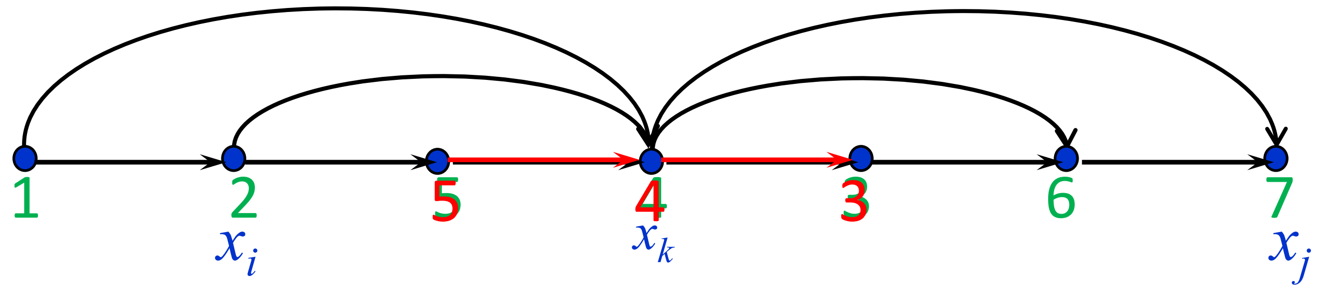
Proof: If a list is ε-far from sorted, it has $\geq \varepsilon n$ **bad** numbers. (Claim 1)

- Each **violated** edge contributes 2 **bad** numbers.
- 2-spanner has $\geq \frac{\varepsilon n}{2}$ **violated** edges out of $n \log n$.

Is a List Sorted or ε-far from Sorted?

Test [Dodis Goldreich Lehman Raskhodnikova Ron Samorodnitsky 99]

Pick a random edge (i, j) from the 2-spanner and **reject** if $x_i > x_j$.



Analysis:

- Call an edge (i, j) **violated** if $x_i > x_j$, and **satisfied** otherwise.

Claim 2. An ε-far list **violates** $\geq \varepsilon/(2 \log n)$ fraction of edges in 2-spanner.

By Witness Lemma, it suffices to sample $(4 \log n)/\varepsilon$ edges from 2-spanner.

Algorithm

Sample $\frac{4 \log n}{\varepsilon}$ edges (i, j) from the 2-spanner and **reject** if $x_i > x_j$.

Guarantee: All sorted lists are accepted. ✓

All lists that are ε-far from sorted are rejected with probability ≥ 2/3. ✓

Time: O((log n)/ ε) ✓

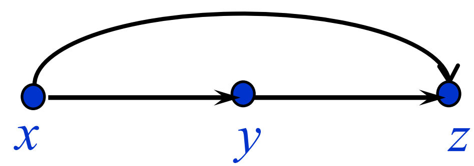
Generalization

Observation:

The same test/analysis apply to any edge-transitive property of a list of numbers that allows extension.

- A property is edge-transitive if
  1) it can be expressed in terms conditions on ordered pairs of numbers

  2) it is transitive: whenever (x, y) and (y, z) satisfy (1), so does (x, z)

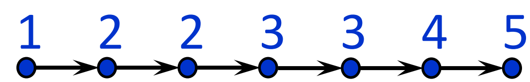


- A property allows extension if
  3) any function that satisfies (1) on a subset of the numbers can be extended to a function with the property

Testing if a Function is Lipschitz [Jha R]

A function $f : D \rightarrow R$ is Lipschitz if it has Lipschitz constant 1:
that is, if for all x,y in *D*,
$distance_R(f(x),f(y)) \leq distance_D(x,y)$.

Consider $f : \{1,\ldots,n\} \rightarrow \mathrm{R}$:



nodes = points in the domain; edges = points at distance 1

node labels = values of the function

The Lipschitz property is *edge-transitive*:

1. a pair (x,y) is *good* if $|f(y)-f(x)| \leq |y-x|$
2. (x,y) and (y,z) are *good*) (x,z) is *good*

It also allows extension for the range R.

Testing if a function $f : \{1,\ldots,n\} \rightarrow \mathrm{R}$ is Lipschitz takes O((log *n*)/²) time.

Properties of a List of n Numbers

- Sorted or ε-far from sorted?

- Lipschitz (does not change too drastically) or ε-far from satisfying the Lipschitz property?

$$O\left(\frac{\log n}{\varepsilon}\right) \text{ time}$$ ✓

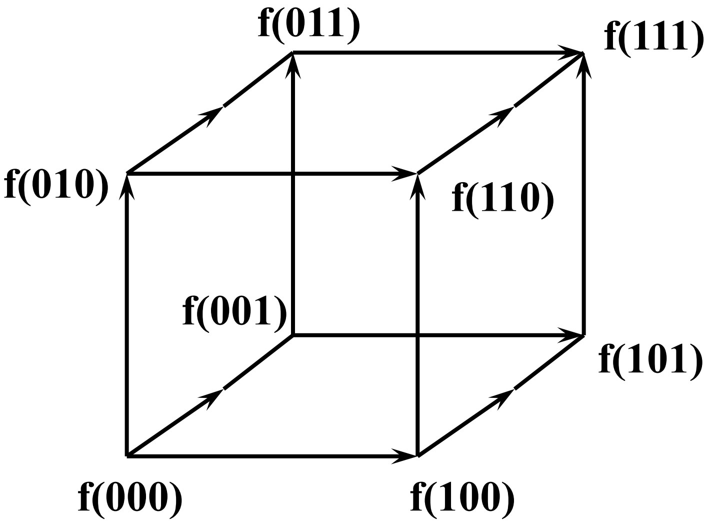
Tight bound: $\Theta\left(\frac{\log(\varepsilon n)}{\varepsilon}\right)$ [Chakrabarty Dixit Jha Seshadhri 15, Belovs 18]

