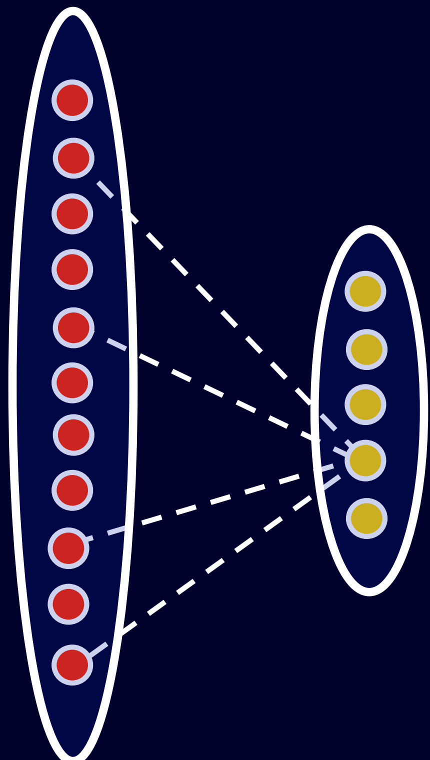
Expander codes, Euclidean sections, and compressed sensing



VENKATESAN GURUSWAMI
University of Washington
(visiting Carnegie Mellon University)

Based on joint works with

James Lee (U. Washington), *Alexander Razborov* (U. Chicago), *Avi Wigderson* (IAS)

random Euclidean sections of L_1^N

- For $x \in \mathbb{R}^N$ we have $||x||_2 \leq ||x||_1 \leq \sqrt{N}\, ||x||_2$
- [Kashin 77, Figiel-Lindenstrauss-Milman 77]:

  For a *random* subspace $X \subseteq \mathbb{R}^N$ with $\dim(X) = N/2$,

  L_2 and L_1 norms are equivalent up to universal factors

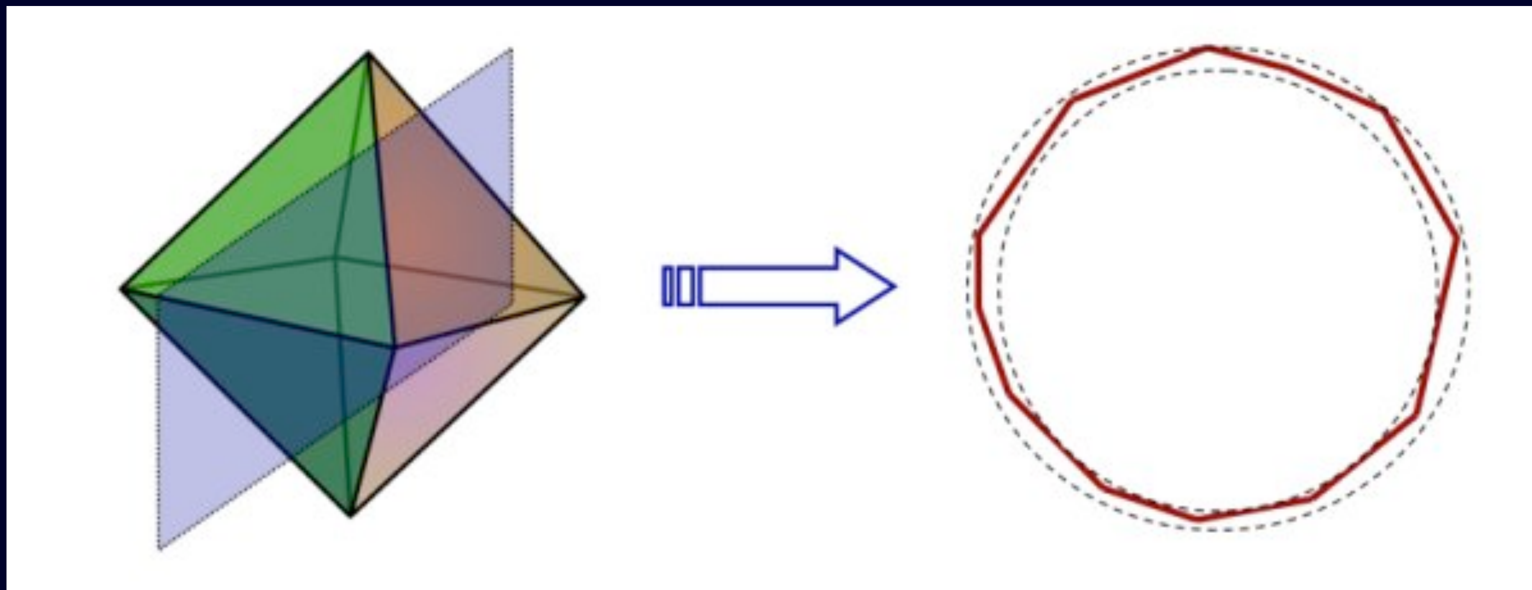
$$\|x\|_1 = \Theta(\sqrt{N})\|x\|_2 \quad \forall x \in X$$

