Fast Dimension Reduction

MMDS 2008

Nir Ailon

Google Research NY

□□

- "Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes" (with Edo Liberty)

- The Fast Johnson Lindenstrauss Transform (with Bernard Chazelle)

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).
- If we were to do the same for Euclidean space, it was evident that improving ***run time*** of ***Johnson-Lindenstrauss*** was key.

Later Motivation

- Provide more elegant proof, use modern techniques.
- Improvement obtained as bonus.
- Exciting use of
  - Talagrand concentration bounds
  - Error correcting codes

Random Dimension Reduction

- Sketching [Woodruff, Jayram, Li]
- (Existential) 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)
- Computational aspects:
  - Time [A, Chazelle, Liberty] + {Sketching Community} + {Fast Approximate Linear Algebra Community}
  - randomness {Functional Analysis community}

Theoretical Challenge

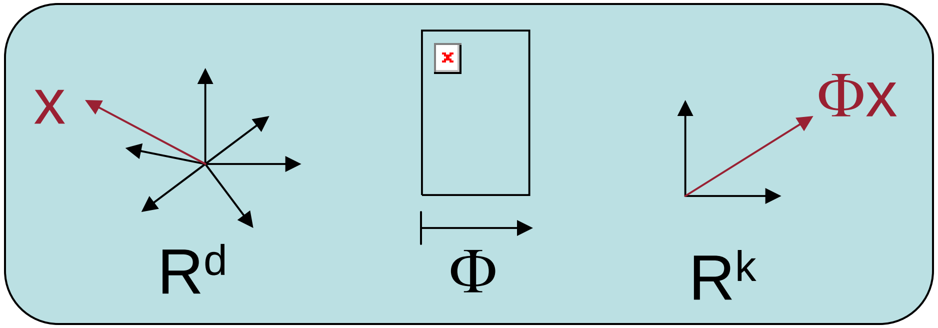
Find random projection Φ from R^d to R^k

(d big, k small)

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

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

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



Usage

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

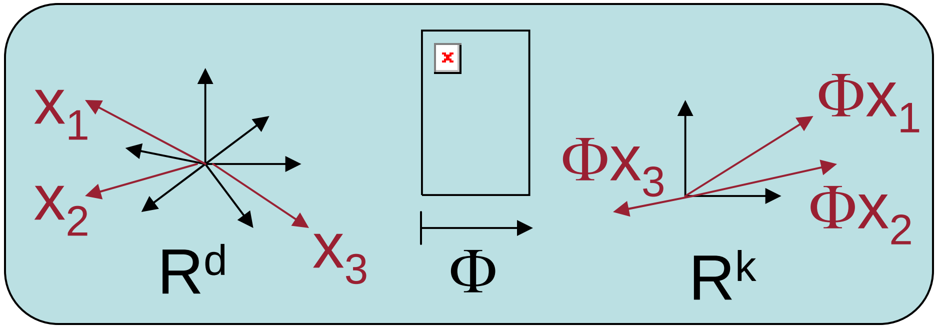
set $k=O(\varepsilon^{-2}\log n)$

by union bound:

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

low-distortion metric embedding

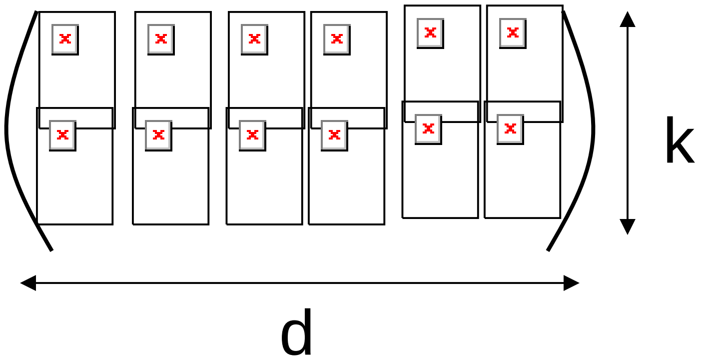
"tight"



Solution:
Johnson-Lindenstrauss (JL)

"dense random matrix"

