Compressed Sensing: A Tutorial

IEEE Statistical Signal Processing Workshop
Madison, Wisconsin
August 26, 2007

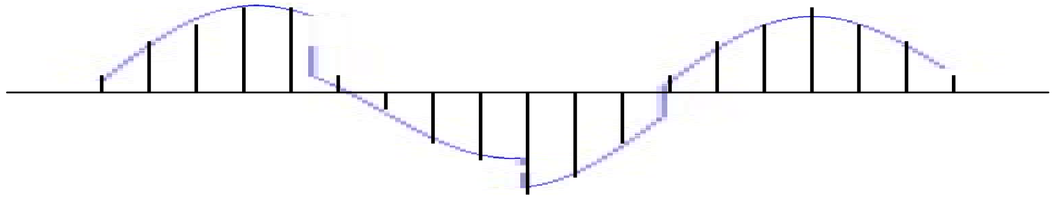
Justin Romberg
School of ECE
Georgia Tech

Michael Wakin
EECS Department
University of Michigan

Download at: `http://users.ece.gatech.edu/∼justin/ssp2007`

Data Acquisition

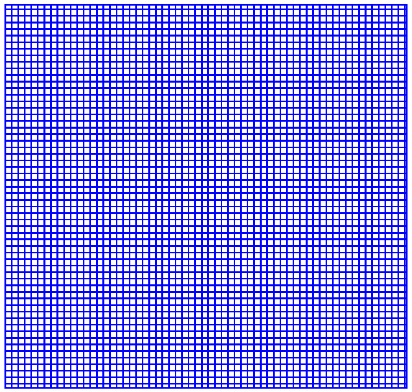
- Shannon-Nyquist sampling theorem:
  no information loss if we sample at 2x the bandwidth

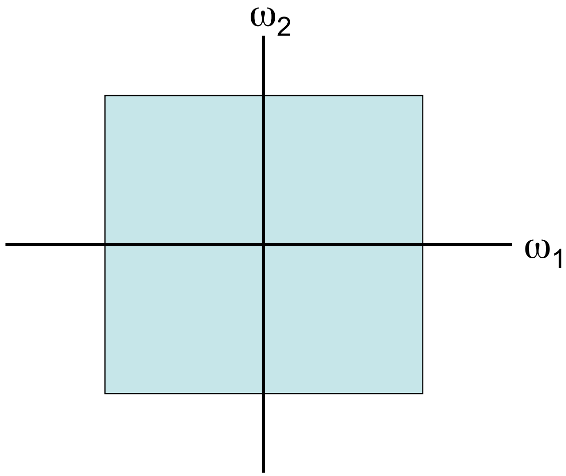


- DSP revolution: sample then process
- Trends (demands):
  - faster sampling
  - larger dynamic range
  - higher-dimensional data
  - lower energy consumption
  - new sensing modalities

Nyquist Sampling

- RF applications: to acquire an EM pulse containing frequencies at f_{max}, we need to sample at rate $\sim f_{max}$
- Pixel imaging: to get n-pixel resolution, we need n sensors
  Fourier imaging (MRI): need dense sampling out to freqs $\sim n$





- *Resolution* determines the measurement complexity
- Makes sense, but we know many times signals are much simpler . . .

Signal and Image Representations

- Fundamental concept in DSP: *Transform-domain processing*
- Decompose f as superposition of atoms (orthobasis or tight frame)

$$f(t) = \sum_i \alpha_i \psi_i(t) \quad \text{or} \quad f = \Psi\alpha$$

  e.g. sinusoids, wavelets, curvelets, Gabor functions,...

- Process the coefficient sequence α

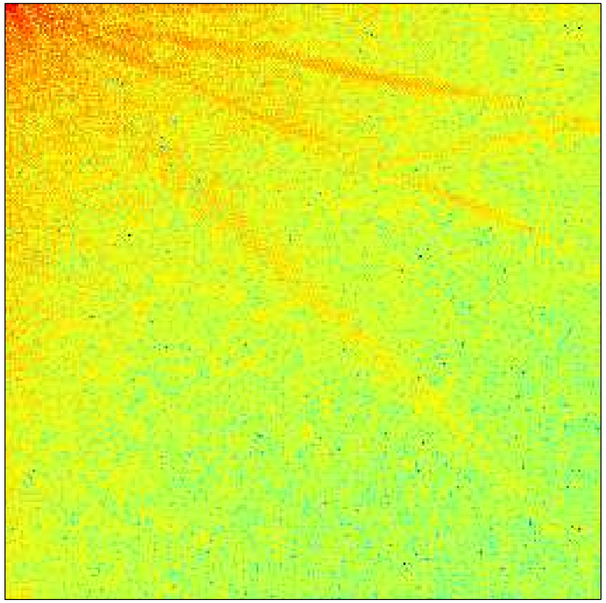
$$\alpha_i = \langle f, \psi_i \rangle, \quad \text{or} \quad \alpha = \Psi^T f$$

- Why do this?
  If we choose Ψ wisely, $\{\alpha_i\}$ will be "simpler" than $f(t)$

Classical Image Representation: DCT

- Discrete Cosine Transform (DCT)
  Basically a real-valued Fourier transform (sinusoids)

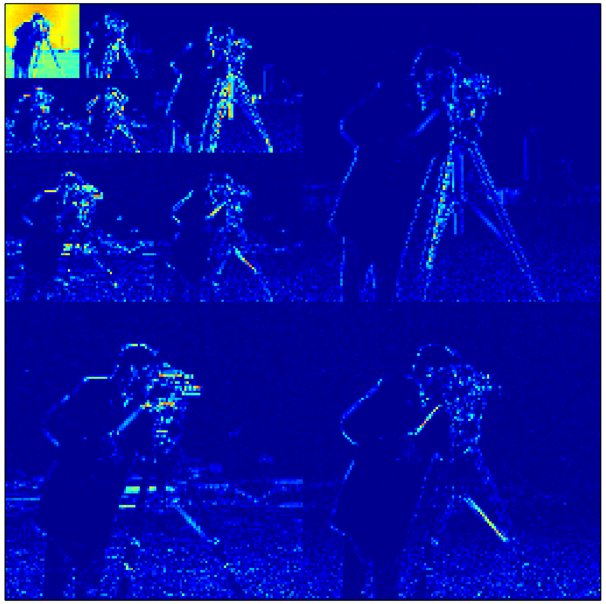
- Model: most of the energy is at low frequencies

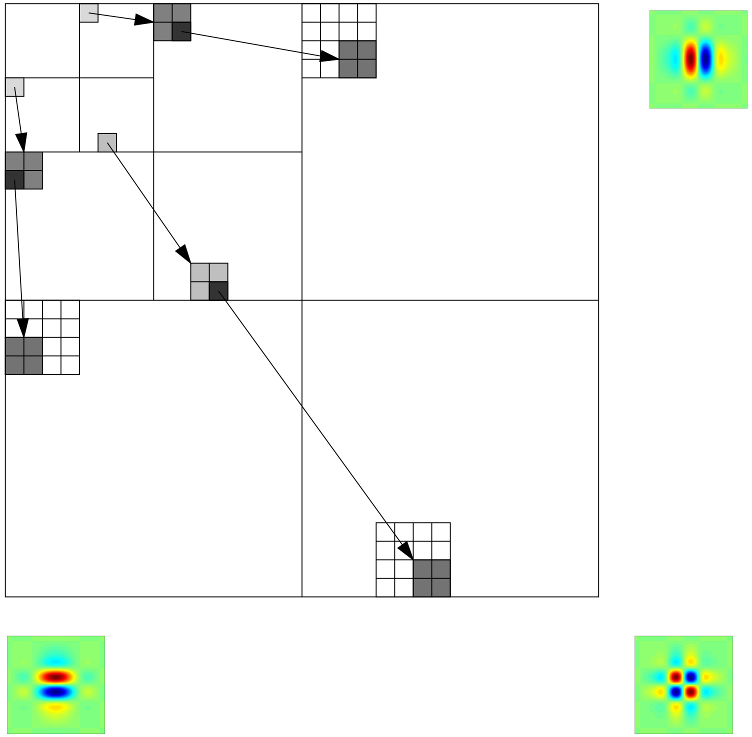


- Basis for JPEG image compression standard

- DCT approximations: smooth regions great, edges blurred/ringing

Modern Image Representation: 2D Wavelets



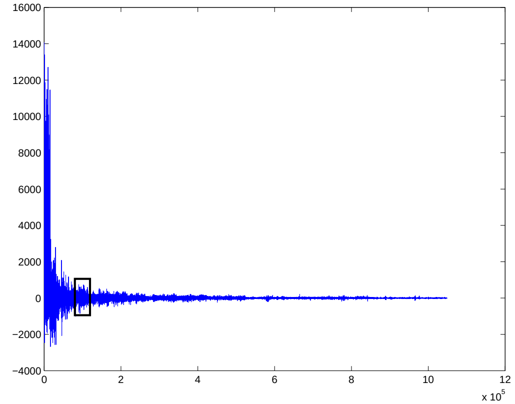


- Sparse structure: few large coeffs, many small coeffs
- Basis for JPEG2000 image compression standard
- Wavelet approximations: smooths regions great, edges much sharper
- *Fundamentally better than DCT for images with edges*

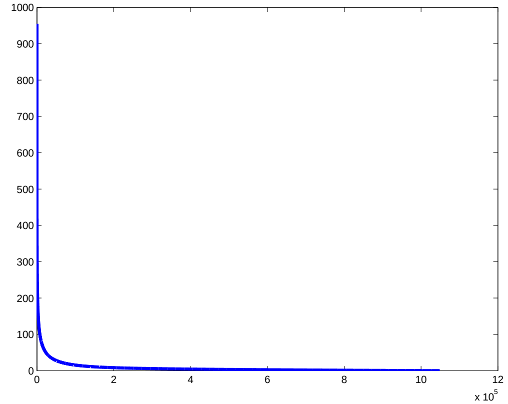
Wavelets and Images

1 megapixel image

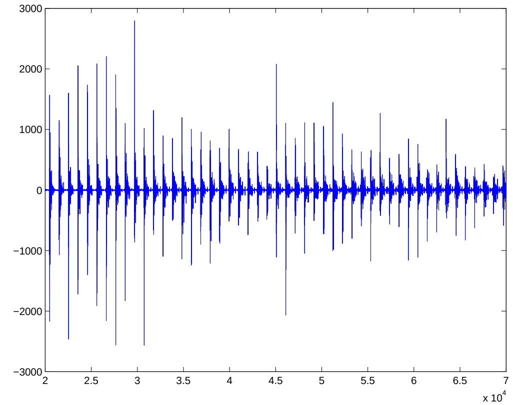
wavelet coeffs



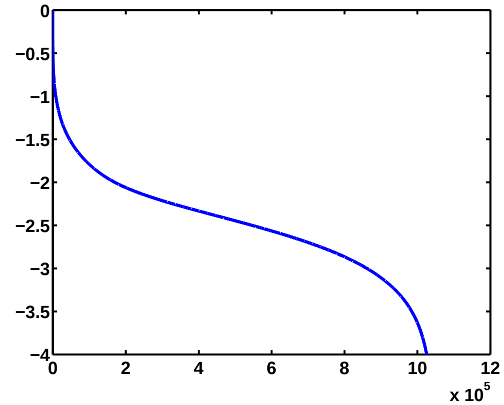
(sorted)



⇓



zoom in

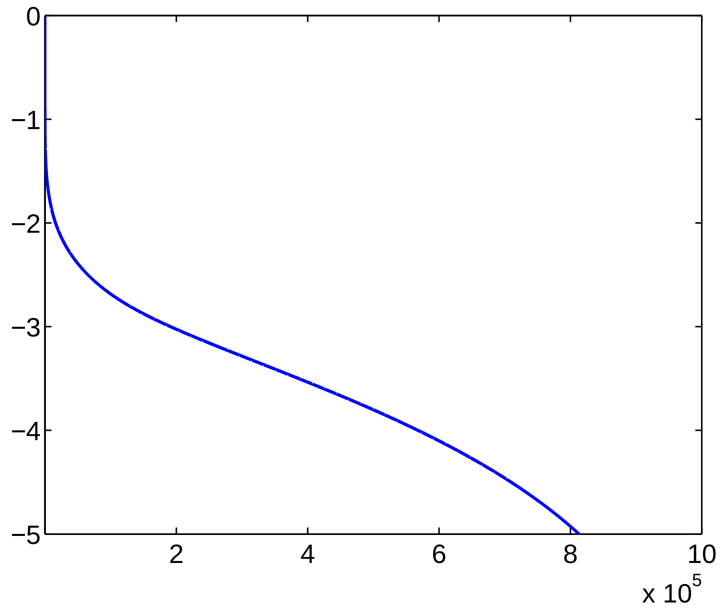


($\log_{10}$ sorted)

Wavelet Approximation

1 megapixel image

25k term approx



B-term approx error

- Within 2 digits (in MSE) with $\approx 2.5\%$ of coeffs
- Original image = f, K-term approximation = f_K

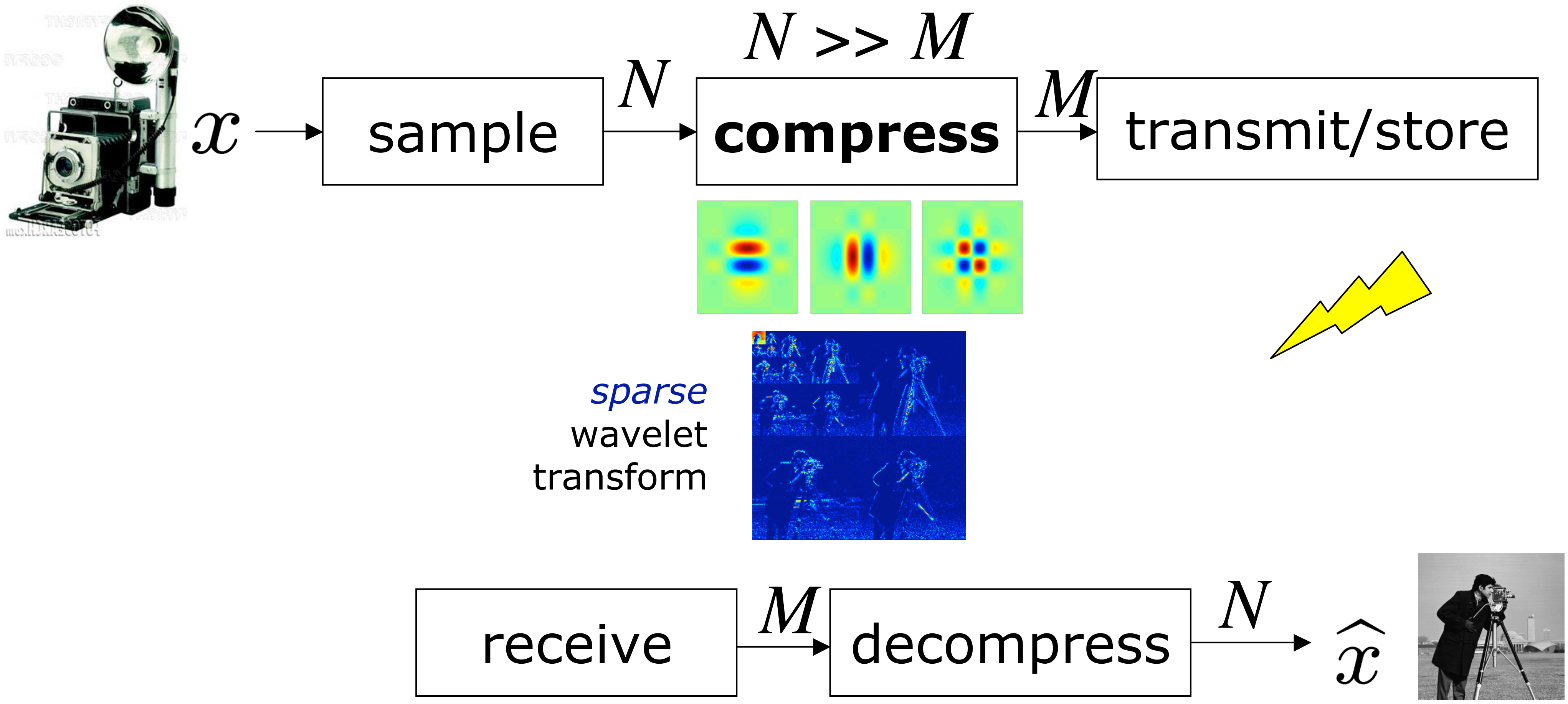
$$\|f - f_K\|_2 \approx .01 \cdot \|f\|_2$$

Computational Harmonic Analysis

- Sparsity plays a *fundamental role* in how well we can:
  - Estimate signals in the presence of noise (shrinkage, soft-thresholding)
  - Compress (transform coding)
  - Solve inverse problems (restoration and imaging)
- Dimensionality reduction facilitates modeling:
  simple models/algorithms are effective
- *This talk:*
  Sparsity also determines how quickly we can acquire signals
  non-adaptively

Sample then Compress

- Established paradigm for data acquisition:
  - *sample* data (A/D converter, photo detector,. . .)
  - *compress* data (exploit structure, nonlinear)



Coded Acquisition

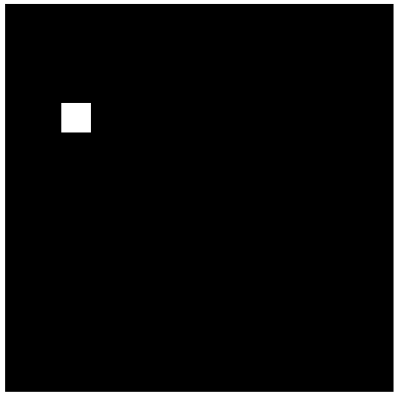
- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

$$y = \Phi f$$

- Equivalent to transform domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: big pixels

$y_m = \Big\langle$  ,  $\Big\rangle$

Coded Acquisition

- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

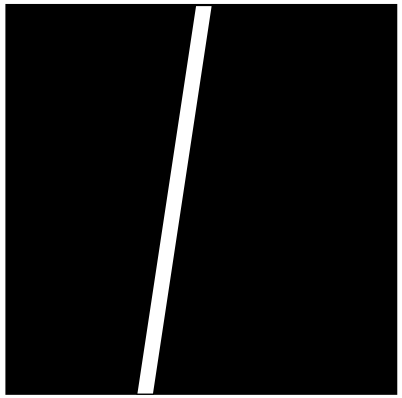
$$y = \Phi f$$

- Equivalent to transform domain sampling, $\{\phi_m\}$ = basis functions

- Example: line integrals (tomography)

$$y_m = \Bigg\langle$$

,



$$\Bigg\rangle$$

Coded Acquisition

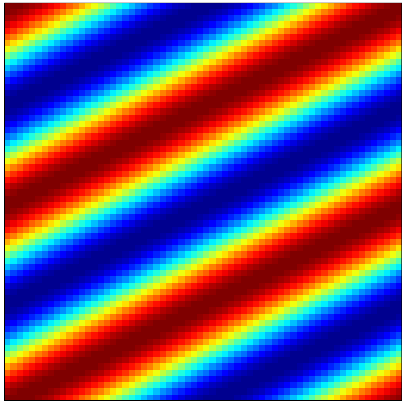
- Instead of pixels, take *linear measurements*

$$y_1 = \langle f, \phi_1 \rangle, \quad y_2 = \langle f, \phi_2 \rangle, \quad \ldots, y_M = \langle f, \phi_M \rangle$$

$$y = \Phi f$$

- Equivalent to transform domain sampling,
  $\{\phi_m\}$ = basis functions

- Example: sinusoids (MRI)

$y_m = \Big\langle$  ,  $\Big\rangle$

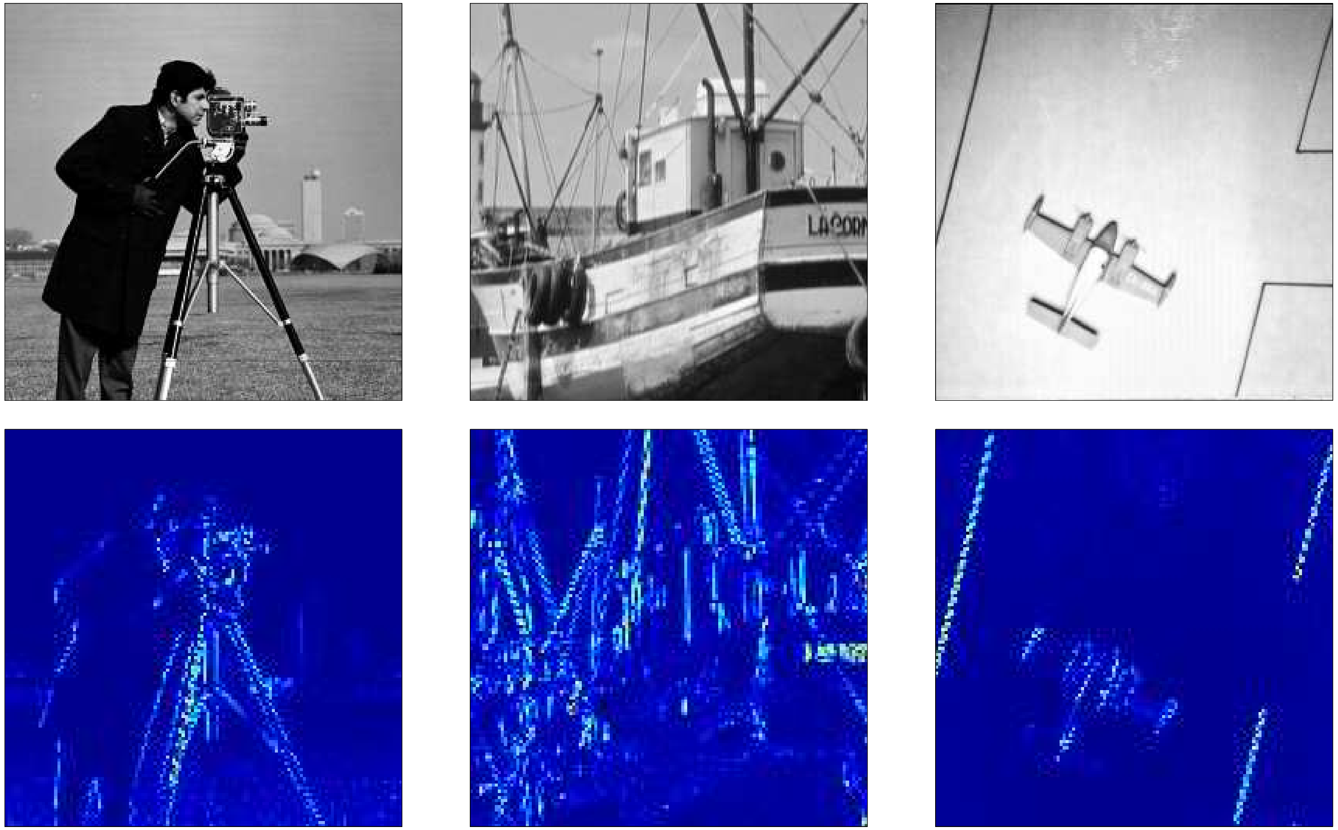
Sampling Domain

$y_k = \Big\langle$  $, \quad ? \Big\rangle$

- Which ϕ_m should we use to minimize the number of samples?
- Say we use a sparsity basis for the ϕ_m:
  M measurements = M-term approximation
- So, should we measure wavelets?

Wavelet Imaging?

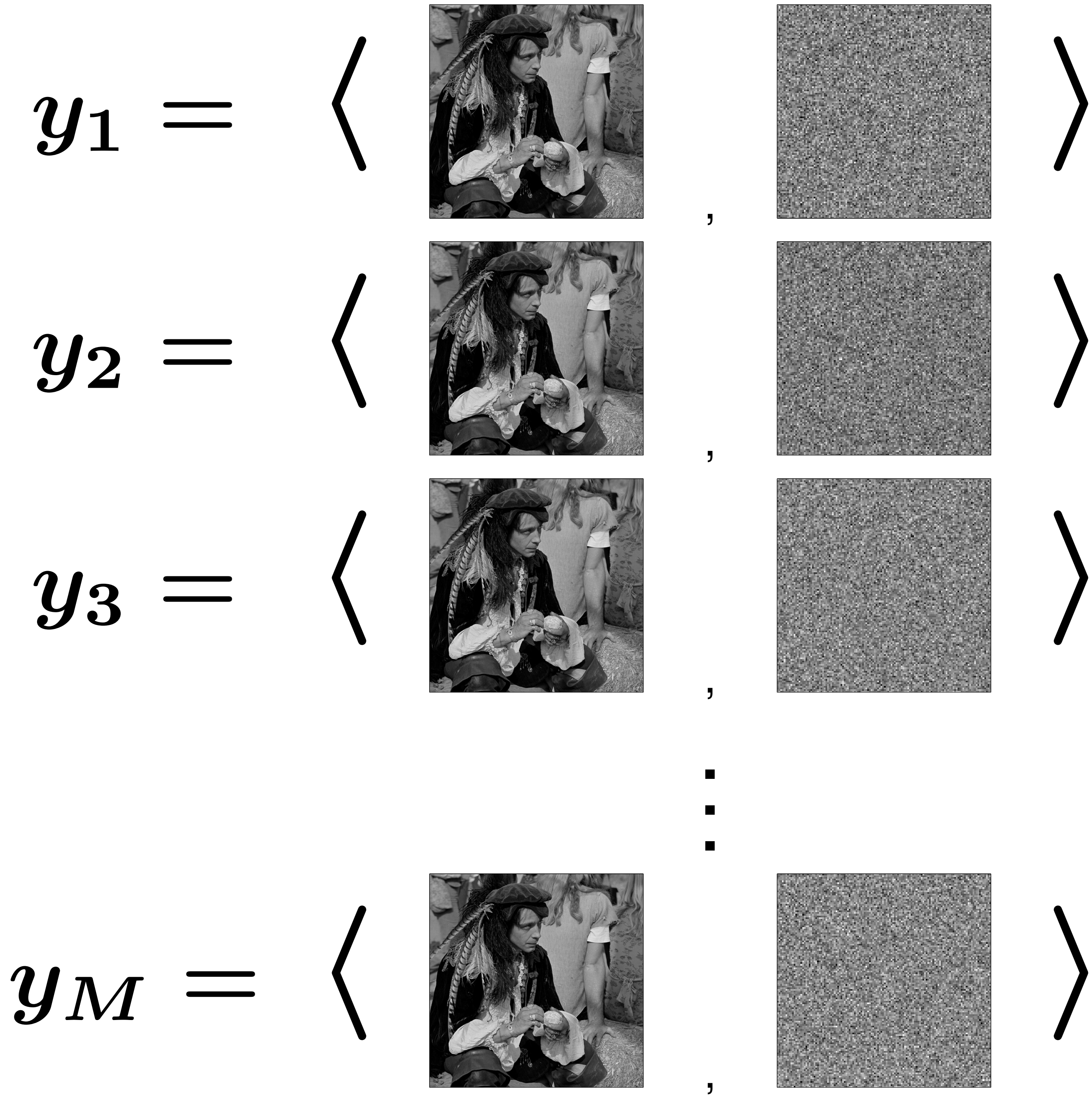
- Want to measure wavelets, but which ones?



The Big Question

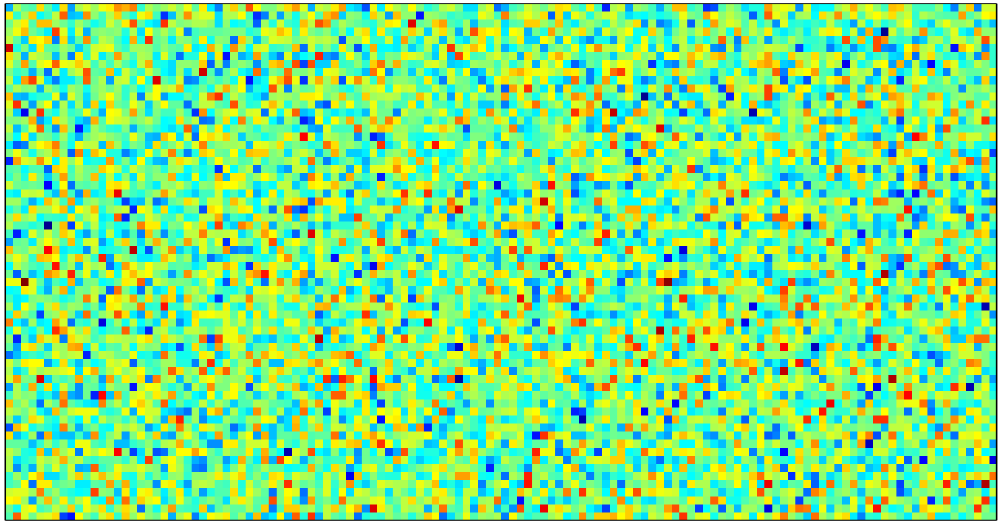
Can we get adaptive approximation performance from a fixed set of measurements?

