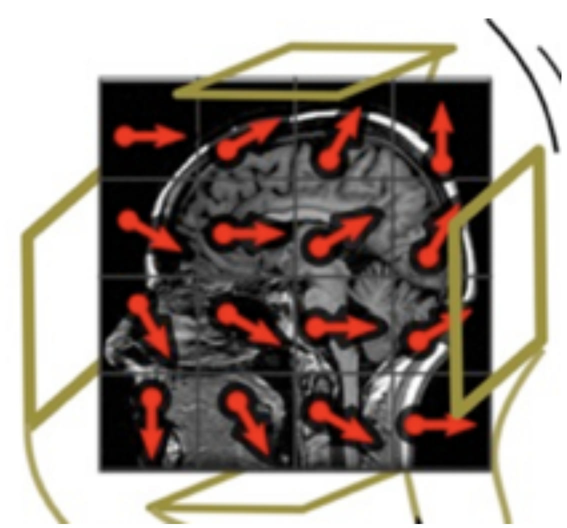
An introduction to Compressive Sampling

Kévin Polisano



Based on these two following articles :

Robust Uncertainty Principle : Exact reconstruction from highly incomplete frequency information (2004)

&

An introduction to compressive sampling (2008)

Pressure is on Digital Sensors

- Success of digital data acquisition is placing increasing pressure on signal/image processing hardware and software to support

higher resolution / denser sampling

» ADCs, cameras, imaging systems, microarrays, …

x

large numbers of sensors

» image data bases, camera arrays, distributed wireless sensor networks, …

x

increasing numbers of modalities

» acoustic, RF, visual, IR, UV

=

deluge of data

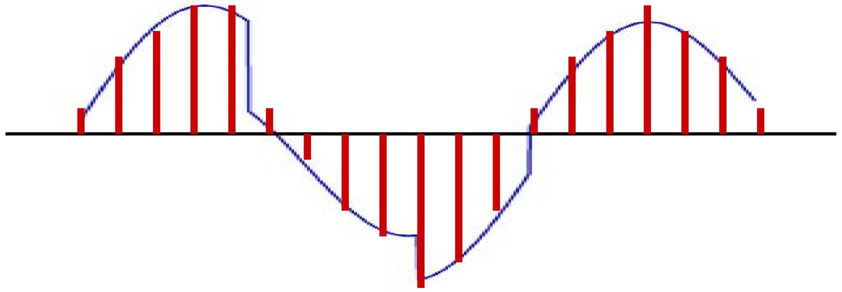
» how to acquire, store, fuse, process efficiently?

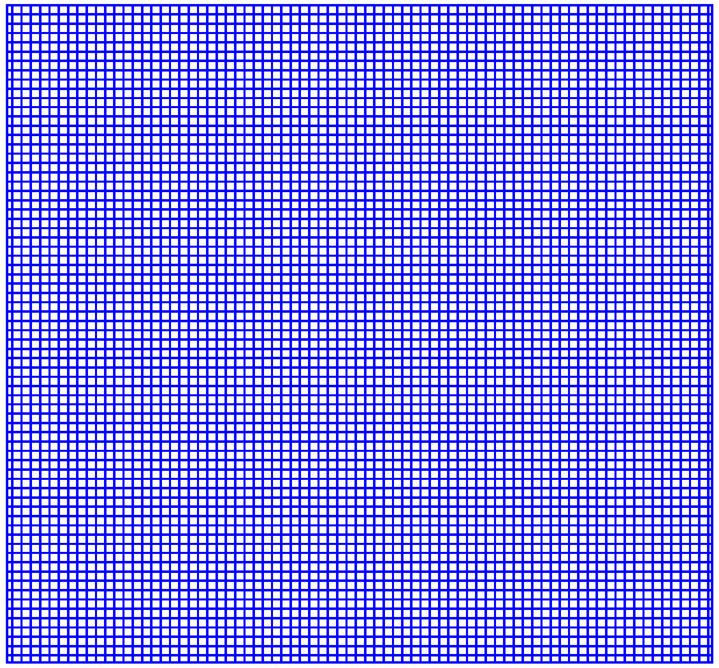
Source : Ron De Vore & Richard Baraniuk

Digital Data Acquisition

- Foundation: *Shannon sampling theorem*

  "if you sample densely enough (at the Nyquist rate), you can perfectly reconstruct the original data"





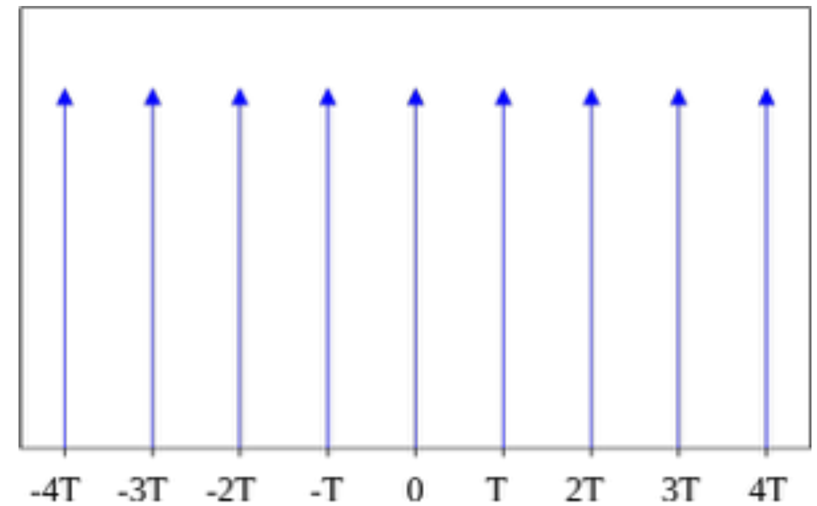
time

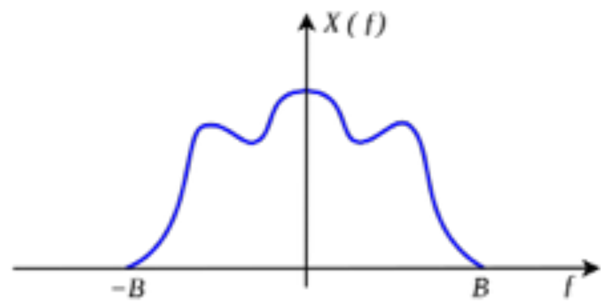
space

Source : Ron De Vore & Richard Baraniuk

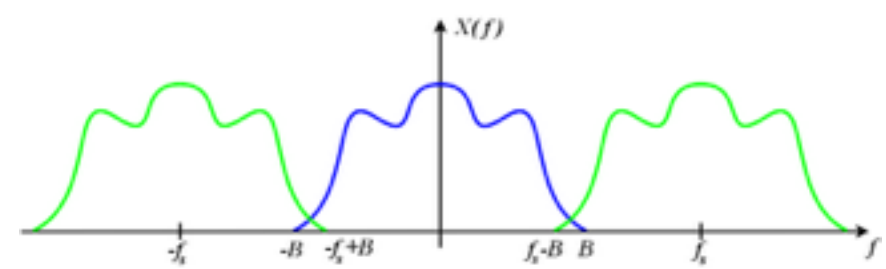
Nyquist–Shannon sampling theorem

$$s^*(t) = Ts(t) \cdot \delta_T(t)$$

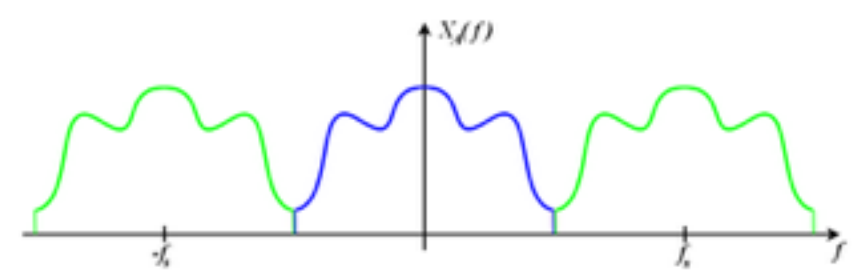




$$\hat{s}^*(f) = \hat{s}(f) * \sum_{n=-\infty}^{\infty} \delta(f - n/T)$$

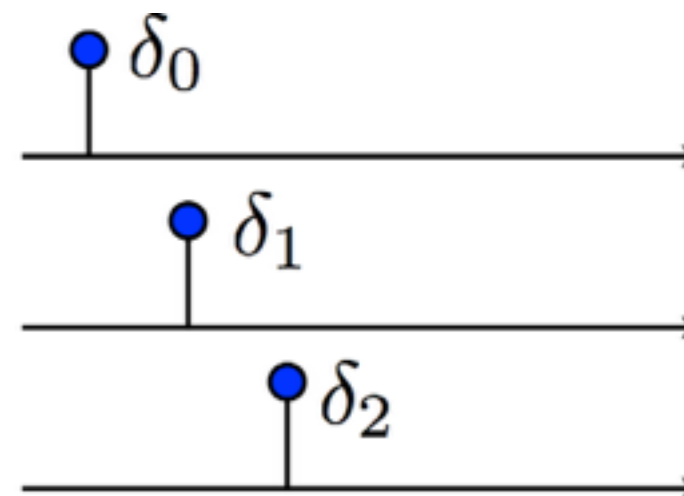


$$\hat{s}^*(f) = \sum_{n=-\infty}^{\infty} \hat{s}(f - n/T)\delta(f - n/T)$$



Sensing

Data aquisition: $f[i] = f(i/N) = \langle f, \delta_i \rangle$



Sensors $(\delta_i)_i$

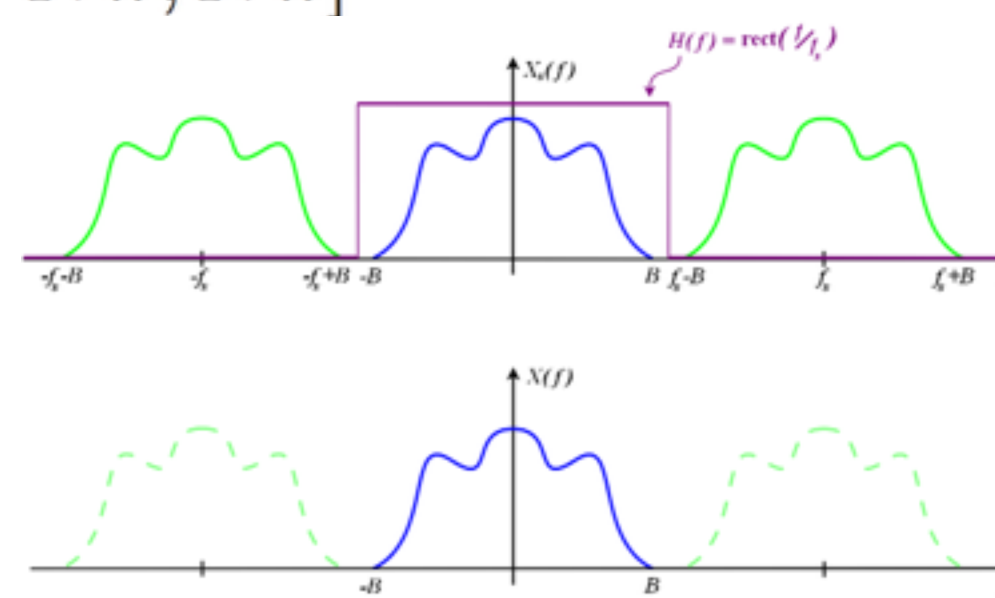
(Diracs)

$\tilde{f} \in \mathrm{L}^2$ $\qquad$ $f \in \mathbb{R}^N$

Shannon interpolation: if $\mathrm{Supp}(\hat{\tilde{f}}) \subset [-N\pi, N\pi]$

$$\tilde{f}(t) = \sum_i f[i] h(Nt - i)$$

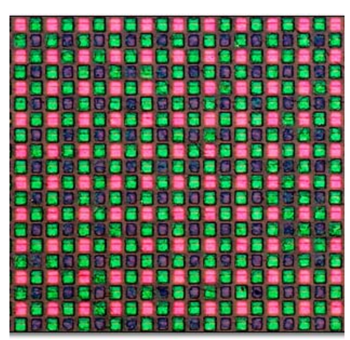
$$\text{where} \quad h(t) = \frac{\sin(\pi t)}{\pi t}$$



Source : Gabriel Peyré

Sensing by *Sampling*

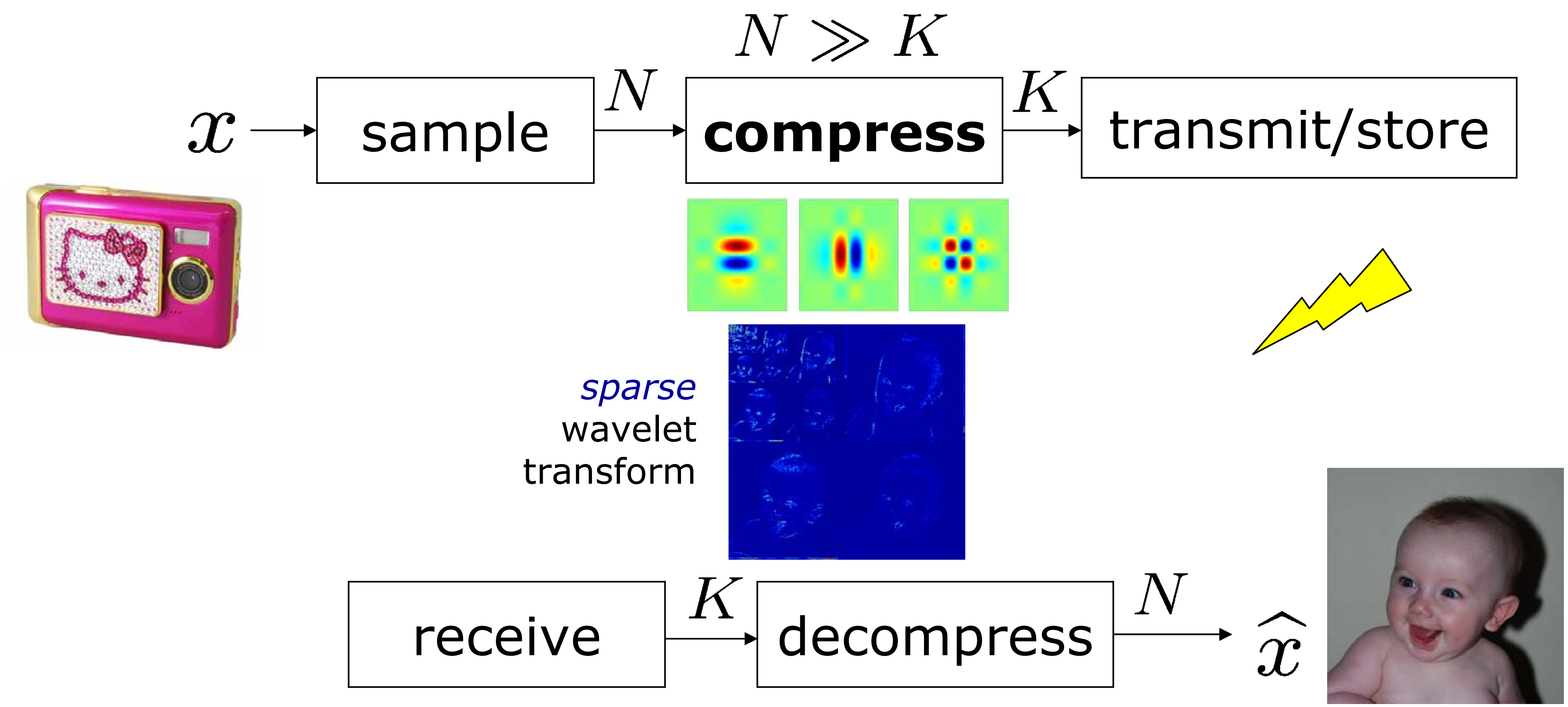
- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)



