RICE UNIVERSITY

The Johnson-Lindenstrauss Lemma Meets Compressed Sensing

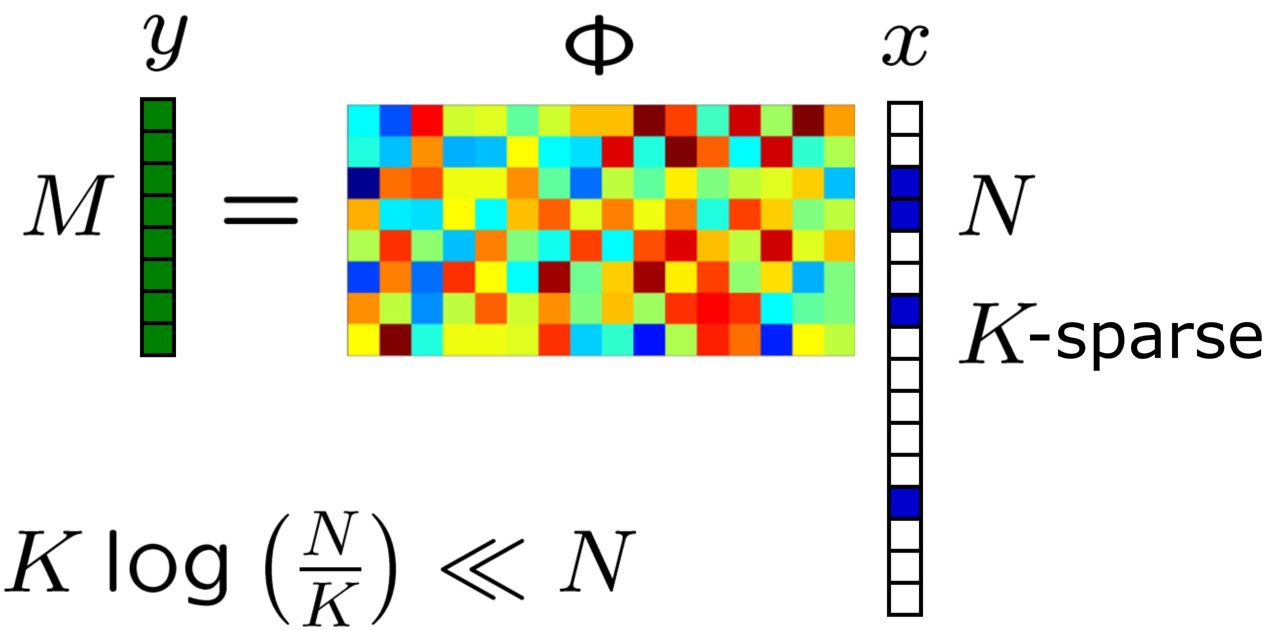
Mark Davenport

Richard Baraniuk
Ron DeVore
Michael Wakin

dsp.rice.edu/cs

Compressed Sensing (CS)

- Observe $y = \Phi x$



$$M \approx K \log\left(\frac{N}{K}\right) \ll N$$

- *Random* measurements

Randomness in CS

New signal models

New applications

Restricted Isometry Property

- What makes a “good” CS matrix?

Restricted Isometry Property

- What makes a "good" CS matrix?
- Let Σ_K denote the set of all *K*-sparse signals.

Φ has RIP of order *K* if there exists $\epsilon \in (0, 1)$ such that

$$(1-\epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1+\epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$

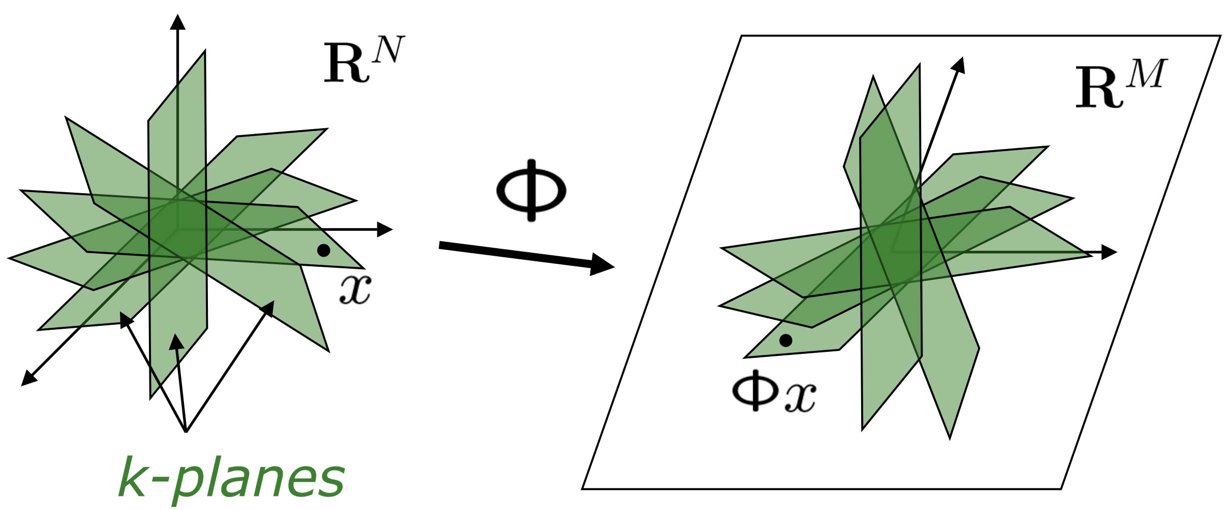
Restricted Isometry Property

- What makes a "good" CS matrix?
- Let Σ_K denote the set of all *K*-sparse signals.