- Surprisingly: yes.
- More surprising: measurements should *not* match image structure at all
- The measurements should look like random noise


$y_1 =$
$y_2 =$
$y_3 =$
$y_M =$

$$\begin{bmatrix} y \end{bmatrix} = \begin{bmatrix} \quad \Phi \quad \end{bmatrix} \begin{bmatrix} f \end{bmatrix}$$

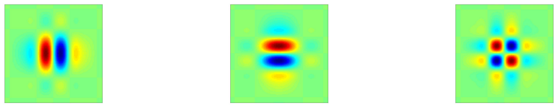
$$\begin{bmatrix} y \end{bmatrix} =$$

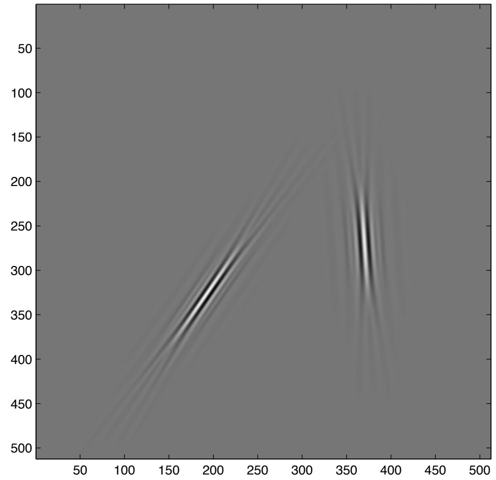


$$\begin{bmatrix} f \end{bmatrix}$$

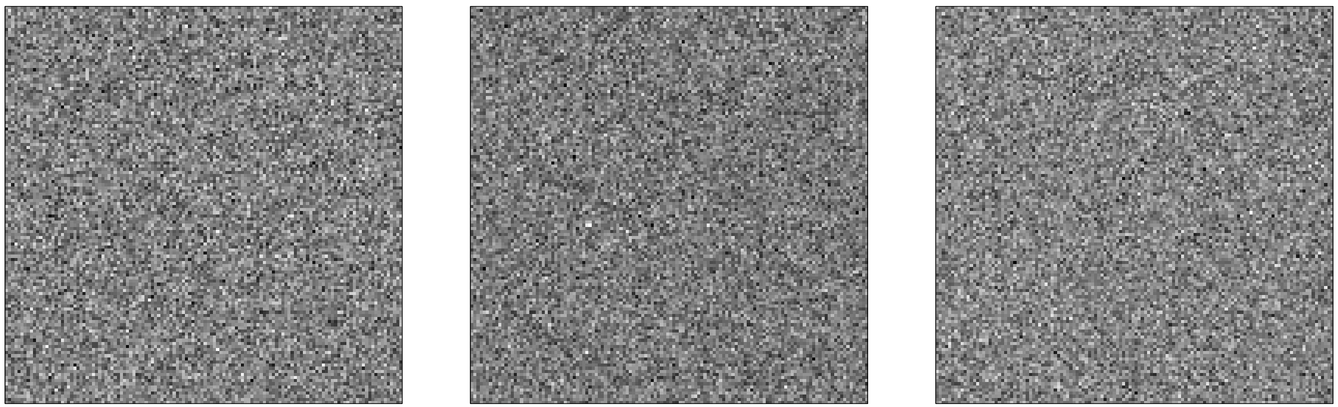
Representation vs. Measurements

- Image structure: *local, coherent*
  Good basis functions:



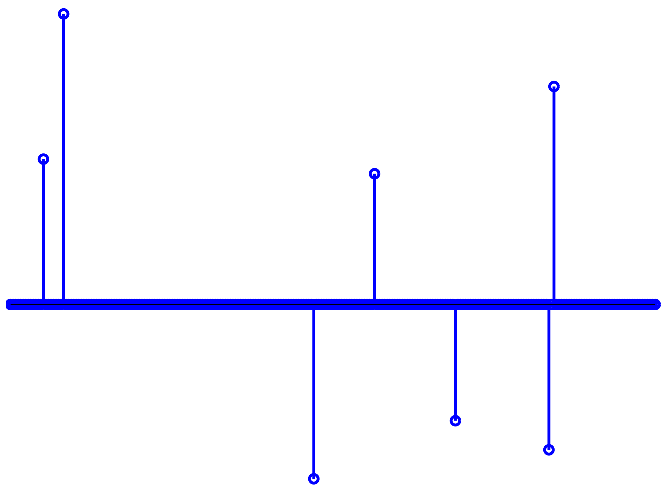


- Measurements: *global, incoherent*
  Good test functions:

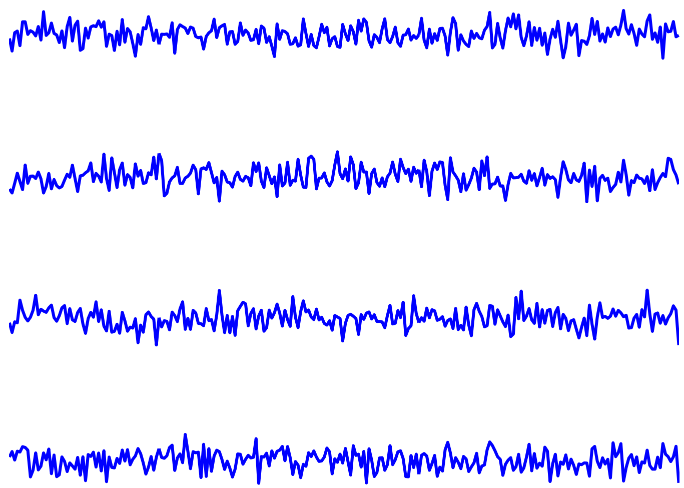


Motivation: Sampling Sparse Coefficients

concentrated vector



incoherent measurements



- Signal is local, measurements are global
- Each measurement picks up a little information about each component
- Triangulate significant components from measurements
- Formalization: Relies on uncertainty principles between sparsity basis and measurement system

Theory, Part I

The Uniform Uncertainty Principle

- Φ obeys a UUP for sets of size K if

$$0.8 \cdot \frac{M}{N} \cdot \|f\|_2^2 \leq \|\Phi f\|_2^2 \leq 1.2 \cdot \frac{M}{N} \cdot \|f\|_2^2$$

  for every K-sparse vector f

- Examples: Φ obeys UUP for $K \lesssim M/\log N$ when
  - ϕ_m = random Gaussian
  - ϕ_m = random binary
  - ϕ_m = randomly selected Fourier samples (extra $\log$ factors apply)

- We call these types of measurements *incoherent*

UUP and Sparse Recovery

- UUP for sets of size $2K \Rightarrow$ there is only one K-sparse explanation for y (almost automatic)
- Say f_0 is K-sparse, and we measure $y = \Phi f_0$
  If we search for the sparsest vector that explains y, we will find f_0:

$$\min_f \#\{t : f(t) \neq 0\} \quad \text{subject to} \quad \Phi f = y$$

- This is nice, but impossible (combinatorial)
- But, we can use the ℓ_1 norm as a *proxy* for sparsity

Sparse Recovery via ℓ_1 Minimization

- Say f_0 is K-sparse, Φ obeys UUP for sets of size $4K$
- Measure $y = \Phi f_0$
- Then solving

$$\min_f \ \|f\|_{\ell_1} \quad \text{subject to} \quad \Phi f = y$$

  will recover f_0 exactly

- We can recover f_0 from

$$M \ \gtrsim \ K \cdot \log N$$

  incoherent measurements by solving a *tractable* program

- *Number of measurements* $\approx$ *number of active components*

Example: Sampling a Superposition of Sinusoids

- Sparsity basis = Fourier domain, Sampling basis = time domain:

$$\hat{f}(\omega) = \sum_{i=1}^{K} \alpha_i \delta(\omega_i - \omega) \quad \Leftrightarrow \quad f(t) = \sum_{i=1}^{K} \alpha_i e^{i\omega_i t}$$

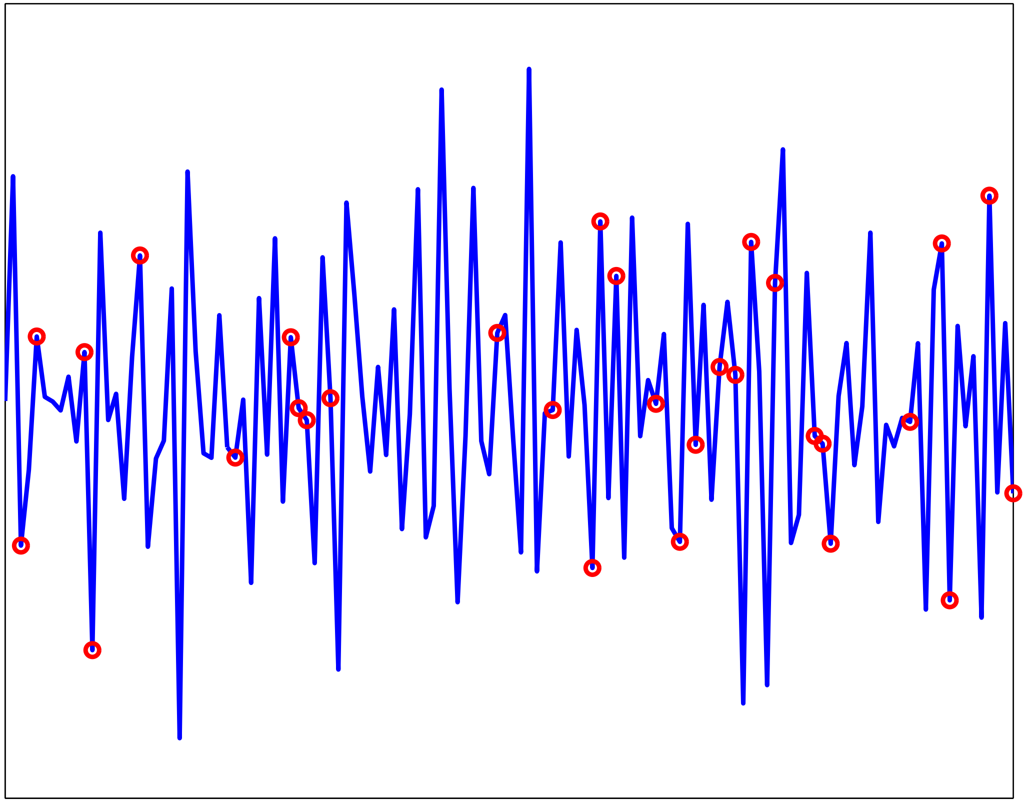
  f is a superposition of K complex sinusoids

- Recall: frequencies $\{\omega_i\}$ and amplitudes $\{\alpha_i\}$ are *unknown*.

- Take M samples of f at locations $t_1, \ldots, t_M$

Sampling Example

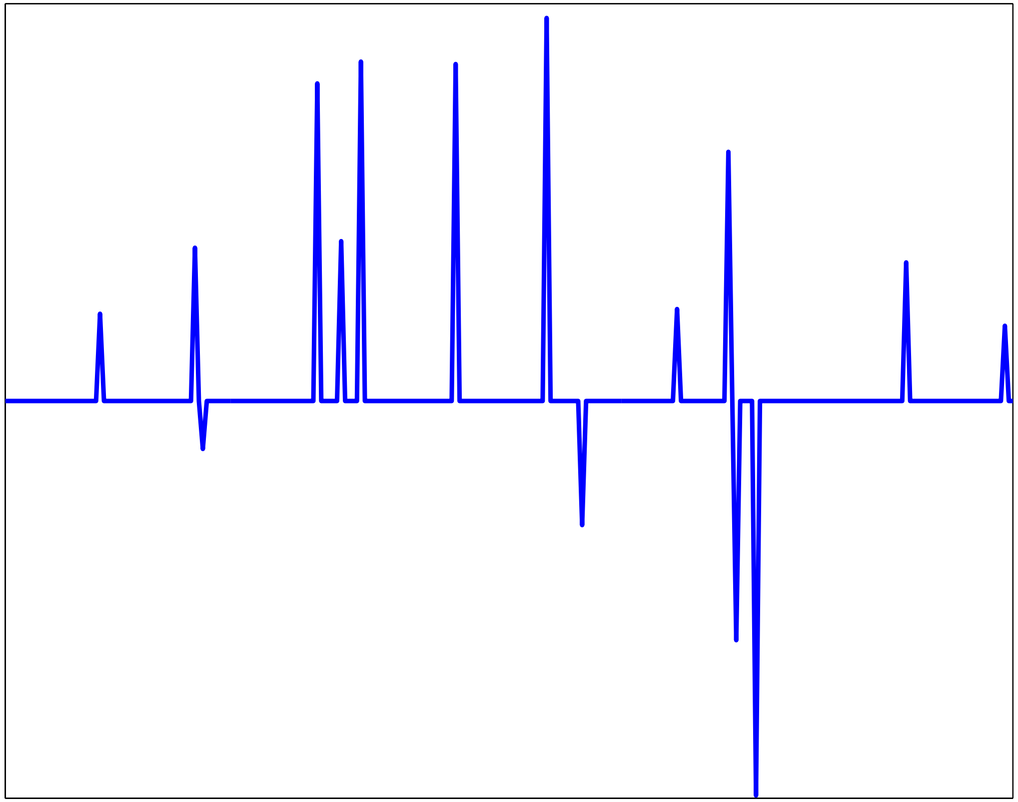
Time domain $f(t)$



Measure M samples

(red circles = samples)

Frequency domain $\hat{f}(\omega)$



K nonzero components

$\#\{\omega : \hat{f}(\omega) \neq 0\} = K$

A Nonlinear Sampling Theorem

- Suppose $\hat{f} \in \mathbb{C}^n$ is supported on set of size K
- Sample at m locations $t_1, \ldots, t_M$ in time-domain
- For the vast majority of sample sets of size

$$M \gtrsim K \cdot \log N$$

solving

$$\min_g \|\hat{g}\|_{\ell_1} \quad \text{subject to} \quad g(t_m) = y_m, \quad m = 1, \ldots, M$$

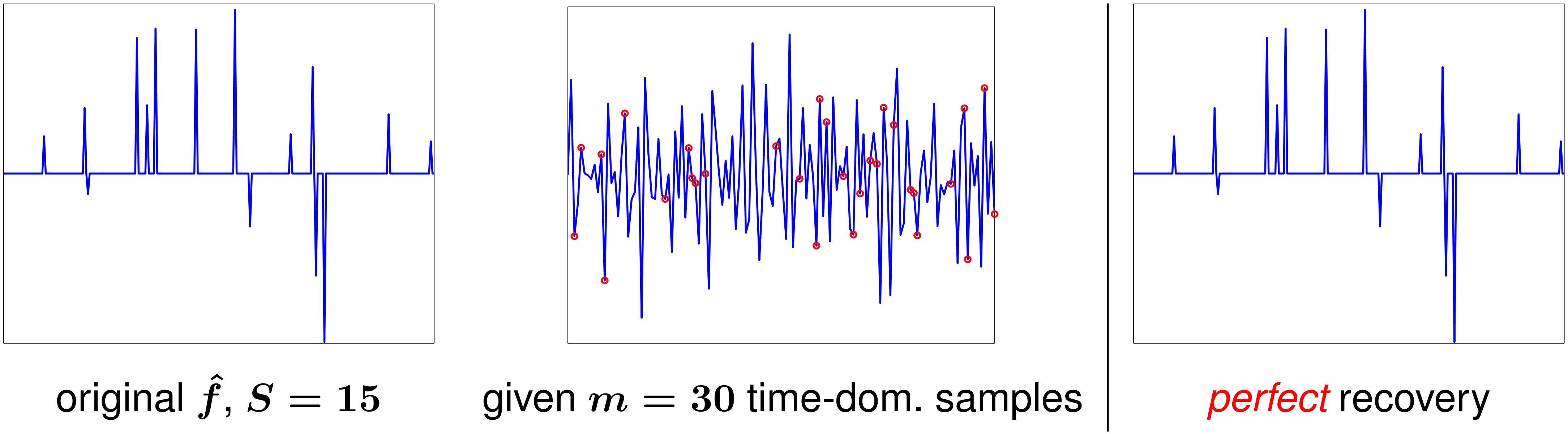
recovers f *exactly*

- In theory, $\mathrm{Const} \approx 20$
- In practice, perfect recovery occurs when $M \approx 2K$ for $N \approx 1000$.
- *# samples required* $\approx$ *# active components*
- Important frequencies are “discovered” during the recovery

ℓ_1 Reconstruction

Reconstruct by solving

$$\min_{g} \|\hat{g}\|_{\ell_1} := \min \sum_{\omega} |\hat{g}(\omega)| \quad \text{subject to} \quad g(t_m) = f(t_m), \; m = 1, \ldots, M$$



original $\hat{f}$, $S = 15$ | given $m = 30$ time-dom. samples | *perfect* recovery

Nonlinear sampling theorem

- $\hat{\boldsymbol{f}} \in \mathbb{C}^{\boldsymbol{N}}$ supported on set $\boldsymbol{\Omega}$ in Fourier domain
- Shannon sampling theorem:
  - $\boldsymbol{\Omega}$ is a known connected set of size $\boldsymbol{K}$
  - exact recovery from $\boldsymbol{K}$ equally spaced time-domain samples
  - linear reconstruction by sinc interpolation
- Nonlinear sampling theorem:
  - $\boldsymbol{\Omega}$ is an *arbitrary and unknown* set of size $\boldsymbol{K}$
  - exact recovery from $\sim \boldsymbol{K} \log \boldsymbol{N}$ (almost) arbitrarily placed samples
  - nonlinear reconstruction by convex programming

Transform Domain Recovery

- Sparsity basis $\boldsymbol{\Psi}$ (e.g. wavelets)
- Reconstruct by solving

$$\min_{\boldsymbol{\alpha}} \ \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{y}$$

- Need measurement to be incoherent in the $\boldsymbol{\Psi}$ domain
  - Random Gaussian: still incoherent (exactly the same)
  - Random binary: still incoherent
  - General rule: just make $\boldsymbol{\Phi}$ unstructured wrt $\boldsymbol{\Psi}$

Random Sensing “Acquisition Theorem”

- Signal/image $\boldsymbol{f} \in \mathbb{C}^{\boldsymbol{N}}$ is $\boldsymbol{S}$-sparse in $\boldsymbol{\Psi}$ domain
- Take

$$\boldsymbol{M} \;\gtrsim\; \boldsymbol{K} \cdot \log \boldsymbol{N}$$

measurements

$$\boldsymbol{y_1} = \langle \boldsymbol{f}, \boldsymbol{\phi_1} \rangle, \; \ldots, \; \boldsymbol{y_M} = \langle \boldsymbol{f}, \boldsymbol{\phi_M} \rangle$$

$\boldsymbol{\phi_m}$ = random waveform

- Then solving

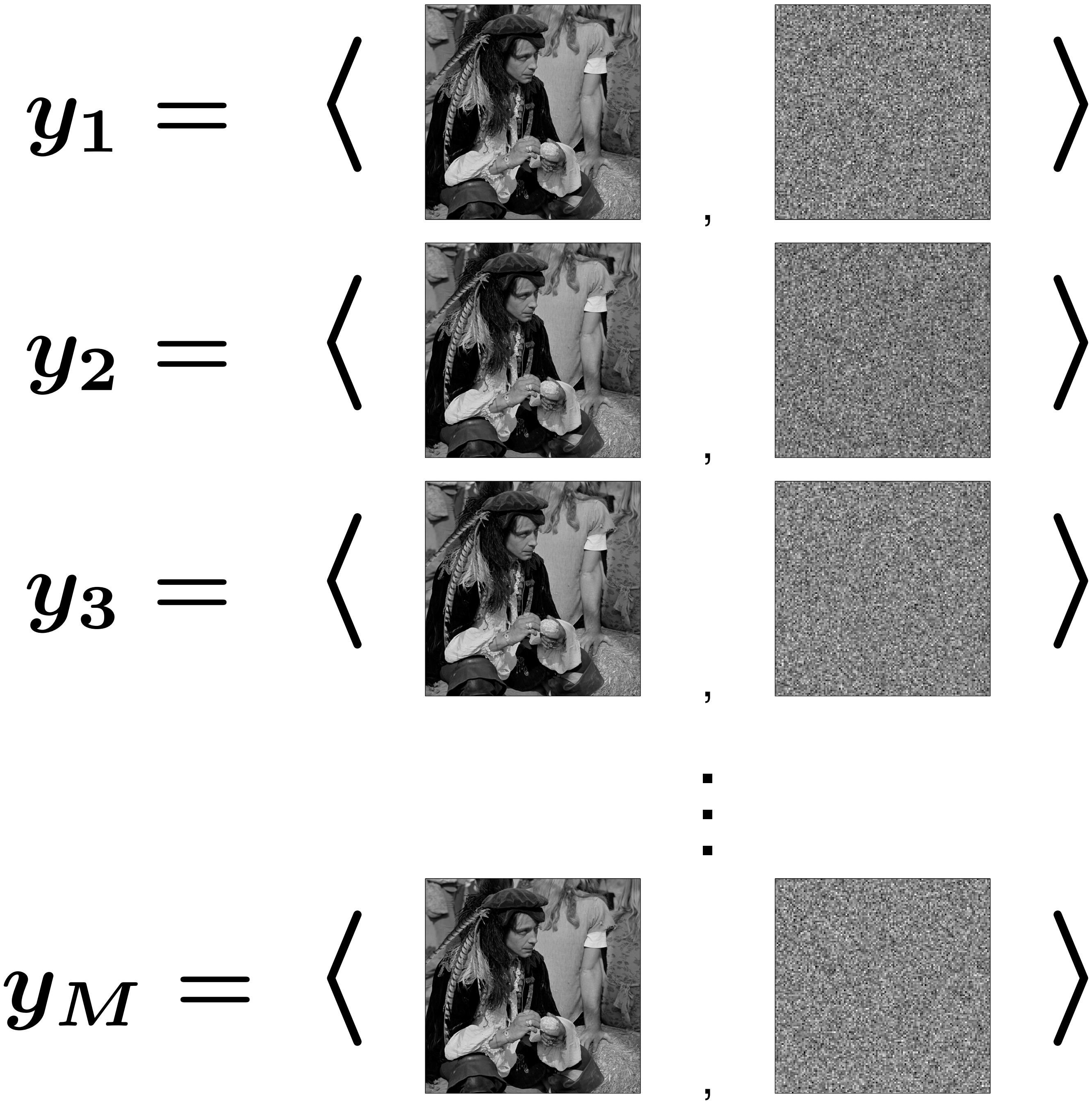
$$\min_{\boldsymbol{\alpha}} \; \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi\Psi\alpha} = \boldsymbol{y}$$

will recover (the transform coefficients) of $\boldsymbol{f}$ exactly

- In practice, it seems that

$$\boldsymbol{M} \;\approx\; \boldsymbol{5K}$$

measurements are sufficient


$y_1 =$
$y_2 =$
$y_3 =$
$y_M =$

Example: Sparse Image

- Take $M = 100,000$ incoherent measurements $y = \Phi f_a$
- f_a = wavelet approximation (perfectly sparse)
- Solve

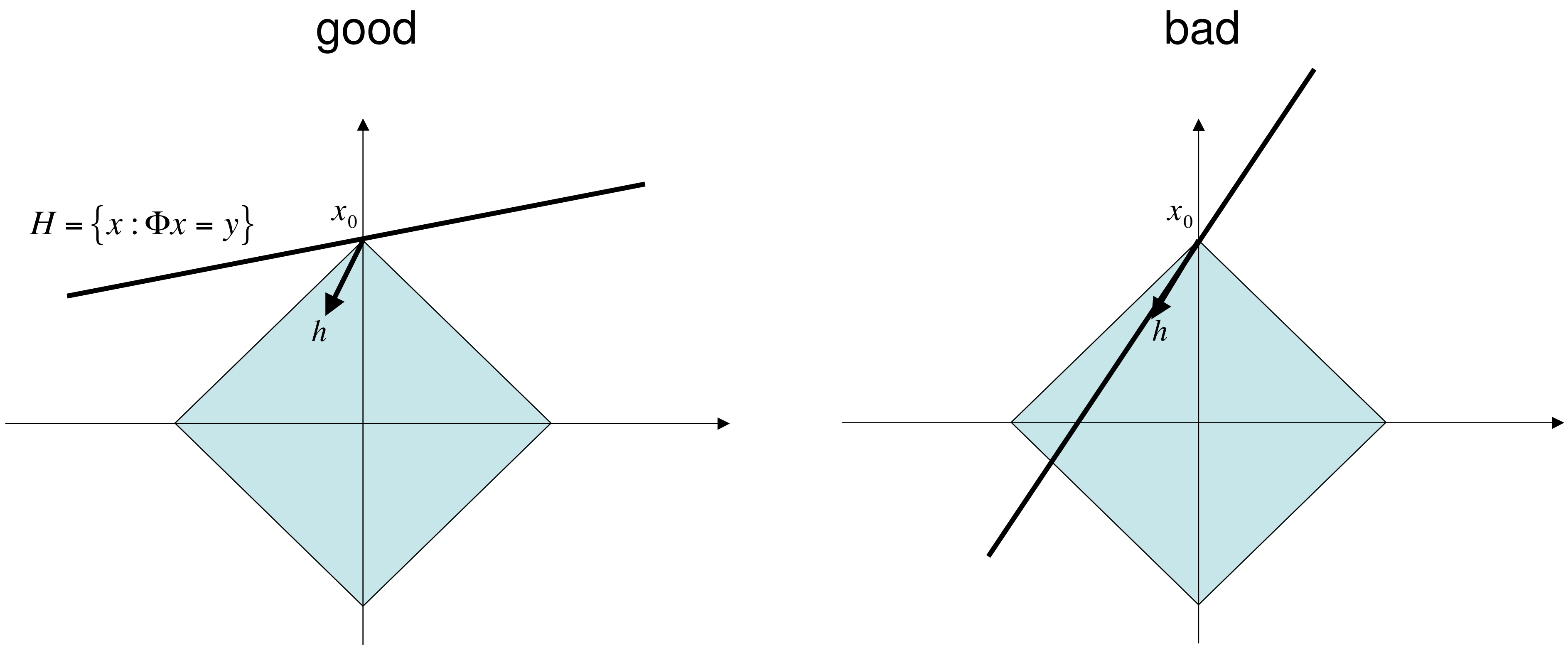
$$\min \ \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \Phi\Psi\alpha = y$$

Ψ = wavelet transform

original (25k wavelets)

perfect recovery

Geometrical Viewpoint



- Consider and "ℓ_1-descent vectors" h for feasible f_0:

$$\|f_0 + h\|_{\ell_1} < \|f_0\|_{\ell_1}$$

- f_0 is the solution if

$$\Phi h \neq 0$$

  for all such descent vectors

Stability

- Real images are not exactly sparse
- For Φ' obeying UUP for sets of size $4K$, and *general* α, recovery obeys

$$\|\alpha_0 - \alpha^*\|_2 \lesssim \frac{\|\alpha_0 - \alpha_{0,K}\|_{\ell_1}}{\sqrt{K}}$$

  $\alpha_{0,S} =$ best S-term approximation

- Compressible: if transform coefficients decay

$$|\alpha_0|_{(m)} \lesssim m^{-r}, \quad r > 1$$

  $|\alpha_0|_{(m)} = m$th largest coefficient, then

$$\|\alpha_0 - \alpha_{0,K}\|_2 \lesssim K^{-r+1/2}$$
$$\|\alpha_0 - \alpha^*\|_2 \lesssim K^{-r+1/2}$$

- *Recovery error* $\sim$ *adaptive approximation error*

Stability

- What if the measurements are noisy?

$$y = \Phi' \alpha_0 + e, \qquad \|e\|_2 \le \epsilon$$

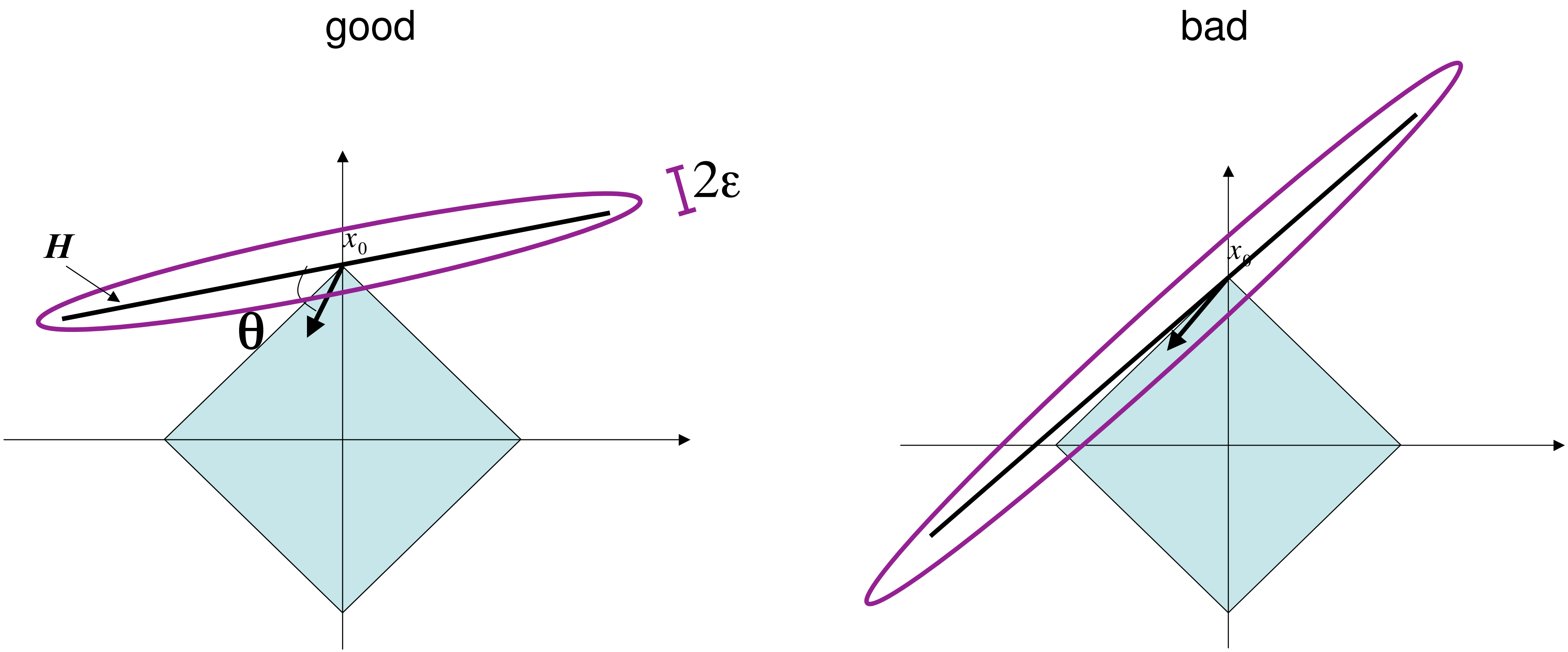
- *Relax* the recovery program; solve

$$\min_{\alpha} \ \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \|\Phi'\alpha - y\|_2 \le \epsilon$$

- The recovery error obeys

$$\|\alpha_0 - \alpha^*\|_2 \ \lesssim \ \sqrt{\frac{N}{M}} \cdot \epsilon \ + \ \frac{\|\alpha_0 - \alpha_{0,K}\|_{\ell_1}}{\sqrt{K}}$$

measurement error $+$ *approximation error*



- Solution will be within ϵ of $\boldsymbol{H}$
- Need that not too much of the ℓ_1 ball near f_0 is feasible

