Sparsity, Randomness and Compressed Sensing

Petros Boufounos
Mitsubishi Electric Research Labs
petrosb@merl.com

Sparsity

Why Sparsity

- Natural data and signals exhibit **structure**
- **Sparsity** often captures that **structure**
- Very **general** signal model
- Computationally **tractable**
- Wide range of applications in signal **acquisition**, **processing**, and **transmission**

Signal Representations

Signal example: Images

- 2-D function f

- Idealized view

$f \in$ some function space defined over $[0, 1] \times [0, 1]$

Signal example: Images

- 2-D function f

- Idealized view

$f \in$ some function
space defined
over $[0,1] \times [0,1]$

- In practice

$f \in \mathbb{R}^{N \times N}$

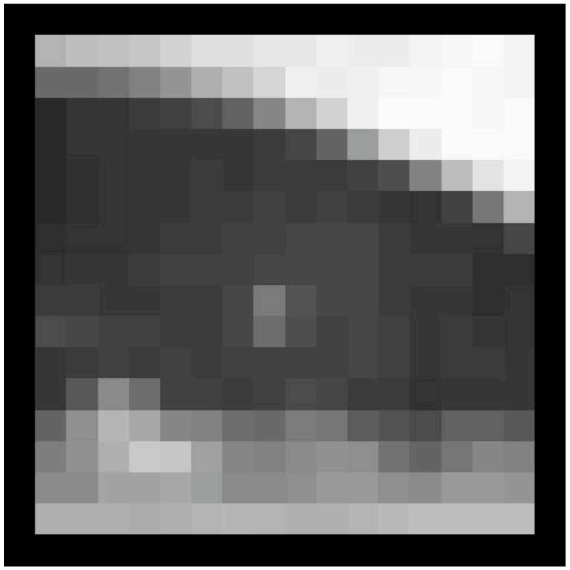
ie: an $N \times N$ matrix

Signal example: Images

- 2-D function f

- Idealized view

$f \in$ some function space defined over $[0,1] \times [0,1]$

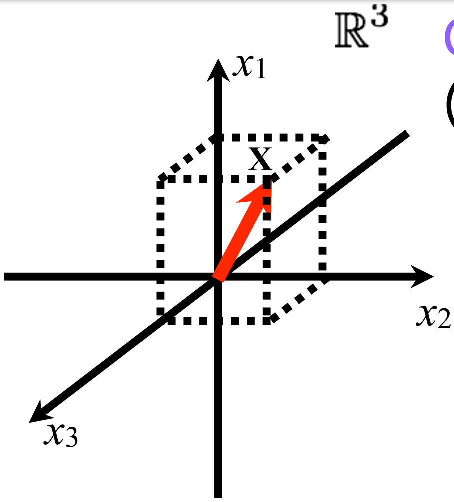


- In practice

$$f \in \mathbb{R}^{N \times N}$$

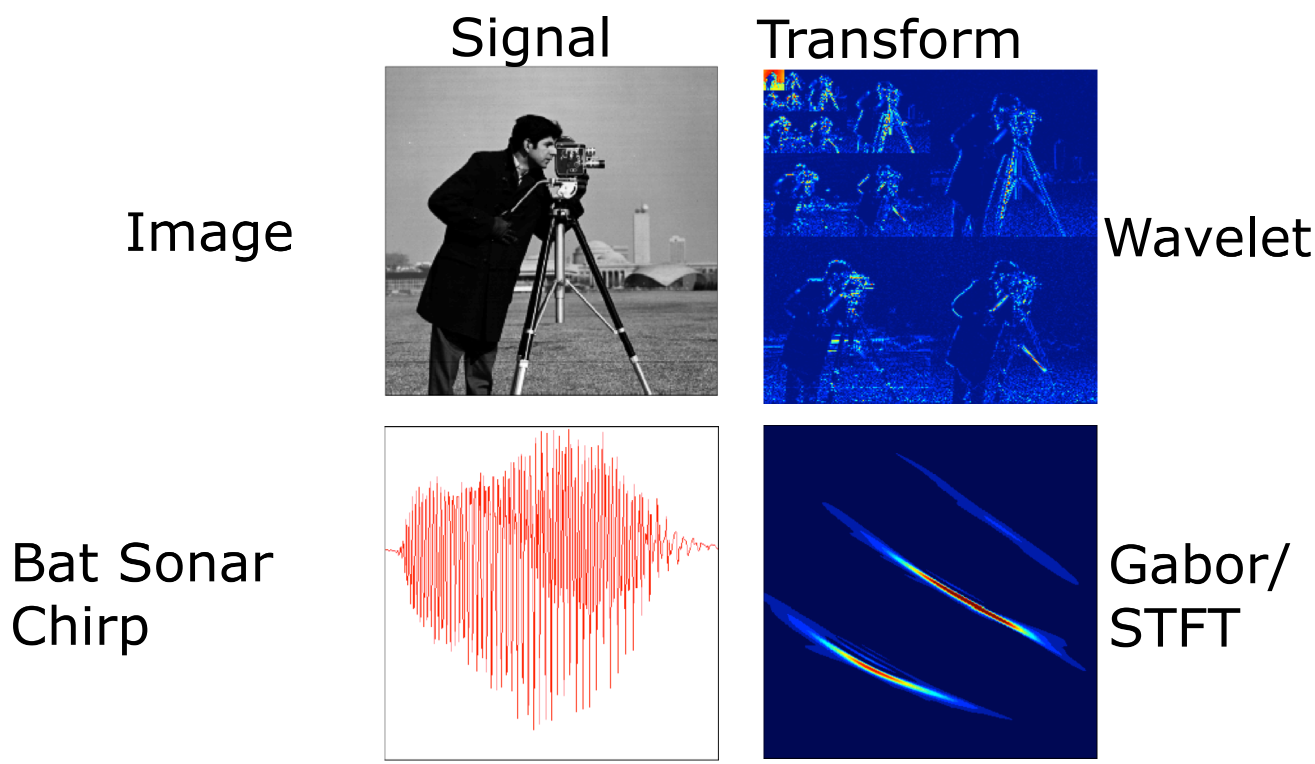
ie: an $N \times N$ matrix (pixel average)

Signal Models



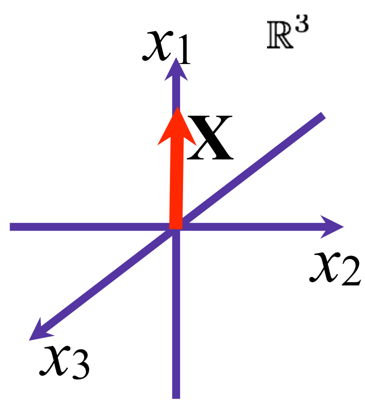
Classical Model: Signal lies in a linear vector space (e.g. bandlimited functions)

Sparse Model: Signals of interest are often sparse or compressible

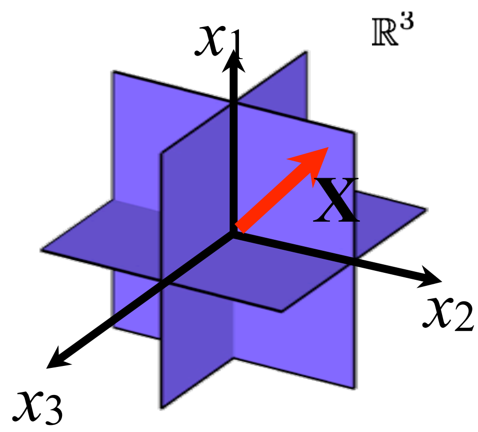


i.e., very few large coefficients, many close to zero.

Sparse Signal Models

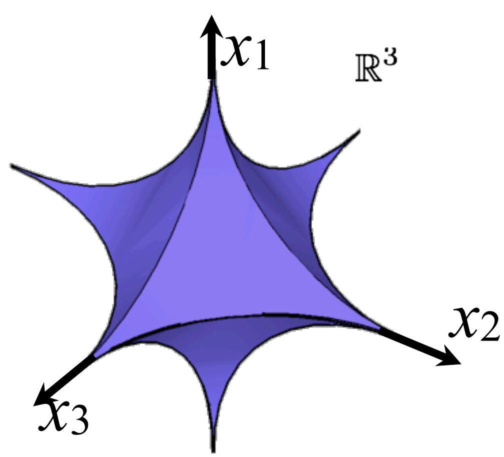


1-sparse



2-sparse

Sparse signals have few non-zero coefficients



Compressible (ℓ_p ball, p<1)

Compressible signals have few significant coefficients.
The coefficients decay as a power law.

Sparse Approximation

Computational Harmonic Analysis

- Representation $\qquad f = \sum_k \underset{\uparrow}{a_k} \underset{\uparrow}{\mathbf{b}_k}$

  coefficients $\qquad$ basis, frame

- **Analysis:** study f through *structure* of $\{a_k\}$

  $\{\mathbf{b}_k\}$ should *extract features* of interest

- **Approximation:** $\widehat{f}_N$ uses just a few terms N

  exploit *sparsity* of $\{a_k\}$

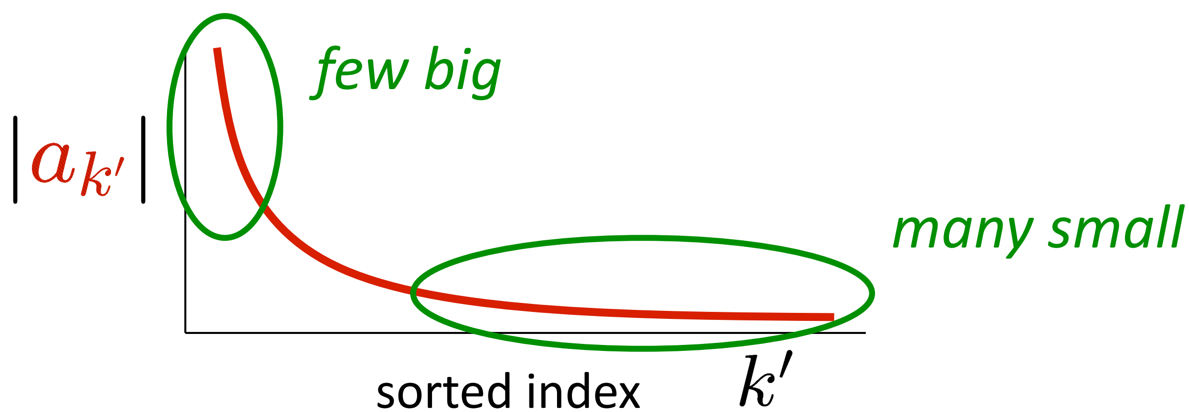
Wavelet Transform Sparsity

$$f = \sum_k a_k \mathbf{b}_k$$

- Many $a_k \approx 0$ (blue)

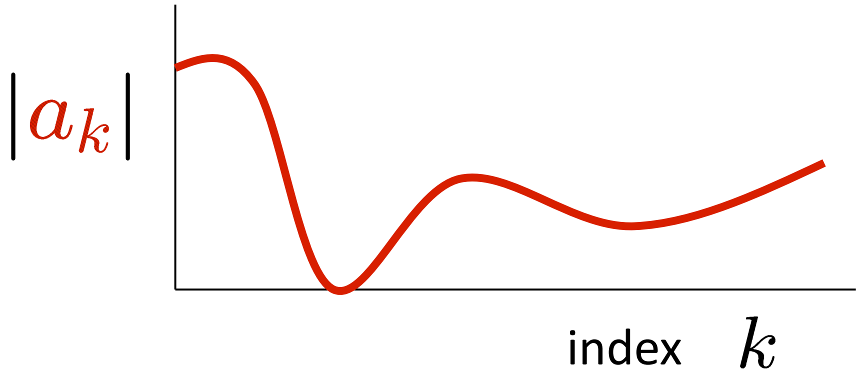
Sparseness $\Rightarrow$ Approximation

$$f = \sum_k a_k \, \mathbf{b}_k$$



Linear Approximation

$$f = \sum_k a_k \mathbf{b}_k$$

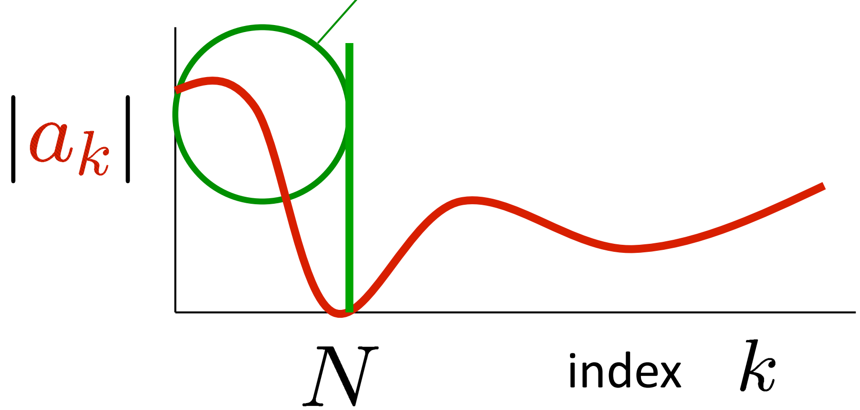


Linear Approximation

$$f \;=\; \sum_k a_k \, \mathbf{b}_k$$

- N-*term approximation*: use "*first*" a_k

$$\widetilde{f}_N \;:=\; \sum_{k=1}^{N} a_k \, \mathbf{b}_k$$



Nonlinear Approximation

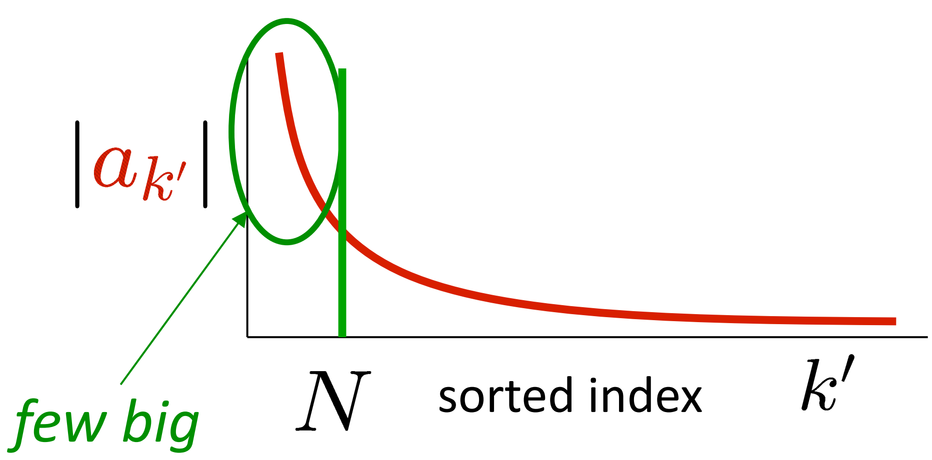
$$f \;=\; \sum_k a_k \, \mathbf{b}_k$$

- N-*term approximation*:

use *largest* a_k *independently*

$$\widehat{f}_N \;:=\; \sum_{k'=1}^{N} a_{k'} \, \mathbf{b}_{k'}$$

- Greedy / *thresholding*



Error Approximation Rates

$$f = \sum_k a_k \, \mathbf{b}_k$$

$$\widehat{f}_N = \sum_{k'=1}^{N} a_{k'} \, \mathbf{b}_{k'}$$

$$\|f - \widehat{f}_N\|_2^2 < C\, N^{-\alpha} \qquad \text{as} \quad N \longrightarrow \infty$$

