Foundations of Compressed Sensing

Mike Davies
Edinburgh Compressed Sensing research group (E-CoS)
Institute for Digital Communications
University of Edinburgh

Part I: Foundations of CS

- Introduction to sparse representations & compression
- Compressed sensing – motivation and concept
- Information preserving sensing matrices
- Practical sparse reconstruction
- Summary & engineering challenges

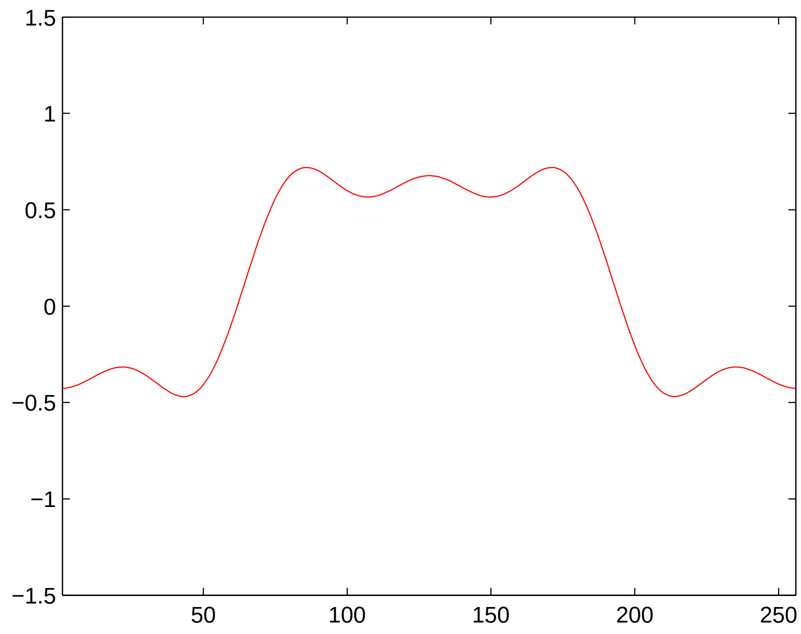
Sparse representations and compression

Fourier Representations

The Frequency viewpoint (Fourier, 1822):

Signals can be built from the sum of harmonic functions (sine waves)

Joseph Fourier



Atomic representation:

$$x(t) = \sum_k c_k e^{\omega_0 k t} = F\boldsymbol{c}$$

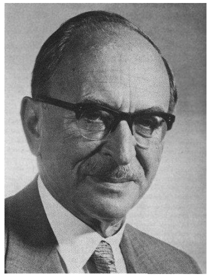
signal

Fourier coefficients

Harmonic functions

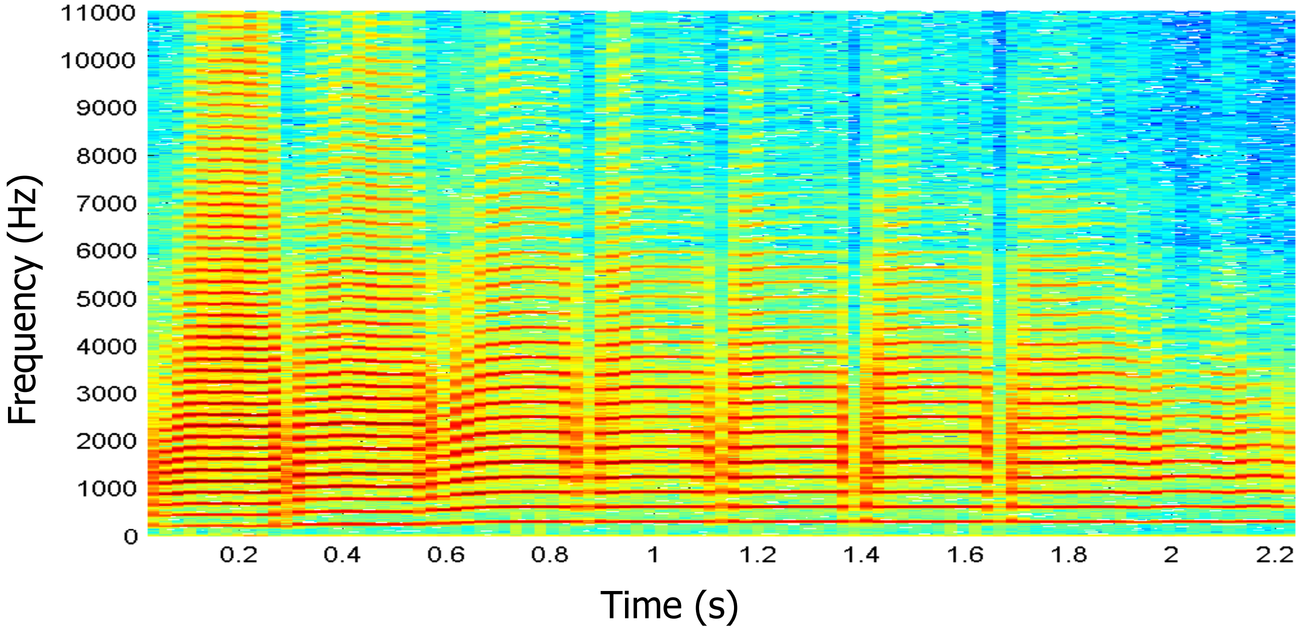
Time-Frequency representations

Time and Frequency (Gabor)



"Theory of Communication," J. IEE (London) , 1946

"… a new method of analysing signals is presented in which time and frequency play symmetrical parts…"



Atomic (dictionary) representation:

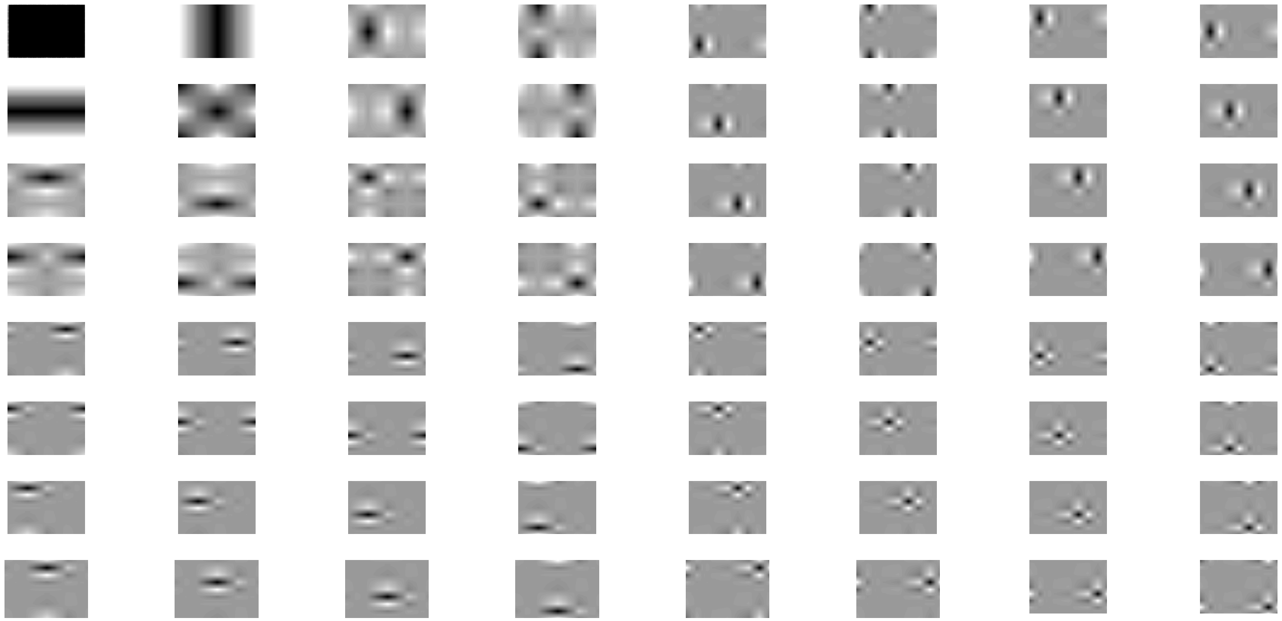
$$x(t) = \sum_n \sum_k c_{n,k} \times g(t - n\tau) e^{j\omega kn} = \Phi \boldsymbol{c}$$

Space-Scale representations

the wavelet viewpoint:

"Daubechies, Ten Lectures on Wavelets**," SIAM 1992**

Images can be built of sums of *wavelets*. These are multi-resolution edge-like (image) functions.



and many other representations

… more recently:

chirplets,

curvelets,

edgelets,

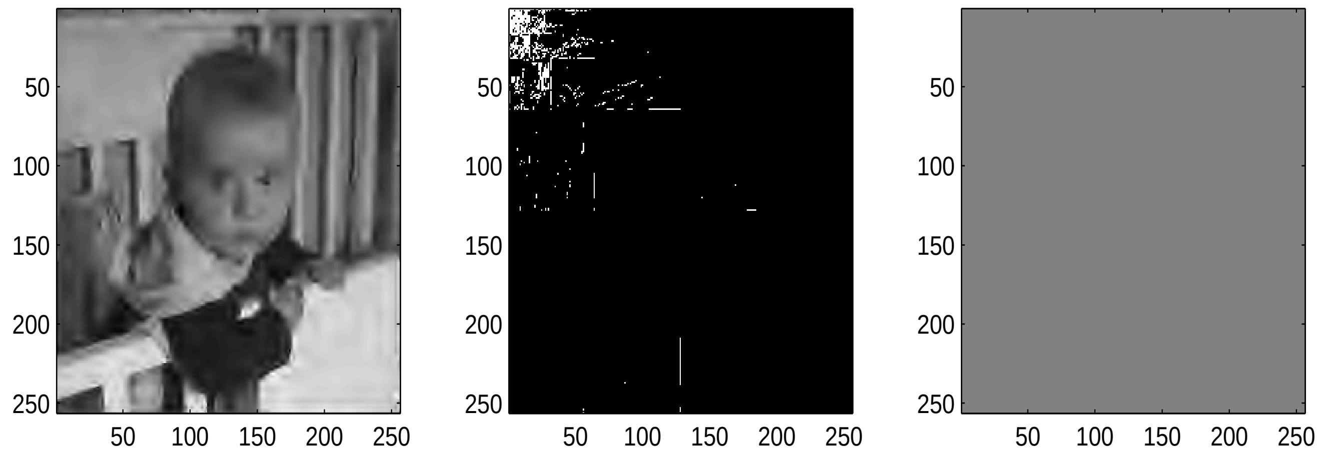
wedgelets, …

dictionary learning...

Coding signals of interest

What is the difference between quantizing a signal/image in the transform domain rather than the signal domain?

Compressed to 0.1 bits per pixel



Quantization in wavelet domain

Tom's nonzero wavelet coefficients

Quantization in pixel domain

Good representations are efficient – e.g. sparse!

Sparsity & Compression

A vector x is k-sparse, if only k of its elements are non-zero.

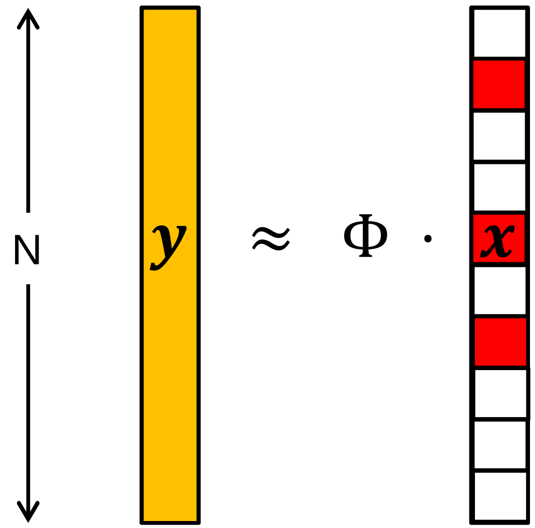
$$[0\ 0.5\ 0\ 0\ 0.1\ 0\ -0.2\ 0\ 0\ 0\ 0\ 0]^T$$

Such vectors have only k-degrees of freedom (k-dimensional) and there are “N choose k”, $\binom{N}{k}$, possible combinations of nonzero coefficients.

