Explicit Constructions in High-Dimensional Geometry

Piotr Indyk
MIT

“High-level Picture”

Compressed Sensing

- Random Projections
- L1 minimization
- (Uniform) UP
- …

Data Stream / Sublinear Algorithms

- (Pseudo)random Projections
- Isolation/Group Testing
- …

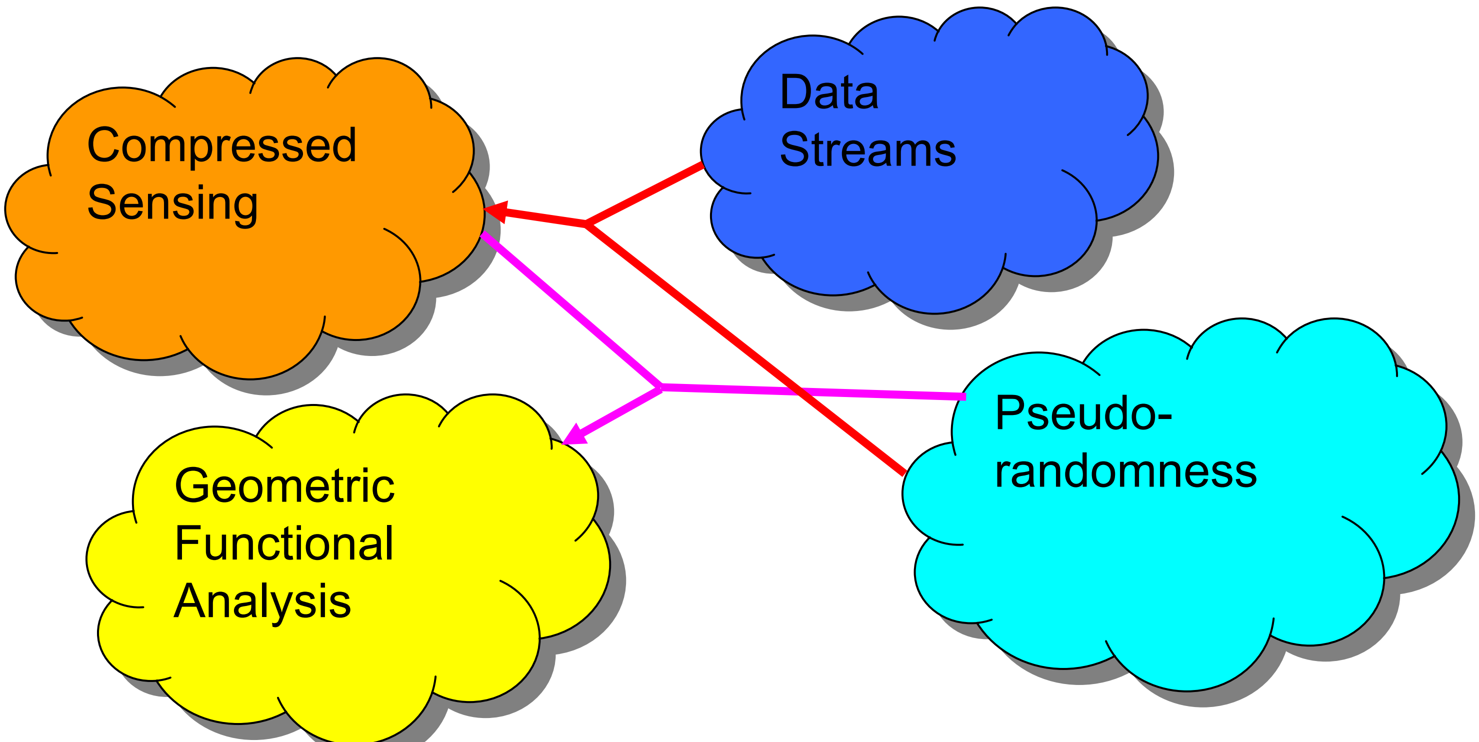
Geometric Functional Analysis (Approximation Theory)

- Concentration of Measure
- Low distortion embeddings
- …

Pseudorandomness

- Derandomization
- Explicit constructions
- Expanders/extractors

This talk



- Two explicit constructions:
  - A “low-distortion” embedding $A: R^n \to R^m$, $m = n^{1+o(1)}$, such that for any x
    $$\|Ax\|_1 = (1 \pm \varepsilon) \|x\|_2$$
    (a.k.a. Dvoretzky’s Theorem for l_1)
  - A “nice” measurement matrix $B: R^n \to R^m$, $m = k\, n^{o(1)}$, such that for any k-sparse x, one can efficiently reconstruct x from Bx
    (several matrices with $>k^2$ measurements known)

