Efficient Dimension Reduction

Nir Ailon

Google Research NY

joint with Bernard Chazelle, Edo Liberty

Original Motivation: Nearest Neighbor Searching

- Wanted to improve algorithm by Indyk, Motwani for approx. NN searching in Euclidean space.
- Evidence for possibility to do so came from improvement on algorithm by Kushilevitz, Ostrovsky, Rabani for approx NN searching over GF(2) (using linear algebraic tools).

Random Dimension Reduction

- “Sketching”
- Metric embedding
  - Distance preserving
    - Sets of points, subspaces, manifolds [Clarkson]
  - Volume preserving [Magen, Zouzias]
- Fast approximate linear algebra
  - SVD, linear regression [Muthukrishnan, Mahoney, Drineas, Sarlos], [Woolfe, Liberty, Rokhlin, Tygert]
- Computational aspects:
  - Time
  - Randomness

Question

Question

Find random mapping Φ from R^d to R^k (d >> k)

Such that for every $x \in R^d$, $||x||_2=1$, $0<\varepsilon<1$

With probability $1-\exp\{-k\varepsilon^2\}$

$$||\Phi x||_2 = 1 \pm O(\varepsilon)$$

Typical Usage

If you have n vectors $x_1..x_n \in R^d$:

- Set $k=O(\varepsilon^{-2}\log n)$
- By **union bound**:
  - For all i,j $||\Phi x_i - \Phi x_j|| \approx_\varepsilon ||x_i - x_j||$

Typical Usage

If you have n vectors $x_1..x_n \in R^d$:

Set $k=O(\varepsilon^{-2}\log n)$

By **union bound**:

For all i,j $||\Phi x_i - \Phi x_j|| \approx_\varepsilon ||x_i - x_j||$

"Tight"

Solution: Johnson-Lindenstrauss [JL84]

Φ: projection onto random k dimensional subspace

Also works if Φ is random matrix of independent sub-Gaussian r.v's

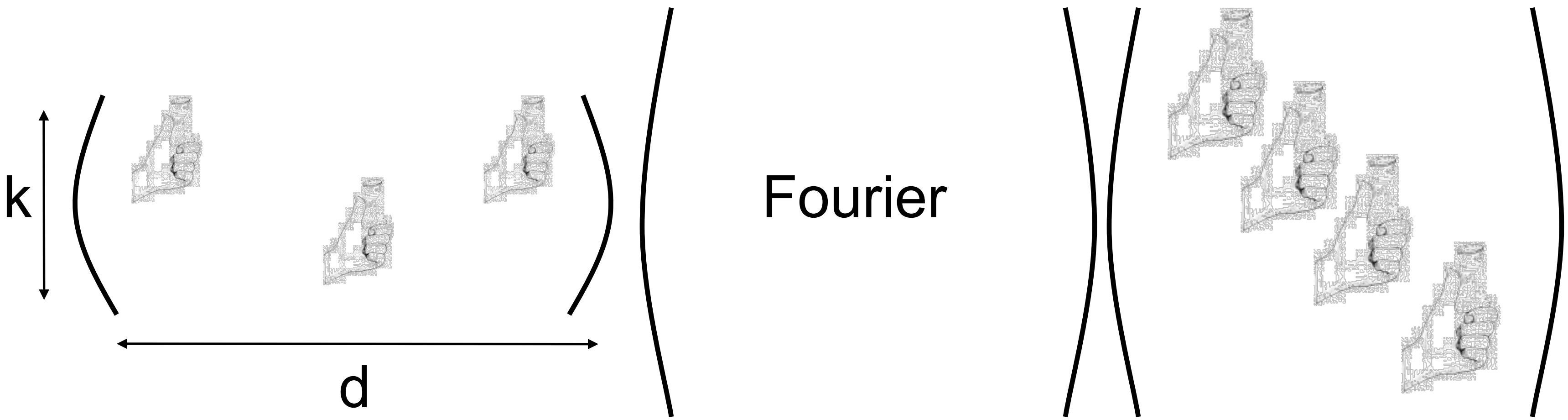
Computer Scientific Question

- Running time of JL $\Omega(kd)$
- Number of random bits $\Omega(kd)$
- Can we do better?