$\Phi =$



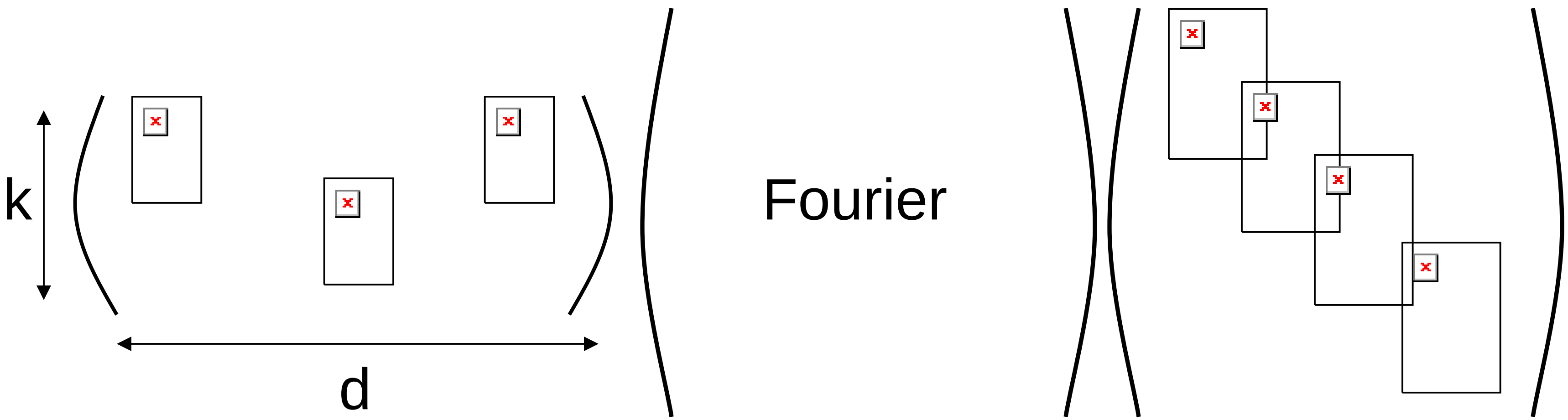
So what's the problem?

- running time $\Omega(kd)$
- number of random bits $\Omega(kd)$
- can we do better?