- Optimize asymptotic *error decay rate* α
- Nonlinear approximation works better than linear

Compression is Approximation

- Lossy compression of an image creates an approximation

$$f \;=\; \sum_k a_k \, \mathbf{b}_k$$

coefficients basis, frame

quantize *to* R *total bits*

$$\widehat{f}_R \;=\; \sum_k a_k^q \, \mathbf{b}_k$$

Sparse approximation ≠ Compression

- Sparse approximation chooses coefficients but does *not quantize* or worry about their *locations*

$$f = \sum_k a_k \mathbf{b}_k$$

$$\Big\downarrow \textit{threshold}$$

$$\widehat{f}_N = \sum_{k'=1}^{N} a_{k'} \mathbf{b}_{k'}$$

Location, Location, Location

- Nonlinear approximation selects N largest a_k to minimize error (easy – threshold)

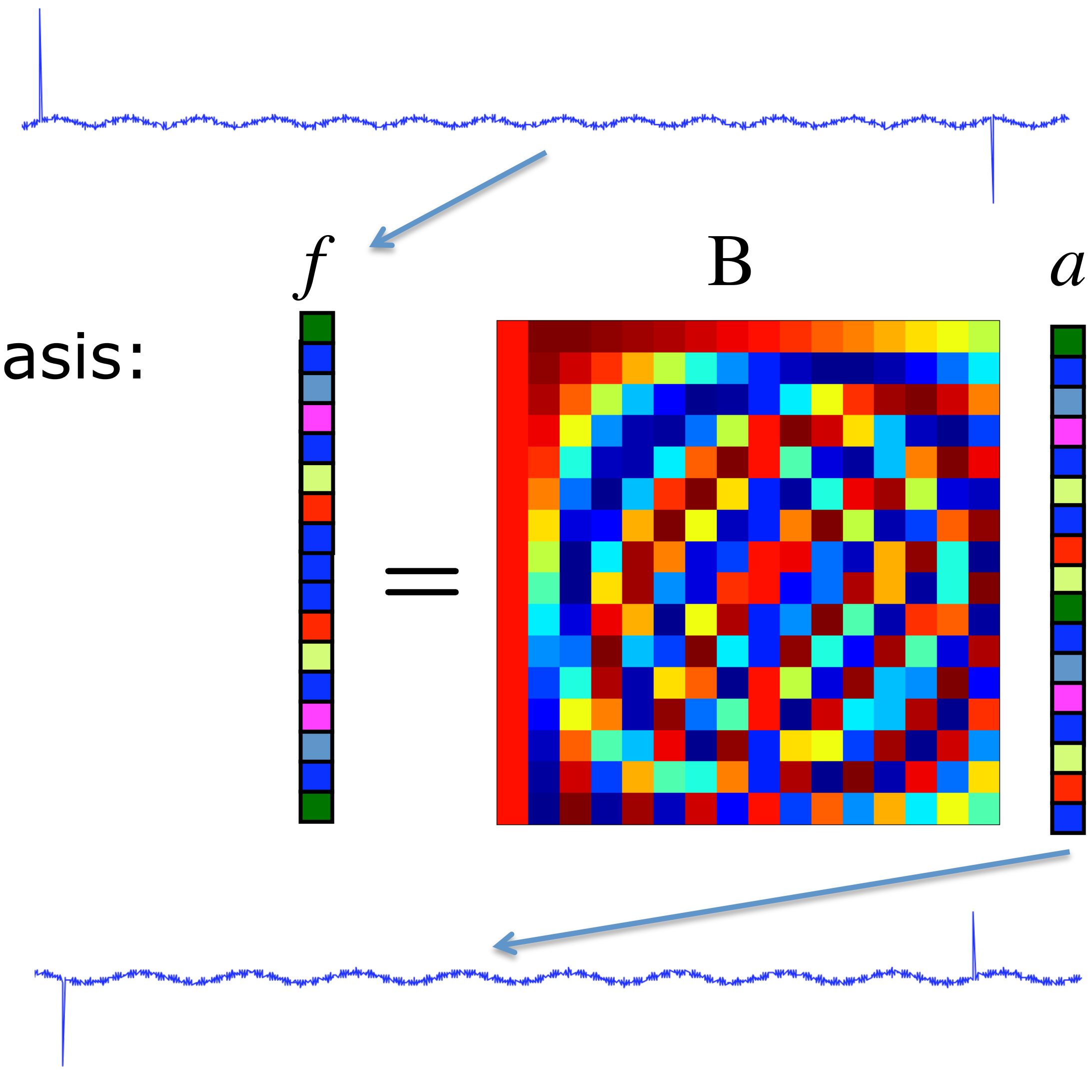
- Compression algorithm must encode *both* a set of a_k *and* their locations (harder)

Exposing Sparsity

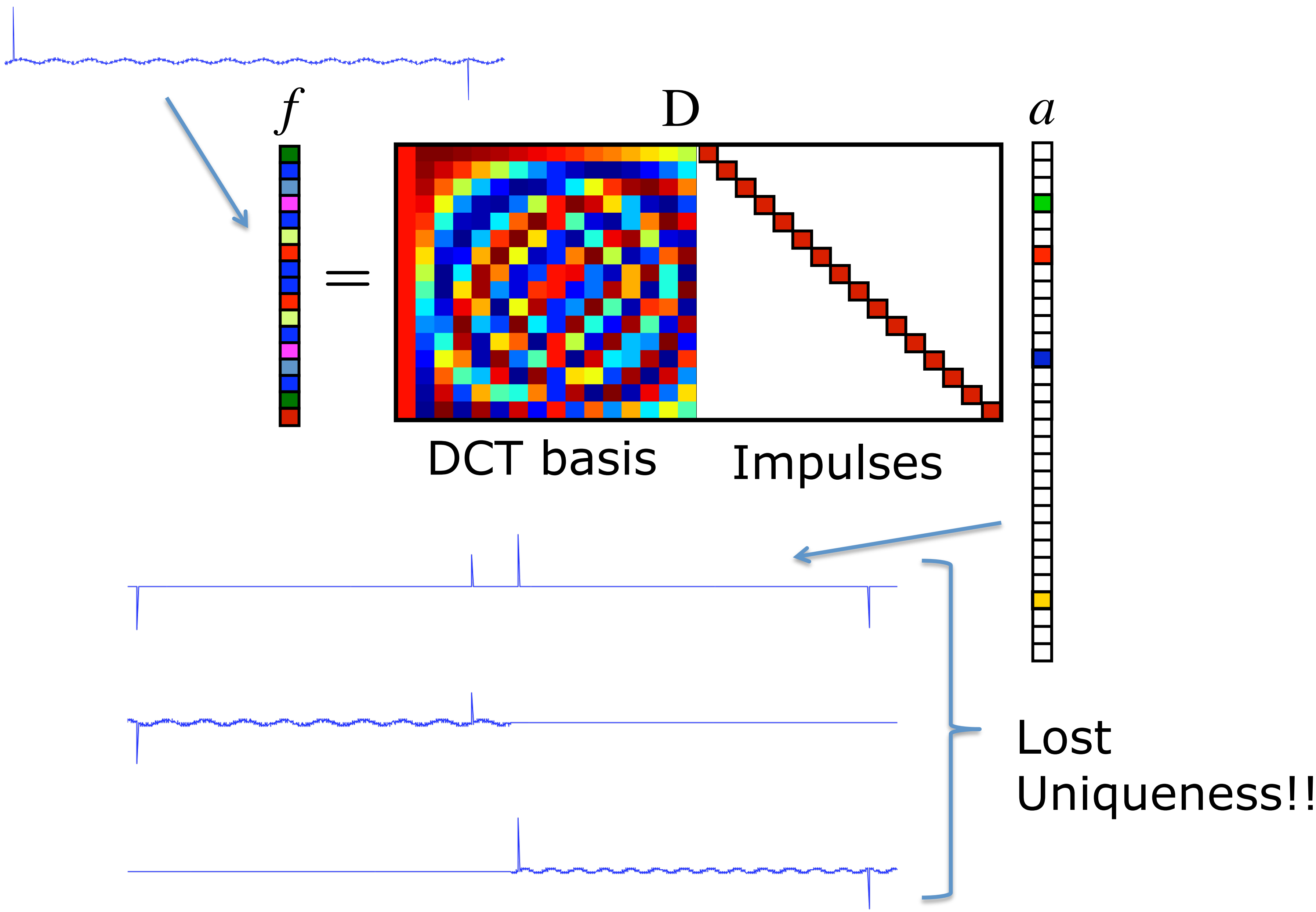
Spikes and Sinusoids example

Example Signal Model: Sinusoidal with a few spikes.

DCT Basis:

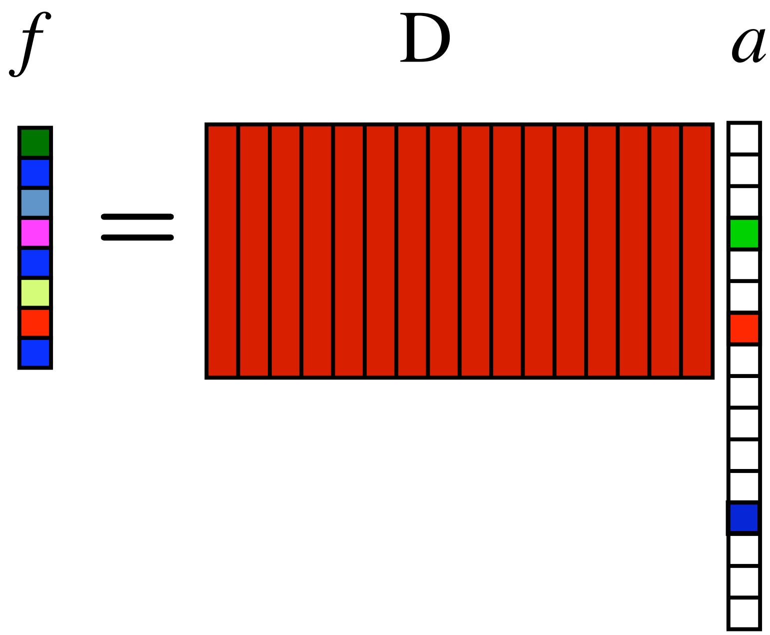


Spikes and Sinusoids Dictionary



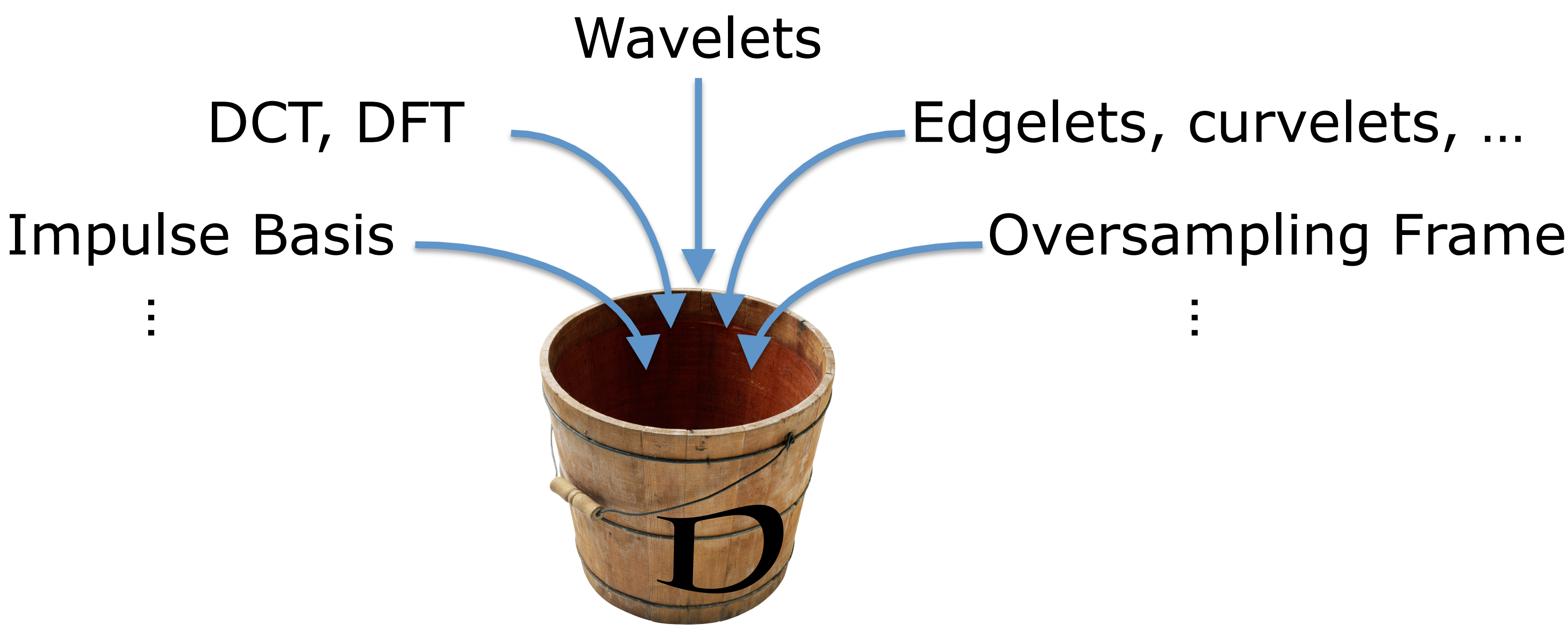
Overcomplete Dictionaries

Strategy: **Improve** sparse approximation
by constructing a large **dictionary.**



How do we **design** a dictionary?

