Compressive Sensing
Part IV: Beyond Sparsity

Mark A. Davenport

Stanford University
Department of Statistics

Beyond Sparsity

- Not all signal models fit neatly into the “sparse” setting

- The concept of “dimension” has many incarnations
  - “degrees of freedom”
  - constraints
  - parameterizations
  - signal families

- How can we exploit these low-dimensional models?

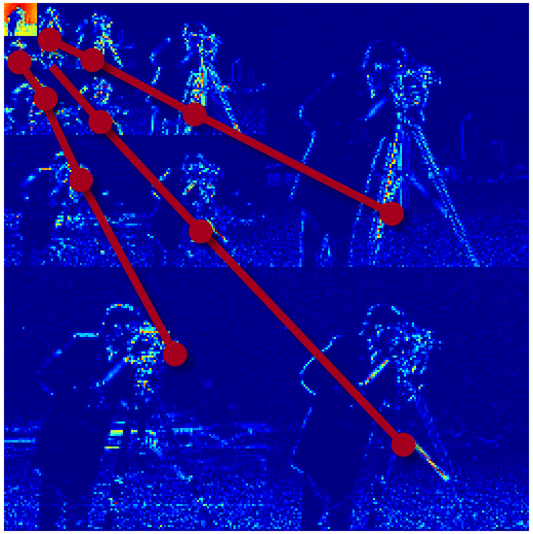
- I will focus primarily on just a few of these
  - *structured* sparsity, finite-rate-of-innovation, manifolds, low-rank matrices

Structured Sparsity

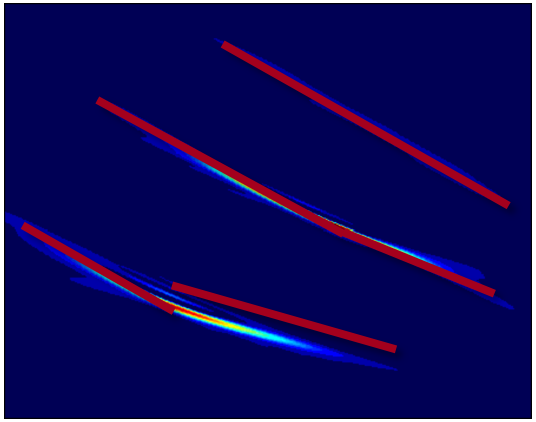
Structured Sparsity

- Sparse signal model captures
  simplistic primary structure

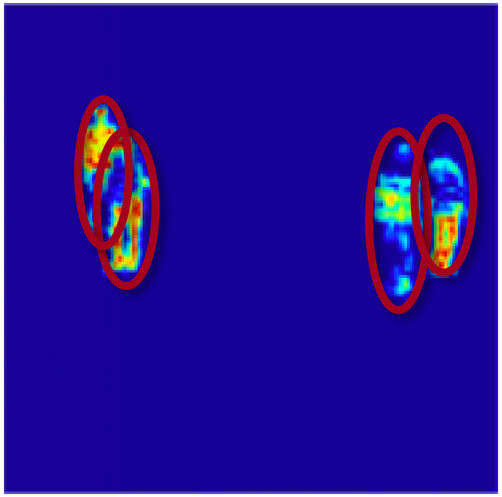
- Modern compression/processing algorithms capture
  richer secondary coefficient structure