Compressed Sensing

- As # measurements increases, error decreases at near-optimal rate

$$\text{best } M\text{-term approximation}: \quad \|\alpha_0 - \alpha_{0,M}\|_2 \lesssim M^{-r}$$

$$\Rightarrow \quad \text{CS recovery}: \quad \|\alpha_0 - \alpha^*_M\|_2 \lesssim (M/\log N)^{-r}$$

- The sensing is *not adaptive*, and is simple
- Compression "built in" to the measurements
- Taking random measurements $=$ universal, analog coding scheme for sparse signals

Compressed Sensing

- As # measurements increases, error decreases at near-optimal rate
- Democratic and robust:
  - all measurement are equally (un)important
  - losing a few does not hurt
- The recovery is flexible, and independent of acquisition

$$\min \ \|\boldsymbol{\alpha}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} = \boldsymbol{y}$$

  Different $\boldsymbol{\Psi}$ yield different recoveries from same measurements

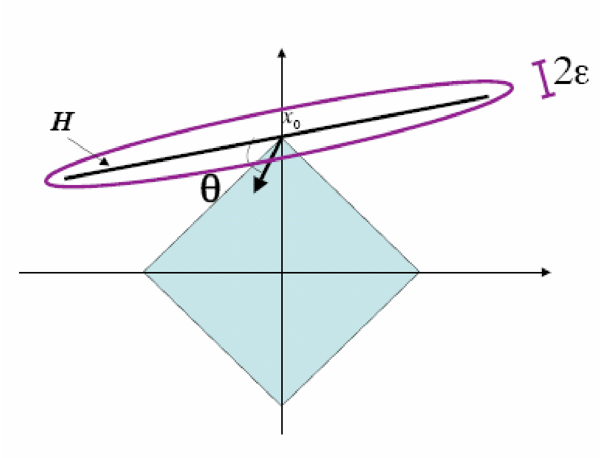
- Use *a posteriori* computing power to reduce *a priori* sampling complexity

Theory, Part II

The Geometry of CS

Geometry in CS

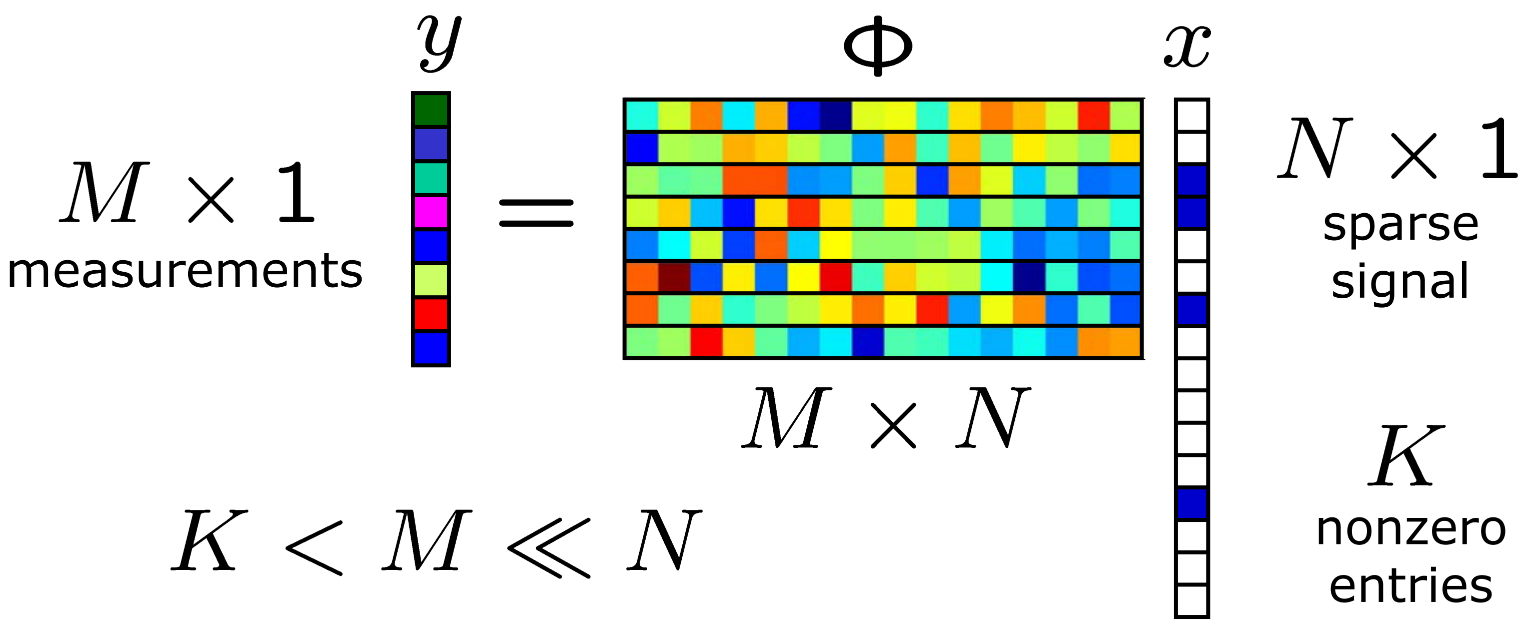
- Major *geometric themes*:
  - where signals live in ambient space
    - before and after projection
  - implications of sparse models
  - mechanics of l_1 recovery



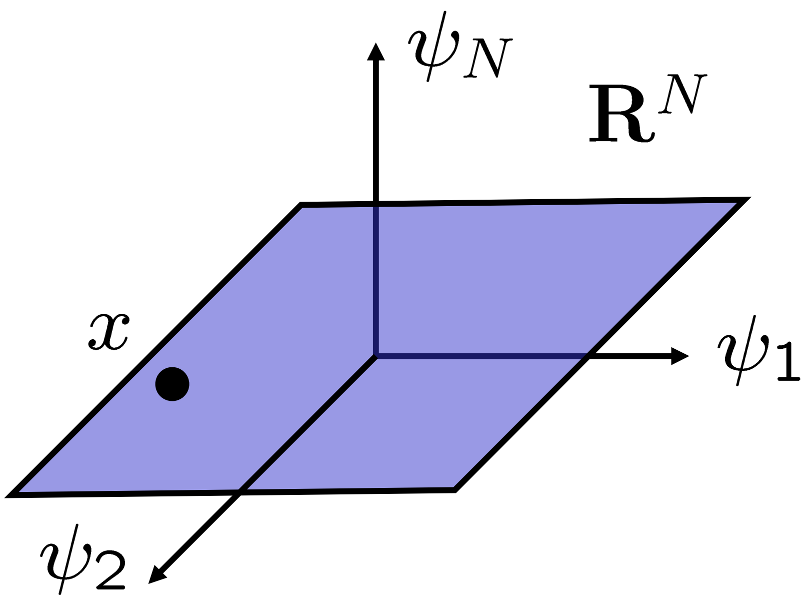
- Important questions:
  - how and why can signals be recovered?
  - how many measurements are really needed?
  - how can all this be extended to other signal models?

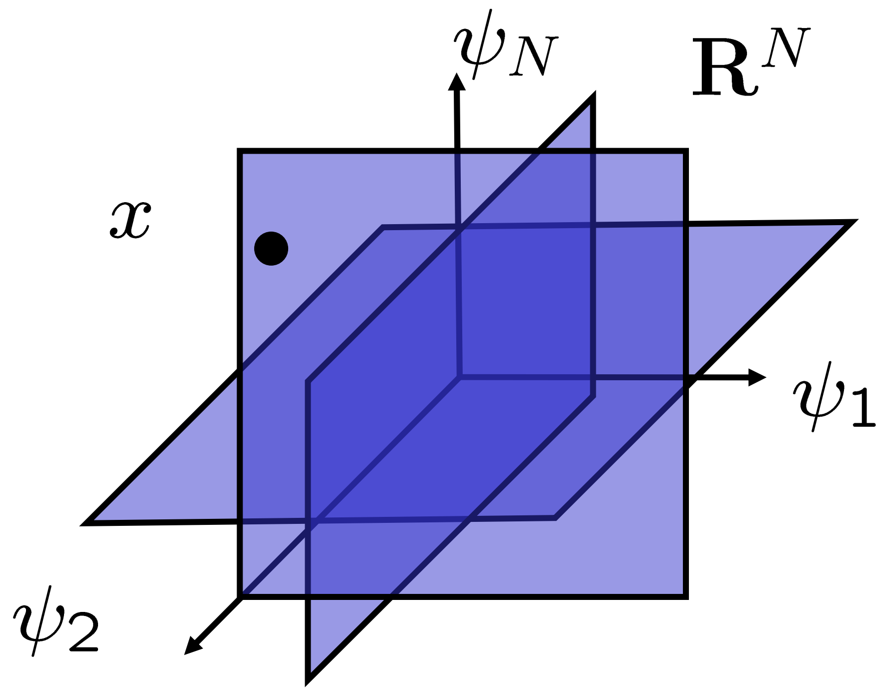
One Simple Question

- When is it possible to recover K-sparse signals?
  - require $\Phi x_1 \neq \Phi x_2$ for all K-sparse $x_1 \neq x_2$
- Necessary: Φ must have at least 2K rows
  - otherwise there exist K-sparse x_1, x_2 s.t. $\Phi(x_1 - x_2) = 0$
- Sufficient: Gaussian Φ with 2K rows (w.p. 1)
  - moreover, L_0 minimization will work for recovery



Geometry of Sparse Signal Sets





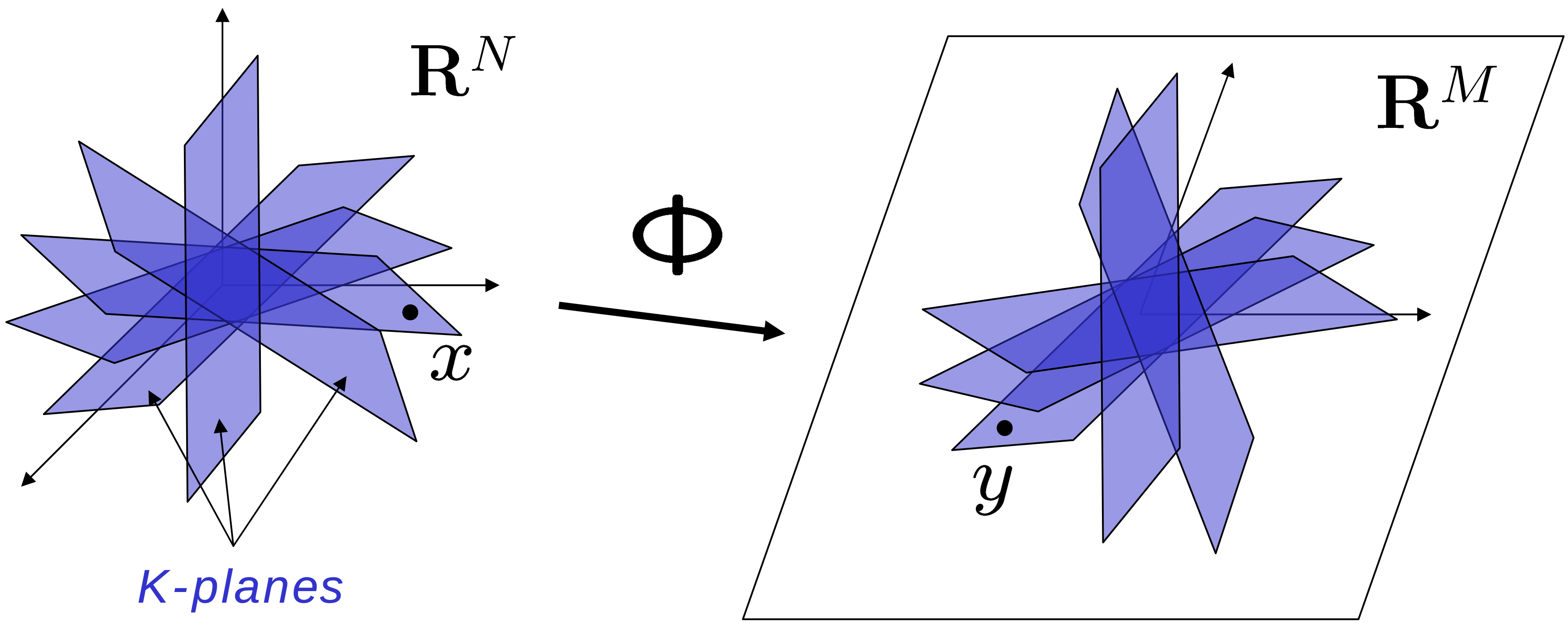
Linear

K-plane

Sparse, Nonlinear

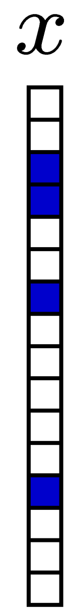
Union of K-planes

Geometry: Embedding in R^M



- Φ(K-plane) = K-plane in general
- M ≥ 2K measurements
  - necessary for injectivity
  - sufficient for injectivity when Φ Gaussian
  - but not enough for efficient, robust recovery
- (PS - can distinguish *most* K-sparse x with as few as M=K+1)