Coding cost:

N floats

$= \mathcal{O}(N)$ bits



Coding cost:

k floats + $\log_2 \binom{N}{k}$ bits

$= \mathcal{O}\left(k \log_2(N/k)\right)$ bits

$$\boldsymbol{y} \approx \Phi \boldsymbol{x}$$

Compressed sensing: motivation and concepts

Generalized Sampling

Different ways to measure…

Equivalent to inner product with various functions

pointwise sampling, tomography, coded aperture,…

Generalized Sampling

Different ways to measure…

Equivalent to inner product with various functions

$\langle$  ,  $\rangle$

pointwise sampling, tomography, coded aperture,…

Generalized Sampling

Different ways to measure…

Equivalent to inner product with various functions

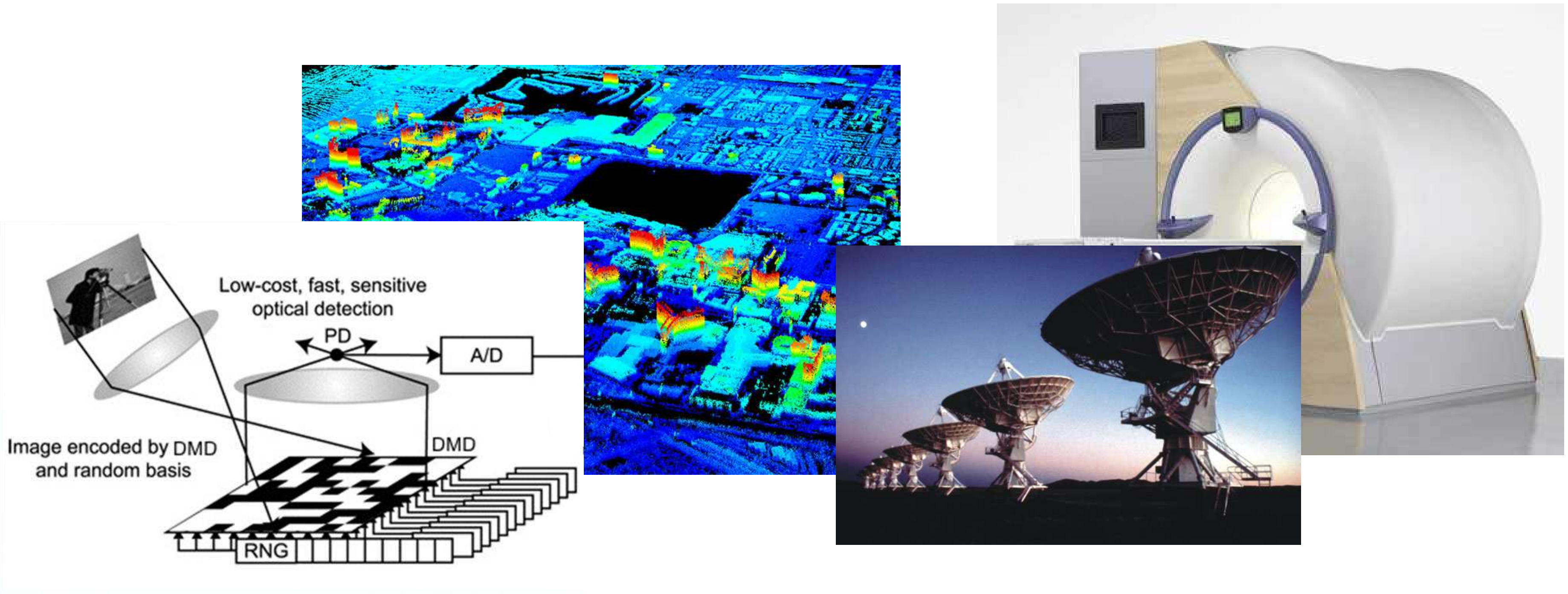
⟨  ,  ⟩

pointwise sampling, tomography, coded aperture,…

New Challenges

Challenge #1: Insufficient Measurements

Complete measurements can be costly, time consuming and sometimes just impossible!



New Challenges

Challenge #2: Too much data

e.g.

DARPA ARGUS-IS

1.8 Gpixel image sensor

15cm resolution, 12 frames a second

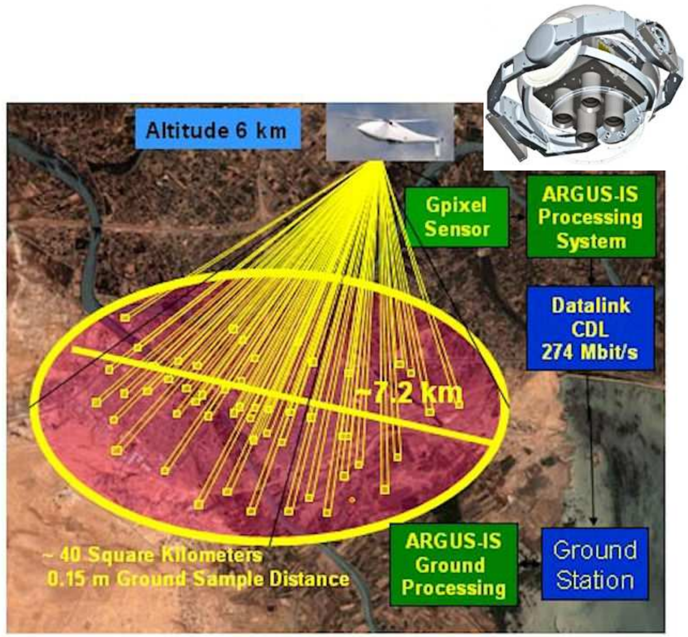
Giving a video rate output:

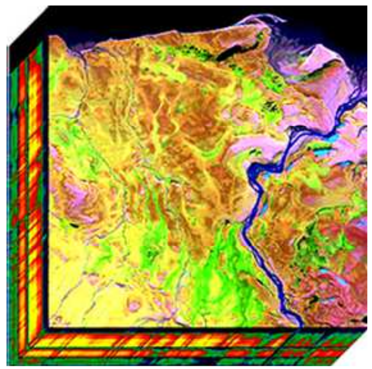
444 Gbits/s

… but the comms link data rate is:

274 Mbits/s

Currently visible spectrum. What about hyperspectral?…





The new hope: Compressed Sensing

E. Candès, J. Romberg, and T. Tao, **"Robust Uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,"** IEEE Trans. Information Theory, 2006

D. Donoho, **"Compressed sensing,"** IEEE Trans. Information Theory, 2006

Why can't we just sample signals at the "Information Rate"?

When compressing a signal we typically take lots of samples (sampling theorem), move to a transform domain, and then throw most of the coefficients away! Can we just sample what we need?

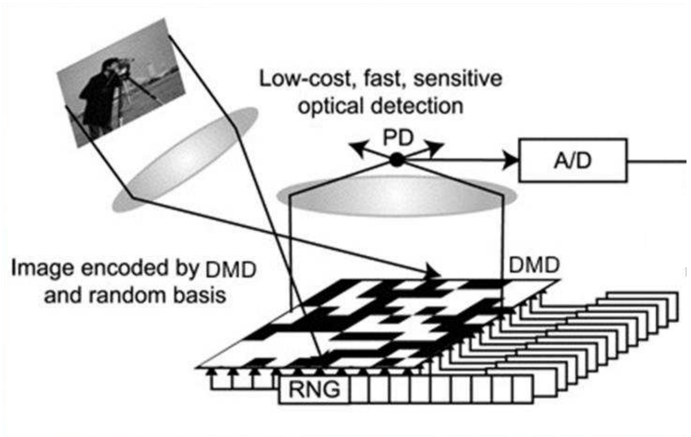
Yes! …and more surprisingly we can do this non-adaptively.

Potential applications

Compressed Sensing provides a new way of thinking about signal acquisition.

Applications areas already include:

- Medical imaging
- Hyperspectral imaging
- Astronomical imaging
- Distributed sensing
- Radar sensing
- Geophysical (seismic) exploration
- High rate A/D conversion



Rice University single pixel camera

Compressed sensing Overview

Observe $\boldsymbol{x} \in \mathbb{R}^N$ via $m \ll N$ measurements, $\boldsymbol{y} \in \mathbb{R}^m$ where $\boldsymbol{y} = \Phi \boldsymbol{x}$

Compressed Sensing assumes a compressible set of signals, i.e. approximately k-sparse.