wavelets:
natural images



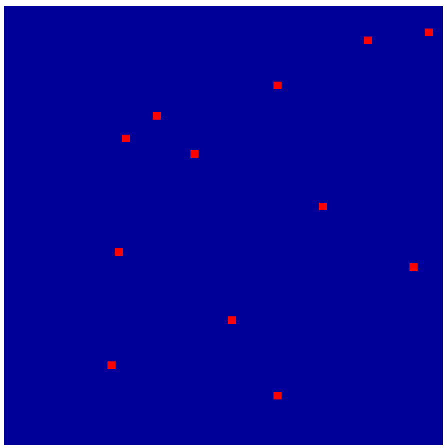
Gabor atoms:
chirps/tones

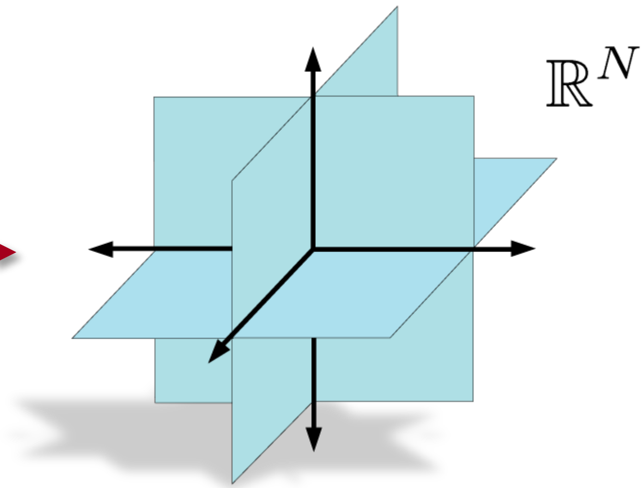


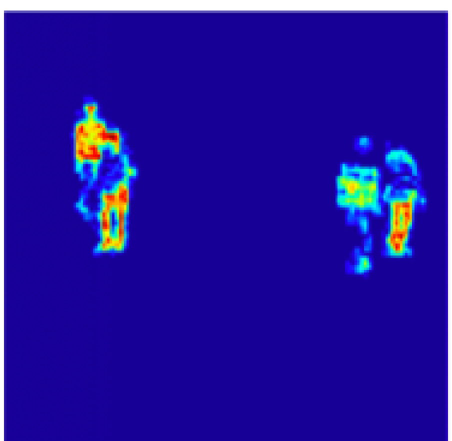
pixels:
background subtracted images

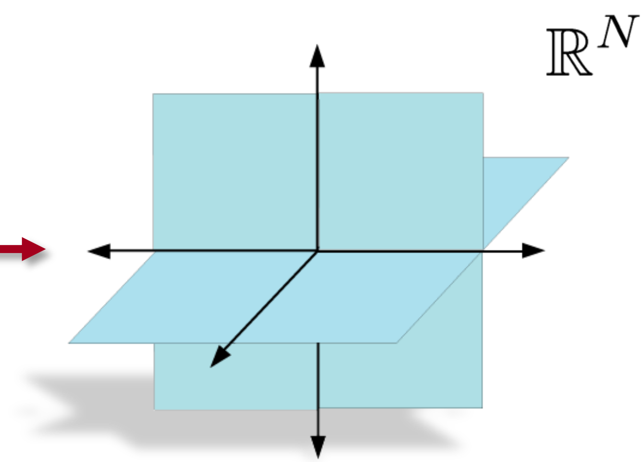
Sparse Signals

Traditional sparse models allow all possible S-dimensional subspaces



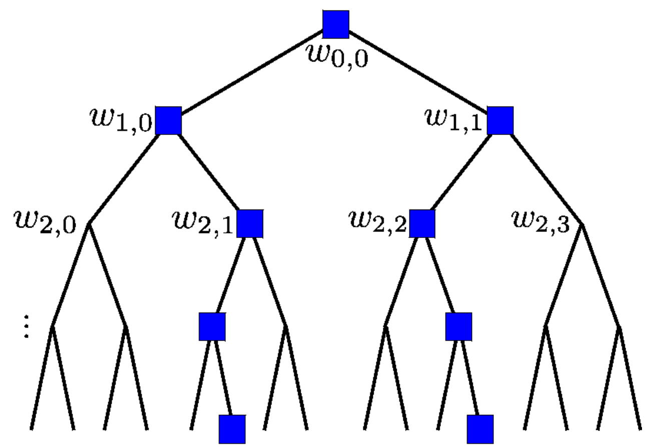


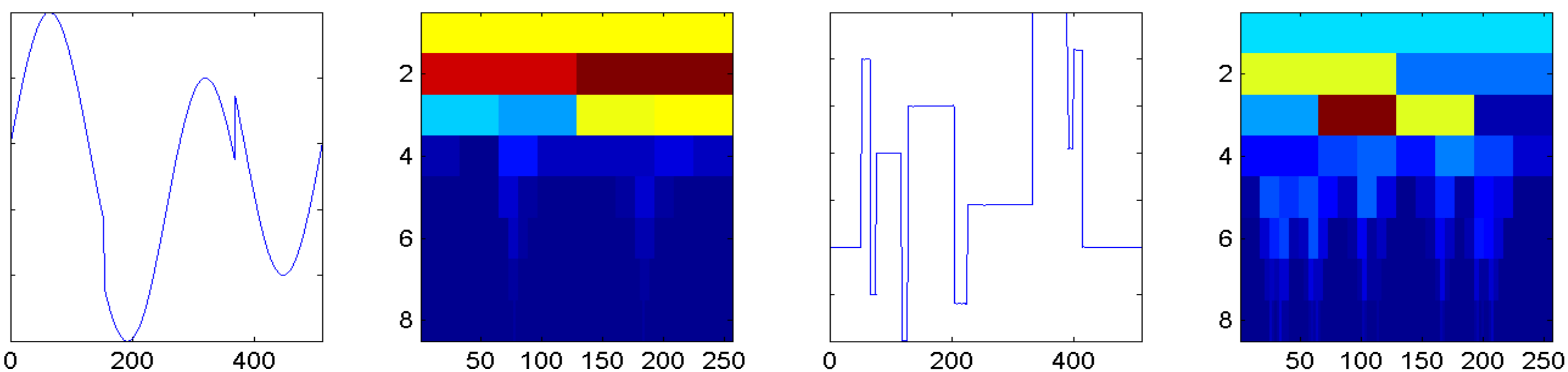




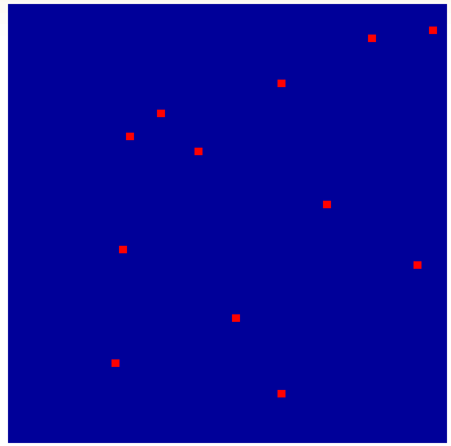
Wavelets and Tree-Sparse Signals

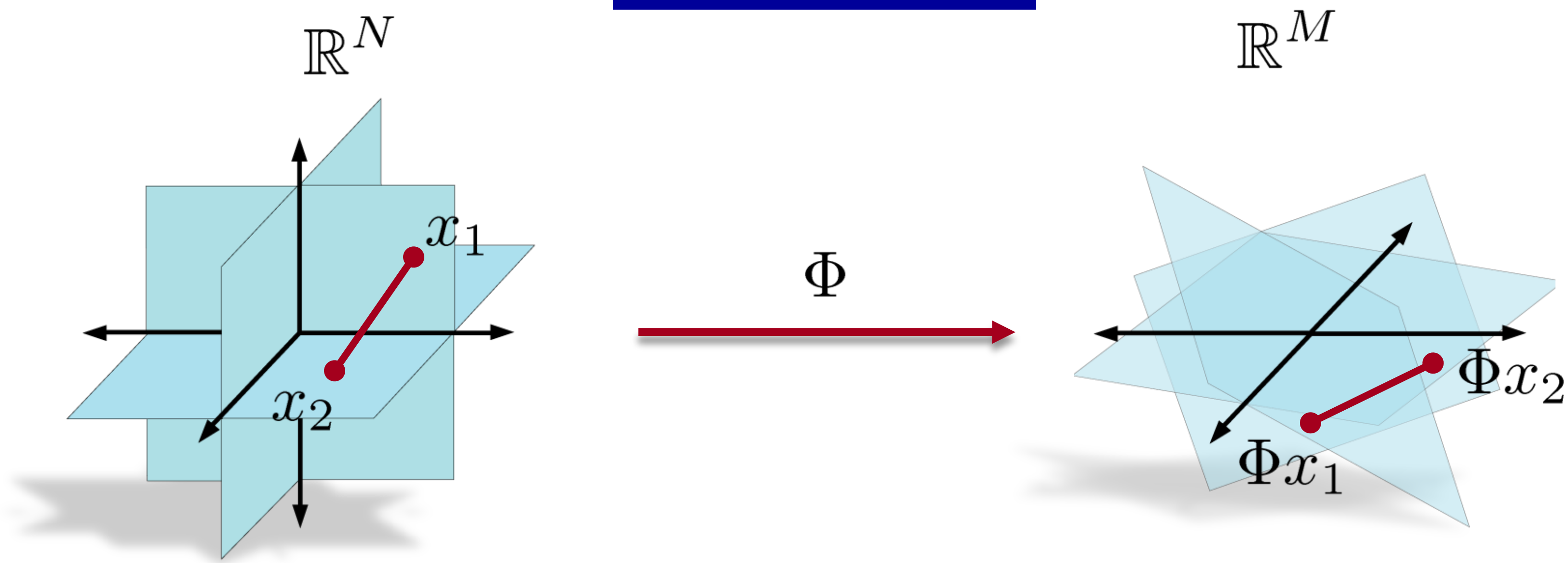
Model: S nonzero coefficients lie on a connected tree





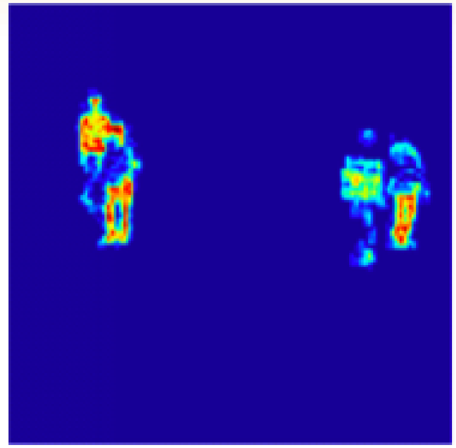
Wavelet Sparse

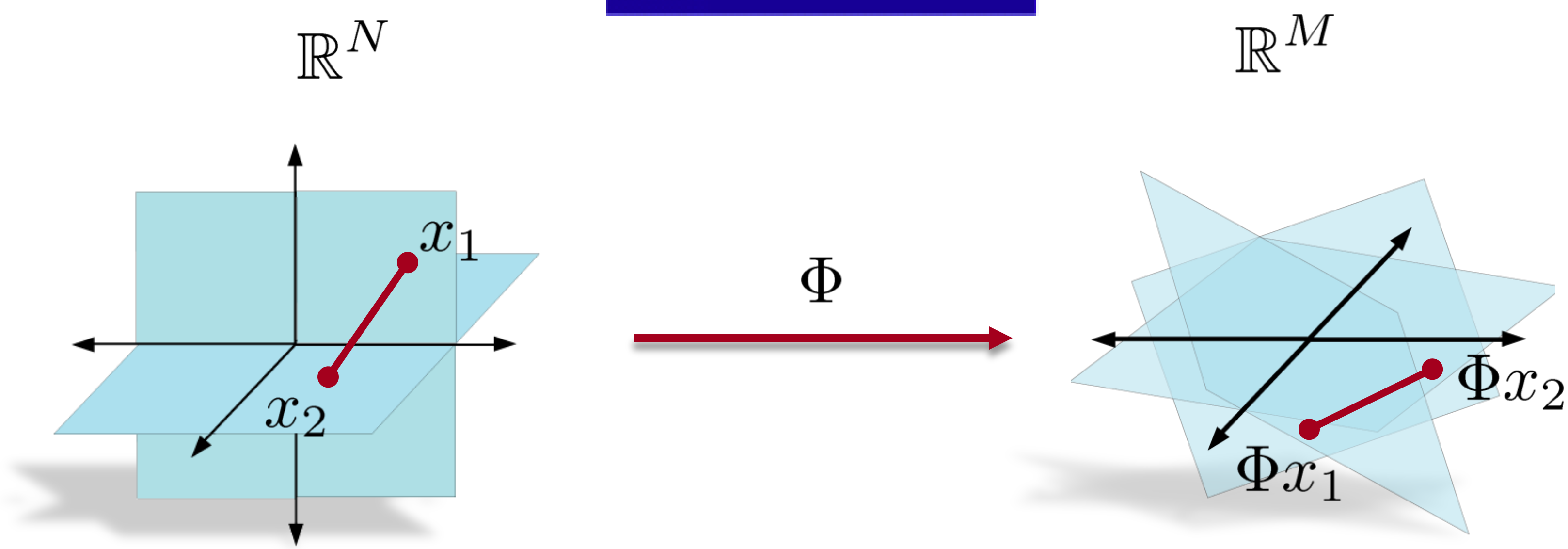




$$M = O(S \log(N/S))$$

Tree Sparse





$$M = O(S) < O(S \log(N/S))$$

Recall: CoSaMP

- The heart of CoSaMP (and many other algorithms) is *hard thresholding*

$$\widehat{x} = \mathrm{hard}(x, S)$$

- This can be viewed as a projection onto the set of all possible S-sparse signals