Fast JL

A, Chazelle 2006

Φ = Sparse · Hadamard · Diagonal



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

beats JL $\Omega(kd)$ bound for: $\log d < k < d^{1/3}$

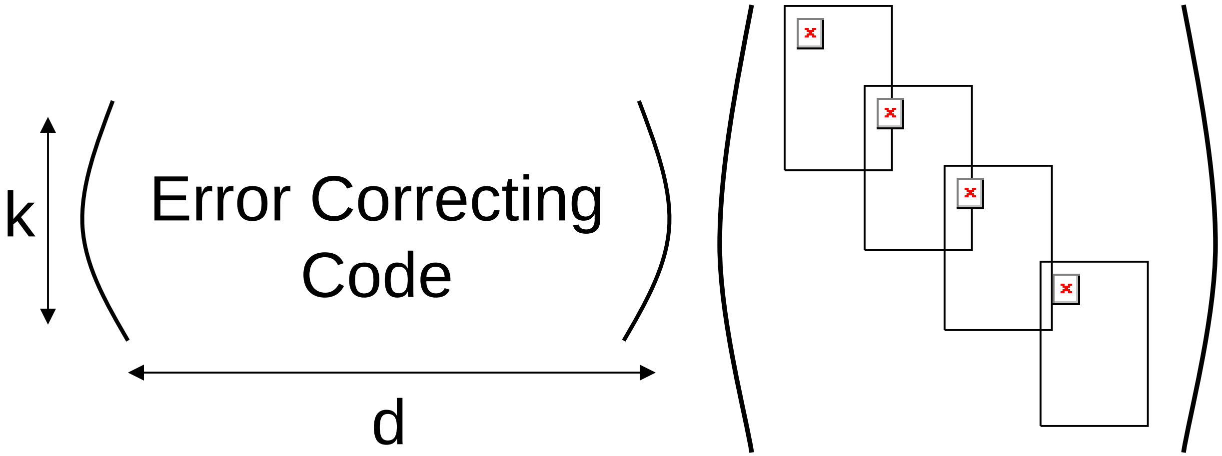
Improvement on FJLT

- O(d logk) for $k < d^{1/2}$
- beats JL up to $k < d^{1/2}$
- O(d) random bits

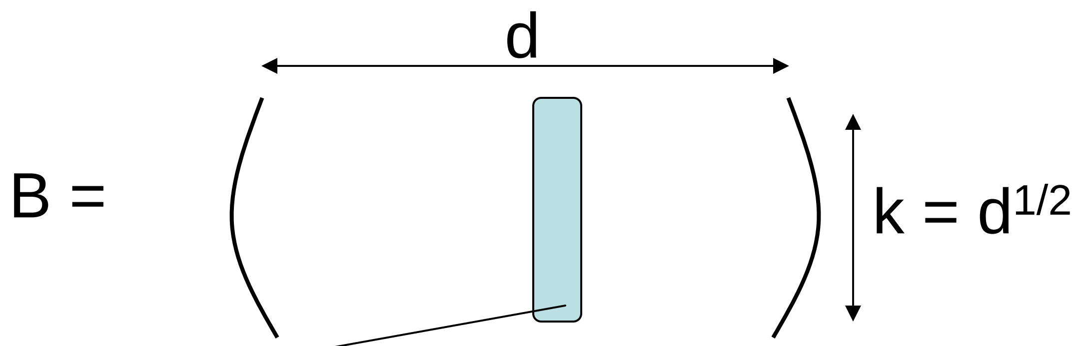
Algorithm ($k=d^{1/2}$)

A, Liberty 2007

Φ = BCH · Diagonal · ....



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

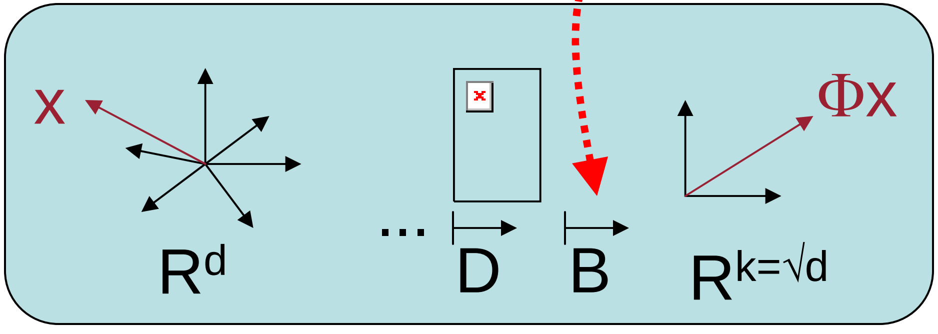
Error Correcting Codes

Fact (easily from properties of dual BCH):

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

for $y^t \in R^k$ with $||y||_2 = 1$:

$||yB||_4 = O(1)$

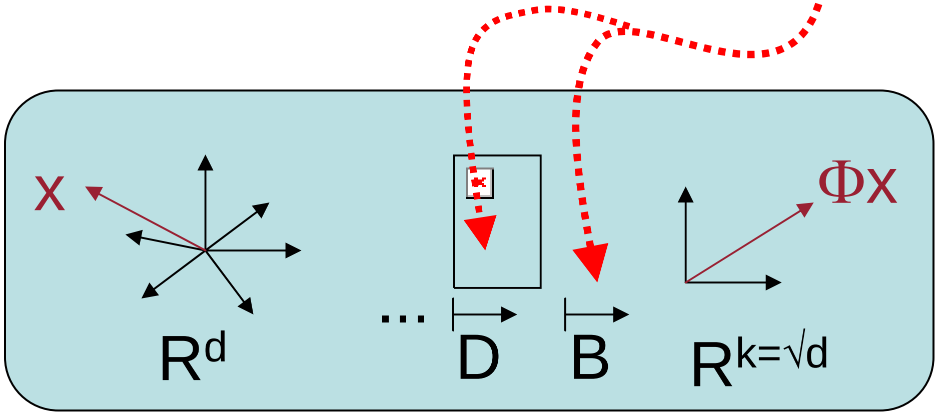


Rademacher Series on Error Correcting Codes

look at r.v. $BDx \in (R^k, l_2)$

$BDx = \sum D_{ii} x_i B_{\bullet i}$ $\quad D_{ii} \in_R \{\pm 1\} \quad i=1...d$