Φ has RIP of order *K* if there exists $\epsilon \in (0,1)$ such that

$$(1-\epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1+\epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$



Proof of RIP

- Random matrix will satisfy RIP for largest possible *K* with high probability
  [Candès, Tao; Donoho]

Proof of RIP

- Random matrix will satisfy RIP for largest possible *K* with high probability
  [Candès, Tao; Donoho]

- Appeal to known results on singular values of random matrices [Davidson, Szarek; Litvak, Pajor, Rudelson, Tomczak-Jaegermann]

Proof of RIP

- Random matrix will satisfy RIP for largest possible *K* with high probability
  [Candès, Tao; Donoho]

- Appeal to known results on singular values of random matrices [Davidson, Szarek; Litvak, Pajor, Rudelson, Tomczak-Jaegermann]

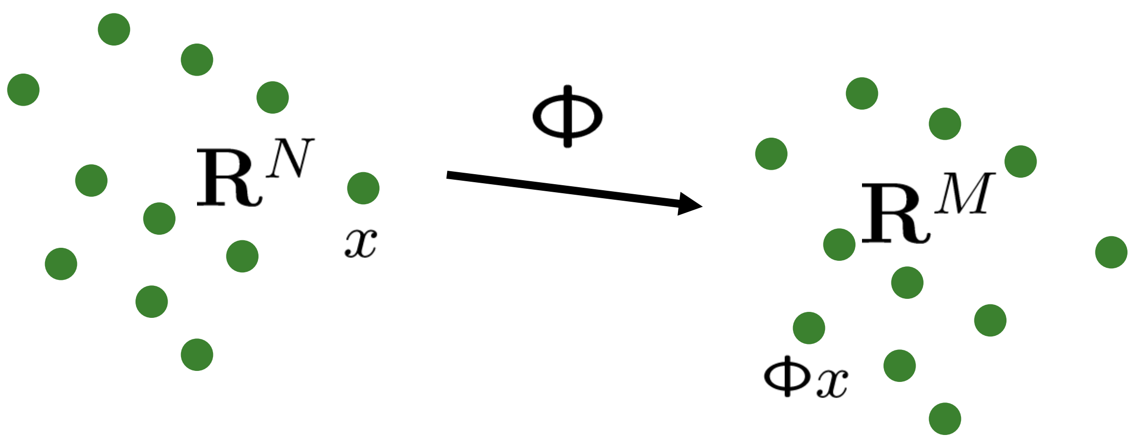
- This is not light reading…

“Proof” of RIP

“It uses a lot of newer mathematical techniques, things that were developed in the 80's and 90's. Noncommutative geometry, random matrices … the proof is very… hip.” - Hal

Dimensionality Reduction

- Point dataset lives in high-dimensional space
- Number of data points is small
- Compress data to few dimensions
- We do not lose information – can *distinguish data points*



Johnson-Lindenstrauss Lemma

Let $\epsilon \in (0,1)$ be given. For every set Q of $|Q|$ points in $\mathbb{R}^N$, if

$$M = O\left(\frac{\log(|Q|/\delta)}{\epsilon^2}\right),$$

a randomly drawn $M \times N$ matrix Φ will satsify

$$(1-\epsilon)\|u-v\|_{\ell_2^N}^2 \leq \|\Phi u - \Phi v\|_{\ell_2^M}^2 \leq (1+\epsilon)\|u-v\|_{\ell_2^N}^2$$

for all $u, v \in Q$ with probability at least $1-\delta$.

Johnson-Lindenstrauss Lemma

Let $\epsilon \in (0,1)$ be given. For every set Q of $|Q|$ points in $\mathbb{R}^N$, if

$$M = O\left(\frac{\log(|Q|/\delta)}{\epsilon^2}\right),$$

a randomly drawn $M \times N$ matrix Φ will satsify

$$(1-\epsilon)\|u-v\|_{\ell_2^N}^2 \leq \|\Phi u - \Phi v\|_{\ell_2^M}^2 \leq (1+\epsilon)\|u-v\|_{\ell_2^N}^2$$

for all $u, v \in Q$ with probability at least $1-\delta$.

- Proof relies on a simple concentration of measure inequality

$$\mathbf{P}(|\,\|\Phi x\|_{\ell_2^M}^2 - \|x\|_{\ell_2^N}^2| \geq \epsilon \|x\|_{\ell_2^N}^2) \leq 2e^{-M\epsilon^2/4}$$

Favorable JL Distributions

- Gaussian

$$\phi_{i,j} \sim \mathcal{N}\left(0, \frac{1}{M}\right)$$

- Bernoulli [Achlioptas]

$$\phi_{i,j} := \begin{cases} +\frac{1}{\sqrt{M}} & \text{with probability} \quad \frac{1}{2}, \\ -\frac{1}{\sqrt{M}} & \text{with probability} \quad \frac{1}{2} \end{cases}$$

Favorable JL Distributions

- "Database-friendly" [Achlioptas]

$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6} \end{cases}$$

- Fast JL Transform [Ailon, Chazelle]

$$\Phi = PHD$$

P : Sparse Gaussian matrix
H : Fast Hadamard transform
D : Random modulation

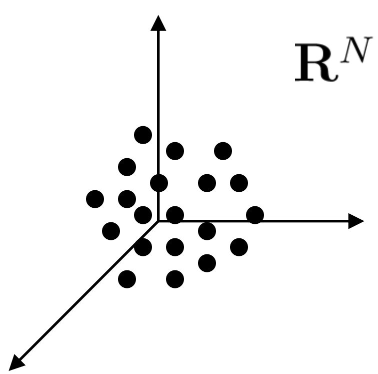
JL Meets CS [Baraniuk, DeVore, Davenport, Wakin]

- **Theorem**: Supposing Φ is drawn from a JL-favorable distribution, then with probability at least 1 - δ, Φ meets the RIP with $M = O(K \log(N/K))$.

