Near Optimal Dimensionality Reductions that Preserve Volumes

RANDOM/APPROX 2008

Avner Magen **Anastasios Zouzias**

University of Toronto

August, 2008

Dimension Reduction

$P \subseteq \mathbb{R}^t$: set of n points

Goal: Find $f : P \to \mathbb{R}^d$ $(d \ll n, t)$ s.t. *some property* is preserved.

Measure of quality (Distance)

f has *distortion* $1 + \varepsilon$ if

$$\forall p, q \in P \quad \|p - q\| \leq \|f(p) - f(q)\| \leq (1 + \varepsilon)\,\|p - q\|\,.$$

Measure of quality (Volume)

f has *volume distortion* $1 + \varepsilon$ if

$$\forall S \subset P, |S| \leq k \quad 1 \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{|S|-1}} \leq 1 + \varepsilon.$$

Dimension Reduction

$P \subseteq \mathbb{R}^t$: set of n points

Goal: Find $f : P \to \mathbb{R}^d$ ($d \ll n, t$) s.t. *some property* is preserved.

Measure of quality (Distance)

f has *distortion* $1 + \varepsilon$ if

$$\forall p, q \in P \quad \|p - q\| \leq \|f(p) - f(q)\| \leq (1 + \varepsilon) \|p - q\| .$$

Measure of quality (Volume)

f has *volume distortion* $1 + \varepsilon$ if

$$\forall S \subset P, |S| \leq k \quad 1 \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}} \right)^{\frac{1}{|S|-1}} \leq 1 + \varepsilon.$$

Dimension Reduction

$P \subseteq \mathbb{R}^t$: set of n points

Goal: Find $f : P \to \mathbb{R}^d$ ($d \ll n, t$) s.t. *some property* is preserved.

Measure of quality (Distance)

f has *distortion* $1+\varepsilon$ if

$$\forall p, q \in P \quad \|p-q\| \leq \|f(p)-f(q)\| \leq (1+\varepsilon)\|p-q\|.$$

Measure of quality (Volume)

f has *volume distortion* $1+\varepsilon$ if

$$\forall S \subset P, |S| \leq k \quad 1 \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{|S|-1}} \leq 1+\varepsilon.$$

Dimension Reduction

$P \subseteq \mathbb{R}^t$: set of n points

Goal: Find $f : P \to \mathbb{R}^d$ ($d \ll n, t$) s.t. *some property* is preserved.

Measure of quality (Distance)

f has *distortion* $1+\varepsilon$ if

$$\forall p, q \in P \quad \|p-q\| \leq \|f(p)-f(q)\| \leq (1+\varepsilon)\,\|p-q\|.$$

Measure of quality (Volume) (This talk)

f has *volume distortion* $1+\varepsilon$ if

$$\forall S \subset P, |S| \leq k \quad 1 \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{|S|-1}} \leq 1+\varepsilon.$$

Johnson Lindenstrauss Lemma

Lemma (Distances)

Let P an n-point subset of Euclidean space. There exists a mapping f from P into $\mathbb{R}^d, d = O(\varepsilon^{-2} \log n)$ such that

$$\forall x, y \in P \quad (1-\varepsilon)\|x-y\| \leq \|f(x) - f(y)\| \leq (1+\varepsilon)\|x-y\|$$

Johnson Lindenstrauss Lemma

Lemma (Distances)

Let P an n-point subset of Euclidean space. There exists a mapping f from P into $\mathbb{R}^d, d = O(\varepsilon^{-2}\log n)$ such that

$$\forall x, y \in P \quad (1-\varepsilon)||x-y|| \leq ||f(x)-f(y)|| \leq (1+\varepsilon)||x-y||$$

Almost tight Lower bound $\Omega(\varepsilon^{-2}\log n / \log(1/\varepsilon))$ [Alon, 2003].

