

Compressed Sensing: The Next “Big Idea” in Signal Processing

Dr Garry Newsam

Intelligence Surveillance and Reconnaissance Division

DSTO

Acknowledgement: Rich Baraniuk

Present Sensing Paradigm

❖ Sensing by sampling

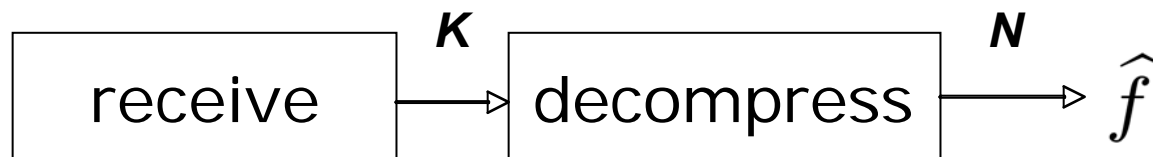
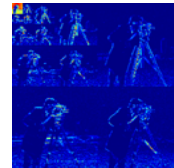
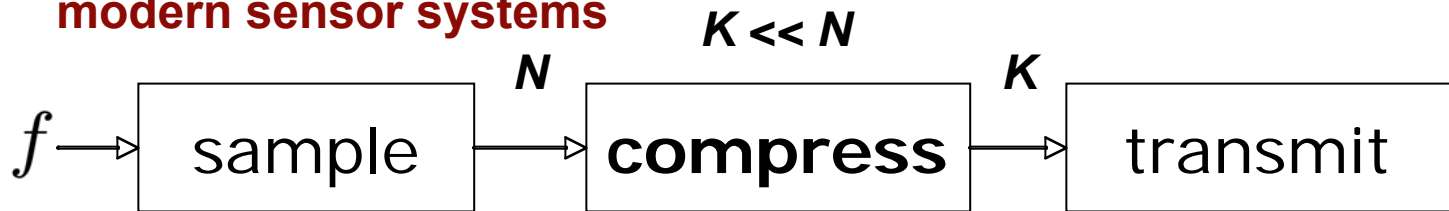
◆ *Sample* data at sensor

- A/D converter, camera back plane

◆ *Compress, transmit, decompress, process*

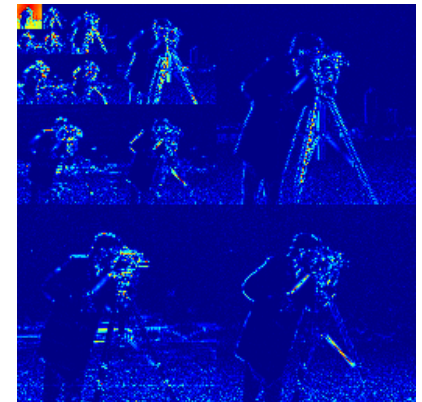
- Nonlinear, signal dependent

◆ Combination produces a “brick wall” faced by many many modern sensor systems



Compressed Sensing

- ❖ **Observation: most signals are sparse (or nearly sparse) in some representation**
 - ◆ Only K stars in N^2 pixel image of night sky
 - ◆ Only K significantly non-zero coefficients in wavelet representation of N^2 pixel image
 - ◆ Only K targets in N^d sensor measurements where $K \ll N$



- ❖ **Vision:** develop sensors that only collect minimum amount of information necessary to reconstruct or process signal

- ❖ **Problems:**
 - ◆ Don't know in advance which locations contain target / which wavelet coefficients are non-zero
 - ◆ Can't build instruments that directly sense in desired basis
 - Can't directly measure wavelet coefficients

- ❖ **Key result (1):** $O(K \log N/K)$ i.i.d. noisy random projections suffice to reconstruct a nearly sparse signal
 - ◆ Do so by solving the convex, mixed-norm optimisation problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \sigma^2$$

- ❖ **Key result (2): fast algorithms have been derived to solve this optimisation problem**