JL Meets CS [Baraniuk, DeVore, Davenport, Wakin]

- **Theorem**: Supposing Φ is drawn from a JL-favorable distribution, then with probability at least 1 - δ, Φ meets the RIP with $M = O(K \log(N/K))$.

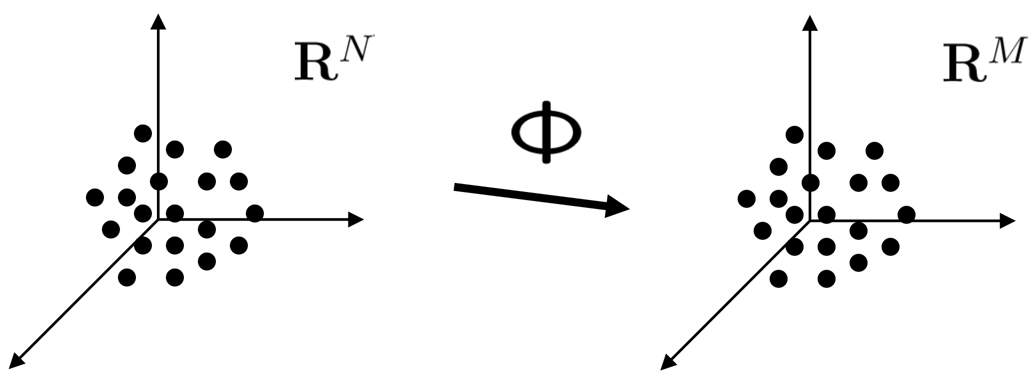
- Key idea
  - construct a set of points Q



JL Meets CS [Baraniuk, DeVore, Davenport, Wakin]

- **Theorem**: Supposing Φ is drawn from a JL-favorable distribution, then with probability at least 1 - δ, Φ meets the RIP with $M = O(K \log(N/K))$.

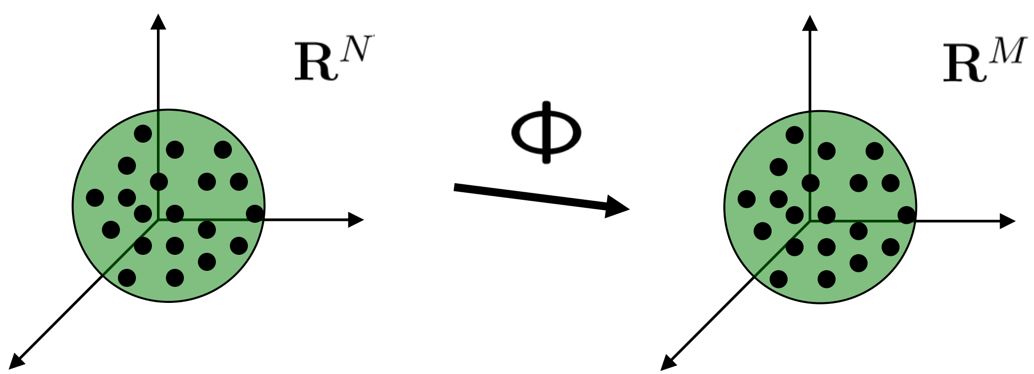
- Key idea
  - construct a set of points Q
  - apply JL lemma (union bound on concentration of measure)



JL Meets CS [Baraniuk, DeVore, Davenport, Wakin]

- **Theorem**: Supposing Φ is drawn from a JL-favorable distribution, then with probability at least 1 - δ, Φ meets the RIP with $M = O(K \log(N/K))$.

- Key idea
  - construct a set of points Q
  - apply JL lemma (union bound on concentration of measure)
  - show that isometry on Q extends to isometry on Σ_K



Recall: RIP

Φ has RIP of order *K* if there exists $\epsilon \in (0, 1)$ such that

$$(1 - \epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1 + \epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$

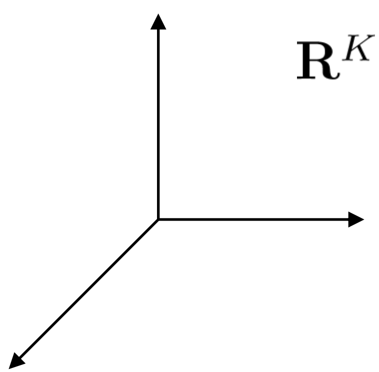
Recall: RIP

Φ has RIP of order *K* if there exists $\epsilon \in (0,1)$ such that

$$(1-\epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1+\epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$

- Fix a *K*-dimensional subspace



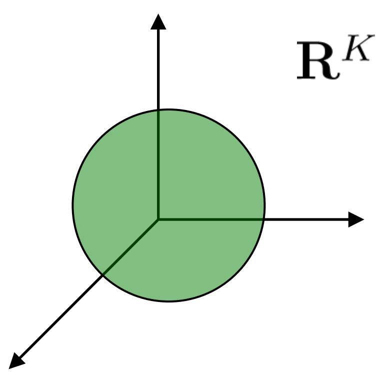
Recall: RIP

Φ has RIP of order *K* if there exists $\epsilon \in (0,1)$ such that

$$(1-\epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1+\epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$