The Geometry of L_1 Recovery

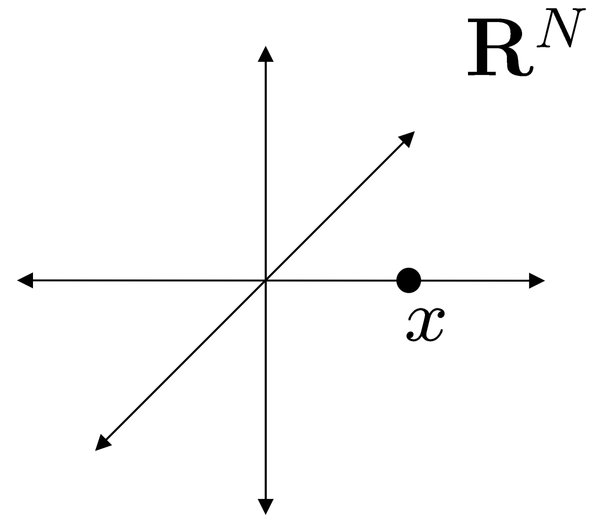


$N \times 1$

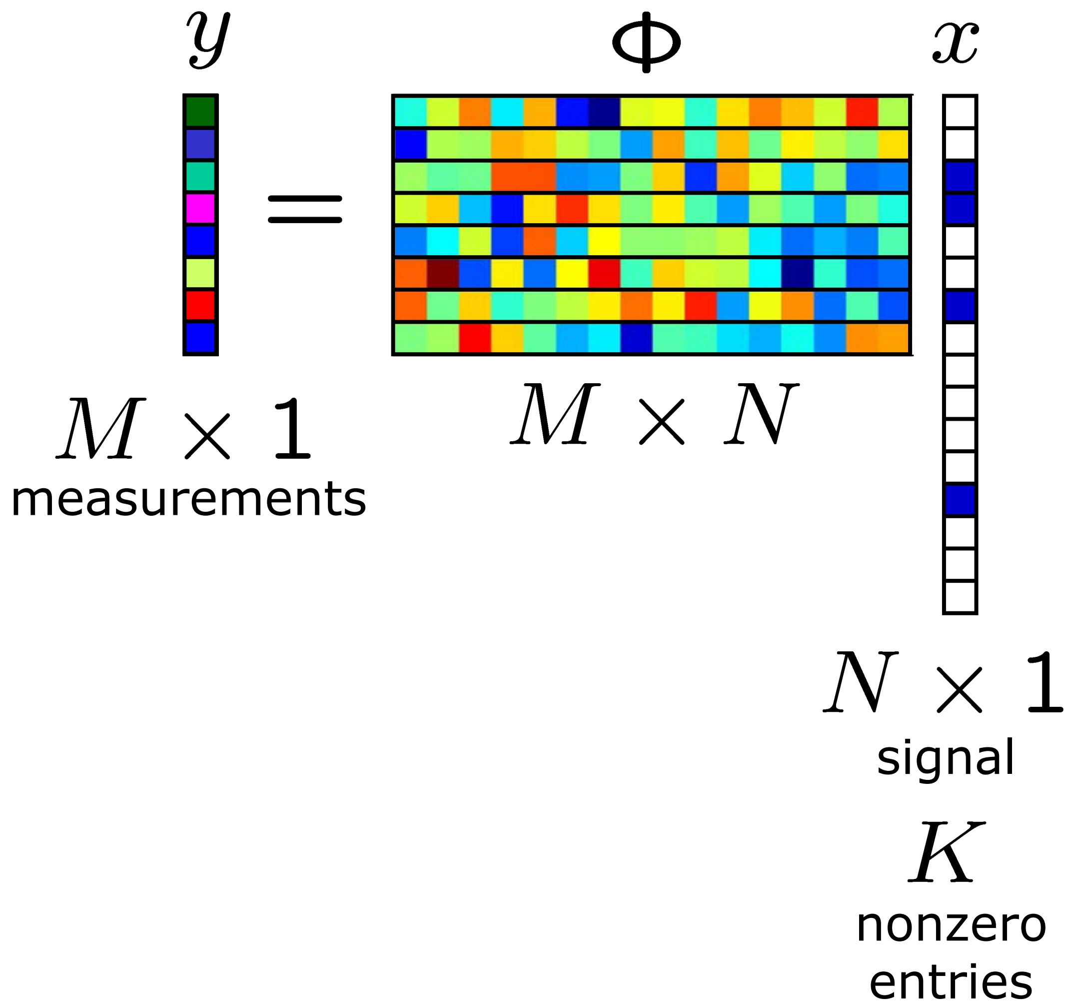
signal

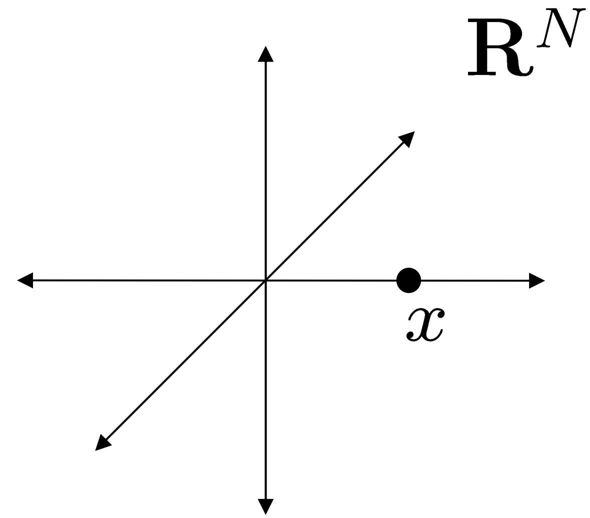
K

nonzero
entries

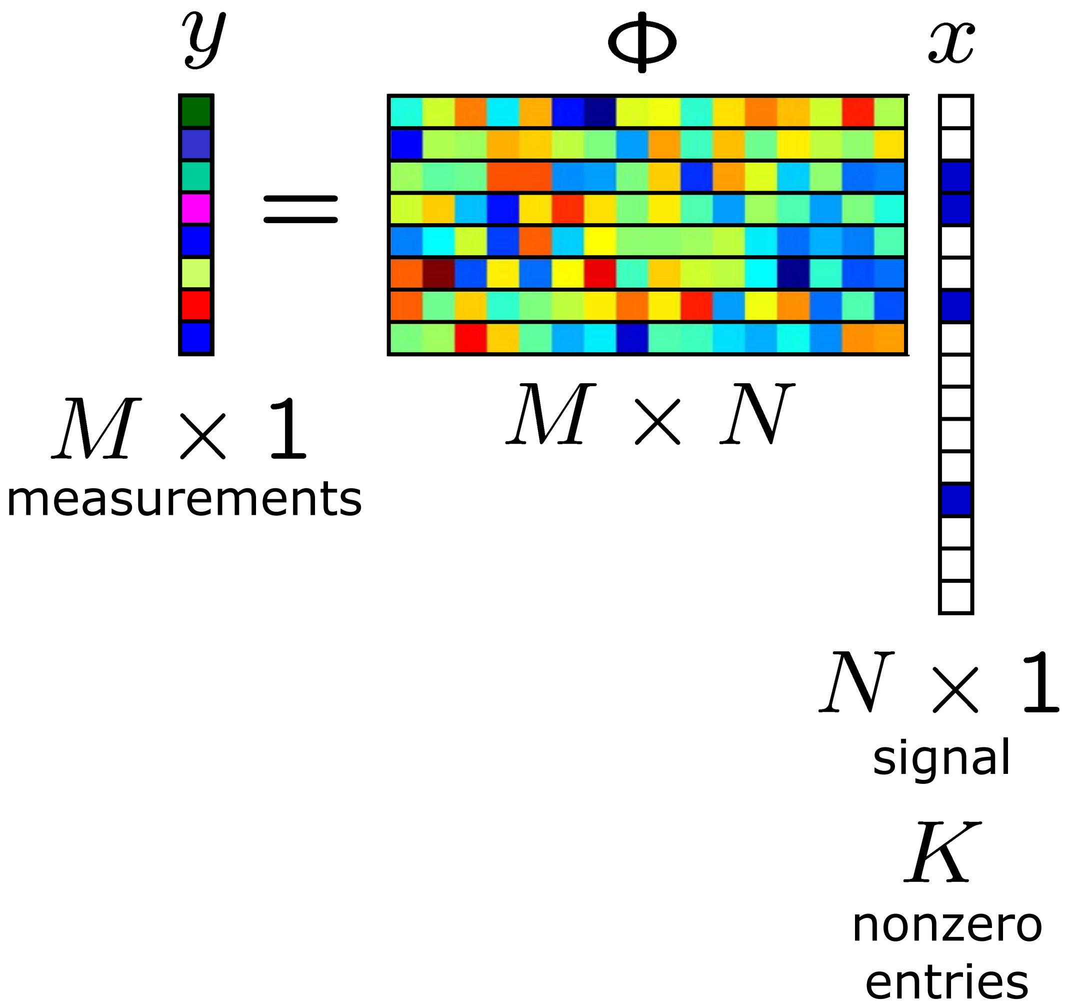


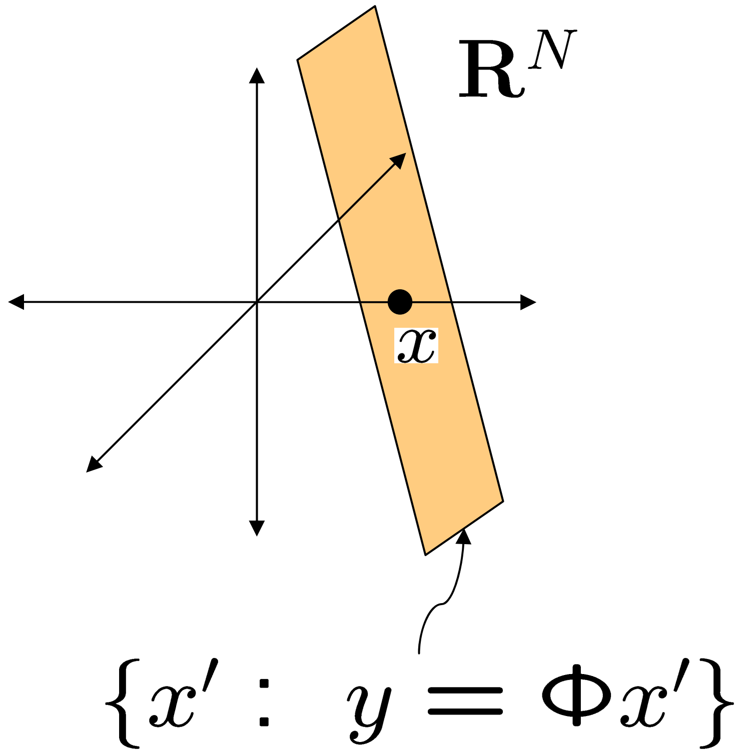
The Geometry of L_1 Recovery



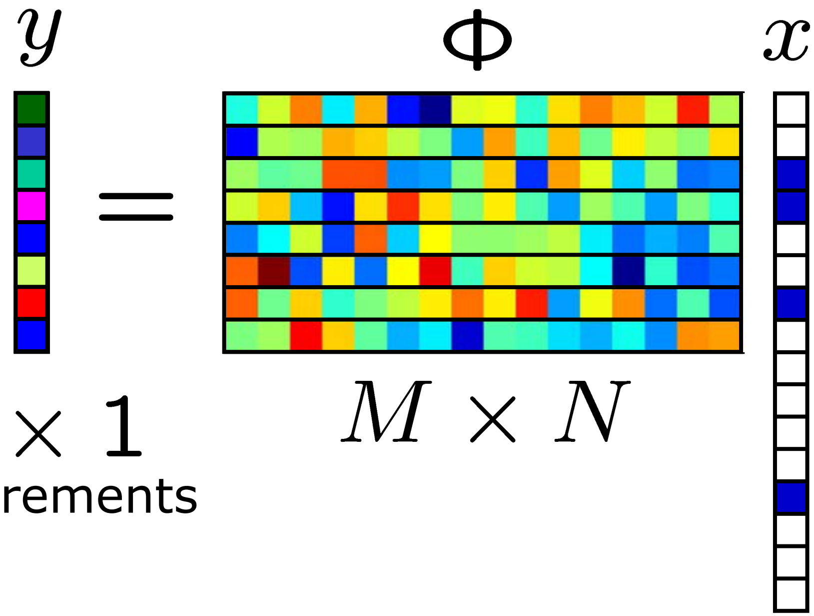


The Geometry of L_1 Recovery





The Geometry of L_1 Recovery

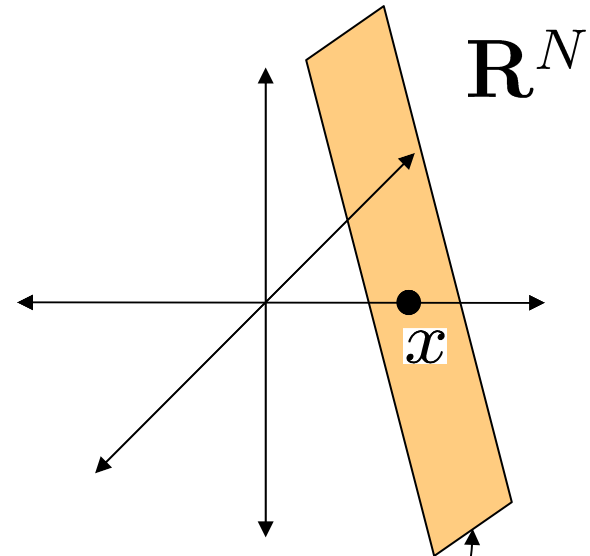


$M \times 1$
measurements

$M \times N$

$N \times 1$
signal

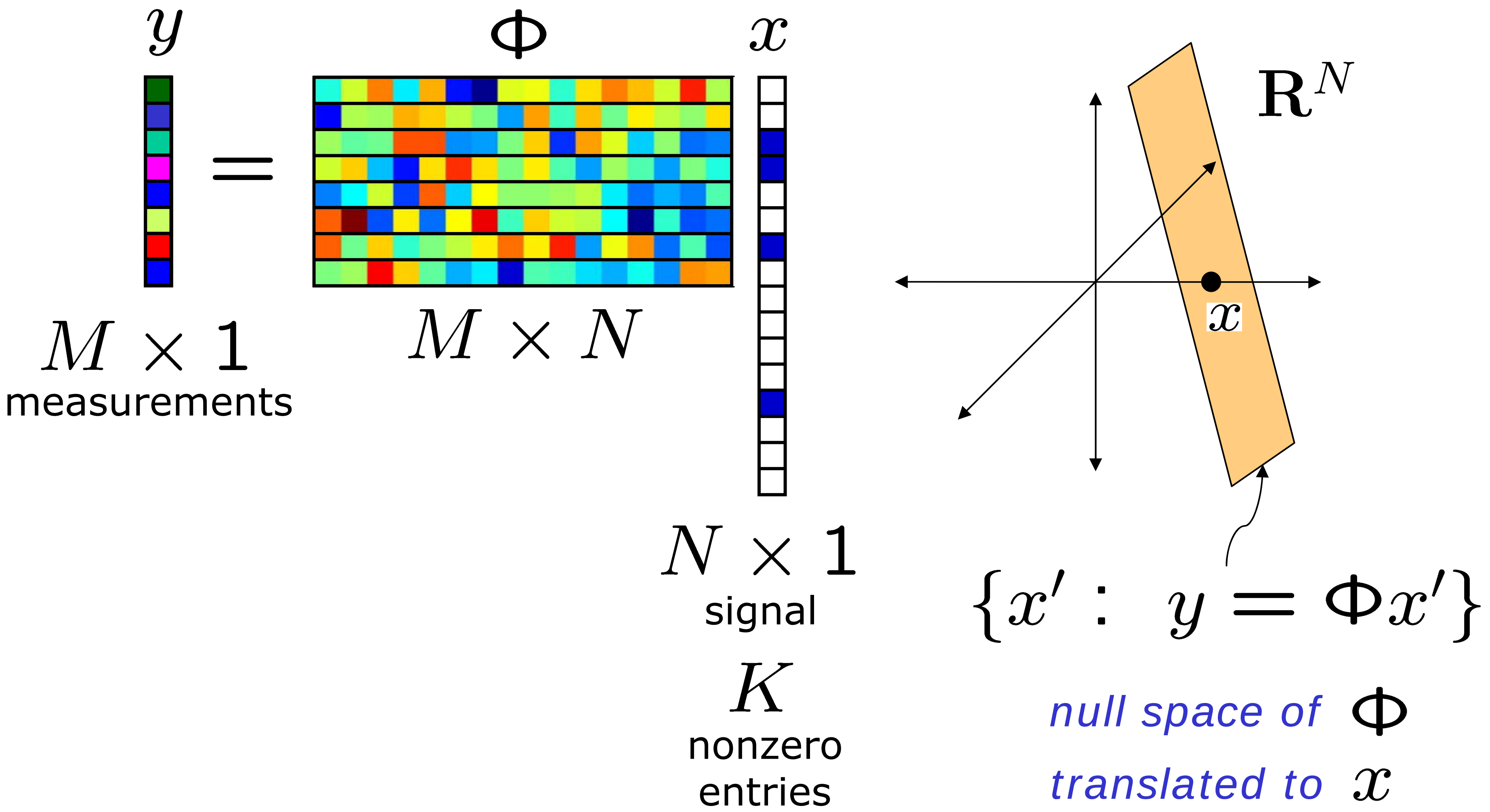
K
nonzero
entries



$$\{x' : \ y = \Phi x'\}$$

null space of Φ
translated to x

The Geometry of L_1 Recovery



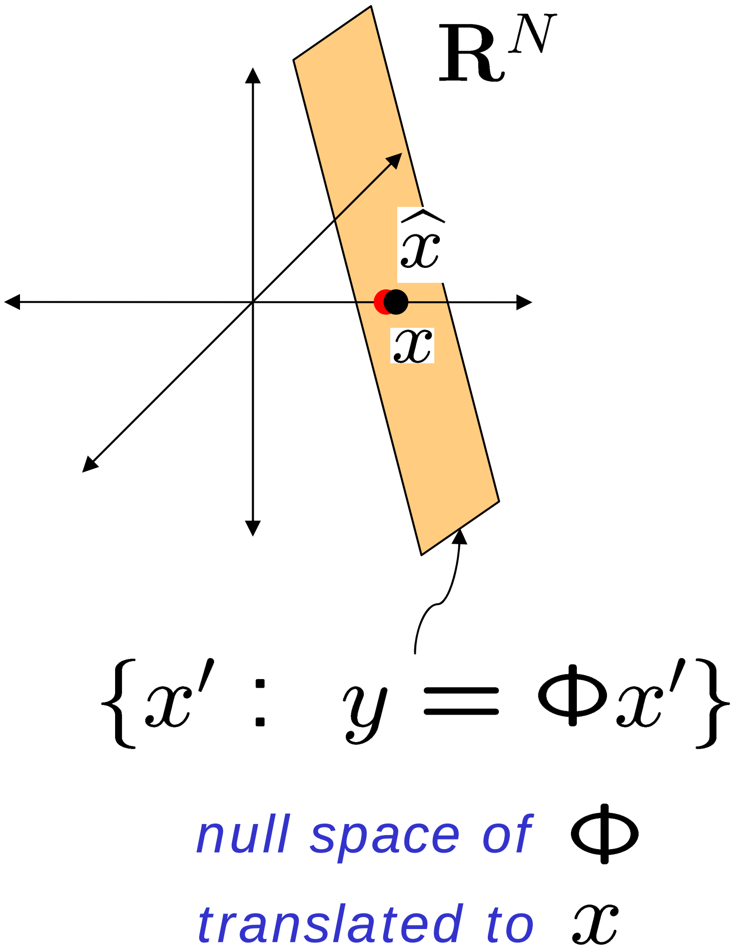
null space of Φ

translated to x

random orientation
dimension N-M

L_0 Recovery Works

$$\widehat{x} = \arg\min_{y=\Phi x'} \|x'\|_0$$

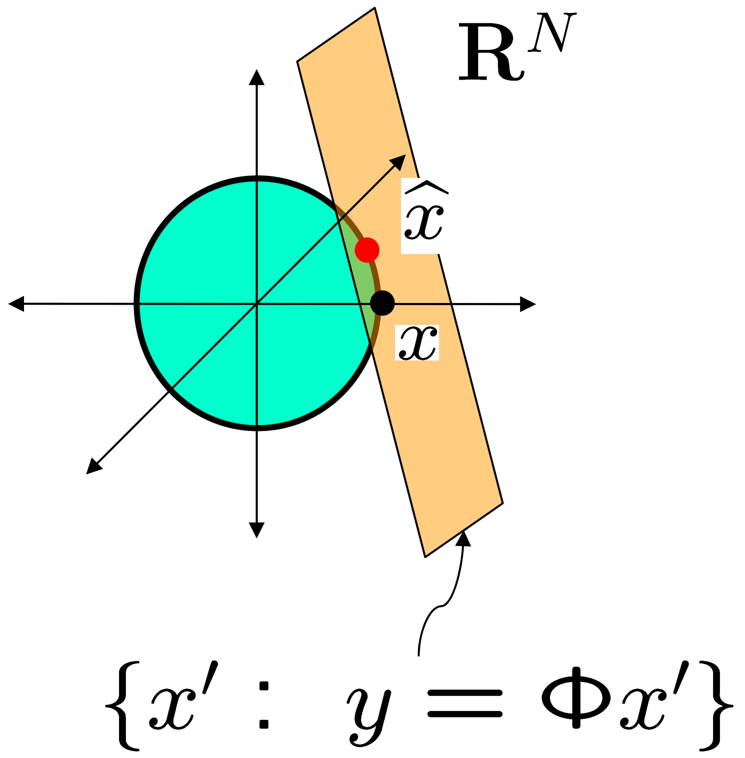


minimum L_0 solution correct
if $M \geq 2K$

(w.p. 1 for Gaussian Φ)

Why L_2 Doesn't Work

$$\widehat{x} = \arg \min_{y=\Phi x'} \|x'\|_2$$



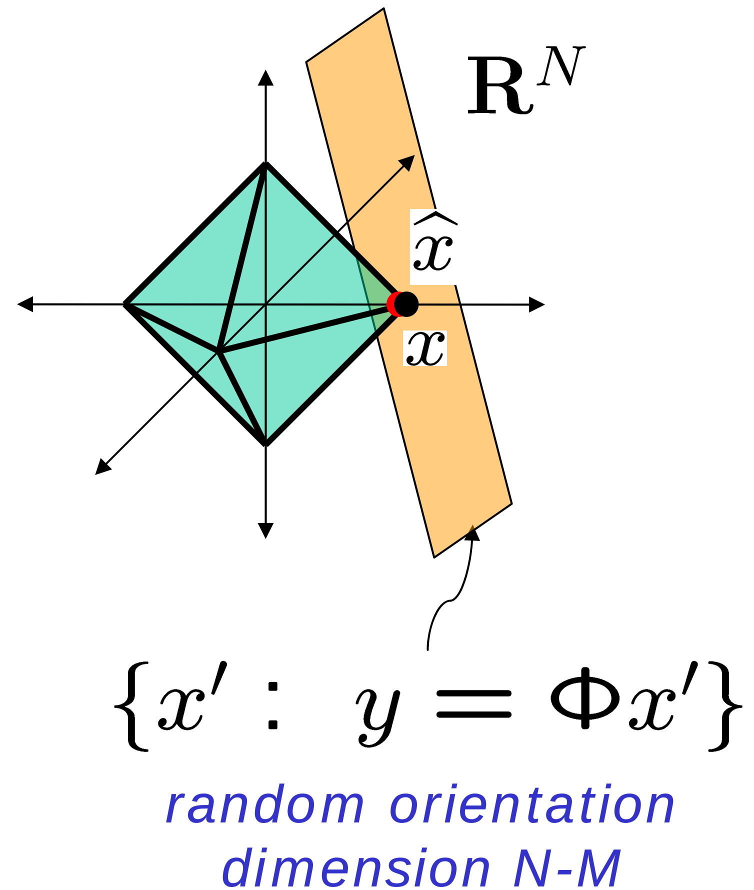
least squares,
minimum L_2 solution
is almost **never sparse**

Why L_1 Works

$$\widehat{x} = \arg\min_{y=\Phi x'} \|x'\|_1$$

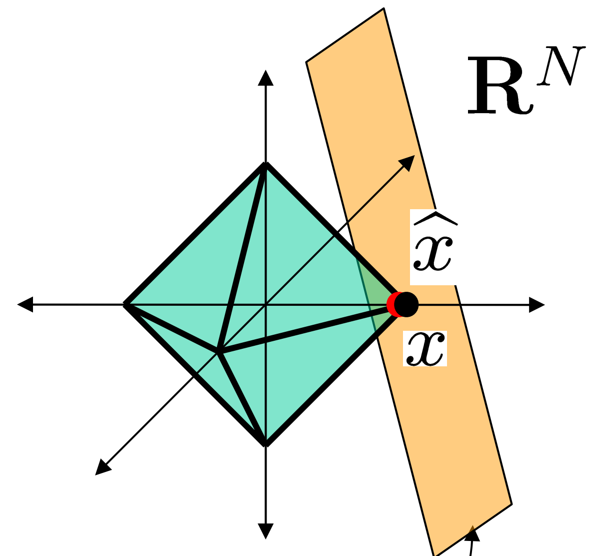
minimum L_1 solution
= L_0 sparsest solution if

$$M \approx K \log N \ll N$$



Why L_1 Works

$$\widehat{x} = \arg \min_{y=\Phi x'} \|x'\|_1$$



$$\{x' : \ y = \Phi x'\}$$

random orientation dimension N-M

Criterion for success:

Ensure with high probability that a *randomly oriented (N-M)-plane*, *anchored on a K-face* of the L_1 ball, *will not intersect* the ball.

Want K small, (N-M) small (i.e., M large)

L_0/L_1 Equivalence
[Donoho, Tanner]

Theorem.

For Gaussian Φ, require

$$M \sim 2eK \log\left(\frac{N}{M\sqrt{\pi}}\right)$$

measurements to recover every K-sparse signal and

$$M \sim 2K \log\left(\frac{N}{M}\right)$$

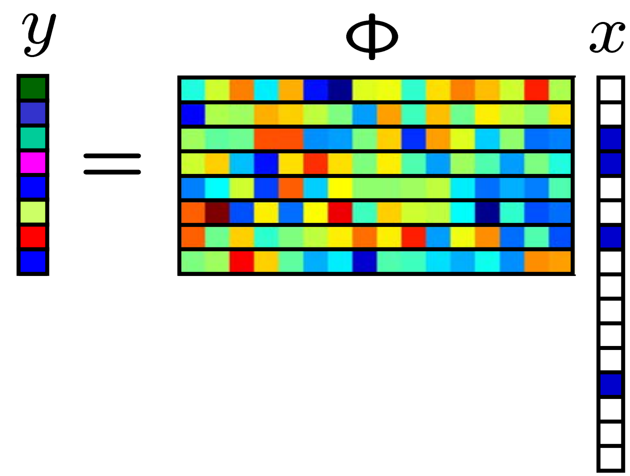
measurements to recover a large majority of K-sparse x. (These bounds are sharp asymptotically.)

Proof (geometric): Face-counting of randomly projected polytopes

Restricted Isometry Property (aka UUP)

[Candès, Romberg, Tao]

- Measurement matrix Φ has ***RIP of order K*** if



$$(1-\delta_K) \leq \frac{\|\Phi x\|_2^2}{\|x\|_2^2} \leq (1+\delta_K)$$

for all K-sparse signals x.

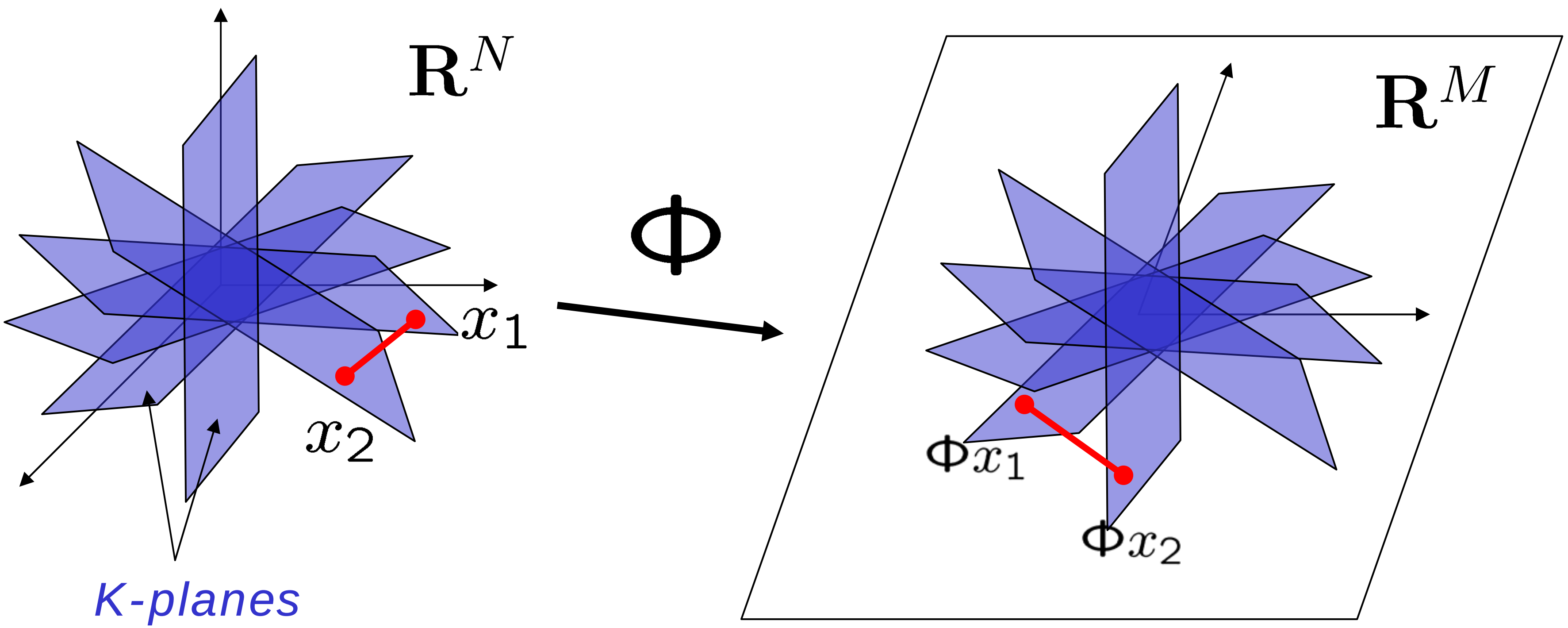
- Does *not* hold for $K > M$; may hold for smaller K.

- Implications: tractable, stable, robust recovery

RIP as a "Stable" Embedding

- RIP of order $2K$ implies: for all K-sparse x_1 and x_2,

$$(1-\delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta_{2K})$$



(if $\delta_{2K} < 1$ have injectivity; smaller δ_{2K} more stable)

Implications of RIP

[Candès (+ et al.); see also Cohen et al., Vershynin et al.]

If δ_{2K} < 0.41, ensured:

1. *Tractable recovery: All K-sparse x* are perfectly recovered via ℓ_1 minimization.

2. *Robust recovery: For any $x \in R^N$,*

$$\|x-\widehat{x}\|_{\ell_1} \le C\|x-x_K\|_{\ell_1} \text{ and } \|x-\widehat{x}\|_{\ell_2} \le C\frac{\|x-x_K\|_{\ell_1}}{K^{1/2}}.$$

3. *Stable recovery:* Measure $y = \Phi x + e$, with $||e||_2 < \varepsilon$, and recover

$$\widehat{x} = \arg\min \|x'\|_1 \ s.t.\ \|y-\Phi x'\|_2 \le \epsilon.$$

Then *for any $x \in R^N$,*

$$\|x-\widehat{x}\|_{\ell_2} \le C_1\frac{\|x-x_K\|_{\ell_1}}{K^{1/2}} + C_2\epsilon.$$

Verifying RIP:
How Many Measurements?

- Want RIP of order 2K (say) to hold for MxN Φ
  - difficult to verify for a given Φ
  - requires checking eigenvalues of each submatrix

- Prove ***random*** Φ will work
  - ***iid Gaussian entries***
  - ***iid Bernoulli entries (+/- 1)***
  - ***iid subgaussian entries***
  - ***random Fourier ensemble***
  - ***random subset of incoherent dictionary***

- In each case, ***$M = O(K \log N)$*** suffices
  - with very high probability, usually $1-O(e^{-CN})$
  - slight variations on log term
  - some proofs complicated, others simple (more soon)

Optimality

[Candès; Donoho]

- Gaussian Φ has RIP order 2K (say) with ***M = O(K log(N/M))***

- Hence, for a given M, for $x \in wl_p$ (i.e., $|x|_{(k)} \sim k^{-1/p}$), $0 < p < 1$, (or $x \in l_1$)

$$
\begin{aligned}
\|x - \widehat{x}\|_{\ell_2} &\leq CK^{-1/2}\|x - x_K\|_{\ell_1} \\
&\leq CK^{1/2-1/p} \\
&\leq C(M/\log(N/M))^{1/2-1/p}
\end{aligned}
$$

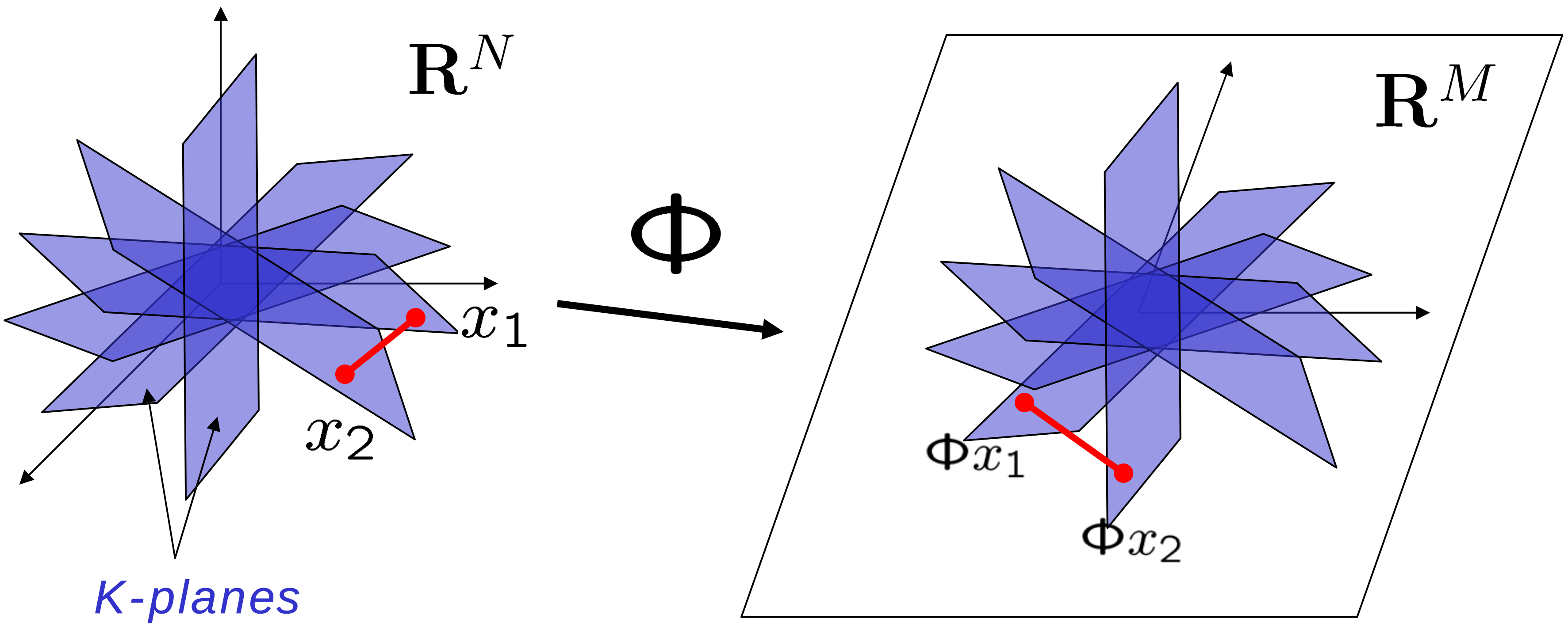
- Up to a constant, these bounds are *optimal*: no other linear mapping to R^M followed by *any* decoding method could yield lower reconstruction error over classes of compressible signals

- <u>Proof (geometric):</u> Gelfand n-widths [Kashin; Gluskin, Garnaev]

Recall: RIP as a "Stable" Embedding

- RIP of order $2K$ implies: for all K-sparse x_1 and x_2,

$$(1-\delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta_{2K})$$

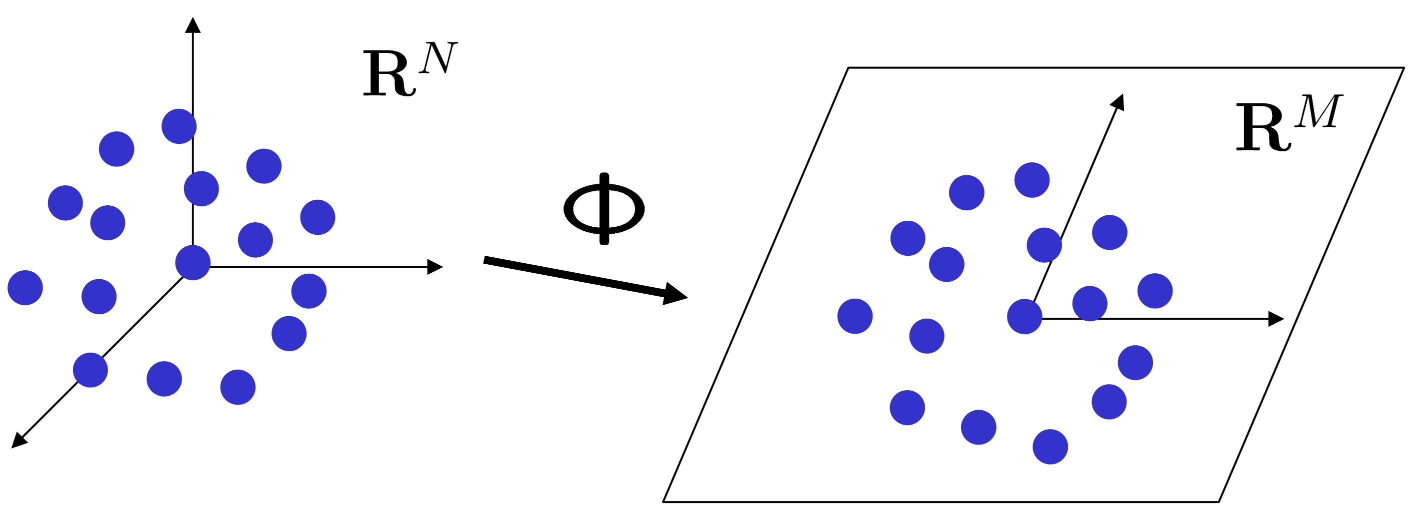


Johnson-Lindenstrauss Lemma

[see also Dasgupta, Gupta; Frankl, Maehara; Achlioptas; Indyk, Motwani]

Consider a point set $Q \subset R^N$ and random* $M \times N$ Φ with $M = O(\log(\#Q)\, \varepsilon^{-2})$. With high prob., for all $x_1, x_2 \in Q$,

$$(1-\epsilon) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\epsilon).$$



Proof via *concentration inequality*: For any $x \in R^N$

$$\mathbf{P}(|\,\|\Phi x\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-\frac{M}{2}(\epsilon^2/2 - \epsilon^3/3)}.$$

Favorable JL Distributions

- Gaussian

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

- Bernoulli/Rademacher [Achlioptas]

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

- "Database-friendly" [Achlioptas]

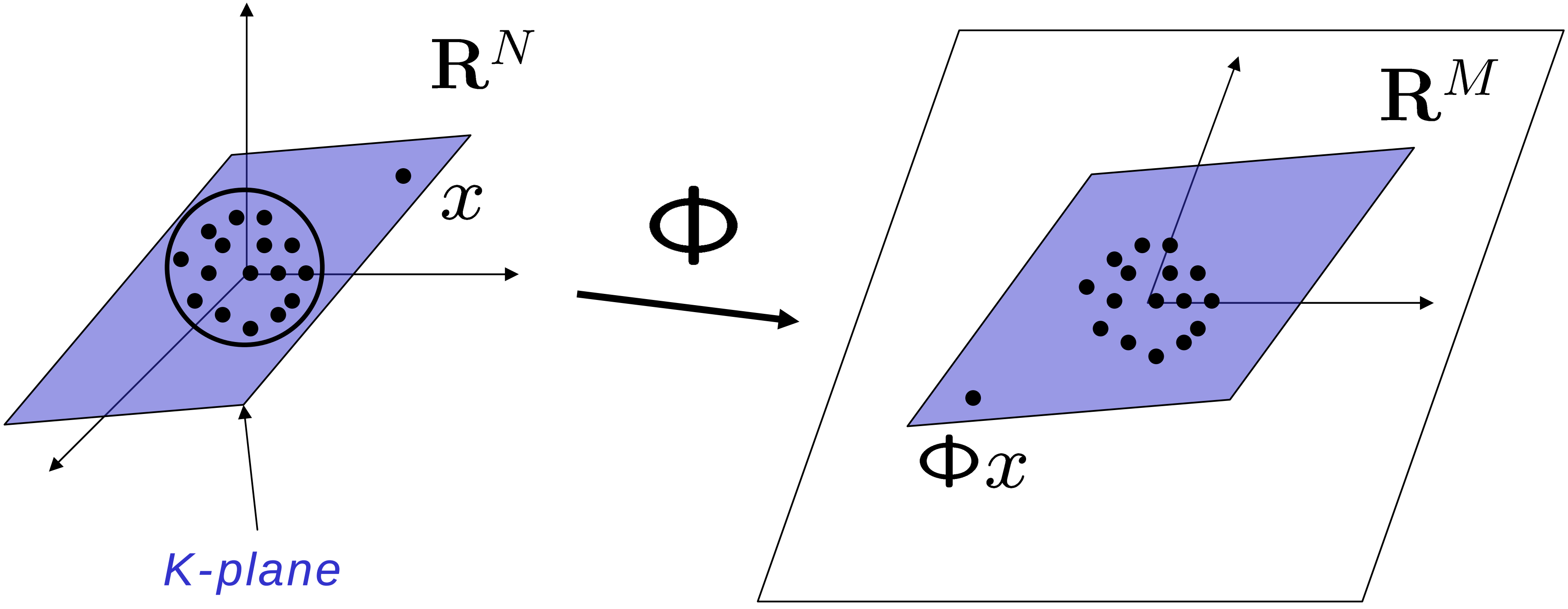
$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability} \quad \frac{1}{6}, \\ 0 & \text{with probability} \quad \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability} \quad \frac{1}{6} \end{cases}$$

- Random Orthoprojection to R^M [Gupta, Dasgupta]

Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

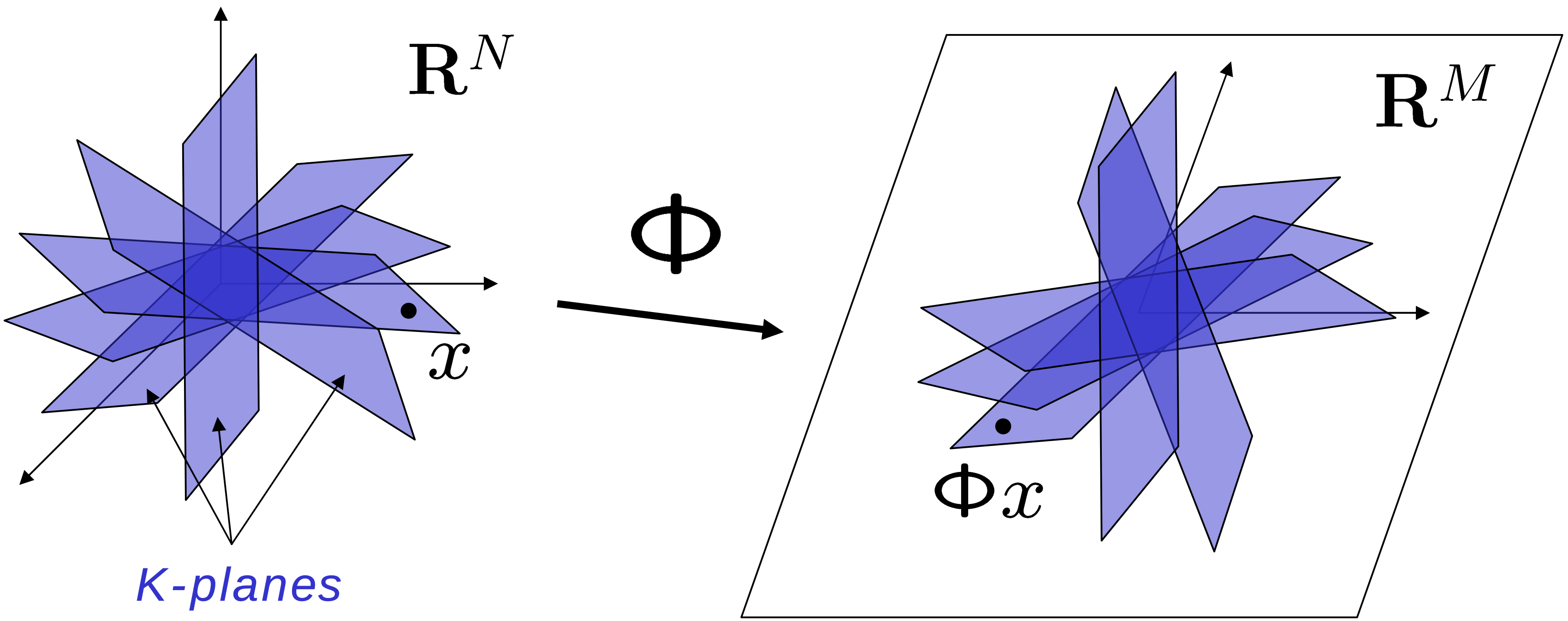
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all $q \in Q$
- extend to isometry for all x in K-plane



Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all q $\in$ Q
- extend to isometry for all x in K-plane
- union bound ➔ isometry for all K-planes



Connecting JL to RIP

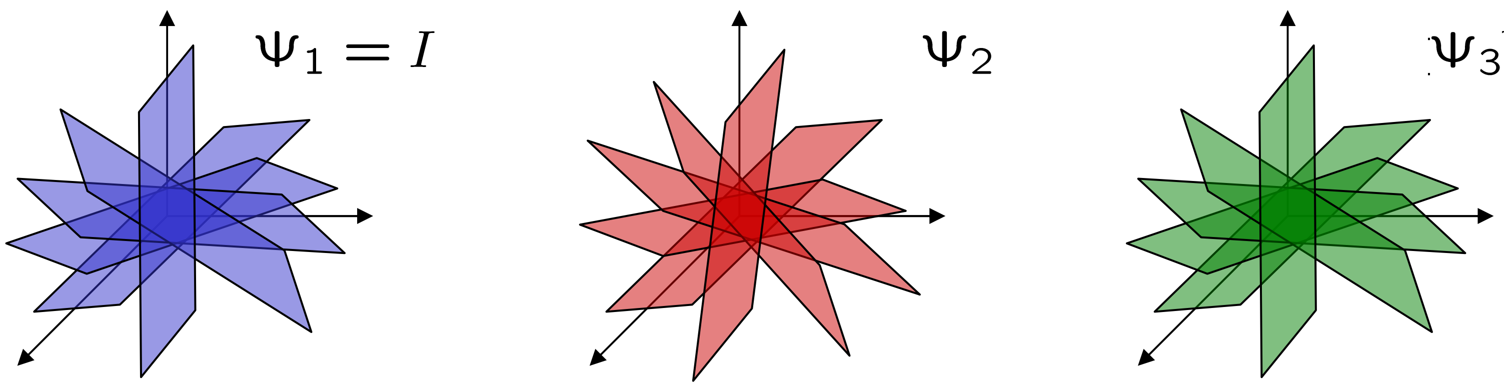
[with R. DeVore, M. Davenport, R. Baraniuk]

- **Theorem:** Supposing Φ is drawn from a JL-favorable distribution,* then with probability at least $1\text{-}e^{-C^*M}$, Φ meets the RIP with

$$K \leq \mathrm{C} \cdot \frac{M}{\log(N/M)+1}.$$

* Gaussian/Bernoulli/database-friendly/orthoprojector

- Bonus: *universality* (repeat argument for any Ψ)

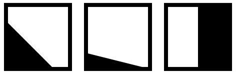


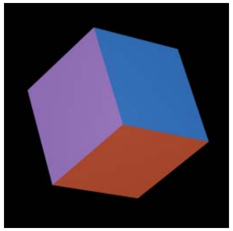
- See also Mendelson et al. concerning subgaussian ensembles

Beyond Sparsity

- *Not all signal models fit into sparse representations*

- Other concise notions
  - constraints
  - degrees of freedom
  - parametrizations
  - articulations
  - signal families





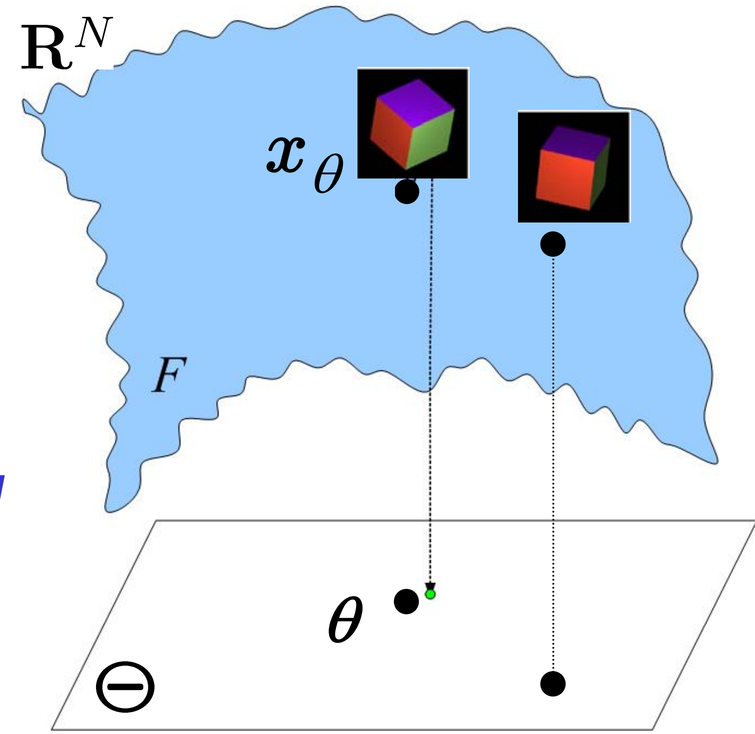
"information level" $\ll$ sparsity level $\ll N$

Challenge: How to exploit these concise models?

