Microsoft Research

Each year Microsoft Research hosts hundreds of influential speakers from around the world including leading scientists, renowned experts in technology, book authors, and leading academics, and makes videos of these lectures freely available.

2016 © Microsoft Corporation. All rights reserved.
Beyond John’s Ellipsoid

Alexandr Andoni (Columbia)
Assaf Naor (Princeton)
Aleksandar Nikolov (Toronto)
Ilya Razenshteyn (Microsoft)
Erik Waingarten (Columbia)

To be submitted to FOCS
John’s ellipsoid

- [John 1948]: for centrally-symmetric convex $K \subseteq \mathbb{R}^d$, there is an ellipsoid $E$ s.t. $E \subseteq K \subseteq \sqrt{d} \cdot E$
- **Lots** of applications:
  - Convex geometry [Ball 1996]
  - Integer programming [Lenstra 1983]
  - Approximating submodular functions [Goemans, Harvey, Iwata, Mirroknii 2009]
  - Communication complexity [Lovett 2014]
Optimality

- The bound $\sqrt{d}$ is tight for:
  - Hypercube $[-1, 1]^d$
  - Cross-polytope $\{x \in \mathbb{R}^d \mid \|x\|_1 \leq 1\}$
  - ...
Analytic form of the John’s theorem

- Each centrally symmetric $\mathcal{K} \subseteq \mathbb{R}^d$ gives rise to a norm
Analytic form of the John’s theorem

- Each centrally symmetric $\mathcal{K} \subseteq \mathbb{R}^d$ gives rise to a norm
  $$\|x\|_{\mathcal{K}} = \inf\{t > 0 | x/t \in \mathcal{K}\}$$

- (John’s theorem, reformulation) For every norm $X = (\mathbb{R}^d, \|\cdot\|_X)$, there is an invertible linear map $T: \mathbb{R}^d \to \mathbb{R}^d$ such that:
  $$\forall x_1, x_2 \in \mathbb{R}^d: \|Tx_1 - Tx_2\|_{\ell_2} \leq \sqrt{d} \cdot \|x_1 - x_2\|_X$$
  $$\forall h_1, h_2 \in \mathbb{R}^d: \|T^{-1}h_1 - T^{-1}h_2\|_X \leq \|h_1 - h_2\|_{\ell_2}$$

- Relation: $\|x\|_\mathcal{K} = \|Tx\|_{\ell_2}$
Banach–Mazur distance

- For two norms $X = (\mathbb{R}^d, \| \cdot \|_X)$ and $Y = (\mathbb{R}^d, \| \cdot \|_Y)$, and a linear map $T: \mathbb{R}^d \to \mathbb{R}^d$, define the operator norm:
  $$\|T\|_{X \to Y} = \min_{\|x\|_X = 1} \|Tx\|_Y$$

- The Banach–Mazur distance $d_{BM}(X, Y)$ is:
  $$d_{BM}(X, Y) = \min_{T: \mathbb{R}^d \to \mathbb{R}^d} \|T\|_{X \to Y} \cdot \|T^{-1}\|_{Y \to X}$$

- John’s theorem, yet another time:
  $$d_{BM}(X, \ell_2^d) \leq \sqrt{d}$$
Going beyond John’s ellipsoid...
The result of Daher

- [Daher 1993]: for a norm \( X = (\mathbb{R}^d, \| \cdot \|_X) \) and \( 0 < \alpha < 0.1 \),
- There exist spaces \( X' = (\mathbb{R}^d, \| \cdot \|_{X'}) \) with \( d_{BM}(X, X') \lesssim d^\alpha \) and \( H' = (\mathbb{R}^d, \| \cdot \|_{H'}) \) with \( d_{BM}(H', \ell_2^d) \lesssim 1 \)
- And a bijection between unit spheres \( F: S_{X'} \to S_{H'} \) such that:
  \( \forall x_1, x_2 \in S_{X'}: \| F(x_1) - F(x_2) \|_{H'} \lesssim \log d \cdot \| x_1 - x_2 \|_{X'}^\alpha \),
  \( \forall h_1, h_2 \in S_{H'}: \| F^{-1}(h_1) - F^{-1}(h_2) \|_{X'} \lesssim \log d \cdot \| h_1 - h_2 \|_{H'}^\alpha \)
The result of Daher

\[ X \approx_{up \to 0(d^\alpha)} X' \]  
\[ H' \approx_{up \to O(1)} \ell_2^d \]