Source : Ron De Vore & Richard Baraniuk

Sensing by *Sampling*

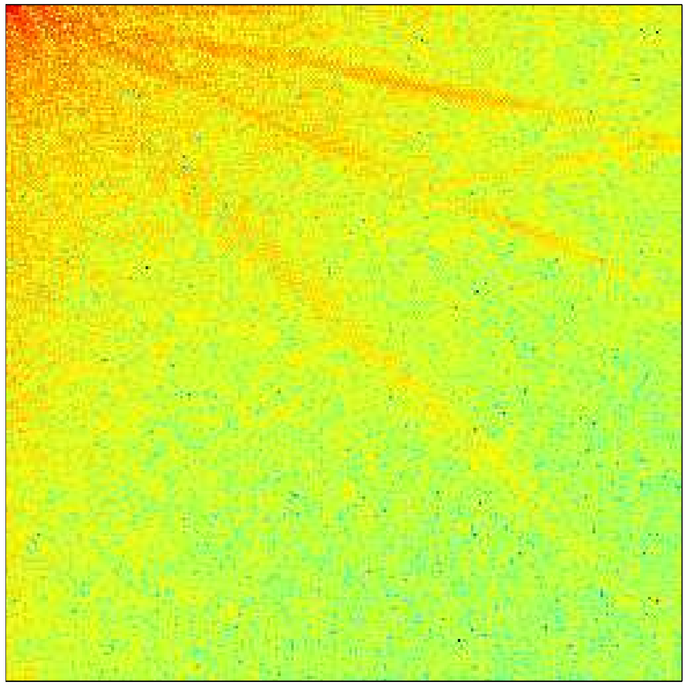
- Long-established paradigm for digital data acquisition
  - uniformly ***sample*** data at Nyquist rate (2x Fourier bandwidth)
  - ***compress*** data (signal-dependent, nonlinear)



Source : Ron De Vore & Richard Baraniuk

Classical Image Representation: DCT

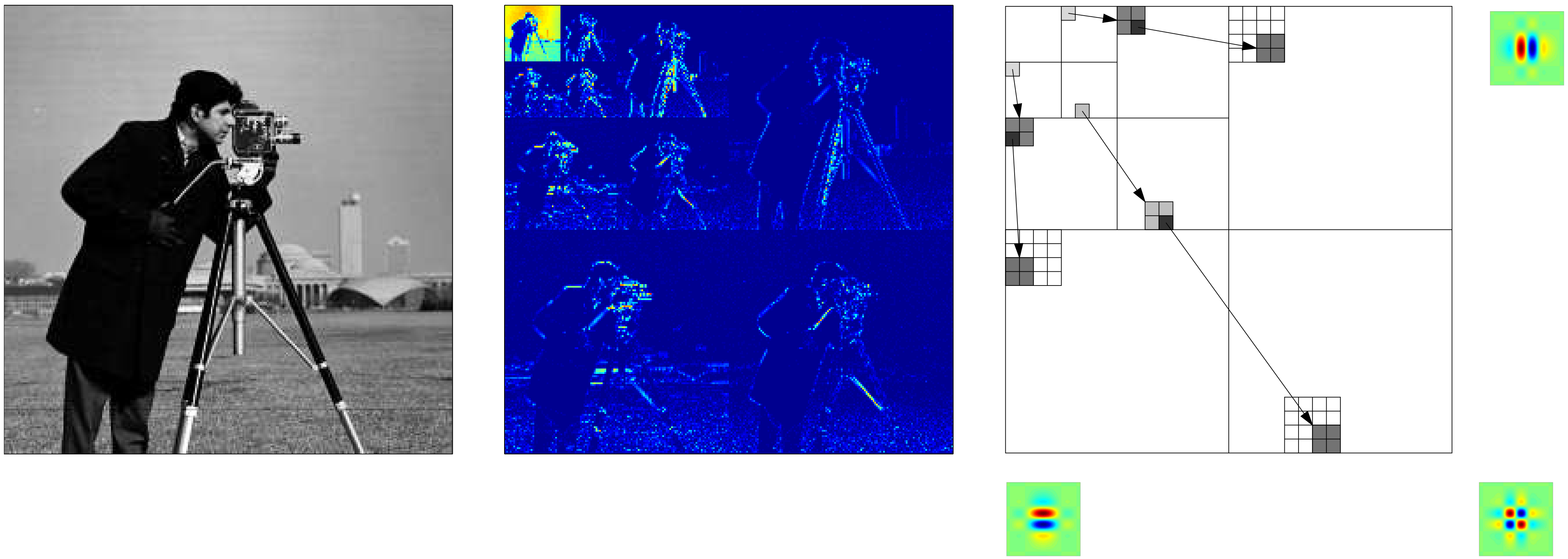
- Discrete Cosine Transform (DCT)
  Basically a real-valued Fourier transform (sinusoids)
- Model: most of the energy is at low frequencies



- Basis for JPEG image compression standard
- DCT approximations: smooth regions great, edges blurred/ringing

Source : Justin Romberg & Michael Wakin

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*

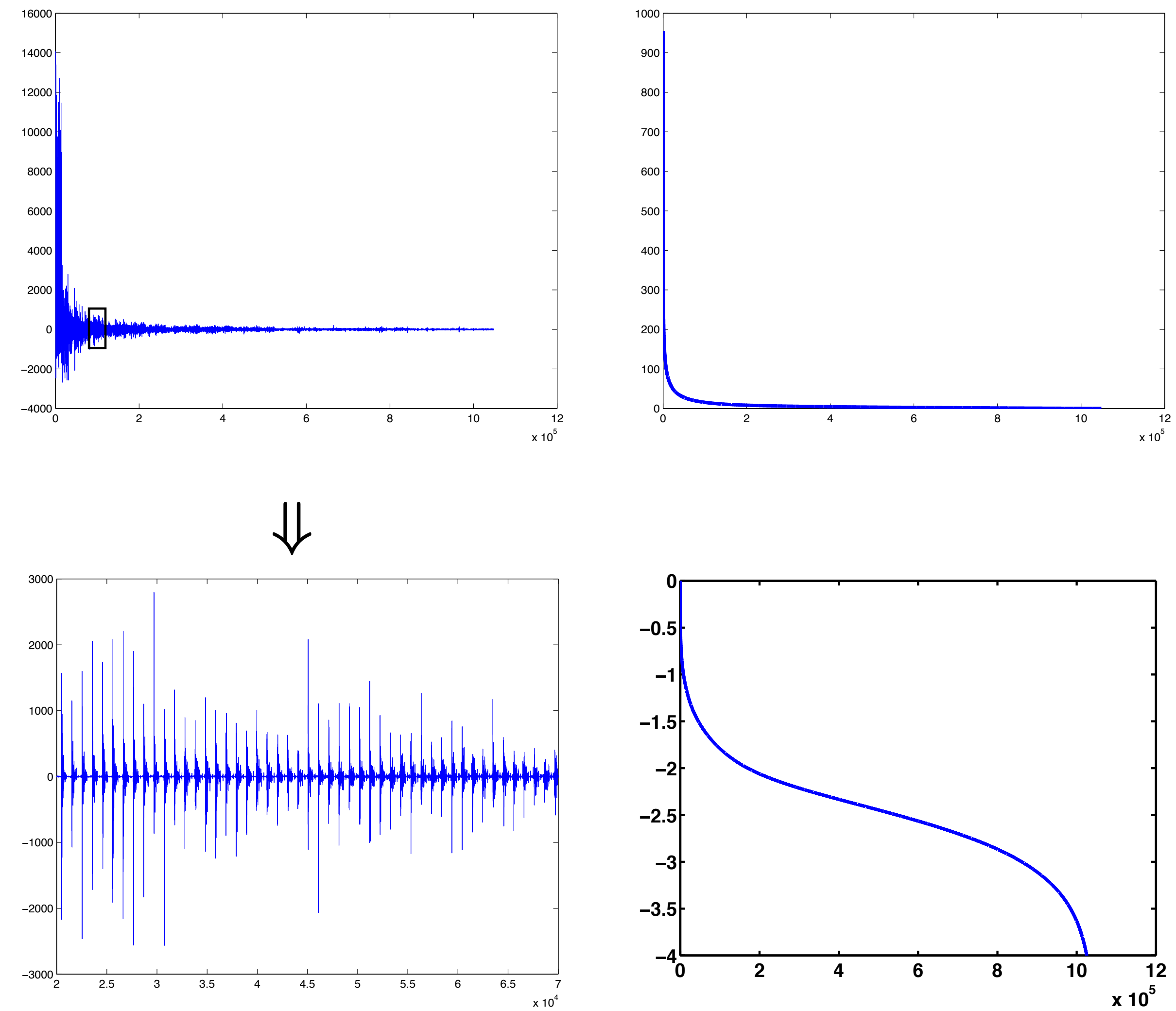
Source : Justin Romberg & Michael Wakin

Wavelets and Images

1 megapixel image

wavelet coeffs

(sorted)



zoom in

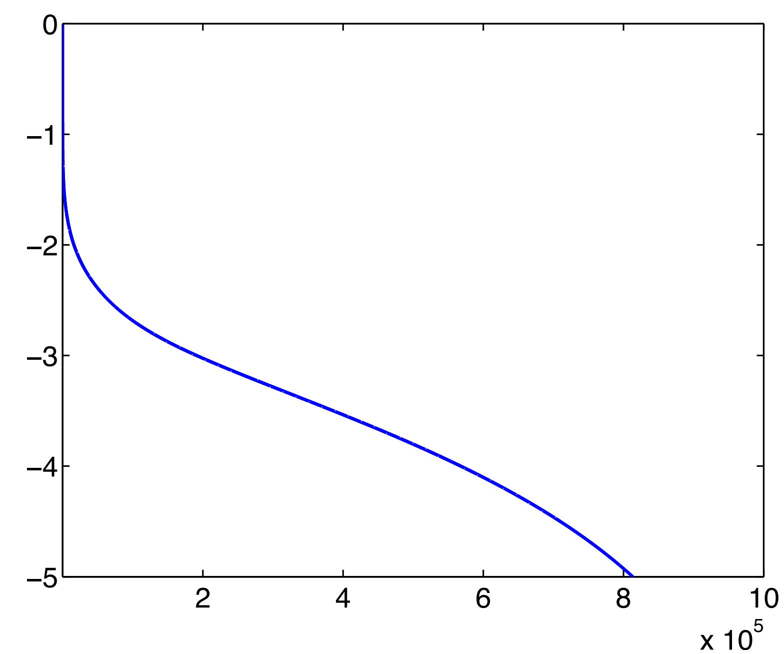
($\log_{10}$ sorted)

Source : Justin Romberg & Michael Wakin

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

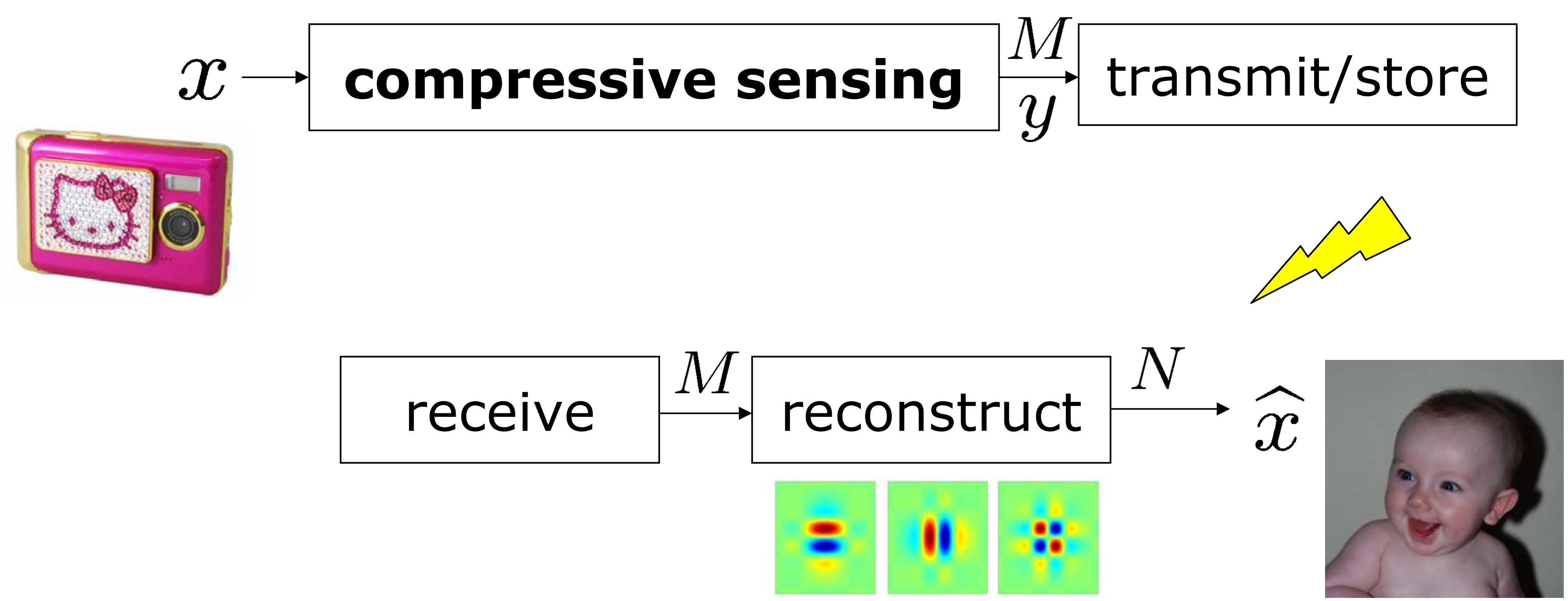
$$\|f - f_K\|_2 \;\approx\; .01 \cdot \|f\|_2$$