Fast JL

A, Chazelle 2006

Φ = Sparse · Hadamard · Diagonal



Time = $O(k^3 + d\log k)$

Beats JL $\Omega(kd)$ bound for: $k < d^{1/2}$

Gives $O(d\log k)$ for $k < d^{1/3}$

Is $O(d\log d)$ right answer for $k > d^{1/3}$?

Proof of Fast JL

- Using Bernstein concentration bounds:

  With high probability
  $||HDx||_\infty < d^{-1/2} \log^{1/2} d$
  $\Rightarrow$ HDx “looks” like random unit vector
- Sparse sampling sufficient for such

Improvement on FJLT?

- A+Chazelle's FJLT: Success of transformation *conditioned* on success of L_∞-statistic of HDx
- This information alone requires k^3 samples from vector HDx in the worst case [Matousek 2007]

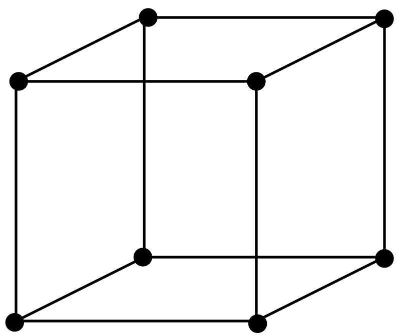
Improvement on FJLT

- [A,Liberty 2007]
- O(d logk) for $k < d^{1/2}$
- O(d) random bits

Improvement on FJLT

- Instead, consider r.v. $||BDx||_2$ *directly* (previously B=SH)
- D random diagonal $D_{ii} \in_R \{\pm 1\}$
- Write as
  $||BDx|| = ||\Sigma D_{ii} x_i B_{\bullet i}||$
  $= ||\Sigma D_{ii} M_{\bullet i}||$ where $M_{\bullet i} = x_i B_{\bullet i}$
- Let $f(D_{11}, \ldots, D_{dd}) = ||\Sigma D_{ii} M_{\bullet i}||$
- Want concentration bound on f…

Talagrand bounds

f:  → R

Many concentration bound results obtained by bounding change in f on Hamming neighbors

Talagrand: If f can be extended to $[-1,1]^d$ convexly then

$Pr[|f(D_1,..,D_d) - \mu| > \varepsilon] < \exp\{ -(\varepsilon / 2||f||_{lip})^2 \}$

Talagrand's Concentration Bound for Rademacher Series

$f = ||\Sigma D_{ii} M_{\cdot i}||_p$ (in this case p=2)

M defined using x (input) and B

$||M||_{2\to 2} \leq ||x||_4 ||B^t||_{2\to 4}$ (Cauchy-Schwartz)

$\Rightarrow$ Find B with small $2\to 4$ norm

Talagrand's Concentration Bound for Rademacher Series

$f = ||\Sigma D_{ii} M_{\cdot i}||_p$ (in this case p=2)

$$\Pr[\,|f-\mu| > \varepsilon\,] = O(\exp\{-\varepsilon^2/4||M||_{2\to p}^{\;2}\})$$

M defined using x (input) and B

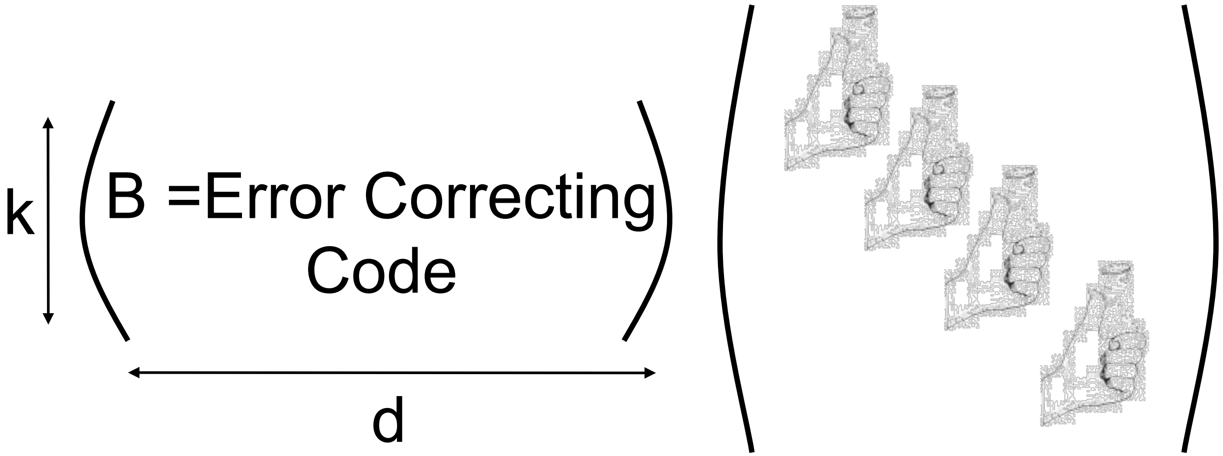
$||M||_{2\to 2} \leq ||x||_4 ||B^t||_{2\to 4}$ (Cauchy-Schwartz)