Basic Properties of Functions

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

Graph representation:

n-dimensional hypercube



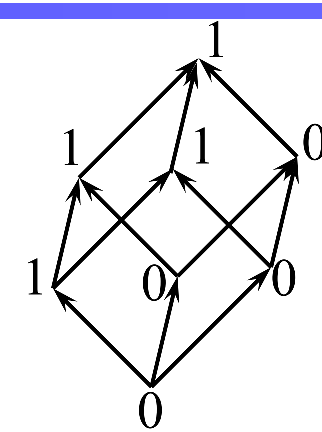
- vertices: bit strings of length n
- edges: (x, y) is an edge if y can be obtained from x by increasing one bit from 0 to 1

x	001001
y	011001

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

Monotonicity of Functions

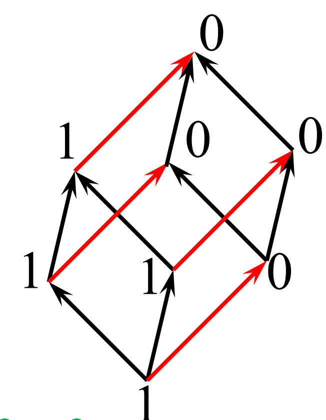
[Goldreich Goldwasser Lehman Ron Samorodnitsky,
Dodis Goldreich Lehman Raskhodnikova Ron Samorodnitsky
Fischer Lehman Newman Raskhodnikova Rubinfeld Samorodnitsky]



monotone

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

- Is f monotone or ε-far from monotone (f has to change on many points to become monontone)?
  - 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
- Advanced techniques: $\Theta(\sqrt{n}/\varepsilon^2)$ for nonadaptive tests, $\Omega(\sqrt[3]{n})$

  [Khot Minzer Safra 15, Chen De Servidio Tang 15, Chen Waingarten Xie 17]

Monotonicity Test [GGLRS, DGLRRS]

Idea: Show that functions that are far from monotone violate many edges.

EdgeTest (f, ε)

1. Pick $2n/\varepsilon$ edges (x, y) uniformly at random from the hypercube.
2. **Reject** if some (x, y) is violated (i.e. $f(x) > f(y)$). Otherwise, **accept**.

Analysis

- If f is monotone, EdgeTest always accepts.
- If f is ε-far from monotone, by Witness Lemma, it suffices to show that $\geq \varepsilon/n$ fraction of edges (i.e., $\frac{\varepsilon}{n} \cdot 2^{n-1} n = \varepsilon 2^{n-1}$ edges) are violated by f.
  - Let $V(f)$ denote the number of edges violated by f.

Contrapositive: If $V(f) < \varepsilon\, 2^{n-1}$,

f can be made monotone by changing $< \varepsilon\, 2^n$ values.

Repair Lemma

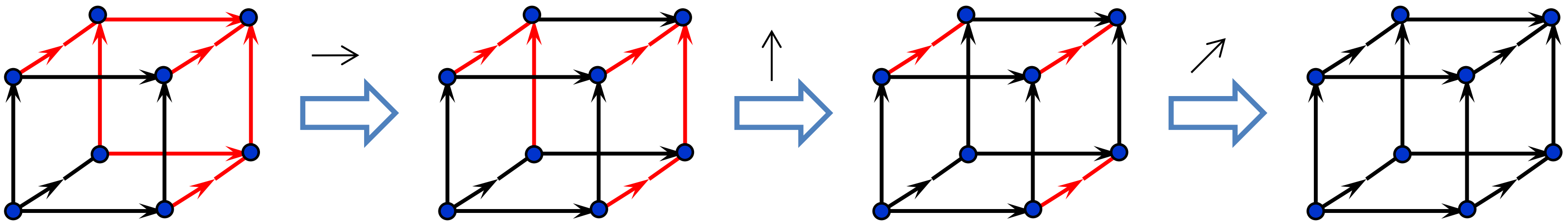
f can be made monotone by changing $\leq 2 \cdot V(f)$ values.

Repair Lemma: Proof Idea

Repair Lemma

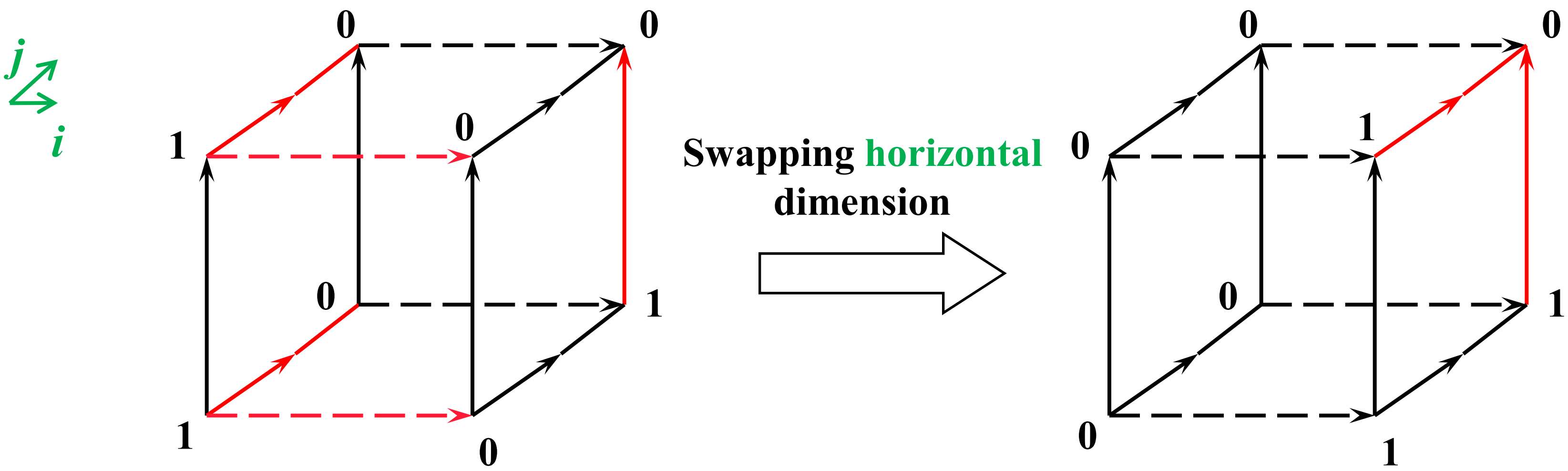
f can be made monotone by changing $\leq 2 \cdot V(f)$ values.