- Fix a *K*-dimensional subspace
- Consider only $\|x\|_{\ell_2} \leq 1$



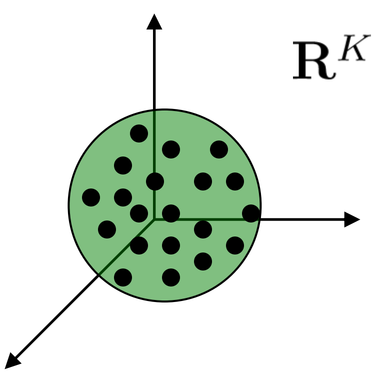
Recall: RIP

Φ has RIP of order *K* if there exists $\epsilon \in (0,1)$ such that

$$(1-\epsilon)\|x\|_{\ell_2}^2 \leq \|\Phi x\|_{\ell_2}^2 \leq (1+\epsilon)\|x\|_{\ell_2}^2$$

for all $x \in \Sigma_K$

- Fix a *K*-dimensional subspace
- Consider only $\|x\|_{\ell_2} \leq 1$



Pick *Q* such that for any x there exists a q such that

$$\|x-q\|_{\ell_2} \leq \frac{\epsilon}{4}$$

Bootstrapping

- Apply JL to get

$$(1-\epsilon/2)\|q\|_{\ell_2} \leq \|\Phi q\|_{\ell_2} \leq (1+\epsilon/2)\|q\|_{\ell_2}$$

Bootstrapping

- Apply JL to get

$$(1-\epsilon/2)\|q\|_{\ell_2} \leq \|\Phi q\|_{\ell_2} \leq (1+\epsilon/2)\|q\|_{\ell_2}$$

- Define *A* to be the smallest number such that

$$\|\Phi x\|_{\ell_2} \leq (1+A)\|x\|_{\ell_2}$$

for all x with $\|x\|_{\ell_2} \leq 1$

Bootstrapping

- Apply JL to get

$$(1-\epsilon/2)\|q\|_{\ell_2} \leq \|\Phi q\|_{\ell_2} \leq (1+\epsilon/2)\|q\|_{\ell_2}$$

- Define *A* to be the smallest number such that

$$\|\Phi x\|_{\ell_2} \leq (1+A)\|x\|_{\ell_2}$$

for all x with $\|x\|_{\ell_2} \leq 1$

- For any x, pick the closest q

$$\begin{aligned}\|\Phi x\|_{\ell_2} &\leq \|\Phi q\|_{\ell_2} + \|\Phi(x-q)\|_{\ell_2} \\ &\leq 1+\epsilon/2+(1+A)\epsilon/4\end{aligned}$$

Bootstrapping

- Apply JL to get

$$(1-\epsilon/2)\|q\|_{\ell_2} \leq \|\Phi q\|_{\ell_2} \leq (1+\epsilon/2)\|q\|_{\ell_2}$$

- Define *A* to be the smallest number such that

$$\|\Phi x\|_{\ell_2} \leq (1+A)\|x\|_{\ell_2}$$

for all x with $\|x\|_{\ell_2} \leq 1$

- For any x, pick the closest q

$$\begin{aligned}\|\Phi x\|_{\ell_2} &\leq \|\Phi q\|_{\ell_2} + \|\Phi(x-q)\|_{\ell_2} \\ &\leq 1+\epsilon/2+(1+A)\epsilon/4\end{aligned}$$

- Hence $1+A \leq 1+\epsilon/2+(1+A)\epsilon/4 \;\Rightarrow\; A \leq \epsilon$

How Many Points?

- For each *K*-dimensional space, we need $(12/\epsilon)^K$

How Many Points?

- For each *K*-dimensional space, we need $(12/\epsilon)^K$
- $\binom{N}{K} \leq \left(\frac{eN}{K}\right)^K$ spaces to consider

How Many Points?

- For each *K*-dimensional space, we need $(12/\epsilon)^K$
- $\binom{N}{K} \leq \left(\frac{eN}{K}\right)^K$ spaces to consider
- How many measurements do we need to get RIP with probability at least $1 - \delta$?

How Many Points?

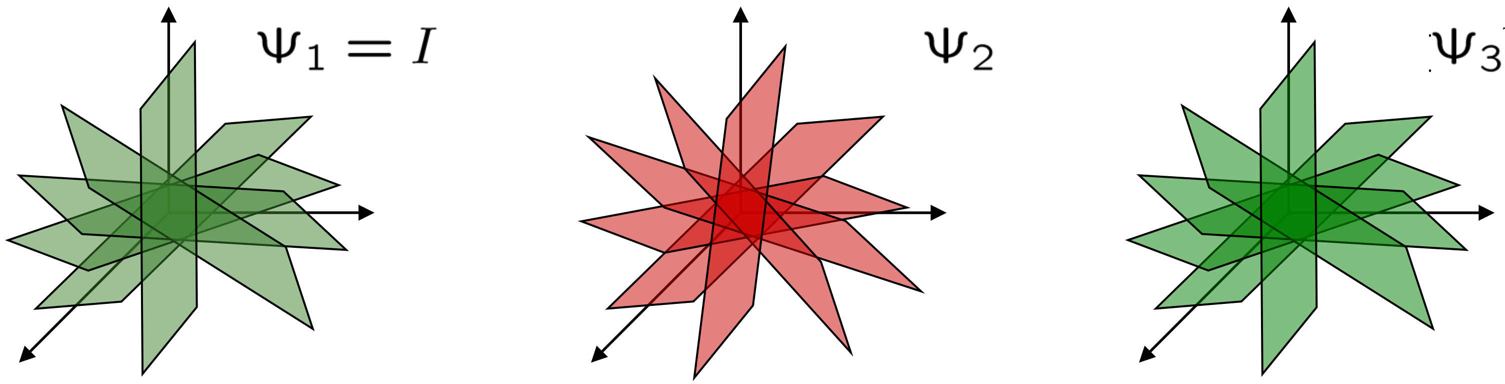
- For each K-dimensional space, we need $(12/\epsilon)^K$
- $\binom{N}{K} \leq \left(\frac{eN}{K}\right)^K$ spaces to consider
- How many measurements do we need to get RIP with probability at least $1-\delta$?