- The L_2 mass of x is spread across many coordinates

$$\#\left\{i : |x_i| \approx N^{-\frac{1}{2}}\|x\|_2\right\} = \Omega(N)$$

- Compare with error-correcting codes: Subspace C of F_2^N such that every nonzero $c \in C$ has $\Omega(N)$ Hamming weight.

Euclidean sections, embeddings



- $L_2 \rightarrow L_1$ embeddings: Write $X = \{ G\,y : y \in \mathbb{R}^{N/2} \}$ for an $N \times N/2$ matrix G with orthonormal columns
  - The map $y \rightarrow (G\,y)/\sqrt{N}$ gives an $O(1)$ distortion embedding from $L_2^{N/2}$ to L_1^N

existential vs. constructive results

- Prominent example of ubiquitous use of probabilistic method in asymptotic convex geometry
- Dilemma we know well:
  - Almost all subspaces are good, except we can't pinpoint even one!
- Question [Szarek, ICM'06; Milman, GAFA'01; Johnson-Schechtman, handbook'01]: Can we find an *explicit* subspace where L_1 and L_2 norms are equivalent?
  - Natural, fundamental question
  - Gain in recent popularity due to ever growing connections to combinatorics and theory CS

Computer Science connections

- explicit embeddings of L_2 to L_1 for nearest-neighbor search [Indyk]
- explicit compressed sensing maps $M : \mathbb{R}^N \rightarrow \mathbb{R}^k$ (for $k << N$) [Devore] *(more on this soon)*
- Coding over reals [Candes-Tao, Dwork-McSherry-Talwar]
- dimension reduction [Ailon-Chazelle]

Explicitness (or derandomization) has many benefits:

- Better understanding of underlying geometric structure
- Faster algorithms
- Certifiability

Distortion

For a subspace $X \subseteq \mathbb{R}^N$, define the distortion of X by

$$\Delta(X) = \max_{0 \neq x \in X} \frac{\sqrt{N}\|x\|_2}{\|x\|_1}$$

Clearly, $1 \leq \Delta(X) \leq \sqrt{N}$

Our goal: **low** distortion subspaces of **large** dimension

Random construction: For a random $X \subseteq \mathbb{R}^N$ with
$\dim(X) = \Omega(N)$, w.h.p $\Delta(X) = O(1)$

- (will mention exact trade-off shortly)

Main results

- [G.-Lee-Razborov'08] Explicit subspace $\mathbf{X} \subseteq \mathbb{R}^{\mathbf{N}}$ with $\dim(X) = N - o(N)$ & $\Delta(X) = (\log N)^{O(\log\log\log N)}$
- [G.-Lee-Wigderson'08] With N^{δ} random bits, can construct subspace X with $\dim(X) = N/2$ and $\Delta(X) = O(1)$ $(= \exp(1/\delta))$
- Subspaces specified as kernel of sign matrix

previous explicit results

Sub-linear dimension (and constant distortion):

- Rudin'60 (and later LLR'94) achieved **$\dim(X) \approx N^{1/2}$** and **$\Delta(X) \leq 3$** (X = span {4-wise independent vectors})
- Indyk'00: $\dim(X) \approx 2^{\sqrt{\log N}}$ and $\Delta(X) \leq 1+o(1)$
- Indyk'07: $\dim(X) \approx N / \exp((\log \log N)^2)$ and $\Delta(X) \leq 1+o(1)$

For $\dim(X) = \Omega(N)$:

- NO explicit construction known with $\Delta(X)$ smaller than $N^{1/4}$
- $\{ (x, Hx) : x \in \mathbb{R}^{N/2} \}$ where H is the $N/2 \times N/2$ Hadamard matrix has distortion $N^{1/4}$
  - Uncertainty principle
- Regime of interest for error-correction over reals
  - constant rate codes

“derandomization” results

- Distortion-dimension trade-off of random subspaces [Kashin’77, Garnaev-Gluskin’84]
  - For a random k x N sign matrix $A_{k,N}$, almost surely

$$\Delta\left(\ker\left(A_{k,N}\right)\right) \lesssim \sqrt{\frac{N}{k}}\,\mathrm{polylog}\left(\frac{N}{k}\right)$$

(and of course **dim(ker($A_{k,N}$)) $\geq$ N – k**)

- Construction with O(N log N) random bits [Arstein-Milman’06]
- Construction with O(N) random bits [Lovett-Sodin 07]

rest of talk

- Connection to compressed sensing
- Subspaces from expander/Tanner codes
- Constant distortion construction
- Explicit construction
  - “Spread-boosting” theorem
  - Using spread boosting: ingredients & analysis
- Conclusions

compressed sensing

Typical camera algorithm:

1. Captures an image, a signal in $\mathbb{R}^N$
2. Compresses image **Comp** : $\mathbb{R}^N \rightarrow \mathbb{R}^r$

Compression works because image is r-sparse in some basis (eg. Wavelet)
Camera still makes N >> r measurements since it doesn't a priori know which ones to make!

Compressed sensing: Find a map $A : \mathbb{R}^N \rightarrow \mathbb{R}^k$ s.t. any r-sparse $x \in \mathbb{R}^N$ can be recovered (efficiently & uniquely) from Ax (ideally $k = r \log N$)

HOT topic: http://www.dsp.ece.rice.edu/cs/ [Donoho], [Candes-Tao], [Rudelson-Vershynin], [Candes-Romberg-Tao], , …. Two talks in ICM'06.

Properties of random matrices (restricted isometry, Gelfand width, distortion) play crucial role in these developments.

relation to distortion

- [Kashin-Temlyakov'07] make explicit a simple connection between distortion of ker(A) & compressed sensing using A:
  - Can uniquely and efficiently recover any r-sparse signal for $r < N / (4 \Delta(\ker(A))^2)$
- Algorithm is basis pursuit or L_1 minimization:
  - given data y, output x that minimizes $||x||_1$ subject to $Ax=y$
  - easy by linear programming (L_0 minimization is NP-hard)
  - handles almost sparse x (very important in practice!)
- Plugging in optimal distortion bound $(N/k \log(N/k))^{1/2}$ gives (optimal) $k \approx r \log N$ measurements!
  - Explicit construction open

proof of connection (for zero noise case)

Let $X = \ker(A)$.

Suppose $y = Ax$ and $|\mathrm{supp}(x)| \le r < N/\, 4\Delta(X)^2$

Need to prove: For any nonzero $u \in X$, $||x+u||_1 > ||x||_1$

Let $S = \mathrm{supp}(x)$, and $T=S^c$

$$||x+u||_1 \ge ||x||_1 - \sum_{j \in S} |u_j| + \sum_{j \in T} |u_j|$$
$$\ge ||x||_1 + ||u||_1 - 2 \sum_{j \in S} |u_j|$$

<u>Claim</u>: For nonzero u, $\sum_{j \in S} |u_j| < ||u||_1/2$

<u>Proof</u>: LHS $\le \sqrt{|S|}\; ||u||_2 \le \sqrt{r}\; (\; ||u||_1\, \Delta(X)/ \sqrt{N}) < ||u||_1/2$