$\Rightarrow$ Find B with small $2\to 4$ norm

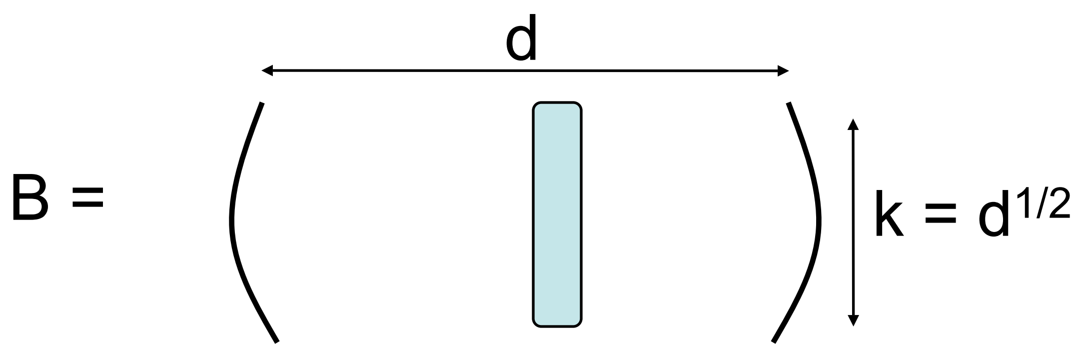
Algorithm ($k=d^{1/2}$)

A, Liberty 2007

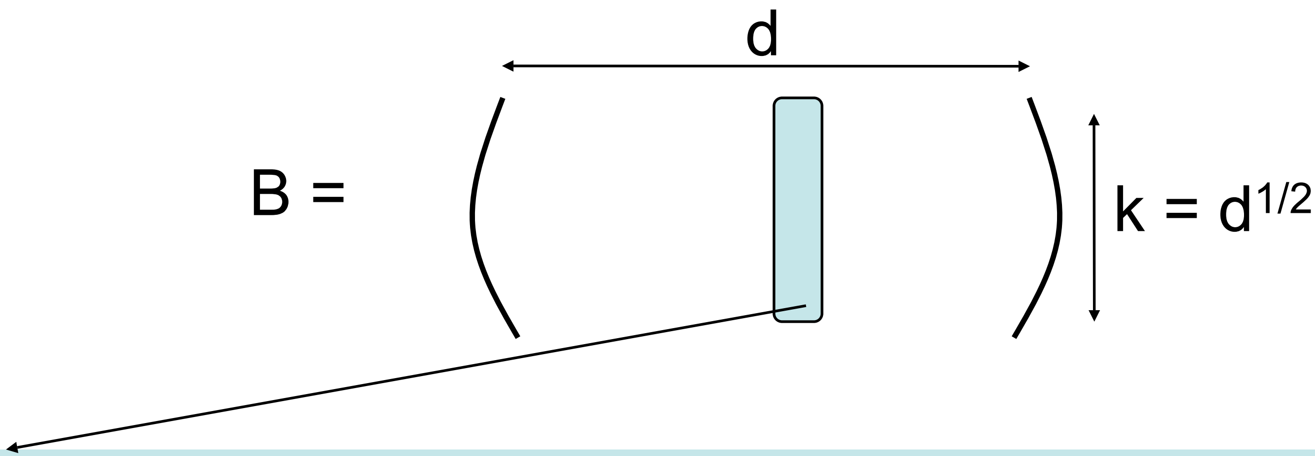
Φ = **B**CH · **D**iagonal · ....