- [Daher 1993]: for a norm \( X = (\mathbb{R}^d, \| \cdot \|_X) \) and \( 0 < \alpha < 0.1 \),
- There exist spaces \( X' = (\mathbb{R}^d, \| \cdot \|_{X'}) \) with \( d_{BM}(X, X') \leq d^\alpha \) and \( H' = (\mathbb{R}^d, \| \cdot \|_{H'}) \) with \( d_{BM}(H', \ell_2^d) \leq 1 \)
- And a bijection between unit spheres \( F: S_{X'} \rightarrow S_{H'} \) such that:
  \[ \forall x_1, x_2 \in S_{X'}: \| F(x_1) - F(x_2) \|_{H'} \lesssim \log d \cdot \| x_1 - x_2 \|_X^\alpha, \]
  \[ \forall h_1, h_2 \in S_{H'}: \| F^{-1}(h_1) - F^{-1}(h_2) \|_{X'} \lesssim \log d \cdot \| h_1 - h_2 \|_{H'}^\alpha, \]
- Pros: \( \log d \) (+ perturbation by \( d^\alpha \)) instead of \( \sqrt{d} \)
- Cons: only for unit vectors, \( \alpha \)-Hölder instead of Lipschitz, nonlinear
Making the embedding algorithmic

- **The main tool**: complex interpolation between normed spaces [Calderon 1964]
- Given two norms $X = (\mathbb{R}^d, \| \cdot \|_X)$ and $Y = (\mathbb{R}^d, \| \cdot \|_Y)$, build a family $[X, Y]_\theta$ indexed by $0 \leq \theta \leq 1$ such that $[X, Y]_0 = X$, $[X, Y]_1 = Y$ and $[X, Y]_\theta$ nicely depends on $\theta$
- $\|x\|_{[X,Y]_\theta}$ is defined as a minimum of a certain functional on an (infinite-dimensional) space of holomorphic functions
- We show how to compute $\|x\|_{[X,Y]_\theta}$ (approximately) using the ellipsoidal method
Applications: finding nearest neighbors

• Given:
  • Dataset: \( n \) points \( P \) from a metric space \( M = (X, D) \)
  • Approximation \( c > 1 \)
Applications: finding nearest neighbors

- **Given:**
  - Dataset: $n$ points $P$ from a metric space $M = (X, D)$
  - Approximation $c > 1$

- **Query:**
  - A point $q \in X$

- **Goal:**
  - A data point $p \in P$ s.t. $D(q, p) \leq c \cdot \min_{p^* \in P} D(q, p^*)$

- **Parameters:** space, query time
- **An important special case:** $M = (\mathbb{R}^d, \| \cdot \|)$
Efficient data structures

Data structures with:
- Space **polynomial** in $n$ and $d$
- Query time **sub-linear** in $n$ and **polynomial** in $d$
The $\ell_1/\ell_2$ case

- Most ANN data structures are designed for $\ell_1$ or $\ell_2$ distances
- Started with: [Indyk, Motwani 1998] (LSH), [Kushilevitz, Ostrovsky, Rabani 1998]
  - Space $n^{1+\rho+o(1)}$, query time $n^{\rho+o(1)}$, where:
    - $\rho = \frac{1}{2c-1}$ for $\ell_1$; $\rho = \frac{1}{2c^2-1}$ for $\ell_2$
    - 0.962 for $c = 1.01$
    - 0.143 for $c = 2$...
Distances beyond $\ell_1/\ell_2$?