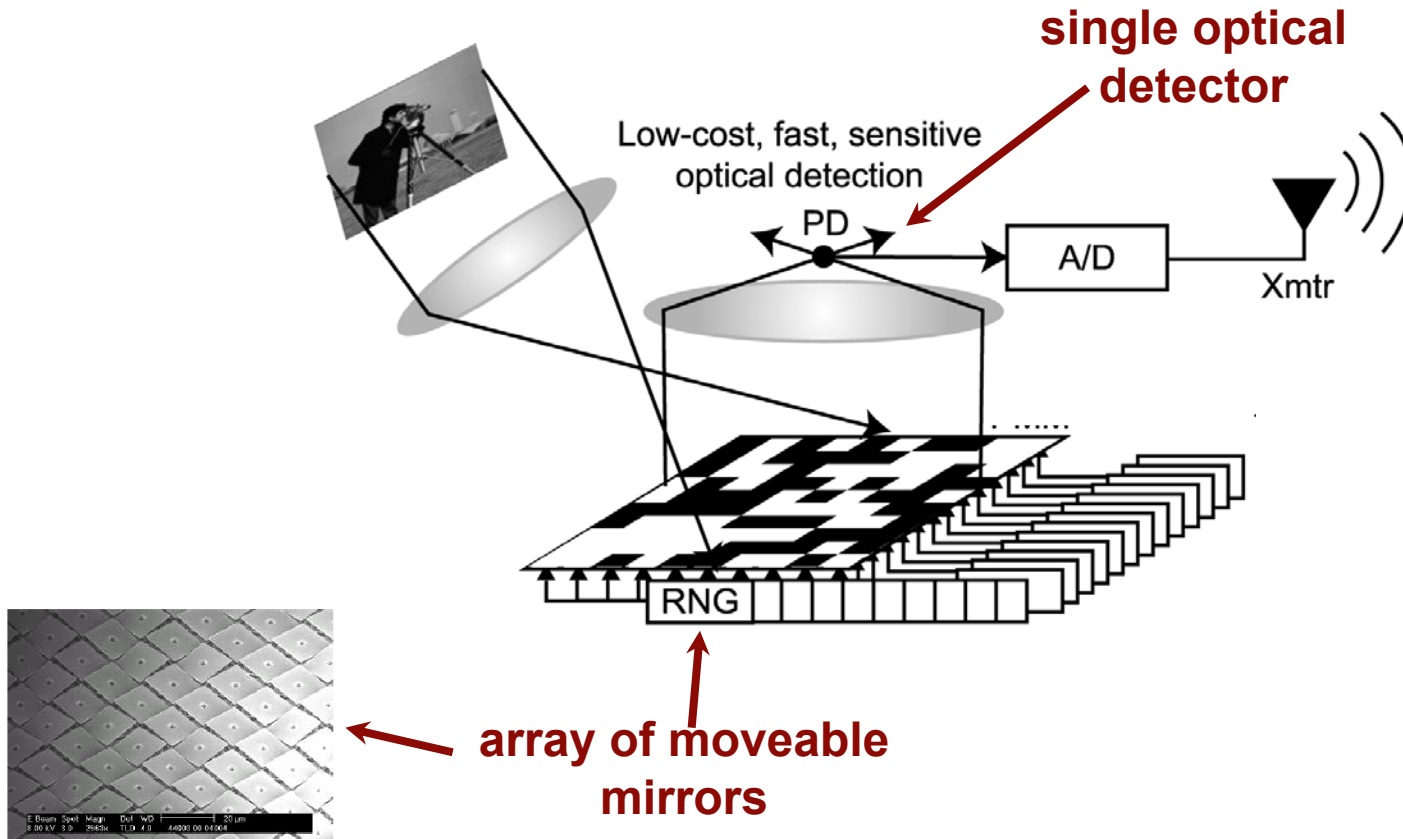
- ❖ **“Ah-ha” insight: i.i.d. random projections remain i.i.d. random projections under orthogonal change of basis**

- ❖ **Practical implications:**

- ◆ **Build sensor that takes small number of cheap random projections of data in convenient reference frame**
- ◆ **Digitally convert projections to known sparse reference frame, then reconstruct or further process signal**

Example of Compressed Sensor

- ❖ Single pixel camera developed by group at Rice University



History and Current State of Play

- ❖ **Background results and drivers**
 - ◆ Sparse reconstructions achievable via ℓ_1 minimisation
 - ◆ Characterisation of conditions under which functions of limited bandwidth can be recovered from limited samples
 - ◆ Desire to work with compressed rather than redundant representations

- ❖ **~ 2004:** proofs start emerging that sparse signals can be accurately reconstructed from minimally redundant set of randomly sampled data under very general conditions
 - ◆ Candes, Donoho and Tao

- ❖ **Explosion of papers**
 - ◆ **Multiple sessions in SPIE conferences**

- ❖ **DARPA programs**
 - ◆ **Analog-to-Information (Dennis Healey)**

- ❖ **Extensions from signal reconstruction to**
 - ◆ **Sensor construction**
 - ◆ **Detection and classification**
 - ◆ **Information theory and coding**
 - ◆ **Network analysis and monitoring**
 - ◆ **Database queries**

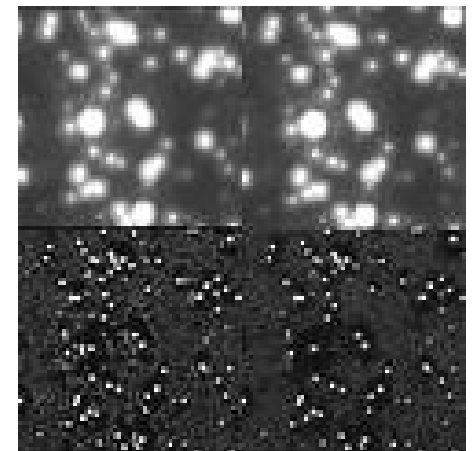
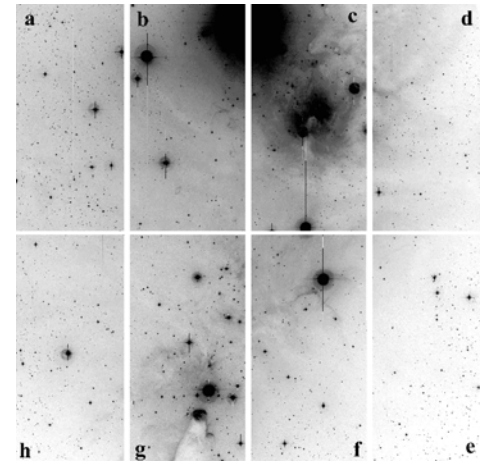
- ❖ **Research groups and web sites**
 - ◆ **Rice University: <http://www.dsp.ece.rice.edu/cs/>**

A Little History

- ❖ **Image processing problems in astronomy**
 - ◆ stars are point sources but atmospheric and optical blur always present

- ❖ **Plethora of de-blurring algorithms**

- ❖ **CLEAN algorithm**
 - ◆ (Hogborn, 1974)
 - ◆ Name comes from author's characterisation of observed image as "dirty" and desired image as "clean",