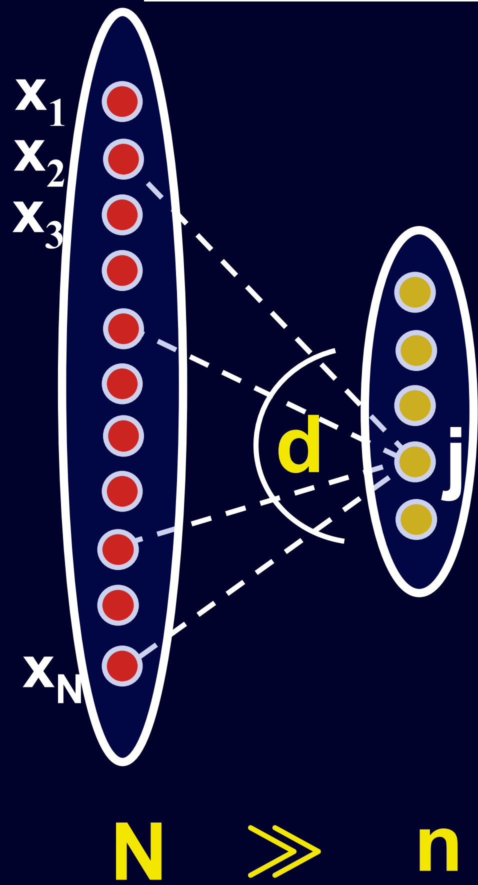
outline

- Connection to compressed sensing
- Subspaces from expander/Tanner codes
- Constant distortion construction
- Explicit construction
  - “Spread-boosting” theorem
  - Using spread boosting: ingredients & analysis
- Conclusions

Expander/LDPC code construction

bipartite graph $G=([N], [n], E)$; d right-regular, $L \subseteq \mathbb{R}^d$

$$X(G,L) = \left\{ x \in \mathbb{R}^N : x_{\Gamma(j)} \in L \quad \forall j \in [n] \right\}$$



Continuous analog of Gallager LDPC codes and extension by Tanner

- Global structure from local constraints
- Like in Sipser-Spielman analysis, ***expansion*** of G plays a crucial role

We show: if L is *good*, and G is an *expander*, then X(G,L) is good (or even *better* in some parameters)

spread subspaces

Key notion: $L \subseteq \mathbb{R}^d$ is **(t, ε)**-spread if every $x \in L$ satisfies

$$\min_{|S| \leq t} \|x_{\bar{S}}\|_2 \geq \varepsilon \cdot \|x\|_2$$

"No t coordinates hog most of the mass"

Equivalent notion to distortion (easier to work with)

- O(1) distortion ⇔ (Ω(d), Ω(1))-spread
- (t, ε)-spread ⇒ distortion $O(\varepsilon^{-2} \cdot (d/t)^{1/2})$

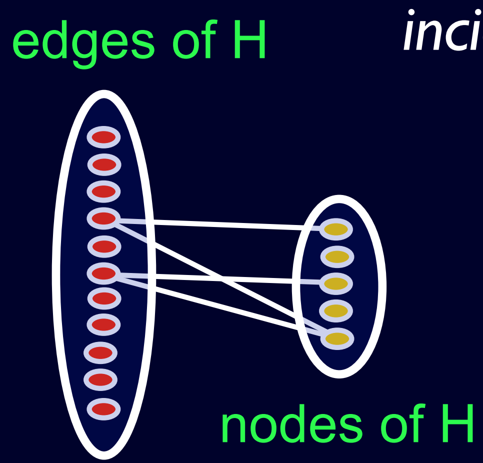
Note: Every subspace is trivially (1/2, 1)-spread.

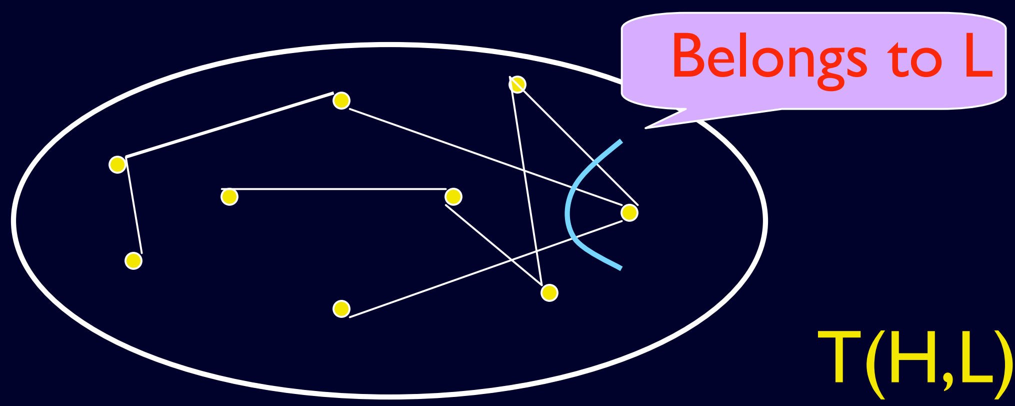
Goal: Increase t while not losing too much mass.