Source : Justin Romberg & Michael Wakin

Compressive Sensing

- Directly acquire "*compressed*" data
- Replace samples by more general "measurements"

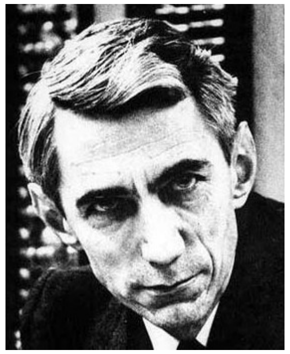
$$K < \underline{M} \ll N$$



Source : Ron De Vore & Richard Baraniuk

Compressive Sensing (CS)

- Recall Shannon/Nyquist theorem
  - Shannon was a *pessimist*
  - 2x oversampling Nyquist rate is a worst-case bound for *any* bandlimited data
  - sparsity/compressibility irrelevant
  - Shannon sampling is a linear process while compression is a nonlinear process

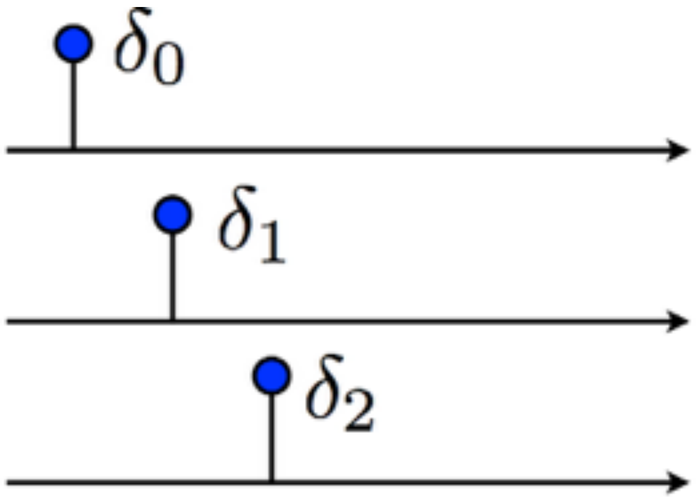


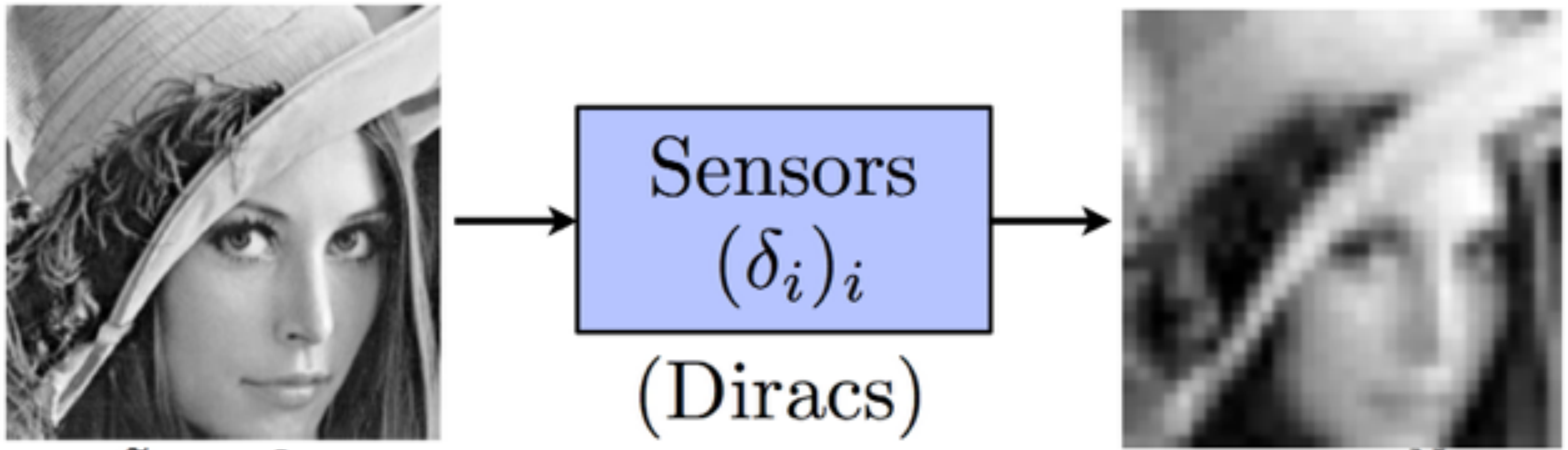
- **Compressive sensing**
  - new sampling theory that *leverages compressibility*
  - based on new *uncertainty principles*
  - *randomness* plays a key role

Source : Ron De Vore & Richard Baraniuk

Sensing

Data aquisition: $f[i] = f(i/N) = \langle f, \delta_i \rangle$





$\tilde{f} \in \mathrm{L}^2$ $\qquad$ $f \in \mathbb{R}^N$

Shannon interpolation: if $\mathrm{Supp}(\hat{\tilde{f}}) \subset [-N\pi, N\pi]$

$$\tilde{f}(t) = \sum_i f[i] h(Nt - i)$$

$$\text{where} \quad h(t) = \frac{\sin(\pi t)}{\pi t}$$

Source : Gabriel Peyré

Coded Acquisition

- Instead of pixels, take *linear measurements*

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

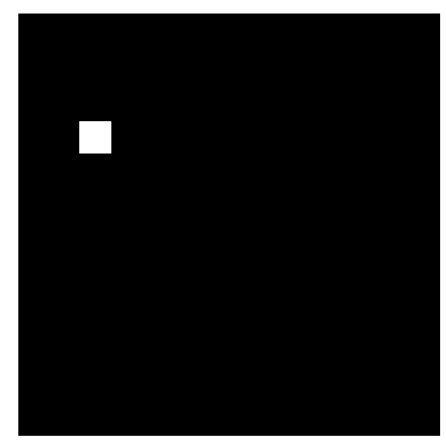
$$y = \Phi f$$

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

- Example: big pixels

$$y_m = \Big\langle$$

,



$$\Big\rangle$$

Source : Justin Romberg & Michael Wakin

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)

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

Coded Acquisition

- Instead of pixels, take *linear measurements*

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

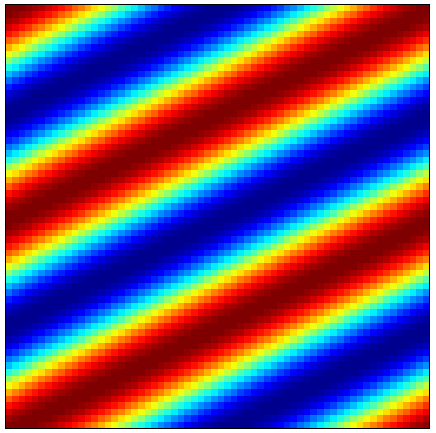
$$y = \Phi f$$

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

- Example: sinusoids (MRI)

$y_m = \Big\langle$

,



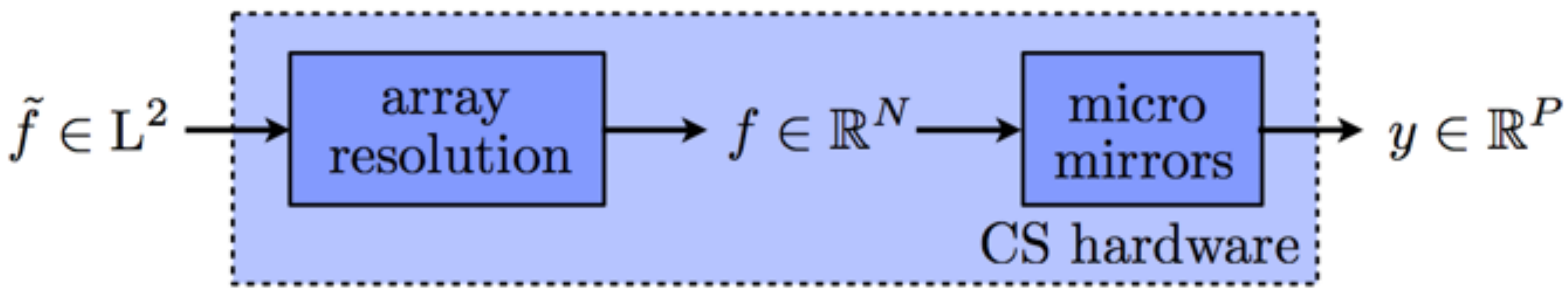
$\Big\rangle$

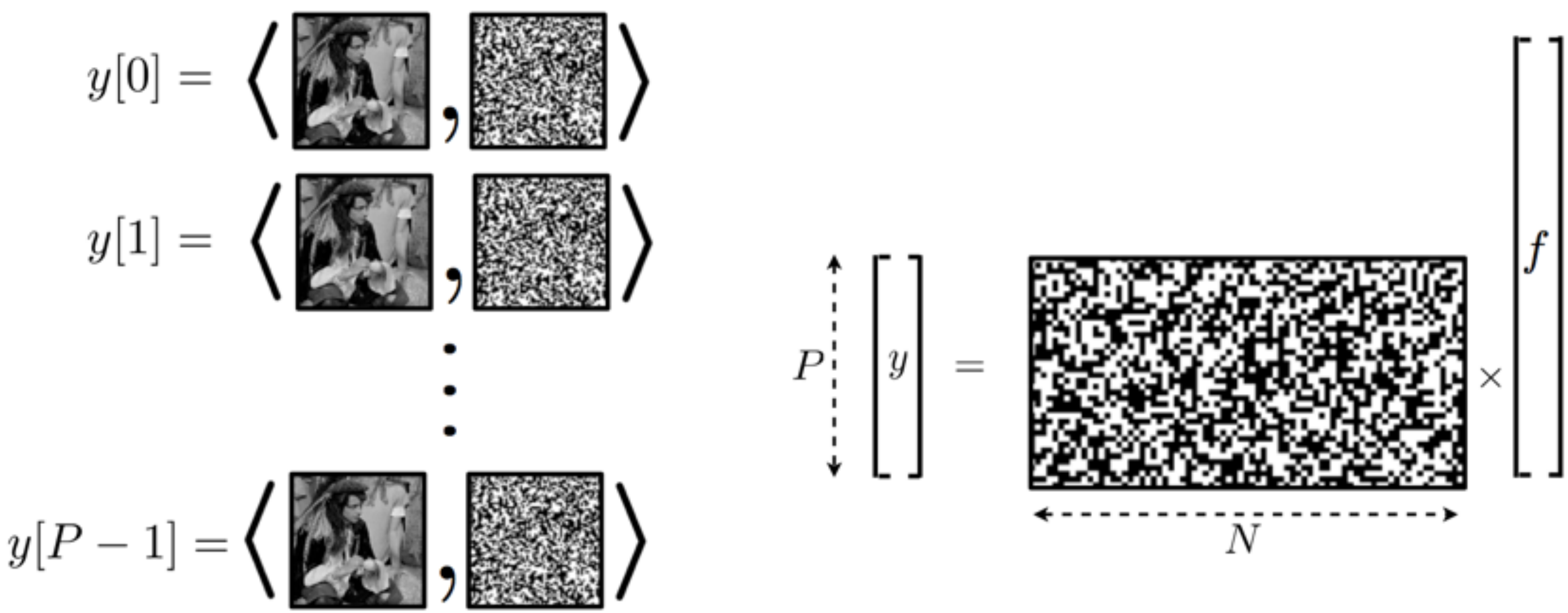
Source : Justin Romberg & Michael Wakin

Random sensing

CS is about designing hardware: input signals $\tilde{f} \in \mathrm{L}^2(\mathbb{R}^2)$.

Physical hardware resolution limit: target resolution $f \in \mathbb{R}^N$.





Source : Gabriel Peyré

Algebraic formulation

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

- Let define the sensing matrix as the following orthobasis

$$\Phi = \begin{pmatrix} \phi_1 & \phi_2 & \ldots & \phi_m \end{pmatrix}^T \in \mathcal{M}_{m \times n}$$

- The process of recovering $f \in \mathbb{R}^n$ from $y = \Phi f \in \mathbb{R}^m$ is ill-posed in general when $m < n$

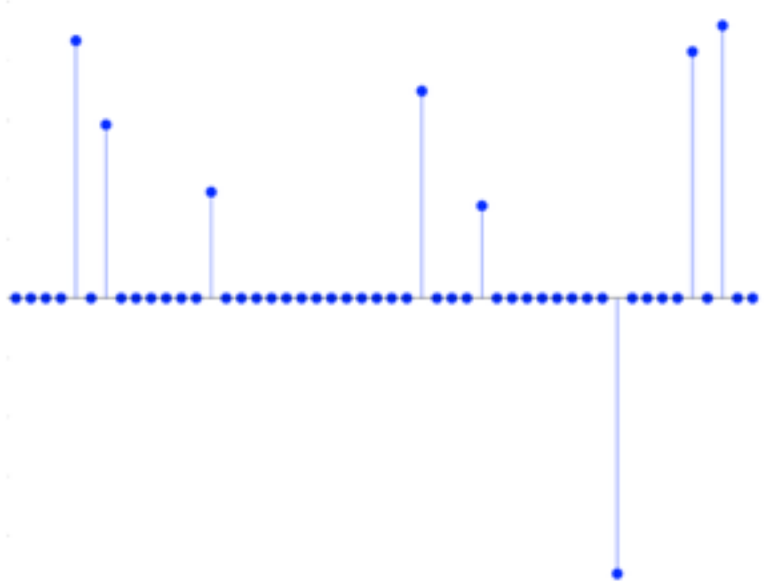
- But one can recover the object if it has a **sparse** representation in another basis functions Ψ which is orthogonal and **incoherent** with the basis Φ .