Dictionary Design



Dictionary D

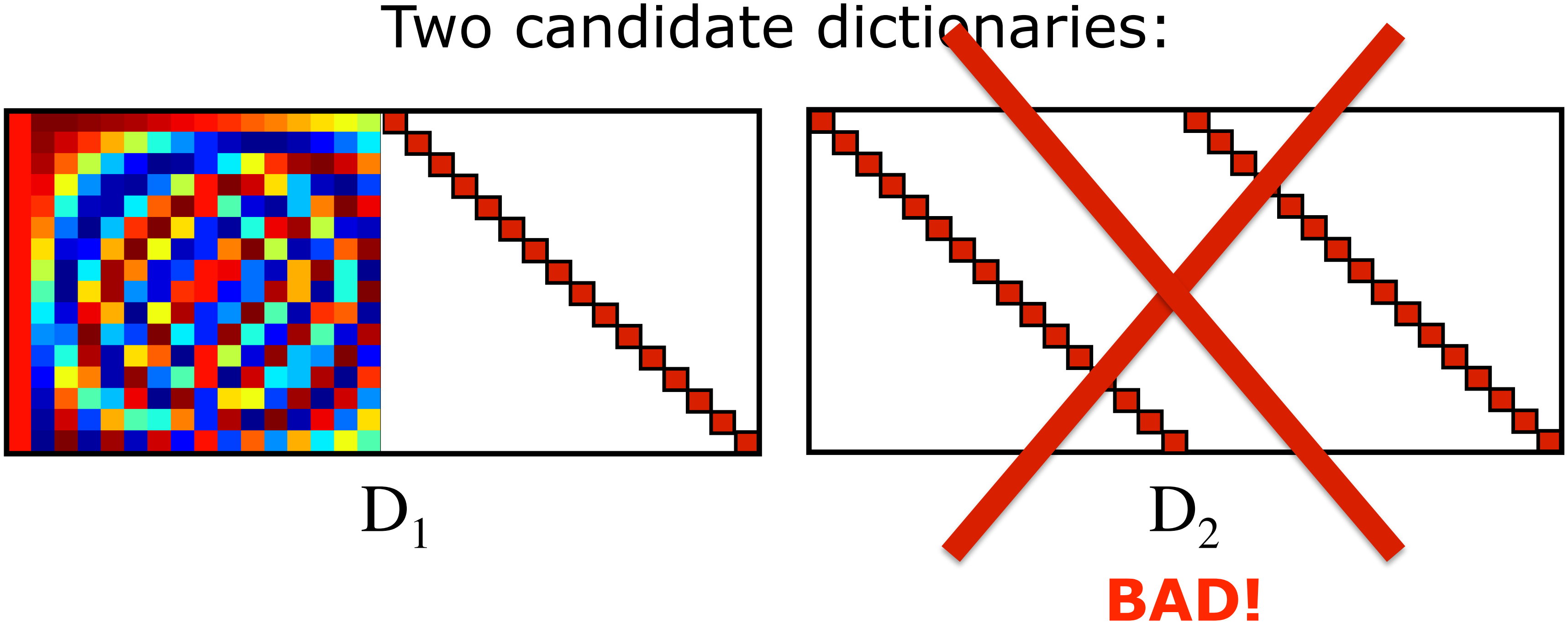
Can we just throw in the bucket **everything** we know?

Dictionary Design Considerations

- Dictionary Size:
  - **Computation** and **storage** **increases** with size
- Fast Transforms:
  - FFT, DCT, FWT, etc. dramatically **decrease** **computation** and **storage**
- Coherence:
  - **Similarity** in elements makes solution **harder**

Dictionary Coherence

Two candidate dictionaries:



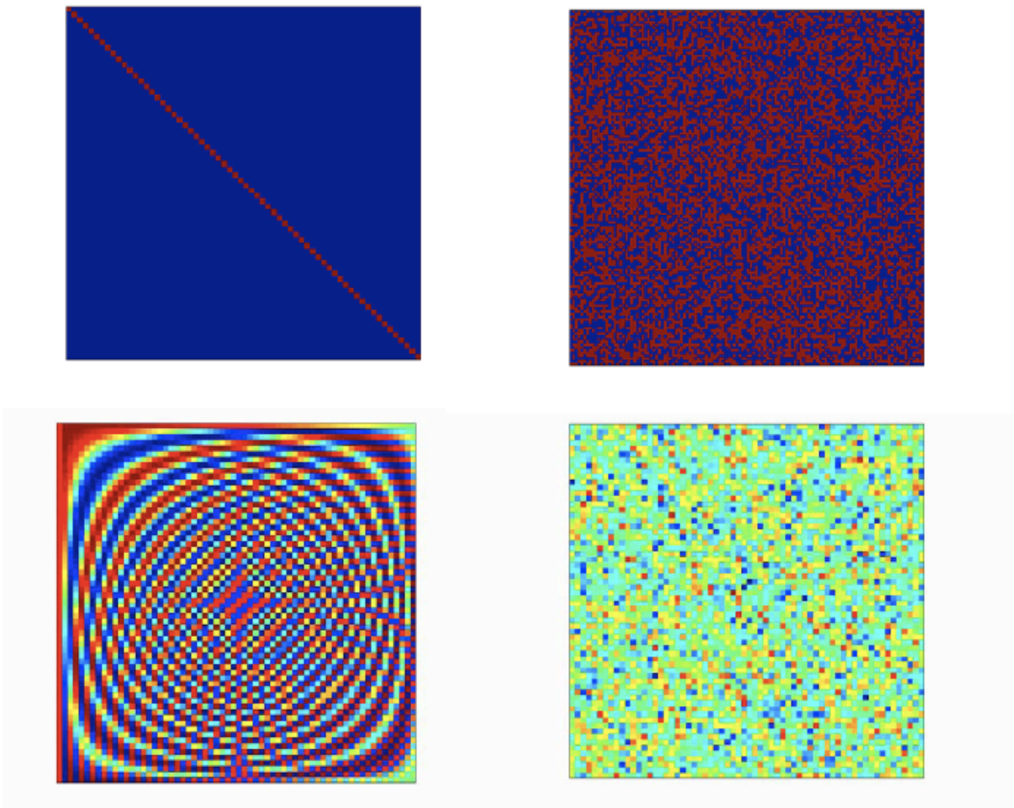
Intuition: D_2 has too many **similar** elements.
It is very **coherent**.

Coherence (similarity) between elements: $|\langle d_1, d_2 \rangle|$

Dictionary coherence: $\mu = \max_{i,j} |\langle d_i, d_j \rangle|$

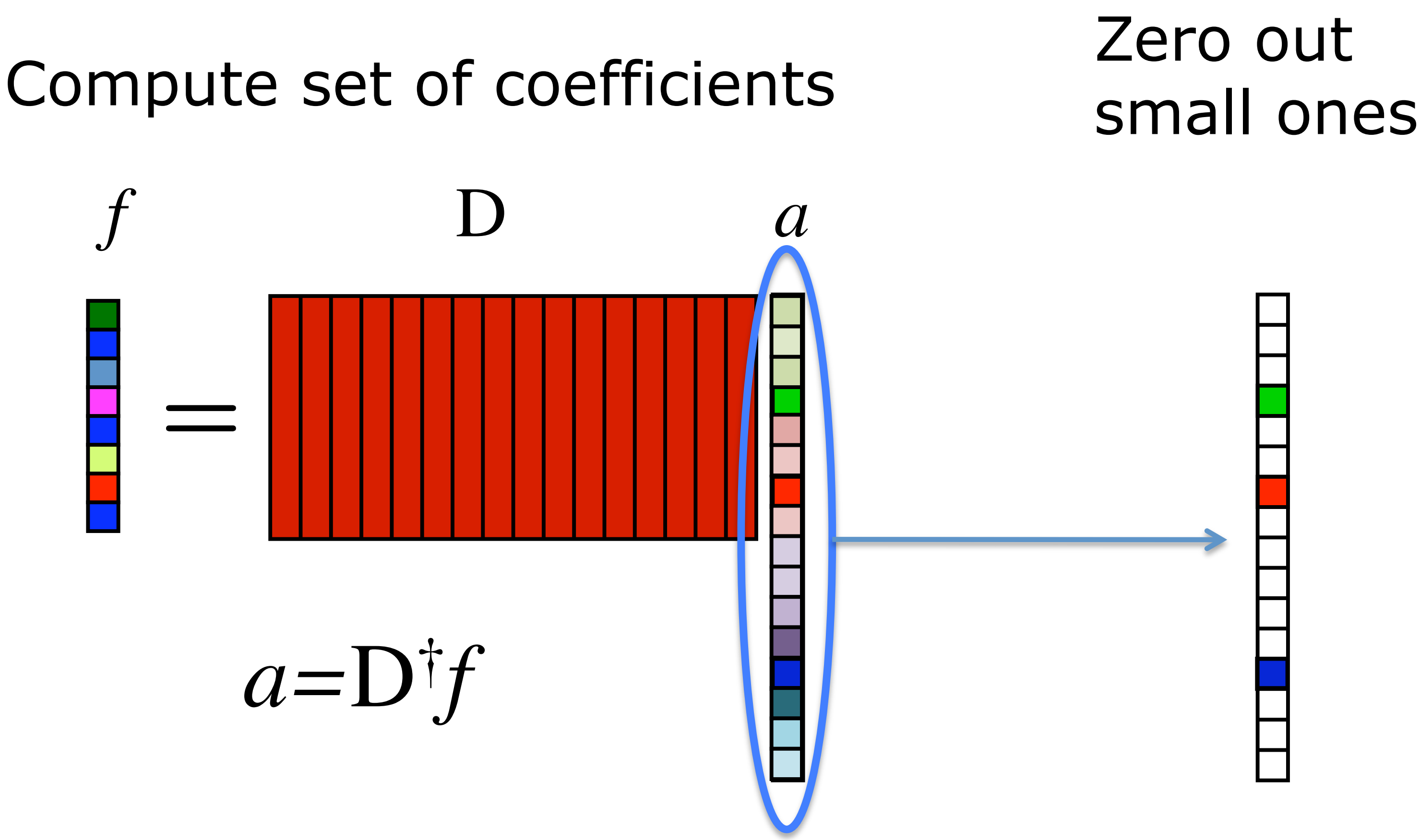
Incoherent Bases

- "Mix" well the signal components
  - Impulses and Fourier Basis
  - Anything and Random Gaussian
  - Anything and Random 0-1 basis



Computing
Sparse Representations

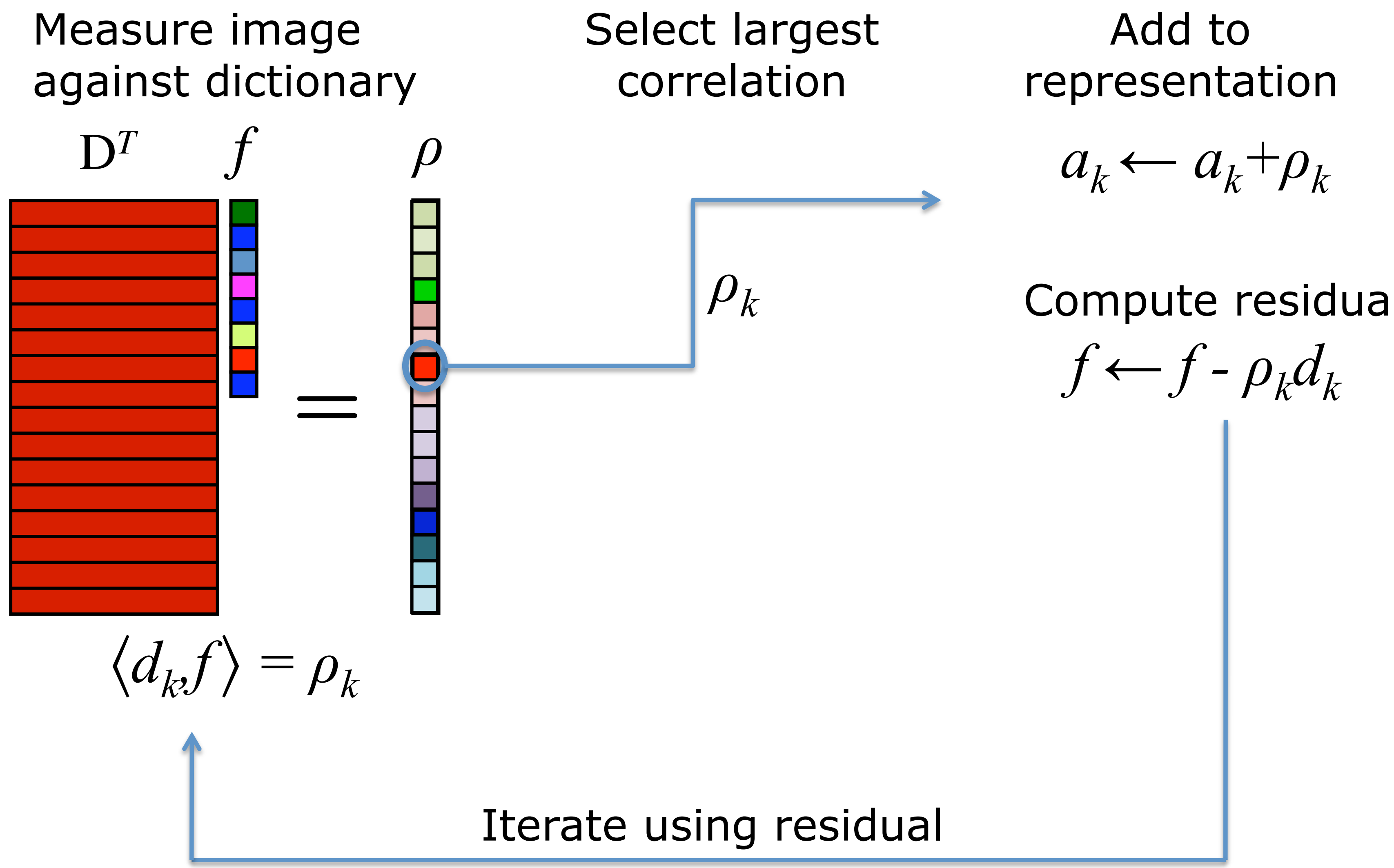
Thresholding



Computationally efficient

Good for small and very incoherent dictionaries

Matching Pursuit



Greedy Pursuits Family

- Several Variations of MP:
  OMP, StOMP, ROMP, CoSaMP, Tree MP, …
  (You can create an **AndrewMP** if you work on it…)