- How does the geometry of a metric space affect the complexity of the ANN problem?
John’s theorem vs. the new result

- For every $d$-dimensional norm with $d = n^{o(1)}$, one can do ANN with:
  - Space $O(n^{1+\varepsilon})$
  - Query time $O(n^\varepsilon)$
  - Approximation $c = O(\sqrt{d/\varepsilon})$

- This work: $c = \exp\left(O_\varepsilon(\sqrt{\log d})\right)$
Overview of the algorithm

Daher’s embedding  [Matousek 1996]  Fixed-point argument

Sparse cuts in embedded graphs

Minimax

“Data-dependent” random space partitions

Standard

ANN data structures
Sparse cuts in embedded graph

• For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?

• For any graph embedded into $X$ with edges of “length” at most 1
  • Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  • Or the graph has an $\varepsilon$-sparse cut
Sparse cuts in embedded graph

- For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?
- For any graph embedded into $X$ with edges of “length” at most 1
  - Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  - Or the graph has an $\varepsilon$-sparse cut
Sparse cuts in embedded graph

• For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?
• For any graph embedded into $X$ with edges of “length” at most 1
  • Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  • Or the graph has an $\varepsilon$-sparse cut
Sparse cuts in embedded graph

- For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?
- For any graph embedded into $X$ with edges of “length” at most 1
  - Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  - Or the graph has an $\varepsilon$-sparse cut
Sparse cuts in embedded graph

- For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?
- For any graph embedded into $X$ with edges of "length" at most 1
  - Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  - Or the graph has an $\varepsilon$-sparse cut
Sparse cuts in embedded graph

- For a given space $X = (\mathbb{R}^d, \| \cdot \|_X)$ what is the smallest $R(\varepsilon) \geq 1$, which makes the following statement true?
- For any graph embedded into $X$ with edges of "length" at most 1
  - Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  - Or the graph has an $\varepsilon$-sparse cut

Ball of radius $R(\varepsilon)$ with $\Omega(n)$ vertices

A cut with conductance at most $\varepsilon$
Sparse cuts in embedded graph

- For a given space $X = (\mathbb{R}^d, \|\cdot\|_X)$ what is the **smallest** $R(\varepsilon) \geq 1$, which makes the following statement true?

- For any graph embedded into $X$ with edges of "length" at most 1
  - Either there is a $X$-ball of radius $R(\varepsilon)$ covering $\Omega(n)$ vertices
  - Or the graph has an $\varepsilon$-sparse cut

- For $\ell_2^d$, $D(\varepsilon) \leq 1/\varepsilon$ (corollary of Cheeger)

- [Naor 2017], in general, $D(\varepsilon) \leq \log d/\varepsilon^2$
Nice cuts

• The argument from [Naor 2017] gives no control over a promised sparse cut! As a result, not so useful for algorithms...

• The bound $D(\epsilon) \leq 1/\epsilon$ for $\ell_2$ comes with a benefit: a sparse cut can be assumed to be a coordinate cut

• For general norms, can achieve $D(\epsilon) \leq \sqrt{d}/\epsilon$ with coordinate cuts using John’s theorem

• This work: $D(\epsilon) \leq \exp\left(O_\epsilon(\sqrt{\log d})\right)$ with coordinate cuts after applying (a radial extension of) the Daher’s map
Conclusions and open problems

- Can approximate a general norm with a Euclidean norm beyond the John’s bound at a cost of weaker guarantees
- Main open problem: **find more applications!**
Nice cuts

- The argument from [Naor 2017] gives no control over a promised sparse cut! As a result, not so useful for algorithms...
- The bound $D(\varepsilon) \leq 1/\varepsilon$ for $\ell_2$ comes with a benefit: a sparse cut can be assumed to be a coordinate cut
- For general norms, can achieve $D(\varepsilon) \leq \sqrt{d}/\varepsilon$ with coordinate cuts using John’s theorem
- This work: $D(\varepsilon) \approx \exp\left(O_{\varepsilon}(\sqrt{\log d})\right)$ with coordinate cuts after applying (a radial extension of) the Daher’s map
Conclusions and open problems

• Can approximate a general norm with a Euclidean norm beyond the John’s bound at a cost of weaker guarantees
• Main open problem: find more applications!

Questions?