Embedding l_2^n into l_1

Distortion	Dim. of l_1	Type	Reference
$1+\sigma$	$O(n/\sigma^2)$	Probabilistic	[Kashin, Figiel-Lindenstrauss-Milman, Gordon]
$O(1)$	$O(n^2)$	Explicit	[Rudin'60,…]
$1+1/n$	$n^{O(\log n)}$	Explicit	[Indyk'00] (cf. LLR'94)
$1+\sigma$	$O(n/\sigma^2)$	Prob., $n\log^2 n$	[Indyk'00]
$1+\sigma$	$O(n/\sigma^2)$	Prob., $n\log n$	[Arstein-Avidan, Milman'06]
$1+\sigma$	$O(n/\sigma^2)$	Prob., n	[Lovett-Sodin'07]
$1+1/\log n$	$n2^{O(\log\log n)^2}$	Explicit	[Indyk'06]
$n^{o(1)}$	$n(1+o(1))$	Explicit'	[Guruswami-Lee-Razborov'07]

Other implications

- Computing Ax takes time $O(n^{1+o(1)})$, as opposed to $O(n^2)$
- Similar phenomenon discovered for Johnson-Lindenstrauss dimensionality reduction lemma [Ailon-Chazelle'06],
  - Applications to approximate nearest neighbor problem, Singular Value Decomposition, etc

  (recall Muthu's talk)

Techniques

- Uncertainty Principles
- Extractors

Compressed Sensing

Pseudo-randomness

Uncertainty principles (UP)

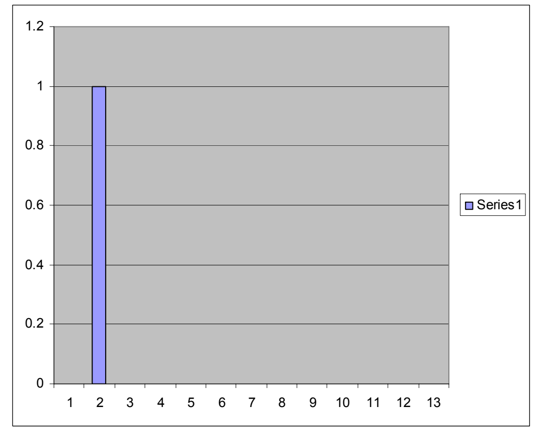
- Consider a vector $x \in R^n$ and a Fourier matrix F
- UP: either x or Fx must have "many" non-zero entries (for $x \neq 0$)
- History:
  - Physics: Heisenberg principle
  - Signal processing [Donoho-Stark'89]:
    - Consider any $2n \times n$ matrix $A=[I\ B]^T$ such that
      - B is orthonormal
      - For any distinct rows A_i , A_j of A we have

        $|A_i * A_j| \leq M$ (coherence)
    - Then for any $x \in R^n$ we have that

      $||Ax||_0 > 1/M$
  - E.g., if $A=[I\ H]^T$, where H is a normalized $n \times n$ Hadamard matrix

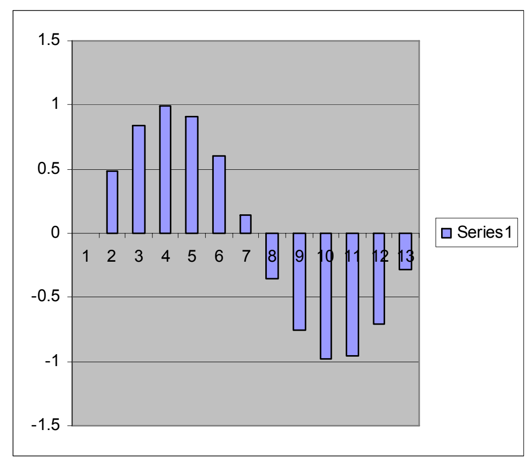
    (orthogonal, entries in $\{-1/n^{1/2}, 1/n^{1/2}\}$):
    - $M=1/n^{1/2}$
    - Ax must have $>n^{1/2}$ non-zero entries
  - We need:
    - $A=[H_1\ H_2 \dots H_L]^T$ with low-coherence (Kerdock codes)
    - Non-zero $\rightarrow$ "significantly non-zero"