Sparsity

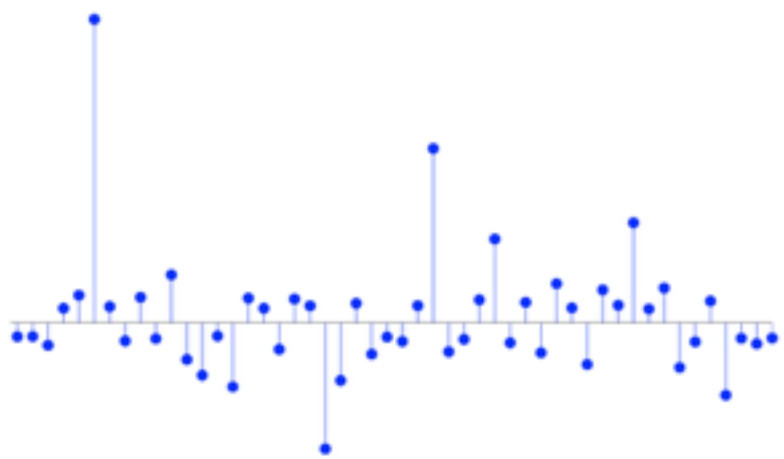
$f \in \mathbb{R}^n$ can be extended in a given basis $\Psi = [\psi_1 \psi_2 \cdots \psi_n]$

$$f(t) = \sum_{i=1}^{n} x_i \psi_i(t)$$

in which a small number of coefficients $x_i = < f, \psi_i >$ are nonzero elements



sparse x



nearly sparse x

Incoherence

Coherence between the sensing basis Φ and the representation basis Ψ is

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leqslant k,j \leqslant n} | < \phi_k, \psi_j > |$$

$$\mu(\Phi, \Psi) \in [1, \sqrt{n}] \text{ since } \forall j, \; \sum_{k=1}^{n} | < \phi_k, \psi_j > |^2 = \|\psi_j\|^2 = 1$$

Examples

$$\phi_k(t) = \delta(t - k) \text{ (spikes basis) and } \psi_j(t) = n^{-1/2} e^{-i2\pi jt/n} \text{ (Fourier basis)}$$

→ $\boxed{\mu(\Phi, \Psi) = 1}$ Maximal incoherence

- Φ = Noiselets, Ψ = Haar $\Rightarrow \mu(\Phi, \Psi) = \sqrt{2}$
- Φ = Noiselets, Ψ = Daubechies D4, D8 $\Rightarrow \mu(\Phi, \Psi) = 2.2, 2.9$
- Φ = Random matrix, Ψ = fixed basis $\Rightarrow \mathbb{E}[\mu(\Phi, \Psi)] = \sqrt{2 \log n}$

→ φ_k i.i.d $\mathcal{N}(0,1)$, ± 1, $\exp(i2\pi\omega_k t)$ ...

Main results

$$y_k =< f, \varphi_k > \forall k \in M \Leftrightarrow y = \Phi f = \Phi\Psi x \text{ with } x \text{ sparse}$$

Recovery

The reconstruction f^* is given by $f^* = \Psi x^*$ where x^* is solution of

$$\min_{\tilde{x}\in\mathbb{R}^n} \|\tilde{x}\|_{\ell_1} \quad \text{subject to} \quad y = \Phi\Psi x$$

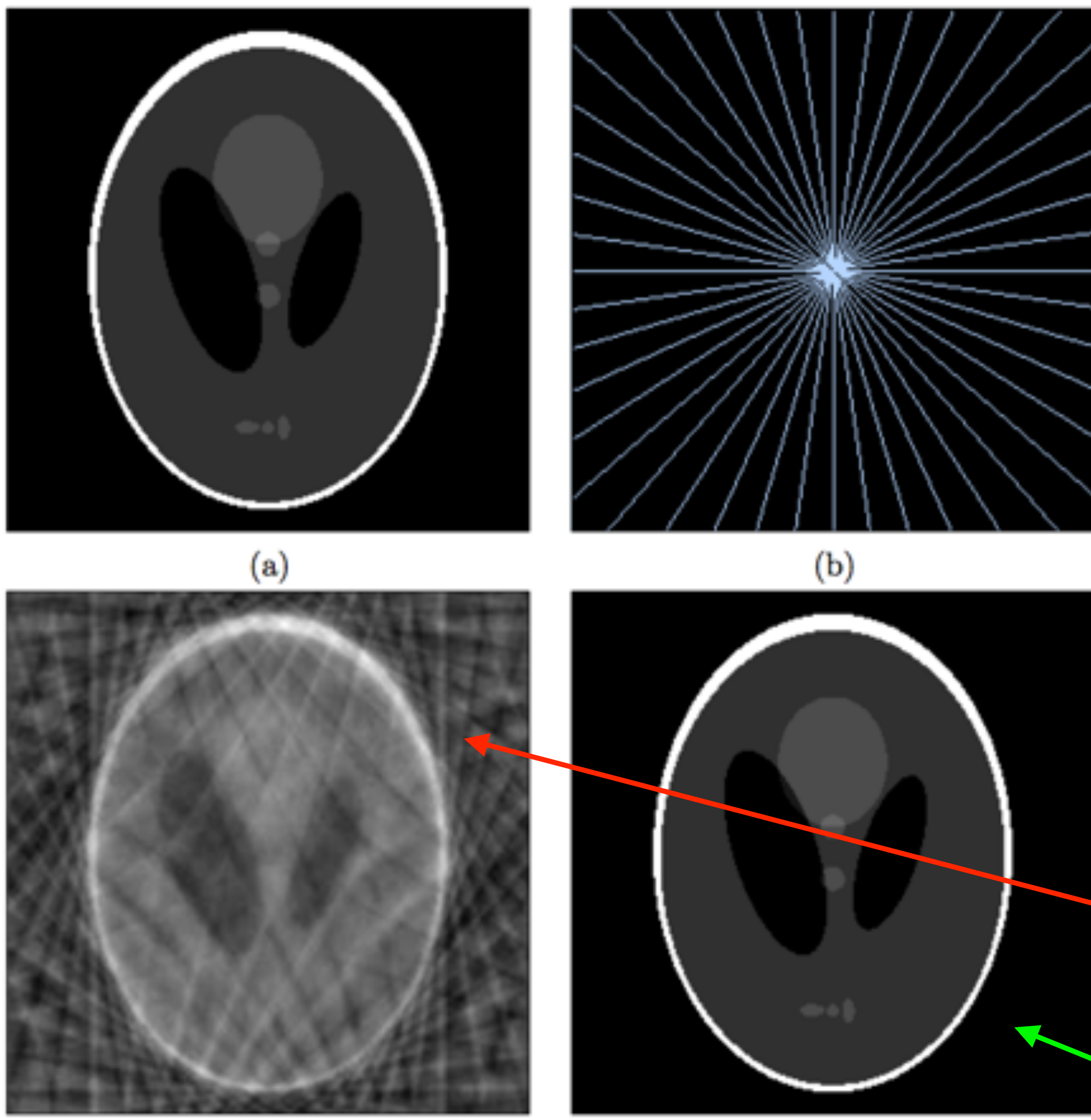
Théorème 1

Fix $f \in \mathbb{R}^n$ and suppose that the coefficient sequence x of f in the basis Ψ is $S-$sparse. Select m measurements in the Φ domain uniformly at random. Then if

$$m \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log(n/\delta)$$

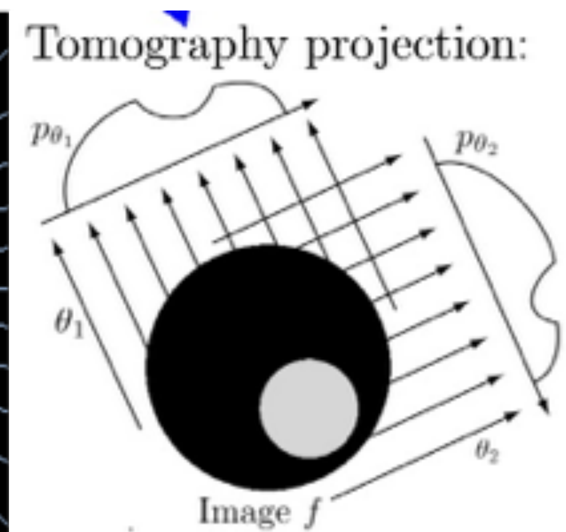
the solution of the convex optimization is exact with probability $1 - \delta$.

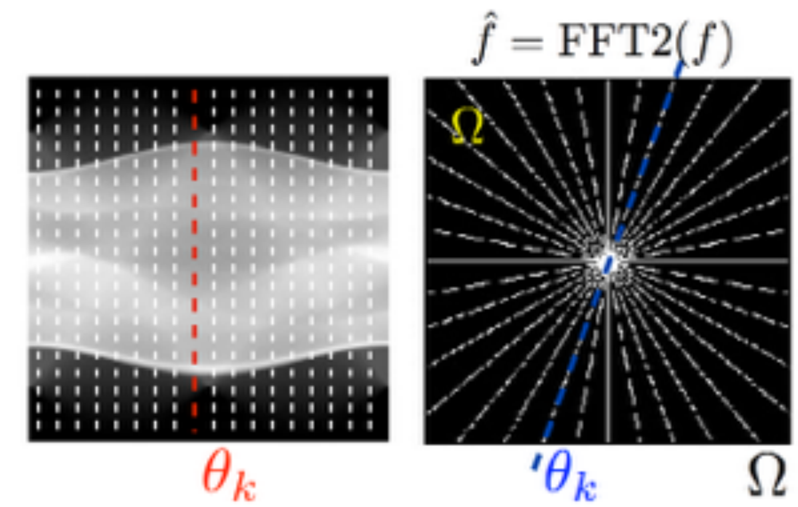
Motivations



(a) (b) (c) (d)

Figure 1: Example of a simple recovery problem. (a) The Logan-Shepp phantom test image. (b) Sampling 'domain' in the frequency plane; Fourier coefficients are sampled along 22 approximately radial lines. (c) Minimum energy reconstruction obtained by setting unobserved Fourier coefficients to zero. (d) Reconstruction obtained by minimizing the total-variation, as in (1.1). The reconstruction is an exact replica of the image in (a).





Fourier Slide Theorem : $\underbrace{\hat{p}_\theta(\rho)}_{\text{1D}} = \underbrace{\hat{f}(\rho\cos(\theta), \rho\sin(\theta))}_{\text{2D Fourier}}$

Partial measurements : $\{p_{\theta_k}(t)\}_{0 \leqslant k < m}$

Equivalent to : $\Phi f = \{\hat{f}[\omega]\}_{\omega \in \Omega}$

Filtred backprojection

VS

Convex minimization

$$\min \|g\|_{BV} \quad \text{subject to} \quad \hat{g}(\omega) = \hat{f}(\omega) \text{ for all } \omega \in \Omega$$

Key points

Hypothesis $\hat{f}(k) = \sum_{t=0}^{N-1} f(t) e^{-i\omega_k t}, \quad \omega_k = \frac{2\pi k}{N}, k = 0, 1, \ldots, N-1$

- Suppose we are only given $\hat{f}|_\Omega$ sampled in some partial subset $\Omega \subset \mathbb{Z}_N$
- Suppose f is supported on a small subset $S \subset \mathbb{Z}_N$: $f = \sum_{t \in S} \alpha_t \delta_t$
- They proved that f can be reconstructed from $\hat{f}|_\Omega$ if $|S| \leqslant |\Omega|/2$. (N prime)
- In principle, we can recover f exactly by solving the optimization

$$(P_0) \quad \min \|g\|_{\ell_0}, \quad \hat{g}|_\Omega = \hat{f}|_\Omega$$

Combinatorial problem → for $|\Omega| \sim N/2 \Rightarrow 4^N \cdot 3^{-3N/4}$ subsets to check !

- Instead one can solved the convex problem

$$(P_1) \quad \min \|g\|_{\ell_1} := \sum_{t \in \mathbb{Z}_N} |g(t)|, \quad \hat{g}|_\Omega = \hat{f}|_\Omega$$

→ (P_0) and (P_1) are equivalent for an overwhelming percentage of the choices for S and Ω with $|S| \leqslant C \cdot |\Omega| / \log N$

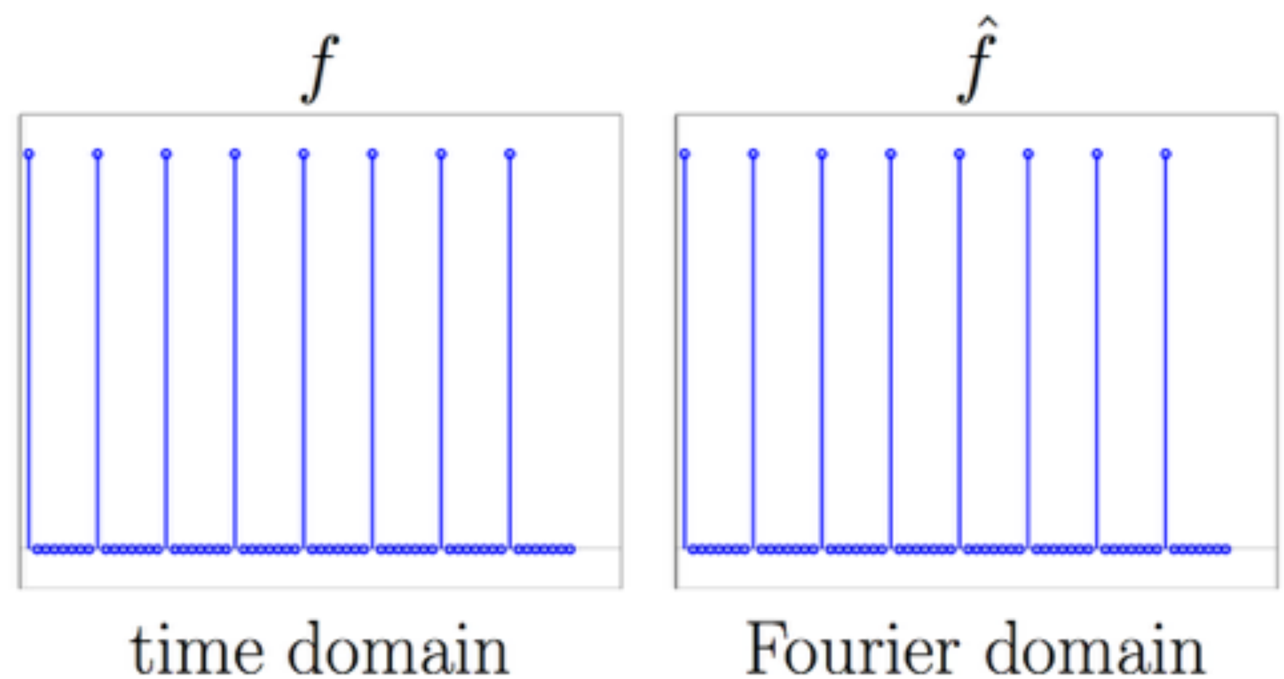
For almost every Ω

- They proved that f can be reconstructed from $\hat{f}|_\Omega$ if $|S| \leqslant |\Omega|/2$. (N prime)
  - for N prime, $\mathcal{F}_{S\to\Omega} f := \hat{f}|_\Omega$ for all $f \in \ell_2(S)$ is injective when $|S| \leqslant |\Omega|$
  - hold for non-prime if S, Ω are not subgroups of $\mathbb{Z}_N$
  - Ω^c must not content a large interval (mostly the case when chosen **randomly**)

- There exist sets Ω and functions f for which the ℓ_1-minimization procedure does not recover f correctly, even if $|\mathrm{supp}(f)|$ is much smaller than $|\Omega|$.