Random Projections

Many ways to generate such a (linear) mapping (encoded by $X \in \mathbb{R}^{n \times d}$):

- $X_{i,j} \sim N(0,1)$
- $X_{i,j} \sim \pm 1$ w.p. $1/2$.
- Sparse Gaussian matrix (with preprocessing)
- Entries with Subgaussian tails
- ECC and Rademacher r.v.
- Lean Walsh Transform. (Next talk, [Liberty et al., 2008])

Random Projections

Many ways to generate such a (linear) mapping (encoded by $X \in \mathbb{R}^{n \times d}$):

- $X_{i,j} \sim N(0,1)$ (This talk)
- $X_{i,j} \sim \pm 1$ w.p. $1/2$.
- Sparse Gaussian matrix (with preprocessing)
- Entries with Subgaussian tails
- ECC and Rademacher r.v.
- Lean Walsh Transform. (Next talk, [Liberty et al., 2008])

Related Work: Extensions of JL to other cases

[Magen, 2002] Preserve volumes of subsets of size up to k *and* affine distances using $\mathcal{O}(\mathbf{k}\varepsilon^{-2}\log n)$ dimensions.

[Sarlos, 2006] Preserve distances of *all* points lying in any k dim. linear subspace by projecting into $\mathcal{O}(k\varepsilon^{-2}\log(k/\varepsilon))$ dimensions.

[Wakin and Baraniuk, 2006, Agarwal et al., 2007, Clarkson, 2008] Moving points, curves, surfaces and manifolds etc.

Related Work: Extensions of JL to other cases

[Magen, 2002] Preserve volumes of subsets of size up to k *and* affine distances using $\mathcal{O}(\mathbf{k}\varepsilon^{-2}\log n)$ dimensions.

[Sarlos, 2006] Preserve distances of *all* points lying in any k dim. linear subspace by projecting into $\mathcal{O}(k\varepsilon^{-2}\log(k/\varepsilon))$ dimensions.

[Wakin and Baraniuk, 2006, Agarwal et al., 2007, Clarkson, 2008] Moving points, curves, surfaces and manifolds etc.

Our Contribution

- Improve Magen's result for volumes, by showing that $\mathcal{O}(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$ dimensions are enough.
- JL Lemma preserves more than distances. It preserves volumes of subsets of size up to $\log n/\varepsilon$.

Our Result

Theorem

Let $P \subset \mathbb{R}^n$. *There* $\exists f : P \to \mathbb{R}^d, d = \mathcal{O}(\max\{\frac{k}{\varepsilon}, \varepsilon^{-2}\log n\})$, *s.t.* $\forall$ *subset* S *of* P, $1 < |S| < k$,

$$1 - \varepsilon \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{|S|-1}} \leq 1 + \varepsilon.$$

Overview of proof:

- There are roughly $\mathcal{O}(n^s)$ sets of size s.
- It suffices to prove the failure probability for a subset of size s is roughly $e^{-\Omega(sd\varepsilon^2)}$.
- Union bound implies that a volume-preserving mapping exists.

Our Result

Theorem

Let $P \subset \mathbb{R}^n$*. There* $\exists f : P \to \mathbb{R}^d, d = \mathcal{O}(\max\{\frac{k}{\varepsilon}, \varepsilon^{-2}\log n\})$*, s.t.* $\forall$ *subset* S *of* P*,* $1 < |S| < k$*,*

$$1 - \varepsilon \leq \left(\frac{\mathrm{vol(f(S))}}{\mathrm{vol(S)}}\right)^{\frac{1}{|S|-1}} \leq 1 + \varepsilon.$$

Overview of proof:

- There are roughly $\mathcal{O}(n^s)$ sets of size s.
- It suffices to prove the failure probability for a subset of size s is roughly $e^{-\Omega(sd\varepsilon^2)}$. (Core of the talk.)
- Union bound implies that a volume-preserving mapping exists.