x



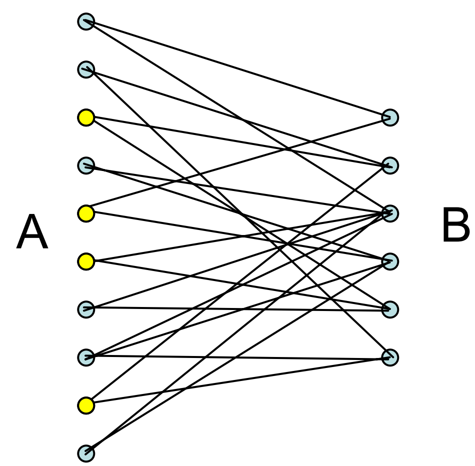
F

Fx



Extractors

- Expander-like graphs:
  - $G=(A,B,E)$, $|A|=a$, $|B|=b$
  - Left degree d
- Property:
  - Consider any distribution $P=(p_1,\ldots,p_a)$ on A s.t. $p_i \leq 1/k$
  - $G(P)$: a distribution on B:
    - Pick i from P
    - Pick t uniformly from $[d]$
    - j is the t-th neighbor of i
  - Then $\|G(P)-\text{Uniform}(B)\|_1 \leq \varepsilon$
- Equivalently, can require the above for $p_i = 1/k$ or 0
- Many explicit constructions
- Holy grail:
  - $k=b$
  - $d=O(\log a)$
- Can achieve bounds close to the above
- Observation: w.l.o.g. one can assume that the right degree is $O(ad/b)$



Overview

- Main idea behind the randomized embedding: "spread the mass" over many coordinates
  - Before:

    $$x = (1, 0, 0, 0, 0, 0, 0, 0, 0, \ldots, 0)$$

  - After:

    $$|Ax| = (\approx 1/m^{1/2}, .., \approx 1/m^{1/2}, \ldots, \ldots, \approx 1/m^{1/2})$$