- Some have **provable guarantees**

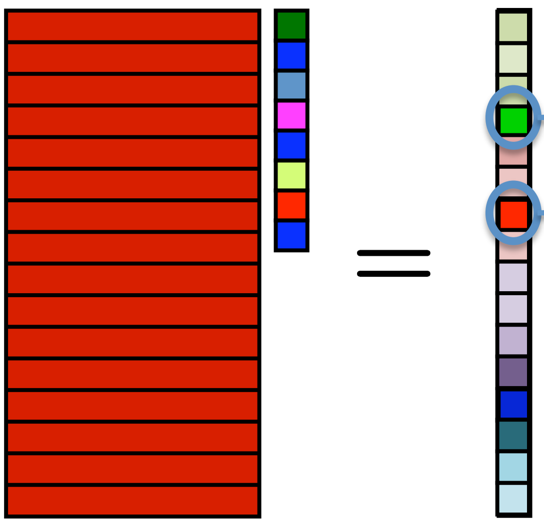
- Some improve **dictionary search**

- Some improve **coefficient selection**

CoSaMP (Compressive Sampling MP)

Measure image against dictionary

$\mathrm{D}^T \quad f \quad \rho$



$\langle d_k, f \rangle = \rho_k$

Select location of largest *2K* correlations

$\mathrm{supp}(\rho|_{2K})$

Add to support set

$$\Omega = \mathrm{supp}(\rho|_{2K}) \cup T$$

Invert over support

$$b = D_\Omega^\dagger f$$

Truncate and compute residual

$$T = \mathrm{supp}(b|_K)$$

$$a = b|_K$$

$$r \leftarrow f - Da$$

Iterate using residual

Optimization (Basis Pursuit)

Sparse approximation:

Minimize non-zeros in representation
s.t.: representation is **close** to signal

$$\min \|a\|_0 \text{ s.t. } f \approx \mathrm{D}a$$

Number of non-zeros
(sparsity measure)

Data Fidelity
(approximation quality)

Combinatorial complexity.
Very hard problem!

Optimization (Basis Pursuit)

Sparse approximation:

Minimize non-zeros in representation
s.t.: representation is **close** to signal

$$\min \ \| a \|_{\cancel{0}} \ \text{s.t.} \ f \approx \mathrm{D}a$$

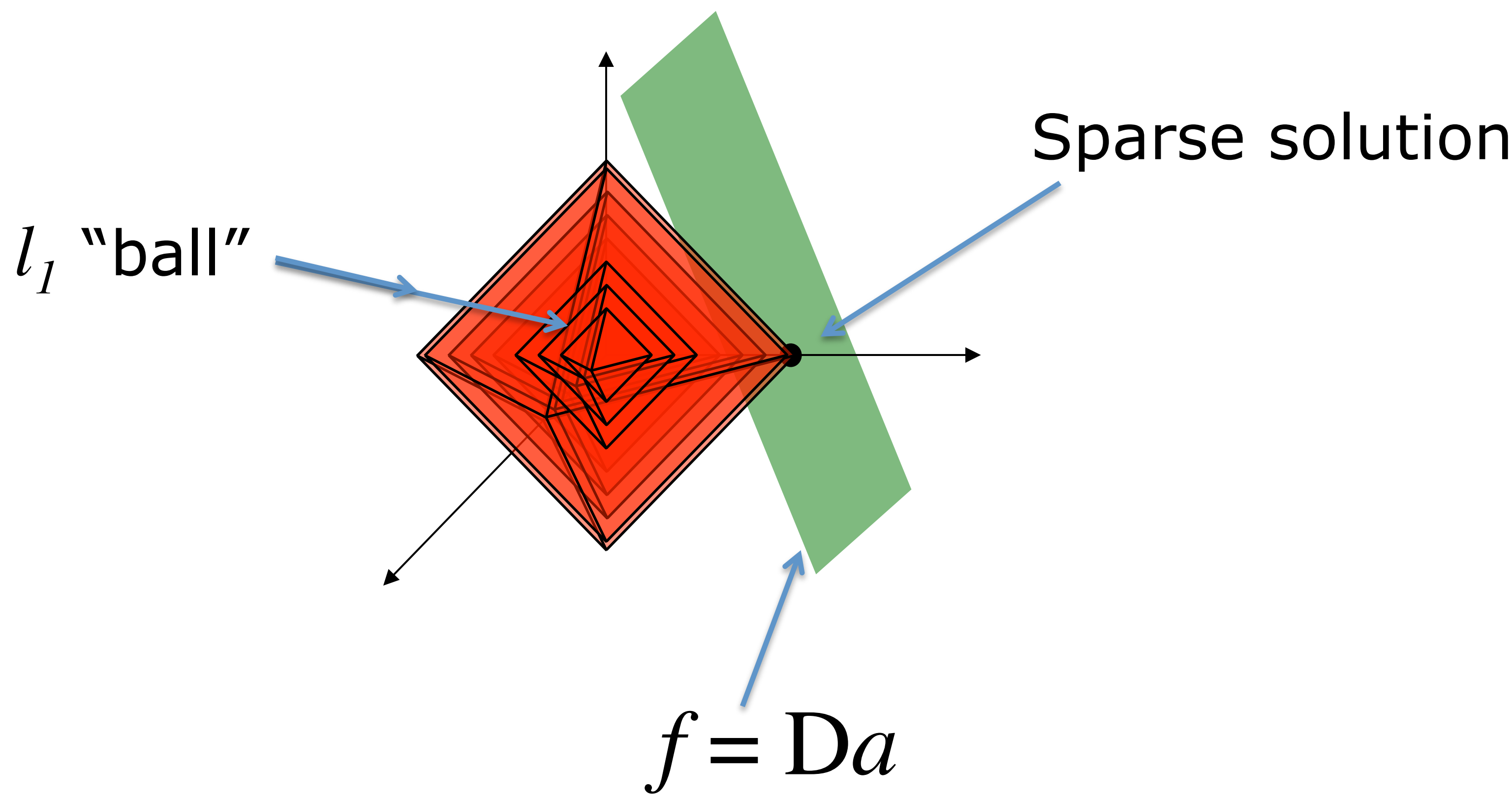
Convex Relaxation

$$\min \ \| a \|_1 \ \text{s.t.} \ f \approx \mathrm{D}a$$

Ploynomial complexity.
Solved using linear programming.

Why l_1 relaxation works

$$\min \|a\|_1 \text{ s.t. } f \approx \mathrm{D}a$$



Basis Pursuits

- Have **provable guarantees**
  - Finds **sparsest** solution for **incoherent** dictionaries

- Several variants in formulation:

  BPDN, LASSO, Dantzig selector, …

- Variations on **fidelity** term and **relaxation** choice

- Several fast algorithms:

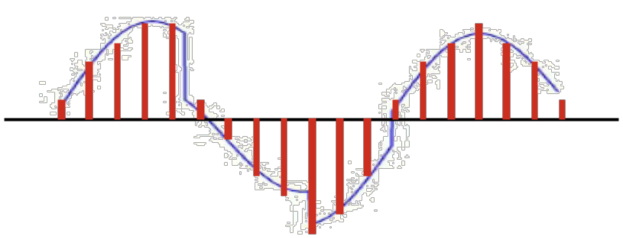
  FPC, GPSR, SPGL, …

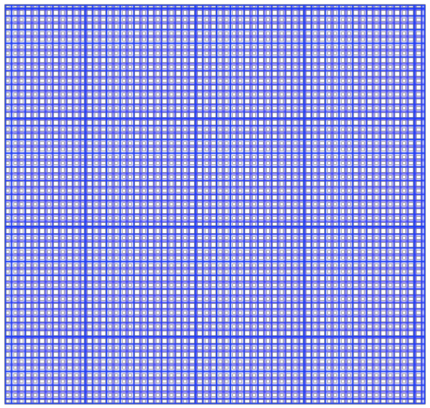
Compressed Sensing:
Sensing, Sampling and
Data Processing

Data Acquisition

- Usual acquisition methods **sample** signals uniformly
  - Time: A/D with microphones, geophones, hydrophones.
  - Space: CCD cameras, sensor arrays.

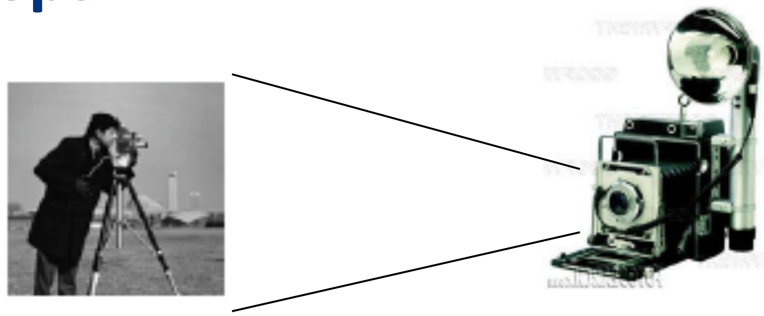
- Foundation: **Nyquist/Shannon sampling theory**
  - Sample at **twice the signal bandwidth.**
  - Generally a **projection to a complete basis** that spans the signal space.





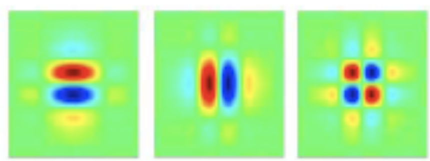
Data Processing and Transmission

- **Data processing steps**:
  - **Sample** Densely



Signal x,
N coefficients

  - **Transform** to an informative domain (Fourier, Wavelet)



sparse
wavelet
transform

K<<N significant
coefficients

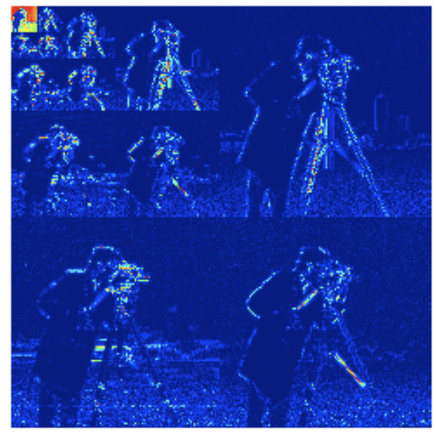
  - **Process/Compress/Transmit**

    Sets small coefficients to zero (sparsification)

Sparsity Model

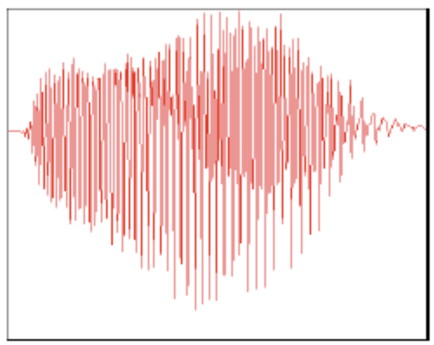
- Signals can usually be **compressed** in some basis

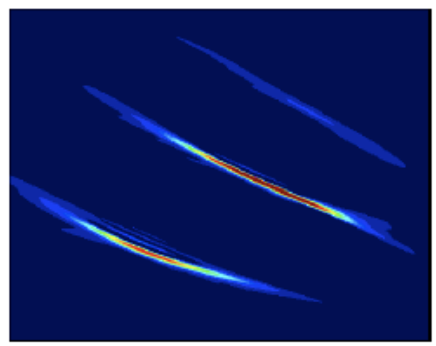
N
pixels



$K \ll N$
large
wavelet
coefficients

N
wideband
signal
samples

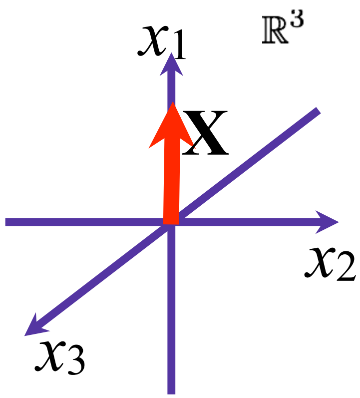




$K \ll N$
large
Gabor
coefficients

- Sparsity: good **prior** in picking from a lot of candidates

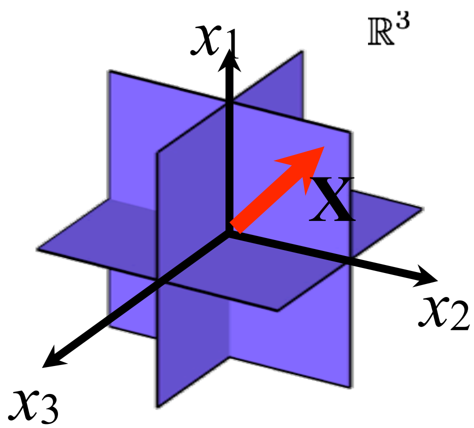
Compressive Sensing Principles



1-sparse

If a signal is sparse, do not waste effort sampling the empty space.

Instead, use fewer samples and allow ambiguity.



2-sparse

Use the sparsity model to reconstruct and uniquely resolve the ambiguity.