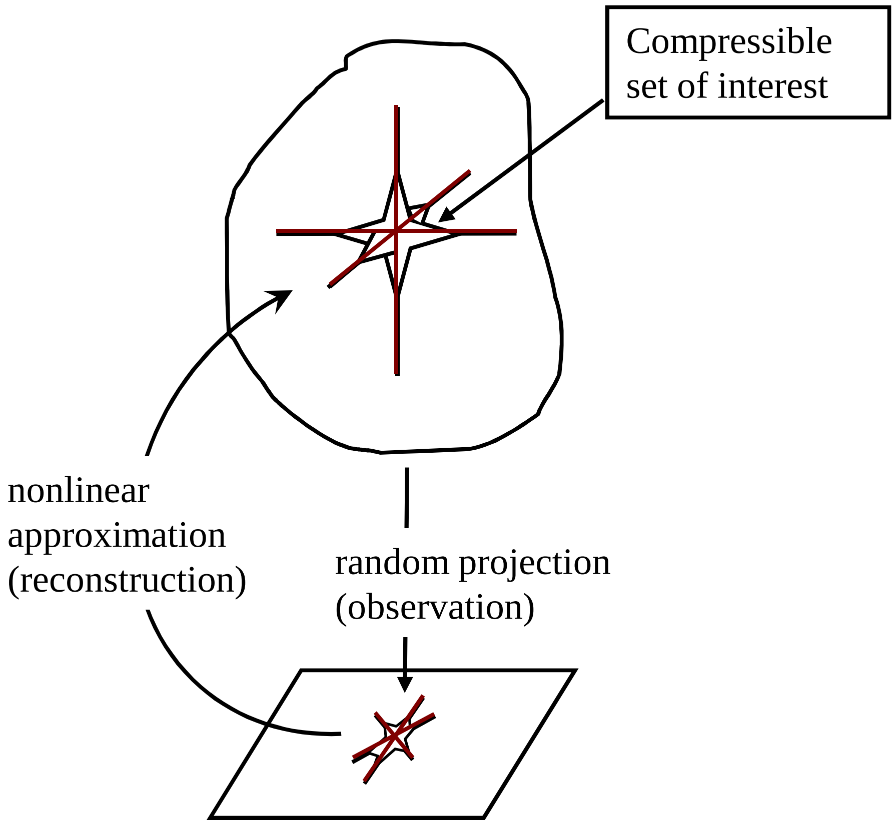
Using approximately

$$m \geq \mathcal{O}\left(k \log_2 \frac{N}{k}\right)$$

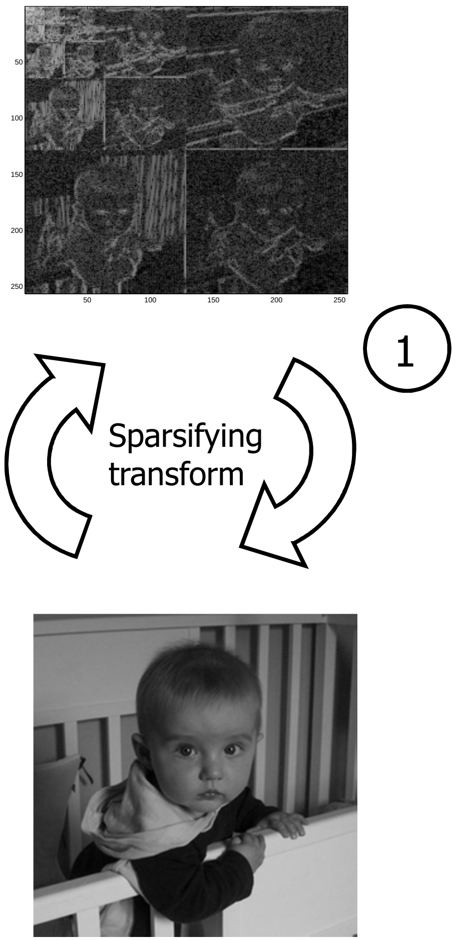
random projections for measurements we have little or no information loss.

Signal reconstruction by a nonlinear mapping.

Many practical algorithms with guaranteed performance e.g. L_1 min., OMP, CoSaMP, IHT.