$$M = O\left(\frac{\log(|Q|/\delta)}{\epsilon^2}\right)$$

$$= C_{\epsilon,\delta} K \log(N/K)$$

Universality

- Easy to see why random matrices are universal with respect to sparsity basis



- Resample your points in new basis – JL provides guarantee for *arbitrary* set of points
  - Gaussian
  - Bernoulli
  - Others…

Summary

- Better understanding of the relevant geometry
  - provides simple proofs of key CS / n-width results

Summary

- Better understanding of the relevant geometry
  - provides simple proofs of key CS / n-width results
- New conditions on what it takes to be a good CS matrix
  - concentration of measure around the mean

Summary

- Better understanding of the relevant geometry
  - provides simple proofs of key CS / n-width results
- New conditions on what it takes to be a good CS matrix
  - concentration of measure around the mean
- New signal models
  - manifolds

Summary

- Better understanding of the relevant geometry
  - provides simple proofs of key CS / *n*-width results
- New conditions on what it takes to be a good CS matrix
  - concentration of measure around the mean
- New signal models
  - manifolds
- Natural setting for studying *information scalability*
  - detection
  - estimation
  - learning

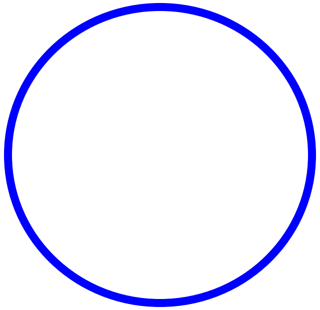
Randomness in CS

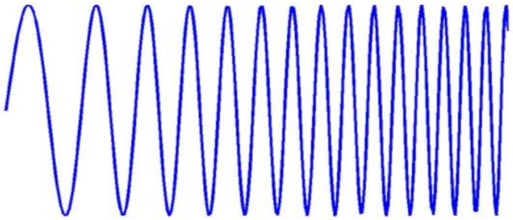
New signal models

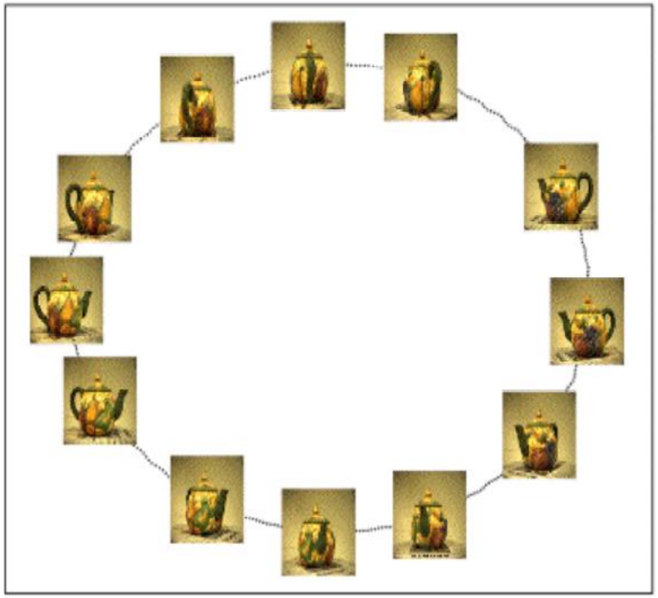
New applications

Manifold Compressive Sensing

- Locally Euclidean topological space
- Typically for signal processing
  - nonlinear K-dimensional "surface" in signal space $\boldsymbol{R}^N$
  - potentially very low dimensional signal model
- Examples (all nonlinear)
  - chirps
  - modulation schemes
  - image articulations







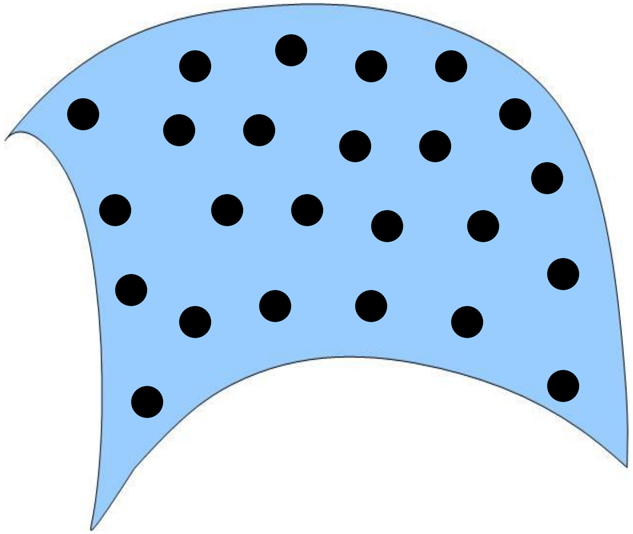
Stable Manifold Embedding

Stability [Wakin, Baraniuk]

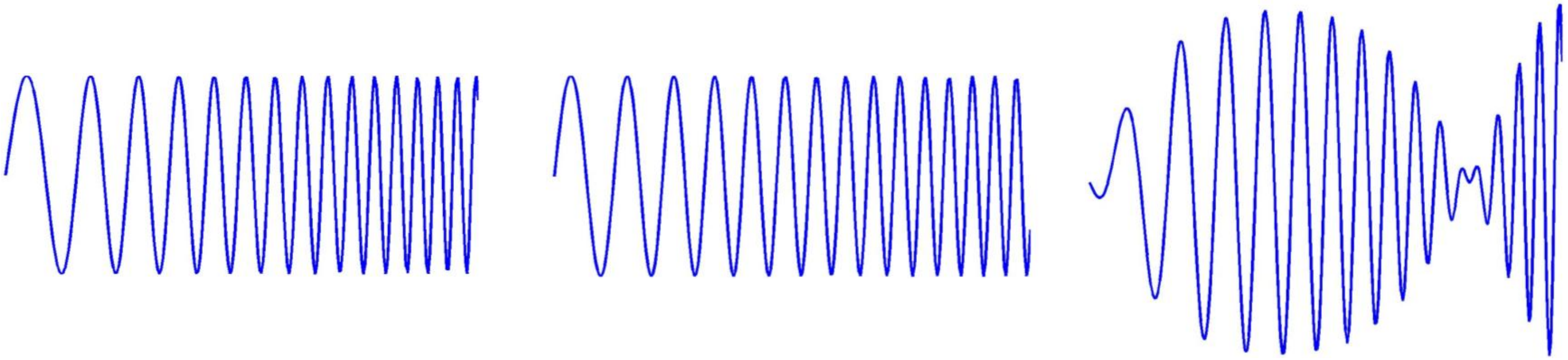
$$(1-\epsilon)\,\|x-y\|_2 \leq \|\Phi x - \Phi y\|_2 \leq (1+\epsilon)\,\|x-y\|_2$$