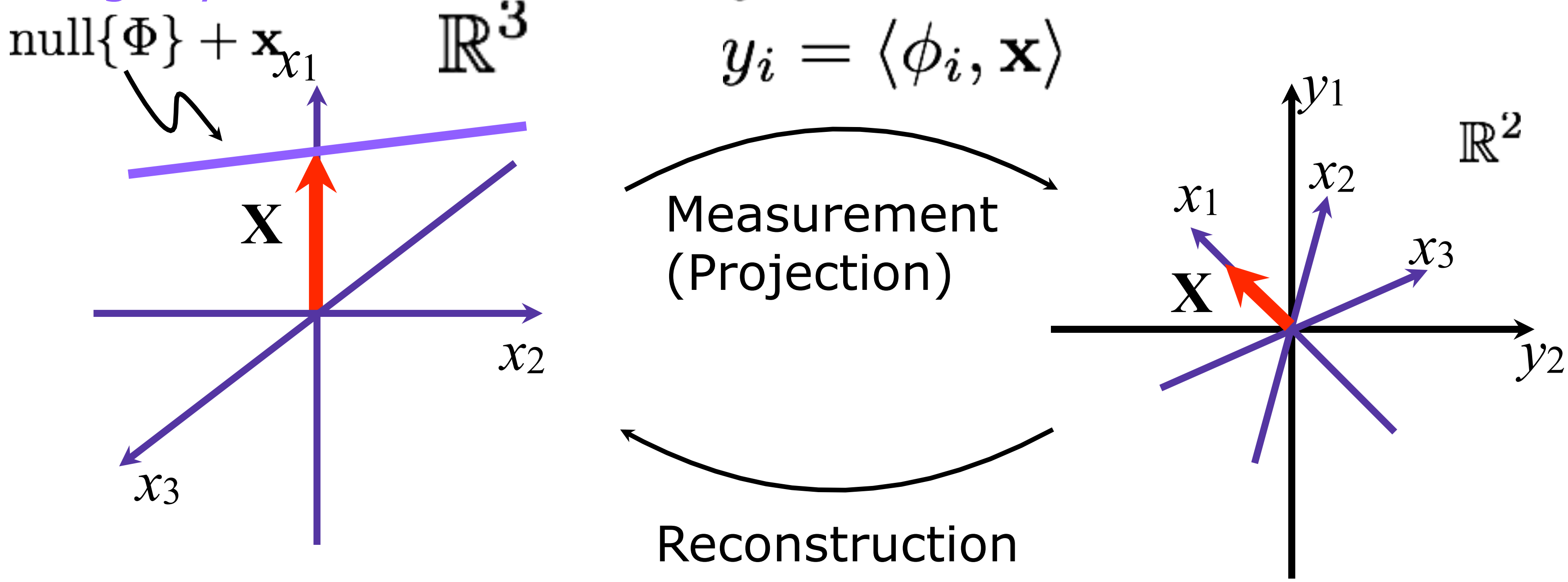
Measuring Sparse Signals

Compressive Measurements

Ambiguity

$$\mathbf{y} = \Phi \mathbf{x}$$

$$y_i = \langle \phi_i, \mathbf{x} \rangle$$



Φ has rank $M \ll N$

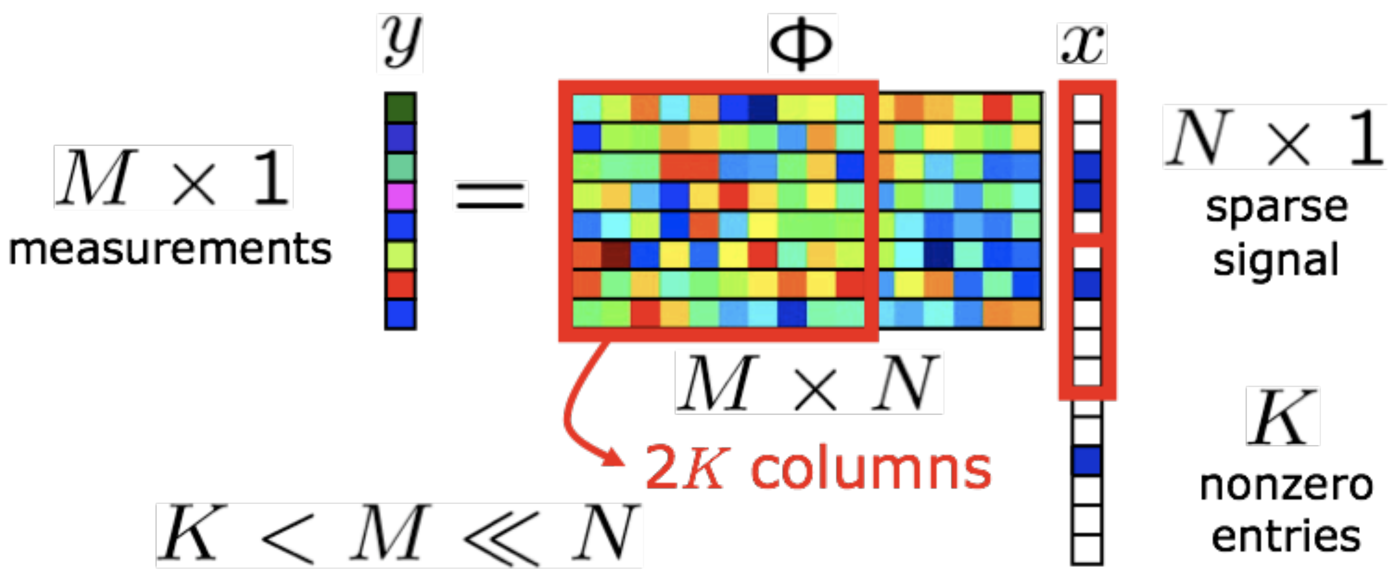
$N =$ Signal dimensionality
$K =$ Signal sparsity

$M =$ Number of measurements
(dimensionality of $\mathbf{y}$)

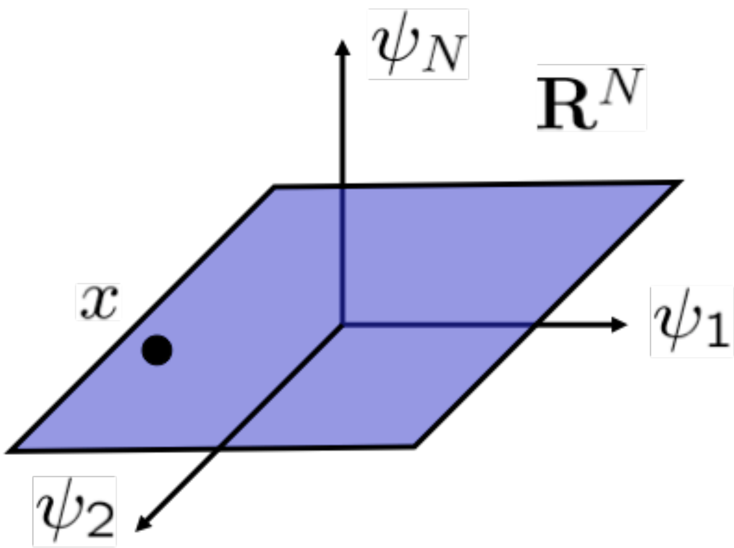
$$N \gg M \gtrsim K$$

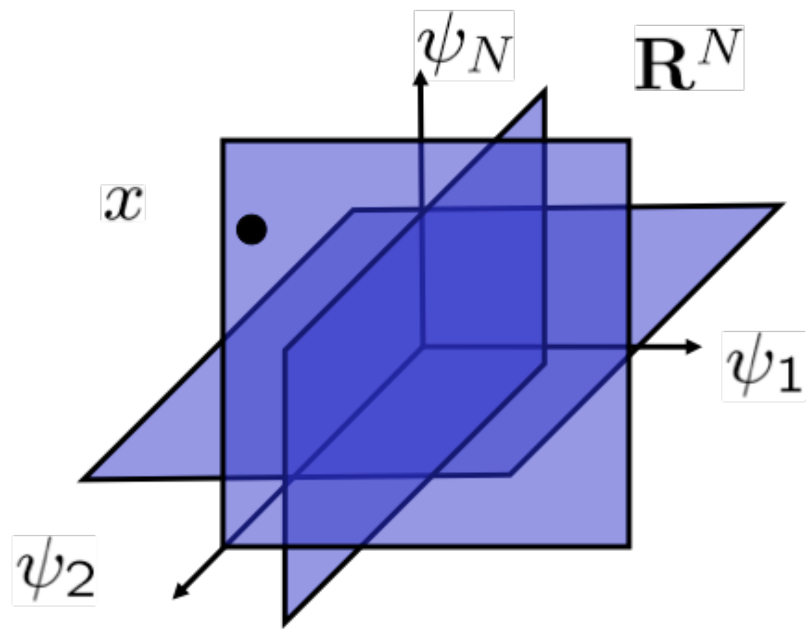
One Simple Question

- When is it possible to distinguish 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



Geometry of Sparse Signal Sets





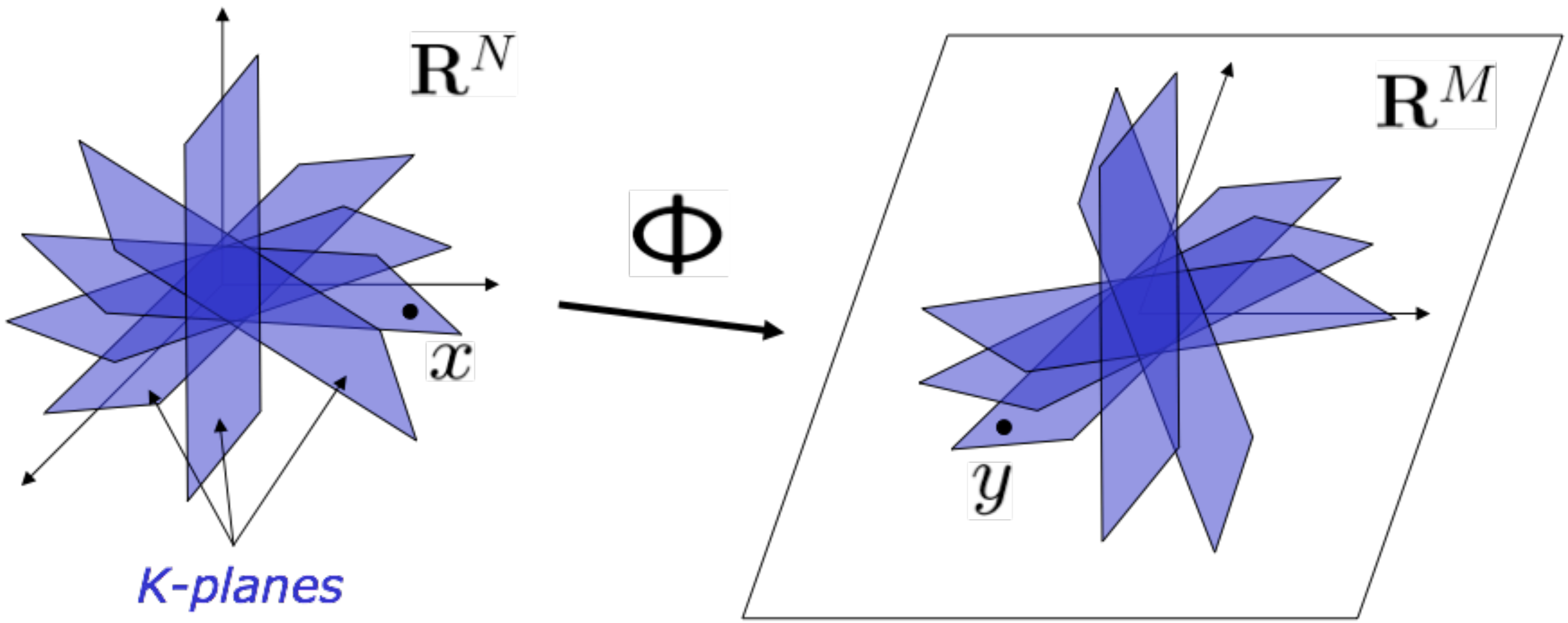
Linear

Sparse, Nonlinear

K-plane

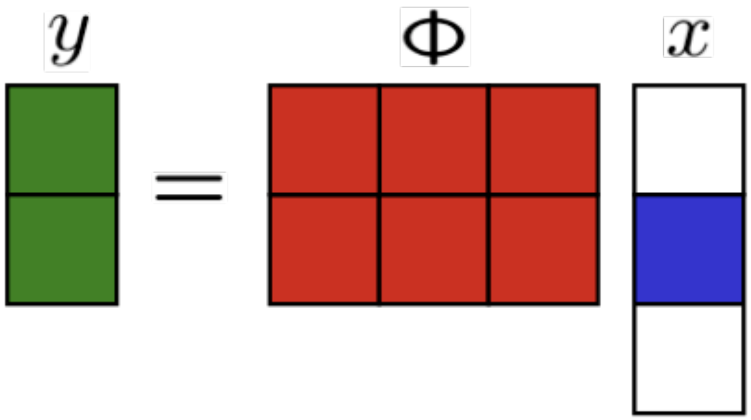
Union of K-planes

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
- See also FROI [Vetterli et al., Lu and Do]

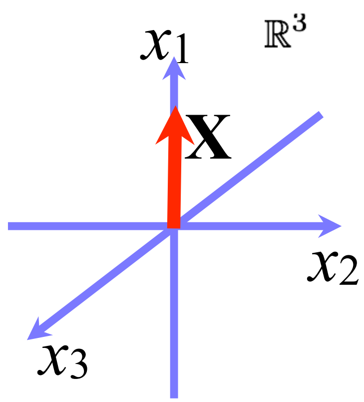
Illustrative Example



N = 3: signal length
K = 1: sparsity
M = 2K = 2: measurements

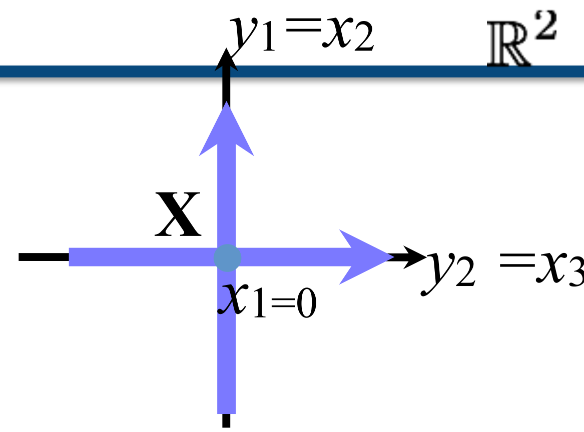
Example: 1-sparse signal

$N=3$
$K=1$
$M=2K=2$



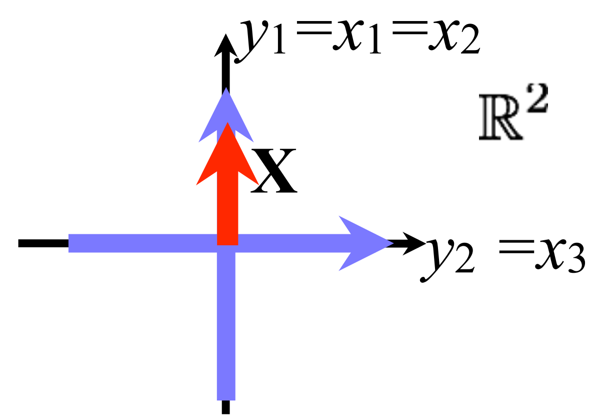
$$\Phi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Bad!



$$\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

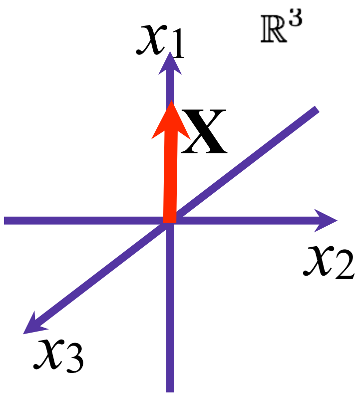
Bad!