Proof idea: Transform f into a monotone function by repairing edges in one dimension at a time.



Repairing Violated Edges in One Dimension

Swap violated edges $1 \to 0$ **in one dimension to** $0 \to 1$**.**



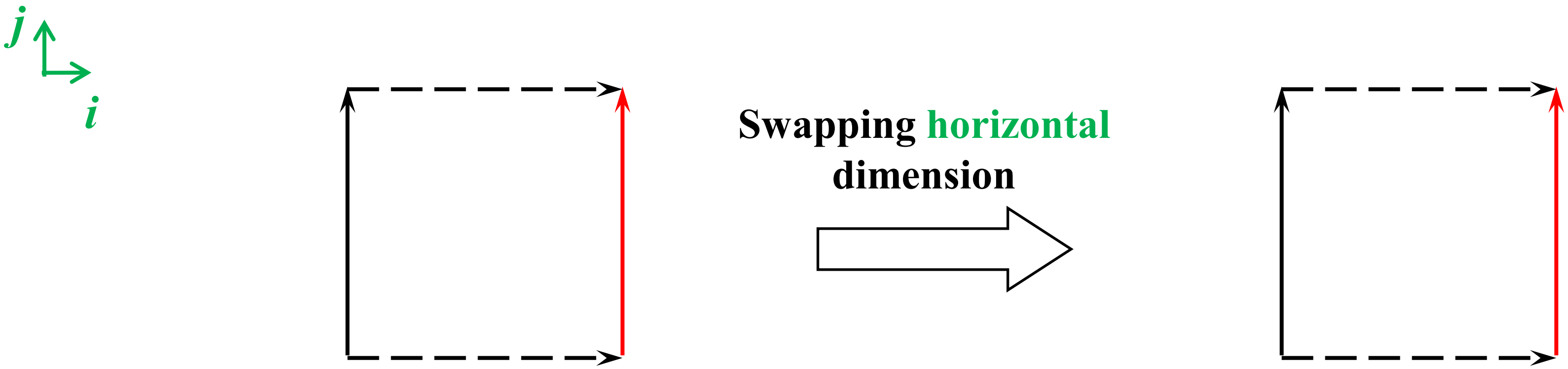
Let V_j = # of violated edges in dimension j

Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$

Enough to prove the claim for squares

Proof of The Claim for Squares

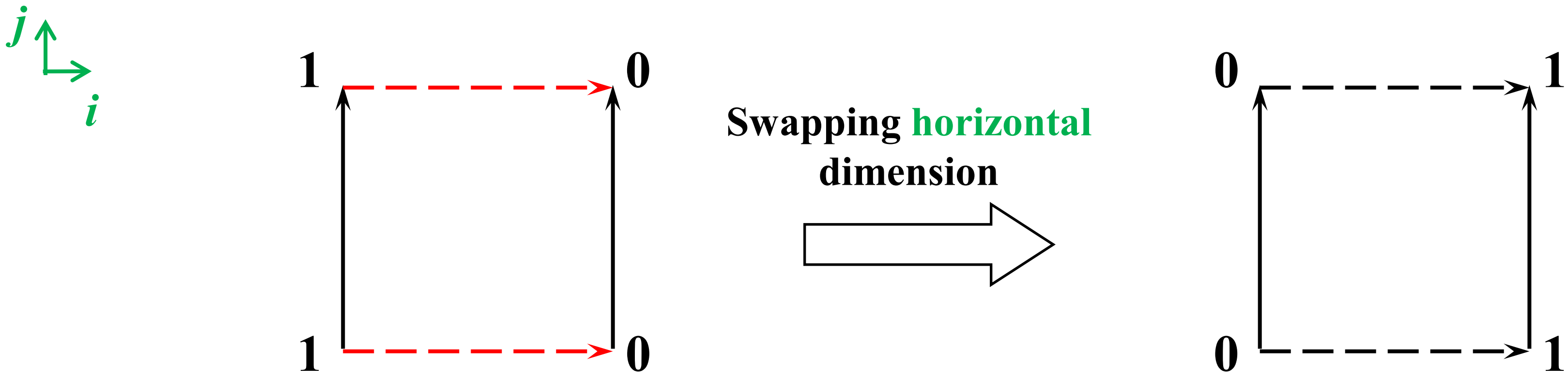
Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$



- If no horizontal edges are violated, no action is taken.