Error Correcting Codes

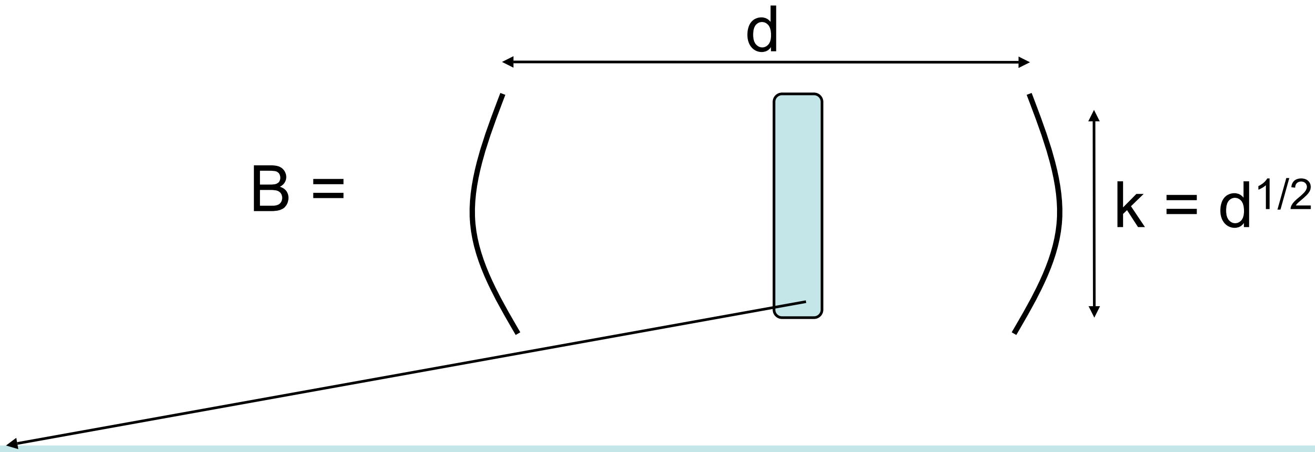


Error Correcting Codes



$\in$ Binary “dual-BCH code of designed distance 4”

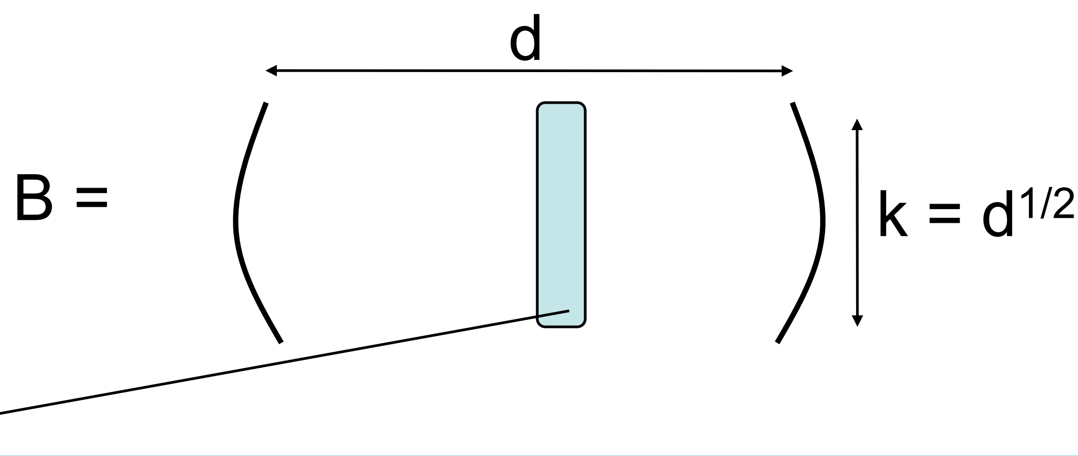
Error Correcting Codes



$\in$ Binary "dual-BCH code of designed distance 4"