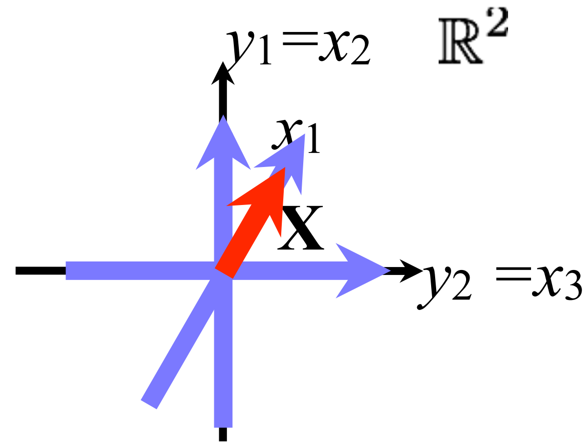
Example: 1-sparse signal

$N=3$
$K=1$
$M=2K=2$

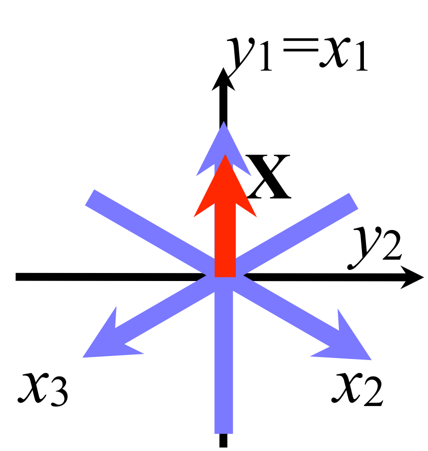
$$\Phi = \begin{bmatrix} 1 & 1 & 0 \\ 1/2 & 0 & 1 \end{bmatrix}$$



Good!



Better!

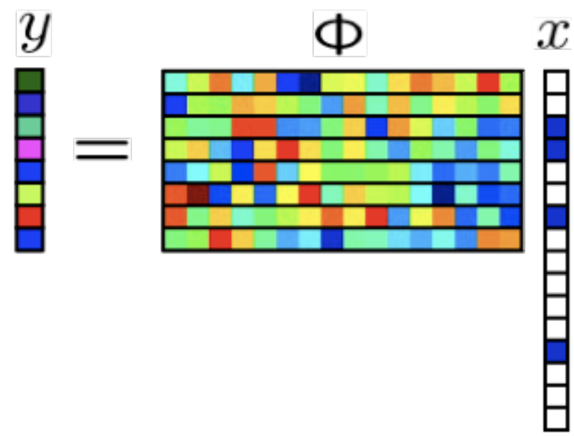


$$\Phi = \begin{bmatrix} 1 & -1/2 & -1/2 \\ 0 & \sqrt{3}/2 & -\sqrt{3}/2 \end{bmatrix}$$

Restricted Isometry Property

[Candès, Romberg, Tao]

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



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

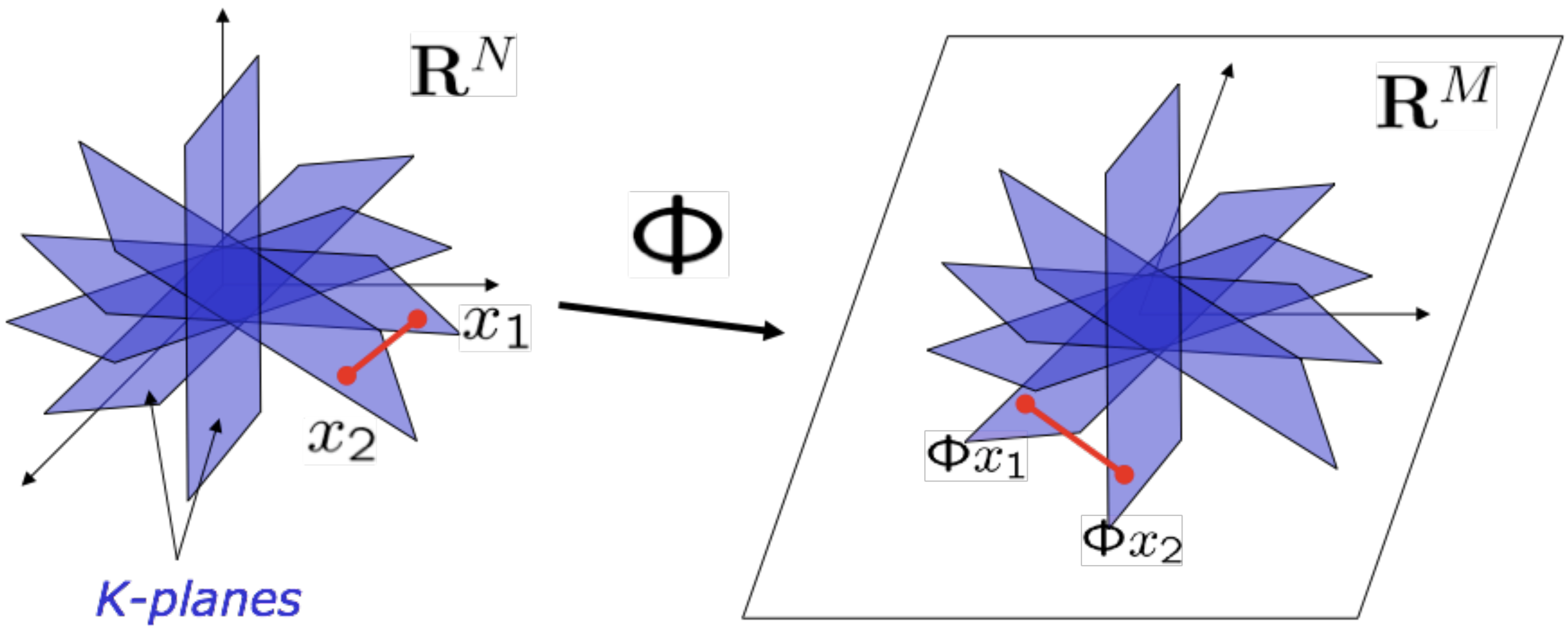
for all K-sparse signals x.

- 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)

Verifying RIP

How Many Measurements?

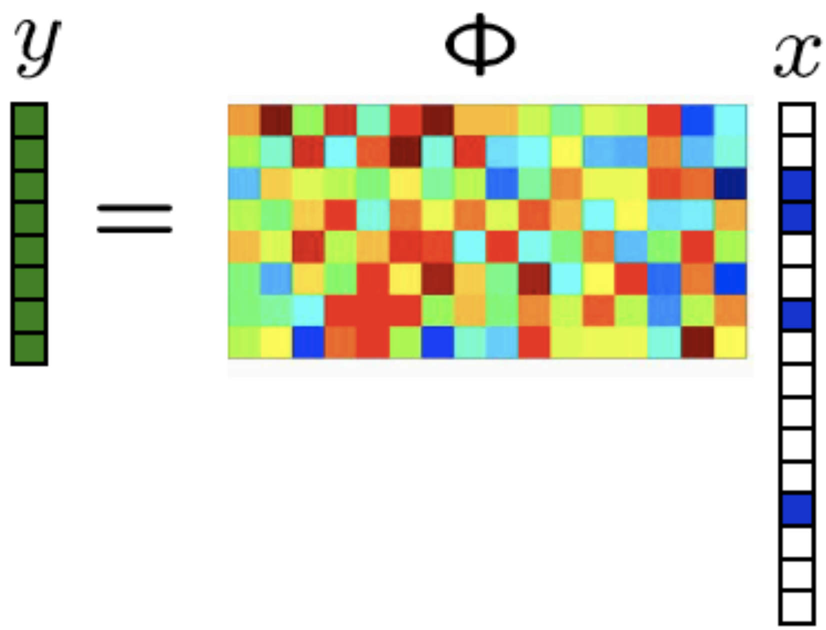
- Want RIP of order 2K (say) to hold for MxN Φ
  - difficult to verify for a given Φ
  - requires checking eigenvalues of each submatrix

- 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

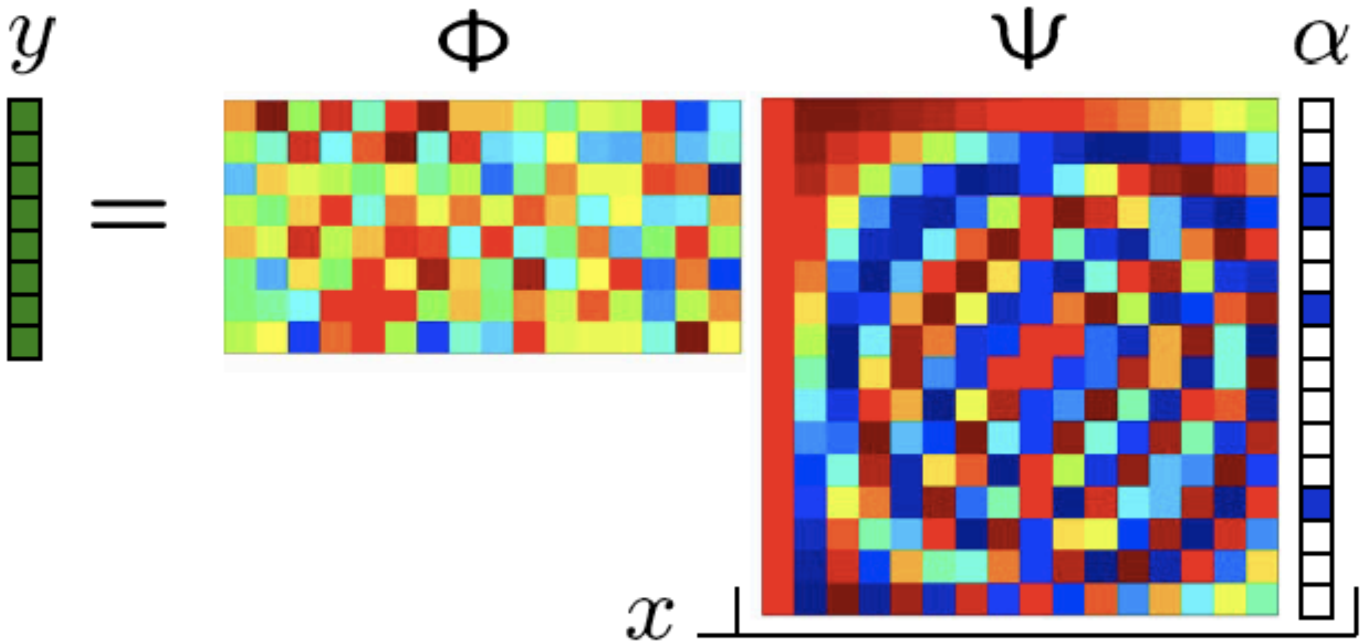
Universality Property

- Gaussian white noise basis is incoherent with ***any*** fixed orthonormal basis (with high probability)
- Signal sparse in time domain: $\Phi = I$



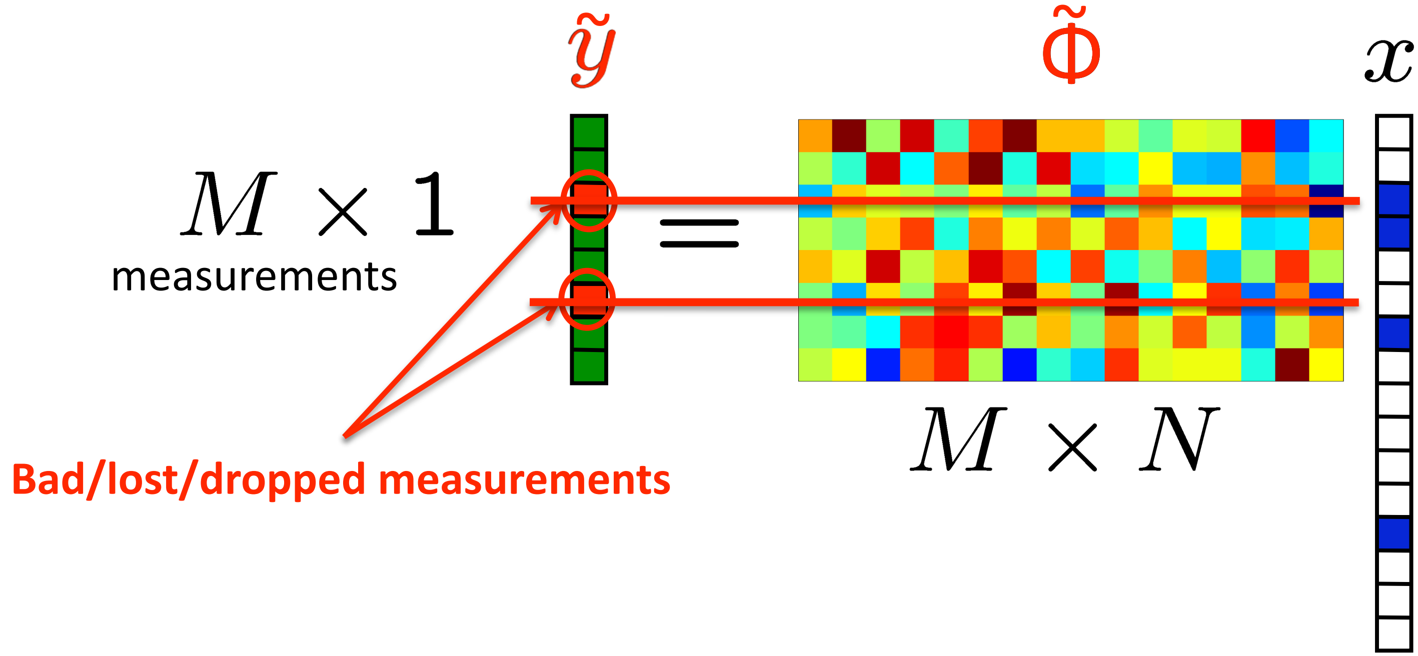
Universality Property

- Gaussian white noise basis is incoherent with ***any*** fixed orthonormal basis (with high probability)
- Signal sparse in frequency domain: $\Psi = \text{idct}$



- Product $\Phi\Psi$ remains Gaussian white noise

Democracy



- Measurements are **democratic** [Davenport, Laska, Boufounos, Baraniuk]
  - They are all equally important
  - We can **loose** some **arbitrarily,** (i.e. an adversary can choose which ones)

- The $\tilde{\Phi}$ still satisfies RIP (as long as we don't drop too many)

Reconstruction

Requirements for Reconstruction

- Let x_1, x_2 be K-sparse signals (I.e. x_1-x_2 is $2K$-sparse):
- Mapping $y=\mathbf{\Phi}x$ is **invertible** for K-sparse signals:

$$\mathbf{\Phi}(x_1-x_2)\neq 0 \text{ if } x_1\neq x_2$$

- Mapping is **robust** for K-sparse signals:

$$||\mathbf{\Phi}(x_1-x_2)||_2\approx|| x_1-x_2||_2$$

  - Restricted Isometry Property (**RIP**):

    $\mathbf{\Phi}$ preserves distance when projecting K-sparse signals

- Guarantees there exists a **unique** K-sparse signal explains the measurements, and is robust to noise.