- ❖ **Model:** image is a sum of a small number of point sources all modified by same known point-spread function

$$y(s_i) = \sum_j x_j \phi(s_i - s_j) \quad , \quad x_j \geq 0$$

$$\mathbf{y} = \Phi \mathbf{x} \quad , \quad [\Phi]_{ij} \equiv \phi(s_i - s_j)$$

- ❖ **Algorithm:** (really simple)

- ◆ **Step 0** $\mathbf{x}^0 = \mathbf{0}$, $\mathbf{r}^0 = \mathbf{y}$ \mathbf{r}^k residual

- ◆ **Step k** $i^k = \arg \max_i [\Phi^T \mathbf{r}^{k-1}]_i \approx \arg \max_i [\mathbf{r}^{k-1}]_i$

$$\mathbf{x}^k = \mathbf{x}^{k-1} + \epsilon^k \mathbf{e}_{i^k}^k \quad \epsilon^k \text{ step length}$$

$$\mathbf{r}^k = \mathbf{r}^{k-1} - \epsilon^k [\Phi]_{i^k}$$

- ◆ **Stopping criterion**

$$\|\mathbf{r}^k\|^2 \leq \sigma^2 \quad \sigma^2 \text{ noise level}$$

What Is It Doing? Why Does It Work?

- ❖ Works well in practice; still widely in use
- ❖ Algorithm can be characterised as optimisation by steepest descent
 - ◆ Marsh & Richardson, 1987 (some time after the event)
- ❖ Trade-off between two objective functions
 - ◆ At k -th step have

Total flux

$$\sum_j x_j^k = \alpha^k$$

$$x_j^k \geq 0$$

$$\|\mathbf{r}^k\|^2 \equiv \|\Phi \mathbf{x}^k - \mathbf{y}\|^2 = \beta^k$$

Mean square reconstruction error

- ❖ **Seek solution that balances flux against reconstruction error**

$$\min_{\delta \mathbf{x}^k} \alpha^k + \lambda \beta^k$$

- ❖ **At each step choose increment to maximise decrease in reconstruction error for prescribed increase in flux, i.e.**

$$\min_{\delta \mathbf{x}^k} \|\mathbf{y} - \Phi(\mathbf{x}^k + \delta \mathbf{x}^k)\|^2 - \|\mathbf{y} - \Phi \mathbf{x}^k\|^2 \approx 2 \delta \mathbf{x}^{kT} \Phi^T (\mathbf{y} - \Phi \mathbf{x}^k)$$

subject to

$$\sum_j (x^k + \delta x^k)_j - \sum_j x_j^k = \delta \mathbf{x}^{kT} \mathbf{e} = \epsilon^k$$

- ❖ **Easy to show solution is**

$$\delta x_i^k = \begin{cases} \epsilon^k & \text{if } i = \arg \max_j [\Phi^T (\mathbf{y} - \Phi \mathbf{x}^k)]_j \\ 0 & \text{otherwise} \end{cases}$$

Reconstructing Sparse Signals

- ❖ **Generic sparse reconstruction problem**

- ◆ **Noise free data**

$$\mathbf{y} = \Phi \mathbf{x} \quad , \quad \mathbf{x} \in \mathbf{R}^N \quad , \quad \mathbf{y} \in \mathbf{R}^K$$

$$x_i \neq 0 \quad \text{iff} \quad i \in \Omega \quad \text{where} \quad |\Omega| = M$$

$$M \leq K \ll N$$

- ❖ **Goal: use knowledge of sparsity to compensate for lack of data**

- ◆ **Problem: don't know structure of Ω *a priori***

Associated Optimisation Problems

- ❖ Consider family of ℓ_p “norms”

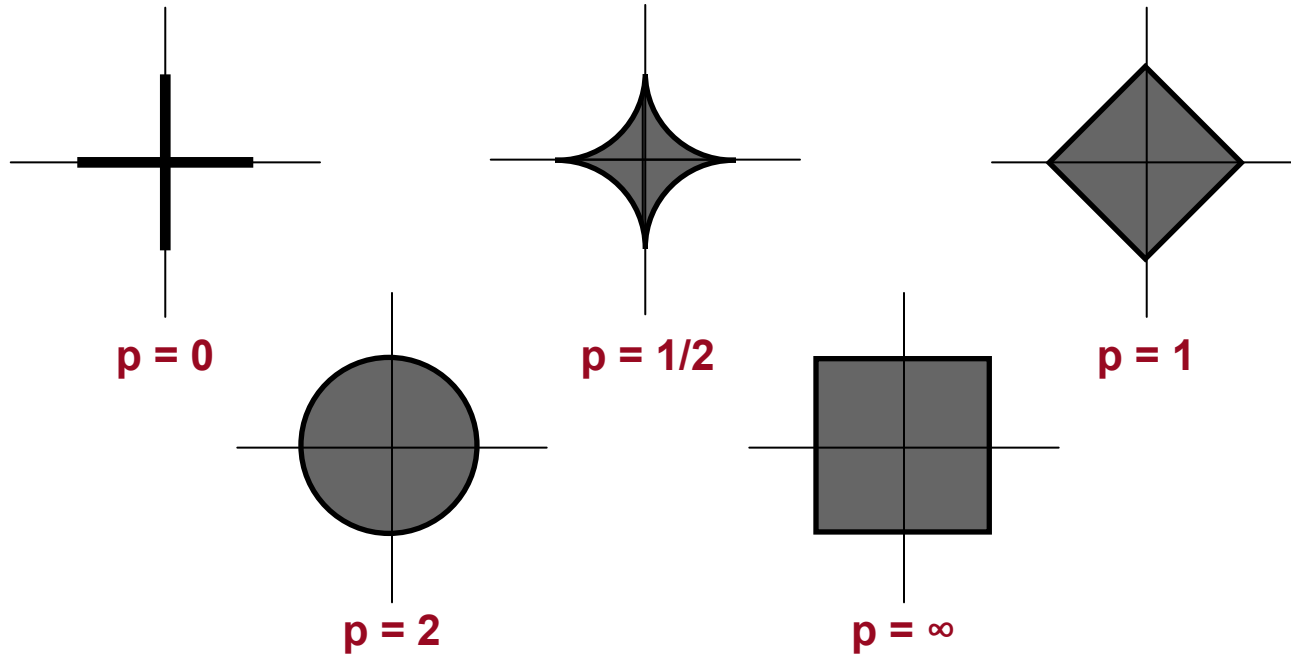
$$\|\mathbf{x}\|_p \equiv \begin{cases} |\{i : x_i \neq 0\}|, & p = 0; \\ \left[\sum_{i=1}^N |x_i|^p \right]^{1/p}, & 0 < p < \infty; \\ \max_i |x_i|, & p = \infty. \end{cases}$$