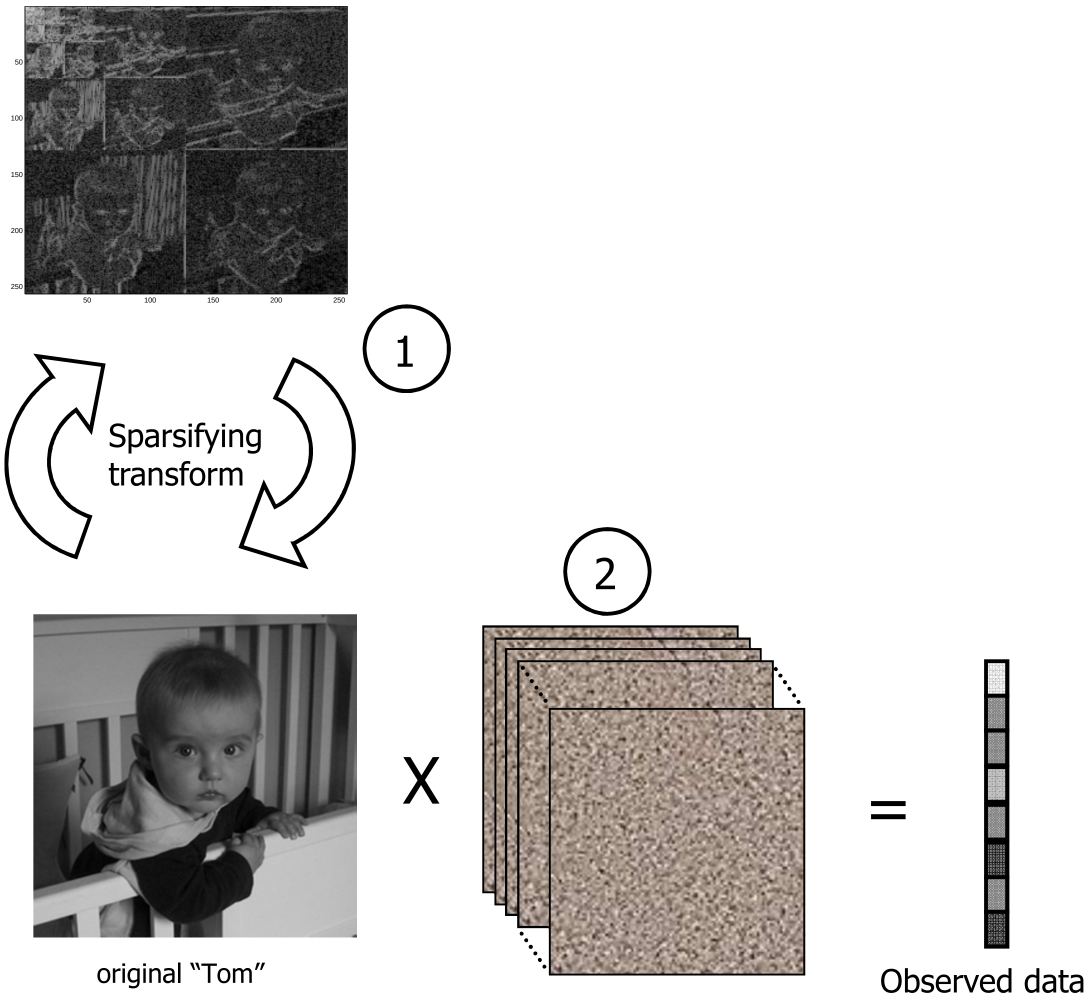


CS acquisition/reconstruction principle

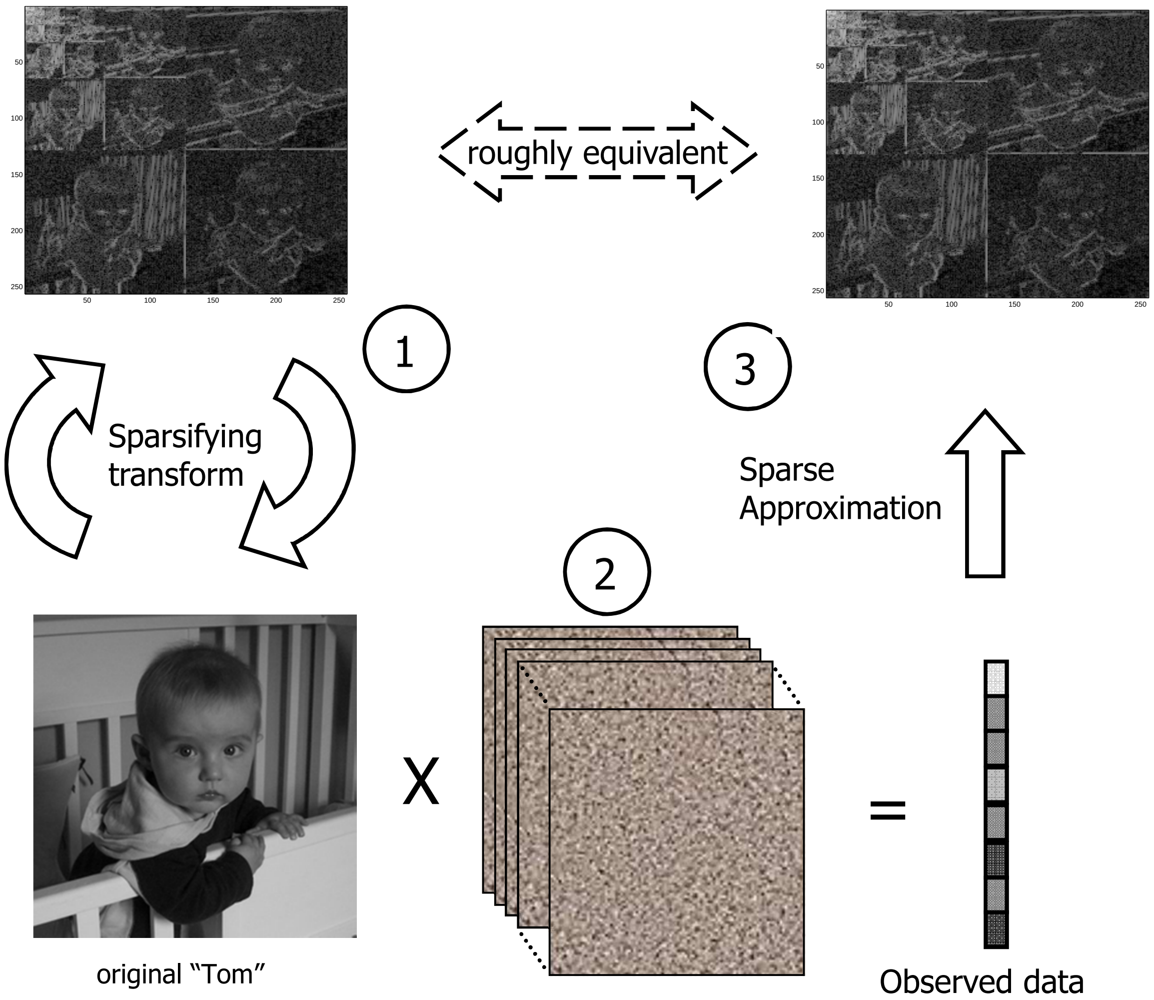


original "Tom"