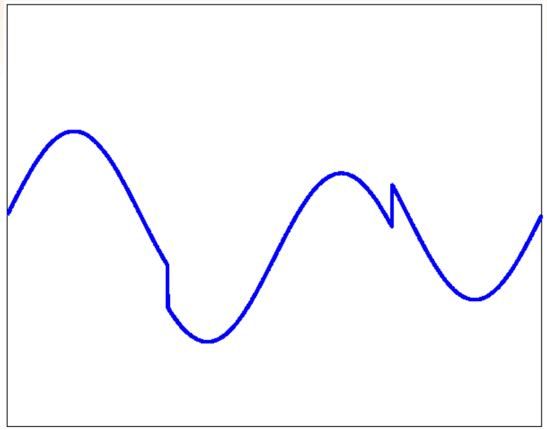
- *"Model-based CoSaMP":* Replace hard thresholding with a more suitable projection

$$\widehat{x} = \mathcal{P}_{\mathcal{M}}(x)$$

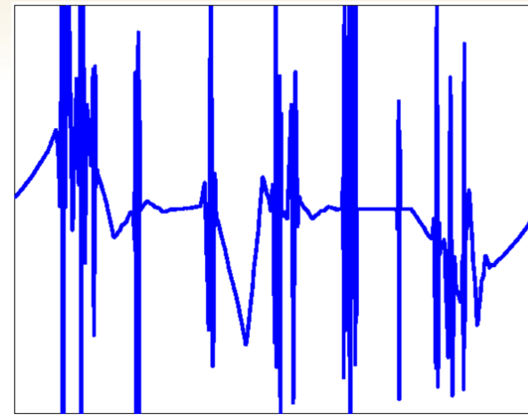
Tree-Sparse Signal Recovery

$N = 1024$

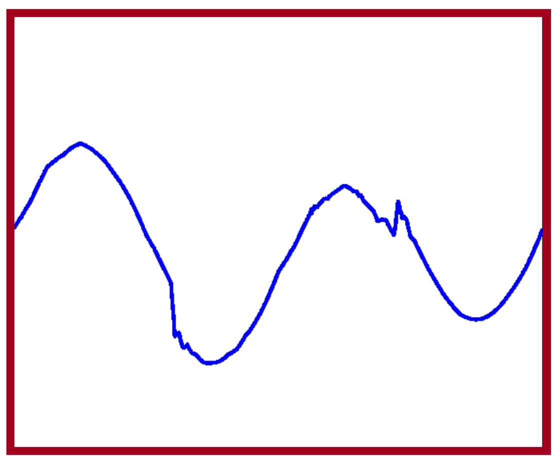
$M = 80$



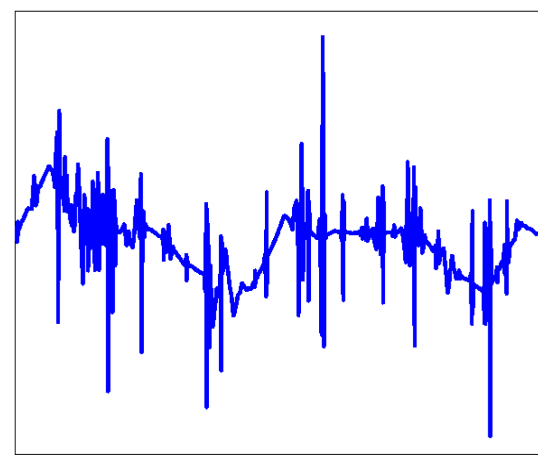
target signal



CoSaMP



Tree-sparse CoSaMP

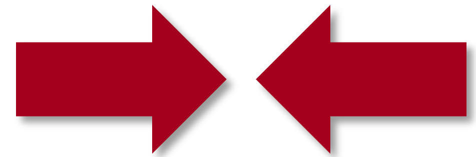


ℓ_1-minimization

[Baraniuk, Cevher, Duarte, and Hegde – 2008]

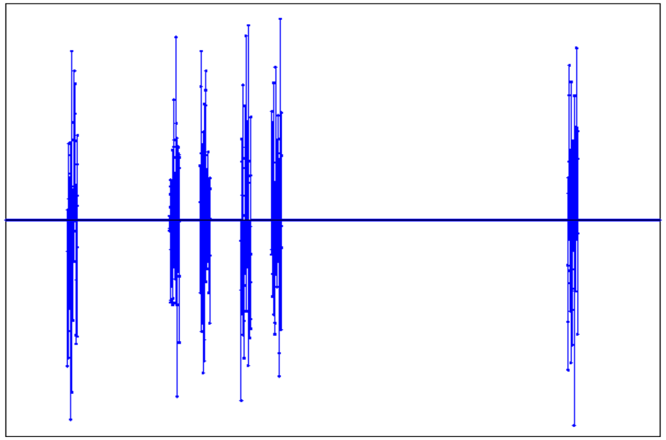
Other Useful Models

- Clustered coefficients
  - tree sparse
  - block sparse
  - Ising models

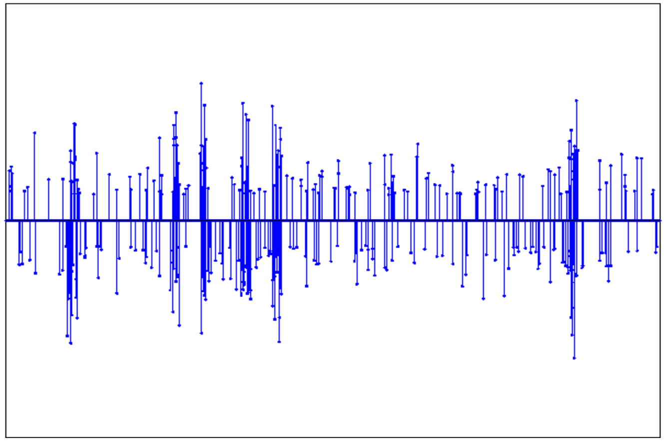


- Dispersed coefficients
  - spike trains
  - pulse trains

Block-Sparsity



target



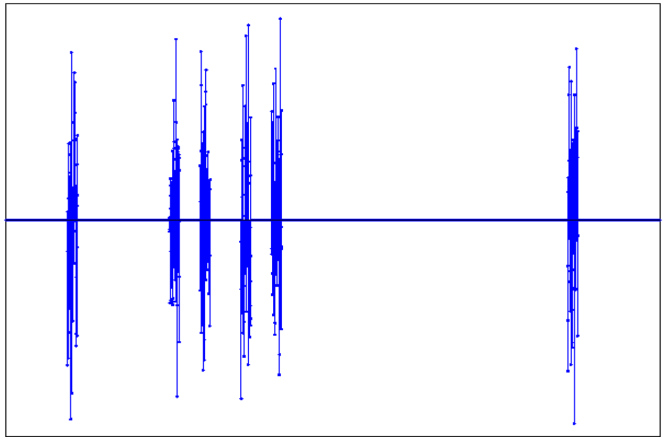
CoSaMP

$N = 4096$

$K = 6$ active blocks

$J = 64$ (block length)