Proof of The Claim for Squares

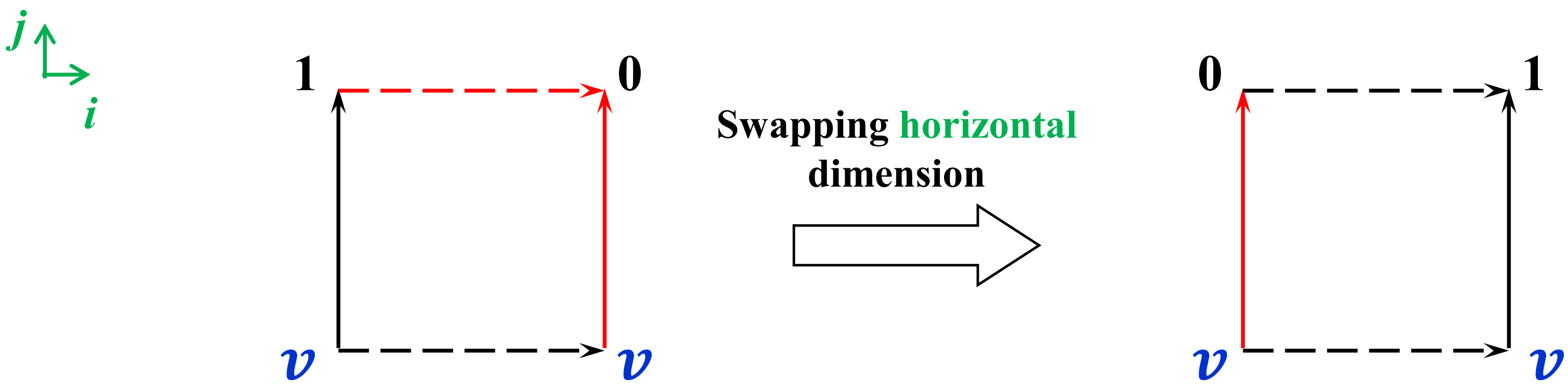
Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$



- If both horizontal edges are violated, both are swapped, so the number of vertical violated edges does not change.

Proof of The Claim for Squares

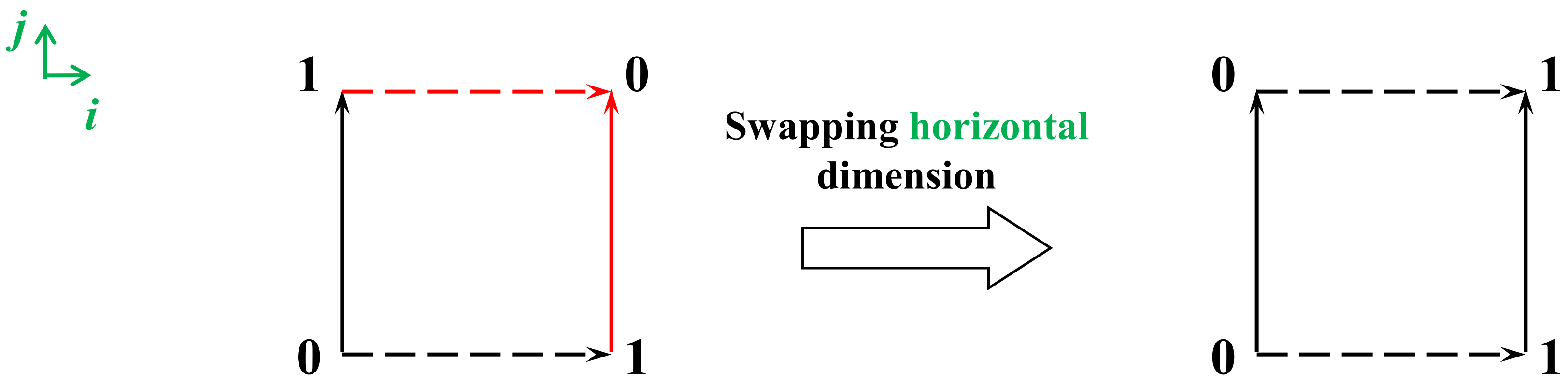
Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$



- Suppose one (say, top) horizontal edge is violated.
- If both bottom vertices have the same label, the vertical edges get swapped.

Proof of The Claim for Squares

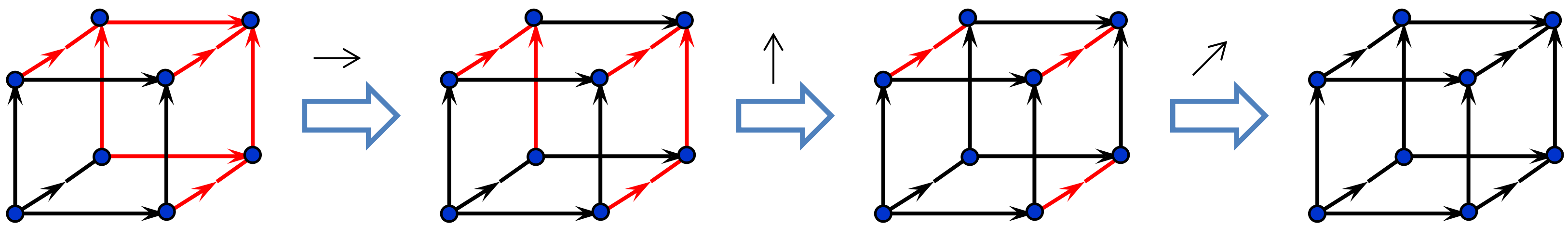
Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$ ✓



- Suppose one (say, top) horizontal edge is violated.
- If both bottom vertices have the same label, the vertical edges get swapped.
- Otherwise, the bottom vertices are labeled 0→1, and the vertical violation is repaired.

Proof of The Claim for Squares

Claim. Swapping in dimension i does not increase V_j for all dimensions $j \neq i$



After we perform swaps in all dimensions:

- f becomes monotone
- # of values changed:

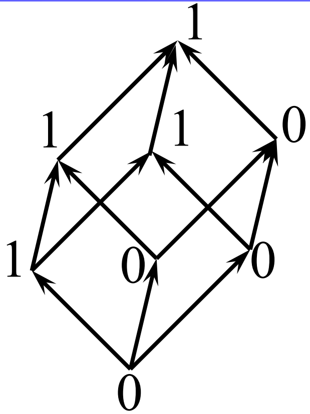
  $2 \cdot V_1 + 2 \cdot$ (# violated edges in dim 2 after swapping dim 1)
  $+ 2 \cdot$ (# violated edges in dim 3 after swapping dim 1 and 2)
  $+ \ldots \leq 2 \cdot V_1 + 2 \cdot V_2 + \cdots 2 \cdot V_n = 2 \cdot V(f)$

Repair Lemma

f can be made monotone by changing $\leq 2 \cdot V(f)$ values.

Improve the bound by a factor of 2.