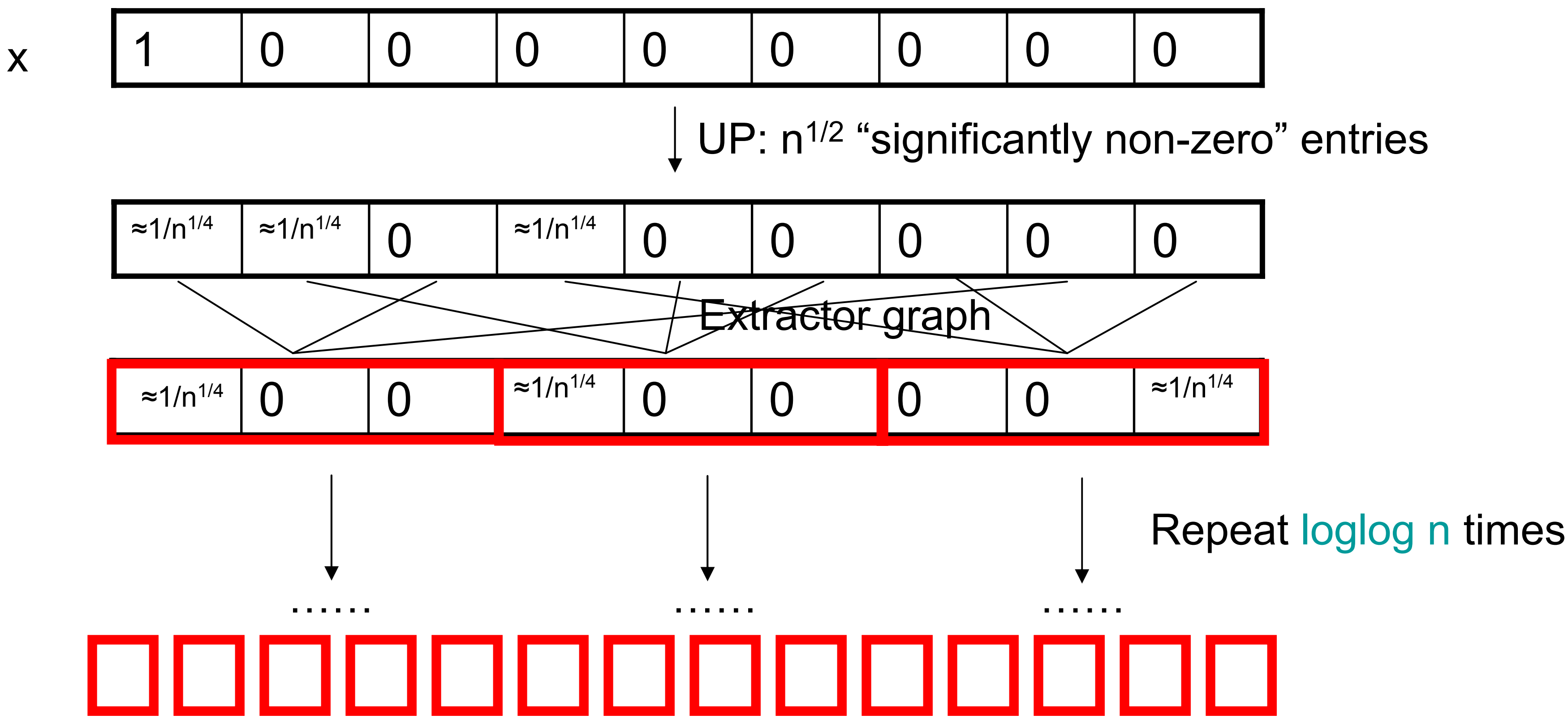
  - Therefore

    $$||Ax||_1 \approx m^{1/2} ||Ax||_2 \approx m^{1/2} ||x||_2$$

Overview, ctd.

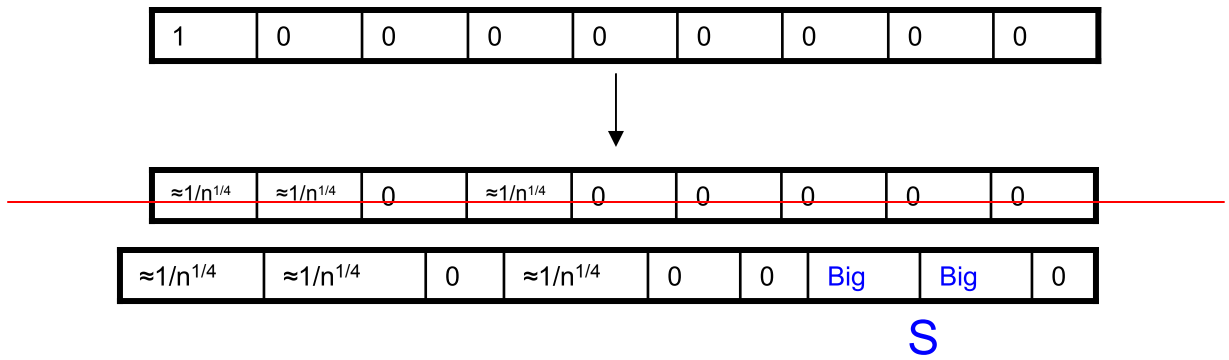
- We would like to obtain something like

  $|Ax| = (\approx 1/m^{1/2}, .., \approx 1/m^{1/2}, \ldots, \ \ldots , \approx 1/m^{1/2})$



Total dimension blowup: $(\log n)^{O(\log \log n)}$

Part I:



- Lemma:
  - Let $A=[\,H_1\ H_2\ \dots\ H_L\,]^T$, such that:
    - Each H_i is an $n\times n$ orthonormal matrix
    - For any two distinct rows A_i, A_j we have $|A_i{*}A_j| \le M$
    - M is called coherence
  - Then, for any $x\in R^n$, and set S of coordinates, $|S|=s$:

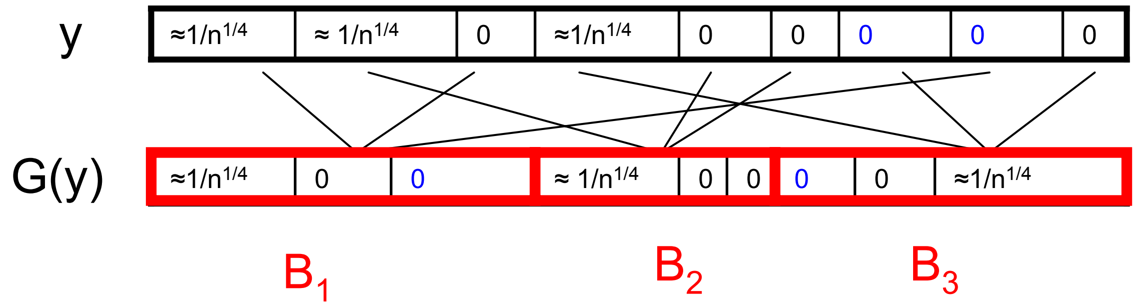
    $$\|(Ax)_{|S}\|_2^2 \le 1+Ms$$

    (note that $\|(Ax)\|_2^2 = L$)