- ❖ Ideal sparse reconstruction would be solution of

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x}$$

- ❖ Problem: non-convex, NP-complete combinatorial optimisation
 - ◆ Impossible in practice

❖ ℓ_p **unit balls** $B_p \equiv \{\mathbf{x} \in \mathbf{R}^N : \|\mathbf{x}\|_p \leq 1\}$



◆ B_p **only convex if** $p \geq 1$

ℓ_1 Minimisation

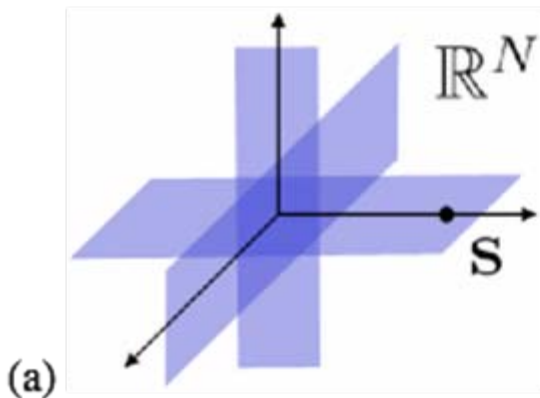
- ❖ Since can't solve ideal but intractable problem, consider associated problem

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x}$$

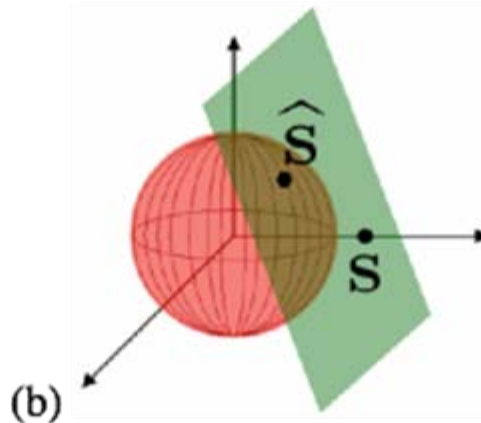
- ❖ Easy to solve numerically
 - ◆ A convex function over a convex set has a unique minimum
 - ◆ Gradient descent methods will converge to minimum
- ❖ Extensive experience in wide range of applications has shown solution is sparse and accurate.
 - ◆ Why?

Intuitive Explanation For Sparseness

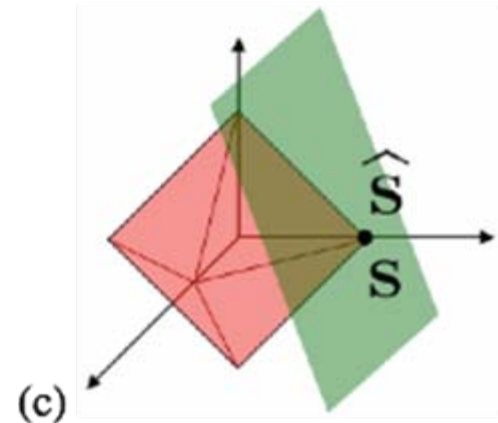
- ❖ In ℓ_1 minimisation solution most likely to occur at corners or edges, not faces
 - ◆ Kuhn-Tucker condition: tangent spaces of ball and subspace have to coincide at solution
 - ◆ Affine subspace of dimension $N-K$ most likely to touch at “edge” defined by K non-zero entries