$M = 2.5JK = 960$



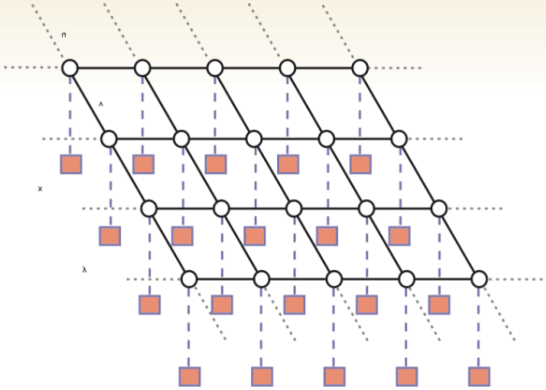
block-sparse CoSaMP

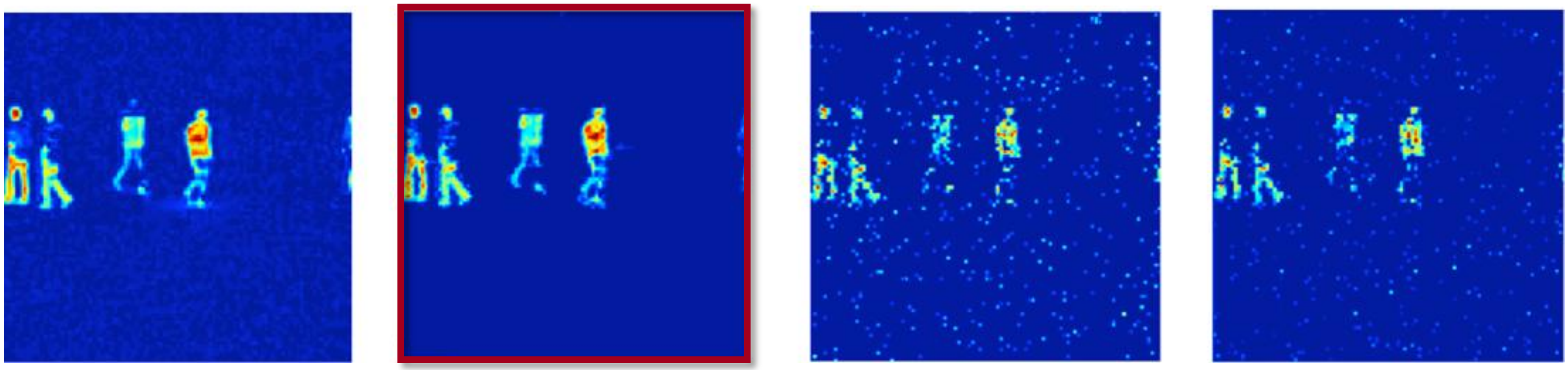
[Baraniuk, Cevher, Duarte, and Hegde – 2008]

Clustered Signals

- Probabilistic approach to modeling clustering of significant pixels using Ising Markov Random Field model

- Ising model projection can be performed efficiently using graph cuts



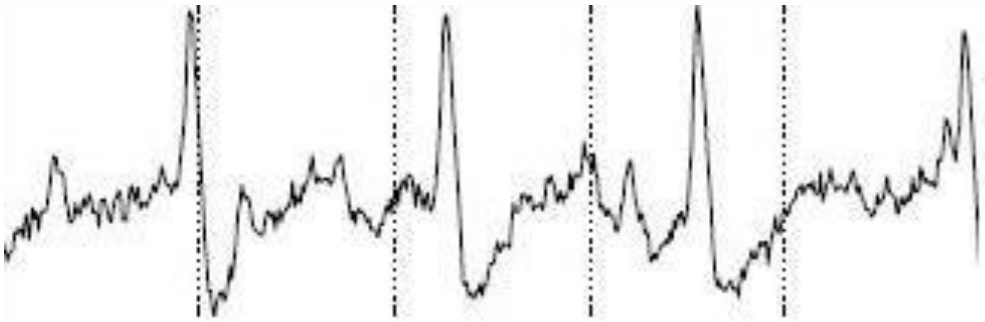


target | Ising-model recovery | CoSaMP recovery | LP (FPC) recovery

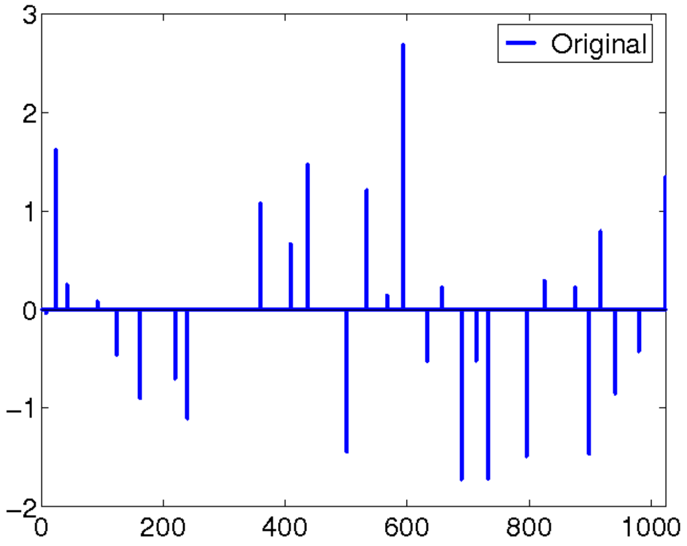
[Baraniuk, Cevher, Duarte, and Hegde – 2008]

Sparse Spike Trains

- Sequence of pulses



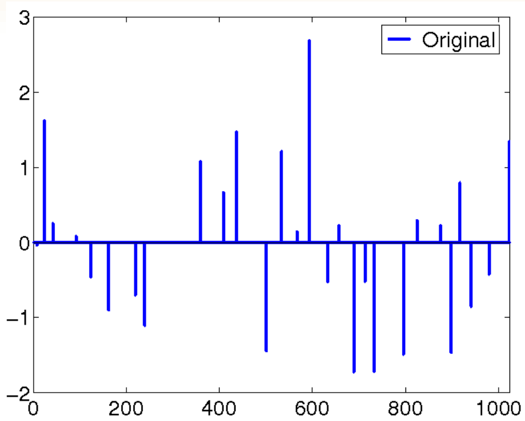
- Simple model:
  - sequence of Dirac pulses
  - *refractory period* Δ between each pulse



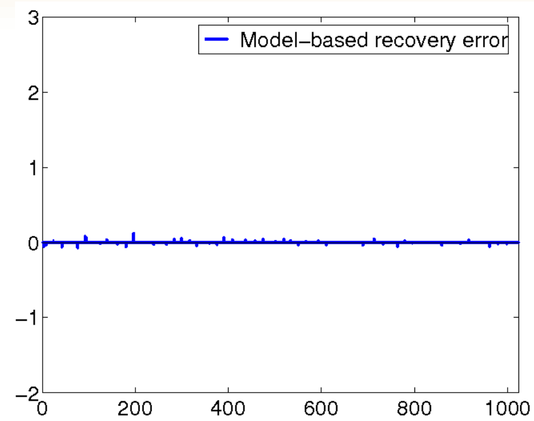
- Model-based RIP if

$$M = O(S \log(N/S - \Delta))$$

Sparse Spike Trains



target



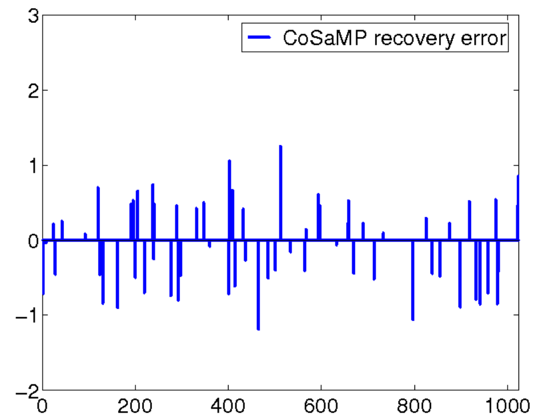
Model-based CoSaMP error

$$N = 1024$$

$$S = 50$$

$$\Delta = 10$$

$$M = 150$$



CoSaMP error

[Hegde, Duarte, and Cevher – 2009]

Sparse Pulse Trains

- More realistic model:
  - spike train convolved with a pulse shape (of length L)
  - refractory period between each pulse of length Δ