Proof

Two steps:

1. Prove it for the regular n-simplex.
2. Reduce the general case to the above case.

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.
- Form a matrix $e_i \to i^{th}$ row, i.e., identity matrix.

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.
- Form a matrix $e_i \rightarrow i^{th}$ row, i.e., identity matrix.
- Random Projection (without normalization)

$$\begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & 1 & 0 \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}_{n\times n} \begin{bmatrix} \\ \\ X_{ij} \sim \mathcal{N}(0,1) \\ \\ \\ \end{bmatrix}_{n\times d}$$

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.
- Projected points are Random Gaussian Vectors in $\mathbb{R}^d$.

$$\begin{bmatrix} \\ \\ X_{ij} \quad \sim \mathcal{N}(0,1) \\ \\ \\ \end{bmatrix}_{n \times d}$$

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.
- Projected points are Random Gaussian Vectors in $\mathbb{R}^d$.
- Pick any subset S, $|S| = s$ of rows of X

$$X_S := \begin{bmatrix} & & \\ & X_{ij} \quad \sim \mathcal{N}(0,1) & \\ & & \end{bmatrix}_{s \times d}$$

The n-simplex

- Assume input points are $\{e_1, \ldots, e_n\}$.
- Projected points are Random Gaussian Vectors in $\mathbb{R}^d$.
- Pick any subset S, $|S| = s$ of rows of X

$$X_S := \begin{bmatrix} & & \\ & X_{ij} \quad \sim \mathcal{N}(0,1) & \\ & & \end{bmatrix}_{s \times d}$$

- $\mathrm{vol}(S \cup \{\mathbf{0}\}) = \sqrt{\det(X_S X_S^\top)}/s!$.

The n-simplex

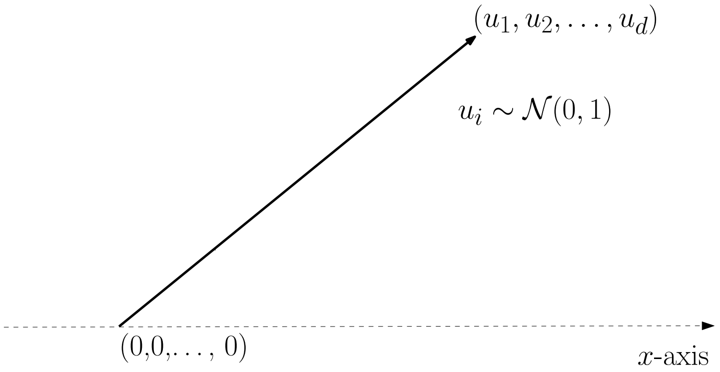
- Assume input points are $\{e_1, \ldots, e_n\}$.
- Projected points are Random Gaussian Vectors in $\mathbb{R}^d$.
- Pick any subset S, $|S| = s$ of rows of X

$$X_S := \begin{bmatrix} & & \\ & X_{ij} \quad \sim \mathcal{N}(0,1) & \\ & & \end{bmatrix}_{s \times d}$$

- $\mathrm{vol}(S \cup \{\mathbf{0}\}) = \sqrt{\det(X_S X_S^\top)}/s!$.
- What's the distribution of $\sqrt{\det(X_S X_S^\top)}$?

Distribution of $\sqrt{\det(X_S X_S^\top)}$ for $s = 2$

1) Pick a random vector

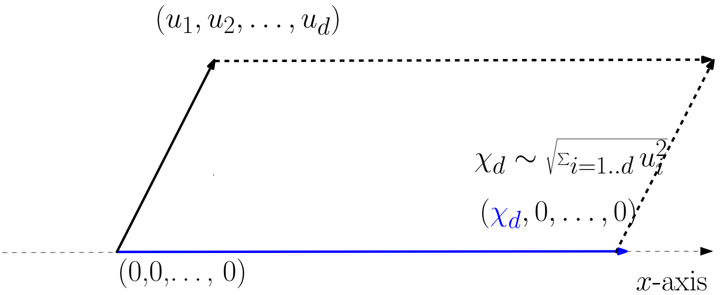


Distribution of $\sqrt{\det(X_S X_S^\top)}$ for $s = 2$

1) Pick a random vector
2) Rotate to x-axis (Rotational invariance).