Manifold Models

- K-dimensional *parameter* $\theta \in \Theta$ captures degrees of freedom in signal $x_\theta \in R^N$

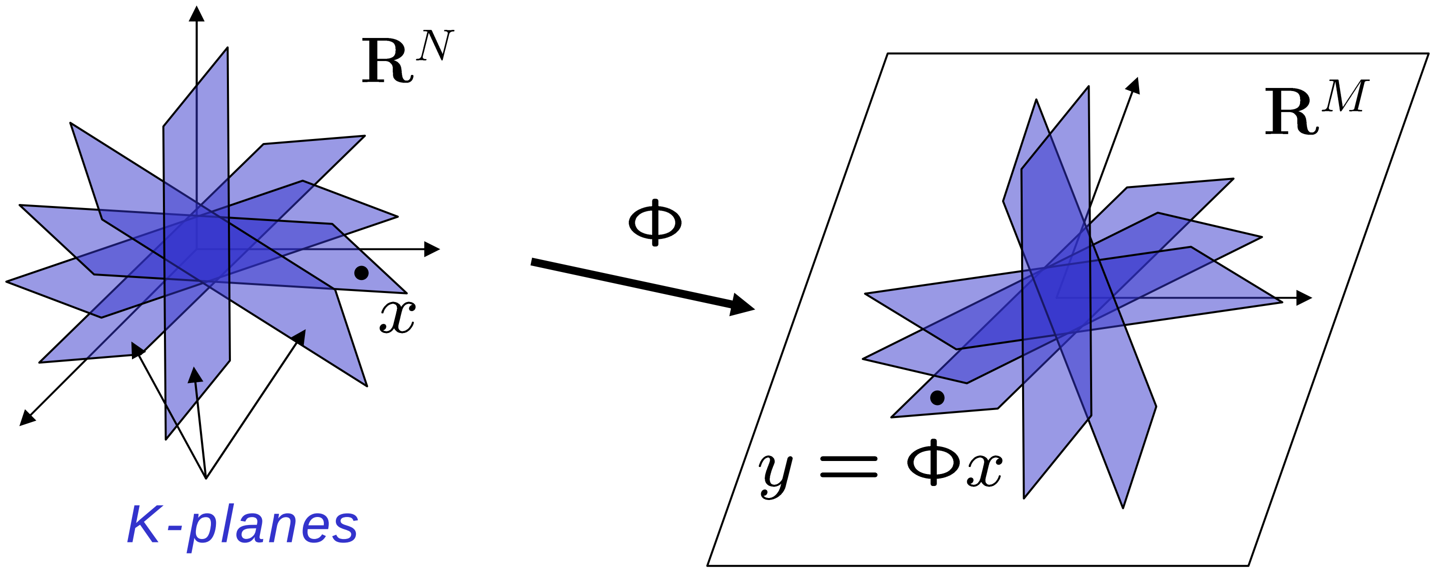
- Signal class $F = \{x_\theta: \theta \in \Theta\}$ forms a K-dimensional *manifold*
  - also nonparametric collections: faces, handwritten digits, shape spaces, etc.



- Dimensionality reduction and manifold learning
  - embeddings [ISOMAP; LLE; HLLE; …]
  - harmonic analysis [Belkin; Coifman; …]

Random Projections

- *Random projections preserve information*
  - Compressive Sensing (sparse signal embeddings)
  - Johnson-Lindenstrauss lemma (point cloud embeddings)

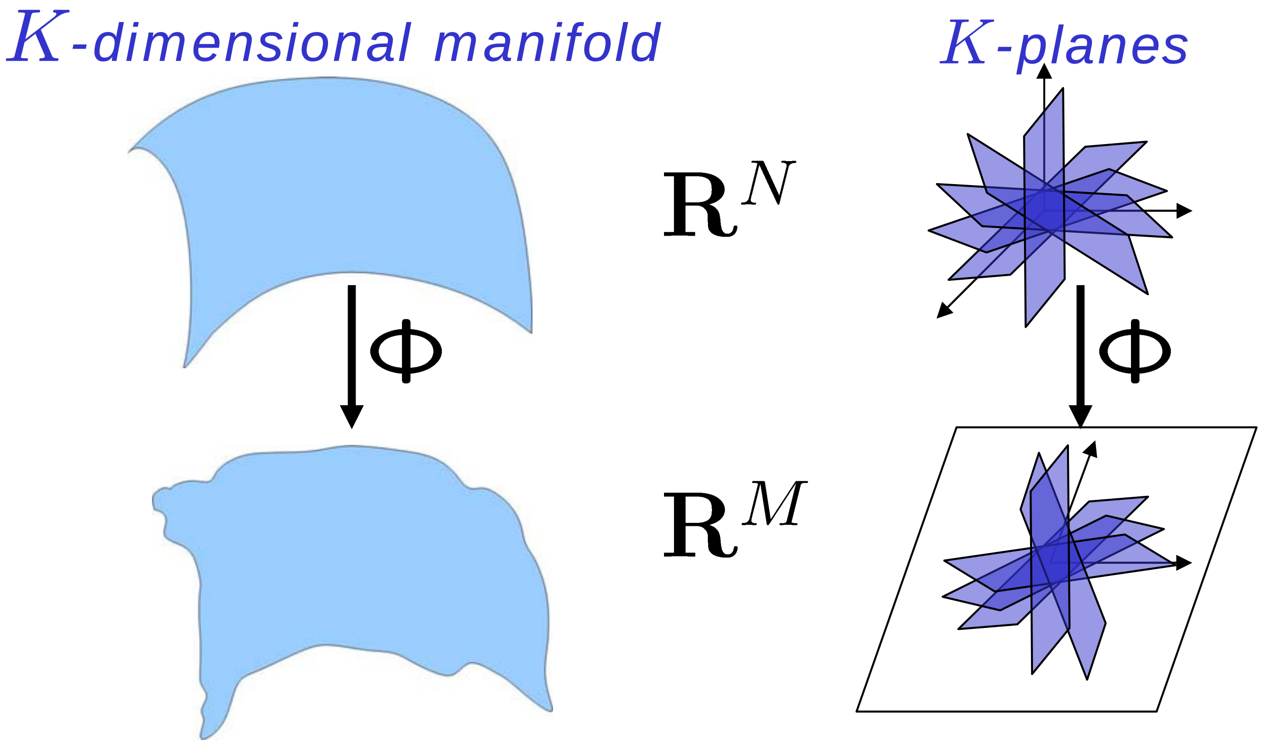


- What about *manifolds*?

Whitney's Embedding Theorem

(1936)

- $M > 2K$ random measurements for embedding



- But… *stable* CS recovery requires $M \sim K \log(N)$:

$$(1-\delta_{2K}) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2^2}{\|x_1 - x_2\|_2^2} \leq (1+\delta_{2K})$$

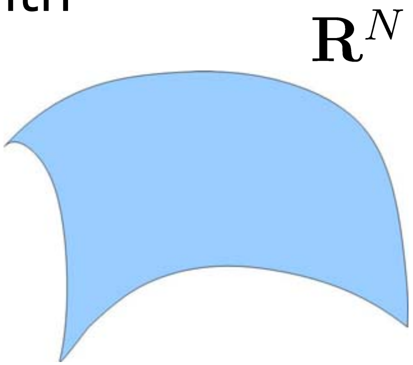
Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

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



Stable Manifold Embedding

[with R. Baraniuk]

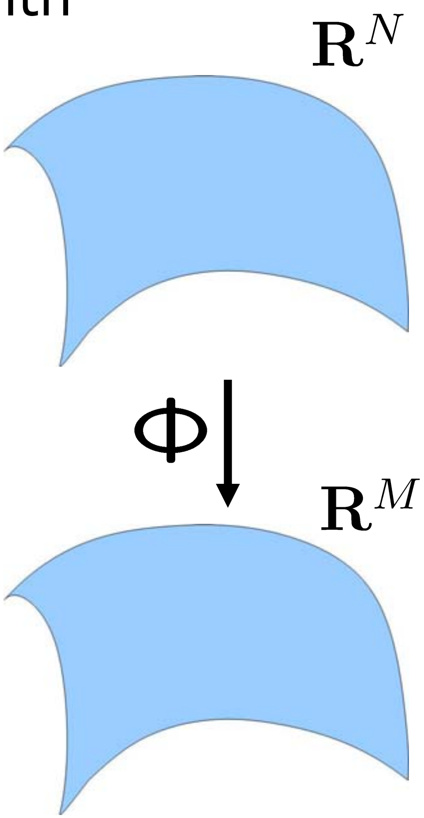
Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

- condition number *1/τ* (curvature, self-avoiding)
- volume V

Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$



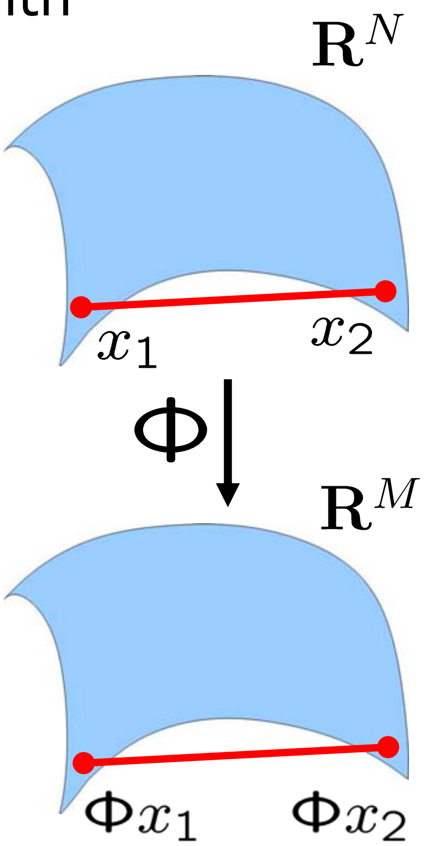
Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

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



Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$

Then with probability at least $1-\rho$, the following statement holds: For every pair $x_1, x_2 \in F$,

$$(1-\epsilon) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon).$$

Stable Manifold Embedding

[with R. Baraniuk]

Theorem:

Let $F \subset R^N$ be a compact *K-dimensional manifold* with

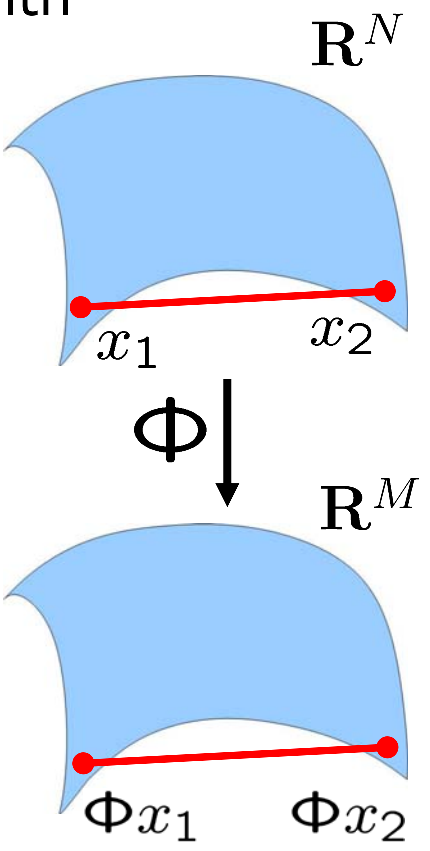
- condition number *1/*τ (curvature, self-avoiding)
- volume V

Let Φ be a random MxN orthoprojector with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right).$$

Then with probability at least *1-*ρ, the following statement holds: For every pair $x_1, x_2 \in F$,

$$(1-\epsilon) \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon).$$



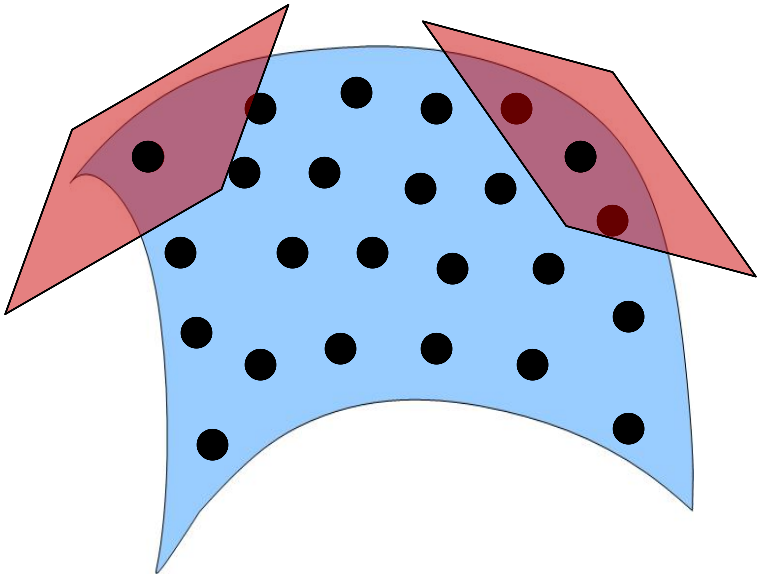
Stable Manifold Embedding

Sketch of proof:

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

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right)$$

- extend to entire manifold



Implications: Many key properties preserved in R^M

- ambient and geodesic distances
- dimension and volume of the manifold
- path lengths and curvature
- topology, local neighborhoods, angles, etc…

Summary – Geometry in CS

- ***Concise models → low-dimensional geometry***
  - bandlimited
  - sparse
  - manifolds

- ***Random Projections***
  - stable embedding thanks to low-dimensional geometry
  - model-based recovery; use the best model available

- ***Compressed Sensing + L_1 minimization***
  - powerful results for explicit, multi-purpose recovery algorithm

- ***Manifolds & other models***
  - specialized algorithms may be required; but apps beyond CS

References – Geometry (1)

L_0 Recovery:

- Dror Baron, Michael Wakin, Marco Duarte, Shriram Sarvotham, and Richard Baraniuk, Distributed compressed sensing. (Preprint, 2005)
- P. Feng and Y. Bresler, "Spectrum-blind minimum-rate sampling and reconstruction of multiband signals," in Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc., Atlanta, GA, 1996, vol. 2, pp. 1689–1692.

L_1 Recovery & Random Polytopes:

- David Donoho and Jared Tanner, Counting faces of randomly-projected polytopes when the projection radically lowers dimension. (Submitted to Journal of the AMS)

Optimality & n-widths:

- David Donoho, Compressed sensing. (IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)
- Emmanuel Candès and Terence Tao, Near optimal signal recovery from random projections: Universal encoding strategies? (IEEE Trans. on Information Theory, 52(12), pp. 5406 - 5425, December 2006)

References – Geometry (2)

RIP/UUP & Implications:

- Emmanuel Candès and Terence Tao, Decoding by linear programming. (IEEE Trans. on Information Theory, 51(12), December 2005)
- David Donoho, For most large underdetermined systems of linear equations, the minimal ell-1 norm solution is also the sparsest solution. (Communications on Pure and Applied Mathematics, 59(6), June 2006)
- Emmanuel Candès, Justin Romberg, and Terence Tao, Stable signal recovery from incomplete and inaccurate measurements. (Communications on Pure and Applied Mathematics, 59(8), August 2006)
- Emmanuel Candès and Terence Tao, The Dantzig Selector: Statistical estimation when p is much larger than n (To appear in Ann. Statistics)
- Rudelson, M., Vershynin, R., "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements." Preprint, 2006.
- Albert Cohen, Wolfgang Dahmen, and Ronald DeVore, Compressed sensing and best k-term approximation. (Preprint, 2006)
- Holger Rauhut, Karin Schass, and Pierre Vandergheynst, Compressed sensing and redundant dictionaries. (Preprint, 2006)
- Ronald A. DeVore, Deterministic constructions of compressed sensing matrices. (Preprint, 2007)
- Deanna Needell and Roman Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. (Preprint, 2007)

References – Geometry (3)

Johnson-Lindenstrausss Lemma:

- D. Achlioptas. Database-friendly random projections. In Proc. Symp. on Principles of Database Systems, pages 274–281. ACM Press, 2001.
- S. Dasgupta and A. Gupta. An elementary proof of the Johnson-Lindenstrauss lemma. Technical Report TR-99-006, Berkeley, CA, 1999.
- P. Frankl and H. Maehara, The Johnson-Lindenstrauss lemma and the sphericity of some graphs, J. Combinatorial Theory Ser. B 44 (1988), no. 3, pp. 355–362.
- P. Indyk and R. Motwani, Approximate nearest neighbors: Towards removing the curse of dimenstionality, Symp. on Theory of Computing, 1998, pp. 604–613.

Geometric proofs of RIP/UUP:

- Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, A simple proof of the restricted isometry property for random matrices. (To appear in Constructive Approximation)
- S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles. (Preprint, 2006)

Manifolds:

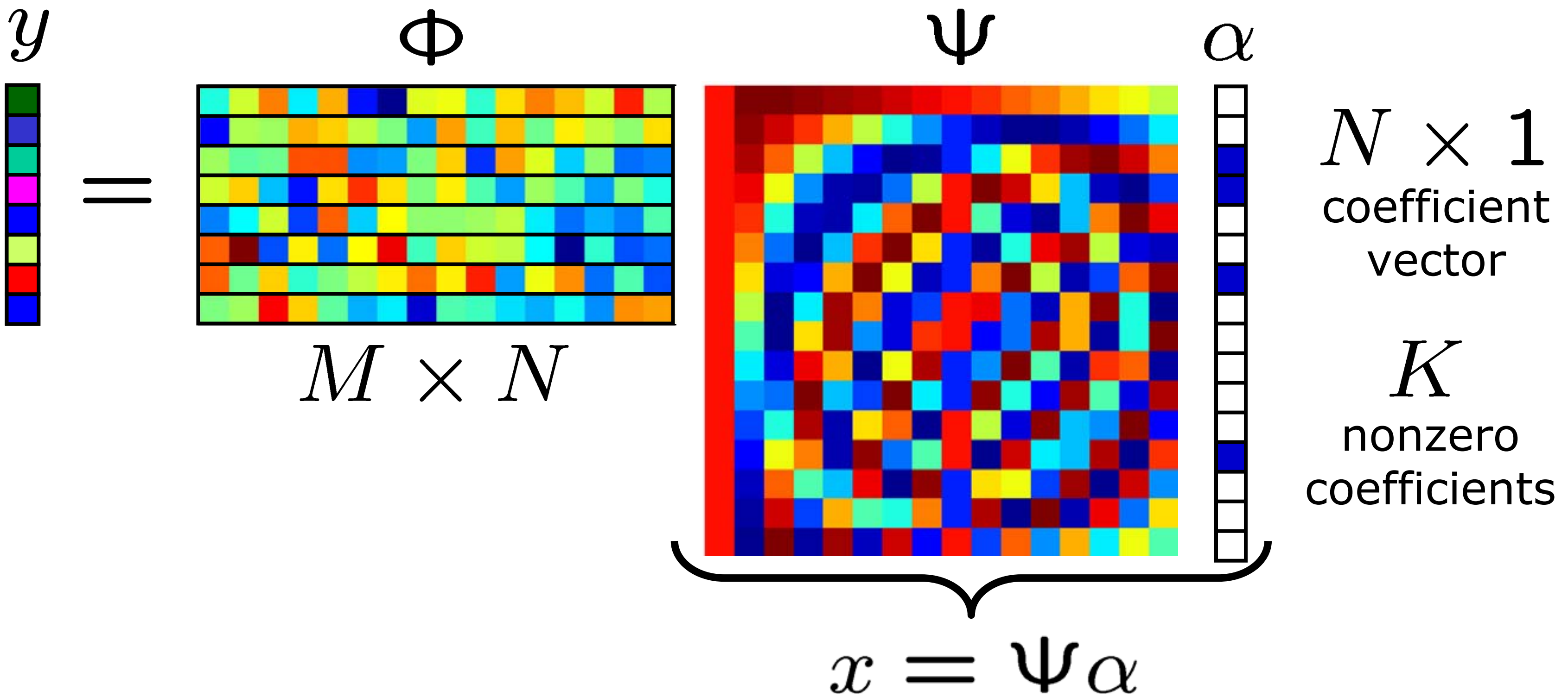
- Richard Baraniuk and Michael Wakin, Random projections of smooth manifolds. (To appear in Foundations of Computational Mathematics)

A Survey of CS Applications

[Thanks to Rich Baraniuk for contributed slides]

CS Paradigm

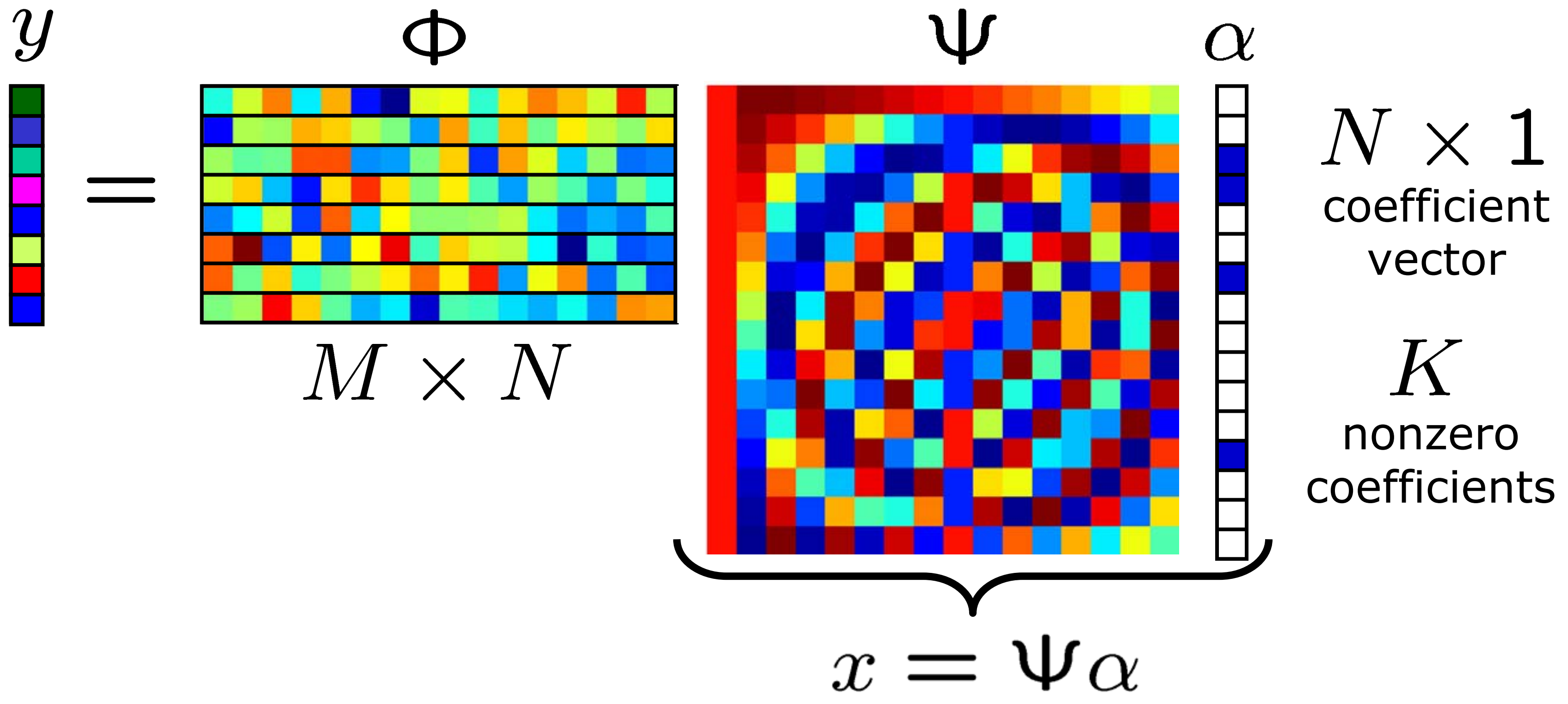
- K-term quality from just $K \log(N)$ measurements



- Robust and widely applicable

1. Data Compression

Idea: Use incoherent/random Φ to compress signal



CS Advantages

- Universal
  - same random projections / hardware can be used for *any* compressible signal class
  - generic and “future proof”: can try different signal models (different bases, statistical models, manifolds, …)

- Democratic
  - each measurement carries the same amount of information
  - simple encoding
  - robust to measurement loss and quantization

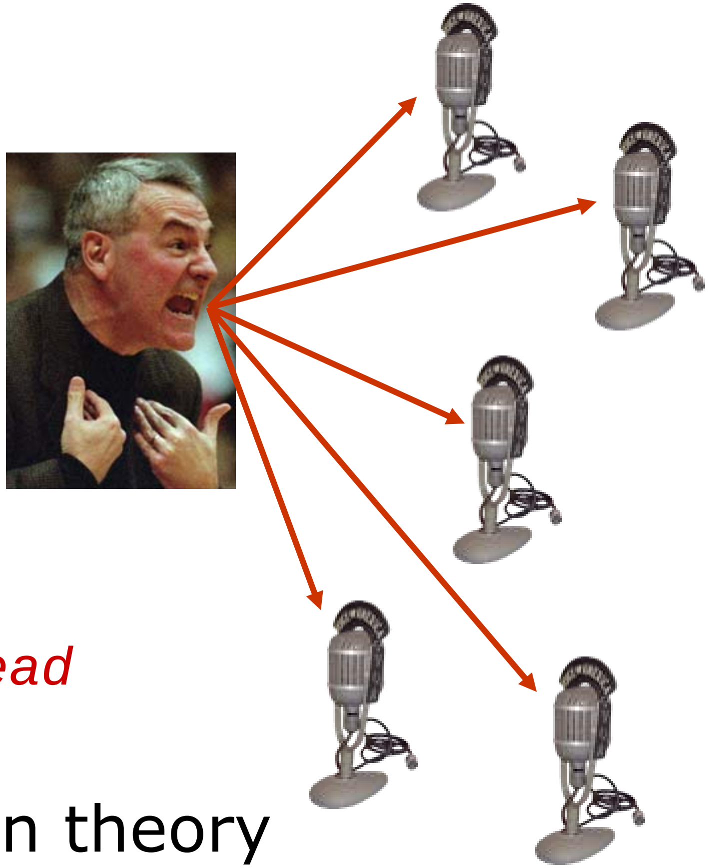
- Asymmetrical (most processing at decoder)