- Dirac's comb

- $\sqrt{N}$ spikes spaced $\sqrt{N}$ apart
- Invariant under Fourier transform ($f = \hat{f}$)
- $|S| + |\Omega| = 2\sqrt{N}$



Measurements : Ω^* all frequencies but the multiples of $\sqrt{N}$, namely $|\Omega^*|$

$\hat{f}|_{\Omega^*} = 0$ → Reconstruction is identically zero

For almost every Ω

- They proved that f can be reconstructed from $\hat{f}|_\Omega$ if $|S| \leqslant |\Omega|/2$. (N prime)
  - for N prime, $\mathcal{F}_{S\to\Omega} f := \hat{f}|_\Omega$ for all $f \in \ell_2(S)$ is injective when $|S| \leqslant |\Omega|$
  - hold for non-prime if S, Ω are not subgroups of $\mathbb{Z}_N$
  - Ω^c must not content a large interval (mostly the case when chosen **randomly**)

■ There exist sets Ω and functions f for which the ℓ_1-minimization procedure does not recover f correctly, even if $|\mathrm{supp}(f)|$ is much smaller than $|\Omega|$.

- Box signals
  - sample size N large
  - $f = \chi_T$ where $T = \{t : -N^{-0.01} < t < N^{0.01}\}$
  - $\Omega = \{k : N/3 < k < 2N/3\}$
  - h a function whose Fourier transform $\hat{h}$ is a non-negative bump function on the interval $\{k : -N/6 < k < N/6\}$ which equals 1 when $-N/12 < k < N/12$

Fourier transform of $|h(t)|^2$ vanishes in Ω

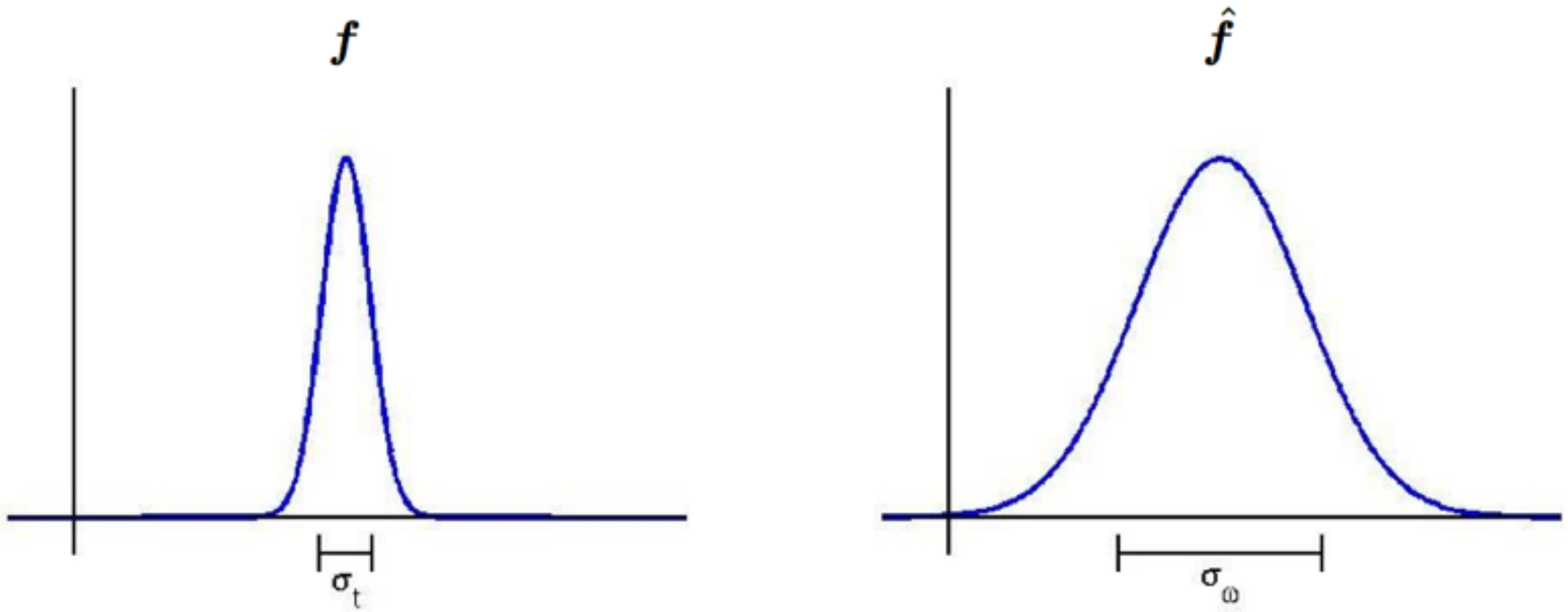
$|h(t)|^2$ rapidly decreases away from $t = 0$: $|h(t)|^2 = O(N^{-100})$ for $t \notin T$

$|h(0)|^2 > c$ for some absolute constant $c > 0$

→ $\mathcal{F}(f - \epsilon|h|^2) = \mathcal{F}(f)$ in Ω and $\|f - \epsilon|h|^2\|_{\ell_1} < \|f\|_{\ell_1}$

Uncertainty Principles

- Heisenberg (1927)
  Uncertainty principle for continuous-time signals

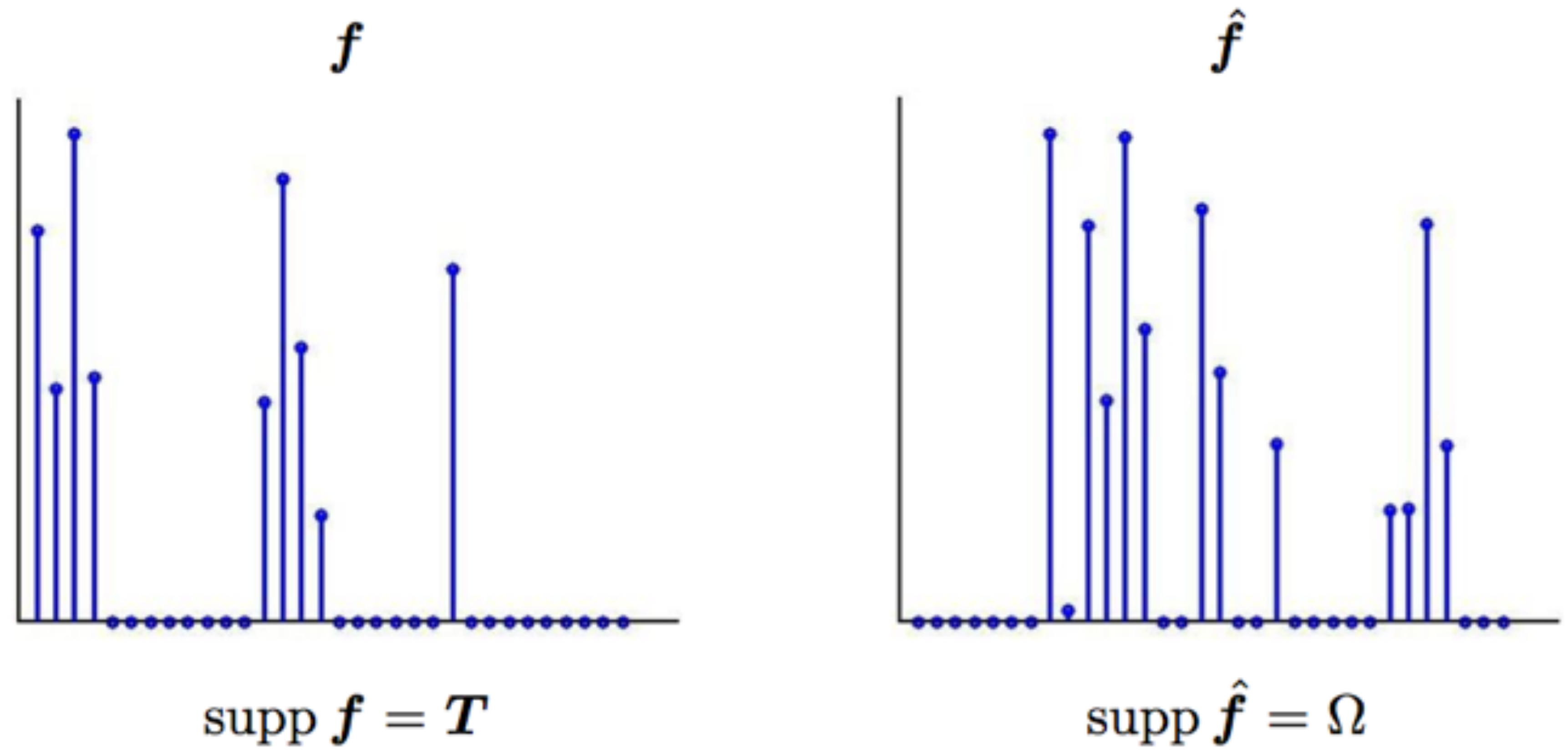


$$\sigma_t \sigma_\omega \geq \frac{1}{4\pi}$$

- Limits *joint resolution* in time and frequency

Uncertainty Principles

- Donoho and Stark (1989)
  Discrete uncertainty principle for $\mathbb{C}^N$



$$|T| + |\Omega| \geq 2\sqrt{N}$$

- Implications: recovery from partial information, unique sparse decompositions
- Generalization to pairs of bases B_1, B_2
  [Donoho,Huo,Elad,Bruckstein,Gribonval,Nielsen]

Relation to the uncertainty principle

Classical arguments show that f is the unique minimizer of (P_1) iff :

$$\sum_{t\in\mathbb{Z}_N} |f(t)+h(t)| > \sum_{t\in\mathbb{Z}_N} |f(t)|, \quad \forall h \neq 0,\ \hat{h}|_\Omega = 0$$

Put $T = \operatorname{supp}(f)$ and apply the triangle inequality

$$\sum_{\mathbb{Z}_N} |f(t)+h(t)| = \sum_{T} |f(t)+h(t)| + \sum_{T^c} |h(t)| \geq \sum_{T} |f(t)| - |h(t)| + \sum_{T^c} |h(t)|.$$

Hence, a sufficient condition to establish that f is our unique solution would be to show

$$\sum_{T} |h(t)| < \sum_{T^c} |h(t)| \qquad \forall h \neq 0,\ \hat{h}|_\Omega = 0.$$

or equivalently $\sum_T |h(t)| < \frac{1}{2}\|h\|_{\ell_1}$. The connection with the uncertainty principle is now explicit; f is the unique minimizer if it is impossible to 'concentrate' half of the ℓ_1 norm of a signal that is missing frequency components in Ω on a 'small' set T.

Robust uncertainty principle

Underlying our analysis is a new notion of uncertainty principle which holds for almost any pair $(\text{supp}(f), \text{supp}(\hat{f}))$. With $T = \text{supp}(f)$ and $\Omega = \text{supp}(\hat{f})$, the classical discrete uncertainty principle [6] says that

$$|T| + |\Omega| \geq 2\sqrt{N}. \tag{1.11}$$

with equality obtained for signals such as the Dirac's comb. As we mentioned above, such extremal signals correspond to very special pairs (T, Ω). However, for most choices of T and Ω, the analysis presented in this paper shows that it is *impossible* to find f such that $T = \text{supp}(f)$ and $\Omega = \text{supp}(\hat{f})$ unless

$$|T| + |\Omega| \geq \gamma(M) \cdot (\log N)^{-1/2} \cdot N, \tag{1.12}$$

which is considerably stronger than (1.11). Here, the statement 'most pairs' says again that the probability of selecting a random pair (T, Ω) violating (1.12) is at most $O(N^{-M})$.

Strategy for proving

(P_0) and (P_1) are equivalent for an overwhelming percentage of the choices for S and Ω with $|S| \leqslant C \cdot |\Omega| / \log N$

- Reformulation with duality theory

Linear program (P1) $\min_{\substack{g^+, g^- \in \mathbb{R}^N \\ g^+, g^- \geq 0}} \sum_{t=0}^{N-1} (g^+(t) + g^-(t)), \quad \mathcal{F}_\Omega(g^+ - g^-) = \hat{f}|_\Omega$ Lagrangien

$$L(g^+, g^-; \lambda, \mu^+, \mu^-) = \sum_{t=0}^{N-1} (g^+(t) + g^-(t)) + \lambda^H (\hat{f}|_\Omega - \mathcal{F}_\Omega(g^+ - g^-)) + \mu^{+*} g^+ + (\mu^-)^* g^-$$

$$\begin{aligned}
\mathcal{F}_\Omega(\tilde{g}^+ - \tilde{g}^-) &= \hat{f}|_\Omega \\
(\mu^+)^* \tilde{g}^+ &= 0 \\
(\mu^-)^* \tilde{g}^- &= 0 \\
\frac{\partial L}{\partial \tilde{g}^+(t)} = I_{\{\tilde{g}^+(t) > 0\}} - \mathcal{F}_\Omega^* \lambda + \mu^+ &= 0 \\
\frac{\partial L}{\partial \tilde{g}_t^-} = I_{\{\tilde{g}^-(t) > 0\}} + \mathcal{F}_\Omega^* \lambda + \mu^- &= 0.
\end{aligned}$$

$$\begin{aligned}
(\mathcal{F}_\Omega^* \lambda)(t) &= \operatorname{sgn}(f)(t) & t \in S \\
1 - (\mathcal{F}_\Omega^* \lambda)(t) - \mu^+ &= 0 & t \in S^c \\
1 + (\mathcal{F}_\Omega^* \lambda)(t) - \mu^- &= 0 & t \in S^c
\end{aligned}$$

Thus, to show that $f^\sharp$ is unique and is equal to f, it suffices to find a trigonometric polynomial P whose Fourier transform is supported in Ω—in other words, which only uses frequencies in Ω—and which matches $\operatorname{sgn}(f)$ on $\operatorname{supp}(f)$, and has magnitude strictly less than 1 elsewhere.

$$\begin{aligned}
P(t) &= \operatorname{sgn}(f)(t) & t \in S \\
|P(t)| &< 1 & t \notin S
\end{aligned}$$

Construction of the polynom P

With $|\Omega| > |T|$, and if $\mathcal{F}_{T\to\Omega}$ is injective, we construct P as follows

$$P := \mathcal{F}_{\Omega}^{*}\mathcal{F}_{T\to\Omega}(\mathcal{F}_{T\to\Omega^{*}}\mathcal{F}_{T\to\Omega})^{-1}\iota^{*}\mathrm{sgn}(f), \tag{3.4}$$

where $\mathcal{F}_{\Omega} = \mathcal{F}_{\mathbb{Z}_N\to\Omega}$ is the Fourier transform followed by a restriction to the set Ω; the the embedding operator $\iota : \ell_2(T) \to \ell_2(\mathbb{Z}_N)$ extends a vector on T to a vector on $\mathbb{Z}_N$ by placing zeros outside of T; and ι^{*} is the dual restriction map $\iota^{*}f = f|_T$

Equivalent representation using matrix

- For a signal $f \in \mathbf{C}^N$, the discrete Fourier transform $\mathcal{F}f = \hat{f} : \mathbf{C}^N \to \mathbf{C}^N$ is defined as

$$\hat{f}(\omega) := \sum_{t=0}^{N-1} f(t)\, e^{-2\pi i \omega t/N}, \quad \omega = 0, 1, \ldots, N-1. \tag{3.5}$$

- The discrete Fourier transform can also be represented using matrix form:

$$\hat{f} = \mathcal{F}f,$$

where $\mathcal{F} \in \mathbf{C}^{\mathcal{N}\times\mathcal{N}}$ and $\mathcal{F}(j,p) = e^{-2\pi i (j-1)(p-1)/N}, 1 \leq j, p \leq N$.