0	→	$k^{-1/2}$
1	→	$-k^{-1/2}$

Error Correcting Codes



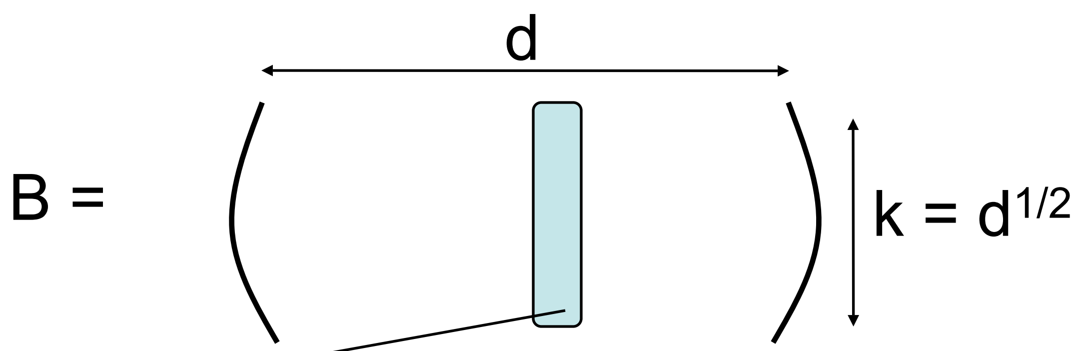
$\in$ Binary "dual-BCH code of designed distance 4"

0 → $k^{-1/2}$

1 → $-k^{-1/2}$

4-wise Independence

Error Correcting Codes



$\in$ Binary "dual-BCH code of designed distance 4"

$0 \rightarrow k^{-1/2}$

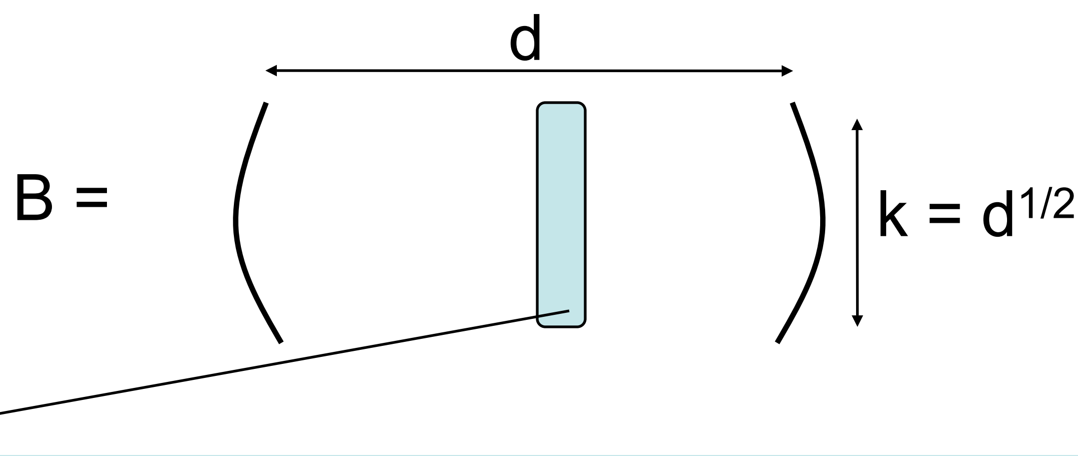
$1 \rightarrow -k^{-1/2}$

4-wise Independence

Columns closed under addition

Row set subset of d x d Hadamard

Error Correcting Codes



$\in$ Binary "dual-BCH code of designed distance 4"

$0 \rightarrow k^{-1/2}$

$1 \rightarrow -k^{-1/2}$

4-wise Independence

Columns closed under addition

Row set subset of d x d Hadamard

Bx computable in time d logd given x

Error Correcting Codes

Error Correcting Codes

Fact (easily from properties of dual BCH):

Error Correcting Codes

Fact (easily from properties of dual BCH):

$||B^t||_{2\to 4} = O(1)$

Error Correcting Codes

Fact (easily from properties of dual BCH):

$||B^t||_{2\to4} = O(1)$

Error Correcting Codes

Fact (easily from properties of dual BCH):

$||B^t||_{2\to4} = O(1)$

Using Talagrand

Error Correcting Codes

Fact (easily from properties of dual BCH):

$||B^t||_{2\to 4} = O(1)$

Using Talagrand

$\Rightarrow \Pr[|\ ||BDx||_2 - 1| > \varepsilon] = O(\exp\{-\varepsilon^2/||x||_4^2\})$

Error Correcting Codes

Fact (easily from properties of dual BCH):

$||B^t||_{2\to4} = O(1)$

Using Talagrand

$\Rightarrow \Pr[|\ ||BDx||_2 - 1| > \varepsilon] = O(\exp\{-\varepsilon^2/||x||_4^2\})$

Must control $||x||_4$ (make it $=k^{-1/2}$)

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\cdot i}||_4$
  $= ||\Sigma D_{ii} M_{\cdot i}||_4$

- by Talagrand:
  $\Pr[\ |Z-\mu_Z| > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}^2\})$