- Random projections weakly encrypted

- Possible application area: sensor networks

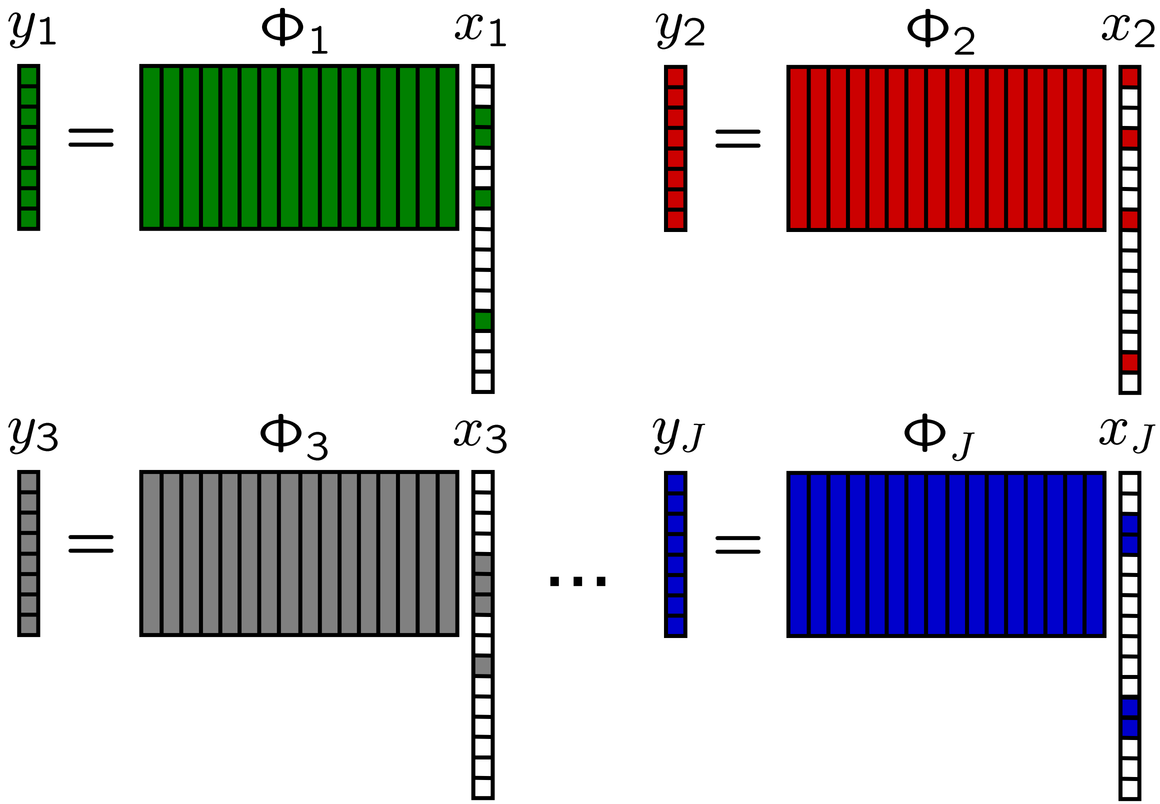
Multi-Signal Compressed Sensing

- Sensor networks:
  intra-sensor and
  inter-sensor correlation
- Can we exploit these to *jointly compress?*
- Popular approach: *collaboration*
  - inter-sensor *communication overhead*
- *Ongoing challenge* in information theory
- Solution: Compressed Sensing



Distributed CS (DCS)

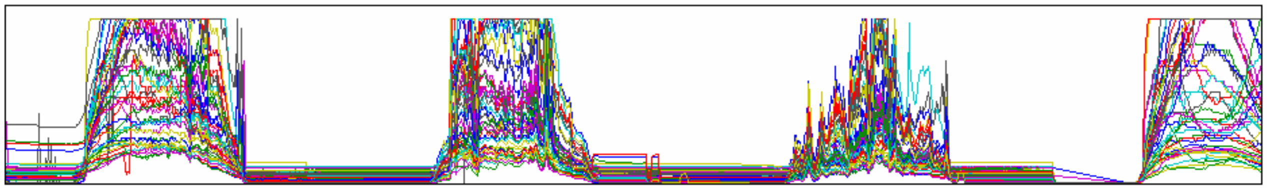
- "Measure separately, reconstruct *jointly*"



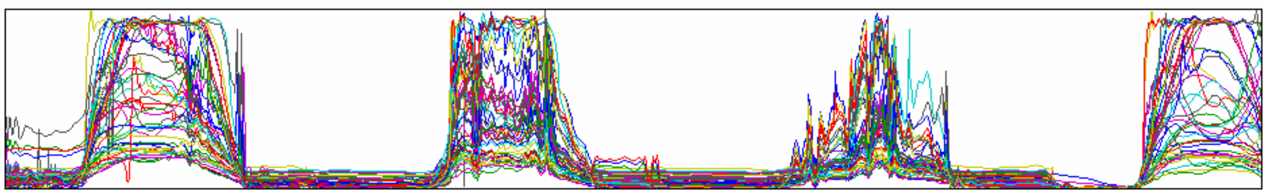
- Zero collaboration, trivially scalable, robust
- Low complexity, universal encoding

Real Data Example

- Light Sensing in Intel Berkeley Lab
- 49 sensors, $N = 1024$ samples each, Ψ = wavelets

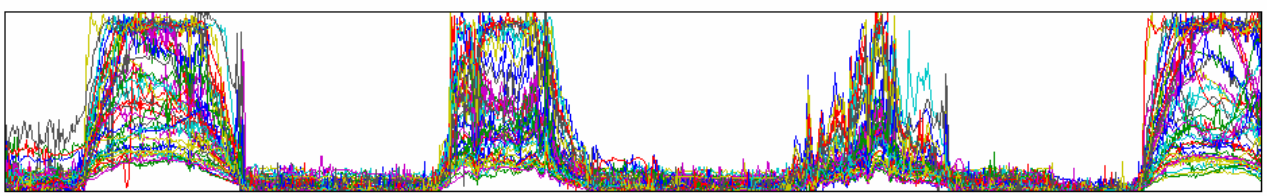


(a) Original



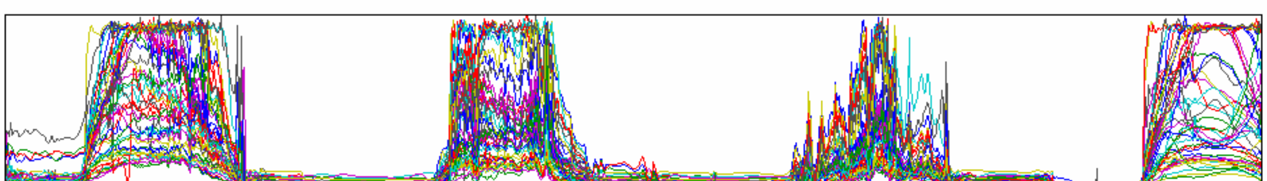
(b) Transform Coding, SNR = 26.4842 dB

K=100



(c) Compressed Sensing, SNR = 21.6426 dB

M=400



(d) Distributed Compressed Sensing, SNR = 27.1906 dB

M=400

Distributed CS (DCS)

- "Measure separately, reconstruct *jointly*"

- Ingredients
  - models for joint sparsity
  - algorithms for joint reconstruction
  - theoretical results for measurement savings

- The power of random measurements
  - *single-signal:* efficiently capture structure without performing the sparse transformation
  - *multi-signal:* efficiently capture joint structure without collaborating or performing the sparse transformation

- One of several CS applications to sensor networks

References – Data Compression (1)

Information Theoretic:

- Emmanuel Candès and Terence Tao, Near optimal signal recovery from random projections: Universal encoding strategies? (IEEE Trans. on Information Theory, 52(12), pp. 5406 - 5425, December 2006)
- David Donoho, Compressed sensing. (IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306, April 2006)
- Emmanuel Candès and Justin Romberg, Encoding the ell-p ball from limited measurements. (Proc. IEEE Data Compression Conference (DCC), Snowbird, UT, 2006)
- Shriram Sarvotham, Dror Baron, and Richard Baraniuk, Measurements vs. bits: Compressed sensing meets information theory. (Proc. Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2006)
- Petros Boufounos and Richard Baraniuk, Quantization of sparse representations. (Rice ECE Department Technical Report TREE 0701 - Summary appears in Proc. Data Compression Conference (DCC), Snowbird, Utah, March 2007)

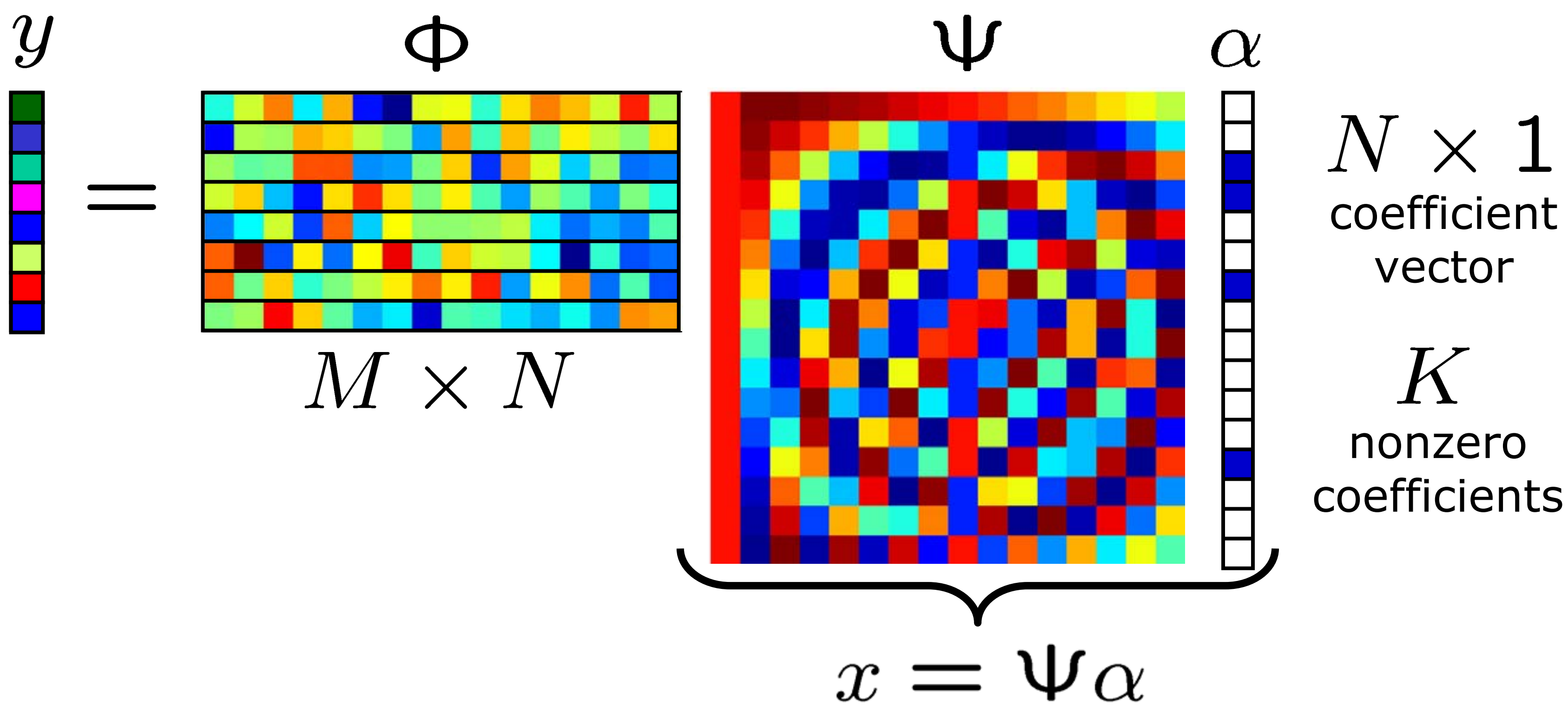
References – Data Compression (2)

Sensor Networks and Multi-Signal CS:

- Dror Baron, Michael Wakin, Marco Duarte, Shriram Sarvotham, and Richard Baraniuk, Distributed compressed sensing. (Preprint, 2005)
- Waheed Bajwa, Jarvis Haupt, Akbar Sayeed, and Rob Nowak, Compressive wireless sensing. (Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Nashville, Tennessee, April 2006)
- Rémi Gribonval, Holger Rauhut, Karin Schnass, and Pierre Vandergheynst, Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms. (Preprint, 2007)
- Wei Wang, Minos Garofalakis, and Kannan Ramchandran, Distributed sparse random projections for refinable approximation. (Proc. Int. Conf. on Information Processing in Sensor Networks (IPSN), Cambridge, Massachusetts, April 2007)

2. Compressive Signal Processing

Idea: Signal inference from compressive measurements



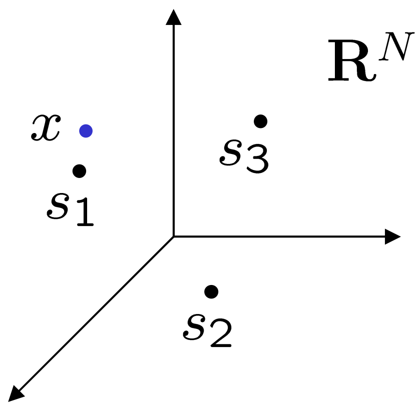
Information Scalability

- If we can ***reconstruct*** a signal from compressive measurements, then we should be able to perform other kinds of statistical signal processing:
  - ***detection***
  - ***classification***
  - ***estimation …***
- Number of measurements should relate to complexity of inference

Multiclass Likelihood Ratio Test (LRT)

- Observe one of P known signals in noise

$$
\begin{aligned}
H_1 &: \; x = s_1 + n \\
H_2 &: \; x = s_2 + n \\
&\;\;\vdots \\
H_P &: \; x = s_P + n
\end{aligned}
$$



- Classify according to: $\displaystyle \arg\max_{j=1,\ldots,P} \; p(x|H_j)$

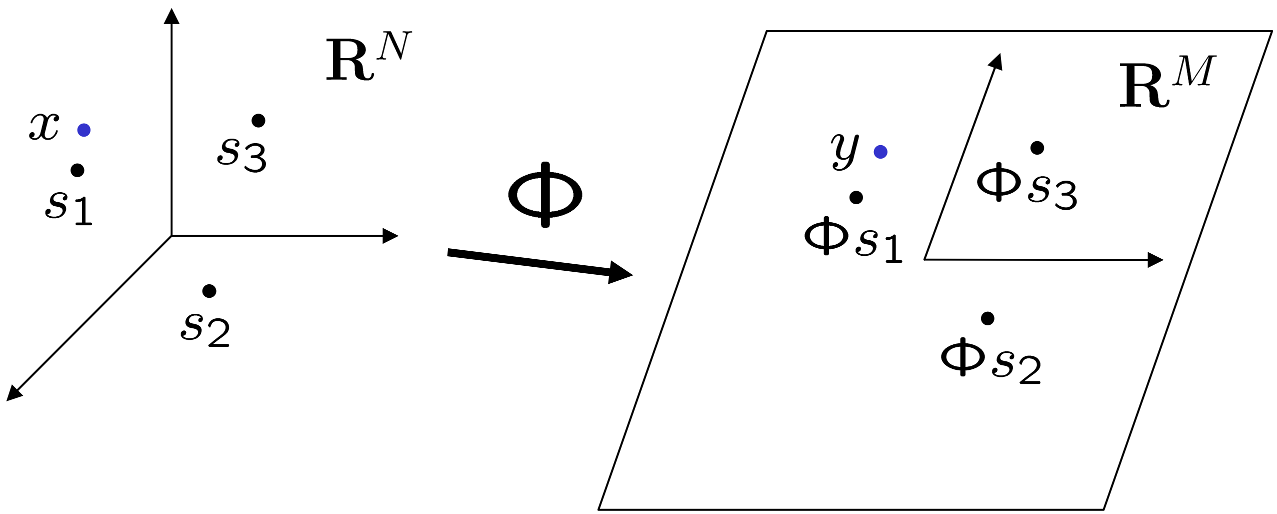
- AWGN: ***nearest-neighbor*** classification

$$\arg\min_{j=1,\ldots,P} \; \|x - s_j\|_2$$

- **Sufficient statistic**: $t_j = \|x - s_j\|_2$

Compressive LRT

- Compressive observations: $H_j \;:\; y = \Phi(s_j + n)$



$$\left.\begin{aligned} t_1 &= \|y - \Phi s_1\|_2 \\ t_2 &= \|y - \Phi s_2\|_2 \\ t_3 &= \|y - \Phi s_3\|_2 \end{aligned}\right\}$$

by the JL Lemma these distances are preserved

[Waagen et al., Davenport et al., Haupt et al.]

Matched Filter

- Signal x belongs to one of J classes
- Observed with some parameterized transformation
  - translation, rotation, scaling, lighting conditions, etc.
  - observation parameter unknown

$$\begin{aligned}
\mathcal{H}_1 &: \quad x = \mathcal{T}_{\theta_1} s_1 + n \\
\mathcal{H}_2 &: \quad x = \mathcal{T}_{\theta_2} s_2 + n \\
&\ \vdots \\
\mathcal{H}_J &: \quad x = \mathcal{T}_{\theta_J} s_J + n
\end{aligned}$$

- *Maximum likelihood* classifier with AWGN

$$\min_{j, \widehat{\theta}_j} \|x - \mathcal{T}_{\widehat{\theta}_j} s_j\|_2$$

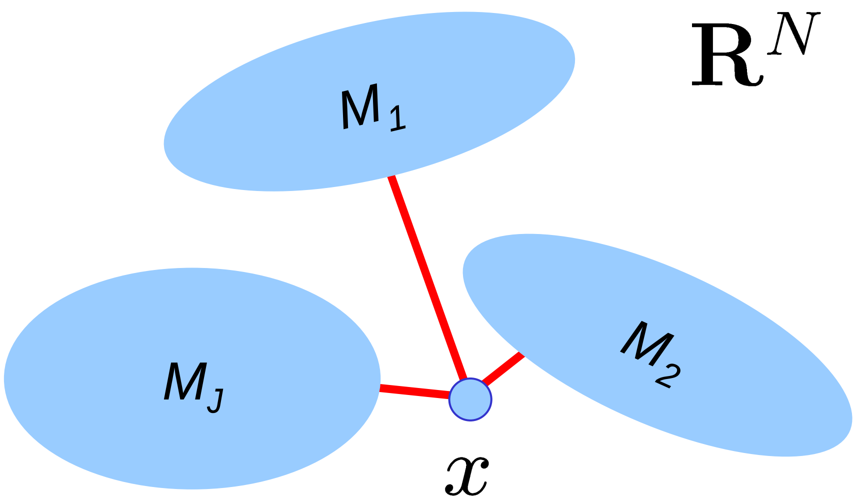
- Solve via convolution when parameter = translation

Matched Filter

- *Maximum likelihood* classifier with AWGN

$$\min_{j,\widehat{\theta}_j} \|x - \mathcal{T}_{\widehat{\theta}_j} s_j\|_2$$

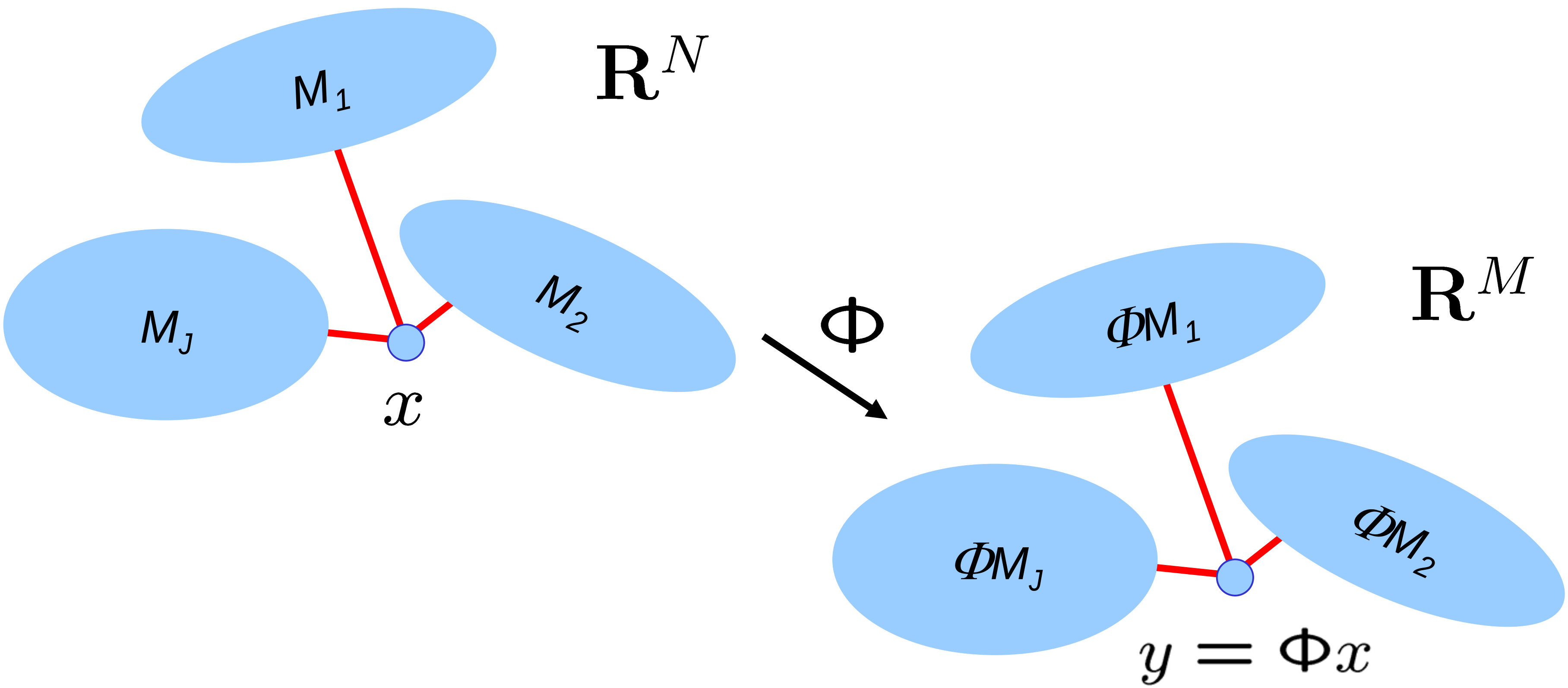
reduces to *nearest neighbor* classification when signal classes form manifolds



"Smashed Filter"

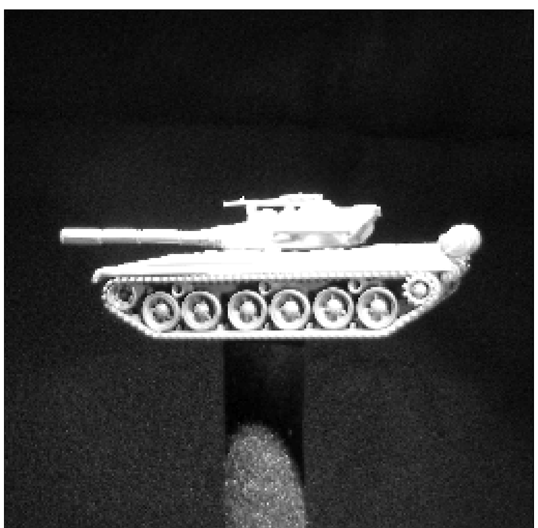
- Solve "nearest manifold" problem using random projections

$$\min_{j,\widehat{\theta}_j} \|\Phi x - \Phi \mathcal{T}_{\widehat{\theta}_j} s_j\|_2$$



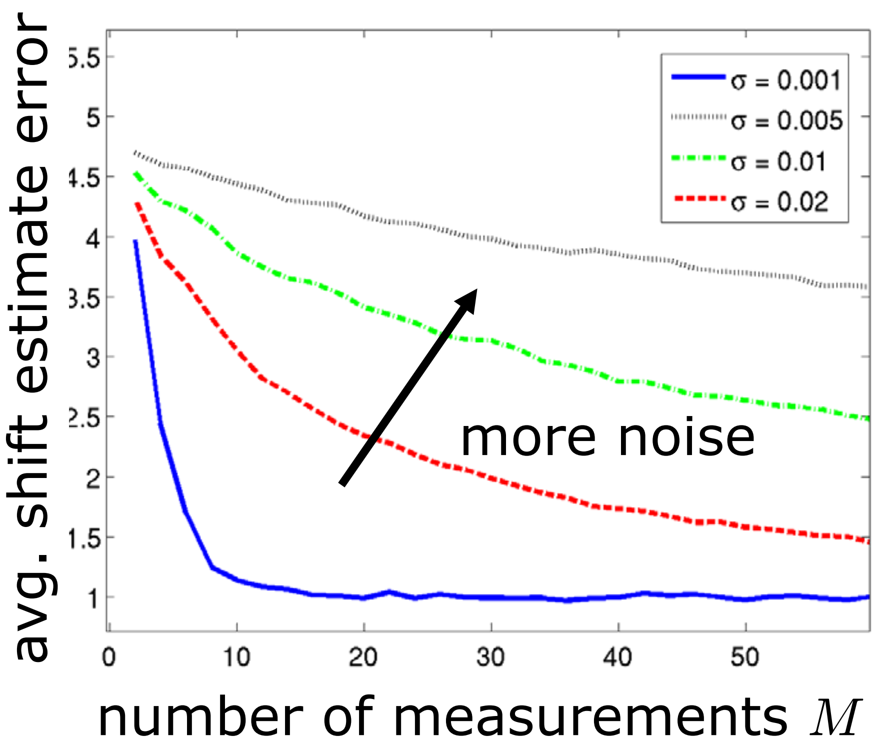
Smashed Filter – Experiments

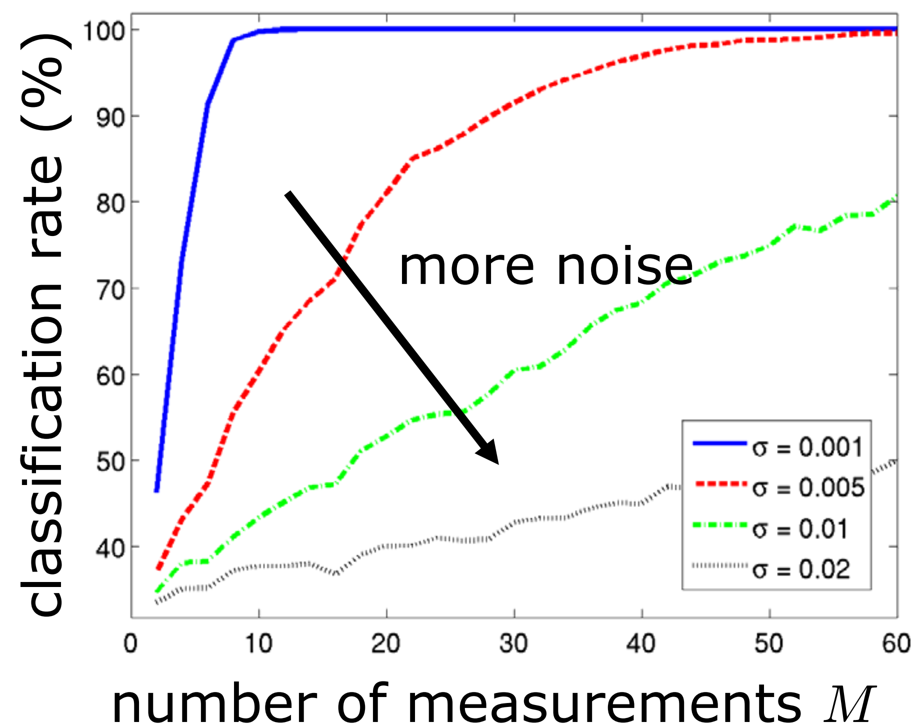
- 3 image classes: tank, school bus, SUV
- N = 65536 pixels
- Imaged using single-pixel CS camera with
  - unknown shift
  - unknown rotation



Smashed Filter – Unknown Position

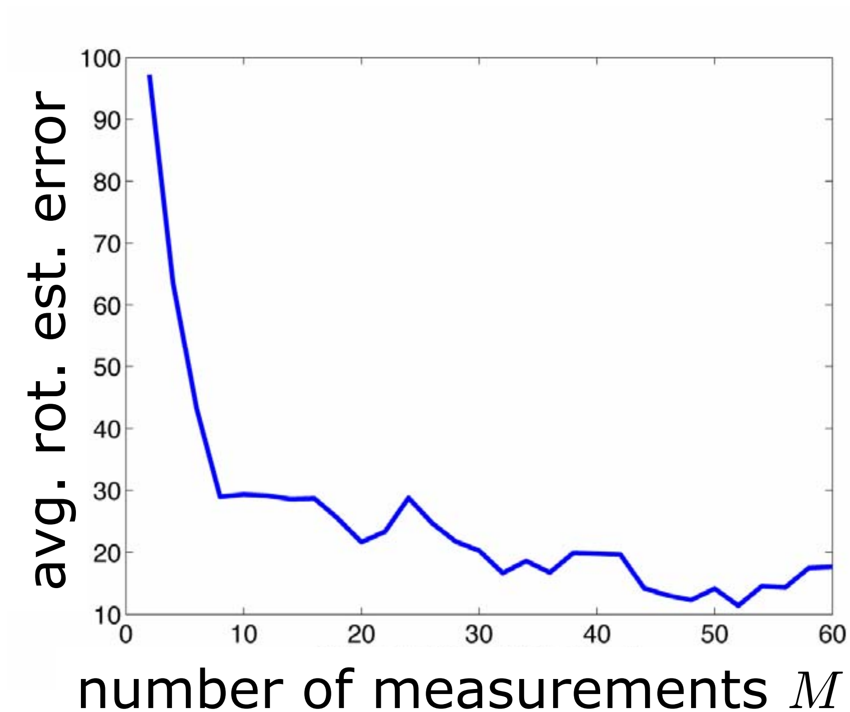
- Object shifted at random (K=2 manifold)
- Noise added to measurements
- Goal: identify most likely position for each image class
  identify most likely class using nearest-neighbor test





Smashed Filter – Unknown Rotation

- Object rotated each 2 degrees
- Goals: identify most likely rotation for each image class
  identify most likely class using nearest-neighbor test
- Perfect classification with as few as 6 measurements
- Good estimates of rotation with under 10 measurements