- Assume we permute the rows/columns of F such that the first $|T|$ columns of $\mathcal{F}$ correspond to the set T, and the first $|\Omega|$ rows of $\mathcal{F}$ correspond to the set Ω. Note that $\mathcal{F}$ is not symmetric any more.

- Also we partition $\mathcal{F}$ in the following form after permutation:

$$\mathcal{F} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix},$$

where $F_{11} \in \mathbf{C}^{|\Omega|\times|T|}$, $F_{12} \in \mathbf{C}^{|\Omega|\times|T^c|}$, $F_{21} \in \mathbf{C}^{|\Omega^c|\times|T|}$, $F_{22} \in \mathbf{C}^{|\Omega^c|\times|T^c|}$, and $T^c = \mathbb{Z}_N - T, \Omega^c = \mathbb{Z}_N - \Omega$.

Equivalent representation using matrix

- The discrete Fourier transform can be represented as

$$\begin{bmatrix} \hat{f}|_\Omega \\ \hat{f}|_{\Omega^c} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} f|_T \\ f|_{T^c} \end{bmatrix}$$

- Thus we have

$$\mathcal{F}_{T\to\Omega} = F_{11}, \mathcal{F}_\Omega = [F_{11}, F_{12}].$$

- The operators ι and ι^* can also be represented in the matrix form

$$\iota = \begin{bmatrix} I_{|T|} \\ 0 \end{bmatrix} \in \mathbf{R}^{|T|\times|\Omega|}, \iota^* = \begin{bmatrix} I_{|T|} & 0 \end{bmatrix} \in \mathbf{R}^{|\Omega|\times|T|}.$$

- Then P can be represented as

$$P = \begin{bmatrix} F_{11}^* \\ F_{12}^* \end{bmatrix} F_{11}(F_{11}^* F_{11})^{-1} \begin{bmatrix} I_{|T|} & 0 \end{bmatrix} \operatorname{sgn}(f).$$

Equivalent representation using matrix

- Note that f is supported on T, thus we have

$$\mathrm{sgn}(f) = \begin{bmatrix} \mathrm{sgn}(f|_T) \\ 0 \end{bmatrix}$$

- Thus P can be simplified:

$$\begin{aligned} P &= \begin{bmatrix} F_{11}^* \\ F_{12}^* \end{bmatrix} F_{11}(F_{11}^* F_{11})^{-1} \begin{bmatrix} I_{|T|} & 0 \end{bmatrix} \mathrm{sgn}(f) \\ &= \begin{bmatrix} I_{|T|} \\ F_{12}^* F_{11} F_{11}^* F_{11})^{-1} \end{bmatrix} \mathrm{sgn}(f|_{|T|}) \end{aligned}$$

- Clearly, we have

$$\iota^* P = \iota^* \mathrm{sgn}(f) = \mathrm{sgn}(f|_{|T|}).$$

Main ideas of the proof

- Fixing f and its support T, we will prove Theorem 1.3 by establishing that if the set Ω is chosen uniformly at random from all sets of size $N_\omega \geq C_M^{-1}|T|\log T$, then we can prove
  - **Invertibility**. The operator $\mathcal{F}_{T\to\Omega}$ is injective, i.e., the matrix $F_{11}^* F_{11}$ is invertible, with probability $1 - O(N^{-M})$.
  - **Magnitude**. The function P obeys $|P(t)| < 1$ for all $t \in T^c$ with probability $1 - O(N^{-M})$.
- Then we can apply Lemma 3.1, and obtain Theorem 1.3 directly.
- A difficulty is how to make use of the argument that Ω of certain size is chosen uniformly at random.
- In this paper, we use Bernoulli probability model for selecting the set Ω, and show how to convert this model to the uniform probability model.

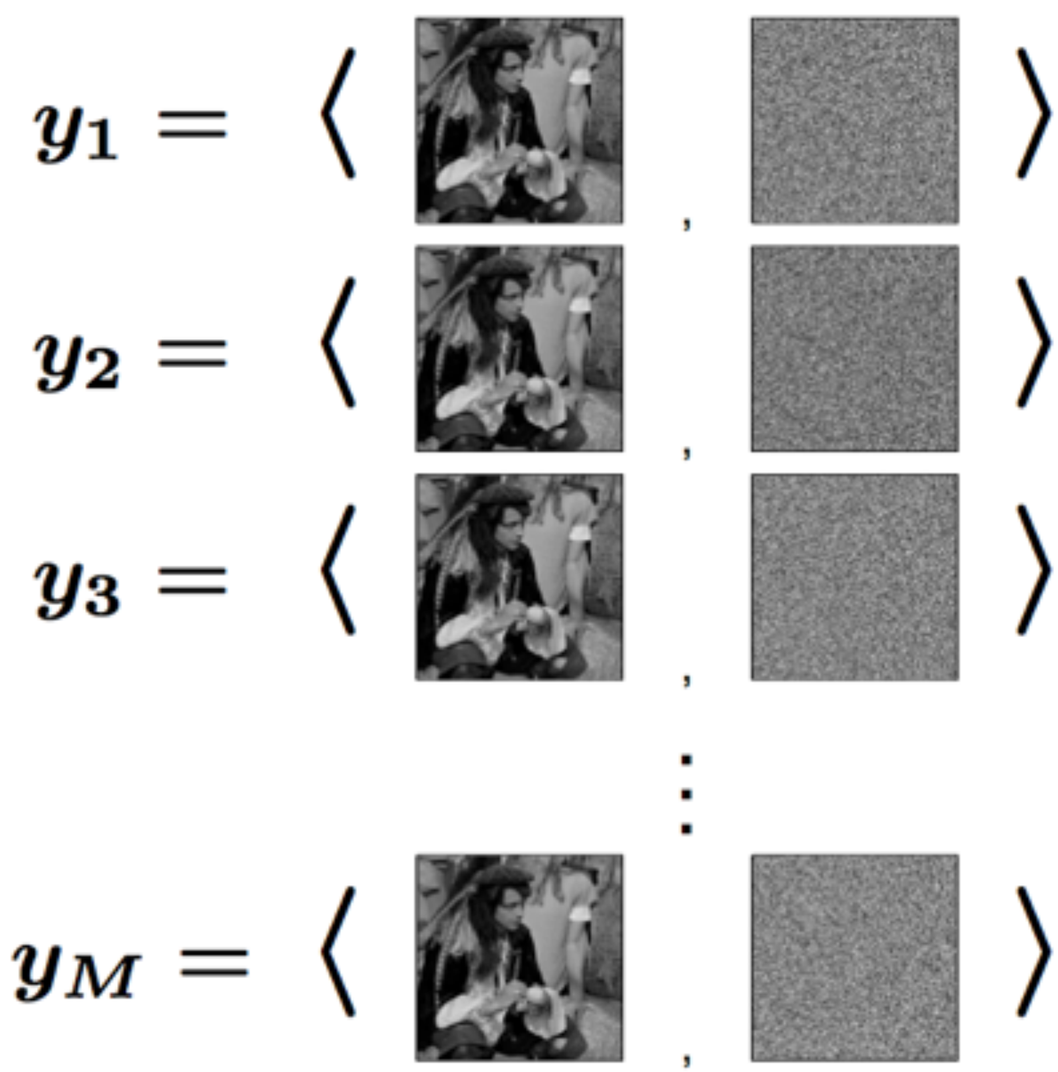
Bernoulli Model

- A set Ω' of Fourier coefficients is sampled using the Bernoulli model with parameter $0 < \tau < 1$ by first creating the sequence

$$I_\omega = \begin{cases} 0 & \text{with prob.} \quad 1-\tau, \\ 1 & \text{with prob.} \quad \tau. \end{cases} \tag{3.6}$$

- Note that the size of Ω' is random, and $\mathbf{E}(\Omega') = \tau N$.
- We will show that the "Invertibility" and "Magnitude" hold with a high probability for the Bernoulli model.

L1 reconstruction of a sparse image



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

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

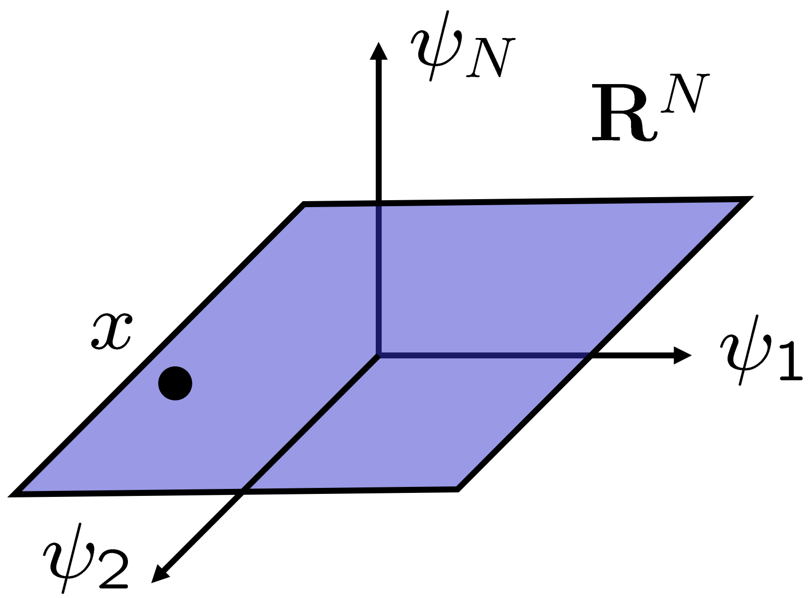
Ψ = wavelet transform

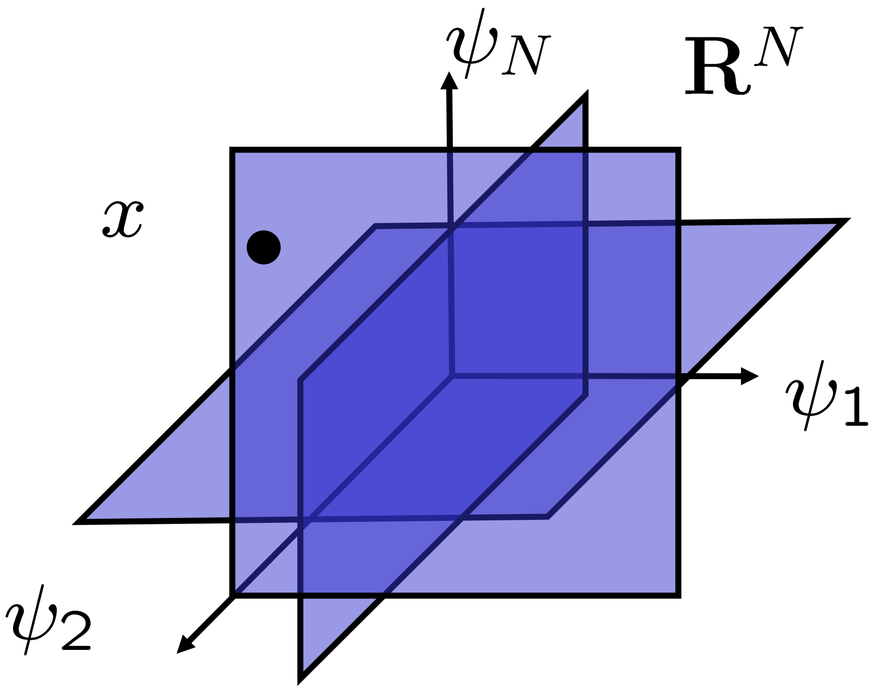
original (25k wavelets)