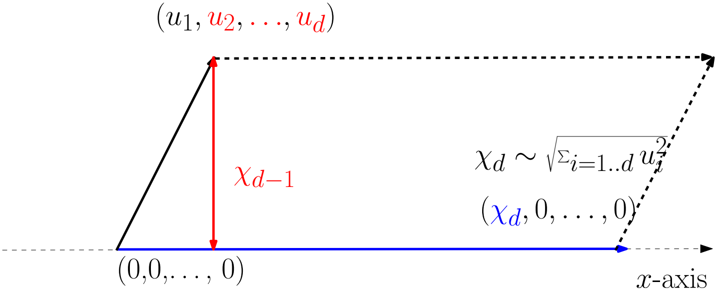
Distribution of $\sqrt{\det(X_S X_S^\top)}$ for $s = 2$

1) Pick a random vector
2) Rotate to x-axis (Rotational invariance).
3) Pick one more.



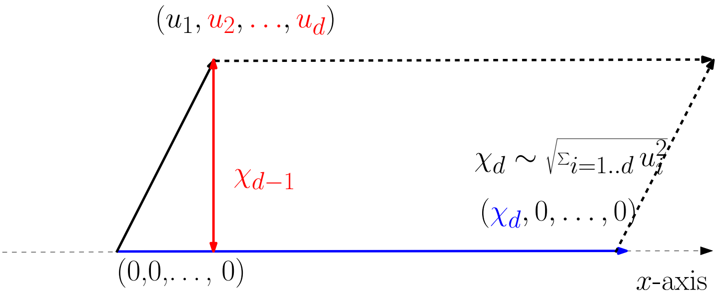
Distribution of $\sqrt{\det(X_S X_S^\top)}$ for $s = 2$

1) Pick a random vector
2) Rotate to x-axis (Rotational invariance).
3) Pick one more.



Distribution of $\sqrt{\det(X_S X_S^\top)}$ for $s = 2$

1) Pick a random vector
2) Rotate to x-axis (Rotational invariance).
3) Pick one more.



$\mathrm{vol}(\diamond) \sim \chi_d \chi_{d-1}$

Random Determinants

- For $s = 2$, $\sqrt{\det(X_S X_S^\top)} = \chi_d \chi_{d-1}$
- Using induction, we can show that:

Let X_S s-by-d ($s \leq d$) Gaussian Random Matrix,

$$\det(X_S X_S^\top) \sim \prod_{i=1}^{s} \chi^2_{d-i+1}$$

where the Chi square r.v. are independent.
Facts about χ^2 distribution:

- $\chi_t^2 = \sum_{i=1}^{t} u_i^2$, $u_i \sim \mathcal{N}(0,1)$.
- χ_t^2 is sharply concentrated around its expected value ($e^{-\Omega(t\varepsilon^2)}$).

$$Pr\left[\left|\chi_t^2 - E[\chi_t^2]\right| > \varepsilon E[\chi_t^2]\right] < e^{-\Omega(t\varepsilon^2)}$$

Random Determinants

- For $s = 2$, $\sqrt{\det(X_S X_S^\top)} = \chi_d \chi_{d-1}$
- Using induction, we can show that:

Let X_S s-by-d ($s \leq d$) Gaussian Random Matrix,

$$\det(X_S X_S^\top) \sim \prod_{i=1}^{s} \chi^2_{d-i+1}$$

where the Chi square r.v. are independent.
Facts about χ^2 distribution:

- $\chi_t^2 = \sum_{i=1}^{t} u_i^2$, $u_i \sim \mathcal{N}(0,1)$.
- χ_t^2 is sharply concentrated around its expected value ($e^{-\Omega(t\varepsilon^2)}$).