- Model-based RIP if $M = O(L + S \log(N/S - \Delta))$

$N = 4096$

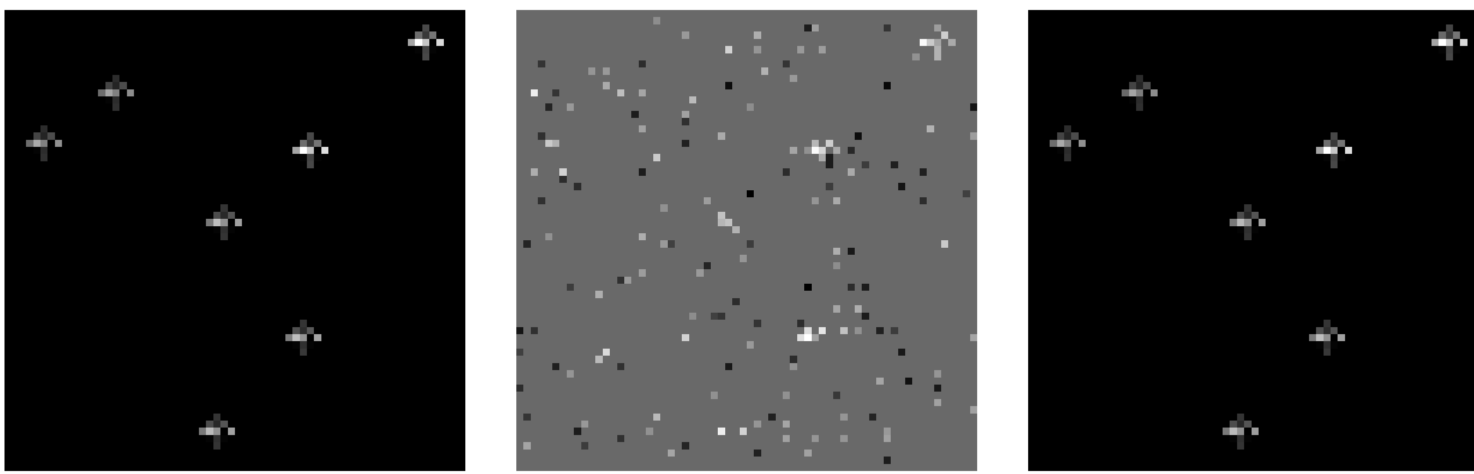
$S = 7$

$L = 25$

$\Delta = 10$

$M = 290$

original | CoSaMP | Model-based CoSaMP



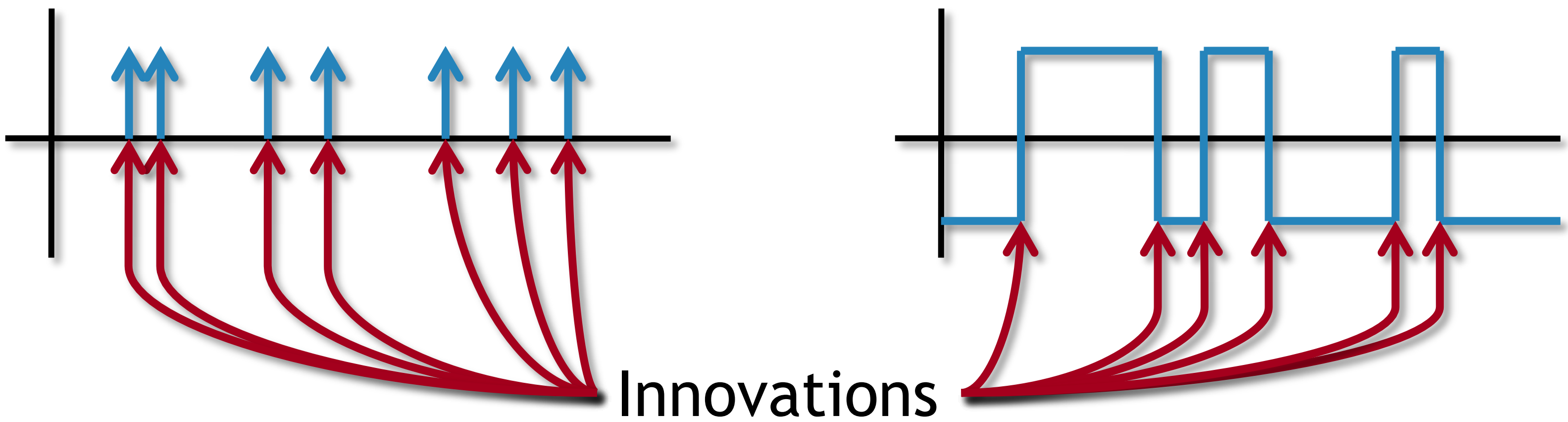
[Hegde and Baraniuk – 2010]

Parametric and Manifold Models

Finite Rate of Innovation

Continuous-time notion of sparsity: "rate of innovation"

Examples:



Rate of innovation:

Expected number of innovations per second

[Vetterli, Marziliano, and, Blu – 2002; Dragotti, Vetterli, and Blu - 2007]

Sampling Signals with FROI

We would like to obtain samples of the form

$$y[m] = \phi(t) * x(t)|_{t=mT_s} = \langle \phi(mT_s - t), x(t) \rangle$$

where we sample at the *rate of innovation.*

Requires *careful construction of sampling kernel* $\phi(t)$ *.*

Drawbacks:

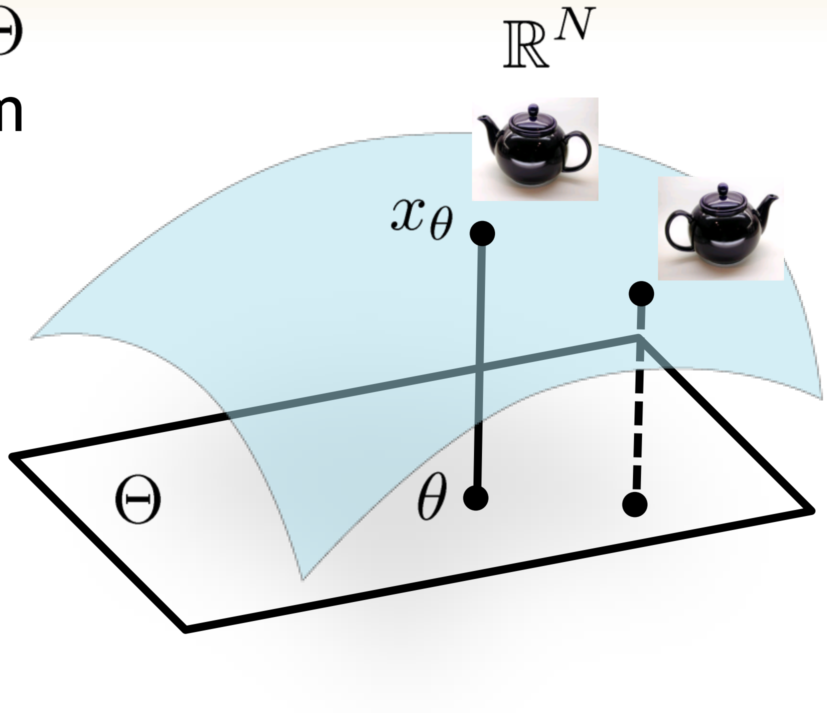
- need to repeat process for each signal model
- stability

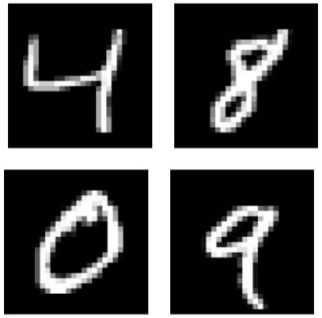
[Vetterli, Marziliano, and, Blu – 2002; Dragotti, Vetterli, and Blu - 2007]

Manifolds

- S-dimensional *parameter* $\theta \in \Theta$ captures the degrees of freedom of signal