$= \sum D_{ii} M_{\bullet i}$



Talagrand's Concentration Bound for Rademacher Series

$Z = ||\sum D_{ii} M_{\cdot i}||_p$ (in our case p=2)

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

Rademacher Series on Error Correcting Codes

look at r.v. $BDx \in (R^k, l_2)$

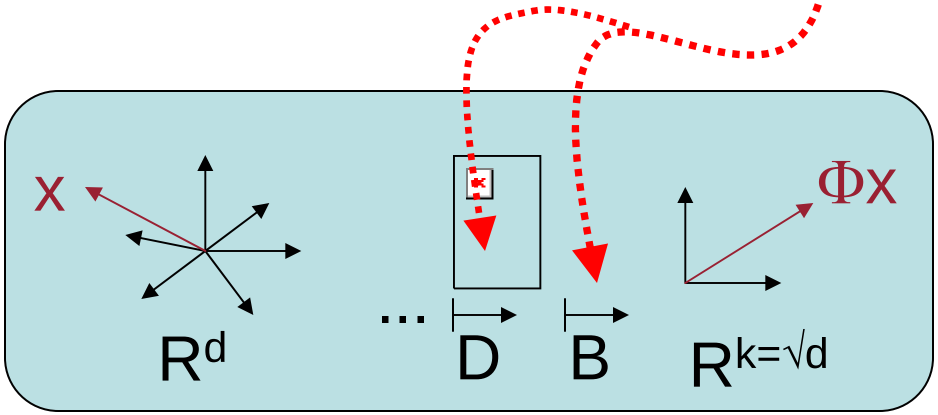
$Z = ||BDx||_2 = ||\Sigma D_{ii} M_{\bullet i}||_2$

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

by ECC properties:

$||M||_{2\to 2} \le ||x||_4 O(1)$ └

trivial: $EZ = ||x||_2 = 1$ □



Rademacher Series on Error Correcting Codes

look at r.v. $BDx \in (R^k, l_2)$

$Z = ||BDx||_2 = ||\Sigma D_{ii} M_{\bullet i}||_2$

$||M||_{2\to 2} = O(||x||_4)$ □

$EZ = 1$

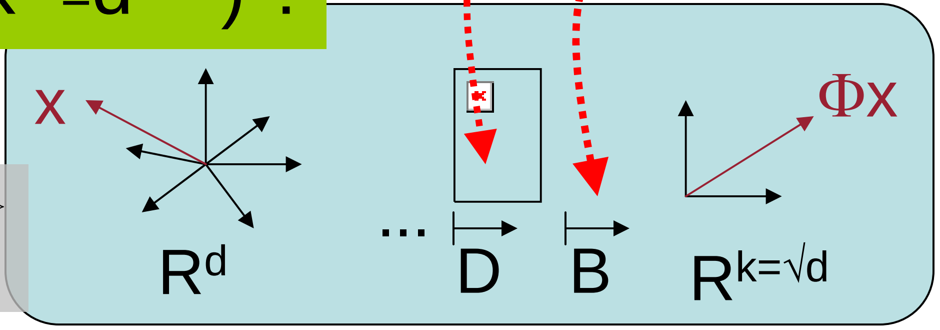
$\Pr[\ |Z-EZ| > \varepsilon\ ] = O(\exp\{-\varepsilon^2/4||M||_{2\to 2}^2\})$

□□

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

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

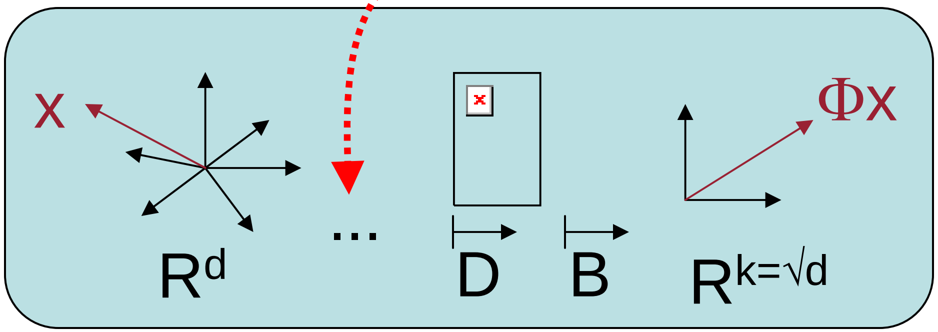
Challenge: w. prob. $\exp\{-k\varepsilon^2\}$ deviation of more than ε