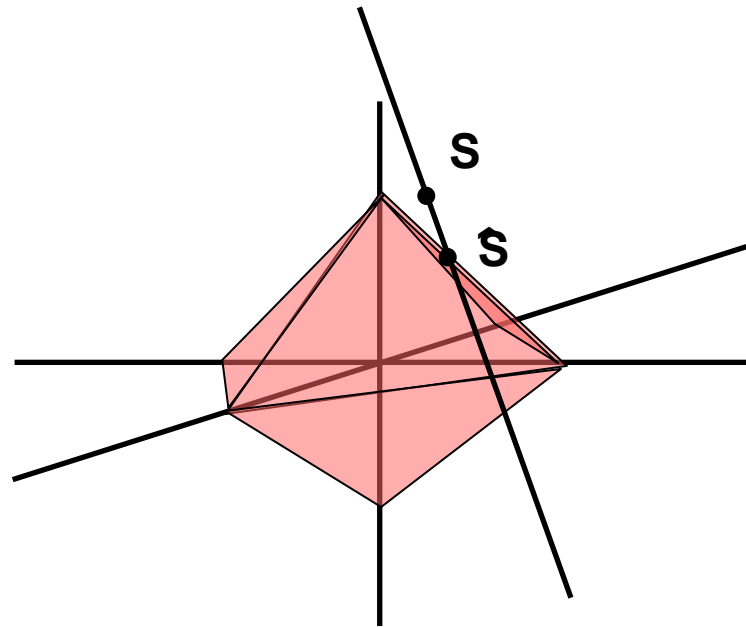
p = 0



p = 2

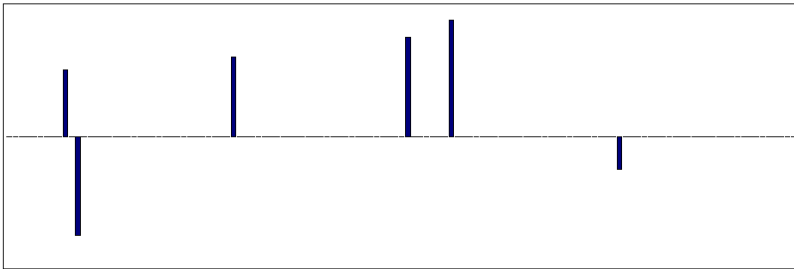


p = 1

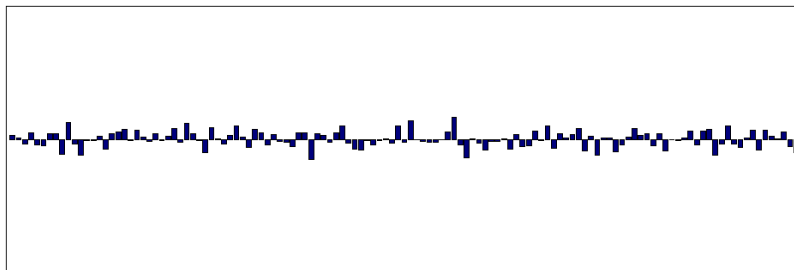


- ❖ Therefore, given K measurements, ℓ_1 minimisation will almost always produce a solution with $N-K$ zeroes
- ❖ Questions
 - ◆ Must this be close to true solution?
 - ◆ If true solution has fewer than K nonzero entries, will ℓ_1 minimiser have same sparsity?

- ❖ In contrast ℓ_2 minimisation tends to spread reconstruction across all coordinates even if true solution is sparse



ℓ_1 solution



ℓ_2 solution $\mathbf{x} = \Phi^T(\Phi\Phi^T)^{-1}\mathbf{y}$

Rigorous Result: I

- ❖ **Theorem (Candes and Tao):** Let \mathbf{x}_M be the best ℓ_1 approximation to \mathbf{x} of length M . Given $K = M \log N/M$ noise-free Gaussian random projections of \mathbf{x} , let \mathbf{x}^* be solution of

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Phi \mathbf{x}$$

Then
$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \frac{1}{\sqrt{M}} \|\mathbf{x} - \mathbf{x}_M\|_1$$

- ❖ **Implications**
 - ◆ ℓ_1 minimisation produces a solution comparable with best approximation
 - ◆ if \mathbf{x} truly sparse (i.e. $\mathbf{x} = \mathbf{x}_M$), ℓ_1 minimisation will reconstruct it exactly from only $O(M)$ measurements

What About Noise?

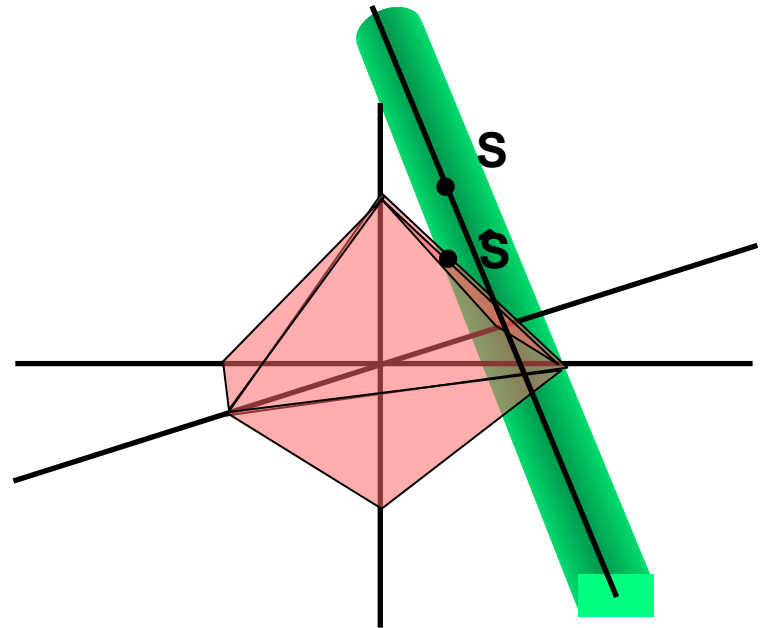
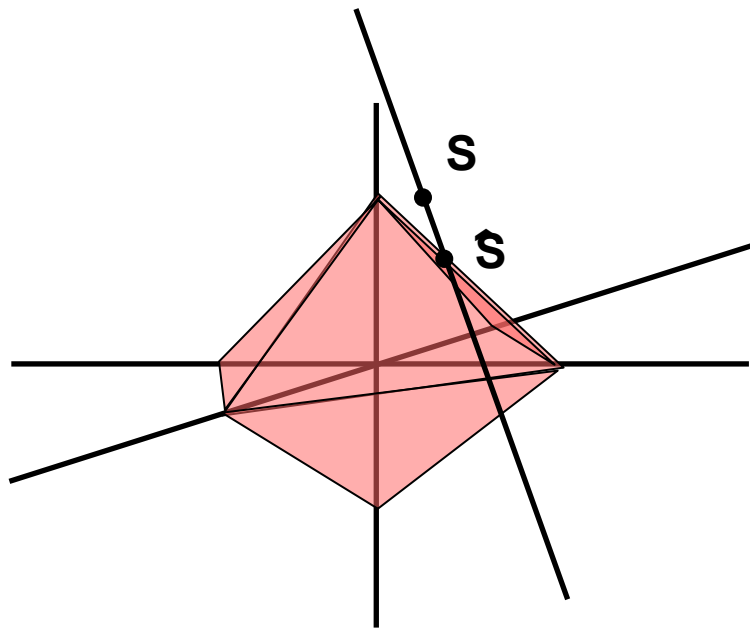
- ❖ Usually data is in error