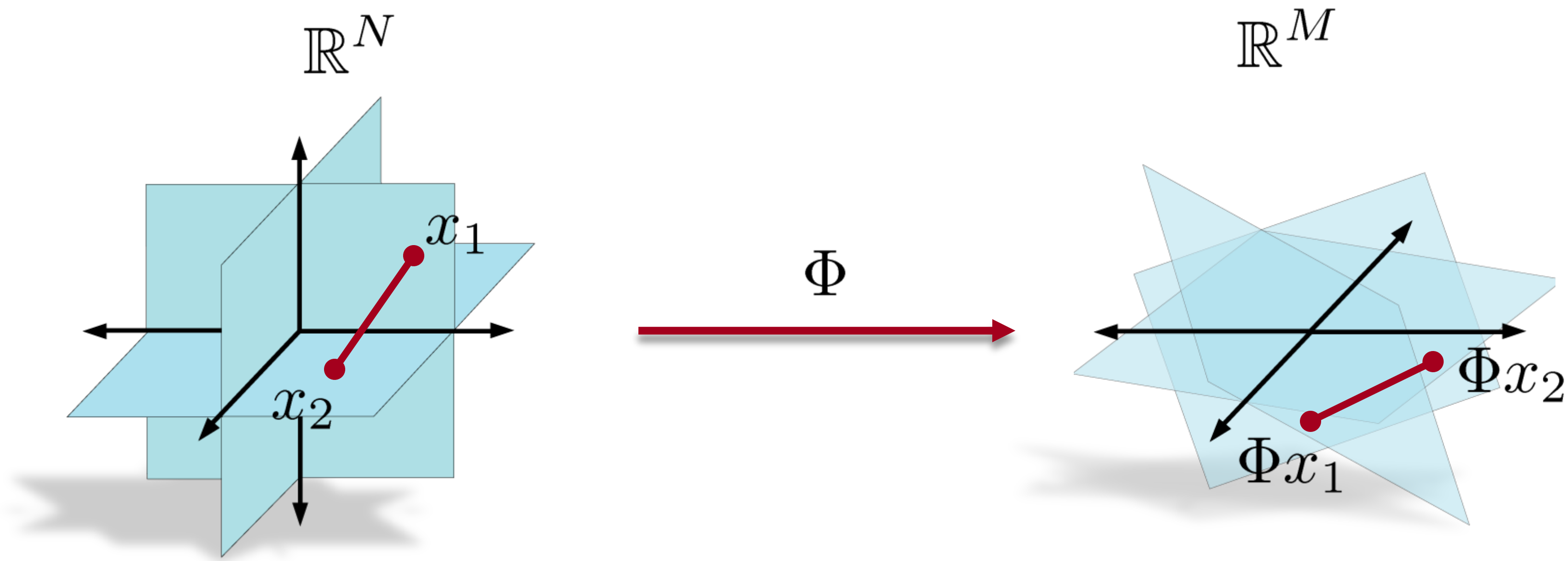
- Signal class forms an S-dimensional *manifold*
  - rotations, translations
  - robot configuration spaces
  - signal with unknown translation
  - sinusoid of unknown frequency
  - faces
  - handwritten digits
  - speech
  - …





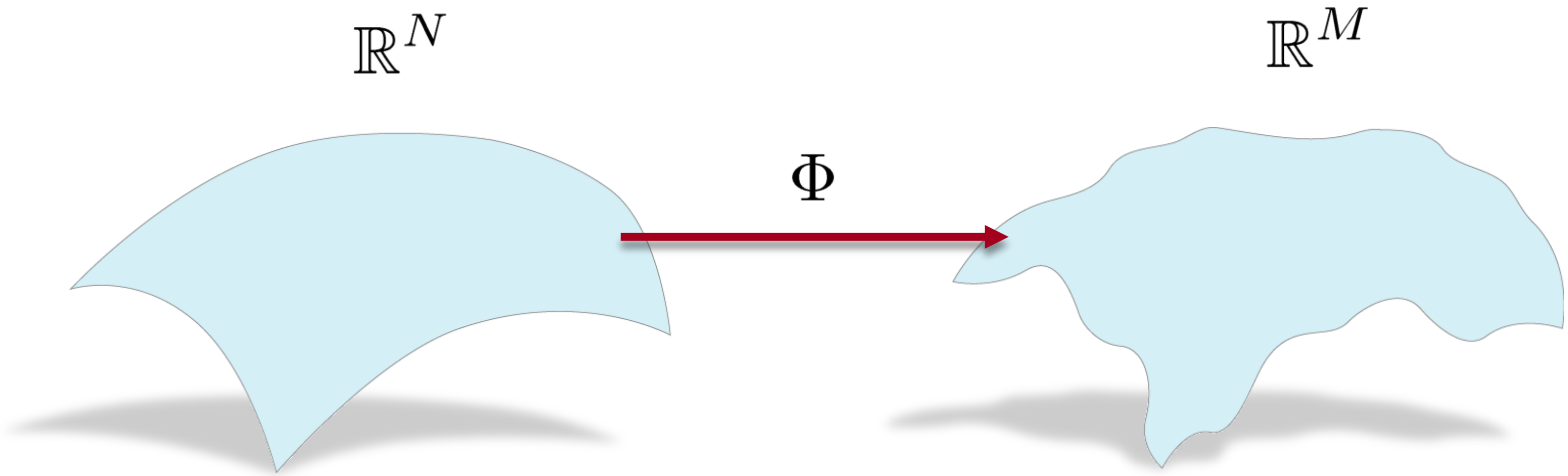
Random Projections

- For sparse signals, random projections preserve geometry



- What about manifolds?

Whitney's Embedding Theorem (1936)



S-dimensional
smooth
compact

$M > 2S$ random projections suffice to embed the manifold…

But *very unstable!*

Stable Manifold Embedding

Theorem:

Let $\mathcal{M} \subseteq \mathbb{R}^N$ be a compact S-dimensional manifold with

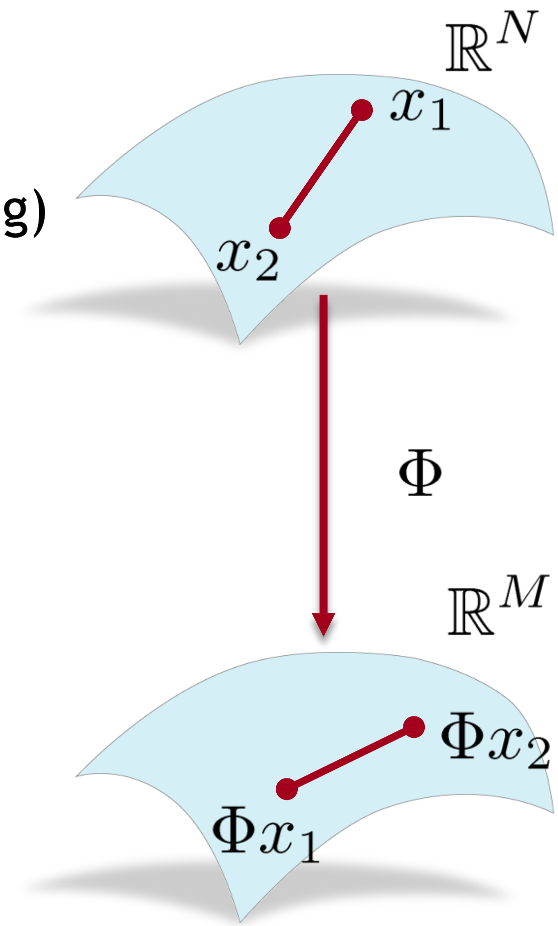
- condition number $1/\tau$ (curvature, self-avoiding)
- volume V