References – Compressive S.P. (1)

Statistical Signal Processing & Information Scalability:

- D. Waagen, N. Shah, M. Ordaz, and M. Cassabaum, “Random subspaces and SAR classification efficacy,” in Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery XII, May 2005.
- Marco Duarte, Mark Davenport, Michael Wakin, and Richard Baraniuk, Sparse signal detection from incoherent projections. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006)
- Mark Davenport, Michael Wakin, and Richard Baraniuk, Detection and estimation with compressive measurements. (Rice ECE Department Technical Report TREE 0610, November 2006)
- Jarvis Haupt, Rui Castro, Robert Nowak, Gerald Fudge, and Alex Yeh, Compressive sampling for signal classification. (Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, California, October 2006)

References – Compressive S.P. (2)

Manifolds, Manifold Learning, Smashed Filter:

- Richard Baraniuk and Michael Wakin, Random projections of smooth manifolds. (To appear in Foundations of Computational Mathematics)
- Mark Davenport, Marco Duarte, Michael Wakin, Jason Laska, Dharmpal Takhar, Kevin Kelly, and Richard Baraniuk, The smashed filter for compressive classification and target recognition. (Proc. of Computational Imaging V at SPIE Electronic Imaging, San Jose, California, January 2007)

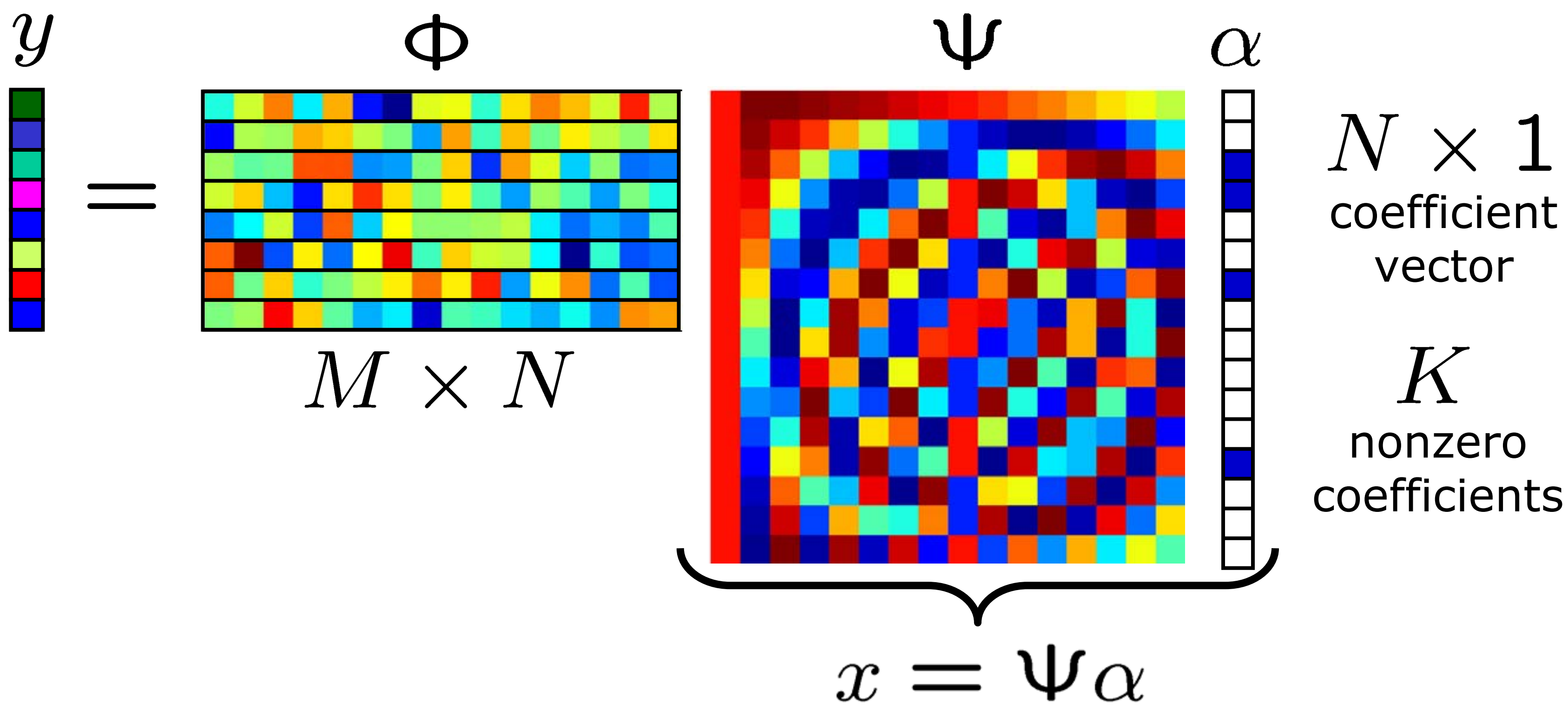
Theoretical Computer Science & Data Streaming Algorithms:

- N. Alon, P. Gibbons, Y. Matias, and M. Szegedy, "Tracking join and self-join sizes in limited storage," in Proc. Symp. Principles of Database Systems (PODS), Philadelphia, PA, 1999.
- Nitin Thaper, Sudipto Guha, Piotr Indyk, and Nick Koudas, Dynamic multidimensional histograms. (Proc. SIGMOD 2002, Madison, Wisconson, June 2002)
- Anna Gilbert, Sudipto Guha, Piotr Indyk, Yannis Kotidis, S. Muthukrishnan, and Martin J. Strauss, Fast small-space algorithms for approximate histogram maintenance. (Proc. 34th Symposium on Theory of Computing, Montréal, Canada, May 2002)
- S. Muthukrishnan, Data Streams: Algorithms and Applications, now, 2005.

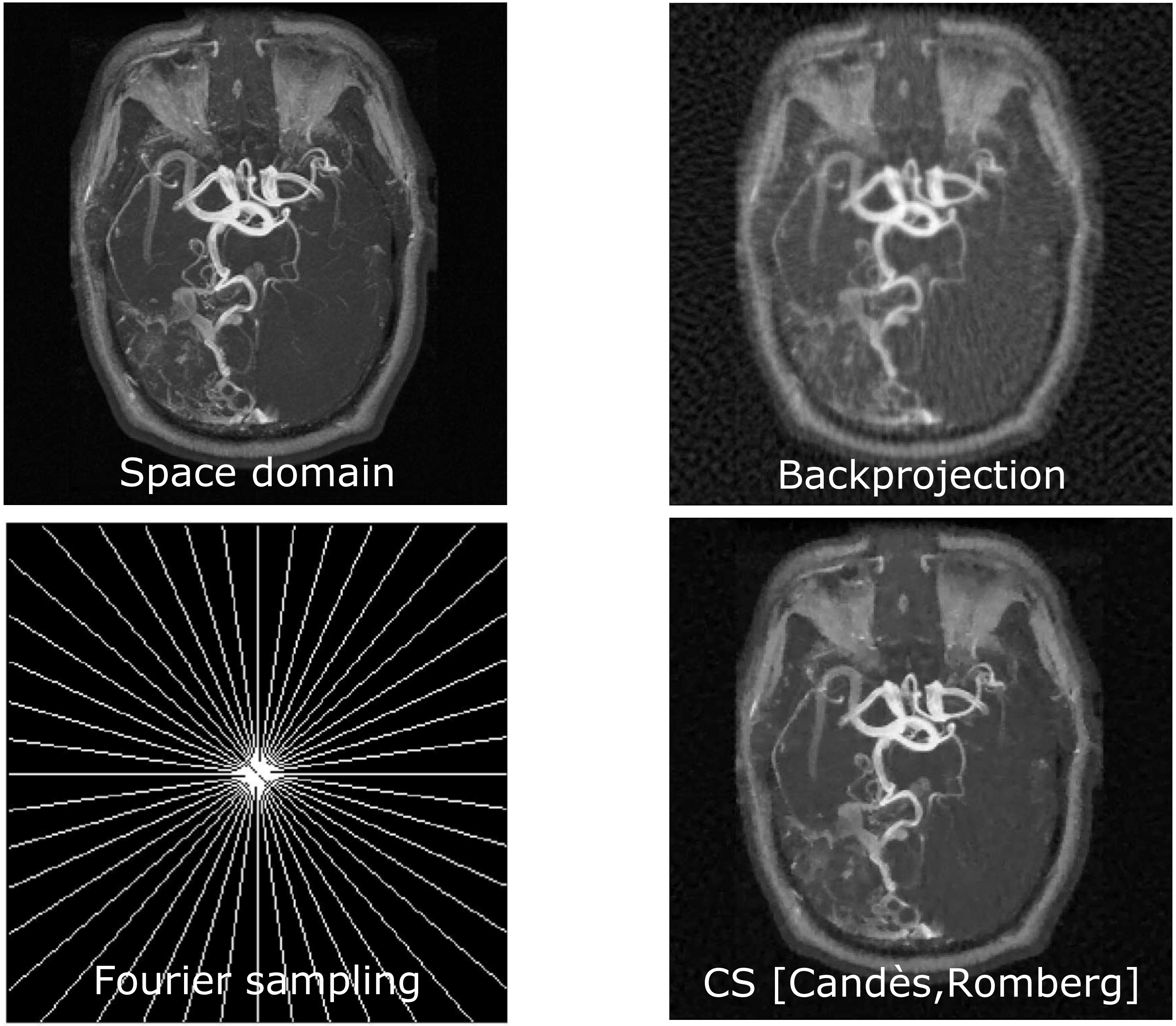
3. Inverse Problems

Idea: Recover signal from available measurements

- little or no control over sensing modality Φ



Magnetic Resonance Imaging



References – Inverse Problems

Medical Imaging:

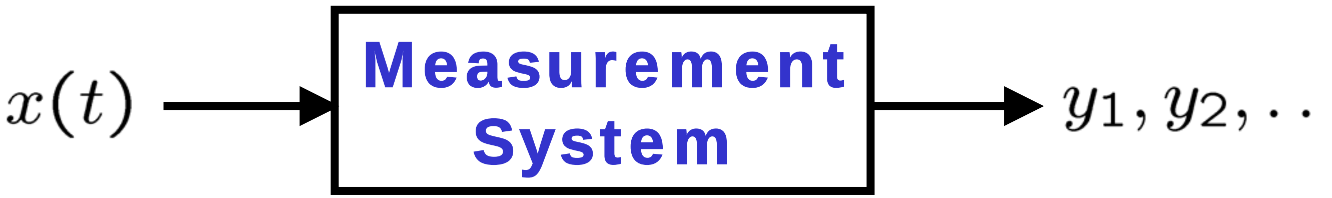
- Emmanuel Candès, Justin Romberg, and Terence Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. (IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, February 2006)
- Michael Lustig, David Donoho, and John M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging. (Preprint, 2007)
- Jong Chul Ye, Compressed sensing shape estimation of star-shaped objects in Fourier imaging (Preprint, 2007)

Other:

- Ingrid Daubechies, Massimo Fornasier, and Ignace Loris, Accelerated projected gradient method for linear inverse problems with sparsity constraints. (Preprint, 2007)
- Mário A. T. Figueiredo, Robert D. Nowak, and Stephen J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. (Preprint, 2007)
- José Bioucas-Dias and Mário Figueiredo, A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. (Preprint, 2007)
- Lawrence Carin, Dehong Liu, and Ya Xue, *In Situ* Compressive Sensing. (Preprint, 2007)

4. Data Acquisition

Idea: "Compressive sampling" of analog signals

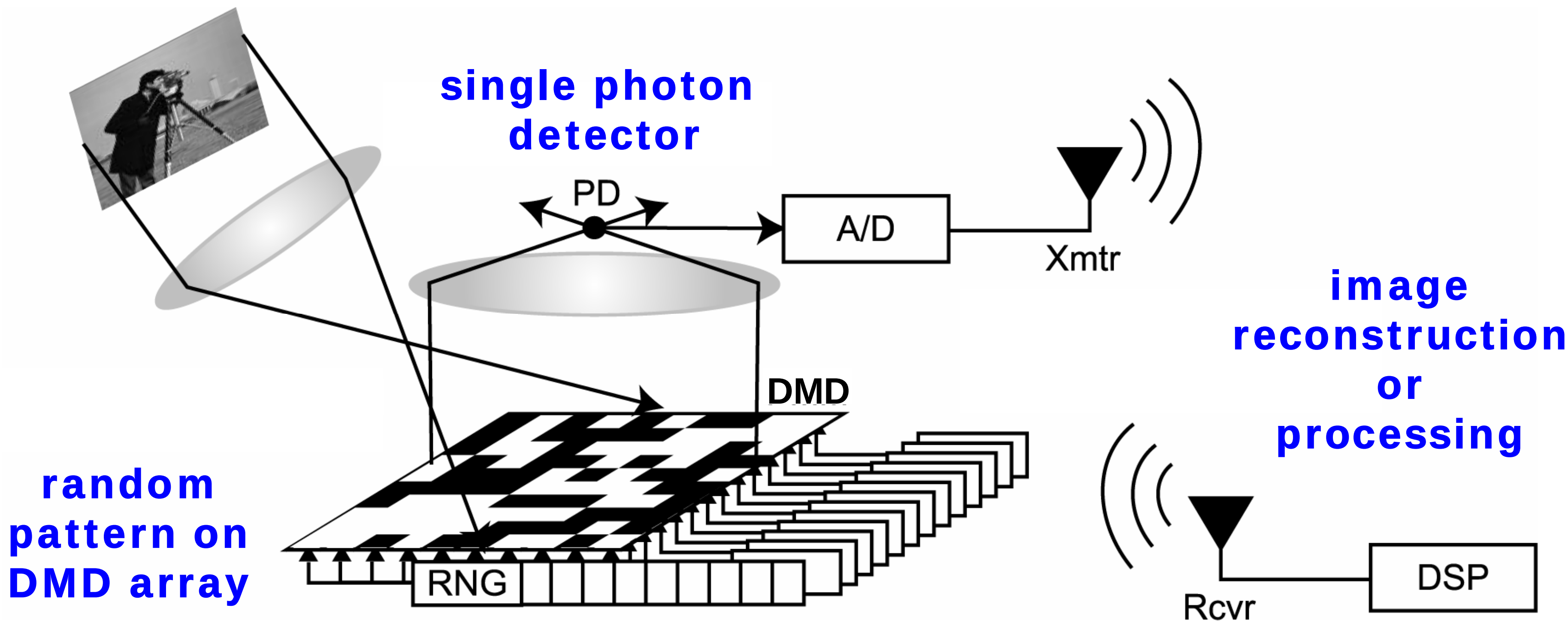


$$x(t) \approx \sum_{k=1}^{K} \alpha_k \psi_k(t)$$

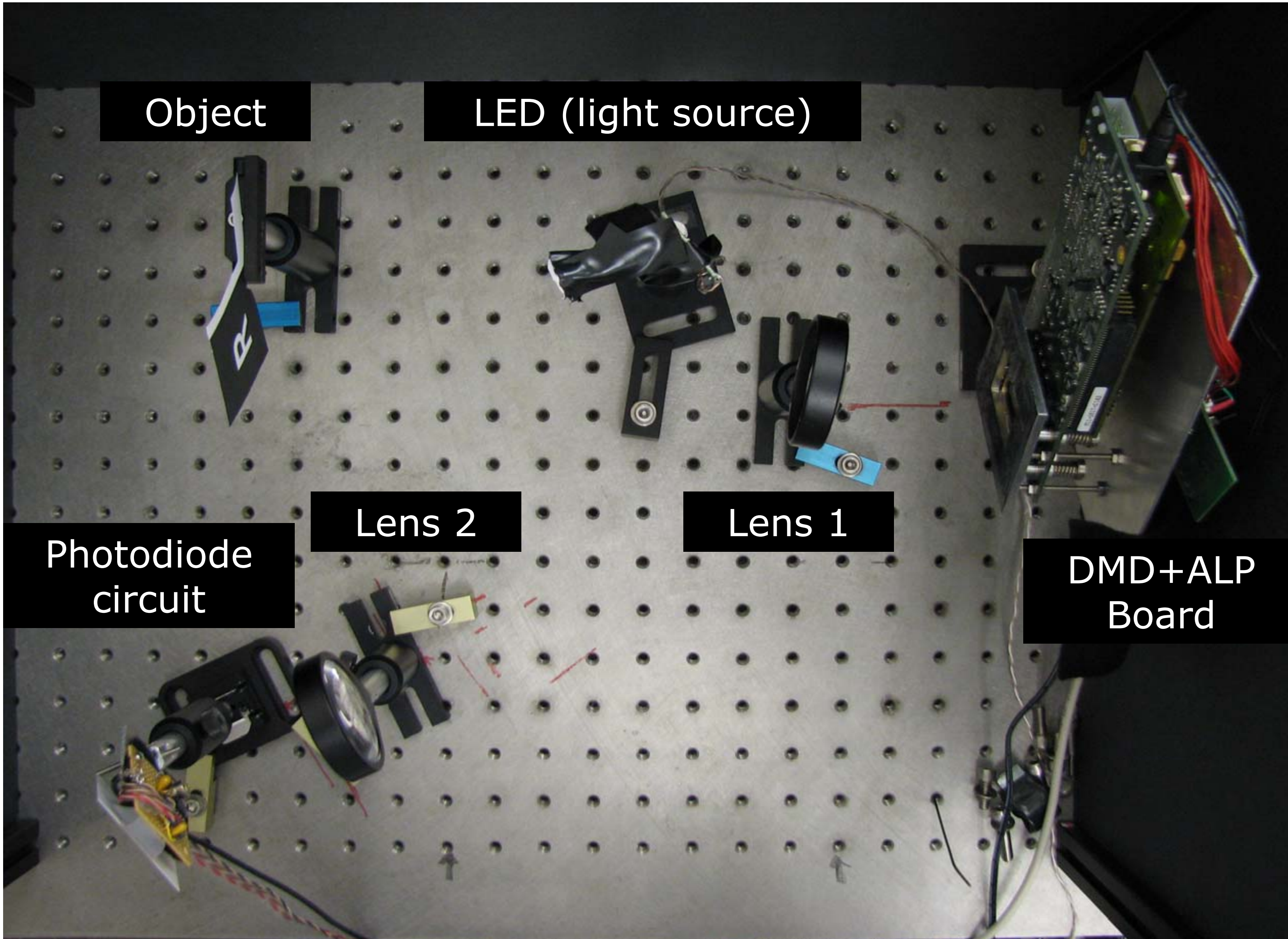
$$y_m \approx \langle x, \phi_m \rangle = \int_{-\infty}^{\infty} x(t) \phi_m(t) dt$$

4a. Single-Pixel CS Camera

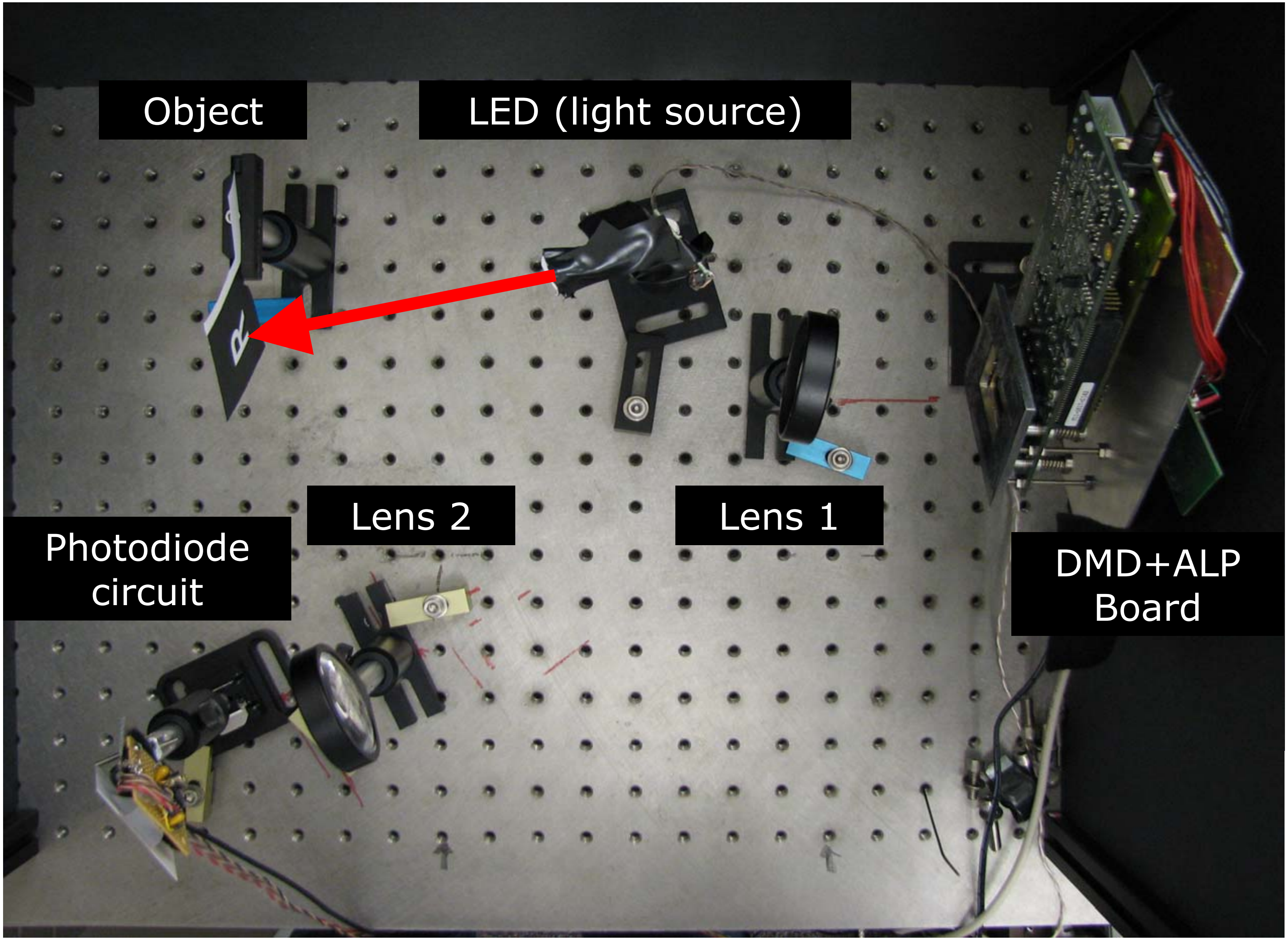
[Baraniuk and Kelly, et al.]