- (t, ε)-spread → (t', ε')-spread

constant distortion construction

Take unbalanced expander to be *edge-vertex incidence* graph of d-regular expander H(V,E)





$$\text{Subspace} = \{x \in \mathbb{R}^E \mid x_{E(v)} \in L \ \forall v \in V\}$$
where $E(v)$ = set of d edges incident on v.

$L \subseteq \mathbb{R}^d$ is a random subspace

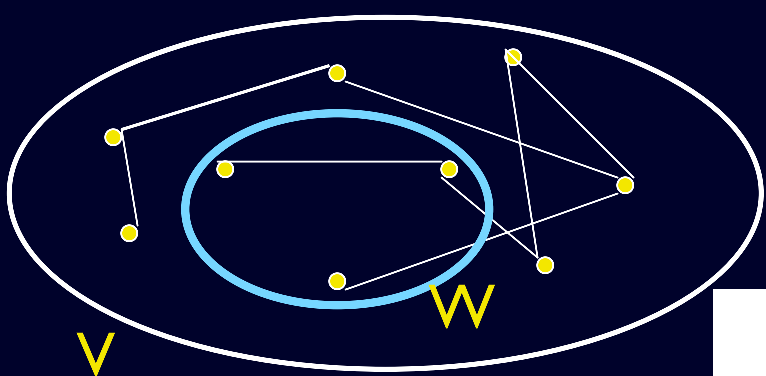
- has O(1) distortion, say is (d/10, 0.1)-spread

For $d = n^{\delta/2}$, can pick L using n^{δ} random bits.

distortion/spread analysis

- Thm: If H is an (n,d,λ)-expander with $\lambda \le d^{0.9}$, and L is $(d/10, 0.1)$-spread, then distortion of T(H,L) is $n^{O(1/\log d)}$
  - $O(1) = \exp(1/\delta)$ distortion with $d = n^{\delta/2}$

- Show T(H,L) is $(N/200, n^{-O(1/\log d)})$-spread
  - ($N = nd/2$ is # edges of H)



Suffices to show:

For unit vector $x \in T(H,L)$ & set W of $< n/20$ vertices

$$\sum_{e \notin E(W)} x_e^2 \geq n^{-O(1/\log d)}$$

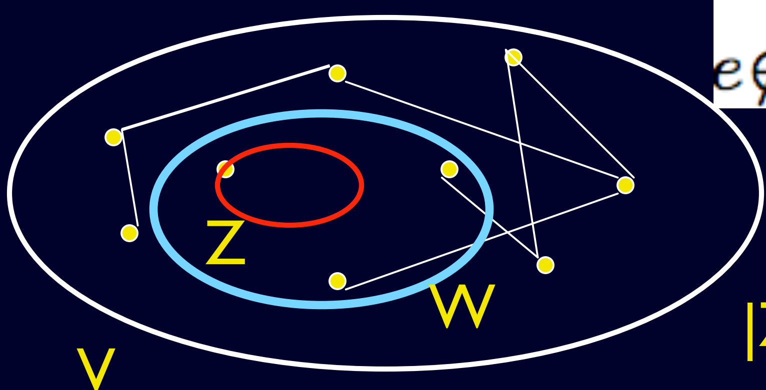
spread outside induced subgraphs

- Define Z = { z ∈ W : z has > d/10 neighbors in W }
- By local (d/10,0.1)-spread property, mass in W \ Z "leaks out"

$$\sum_{e \in E(W,\overline{W})} x_e^2 \geq \frac{1}{100} \sum_{v \in W \setminus Z} \|x_{N(v)}\|_2^2$$

It follows that

$$\sum_{e \notin E(W)} x_e^2 \geqslant \frac{1}{100} \sum_{e \notin E(Z)} x_e^2$$



By expander mixing lemma,

$|Z| < O((\lambda/d)^2)\ |W| < O(|W|/d^{0.2})$

Iterating this $\log_d n$ times, claim follows.