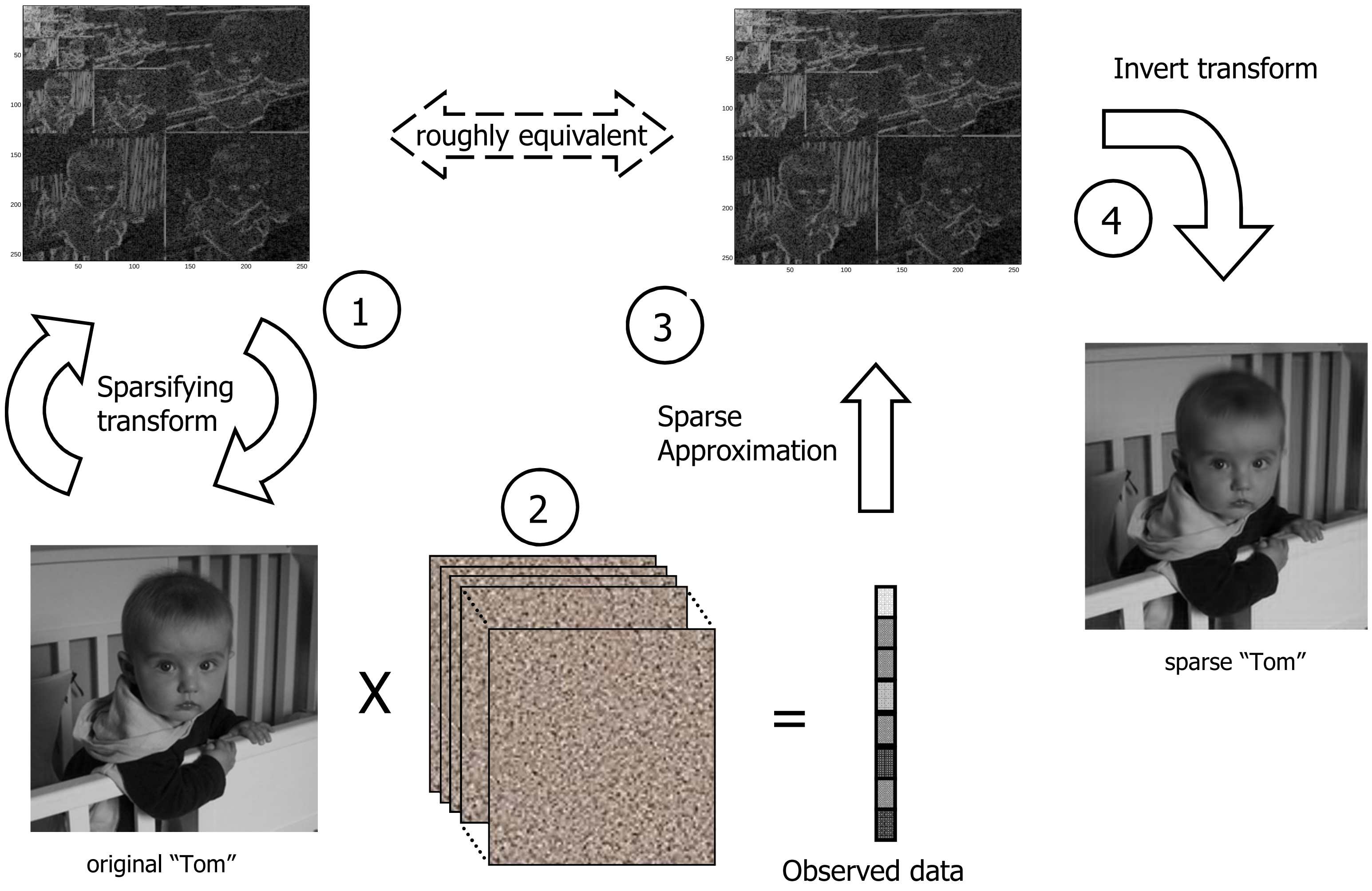
CS acquisition/reconstruction principle



CS acquisition/reconstruction principle



CS acquisition/reconstruction principle

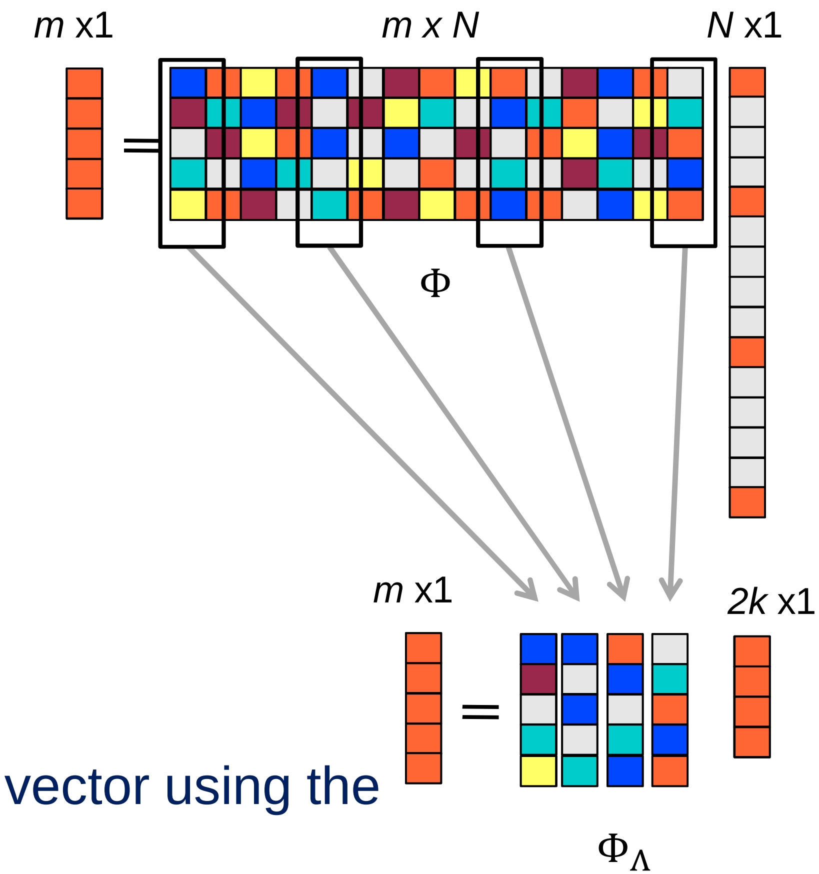


Information preserving sensing matrices

Information preservation

Underdetermined ($m < N$) linear systems are not invertible: $\Phi x = \Phi x' \nRightarrow x = x$

However, they may be invertible restricted to the sparse set: $\Sigma_{\mathrm{k}} \coloneqq \{x : |\mathrm{supp}(x)| \leq k\}$



Uniqueness on Σ_{k} is equivalent to

$$\mathcal{N}(\Phi) \cap \Sigma_{2k} = 0$$

$\mathcal{N}(\Phi) = \{z : \Phi z = 0\}$ is null space of Φ

We can then recover the original k-sparse vector using the following l_0 minimization scheme:

$$\hat{x} = \underset{x}{\mathrm{argmin}} \; \|x\|_0 \text{ subject to } \Phi x = y$$

Robust Null Space Properties

In order to achieve robustness we need to consider stronger NSPs

[Cohen et al. 2009] introduced the notion of Instance Optimality and showed that the following are equivalent up to a change in constant C

1. There exists a reconstruction mapping, Δ, such that for all x:

$$\|\Delta(\Phi x) - x\|_1 \leq C\sigma_k(x)_1$$

   where $\sigma_k(x)_1$ is the L_1 **best k-term approximation error** of x

2. Φ satisfies the following NSP:

$$\|z_\Lambda\|_1 \leq C'\sigma_{2k}(z)_1$$

   for all $z \in \mathcal{N}(\Phi)$ and all k-sparse supports, Λ.

Informally, null space vectors must be relatively flat.

Deterministic Sensing Matrices

Showing the NSP for a given Φ involves combinational computational complexity. The coherence of a matrix provides easily (but crude) computable guarantees.

Coherence

$$\mu(\Phi) = \max_{1 \le i < j \le N} \frac{\left|\left\langle \Phi_i, \Phi_j \right\rangle\right|}{\|\Phi_i\| \left\|\Phi_j\right\|}$$

Using the coherence it is possible to show that Φ is invertible on the sparse set if:

$$k < \frac{1}{2}\left(1 + \frac{1}{\mu(\Phi)}\right)$$

However, this only guarantees that $k \sim \mathcal{O}(\sqrt{m})$.

Restricted Isometry Property

Low Distortion Embeddings

A useful tool in compressed sensing is the *restricted isometry constant (RIC),* the smallest constant δ_k for which:

$$(1-\delta_k)\|x\|_2 \leq \|\Phi x\|_2 \leq (1+\delta_k)\|x\|_2$$

holds for all k-sparse vectors x.

A matrix Φ with $\boldsymbol{\delta_{2k} < 1}$ provides an embedding (one-to-one mapping) for the k-sparse set. $\boldsymbol{\delta_{2k}}$ also quantifies the robustness of the embedding (low distortion).