What about the concentration of the (normalized) product of χ^2?

Normalized Product of Independent χ^2

Theorem ([Gordon, 1989])

$$\left(\prod_{i=1}^{s} \chi^2_{d-i+1}\right)^{\frac{1}{s}} \sim \frac{1}{s}\chi^2_{sd\pm O(s^2)}$$

Reminder: χ^2_t concentrated with $e^{-\Omega(t\varepsilon^2)}$.

Implication:

- Take any subset S of the vertices of the n-simplex, call it I_S.
- $\left(\frac{\mathrm{vol(f(I_S))}}{\mathrm{vol(I_S)}}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\mathrm{det}(X_S X_S^\top)}/s!}{\sqrt{\mathrm{det}(I_S I_S^\top)}/s!}\right)^{\frac{1}{s}} = \left(\mathrm{det}(X_S X_S^\top)\right)^{\frac{1}{2s}}$

Normalized Product of Independent χ^2

Theorem ([Gordon, 1989])

$$\left(\prod_{i=1}^{s} \chi^2_{d-i+1}\right)^{\frac{1}{s}} \sim \frac{1}{s}\chi^2_{sd\pm O(s^2)}$$

Reminder: χ^2_t concentrated with $e^{-\Omega(t\varepsilon^2)}$.

Implication:

- Take any subset S of the vertices of the n-simplex, call it I_S.
- $\left(\frac{\mathrm{vol}(\mathrm{f}(\mathrm{I_S}))}{\mathrm{vol}(\mathrm{I_S})}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\det(\mathrm{X_S X_S^\top})}/s!}{\sqrt{\det(\mathrm{I_S I_S^\top})}/s!}\right)^{\frac{1}{s}} = \left(\det(\mathrm{X_S X_S^\top})\right)^{\frac{1}{2s}}$

Normalized Product of Independent χ^2

Theorem ([Gordon, 1989])

$$\left(\prod_{i=1}^{s} \chi^2_{d-i+1}\right)^{\frac{1}{s}} \sim \frac{1}{s}\chi^2_{sd\pm O(s^2)}$$

Reminder: χ^2_t concentrated with $e^{-\Omega(t\varepsilon^2)}$.

Implication:

- Take any subset S of the vertices of the n-simplex, call it I_S.
- $\left(\frac{\mathrm{vol(f(I_S))}}{\mathrm{vol(I_S)}}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\mathrm{det(X_S X_S^\top)}}/s!}{\sqrt{\mathrm{det(I_S I_S^\top)}}/s!}\right)^{\frac{1}{s}} = \left(\mathrm{det(X_S X_S^\top)}\right)^{\frac{1}{2s}}$

Normalized Product of Independent χ^2

Theorem ([Gordon, 1989])

$$\left(\prod_{i=1}^{s} \chi^2_{d-i+1}\right)^{\frac{1}{s}} \sim \frac{1}{s}\chi^2_{sd \pm O(s^2)}$$

Reminder: χ^2_t concentrated with $e^{-\Omega(t\varepsilon^2)}$.

Implication:

- Take any subset S of the vertices of the n-simplex, call it I_S.
- $\left(\frac{\mathrm{vol}(\mathrm{f}(\mathrm{I_S}))}{\mathrm{vol}(\mathrm{I_S})}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\det(\mathrm{X_S X_S^\top})}/s!}{\sqrt{\det(\mathrm{I_S I_S^\top})}/s!}\right)^{\frac{1}{s}} = \left(\det(\mathrm{X_S X_S^\top})\right)^{\frac{1}{2s}} \approx \chi^2_{sd}$

Normalized Product of Independent χ^2

Theorem ([Gordon, 1989])

$$\left(\prod_{i=1}^{s} \chi^2_{d-i+1}\right)^{\frac{1}{s}} \sim \frac{1}{s}\chi^2_{sd\pm O(s^2)}$$

Reminder: χ^2_t concentrated with $e^{-\Omega(t\varepsilon^2)}$.