Number of measurements required

$$M = C_1 K \log(C_2 N)$$



Example: Linear Chirps



original **initial guess** **initial error**

N = 256

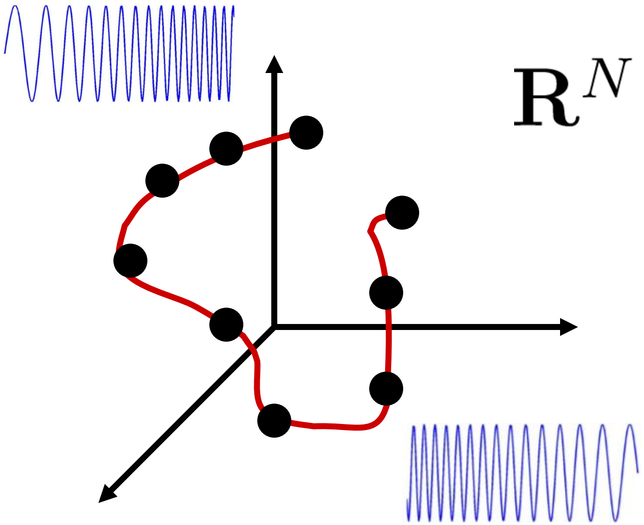
K = 2 (start & end frequencies)

M = 5: 55% success

M = 30: 99% success

Manifold Learning

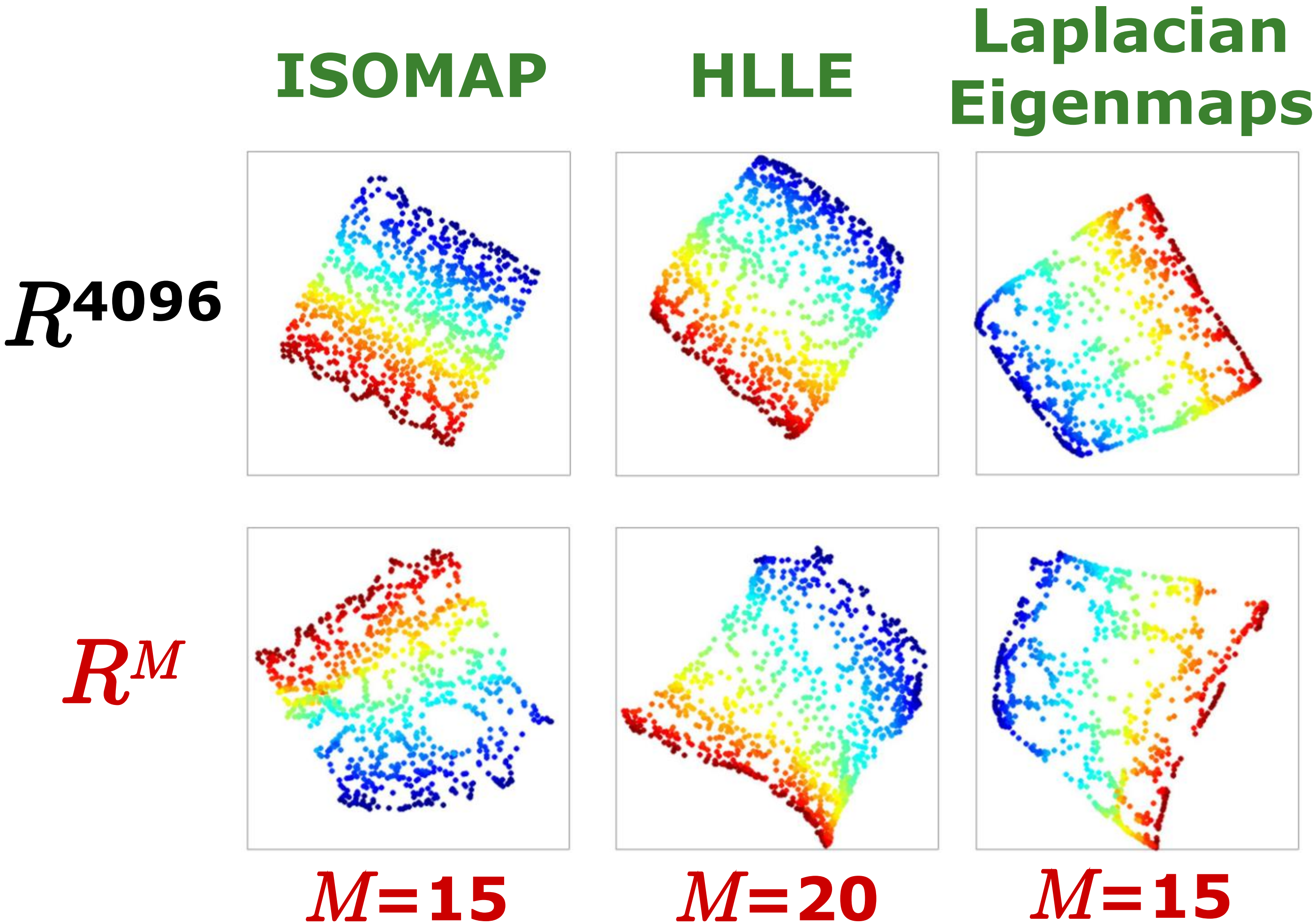
- Manifold learning algorithms for *sampled data in R^N*
  - ISOMAP, LLE, HLLE, etc.