Testing if a Functions $f : \{0,1\}^n \rightarrow \{0,1\}$ *is monotone*



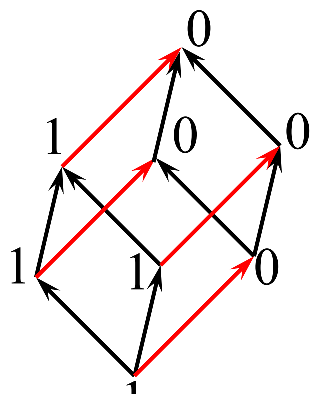
monotone

Monotone or

ε-far from monotone?

O(n/ε) time ✓

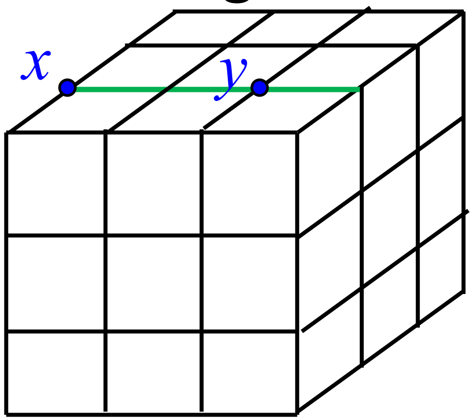
(logarithmic in the size of the input)



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

Testing Properties of High-Dimensional Functions

In polylogarithmic time, we can test a large class of properties of functions $f: \{1, \ldots, n\}^d \to \mathbb{R}$, including:

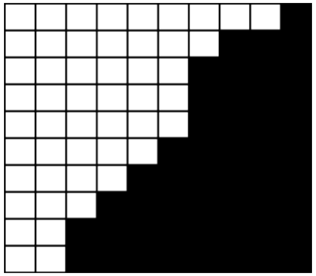


- Lipschitz property [Jha **R**]
- Bounded-derivative properties [Chakrabarty Dixit Jha Seshadhri]
- Unateness [Baleshzar Chakrabarty Pallavoor **R** Seshadhri]

Back to Testing if an Image is a Half-plane

A half-plane or

ε-far from a half-plane?



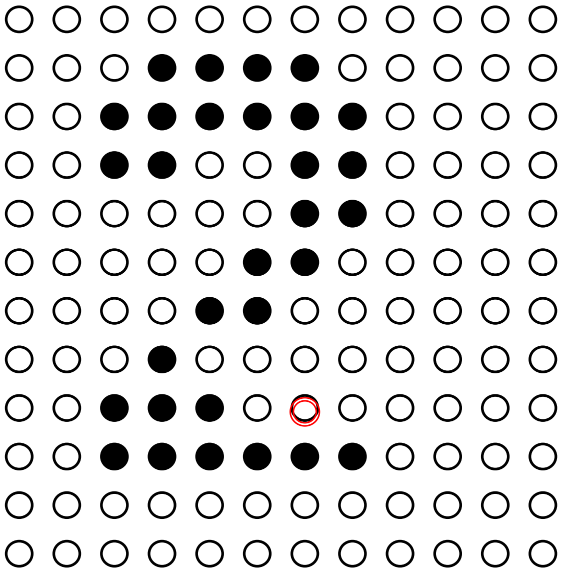
O(1/ε) time [**R** 03]

O(1/ε) time with uniform samples

[Berman Murzabulatov **R** 16]

Pixel Model

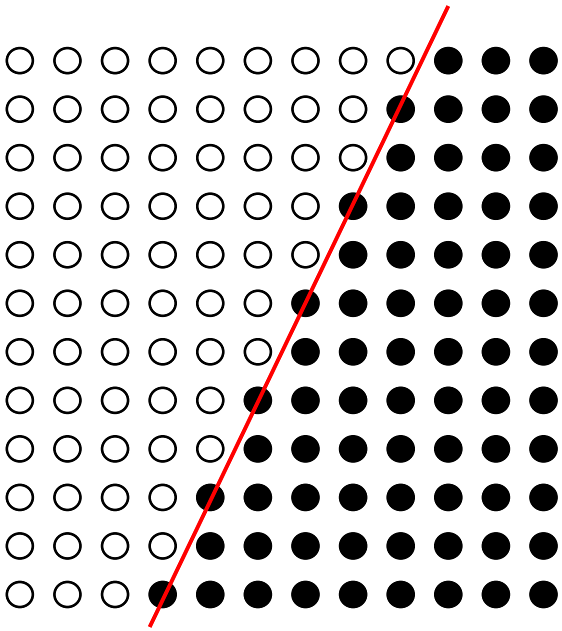
Input: $n \times n$ matrix of pixels
(0/1 values for black-and-white pictures)



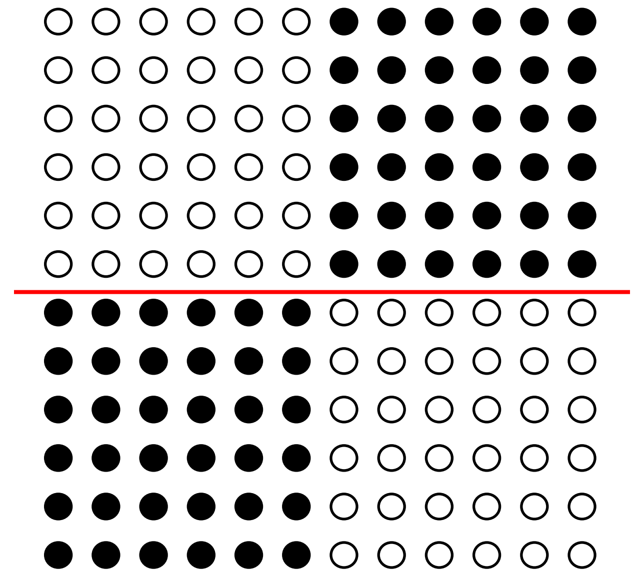
Query: point (i_1, i_2)