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

- ◆ Look for sparsest solution within error bounds
 - ◆ Standard error model is Gaussian / sum-of-squares
- ❖ Relax constraint that reconstruction exactly fits data to reconstruction fits data within noise level

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \sigma^2$$

- ❖ Still get sparse solutions



Rigorous Result: II

- ❖ Assume true signal \mathbf{x} is *compressible*; i.e.

$$\exists p, c \quad \ni \quad |x_{(n)}| \leq cn^{-1/p}$$

- ◆ Weaker condition than true sparseness (case when $p = 0$); coefficients decay quickly in given basis

- ❖ *Theorem (Candes and Tao)*: given K appropriate samples, error induced by approximating \mathbf{x} by solution \mathbf{x}^* of

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \sigma^2$$

is bounded by

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq cd_p \left(\frac{K}{\log N/K} \right)^{1/2-1/p}$$

- ❖ **Bound is optimal up to a constant**
 - ◆ **Error in reconstruction from best possible set of K samples not significantly smaller than error from random set**
- ❖ **Theorem holds for very broad class of random samplers, e.g.**

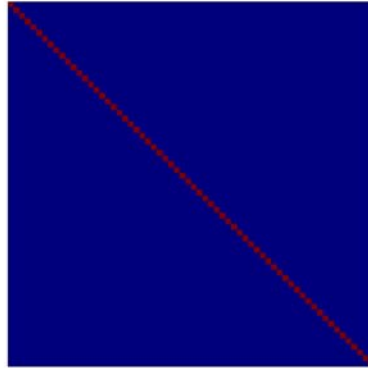
$$[\phi_k]_i \sim \mathcal{N}(0, 1) \quad , \quad [\phi_k]_i \in \{-1, +1\}$$

$$[\phi_k]_n = e^{\frac{2\pi i m_k n}{N}} \quad \text{where} \quad \{m_k\}_{k=1}^K \quad \text{random subset of} \quad \{1, \dots, N\}$$

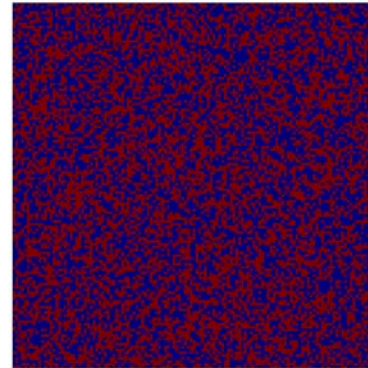
- ◆ **Proof relies on results on distribution of singular values of random matrices**
- ❖ **Latest in line of similar theorems covering variety of situations**
 - ◆ **other players include Donoho and Romberg**

- ❖ Intuitively, sampling vectors must be incoherent with the basis vectors forming sparse representation

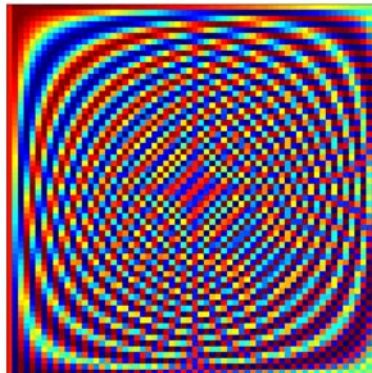
standard basis



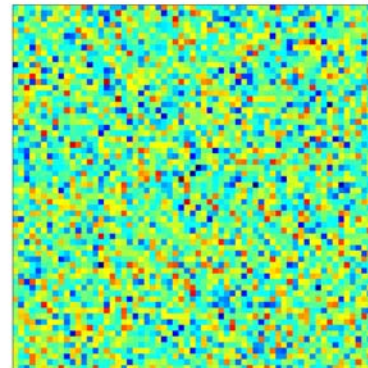
binary code



discrete cosine transform



Gaussian random vectors



“Ah-ha” Insight

- ❖ Do not need to generate random vectors in basis in which \mathbf{x} is sparse
- ❖ Orthogonal transform of i.i.d. vectors remains i.i.d.
- ❖ Therefore generate random vectors in any convenient basis, confident they will transform into appropriate random vectors when used to sample signal

Reconstruction Algorithms

- ❖ **Proposed reconstructions are standard optimisation problems**

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \Phi\mathbf{x} \quad \text{standard linear programming problem}$$

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \leq \sigma^2 \quad \text{quadratically constrained linear programming problem}$$