- *Stable embedding preserves key properties in R^M*
  - ambient and geodesic distances
  - dimension and volume of the manifold
  - path lengths and curvature
  - topology, local neighborhoods, and angles
  - etc…

- Can we learn these properties from *projections in R^M*?
  - savings in computation, storage, acquisition costs

Example: Manifold Learning



Randomness in CS

New signal models

New applications

Detection – Matched Filter

$$H_0 \;:\; x = n$$
$$H_1 \;:\; x = s + n$$

- Testing for presence of a known signal *s*
- *Sufficient statistic* for detecting *s*:

$$t = \langle x, s \rangle$$

Compressive Matched Filter

$$H_0 \ : \ x = n$$
$$H_1 \ : \ x = s + n$$

- Now suppose we have CS measurements $y = \Phi x$
  - when Φ is an orthoprojector, Φn remains white noise
  - new sufficient statistic is simply the ***compressive matched filter (smashed filter?)***

$$t' = \langle y, \Phi s \rangle$$

CMF – Performance

- ROC curve for Neyman-Pearson detector:

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|\Phi s\|_2}{\sigma}\right)$$

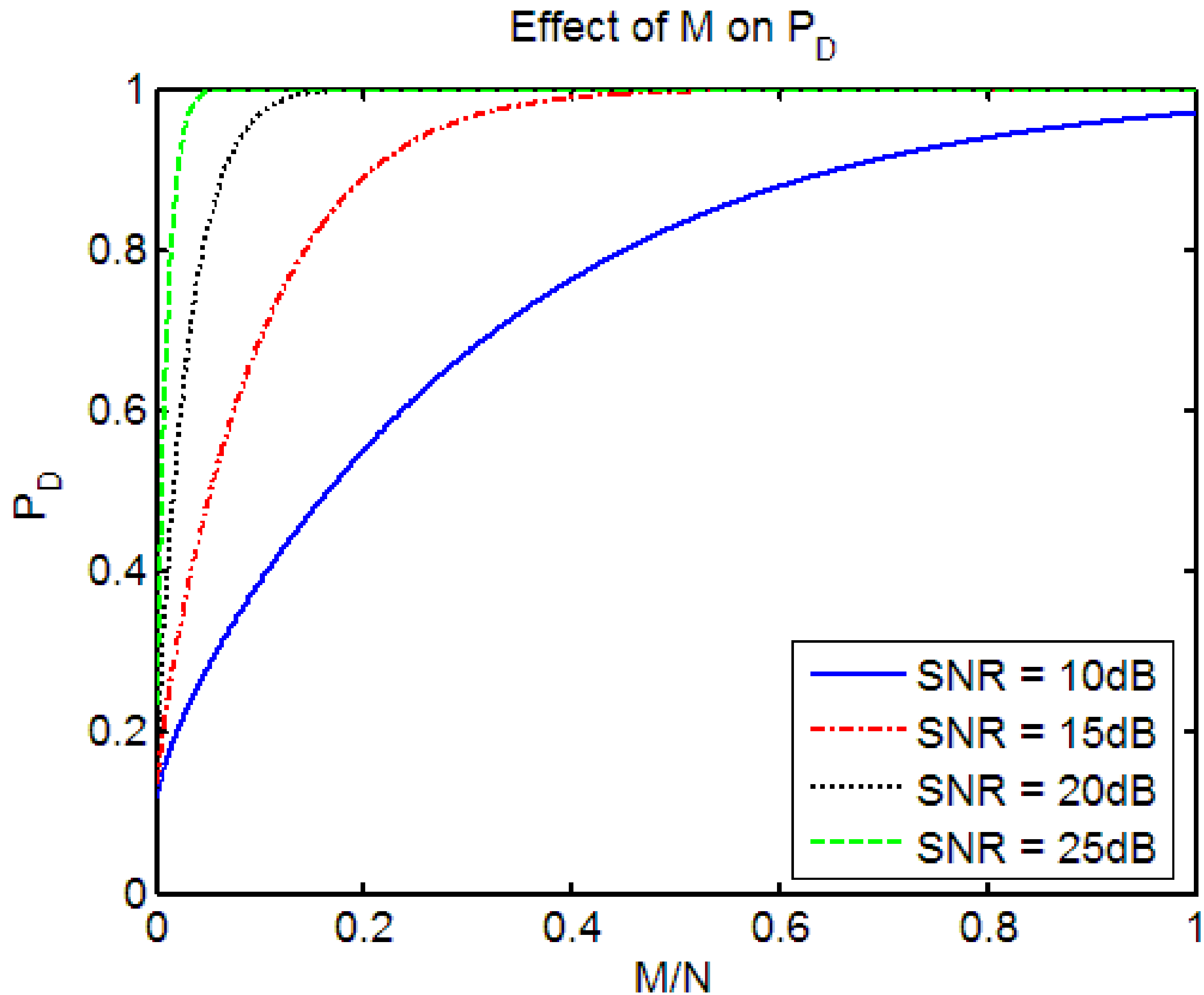
- From JL lemma, for random orthoprojector Φ