Answer: color of (i_1, i_2)

Half-plane Instances



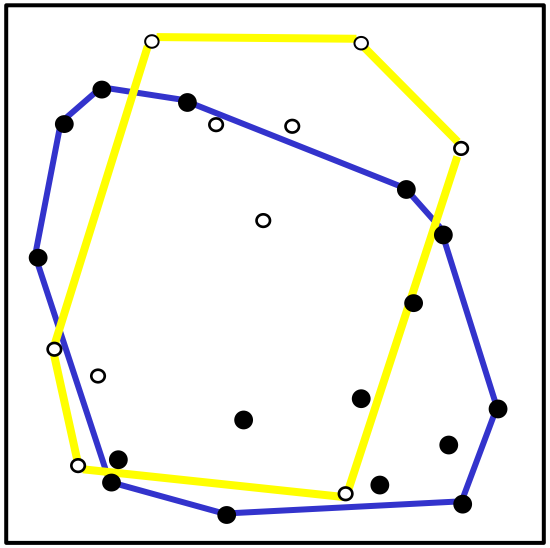
A half-plane



$\frac{1}{4}$-far from a half-plane

Half-Plane Tester

1. Sample $s = \Theta\left(\frac{1}{\varepsilon}\right)$ pixels uniformly and independently.
2. Find convex hull of black samples and convex hull of white samples.
3. If the two hulls intersect, **reject**; otherwise, **accept**.



- The tester always accepts half-plane images.

Correctness Theorem

If an image is ε-far from being a half-plane, it is rejected w.p. $\geq 2/3$.

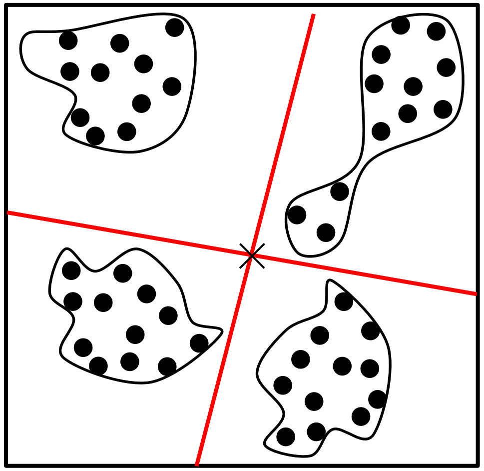
Analysis Idea: Central Points

Some points are likely to end up in the convex hull of black pixels.

- A point does not have to correspond to a pixel.

Definition

A point is **black-central** if it is the intersection of two lines such that each quadrant formed by the lines has $\geq \varepsilon n^2/4$ **black** pixels.



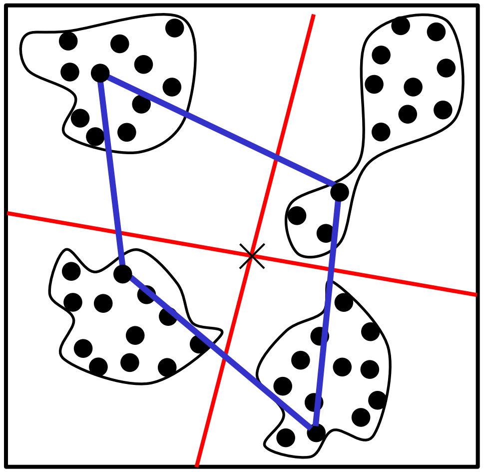
- A white-central point is defined analogously.

Analysis Idea: Central Points

Some points are likely to end up in the convex hull of black pixels.

Definition

A point is **black-central** if it is the intersection of two lines such that each quadrant formed by the lines has $\geq \varepsilon n^2/4$ **black** pixels.



- If we sample a black pixel (``witness'') from each quadrant, then the black-central point is in the convex hull of black pixels.

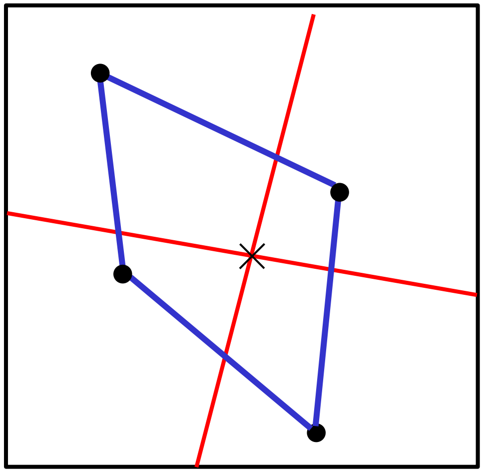
  We say ``we **captured** the black-central point''.

Analysis Idea: Central Points

Some points are likely to end up in the convex hull of black pixels.

Definition

A point is **black-central** if it is the intersection of two lines such that each quadrant formed by the lines has $\geq \varepsilon n^2/4$ **black** pixels.



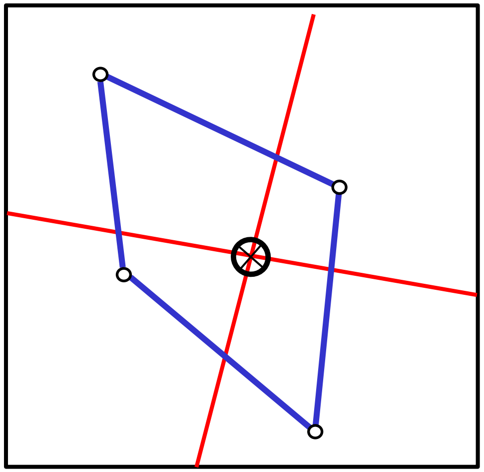
- By Witness Lemma, if we sample $\frac{\ln 100}{\varepsilon/4}$ random pixels, we fail to find a witness from a quadrant w.p. $\leq \frac{1}{100}$.
- By the union bound, we fail to capture a black-central w.p. $\leq \frac{4}{100}$

Analysis Idea: Central Points

Some points are likely to end up in the convex hull of black pixels.