perfect recovery

Source : Justin Romberg & Michael Wakin

Geometry of Sparse Signal Sets





Linear

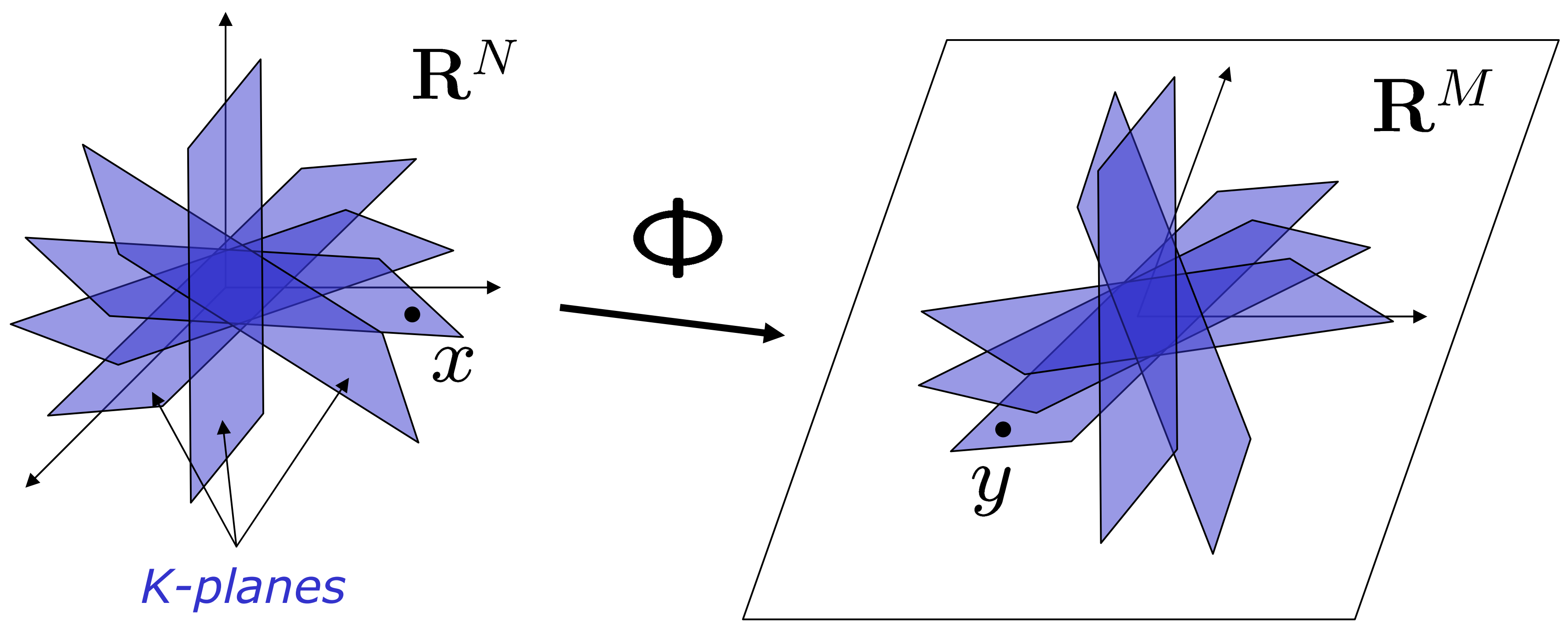
K-plane

Sparse, Nonlinear

Union of K-planes

Source : Justin Romberg & Michael Wakin

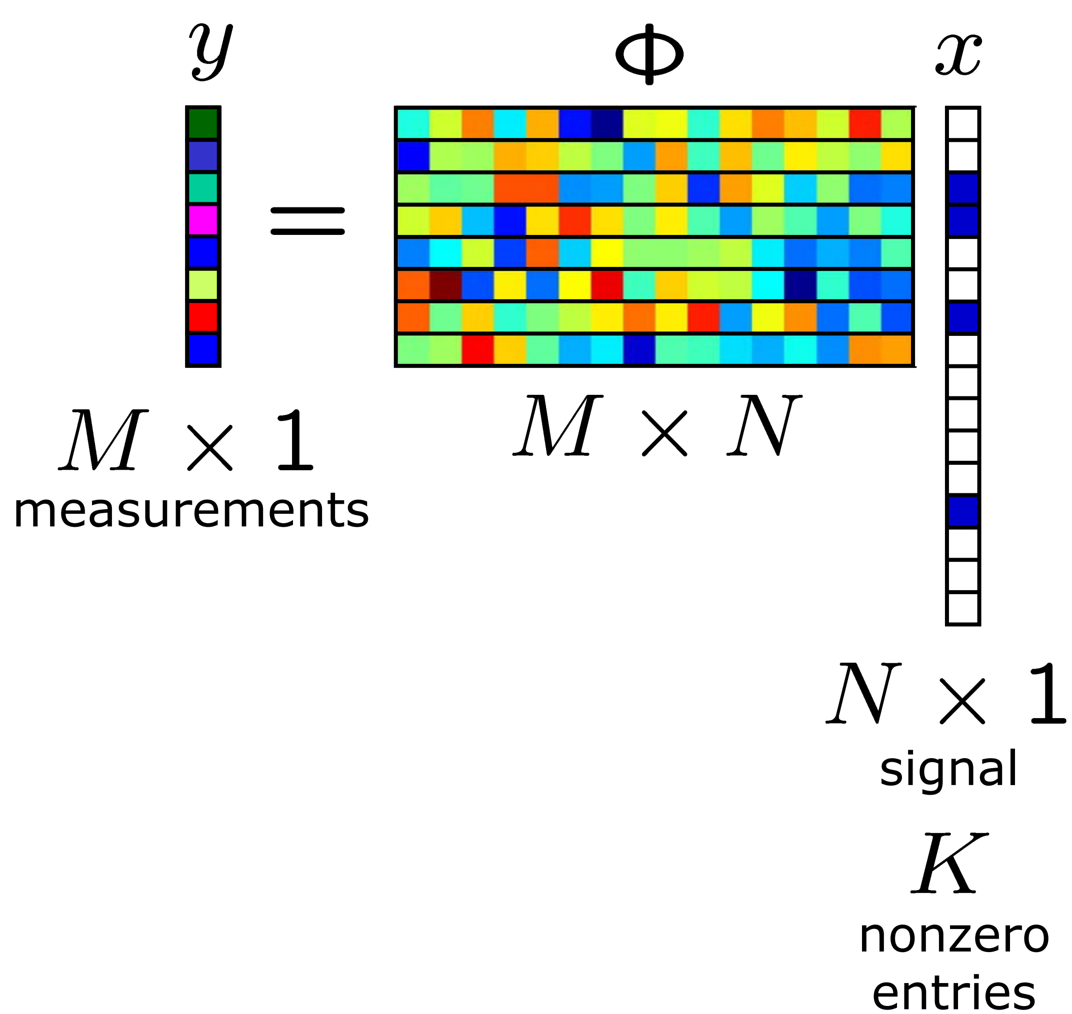
Geometry: Embedding in R^M

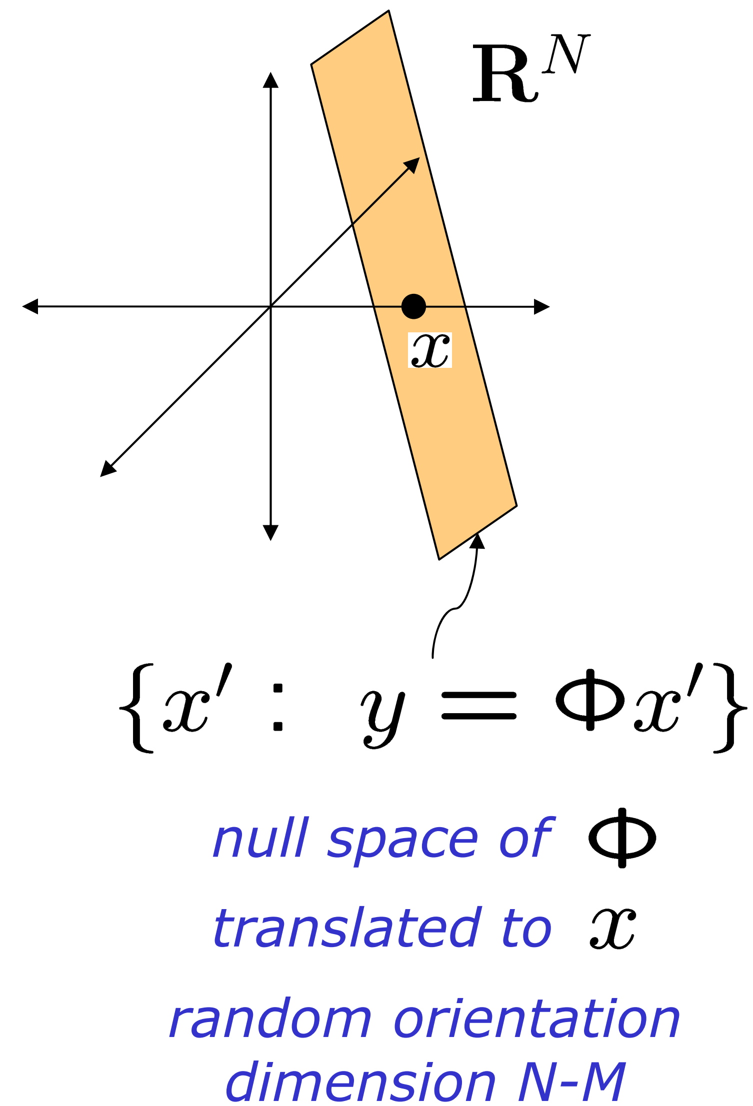


- Φ(K-plane) = K-plane in general
- M ≥ 2K measurements
  - necessary for injectivity
  - sufficient for injectivity when Φ Gaussian
  - but not enough for efficient, robust recovery
- (PS - can distinguish *most* K-sparse x with as few as M=K+1)

Source : Justin Romberg & Michael Wakin

The Geometry of L_1 Recovery

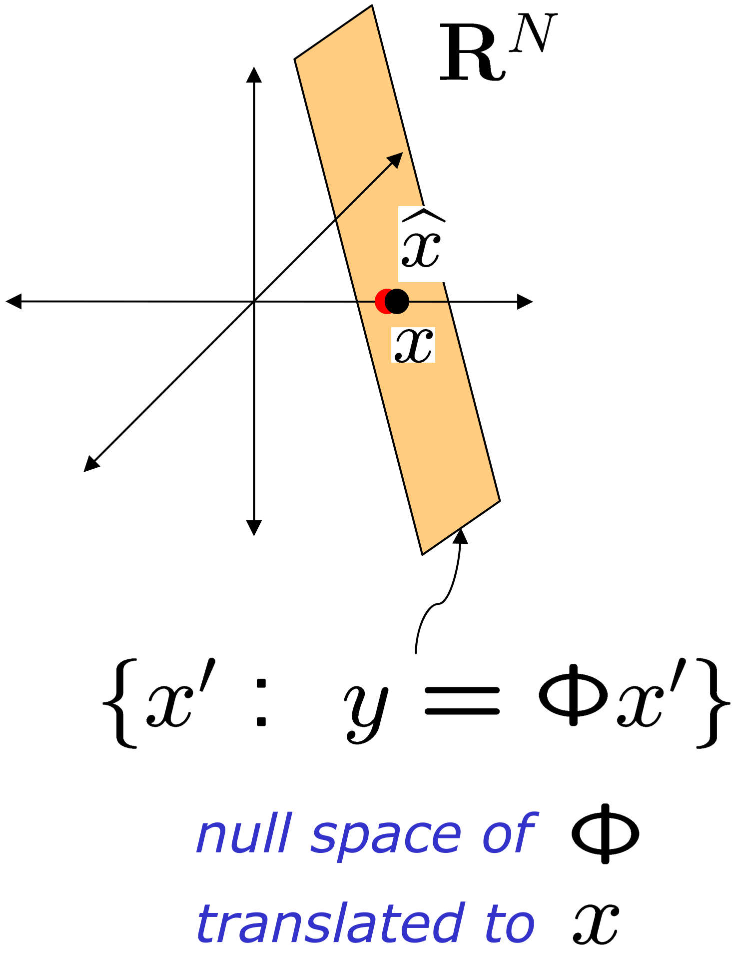




Source : Justin Romberg & Michael Wakin

L_0 Recovery Works

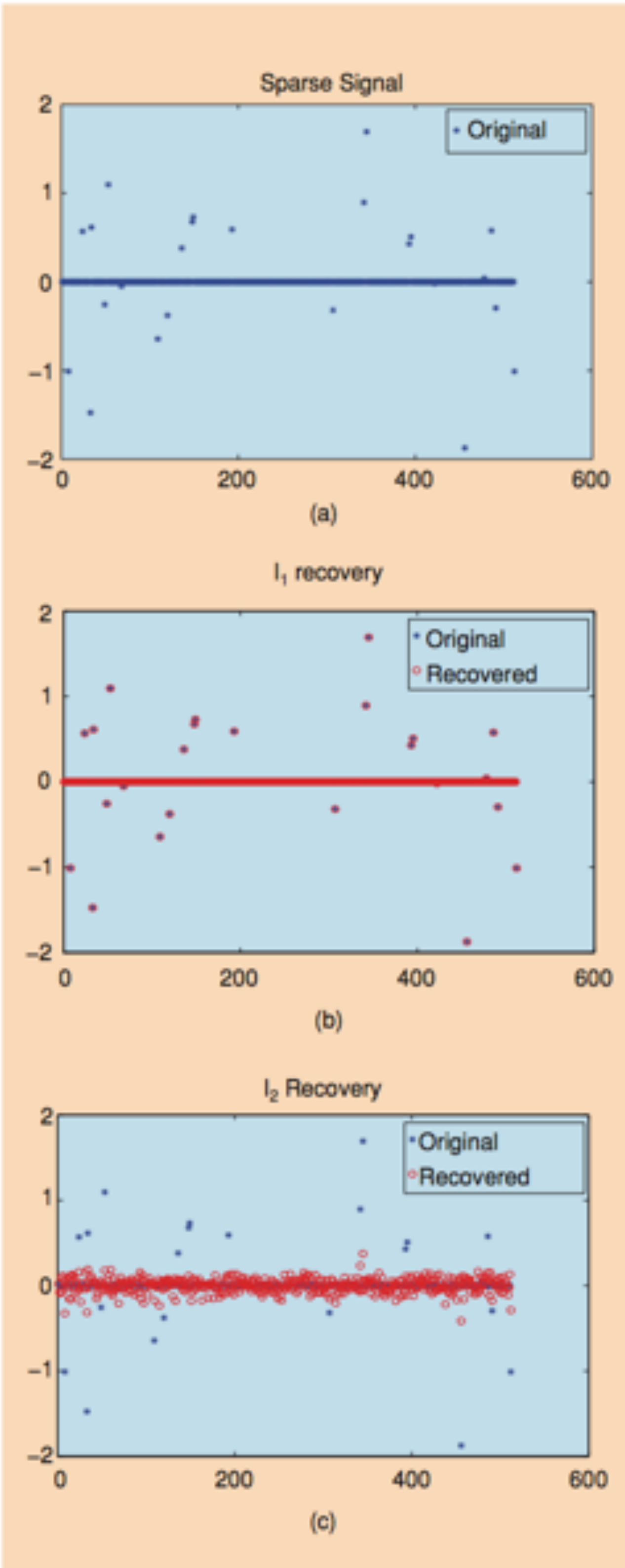
$$\widehat{x} = \arg \min_{y=\Phi x'} \|x'\|_0$$



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

(w.p. 1 for Gaussian Φ)

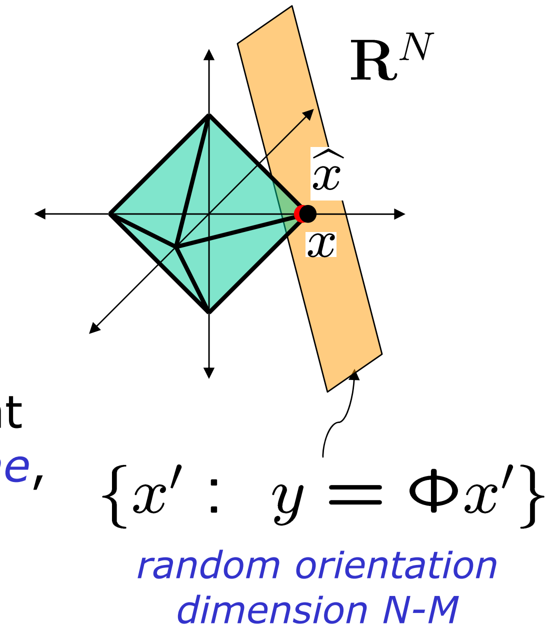
Source : Justin Romberg & Michael Wakin



[FIG2] (a) A sparse real valued signal and (b) its reconstruction from 60 (complex valued) Fourier coefficients by ℓ_1 minimization. The reconstruction is exact. (c) The minimum energy reconstruction obtained by substituting the ℓ_1 norm with the ℓ_2 norm; ℓ_1 and ℓ_2 give wildly different answers. The ℓ_2 solution does not provide a reasonable approximation to the original signal.

Why L_1 Works

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

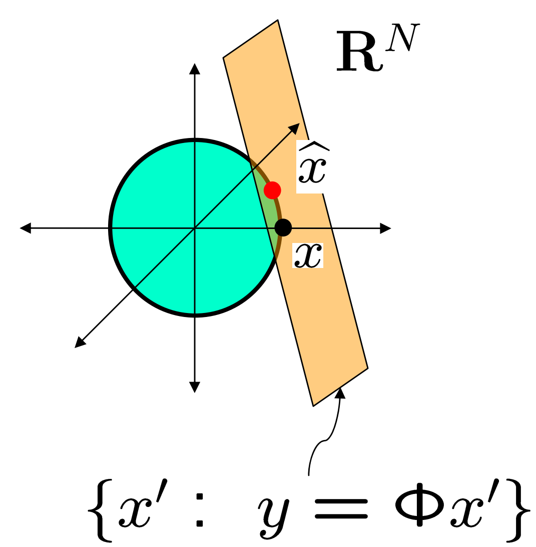


Criterion for success:
Ensure with high probability that a *randomly oriented (N-M)-plane*, *anchored on a K-face* of the L_1 ball, *will not intersect* the ball.