- ❖ **Variety of algorithms**

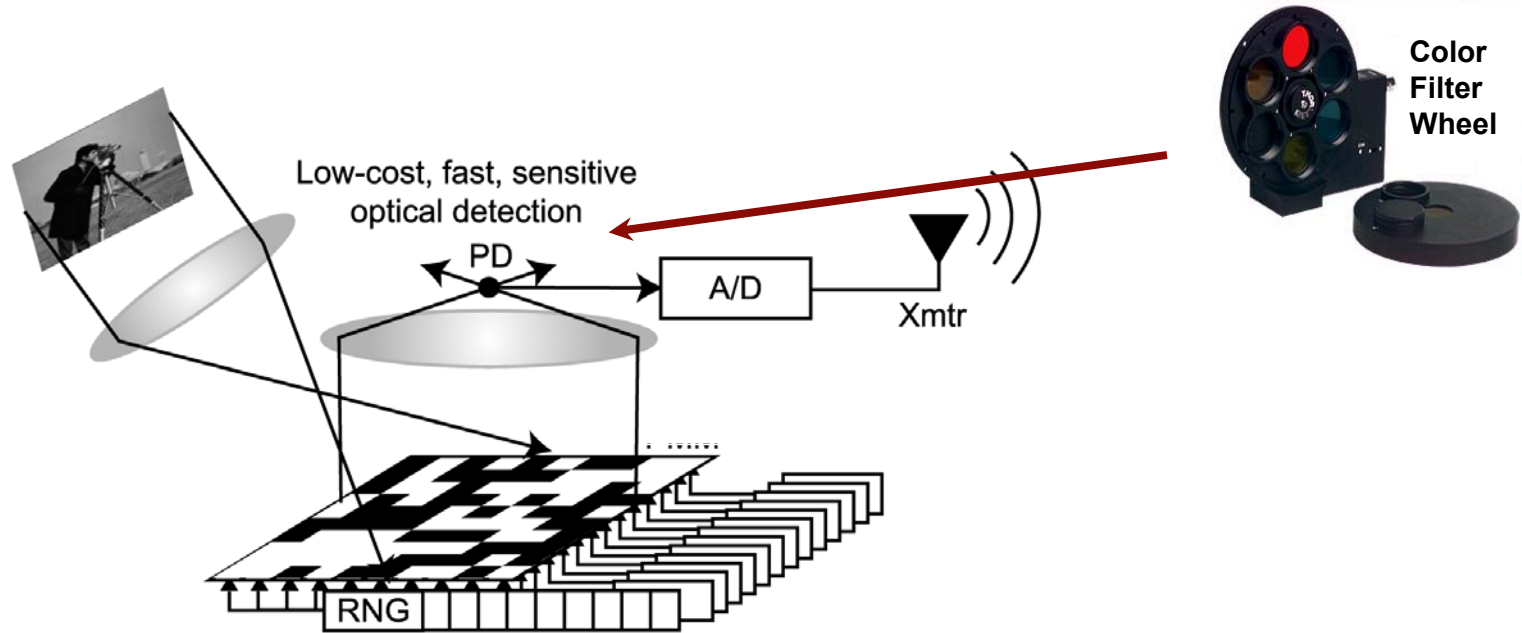
- ◆ **LARS, LASSO, SparseLab, ℓ_1 -Magic, Basis Pursuit, Orthogonal Matching Pursuit**
- ◆ **Common themes**
 - **Convex programming techniques, gradient descent, greedy choice of sparse basis**

- ❖ **Current best performer appears to be GPSR**
 - ◆ **Gradient Projection for Sparse Reconstruction**
 - ◆ **Figueiredo, Nowak and Wright, 2007**
 - ◆ **Simple but fast**

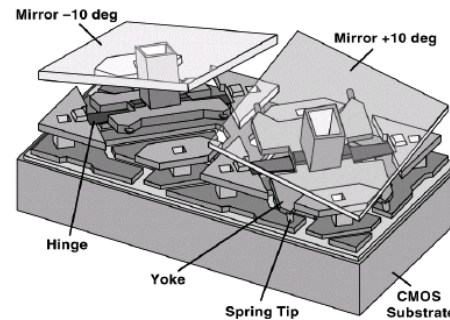
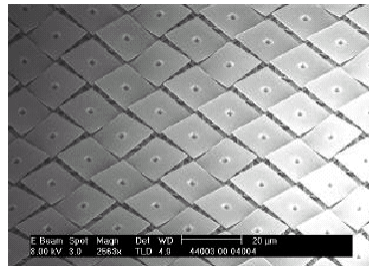
- ❖ **Solves unconstrained mixed norm version of problem**
$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1$$
 - ◆ **Rewrites as a quadratic program with positive variables, solves by steepest descent projected onto positive orthant**