Definition

A point is **black-central** if it is the intersection of two lines such that each quadrant formed by the lines has $\geq \varepsilon n^2/4$ **black** pixels.

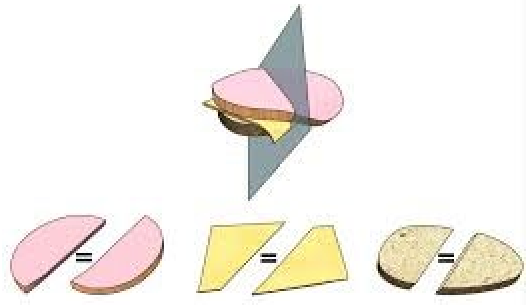


- Analogously, we fail to capture a white-central w.p. $\leq \frac{4}{100}$

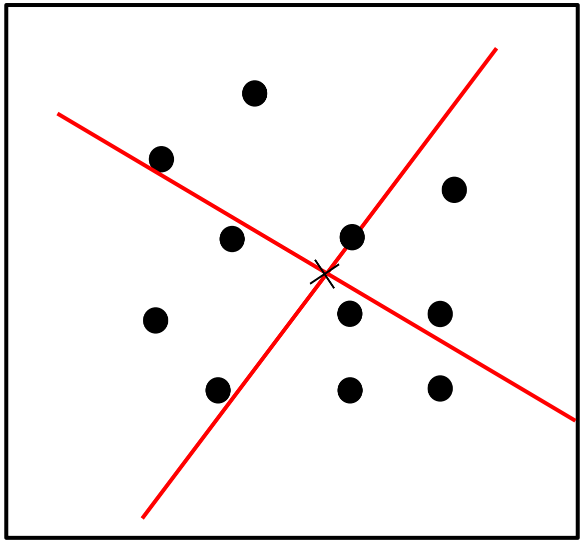
Central Points Exist

The Ham Sandwich Theorem

In n dimensions, any n measurable sets can be simultaneously bisected (w.r.t. their measure) by an $(n-1)$-dimensional hyperplane.



- If an image is ε-far from being a half-plane, it contains at least εn^2 pixels of each color.
- By continuity, there is a line that bisects all pixels of the same color into two sets.
- By the Ham Sandwich Theorem (for $n=2$), there is another line that bisects both sets.

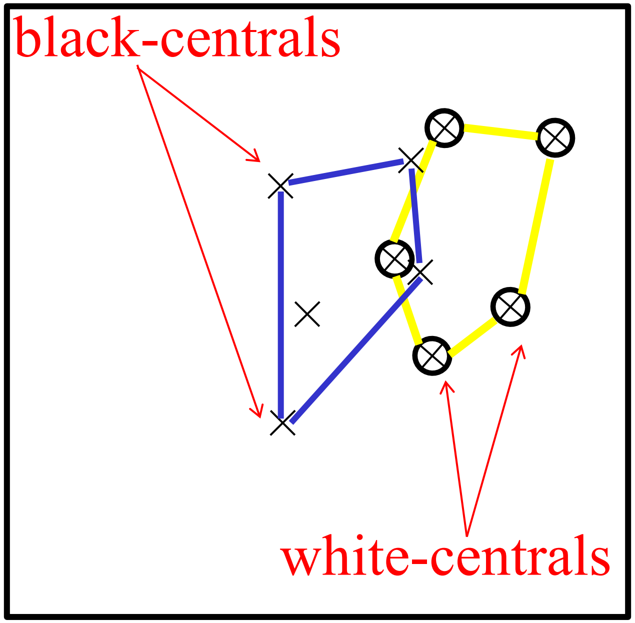


Ham sandwich image credit: *https://apieceofthepi.substack.com/p/ham-sandwiches-and-necklace-splitting*

Hulls of Black- and White-Central Points

Main Lemma

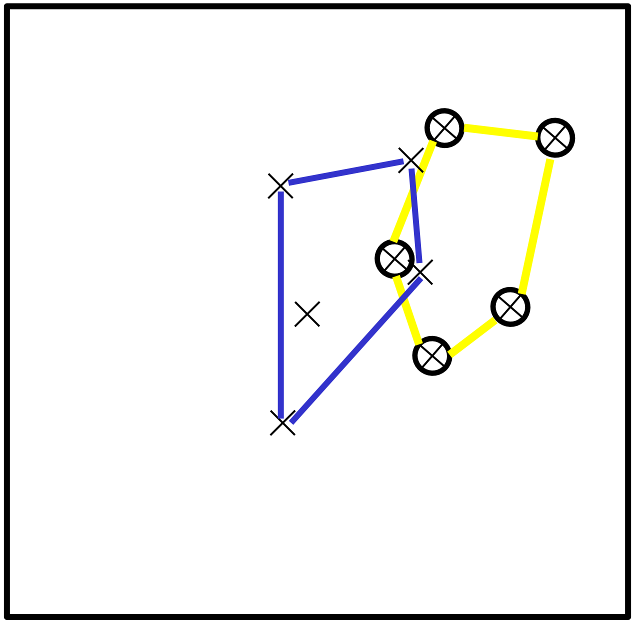
If the image is ε-far from being a half-plane then the convex hull of black-centrals intersects the convex hull of white-centrals.



Hulls of Black- and White-Central Points

Main Lemma

If the image is ε-far from being a half-plane then the convex hull of black-centrals intersects the convex hull of white-centrals.