Let Φ be a random $M \times N$ projection with

$$M = O(S \log(NV/\tau))$$

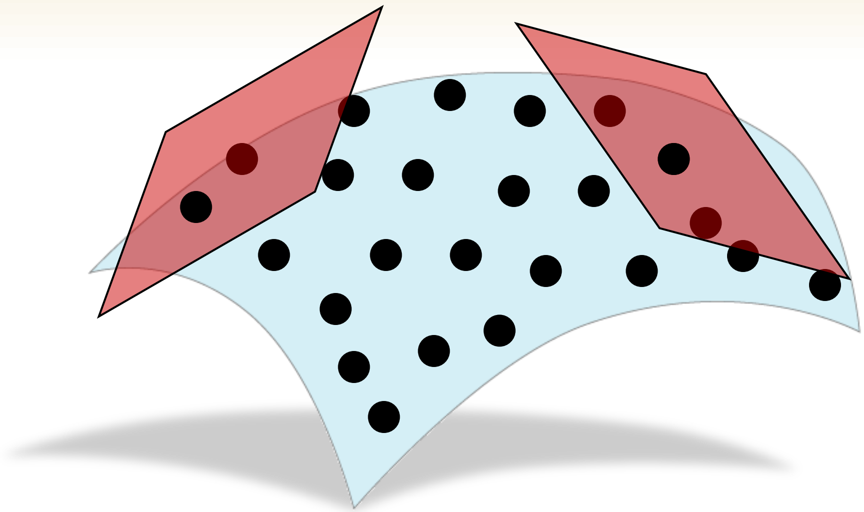
Then with high probability, and any $x_1, x_2 \in \mathcal{M}$

$$1 - \delta \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq 1 + \delta$$



[Baraniuk and Wakin – 2009]

Stable Manifold Embedding



Sketch of proof:

- construct a sampling of points
  - on manifold at fine resolution
  - from local tangent spaces
- apply JL lemma to these points

$$M = O(S \log(NV/\tau))$$

- extend to entire manifold

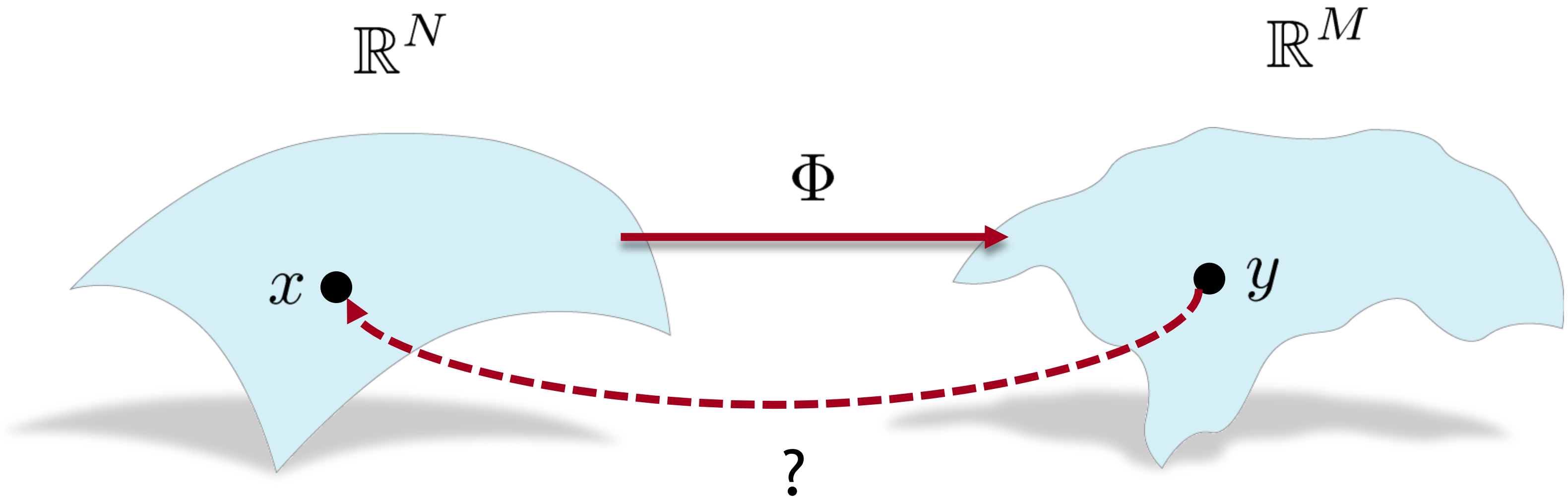
Implications:

Nonadaptive (even random) linear projections can efficiently capture & preserve structure of manifold

See also: Indyk and Naor, Agarwal et al., Dasgupta and Freund

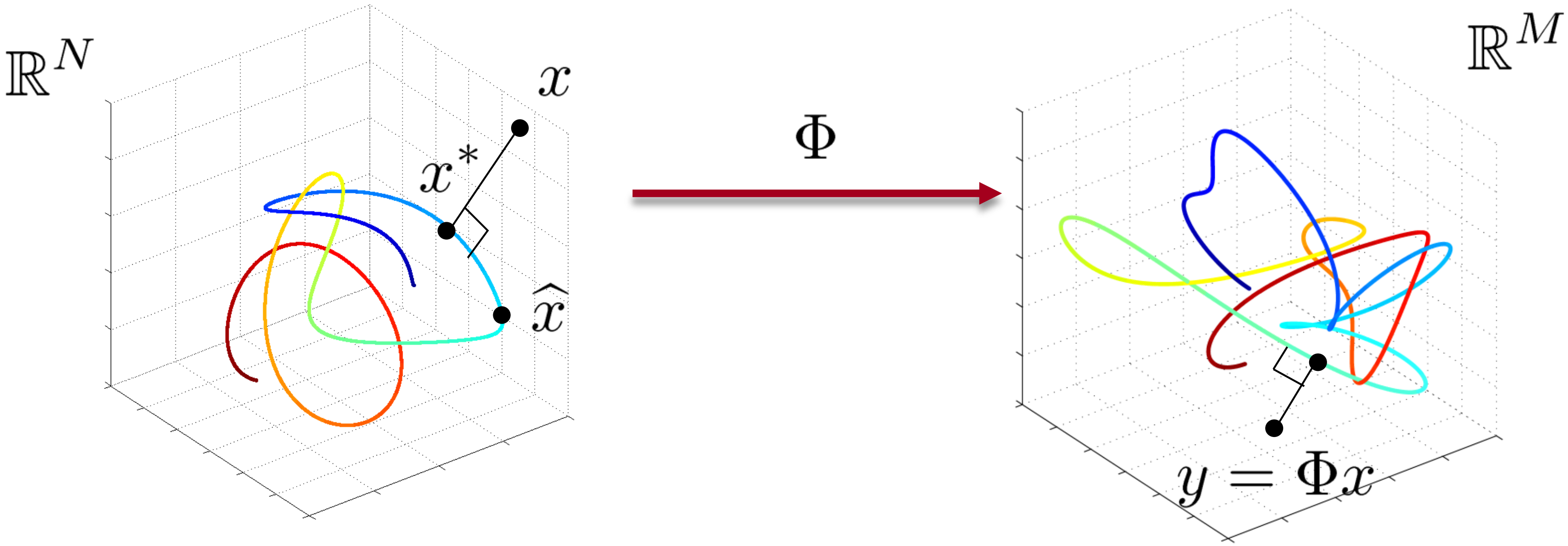
[Baraniuk and Wakin – 2009]

Compressive Sensing with Manifolds



- Same sensing protocols/devices
- Different reconstruction models
- Measurement rate depends on *manifold dimension*
- Stable embedding guarantees robust recovery