$$\|\Phi s\|_2 \approx \sqrt{\frac{M}{N}}\|s\|_2$$

- Thus

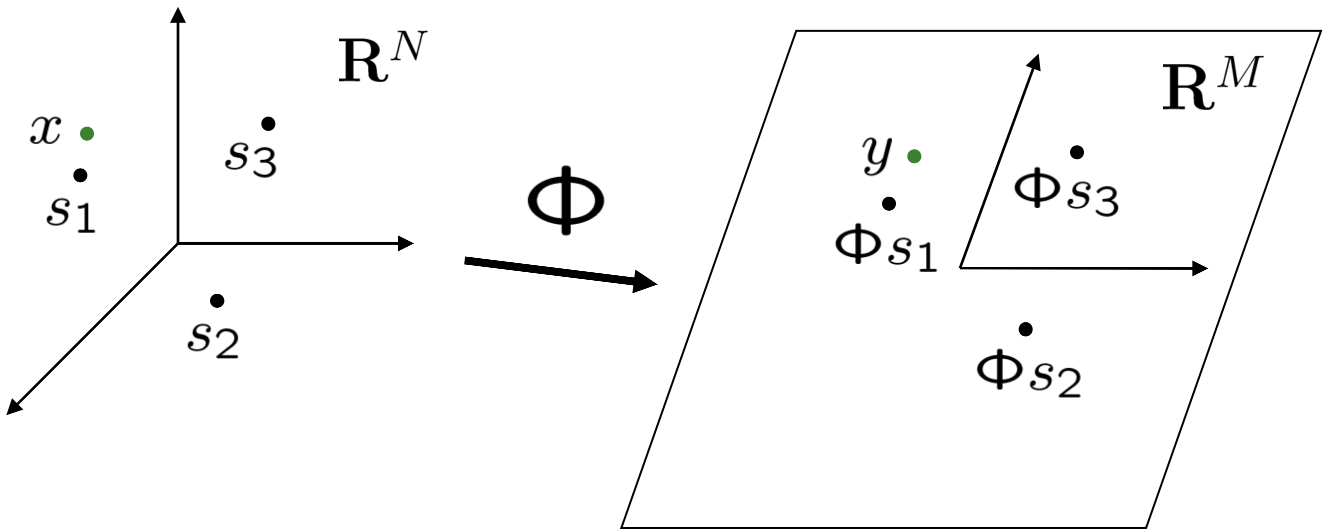
$$P_D(\alpha) \approx Q\left(Q^{-1}(\alpha) - \sqrt{\frac{M}{N}}\frac{\|s\|_2}{\sigma}\right)$$

CMF – Performance



Generalization – Classification

- More generally, suppose we want to classify between several possible signals



$$t_1 = \|y - \Phi s_1\|_2$$
$$t_2 = \|y - \Phi s_2\|_2$$
$$t_3 = \|y - \Phi s_3\|_2$$

by the JL Lemma these distances are preserved

CMF as an Estimator

- How well does the compressive matched filter estimate the output of the true matched filter?

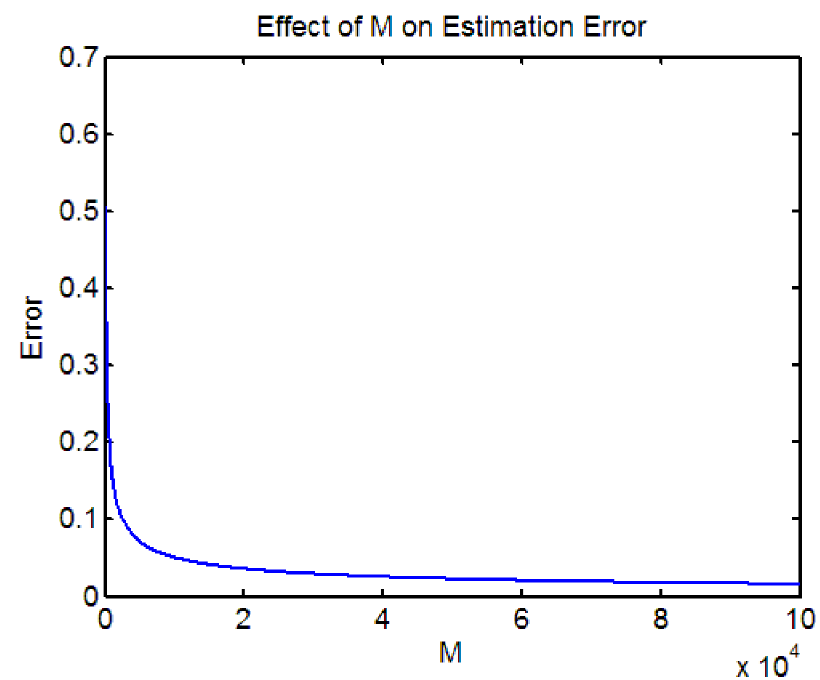
With probability at least $1 - \delta$

$$|\langle \Phi x, \Phi s \rangle - \langle x, s \rangle| \leq \kappa_\delta \frac{\|x\|_2 \, \|s\|_2}{\sqrt{M}}$$

where

$$\kappa_\delta = 2\sqrt{12 \log\left(\frac{6}{\delta}\right)}$$

[Alon, Gibbons, Matias, Szegedy; Davenport, Baraniuk, Wakin]



Conclusions

- Random matrices work for CS for the same reason that they work for the JL lemma

Conclusions

- Random matrices work for CS for the same reason that they work for the JL lemma
- Another way to characterize "good" CS matrices

Conclusions

- Random matrices work for CS for the same reason that they work for the JL lemma
- Another way to characterize “good” CS matrices
- Allows us to extend CS to new signal models
    - manifold/parametric models

Conclusions

- Random matrices work for CS for the same reason that they work for the JL lemma
- Another way to characterize “good” CS matrices
- Allows us to extend CS to new signal models
  - manifold/parametric models
- Allows us to extend CS to new settings
  - detection
  - classification/learning
  - estimation