- ❖ **Seems to run in roughly linear time and be >5 x faster than others**

Compressed Camera Revisited



1024 x 728 array

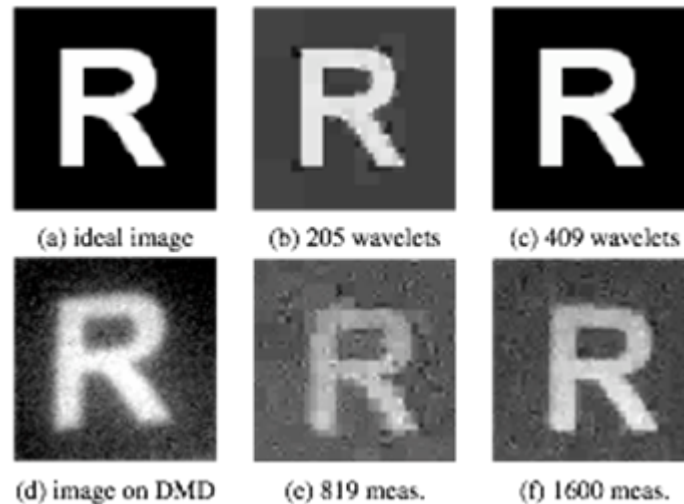


❖ Sampling sequence of random binary masks



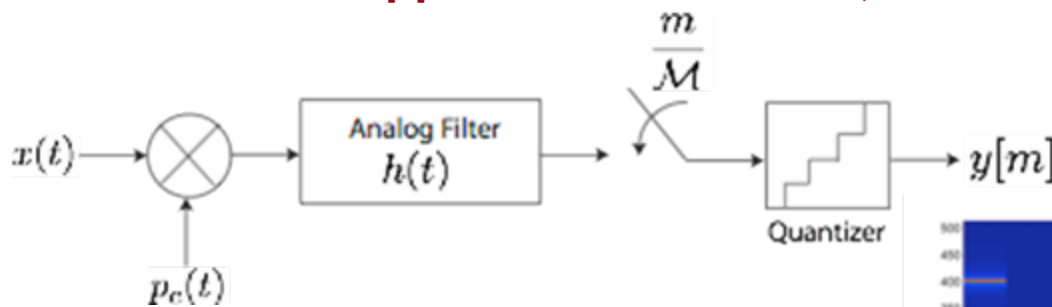
❖ Example reconstruction

image taken at
detector by 320 x 240
pixel camera with
mirrors all flat

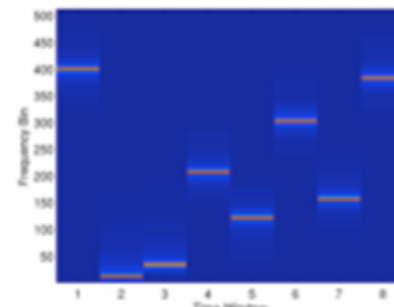


Digitising Wide-band Signals

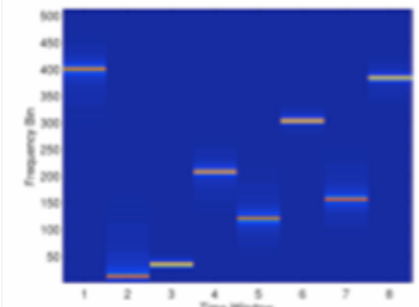
- ❖ Compressed sensing of wide-band signals with sparse time-frequency representations
 - ◆ Chirps, frequency hoppers
- ❖ Team at Rice have designed and simulated compressed sensing digitiser
 - ◆ Approach: randomise, then sub-sample at below Nyquist rate



high frequency
noise



Reconstruction
from samples at
1/4 Nyquist



Developments: Target Detection

- ❖ **Test hypothesis**

$$\mathcal{H}_1 : \mathbf{x} = \mathbf{s} + \mathbf{i} + \mathbf{n} \quad \text{vs.} \quad \mathcal{H}_0 : \mathbf{x} = \mathbf{i} + \mathbf{n}$$

- ❖ **Assume both signal and interference have sparse representations in different, mutually incoherent bases**

$$\mathbf{x} = [\Phi_s : \Phi_i] \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_i \end{bmatrix} + \mathbf{n}$$

- ❖ **Collect random projections, run reconstruction algorithm, declare a detection if $\|\mathbf{x}_s\|$ exceeds specified threshold**

- ❖ **Experiments indicate fewer samples required for reliable estimation of norm than for full reconstruction**

Bayesian Interpretation

- ❖ Laplacian prior introduces ℓ_1 norm in likelihood function, so promotes sparsity
- ❖ Reconstruction by ℓ_1 minimisation now equivalent to MAP (maximum *a posteriori*) estimation

$$\max_{\mathbf{x}} p_G(\mathbf{y} | \mathbf{x}) p_L(\mathbf{x}) \equiv \min_{\mathbf{x}} \left[\frac{\|\mathbf{y} - \Phi\mathbf{x}\|_2^2}{2\sigma^2} + \lambda\|\mathbf{x}\|_1 \right]$$

- ❖ Statistical interpretation provides an alternate paradigm
 - ◆ opens up possibility of improving performance by adaptively choosing sampling strategy to reduce uncertainty in estimates
- ❖ Approach being pursued by Lawrence Carin and group at Duke

Compressed Sensing as Universal Encoding Scheme

- ❖ Encoder / decoder share library of random vectors $\{\phi_k\}_{k=1}^K$
- ❖ Encoder: calculates $\phi_k^T \mathbf{x}$ and quantises to get y_k
- ❖ Decoder: solves $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ subject to $\|\mathbf{y} - \Phi \mathbf{x}\|_2 \leq \sigma^2$
- ❖ Advantages
 - ◆ Sampling and computations are not adaptive
 - ◆ All samples have equal importance
 - Robustness to packet loss, order of receipt not important
 - ◆ Decoding can begin as soon as data stream arrives and continue for as long as data arrives
 - Can take advantage of any scale-invariance properties of signal to compute reconstructions at successive resolutions

Conclusions

- ❖ Theory now provides explanation for why ℓ_1 minimisation gives good reconstructions of solutions known *a priori* to be sparse
 - ◆ Long observed but little understood
- ❖ Provably near optimal performance of random sampling strategy opens up ways of exploiting sparsity by collecting “just enough” data to reconstruct solution
- ❖ Applications seem limited only by researchers’ ingenuity
- ❖ Arousing a lot of interest from a lot of big names
 - ◆ Therefore probably worth looking into