outline

- Connection to compressed sensing
- Subspaces from expander/Tanner codes
- Constant distortion construction
- Explicit construction
  - "Spread-boosting" theorem
  - Using spread boosting: ingredients & analysis
- Conclusions

spread-boosting theorem (general bipartite expanders)

setup: bipartite graph $G=([N], [n], E)$; d right-regular;

(t,ε)-spread local subspace $L \subseteq \mathbb{R}^d$

(left-to-right) expansion profile of G:

$$\Lambda_G(m) = \min\{|\Gamma_G(S)| : |S| \geq m, S \subseteq [N]\}$$

theorem: If $X(G,L)$ is (T,δ)-spread, then $X(G,L)$ is

$$\left(\frac{t}{D}\Lambda_G(T), \frac{\varepsilon\delta}{\sqrt{2D}}\right)\text{-spread}$$

applying the thm: Think ε, D as constants. Want $t\,\Lambda_G(T)/T$ large to get from $(1/2,1)$-spread to say $(\Omega(N), \gamma)$-spread in few iterations

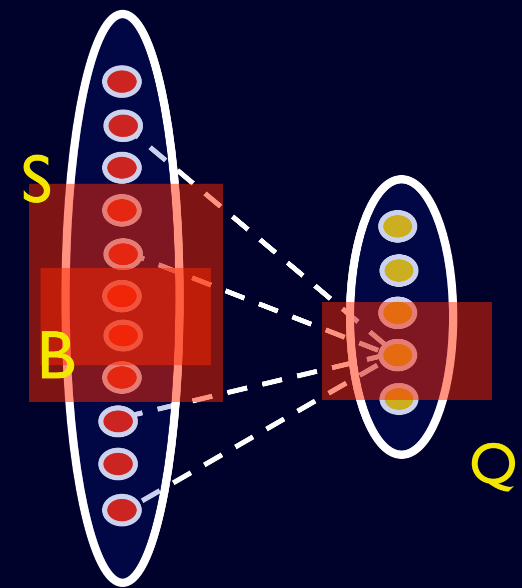
(γ is exponentially small in # iterations)

proving spread-boosting theorem

X(G,L) is (T,δ)-spread ⇒ X(G,L) is $\left(\frac{t}{D}\Lambda_G(T), \frac{\varepsilon\delta}{\sqrt{2D}}\right)$-spread

Let S arbitrary with $|S| \leq t\, \Lambda_G(T)/D$

Idea: S should "leak" L_2 mass outside (since L is spreading and G is an expander), unless most of the mass in S is concentrated on small subset B (impossible by assumption)



Details:

- Q = right nodes with > t neighbors in S
- $|Q| < |S|\, D / t \leq \Lambda_G(T)$
- B = nodes in S whose neighbors are all in Q
- $|B| < T$ so it can't contain too much mass
- Mass in S - B leaks out

outline

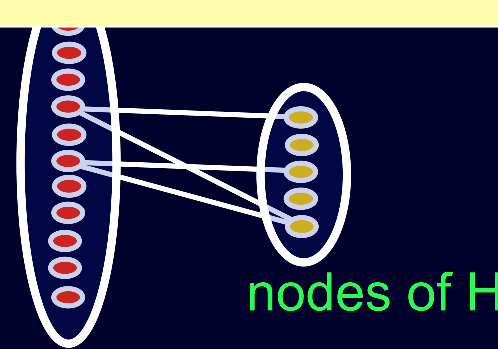
- Connection to compressed sensing
- Subspaces from expander/Tanner codes
- Constant distortion construction
- Explicit construction
  - "Spread-boosting" theorem
  - Using spread boosting: ingredients & analysis
- Conclusions

using spread-boosting

Goal: Construct explicit $(d^{0.51}, \Omega(1))$-spread subspace of large dimension

- [Alon-Chung] (expander mixing lemma):

$$\Lambda_G(m) \approx \min\left\{ \frac{m}{\sqrt{d}}, \frac{\sqrt{Nm}}{d} \right\}$$