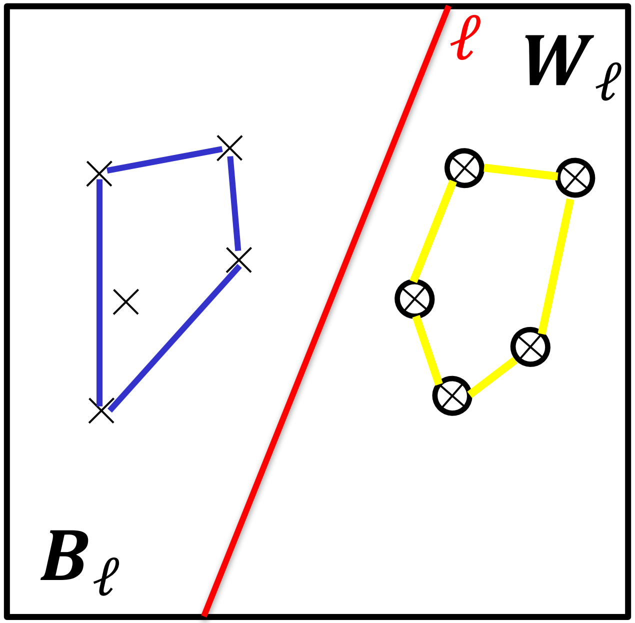
Proof: For the sake of contradiction, assume they do not intersect.

- Then some line ℓ separates white-central and black-central points.

Hulls of Black- and White-Central Points

Main Lemma

If the image is ε-far from being a half-plane then the convex hull of black-centrals intersects the convex hull of white-centrals.



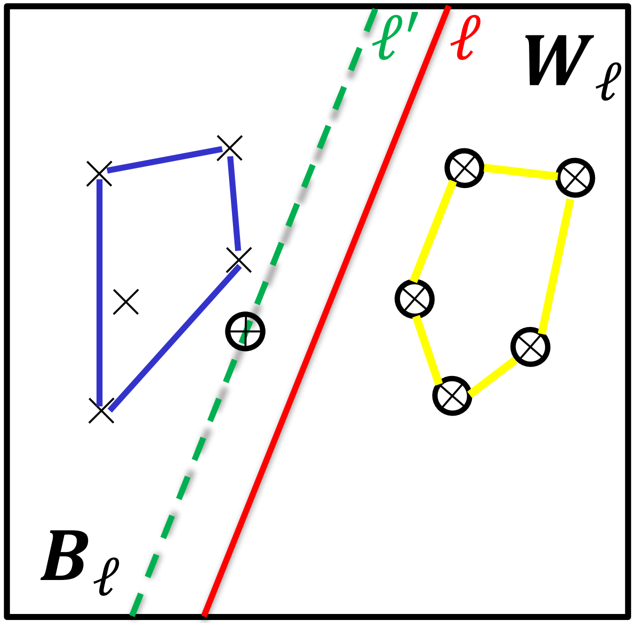
Proof: For the sake of contradiction, assume they do not intersect.

- Then some line ℓ separates white-central and black-central points.
- Let B_ℓ and W_ℓ be the closed half-planes formed by ℓ, with black-central and white-central points, respectively.

Hulls of Black- and White-Central Points

Main Lemma

If the image is ε-far from being a half-plane then the convex hull of black-centrals intersects the convex hull of white-centrals.



Proof: For the sake of contradiction, assume they do not intersect.

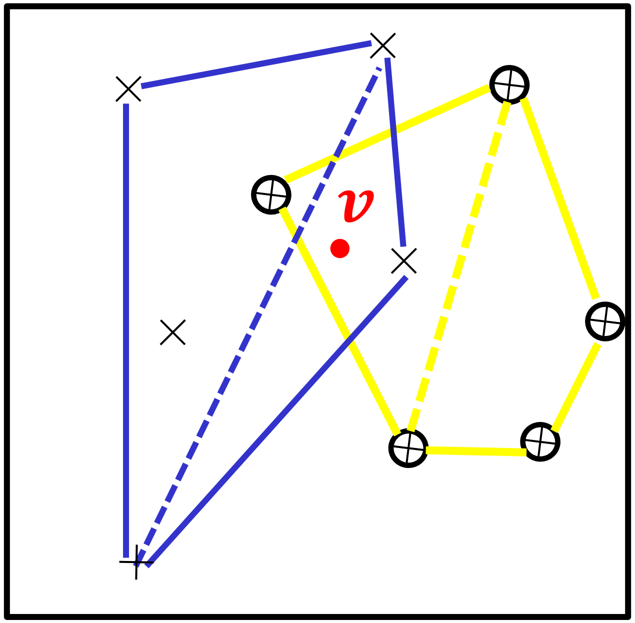
- There are $\geq \frac{\varepsilon n^2}{2}$ black pixels in W_ℓ or white pixels in B_ℓ. W.l.o.g. suppose the latter holds.
- Let ℓ' be the line parallel to ℓ and furthest from ℓ s.t. there $\geq \frac{\varepsilon n^2}{2}$ white pixels in closed half-plane to the left of ℓ'.
- There are $\geq \frac{\varepsilon n^2}{2}$ white pixels in closed half-plane to the right of ℓ'.
- By Ham Sandwich Theorem, there is a white-central point on ℓ'.

Contradiction!

Completing the Analysis

Main Lemma

If the image is ε-far from being a half-plane then the convex hull of black-centrals intersects the convex hull of white-centrals.



- Then some point v is in both hulls.
- Moreover, v is in the convex hull of
  - (at most) 3 black-central points;
  - (at most) 3 white-central points.

- If we capture all 6, then v is in the hull of black samples and in the hull of white samples.
- Recall: we fail to capture a central point w.p. $\leq \frac{4}{100}$
- By union bound, we fail to capture one or more of the 6 central points w.p. $\leq \frac{24}{100} < \frac{1}{3}$.

Summary: Half-plane Testing

- O$(1/\varepsilon)$ uniform samples are sufficient for testing the half-plane property with 1-sided error.
- It is easy to show that $\Omega(1/\varepsilon)$ queries are necessary for even 2-sided error, adaptive testers.

A half-plane or

ε-far from a half-plane?

in O$(1/\varepsilon)$ uniform samples ✓