- Proof:
  - Take A_S
  - $\max_{\|x\|=1} \|A_S x\|_2^2 = \lambda(A_S \times A_S^{\,T})$
  - But $A_S \times A_S = I+E$, $|E_{ij}|\le M$
  - Since E is an $s \times s$ matrix, $\lambda(E) \le Ms$

- Suppose that we have A s.t. $M\le 1/n^{1/2}$. Then:
  - For any $x\in R^n$, $|S|\le n^{1/2}$, we have $\|(Ax)_{|S}\|_2^2 \le 2$
  - At the same time, $\|(Ax)\|_2^2 = L$
  - Therefore, $(1-2/L)$ fraction of the "mass" $\|Ax\|_2^2$ is contained in coordinates i s.t. $(Ax)_i^{\,2}\le 1/n^{1/2}$

Part II:



- Let $y=(y_1, \ldots, y_{n'})$
- Define probability distribution

$$P=(y_1^2/||y||_2^2, \ldots, y_n^2/||y||_2^2)$$

- Extractor properties imply that, for "most" buckets B_i, we have

$$||G(y)_{|Bi}||_2^2 \approx ||G(y)||_2^2 \, / \, \#\text{buckets}$$

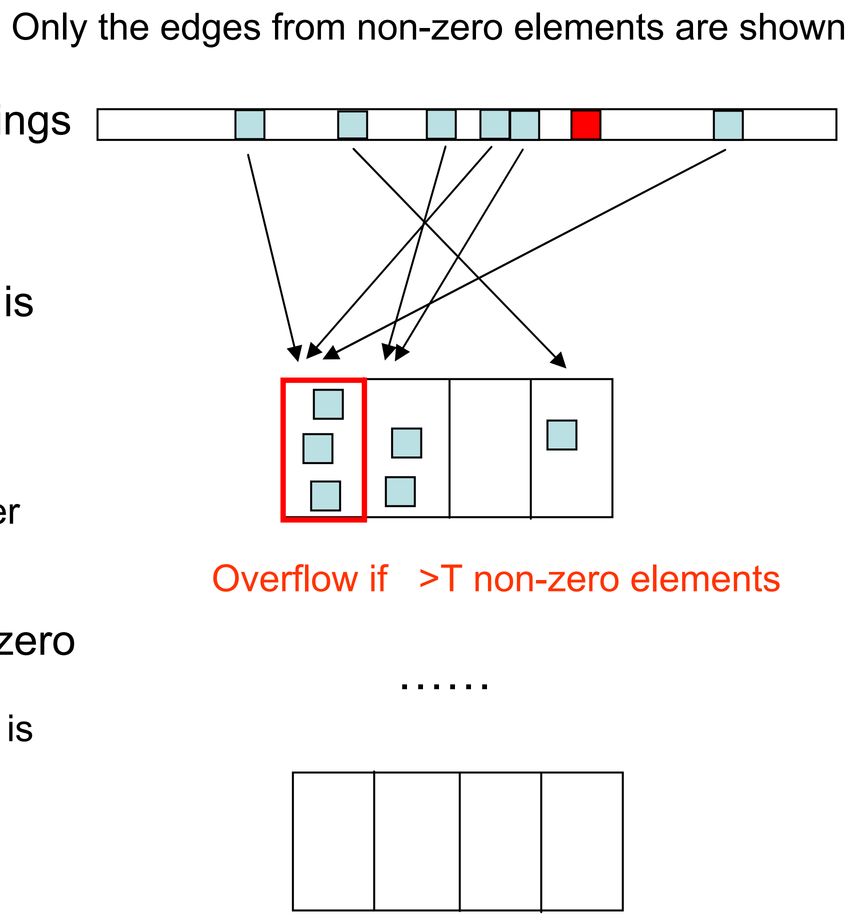
- After $\log \log n$ steps, "most" entries will be around $1/n^{1/2}$