Reconstruction Ambiguity

- Solution should be **consistent** with measurements

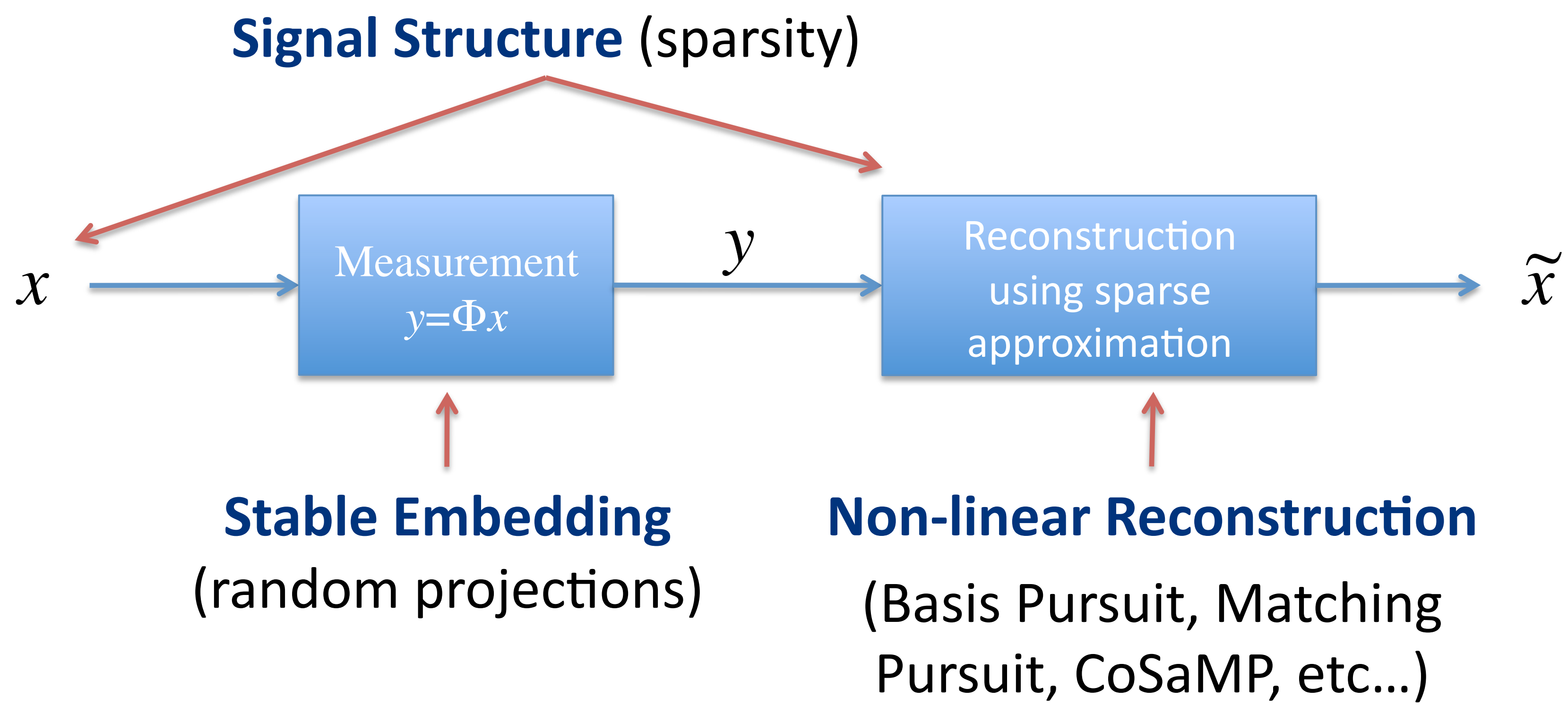
$$\hat{\mathbf{x}} \quad \text{s.t.} \quad \mathbf{y} = \Phi\hat{\mathbf{x}} \quad \text{or} \quad \mathbf{y} \approx \Phi\hat{\mathbf{x}}$$

- Projections imply that an **infinite** number of solutions are consistent!
- Classical approach: use the pseudoinverse (minimize l_2 norm)
- Compressive sensing approach: pick the **sparsest**.
- **RIP guarantee**: sparsest solution **unique** and **reconstructs the signal.**

Becomes a **sparse approximation problem!**

Putting everything together

Compressed Sensing Coming Together

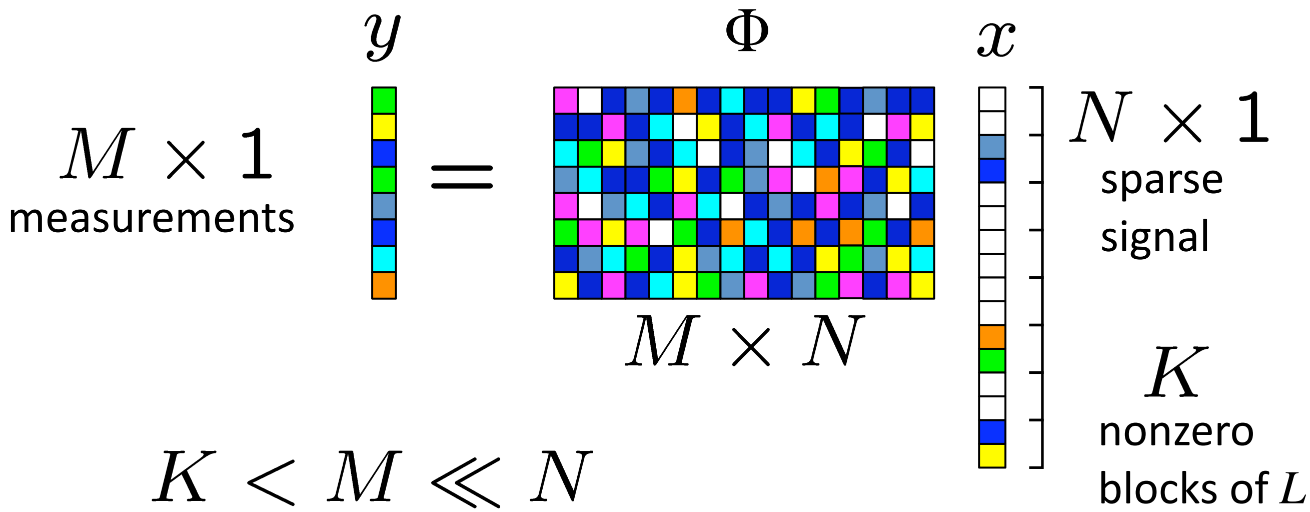


- **Signal model:** Provides prior information; allows undersampling
- **Randomness:** Provides robustness/stability; makes proofs easier
- **Non-linear reconstruction:** Incorporates information through computation

Beyond: Extensions, Connections, Generalizations

Sparsity Models

Block Sparsity



Mixed l_1/l_2 norm—sum of l_2 norms: $\sum_i \|\mathbf{x}_{B_i}\|_2$

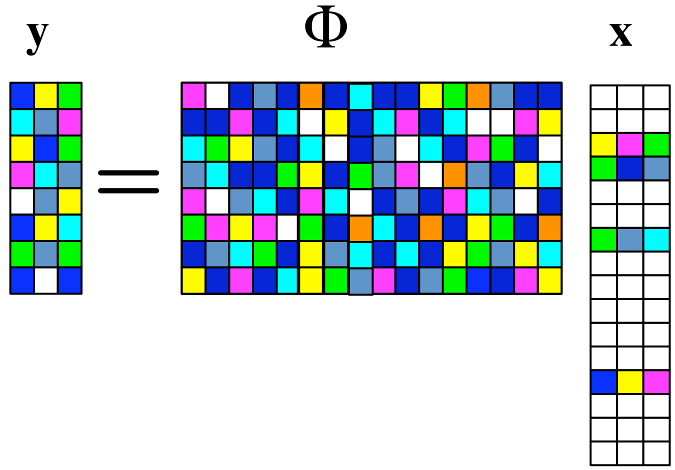
Basis pursuit becomes: $\min_{\mathbf{x}} \sum_i \|\mathbf{x}_{B_i}\|_2 \text{ s.t. } y \approx \Phi x$

Blocks are not allowed to overlap

Joint Sparsity

$M \times L$

measurements



$N \quad L$

L sparse signals in R^N

K

Sparse components per signal with common support

Mixed l_1/l_2 norm—sum of l_2 norms: $\sum_i \|\mathbf{x}_{(i,\cdot)}\|_2$

Basis pursuit becomes: $\min_{\mathbf{x}} \sum_i \|\mathbf{x}_{(i,\cdot)}\|_2 \text{ s.t. } \mathbf{y} \approx \Phi \mathbf{x}$

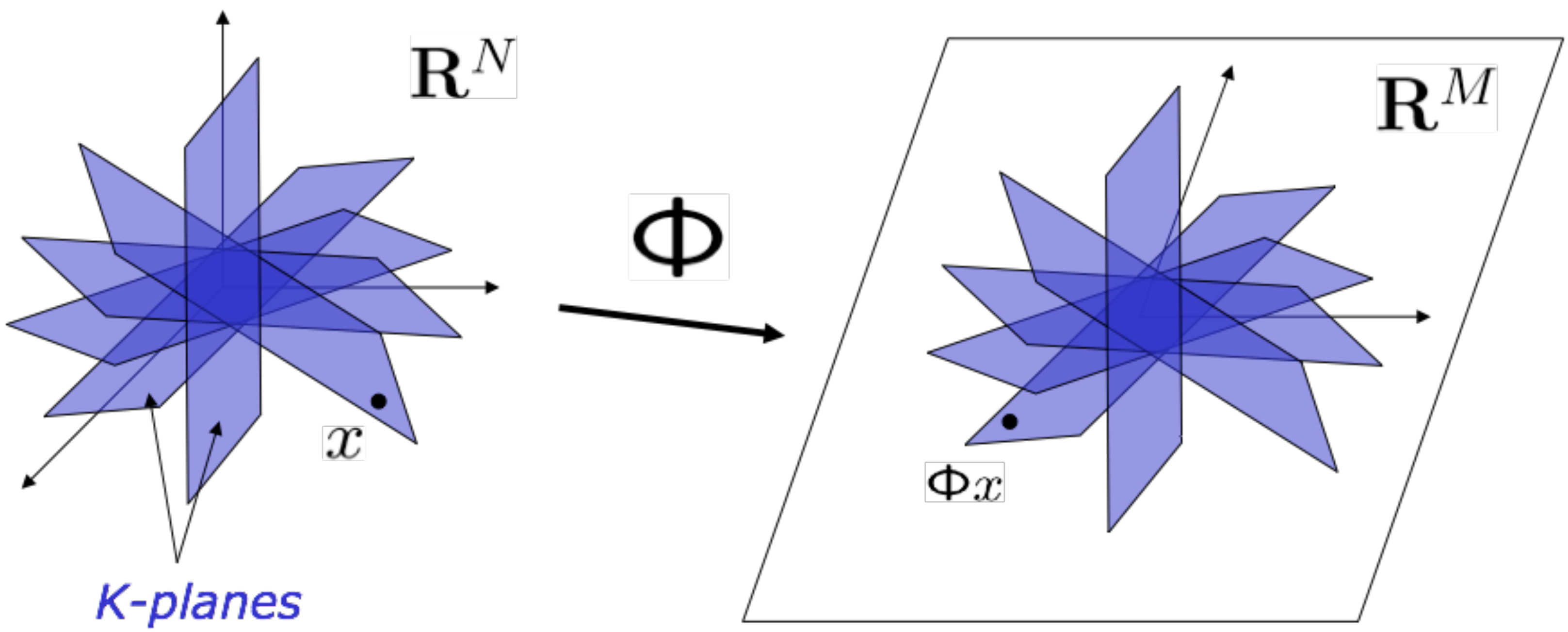
Randomized Embeddings

Stable Embeddings

Recall: RIP

- RIP of order K requires: for all K-sparse x,

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

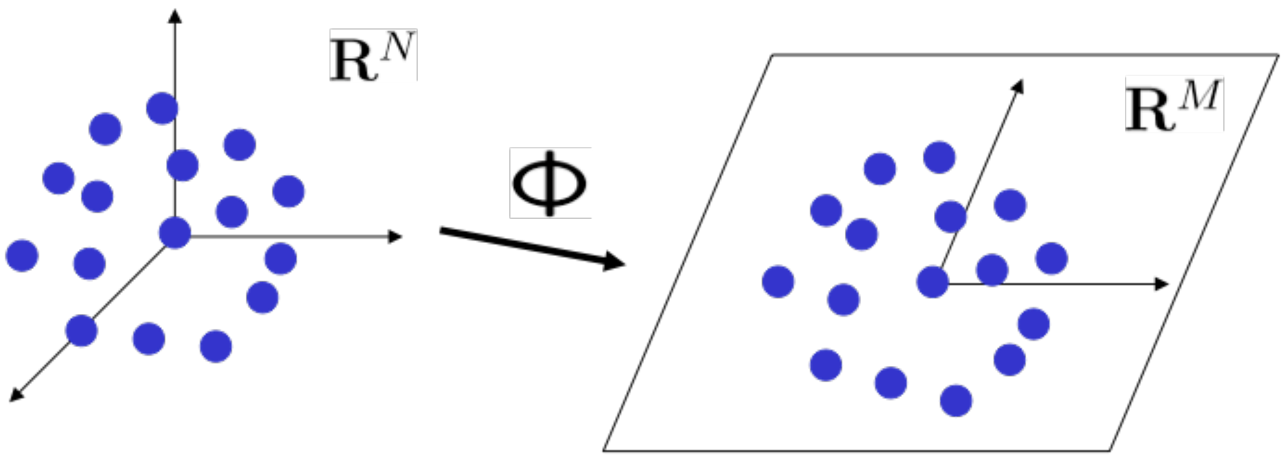


Johnson-Lindenstrauss Lemma

[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]