Random observations – a key insight in compressed sensing is that random matrices have small RICs with high probability whenever:

$$m \sim \mathcal{O}\left(k\delta_{2k}^{-2}\log_2(N/k)\right)$$

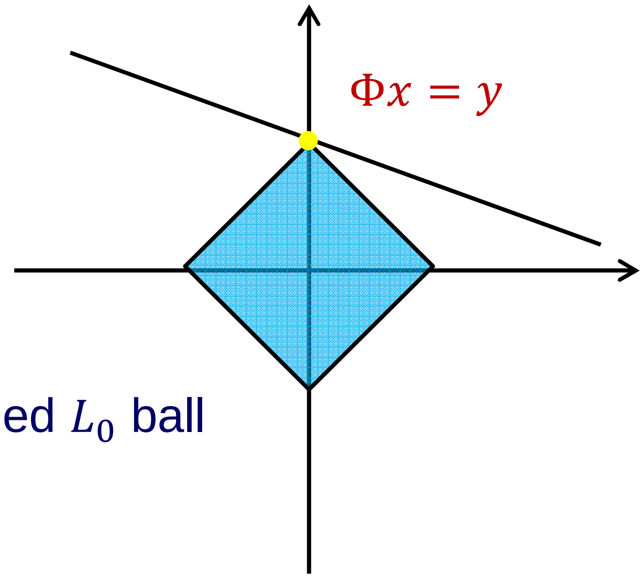
Practical sparse reconstruction

Sparse Recovery via L_1 Minimization

A key advance in Sparse Representations was the use of the L_1 minimization (convex!) as a proxy for L_0 reconstruction:

$$\hat{x} = \underset{x}{\mathrm{argmin}} \ \|x\|_1 \text{ subject to } \Phi x = y$$

where the L_1 norm is defined as: $\|\boldsymbol{x}\|_1 = \sum_i |x_i|$



Intuition:

1. Minimum L_1 solutions - ● - are sparse
2. L_1 ball is the “closest” convex set to the bounded L_0 ball

L_1 Performance Guarantees

For deterministic matrices L_1 minimization guarantees derived from coherence [Donoho & Elad 2003] : $\mathrm{m} \sim \mathcal{O}(k^2)$.

For general matrices [Candes 2008] showed:

Theorem: If Φ has RIP $\delta_{2k} \leq \sqrt{2} - 1 \implies L_1$NSP $\implies$ Instance Optimality:

$$\|\Delta(\Phi x) - x\|_1 \leq C\sigma_k(x)_1$$

Since i.i.d. random matrices are near optimal: $\mathbf{m} \sim \boldsymbol{\mathcal{O}}(\boldsymbol{k}\log(\boldsymbol{N}/\boldsymbol{k}))$

Since then it has been shown [Donoho & Tanner 2009] that $L_1 - L_0$ equivalence for sparse vectors and random Φ if: $\boldsymbol{m} \geq \mathbf{2}\boldsymbol{k}\log(\boldsymbol{N}/\boldsymbol{k})$.

Other Practical Recovery Algorithms

The other main class of practical (polynomial complexity) recovery algorithms are “Greedy methods”: Orthogonal Matching Pursuit, CoSAMP, Iterative Hard Thresholding…

Aim to solve mixed continuous/discrete L_0 minimization problem (non-convex!) using: (1) Least squares minimization and (2) Hard decisions on coefficient selection.

e.g. Iterative Hard Thresholding [Blumensath, D. 2010]: greedy gradient projection

$$x^{\{t+1\}} = P_{\Sigma_k}\left(x^{\{t\}} + \mu\Phi^T\left(y - \Phi x^{\{t\}}\right)\right)$$

Theorem : RIP $\delta_{2k} \leq 1/5 \implies \lim_{t\to\infty} x^{\{t\}} \implies$ Instance Optimality

Performance guarantees come directly from RIP type considerations.

Summary & Engineering Challenges

Sparse Representations provide a powerful nonlinear model for real world signals.

Sparse signals can be sampled and faithfully reconstructed using many fewer samples than predicted by traditional sampling theory.

Engineering Challenges in CS

- What is the right signal model?
  Sometimes obvious, sometimes not. When can we exploit additional structure?
- How can/should we sample?
  Physical constraints; SNR issues; can we randomly sample; exploiting structure; how many measurements?
- What are our application goals?
  Reconstruction? Detection? Estimation?

Selected References

R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. Const. Approx., 28(3):253-263, 2008.

T. Blumensath, M. E. Davies 2009, "Iterative Hard Thresholding for Compressed Sensing", Applied and Computational Harmonic Analysis, vol 27(3), pp 265-274, 2009.

T. Blumensath, M. E. Davies 2010, "Normalised Iterative Hard Thresholding; guaranteed stability and performance", IEEE Journal of Selected Topics in Signal Processing vol 4(2), pp 298-309, 2010.

E. J. Candès and T. Tao, ‚Near optimal signal recovery from random projections: Universal encoding strategies?,‛ IEEE Trans. Info. Theory, vol. 52, no. 12, pp. 5406–5425, Dec. 2006

E. J. Candès, J. K. Romberg, and T. Tao, ‚Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,‛ IEEE Trans. Info. Theory, vol. 52, no. 2, pp. 489–509, 2006

E. Candes. The restricted isometry property and its implications for compressed sensing. Comptes rendus de l'Academie des Sciences, Serie I, 346(9-10):589-592, 2008.

A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. J. Amer. Math. Soc., 22(1):211-231, 2009.

D. L. Donoho and M. Elad, ‚Optimally sparse representation in general (nonorthogonal) dictionaries via L1 minimization,‛ Proc. Nat. Acad. Sci., vol. 100, no. 5, pp. 2197–2202, Mar. 2003

D. L. Donoho, ‚Compressed sensing,‛ IEEE Trans. Info. Theory, vol. 52, no. 4, pp. 1289–1306, Sep. 2006

D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. J. Amer. Math. Soc., 22(1):1-53, 2009.

D. Needell and J. A. Tropp, ‚CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,‛ Appl. Comput. Harmon. Anal., vol. 26, no. 3, pp. 301–321, May 2008

J. Tropp and A. Gilbert. Signal recovery from partial information via orthogonal matching pursuit. IEEE Trans. Inform. Theory, 53(12):4655-4666, 2007.