Incoherent dictionaries

- Can we construct $A=[\ H_1\ H_2\ \ldots\ H_L\ ]^T$ with coherence (close to) $1/n^{1/2}$?
  - For $L=2$, take $A=[I\ H]^T$
  - For $L>2$:
    - Idea I: use method of conditional probabilities
      - Take $H_i = H \times D_i$, D_i has ± 1 on the diagonal and 0's elsewhere
      - Any pair of rows $u \in H_i$ and $v \in H_j$, $i \neq j$, are probabilistically indep.
        $\Rightarrow$ $|u^*v| = O(\ n^{1/2} \log n\ /\ n\ )$ with high probability
      - Derandomize using method of conditional probabilities
    - Idea II: use Google (Scholar)
      - Turns out A is known for L up to $n/2+1$ (Kerdock codes)
      - Take $H_i = H \times D_i$, D_i has ± 1 on the diagonal and 0's elsewhere

Efficient measurement matrix

- Decompose the graph into one-sided matchings / hash functions
  - $d = 2^{O(\log \log d)^2}$ hash functions
  - Each function maps $\{1..n\}$ into $\{1..O(k)\}$
- For each hash function, a non-zero element is isolated if it falls into a bucket that does not overflow
  - Can recover isolated entries using previous results
  - Cost: roughly T^2 measurements per bucket per hash function
  - Possible to set T to be polynomial in d
- Hash function property (*): at most ε of non-zero entries are not isolated
  - (*) satisfied for most hash functions if a graph is an extractor
  - Can use majority vote to determine non-zero elements
- Non-isolated elements lead to incorrect identifications
- Good news: only $O(\varepsilon)$ fraction of incorrect elements
  - We recovered z s.t., $||z-x||_0 \leq O(\varepsilon\, m)$
  - Recurse to recover $z-x$ from $Az-Ax$



Conclusions

- Extractors+UP $\rightarrow$ Embedding of l_2 into l_1
  - Dimension almost as good as for the probabilistic result
  - Near-linear in n embedding time
- Extractors + group testing $\rightarrow$ efficient measurement matrix for sparse vectors
- Questions:
  - Remove $2^{O(\log \log n)^2}$?
  - Making other embeddings/matrices explicit ?
  - Any particular reason why both [AC'06] and this paper use $H \times D_i$ matrices ?

Appendix

Digression

- Johnson-Lindenstrauss'84:
  - Take a "random" matrix $A: R^n \rightarrow R^{m/\varepsilon^2}$ $(m \ll n)$
  - For any $\varepsilon > 0$, $x \in R^n$ we have

$$||Ax||_2 = (1 \pm \varepsilon)||x||_2$$

    with probability $1-\exp(-m)$
  - Ax can be computed in $O(mn/\varepsilon^2)$ time
- Ailon-Chazelle'06:
  - Essentially: take $B = A \times P \times (H \times D_i)$, where
    - H : Hadamard matrix
    - D_i : random ± 1 diagonal matrix
    - P : projection on m^2 coordinates
    - A as above (but n replaced by m/ε^2)
  - Ax can be computed $O(n\log n + m^3/\varepsilon^2)$

(Norm) embeddings

- Metric spaces $M=(X,D)$, $M'=(X',D')$
  (here, $X=R^n$, $X'=R^m$, $D=\|\,.\|_X$ and $D=\|.\|_{X'}$)
- A mapping $F: M \to M'$ is a c-embedding if for any $p\in X$, $q\in X$ we have
  $$D(p,q) \le D'(F(p),F(q)) \le c\, D(p,q)$$
  $$\text{(or, } \|p-q\|_X \le \|F(p-q)\|_{X'} \le c\, \|p-q\|_X \text{)}$$
- History:
  - Mathematics:
    - [Dvoretzky'59]: there exists $m(n,\varepsilon)$ s.t., for any $m>m(n,\varepsilon)$ and any space $M'=(R^m,\|.\|_{X'})$ there exists a $(1+\varepsilon)$-embedding of an n-dimensional Euclidean space l_2^n into M'
    - In general, m must be exponential in n
    - [Milman'71]: proof using concentration of measure methods
    - ……..
    - [Figiel-Lindenstrauss-Milman'77, Gordon]: if $M'=l_1^m$, then $m \approx n/\varepsilon^2$ suffices
      That is, l_2^n $O(1)$-embeds into $l_1^{O(n)}$
      A.k.a. Dvoretzky's theorem for l_1
    - …
  - Computer science:
    - [Linial-London-Rabinovich'94]: [Bourgain'85] for sparsest cut, many other tools
    - …….
    - [Dvoretzky, FLM] used for approximate nearest neighbor [IM'98, KOR'98], hardness of lattice problems [Regev-Rosen'06], etc.