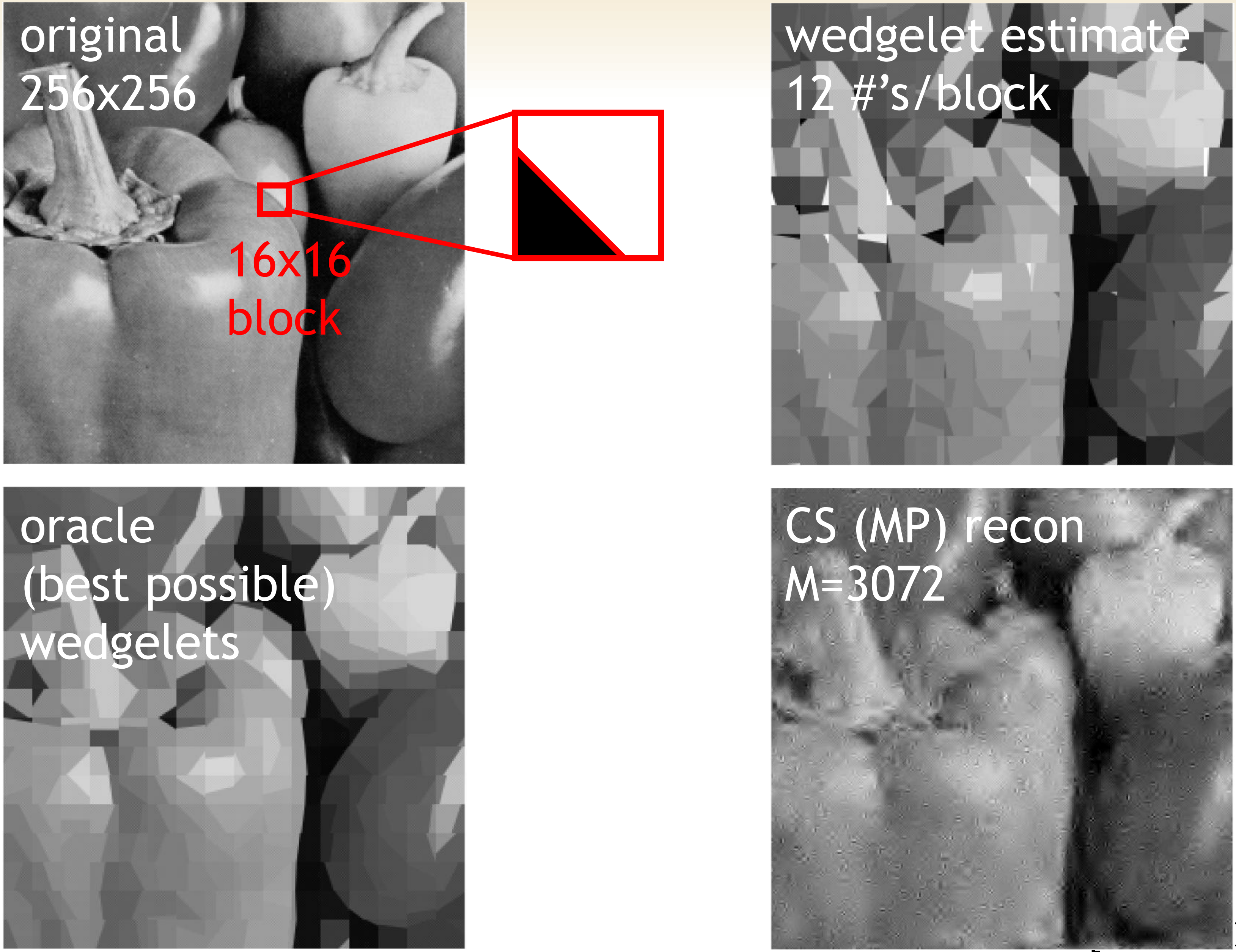
Signal Recovery



$$x^* = \arg\min_{x' \in \mathcal{M}} \|x - x'\|_2$$

$$\widehat{x} = \arg\min_{x' \in \mathcal{M}} \|y - \Phi x'\|_2$$

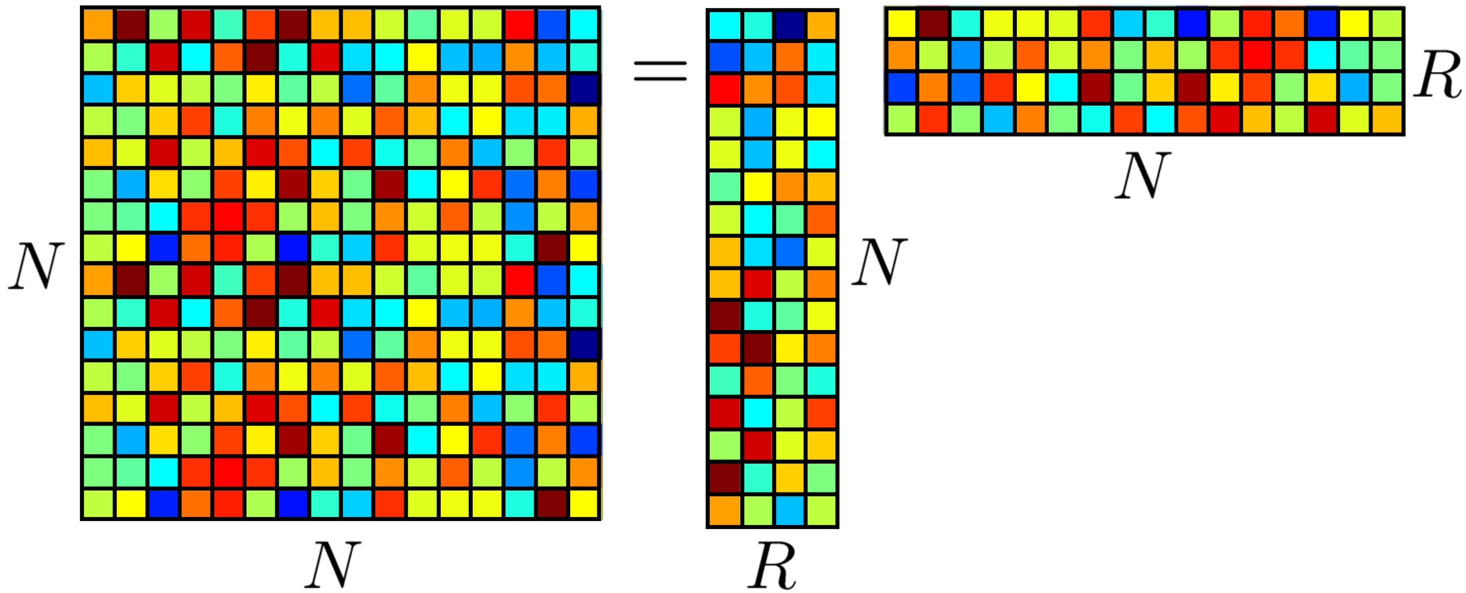
Example: Edges



2007]

Low-Rank Matrices

Low-Rank Matrices

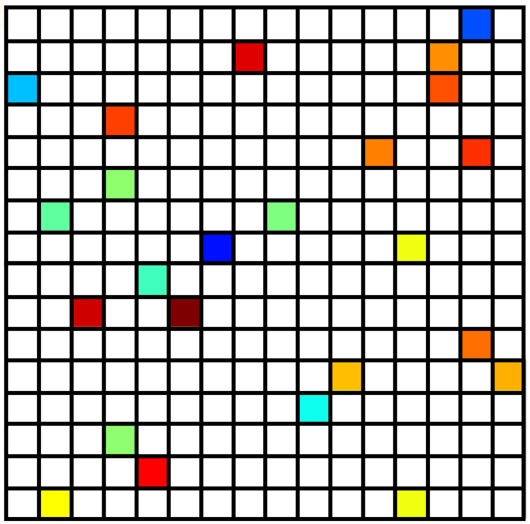


Singular value decomposition:

$$X = U\Sigma V^* \quad \Rightarrow \quad \approx NR \ll N^2$$

degrees of freedom

Matrix Completion



- Collaborative filtering (“Netflix problem”)
- How many samples will we need?

$$M \geq CNR$$

- Coupon collector problem

$$M \geq N \log N$$

Low-Rank Matrix Recovery

Given:

- an $N \times N$ matrix X of rank R
- linear measurements $y = \mathcal{A}(X)$

How can we recover X ?

$$\boxed{\widehat{X} = \underset{X : \mathcal{A}(X) = y}{\arg\inf} \ \operatorname{rank}(X)}$$

Can we replace this with something computationally feasible?

Nuclear Norm Minimization

Convex relaxation!

Replace $\mathrm{rank}(X)$ with $\|X\|_* = \sum_{j=1}^{N} |\sigma_j|$

The "nuclear norm" is just the ℓ_1-norm of the vector of singular values

$$\boxed{\widehat{X} = \underset{X : \mathcal{A}(X) = y}{\arg\inf} \|X\|_*}$$

$$M = O(NR \log N)$$

[Candès, Fazel, Keshavan, Li, Ma, Montanari, Oh, Parrilo, Plan, Recht, Tao, Wright, ...]

Conclusions

- "Conciseness" has many incarnations

- Structured sparsity
  - usually present in practice
  - often allows for *significant improvements*

- Manifolds
  - very common
  - very general

- Low-rank matrices
  - exciting community, *lots of open problems*