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

- "Database-friendly" [Achlioptas]

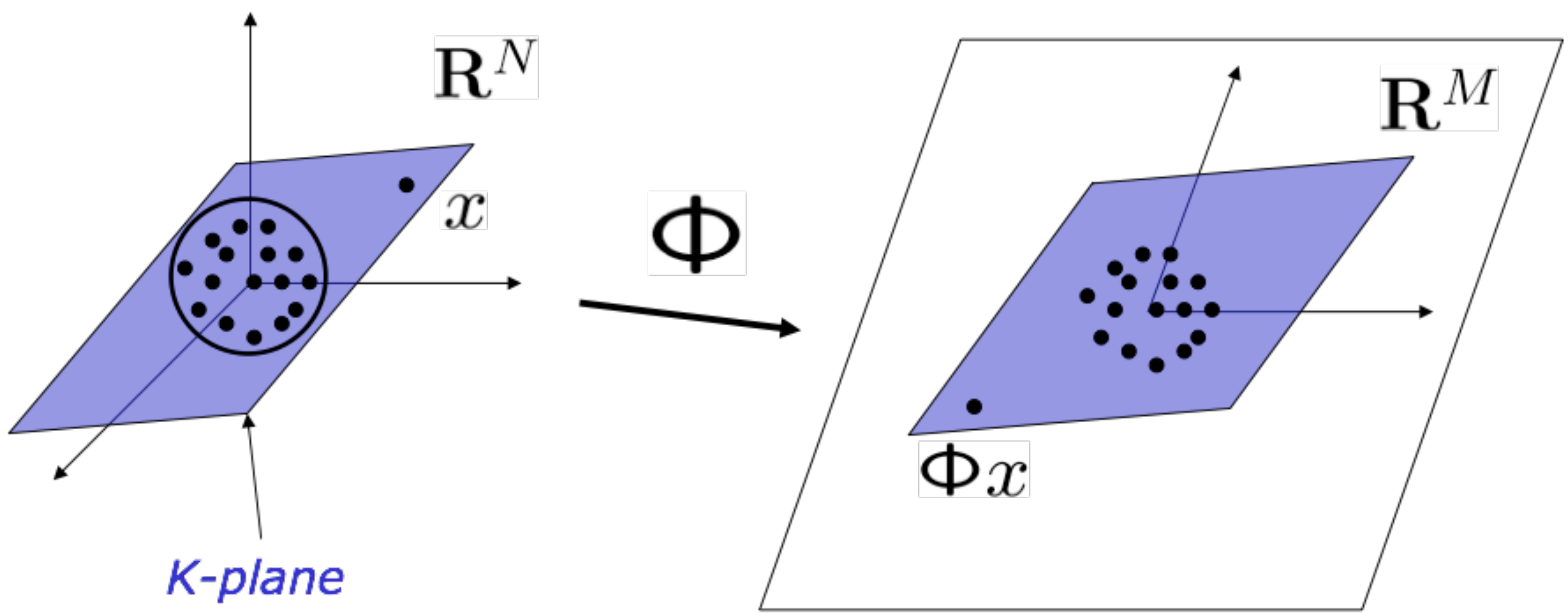
$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6}, \\ 0 & \text{with probability } \frac{2}{3}, \\ -\sqrt{\frac{3}{M}} & \text{with probability } \frac{1}{6} \end{cases}$$

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

Connecting JL to RIP

Consider effect of random JL Φ on each K-plane

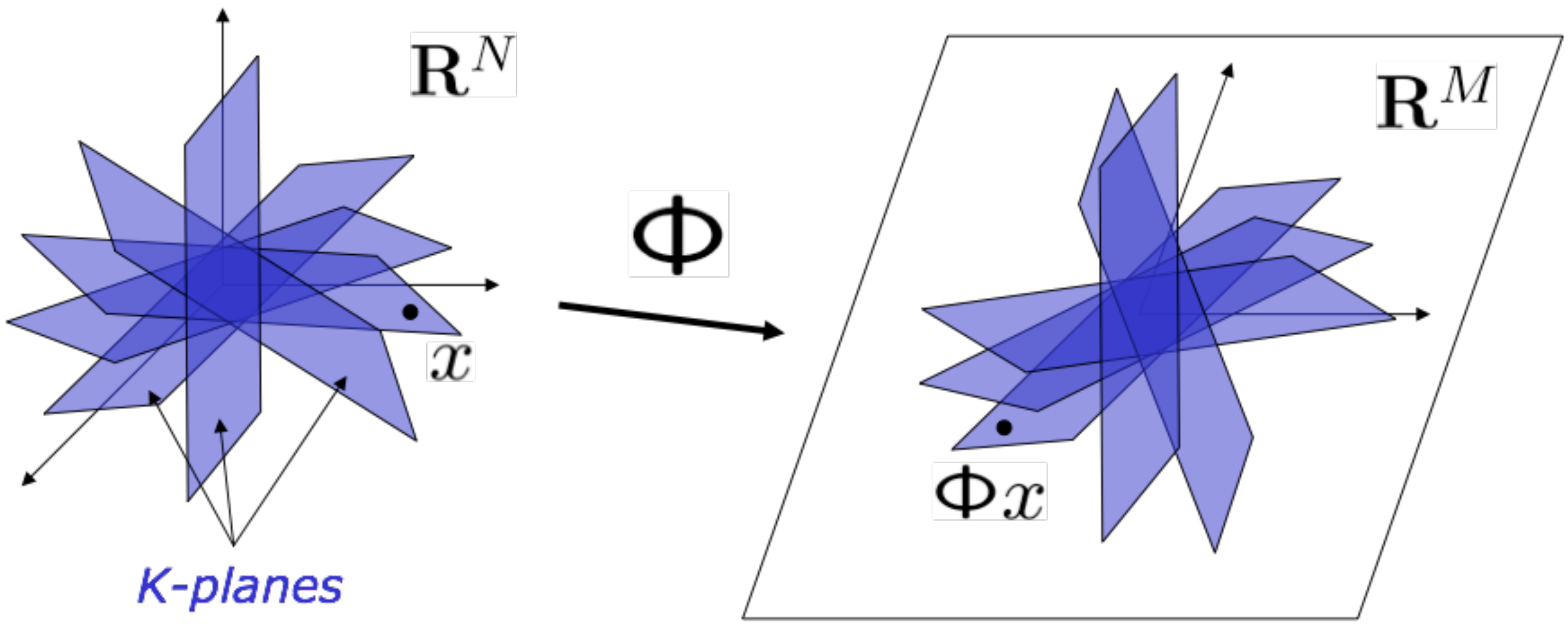
- construct covering of points Q on unit sphere
- JL: isometry for each point with high probability
- union bound ➔ isometry for all $q \in Q$
- extend to isometry for all x in K-plane



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



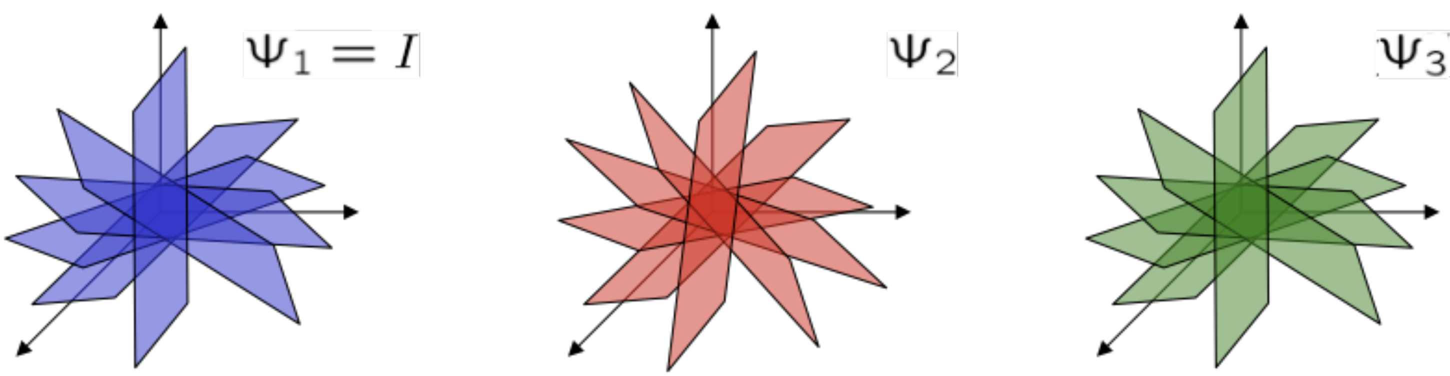
Connecting JL to RIP [Baraniuk, DeVore, Davenport, Wakin]

- **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

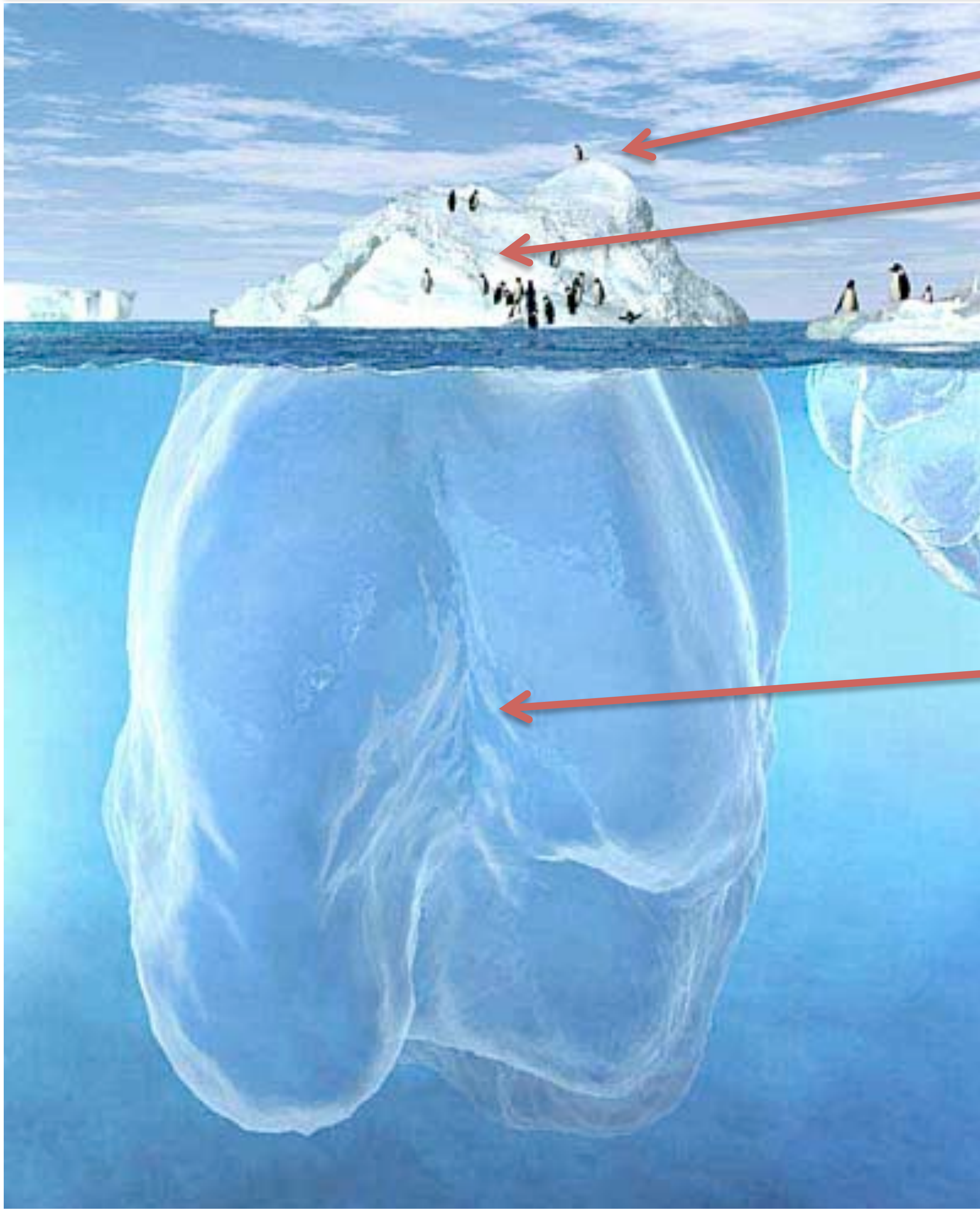
- Bonus: *universality* (repeat argument for any Ψ)



- See also Mendelson et al. concerning subgaussian ensembles

More?

The tip of the iceberg



Today's lecture

Compressive Sensing Repository
dsp.rice.edu/cs

Blog on CS
nuit-blanche.blogspot.com/

Yet to be discovered…
Start working on it ☺