  $\mu_Z = O(d^{-1/4})$ (trivial)
  $||M||_{2\to 4} \le ||H||_{4/3\to 4} ||x||_4$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\bullet i}||_4$
  $= ||\Sigma D_{ii} M_{\bullet i}||_4$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$

$= ||\Sigma D_{ii} x_i H_{\bullet i}||_4$

$= ||\Sigma D_{ii} M_{\bullet i}||_4$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z \quad = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\bullet i}||_4$
  $= ||\Sigma D_{ii} M_{\bullet i}||_4$

- by Talagrand:
  $\Pr[\ |Z-\mu_Z| > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}{}^2\})$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\bullet i}||_4$
  $= ||\Sigma D_{ii} M_{\bullet i}||_4$

- by Talagrand:
  $\Pr[\ |Z-\mu_Z| > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}{}^2\})$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\cdot i}||_4$
  $= ||\Sigma D_{ii} M_{\cdot i}||_4$

- by Talagrand:
  $\Pr[\ |Z-\mu_Z| > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}^2\})$

  $\mu_Z = O(d^{-1/4})$ (trivial)

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

- Can multiply x by orthogonal matrix
- Use matrix HD

(HD used in [AC06] to control $||HDx||_\infty$)

- $Z = ||HDx||_4$
  $= ||\Sigma D_{ii} x_i H_{\cdot i}||_4$
  $= ||\Sigma D_{ii} M_{\cdot i}||_4$

- by Talagrand:
  $\Pr[\ |Z-\mu_Z| > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}^2\})$

  $\mu_Z = O(d^{-1/4})$ (trivial)
  $||M||_{2\to 4} \le ||H||_{4/3\to 4} ||x||_4$

Controlling $||x||_4^2$

how to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

Controlling $||x||_4^2$

how to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

$Z = ||\Sigma D_{ii} M_{\bullet i}||_4$ $\quad M_{\bullet i} = x_i H_{\bullet i}$

Controlling $||x||_4^2$

how to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

$Z = ||\Sigma D_{ii} M_{\bullet i}||_4$ $\quad M_{\bullet i} = x_i H_{\bullet i}$

$\Pr[\ |Z - \mu_Z\ | > t\ ] = O(\exp\{-t^2/4||M||_{2\to4}^2\})$ (Talagrand)

$\mu_Z = O(d^{-1/4})$ (trivial)

$||M||_{2\to4} \leq ||H||_{4/3\to4} ||x||_4$ (Cauchy Schwartz)

Controlling $||x||_4^2$

how to get $||x||_4^2 = O(k^{-1}=d^{-1/2})$?

$Z = ||\Sigma D_{ii} M_{\bullet i}||_4 \qquad M_{\bullet i} = x_i H_{\bullet i}$

$\Pr[\ |Z - \mu_Z\ | > t\ ] = O(\exp\{-t^2/4||M||_{2\to 4}^2\})$ (Talagrand)

$\mu_Z = O(d^{-1/4})$ (trivial)

$||M||_{2\to 4} \leq ||H||_{4/3\to 4} ||x||_4$ (Cauchy Schwartz)

$||H||_{4/3\to 4} \leq d^{-1/4}$ (Hausdorff-Young)

Controlling $||x||_4^2$

how to get $||x||_4^2 = O(k^{-1}{=}d^{-1/2})$?

$Z = ||\Sigma D_{ii} M_{\bullet i}||_4 \qquad M_{\bullet i} = x_i H_{\bullet i}$

$\Pr[\,|Z - \mu_Z\,| > t\,] = O(\exp\{-t^2/4||M||_{2\to4}^2\})$ (Talagrand)

$\mu_Z = O(d^{-1/4})$ (trivial)

$||M||_{2\to4} \le ||H||_{4/3\to4} ||x||_4$ (Cauchy Schwartz)

$||H||_{4/3\to4} \le d^{-1/4}$ (Hausdorff-Young)

$\Rightarrow ||M||_{2\to4} \le d^{-1/4} ||x||_4$

$\Rightarrow \Pr[\,||HDx||_4 > d^{-1/4} + t\,] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Need some slack $k=d^{1/2-\delta}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Need some slack $k=d^{1/2-\delta}$

Max allowed error probability: $\exp\{-k\}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Need some slack $k=d^{1/2-\delta}$