Implication:

- Take any subset S of the vertices of the n-simplex, call it I_S.
- $\left(\frac{\mathrm{vol}(\mathrm{f}(\mathrm{I_S}))}{\mathrm{vol}(\mathrm{I_S})}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\det(\mathrm{X_S X_S^\top})}/s!}{\sqrt{\det(\mathrm{I_S I_S^\top})}/s!}\right)^{\frac{1}{s}} = \left(\det(\mathrm{X_S X_S^\top})\right)^{\frac{1}{2s}} \approx \chi^2_{sd}$
- The normalized volume of any subset I_S of the n-simplex is concentrated with $\exp(-\Omega(sd\varepsilon^2))$

Points in General Position

- Let P a n-point set in general position.

$$\begin{bmatrix} * & * & * & \dots & * \\ * & * & * & \dots & * \\ * & * & P_{ij} & & \vdots \\ \vdots & \vdots & & * & * \\ * & * & \dots & * & * \end{bmatrix}_{n\times n} \begin{bmatrix} & & \\ & X_{ij} \sim \mathcal{N}(0,1) & \\ & & \end{bmatrix}_{n\times d}$$

Points in General Position

- Pick any subset P_S of P

$$\begin{bmatrix} * & * & * & * & * \\ * & * & P_S & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}_{s\times n} \begin{bmatrix} & & \\ & X_{ij} \sim \mathcal{N}(0,1) & \\ & & \end{bmatrix}_{n\times d}$$

Points in General Position

- Pick any subset P_S of P

$$\begin{bmatrix} * & * & * & * & * \\ * & * & P_S & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}_{s \times n} \begin{bmatrix} \\ X_{ij} \sim \mathcal{N}(0,1) \\ \\ \end{bmatrix}_{n \times d}$$

- Let $Y_S = P_S X$.
- Observation 1: Y_S is a correlated Gaussian matrix.
- Using the *stability* property of Gaussian Distribution, we can show that

$$\frac{\det(\mathrm{Y_S Y_S^\top})}{\det(\mathrm{P_S P_S^\top})} \sim \det(\mathrm{X_S X_S^\top})$$

Points in General Position

- $\left(\frac{\mathrm{vol}(Y_S)}{\mathrm{vol}(P_S)}\right)^{\frac{1}{s}} = \left(\frac{\sqrt{\det(Y_S Y_S^\top)}/s!}{\sqrt{\det(P_S P_S^\top)}/s!}\right)^{\frac{1}{s}} = \left(\det(X_S X_S^\top)\right)^{\frac{1}{2s}} \approx \chi^2_{sd}$
- Same distribution as n-simplex case.
- Observation 2: (Normalized) volume distribution independent of P_S.

Conclusion:

- For points in general position we have the desired concentration.
- Therefore, by union bound there exists a volume preserving mapping to $\mathbb{R}^d$.

Why k/ε on $d = \mathcal{O}(\max\{k/\varepsilon, \varepsilon^{-2}\log n\})$?

- Let μ_s be the expected normalized volume of subsets of size s.
- μ_s is decreasing in s, $\mu_s \leq \sqrt{d - s/2}$.
- Also, $\mu_2 \geq \sqrt{d-1}$.
- We must satisfy:

$$\sqrt{\frac{d - k/2}{d-1}} \geq 1 - \mathcal{O}(\varepsilon)$$

- Therefore, $d = \Omega(k/\varepsilon)$.