Controlling $||x||_4^2$

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

- if you think about it for a second...
- "random" x has $||x||_4^2=O(d^{-1/2})$
- but "random" x easy to reduce:
  just output first k dimensions
- are we asking for too much?
- no: truly random x has strong bound on $||x||_p$ for all p>2



Controlling $||x||_4^2$

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

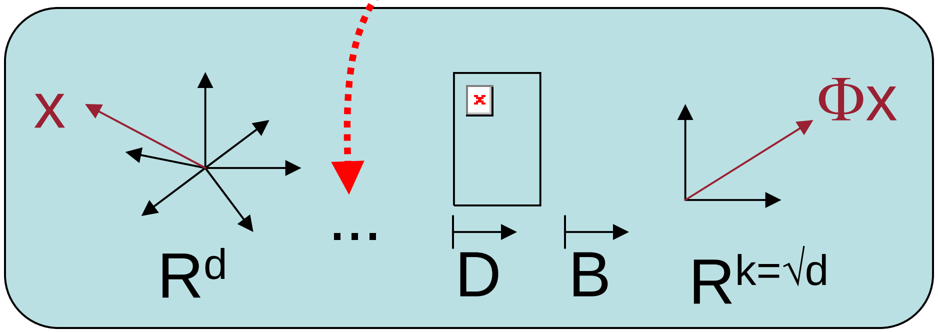
- can multiply x by orthogonal matrix
- try matrix HD

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

  $EZ = O(d^{-1/4})$ (trivial)
  $||M||_{2\to4} \le ||H||_{4/3\to4} ||x||_4$
  (Cauchy Schwartz)



Controlling $||x||_4^2$

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

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

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

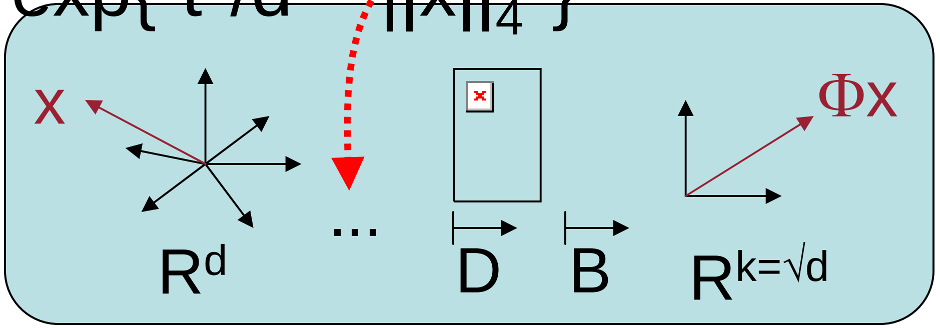
$EZ = O(d^{-1/4})$ (trivial)

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

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

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

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



Controlling $||x||_4^2$

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

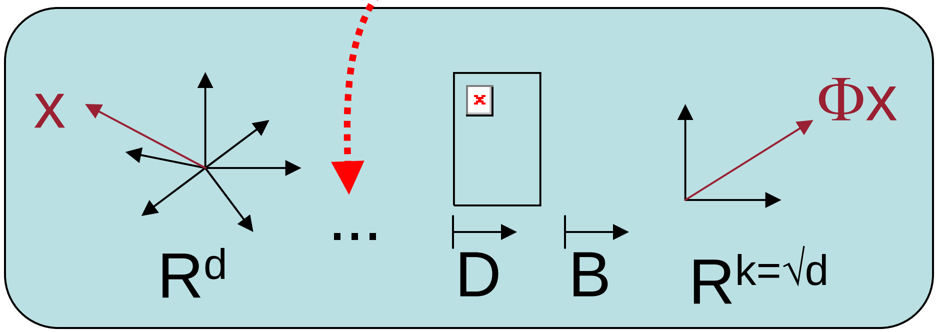
$\Pr[\ ||HDx||_4 > d^{-1/4} + t\ ] = \exp\{-t^2/d^{-1/2}||x||_4^2\}$