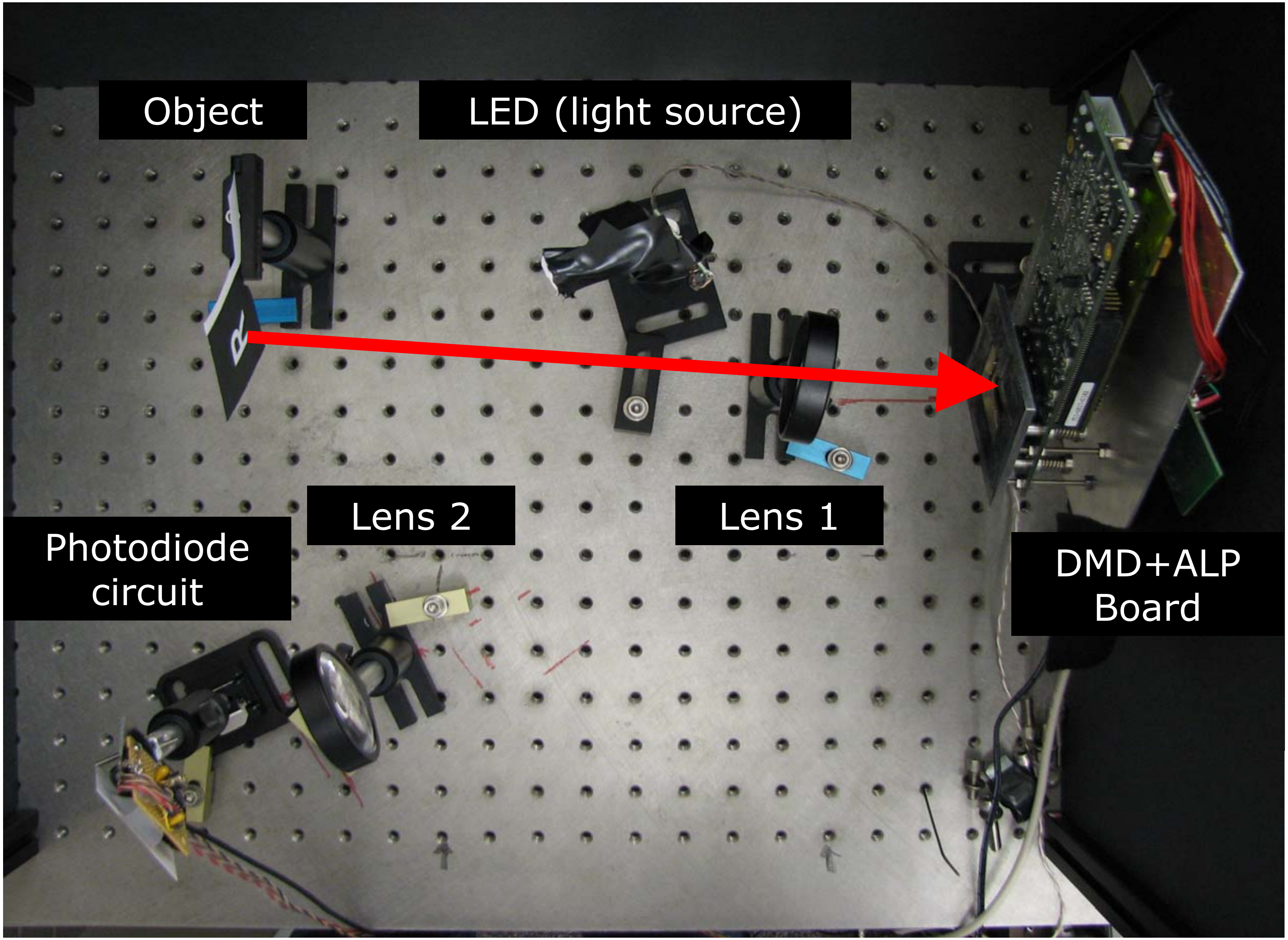
Single Pixel Camera



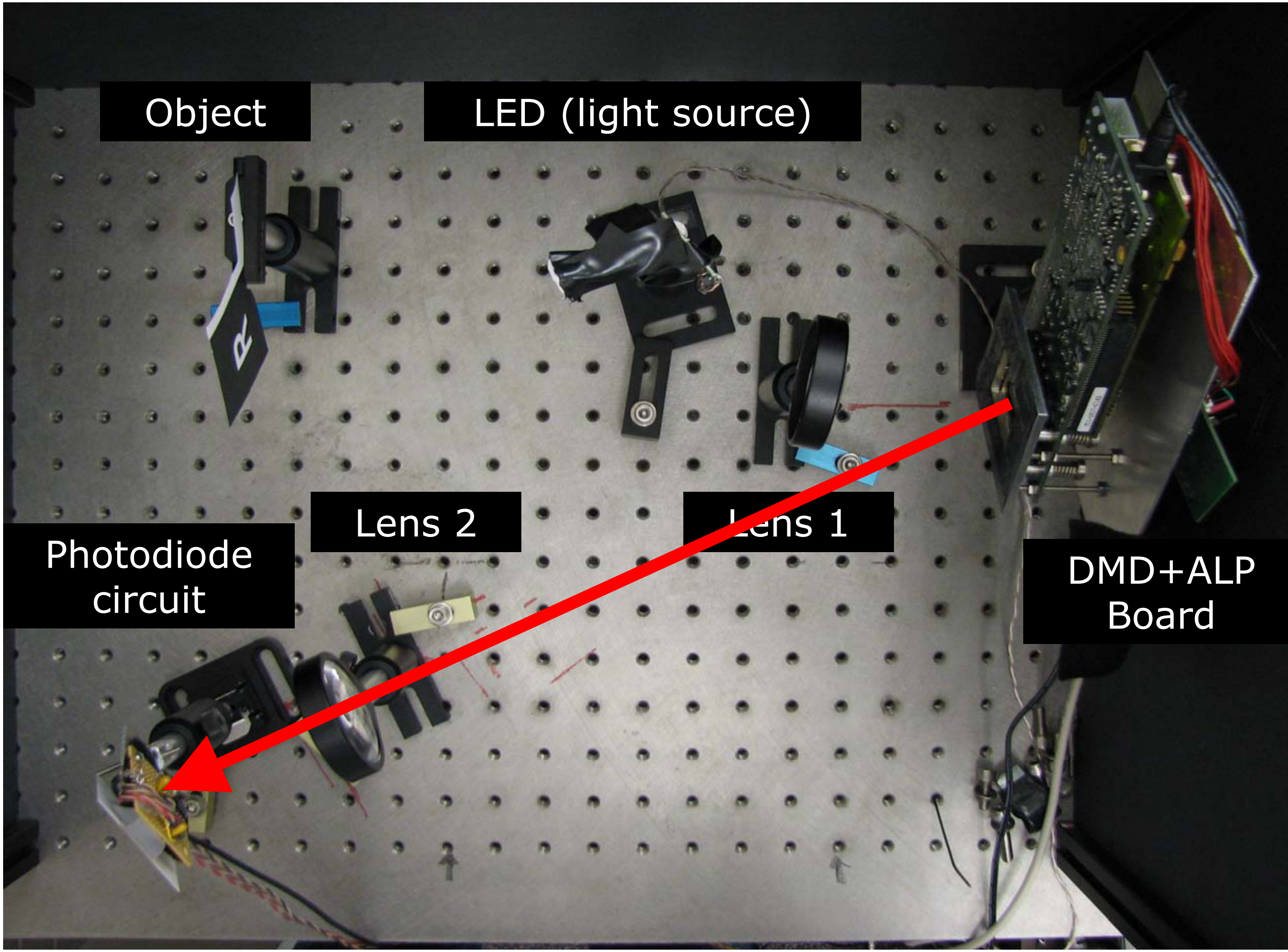
Single Pixel Camera



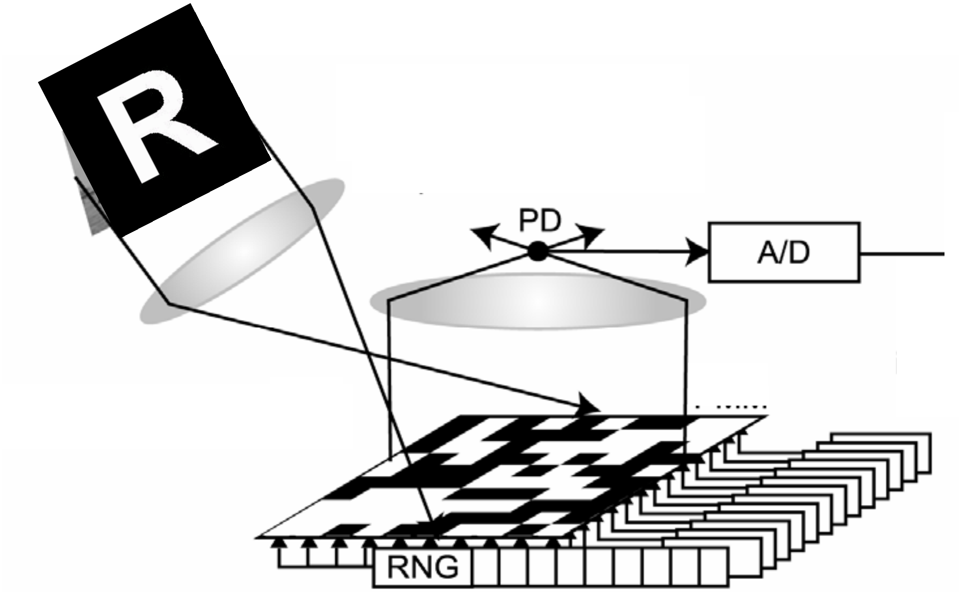
Single Pixel Camera



Single Pixel Camera

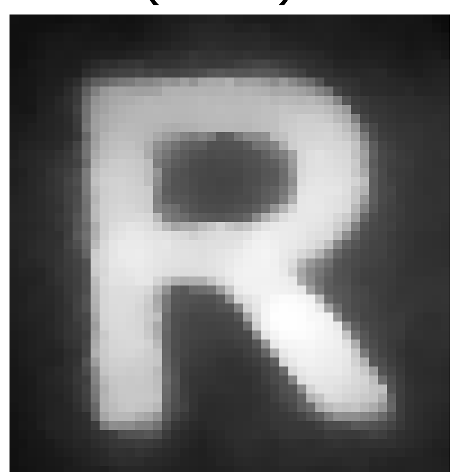


First Image Acquisition



target
65536 pixels

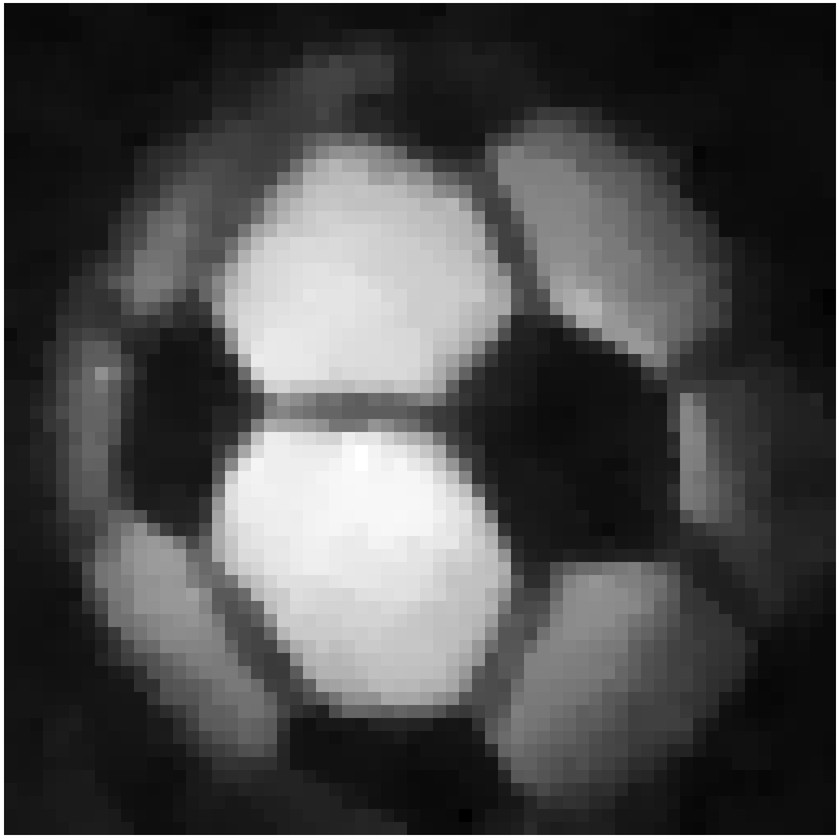
11000 measurements
(16%)



1300 measurements
(2%)

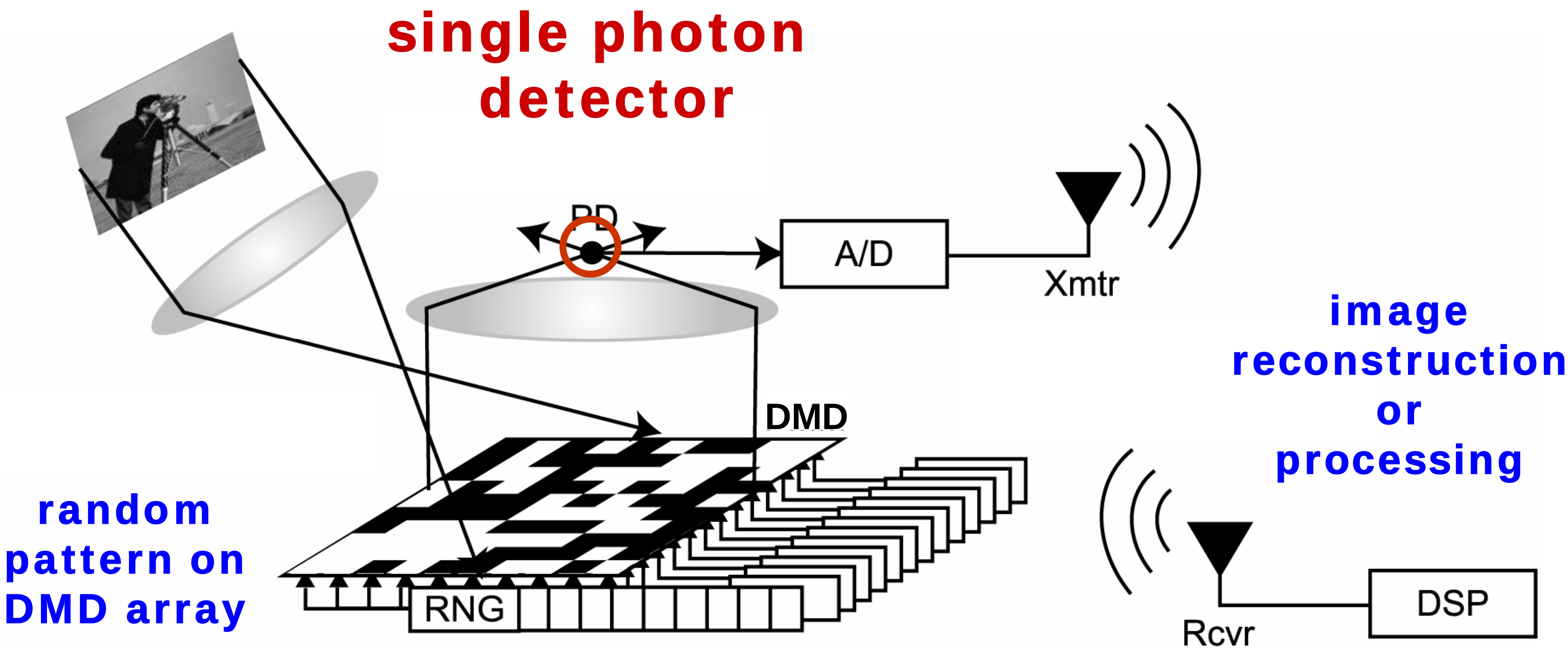
Second Image Acquisition

4096
pixels



500
random measurements

Single-Pixel Camera



Photodetector Options

- Simple photodiode
  - augment with color filters
- Dual photodiode sandwich
  - dual visible and infrared imaging
- Photomultiplier tube for low-light CS

true color low-light imaging

256 x 256 image with 10:1 compression

[Nature Photonics, April 2007]

4b. Analog-to-Information Conversion

[with E. Candès and J. Romberg]

DARPA A/I Project:

Efficient sampling of high-bandwidth signals

- sparse models allow sampling far below Nyquist rate
- new architectures for incoherent measurements

$$x(t) \approx \sum_{k=1}^{K} \alpha_k \psi_k(t)$$

$$y_m \approx \langle x, \phi_m \rangle = \int_{-\infty}^{\infty} x(t) \phi_m(t) dt$$

Two Sampling Architectures

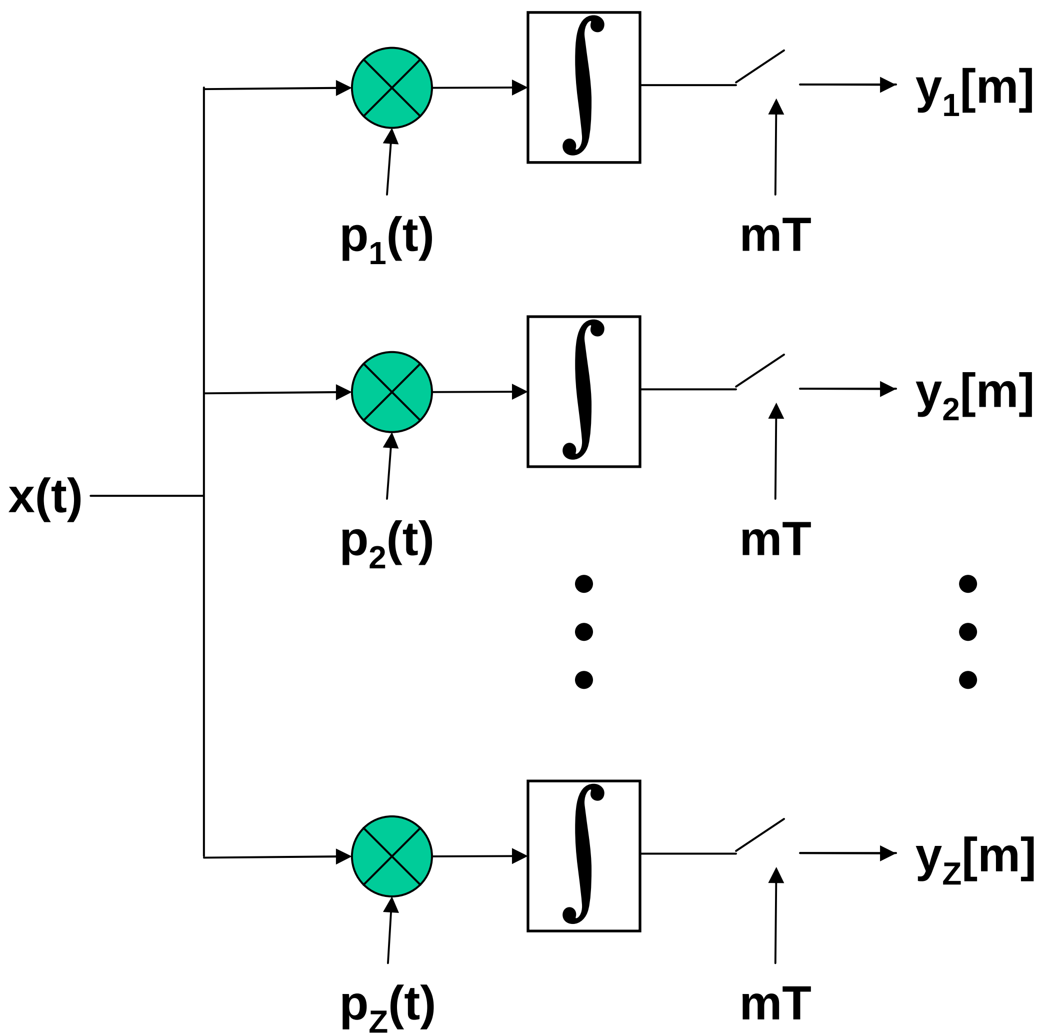
1. Nonuniform sampler (NUS)
   – incoherent measurements for signals with sparse spectra

$$x(t) \longrightarrow \boxed{\textbf{NUS}} \longrightarrow y_1, y_2, \cdots$$

$$y_m = x(t_m) = \langle x, \delta_{t_m} \rangle$$

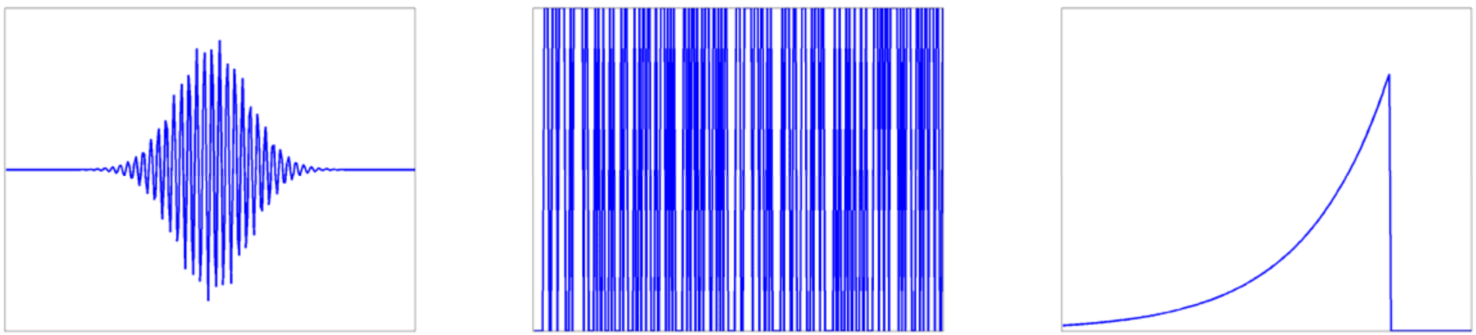
Two Sampling Architectures

2. Random pre-integrator (RPI)
   - more universal incoherent measurement system



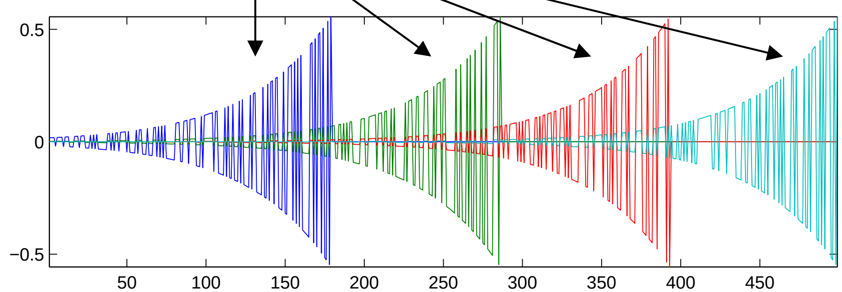
RPI Measurement Functions

$$y[m] = ((x \times p) * h)(t)\big|_{t=mT}$$



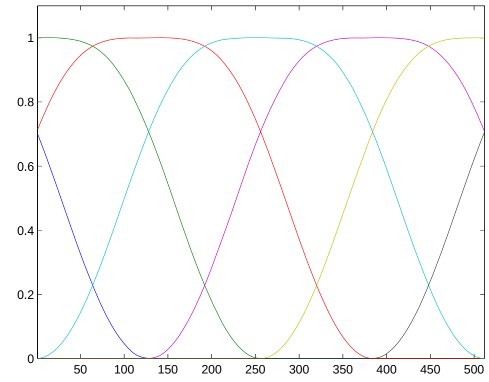
$$h(t) = e^{-at}, \quad t \geq 0$$

$$= \langle x, \phi_m \rangle$$



Multiscale Gabor Dictionary
for Time-Frequency Sparse Signals

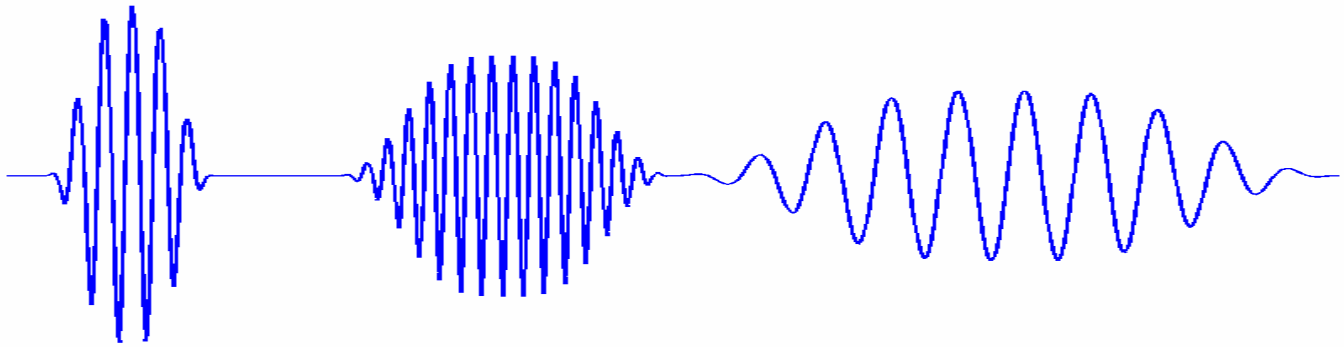
- Windowed sinusoids at a collection of
  - scales/durations
  - frequencies
  - positions



smooth windows

- Overcomplete, efficiently computable
  - size/complexity: c*N*log(N)

- Sparse representation of arbitrary pulse signals



[Thanks to P. Stobbe]

L_1 Synthesis vs. Analysis

- Consider general sparse dictionary Ψ
  - measure $y = \Phi x$
- Canonical approach: *synthesis-based*
  - find sparse coefficients α that would synthesize signal x

$$\min \|\alpha\|_1 \text{ such that } y = \Phi\Psi\alpha$$

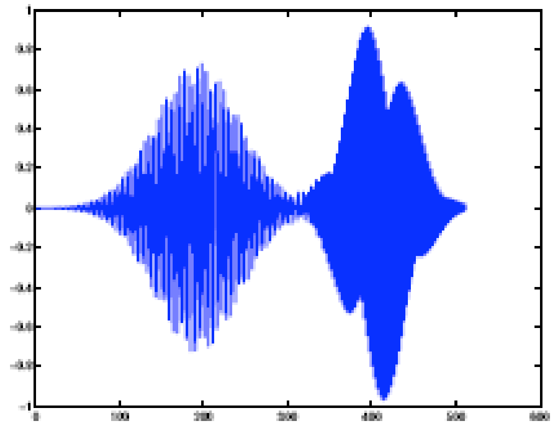
- Alternative approach: *analysis-based*
  - find signal x that has sparse analysis in Ψ

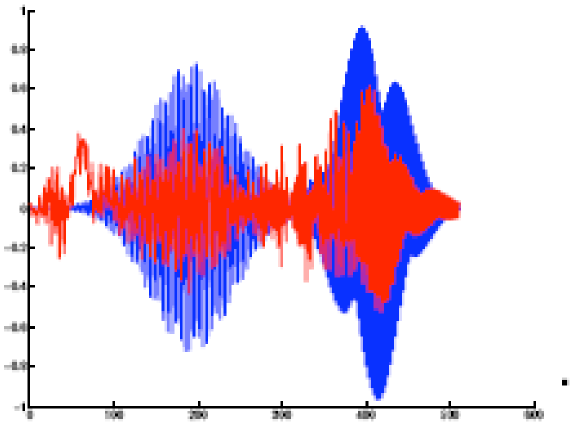
$$\min \|\Psi^T x\|_1 \text{ such that } y = \Phi x$$

- Solutions differ when Ψ is overcomplete [Elad et al.; Starck et al.]
  - chicken-or-egg: x or Ψ?
  - potentially faster to find x than α

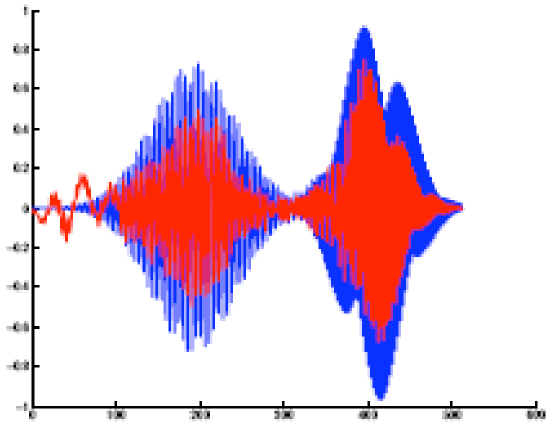
Example Reconstruction

- Two-pulse signal, length N=512
- Sensing matrix random ± 1
- M=30 random measurements
- Gabor dictionary 43x overcomplete

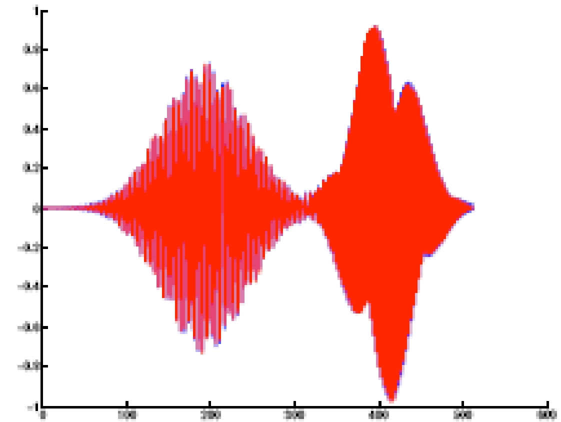




ℓ_1 synthesis
rel. err. = 67%



ℓ_1 analysis
46%



reweighted ℓ_1 analysis
2.2% (4 iter)

References – Data Acquisition (1)

CS Camera:

- Dharmpal Takhar, Jason Laska, Michael Wakin, Marco Duarte, Dror Baron, Shriram Sarvotham, Kevin Kelly, and Richard Baraniuk, A new compressive imaging camera architecture using optical-domain compression. (Proc. of Computational Imaging IV at SPIE Electronic Imaging, San Jose, California, January 2006)
- Duncan Graham-Rowe, Digital cameras: Pixel power, Nature Photonics 1, 211 - 212 (2007).
- CS Camera Website: http://www.dsp.ece.rice.edu/cs/cscamera/

Analog-to-Information Conversion:

- Joel Tropp, Michael Wakin, Marco Duarte, Dror Baron, and Richard Baraniuk, Random filters for compressive sampling and reconstruction. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Toulouse, France, May 2006)
- Jason Laska, Sami Kirolos, Marco Duarte, Tamer Ragheb, Richard Baraniuk, and Yehia Massoud, Theory and implementation of an analog-to-information converter using random demodulation. (Proc. IEEE Int. Symp. on Circuits and Systems (ISCAS), New Orleans, Louisiana, 2007)

References – Data Acquisition (2)

Analog-to-Information Conversion [cont.]:

- Tamer Ragheb, Sami Kirolos, Jason Laska, Anna Gilbert, Martin Strauss, Richard Baraniuk, and Yehia Massoud, Implementation models for analog-to-information conversion via random sampling. (To appear in Proc. Midwest Symposium on Circuits and Systems (MWSCAS), 2007)

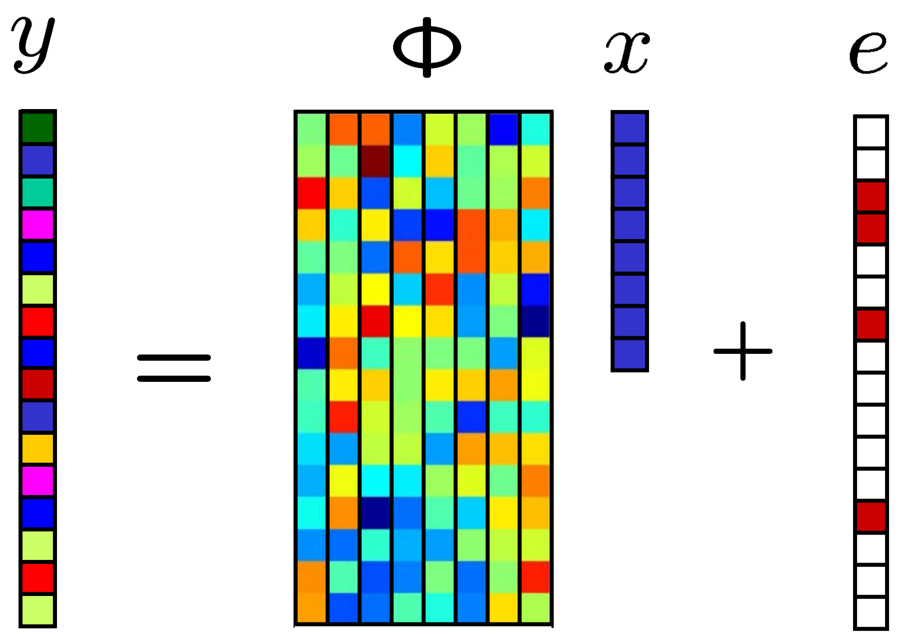
Analysis versus Synthesis in L_1 minimization:

- J.-L. Starck, M. Elad, and D. L. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," Adv. Imaging and Electron Phys., vol. 132, 2004.
- M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," Inverse Problems, vol. 23, pp. 947–968, 2007.

5. Error Correction

Idea: Channel coding using CS principles

- unconstrained minimization problem
- robust to some large and many small errors



$$\hat{x} = \arg\min \|y - \Phi x'\|_1$$

References – Error Correction

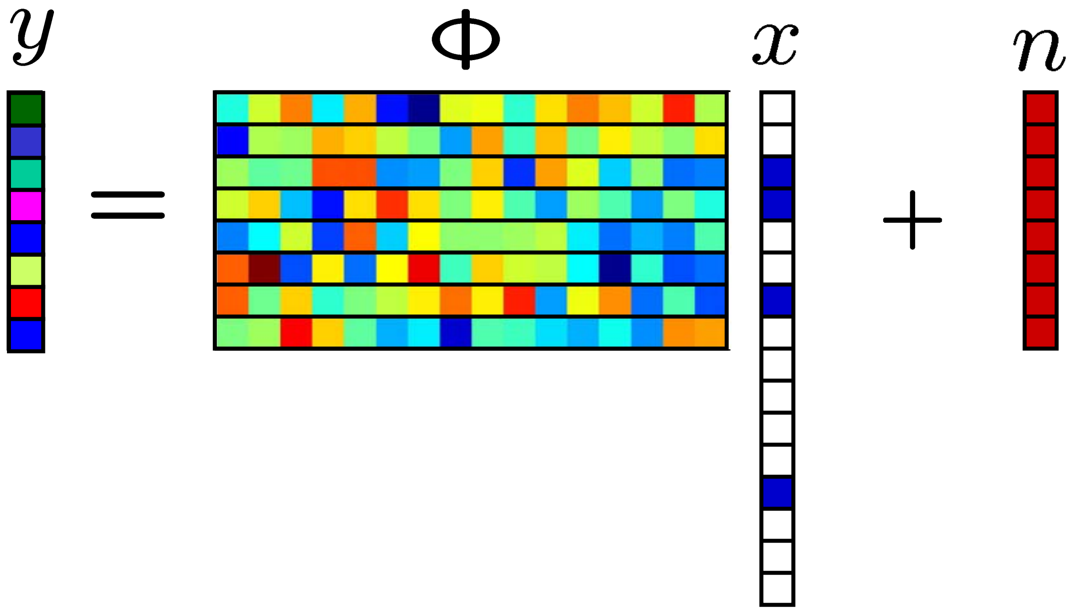
Error Correction

- Emmanuel Candès and Terence Tao, Decoding by linear programming. (IEEE Trans. on Information Theory, 51(12), pp. 4203 - 4215, December 2005)
- Mark Rudelson and Roman Vershynin, Geometric approach to error correcting codes and reconstruction of signals. (International Mathematical Research Notices, 64, pp. 4019 - 4041, 2005)
- Emmanuel Candès and Paige Randall, Highly robust error correction by convex programming. (Preprint, 2006)
- Rick Chartrand, Nonconvex compressed sensing and error correction. (Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Honolulu, Hawaii, April 2007)
- Cynthia Dwork, Frank McSherry, and Kunal Talwar, The price of privacy and the limits of LP decoding. (Proc. Symposium on Theory of Computing (STOC), San Diego, California, June, 2007)

6. Statistical Estimation

Idea: Model selection when #variables $\gg$ #observations

- sparse model provides simple explanation



$$\min \|x'\|_1 \text{ subject to } \|\Phi^*(y - \Phi x')\|_\infty \leq \epsilon$$

References – Statistical Estimation

Dantzig Selector:

- Emmanuel Candès and Terence Tao, The Dantzig Selector: Statistical estimation when p is much larger than n (To appear in Annals of Statistics)

Phase Transition:

- David Donoho and Victoria Stodden, Breakdown Point of Model Selection When the Number of Variables Exceeds the Number of Observations, International Joint Conference on Neural Networks, 2006.

7. Additional References

Related areas:

- Martin Vetterli, Pina Marziliano, and Thierry Blu, Sampling signals with finite rate of innovation. (IEEE Trans. on Signal Processing, 50(6), pp. 1417-1428, June 2002)
- Anna Gilbert, Sudipto Guha, Piotr Indyk, S. Muthukrishnan, and Martin Strauss, Near-optimal sparse Fourier representations via sampling. (Proc. ACM Symposium on Theory of Computing (STOC), 2002)

Other CS applications:

- David Donoho and Yaakov Tsaig, Extensions of compressed sensing. (Signal Processing, 86(3), pp. 533-548, March 2006)
- Mona Sheikh, Olgica Milenkovic, and Richard Baraniuk, Compressed sensing DNA microarrays. (Rice ECE Department Technical Report TREE 0706, May 2007)

More at: **dsp.rice.edu/cs**