Digestion slide

	Distance (JL)	Volume (This work)
	$\lvert S\rvert = s$	
Parameter	$s = 1$	$s < k$
Quantity(Squared & normalized)	Length	Volume
	$\frac{\lVert SX\rVert^2}{\lVert S\rVert^2}$	$\left(\frac{\det(SXX^\top S^\top)}{\det(SS^\top)}\right)^{\frac{1}{s}}$
Random Variable	χ_d^2	$\left(\prod_{i=1}^{s} \chi_{d-i+1}^2\right)^{\frac{1}{s}}$
Concentration	$e^{-\Omega(d\varepsilon^2)}$	$e^{-\Omega(sd\varepsilon^2)}$
Lower Bound	[Alon, 2003]	?

Open Question - Lower Bound

Open Question

Let A be an $n \times n$ positive semidefinite matrix such that the determinant of every $s \times s$ principal minor ($s \leq k$) is between $(1-\varepsilon)^{s-1}$ to 1. Then is the rank of A at least $\min\{\Omega(k/\varepsilon), n\}$?

Summary

We discussed:

- A $O(\max\{\varepsilon^{-2}\log n, k/\varepsilon\})$ upper bound for volume-preserving (to subset of size up to k) dimension reduction.
- JL Lemma preserves more than distances. It preserves volumes of subsets of size up to $\log n/\varepsilon$.
- An open question for the lower bound.

Future work:

- Extension to $\mathbb{R}^d$, $d = o(\log n)$.

Thank You

References I

Agarwal, P. K., Har-Peled, S., and Yu, H. (2007).
Embeddings of surfaces, curves, and moving points in euclidean space.
In *SCG '07: Proceedings of the twenty-third annual symposium on Computational geometry*, pages 381–389, New York, NY, USA. ACM.

Alon, N. (2003).
Problems and results in extremal combinatorics, i.
Discrete Math., (273):31–53.

Clarkson, K. L. (2008).
Tighter bounds for random projections of manifolds.
In *SCG '08: Proceedings of the twenty-fourth annual symposium on Computational geometry*, pages 39–48, New York, NY, USA. ACM.

Gordon, L. (1989).
Bounds for the distribution of the generalized variance.
The Annals of Statistics, 17(4):1684–1692.

Liberty, E., Ailon, N., and Singer, A. (2008).
Fast random projections using lean walsh transforms.
To Appear in RANDOM.

Magen, A. (2002).
Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications.
In *RANDOM '02*, pages 239–253. Springer-Verlag.

Sarlos, T. (2006).
Improved approximation algorithms for large matrices via random projections.
In *FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 143–152, Washington, DC, USA. IEEE Computer Society.

References II

Wakin, M. and Baraniuk, R. (2006).
Random projections of signal manifolds.
Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, 5:V–V.

Extra slides

Concentration bounds for χ^2

Lemma

Let $\chi_t^2 = \sum_{i=1}^{t} X_i^2$*, where* $X_i \sim \mathcal{N}(0,1)$*. Then for every* ε*, with* $0 < \varepsilon \leq 1/2$*, we have that*

$$\Pr\left[\chi_t^2 \leq (1-\varepsilon)E[\chi_t^2]\right] \quad \leq \quad \exp(-t\frac{\varepsilon^2}{6})$$

and

$$\Pr\left[\chi_t^2 \geq (1+\varepsilon)E[\chi_t^2]\right] \quad \leq \quad \exp(-t\frac{\varepsilon^2}{6}).$$

Extra slides

Theorem ([Gordon, 1989])

Let $u_i := \chi^2_{d-i+1}$ be independent Chi-square random variables for $i = 1, 2, \ldots, s$. If u_i are independent, then the following holds for every $s \geq 1$,

$$\chi^2_{s(d-s+1)+\frac{(s-1)(s-2)}{2}} \succeq s\left(\prod_{i=1}^{s} u_i\right)^{1/s} \succeq \chi^2_{s(d-s+1)}. \tag{1}$$

- By $X \succeq Y$ we denote that the r.v. X is stochastically greater than Y.