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

max error probability for challenge: $\exp\{-k\}$

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

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



Controlling $||x||_4^2$

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

first round:

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

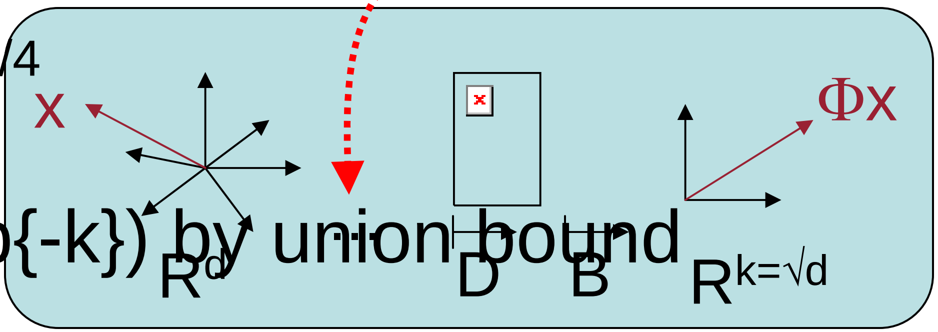
second round:

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

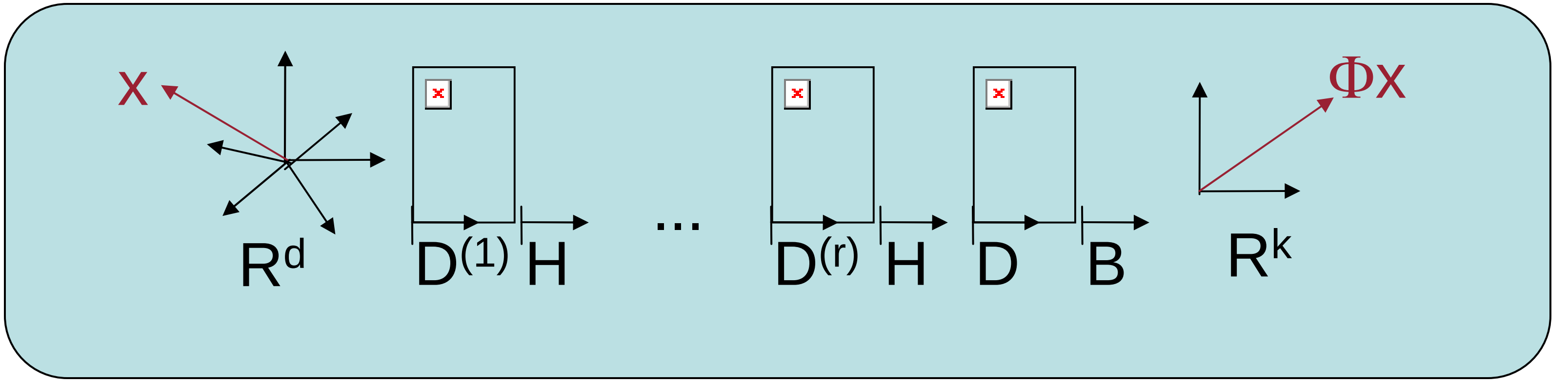
$r=O(1/\delta)$'th round:

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

... with probability 1-O(exp{-k}) by union bound



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



running time O(d logd)
randomness O(d)

Open Problems

- Go beyond $k=d^{1/2}$
  Conjecture: can do $O(d \log d)$ for $k=d^{1-\delta}$
- Approximate linear l_2-regression
  minimize $||Ax-b||_2$ given A,b (overdetermined)
  - State of the art for *general inputs*:
    Õ(linear time) for #{variables} < $\#\{\text{equations}\}^{1/3}$
  - Conjecture: can do Õ(linear time) "always"
- What is the best you can do in linear time?
  [A, Liberty, Singer 08]

Open Problem
Worthy of Own Slide

- Prove that JL onto $k=d^{1/3}$ (say) with distortion $\varepsilon=1/4$ (say) requires $\Omega(d\log(d))$ time
- This would establish similar lower bound for FFT
  - Long standing dormant open problem