- Suppose L is $(t, 0.1)$-spread
- $(T,\delta) \rightarrow (\ \Omega(\ t\ \Lambda_G(T)\ )\ ,\ \Omega(\delta)\ )$-spread
- $(T,\delta) \rightarrow (\ \Omega(\ T\ t\ d^{-1/2}\ )\ ,\ \Omega(\delta)\ )$-spread
- If $t \gg d^{1/2}$, we compensate for the factor $d^{1/2}$ loss in expansion and get increase in T
- Optimal/random L has $t = \Omega(d)$
  - Can construct efficiently only for small d, say $d = \log N$
  - Problem: # iterations $\approx \log_d N$, so need d large (say $N^{0.1}$)

explicit somewhat well-spread subspace

Mutually Unbiased Bases from Kerdock codes [Kerdock'72, Cameron-Seidel'73]:

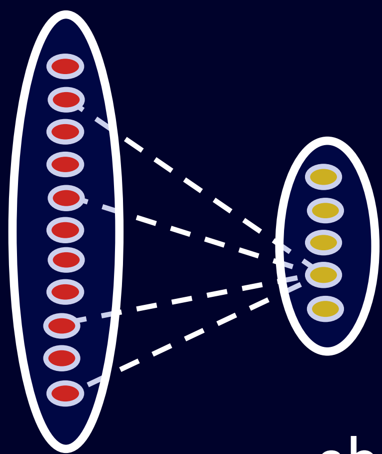
Explicit set of $k/2$ orthonormal bases $B_1, \ldots, B_{k/2} \subseteq k^{-1/2}\{-1,1\}^k$ such that $u \in B_i$ and $v \in B_j$, $i \neq j \Rightarrow |\langle u,v \rangle| = k^{-1/2}$.

- $A = [\, B_1, \ldots, B_m \,]$ for any $1 \leq m \leq k/2$
- One can show that $\ker(A)$ is **$(\Omega(d^{1/2}), \Omega(1))$**-spread
- Gives **$(\Omega(d^{1/2}), \Omega(1))$**-spread subspace of dimension $(1-\varepsilon)d$ for any $\varepsilon > 0$
- But too weak to work with Ramanujan construction :(

sum-product expander

- Need (N,n,d)-expander with expansion factor better than $1/\sqrt{d}$ factor for sets of size up to $N^{0.51}$
- Sum-product theorems [Bourgain-Katz-Tao, …]: For $A \subseteq \mathbb{F}_p$ with $|A| < p^{0.9}$, $|A+A| \geq |A|^{1+\delta_0}$ or $|A \cdot A| \geq |A|^{1+\delta_0}$

$(a,b,c) \in \mathbb{F}_p^3 \longrightarrow (1,a), (2,b), (3,c), (4, a \cdot b+c)$



[Barak-Kindler-Shaltiel-Sudakov-Wigderson, Barak-Impagliazzo-Wigderson]: For some $\xi > 0$

$\Lambda_G(m) \geq \min (p^{0.9} , m^{1/3+\xi})$

For L = Kerdock, G = sum-product expander: above + spread-boosting theorem ⇒ L' = X(G,L) is $(d^{1/2+c}, \Omega(1))$-spread for some $c > 0$.

the final construction

- Now plug L' into X(G',L') with G' = edge-vertex graph of Ramanujan and get non-trivial boosting
- Actual construction is intersection of many such "parallel" constructions
  - To minimize iterations, need right degree d large
  - $1/\sqrt{d}$ expansion stops at N/d, so can't use a single large d
  - Idea: using many graphs with different degrees, each efficiently boosting in a different range of sizes
    - $d_{i+1} = d_i$^(β^i)
    - Degrees reduce from N to log log N in O(log log N) steps (lose factor log log N in L_2 mass in each step)

concluding remarks

- Subspaces of $\mathbb{R}^N$ of dimension $\Omega(N)$
  - Explicit with distortion $(\log N)^{O(\log \log \log N)}$
  - Using N^δ random bits, distortion $O(1)$
  - Continuous analog of expander/Tanner codes
  - Ingredients in explicit construction: Ramanujan graphs, Kerdock codes, sum-product theorem in finite fields
- Some questions:
  - Explicit construction with dim $\Omega(N)$ and distortion $O(1)$
  - Better dependence of distortion on co-dimension (important for compressed sensing application)
  - Iterative near-linear time decoding (for Tanner codes)
    - some results in [Xu-Hassibi, G.-Lee-Wigderson]