Max allowed error probability: $\exp\{-k\}$

$k = t^2/d^{-1/2}||x||_4^2$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

Need some slack $k=d^{1/2-\delta}$

Max allowed error probability: $\exp\{-k\}$

$k = t^2/d^{-1/2}||x||_4^2$

$\Rightarrow t = k^{1/2}d^{-1/4}||x||_4 = ||x||_4 d^{-\delta/2}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

Second round:

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

Second round:

$$||HD''HD'x||_4 < d^{-1/4} + d^{-1/4-\delta/2} + ||x||_4 \, d^{-\delta}$$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

Second round:

$$||HD''HD'x||_4 < d^{-1/4} + d^{-1/4-\delta/2} + ||x||_4 \, d^{-\delta}$$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$

Second round:

$||HD''HD'x||_4 < d^{-1/4} + d^{-1/4-\delta/2} + ||x||_4 \, d^{-\delta}$

$r=O(1/\delta)$'th round:

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

Second round:

$$||HD''HD'x||_4 < d^{-1/4} + d^{-1/4-\delta/2} + ||x||_4 \, d^{-\delta}$$

$r=O(1/\delta)$'th round:

$$||HD^{(r)} \cdots HD'x||_4 < O(r) d^{-1/4}$$

Controlling $||x||_4^2$

How to get $||x||_4^2 = O(d^{-1/2})$?

First round:

$$||HDx||_4 < d^{-1/4} + ||x||_4 \, d^{-\delta/2}$$

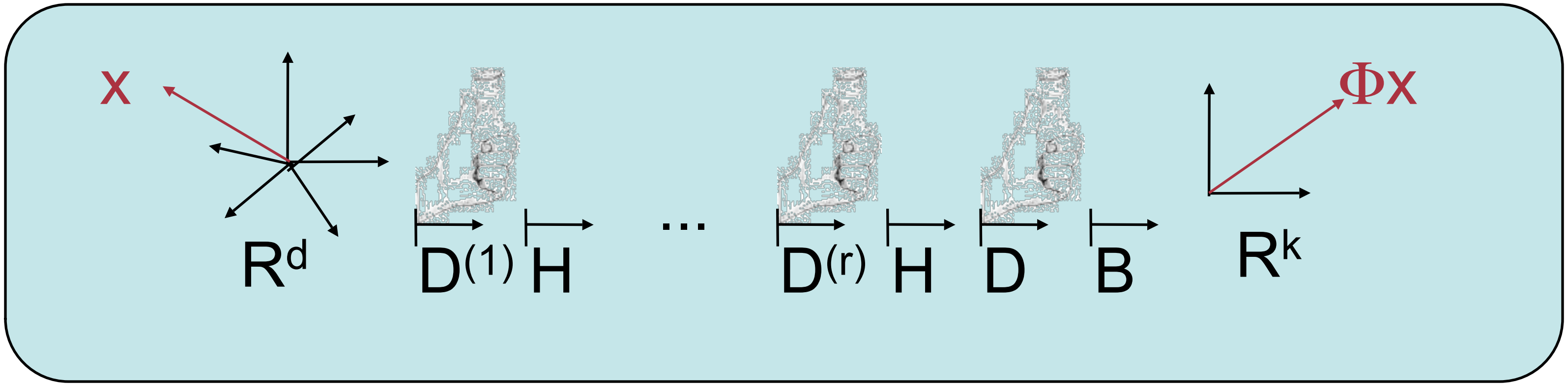
Second round:

$$||HD''HD'x||_4 < d^{-1/4} + d^{-1/4-\delta/2} + ||x||_4 \, d^{-\delta}$$

$r=O(1/\delta)$'th round:

$$||HD^{(r)} \cdots HD'x||_4 < O(r)d^{-1/4}$$

Algorithm for $k=d^{1/2-\delta}$



Running time O(d logd)
Randomness O(d)

Connection to Compressed Sensing

- Random Gaussian matrices, random sub-Fourier matrices used in Compressed Sensing [CT 06], [CRT 06], [R 08], [RV 08]

Open Problems

- Go beyond $k=d^{1/2}$
  Conjecture: can do O(d log d) for $k=d^{1-\delta}$
- Prove that JL onto $k=d^{1/3}$ with distortion $\varepsilon=1/4$ requires $\Omega(d\log(d))$ time
- This would establish similar lower bound for Fourier Transform