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

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**

Source : Justin Romberg & Michael Wakin

CS Signal Recovery

- Reconstruction/decoding: given $y = \Phi x$
  (ill-posed inverse problem) find x

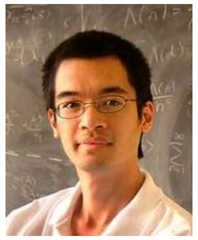
- ***L_2*** fast, wrong $\quad \widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$

- ***L_0*** correct, slow $\quad \widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$

- ***L_1*** **correct, efficient**
  mild oversampling $\quad \widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$
  [Candes, Romberg, Tao; Donoho] ***linear program***

$$M = O(K \log(N/K)) \ll N$$



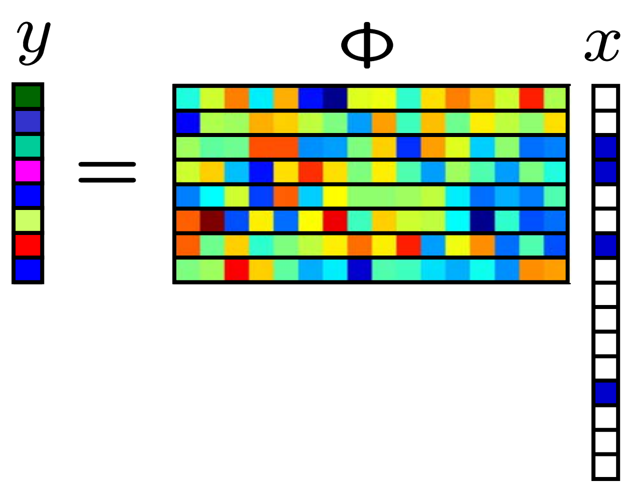



Source : Justin Romberg & Michael Wakin

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.

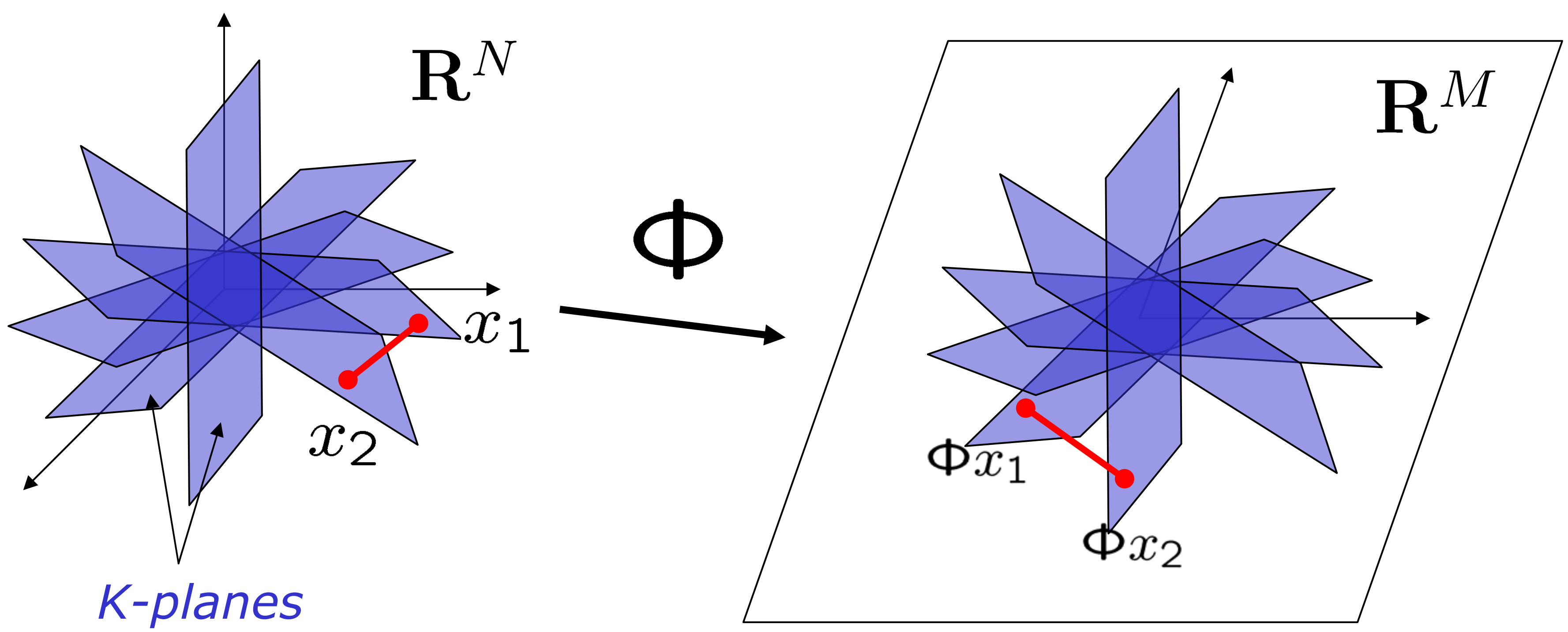
- Implications: tractable, stable, robust recovery

Source : Justin Romberg & Michael Wakin

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)

Source : Justin Romberg & Michael Wakin

Implications of RIP

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

If $\delta_{2K} < 0.41$, ensured:

1. *Tractable recovery: All K-sparse x* are perfectly recovered via $\ell_1$minimization.
2. *Robust recovery: For any $x \in R^N$,*

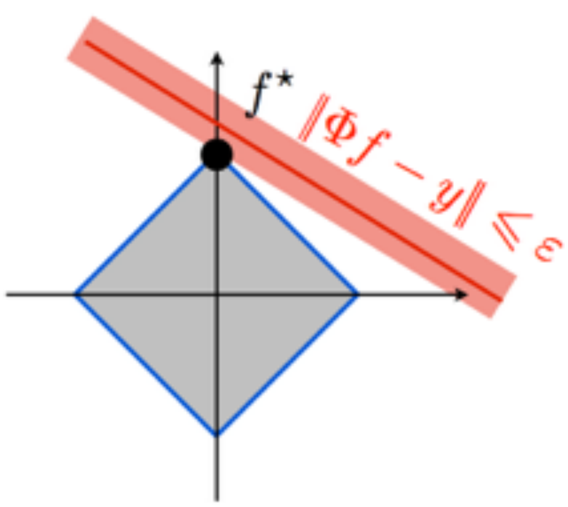
$$\|x - \widehat{x}\|_{\ell_1} \leq C\|x - x_K\|_{\ell_1} \text{ and } \|x - \widehat{x}\|_{\ell_2} \leq C\frac{\|x - x_K\|_{\ell_1}}{K^{1/2}}.$$

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

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

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

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



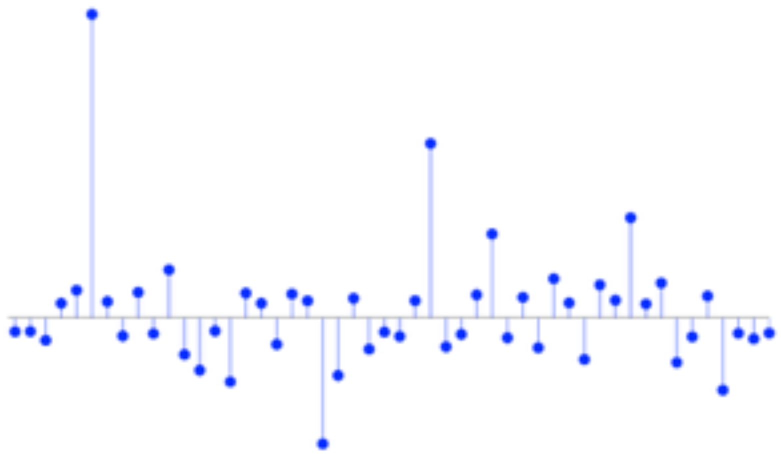
Verifying RIP:
How Many Measurements?

- 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)

Source : Justin Romberg & Michael Wakin



Optimality

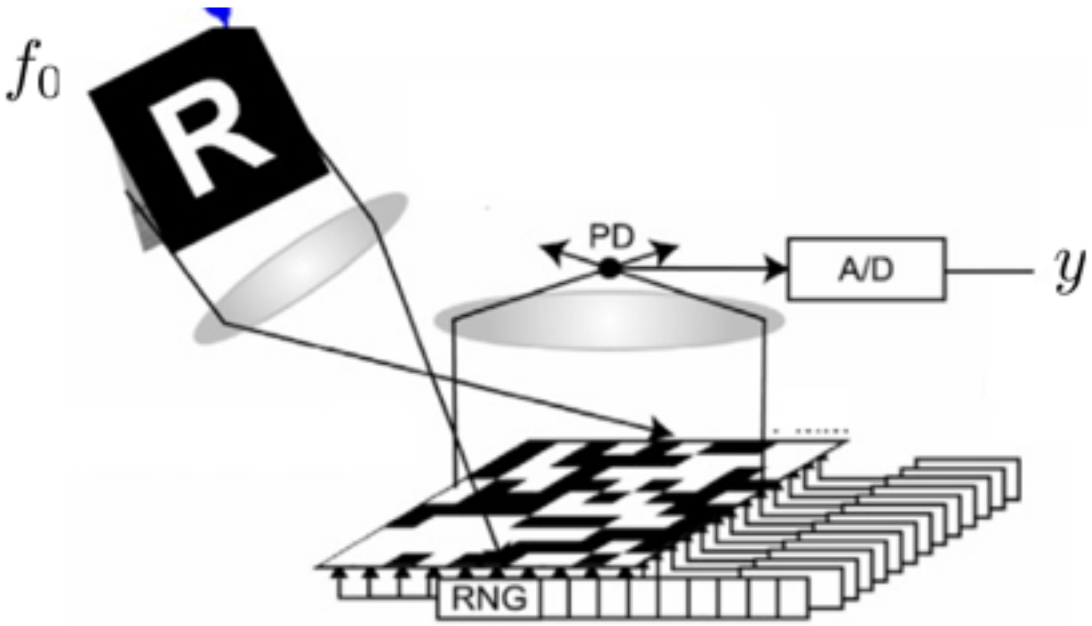
[Candès; Donoho]

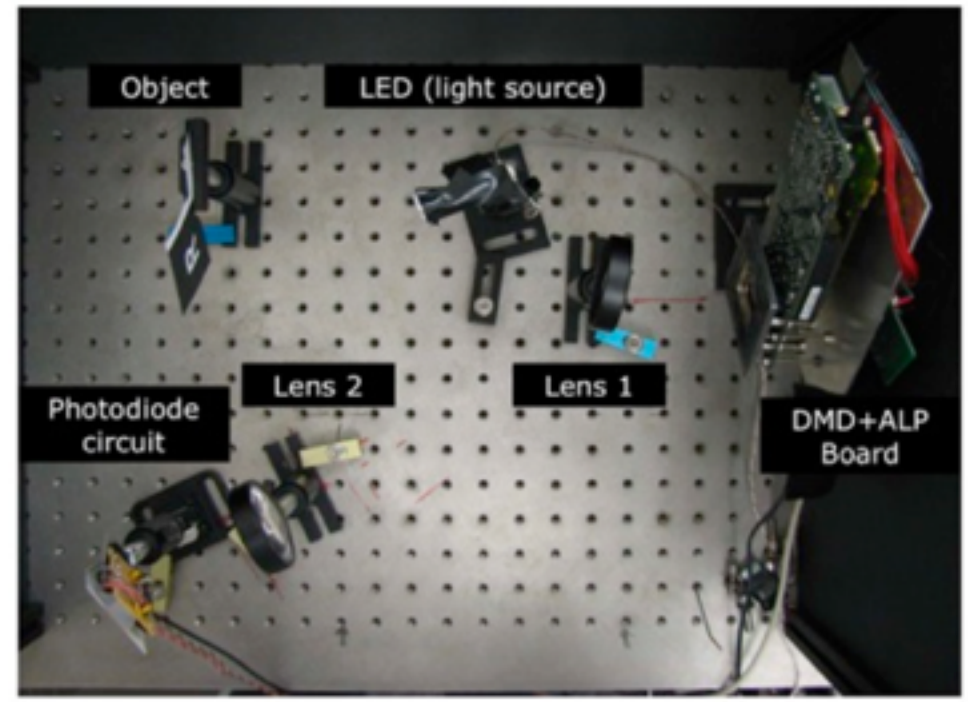
- Gaussian Φ has RIP order 2K (say) with ***M = O(K log(N/M))***
- Hence, for a given M, for $x \in wl_p$ (i.e., $|x|_{(k)} \sim k^{-1/p}$), $0 < p < 1$, (or $x \in l_1$)

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

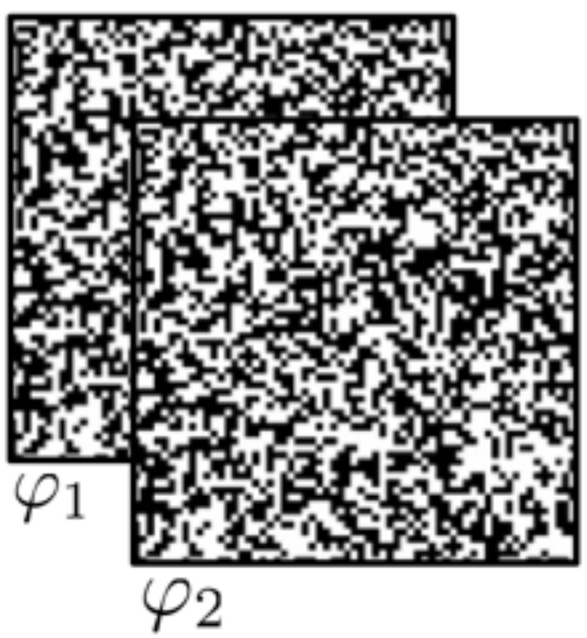
- Up to a constant, these bounds are *optimal*: no other linear mapping to R^M followed by *any* decoding method could yield lower reconstruction error over classes of compressible signals
- Proof (geometric): Gelfand n-widths [Kashin; Gluskin, Garnaev]

Applications : new sensing architectures





$y[i] = \langle f_0, \varphi_i \rangle$



$f_0, N = 256^2$ $\quad f^\star, P/N = 0.16$ $\quad f^\star, P/N = 0.02$

Source : Gabriel Peyré

Thank you for listening !