Compressed Sensing for Radio Astronomers or: Why CLEAN Works

Ludwig Schwardt

SKA Calibration & Imaging Workshop
Kiama, Australia
5 March 2014

The Other (Unrelated) Schwarz

Ulrich J. Schwarz

“Mathematical-statistical description of the iterative beam-removing technique (Method CLEAN)”, A&A 65, 345, 1978.

CLEAN does least-squares fit

What’s in a Name?

Results of Google popularity contest

1. “compressed sensing”: 466,000 + Wikipedia entry (+43%)
2. “compressive sensing”: 304,000 (+22%)
3. “compressive sampling”: 98,600 (+38%)
4. “sparse sampling”: 86,900 (+4%)
5. “sparse approximation”: 63,300 (-63%)
6. “compressed sampling”: 9,020 (+4%)

What's in a Name?

Results of Google popularity contest

1. "compressed sensing": 466,000 + Wikipedia entry (+43%)
2. "compressive sensing": 304,000 (+22%)
3. "compressive sampling": 98,600 (+38%)
4. "sparse sampling": 86,900 (+4%)
5. "sparse approximation": 63,300 (-63%)
6. "CLEAN algorithm": 20,700 (+2%)
7. "compressed sampling": 9,020 (+4%)

Motivation for CS

DIGITAL CAMERA

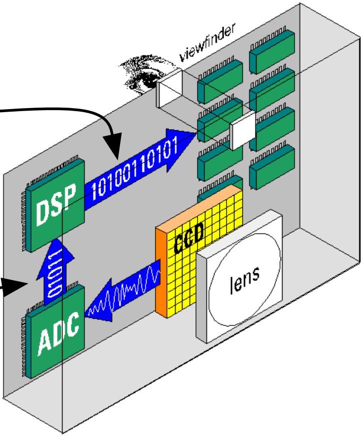
DSP Output:
4 MB JPEG

compress

DSP Input:
36 MB BMP

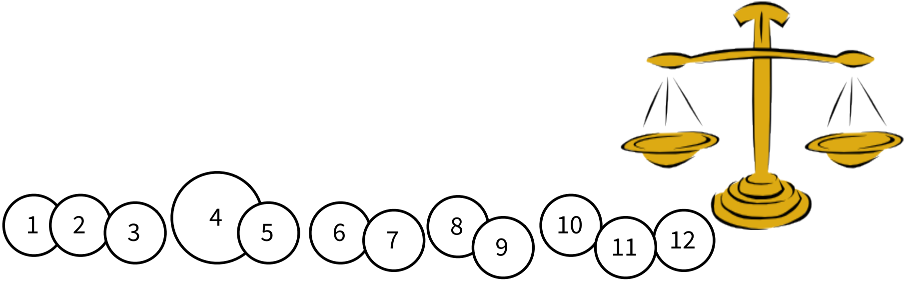
why can't this
be 4 MB?

From Computer Desktop Encyclopedia
© 1998 The Computer Language Co. Inc.

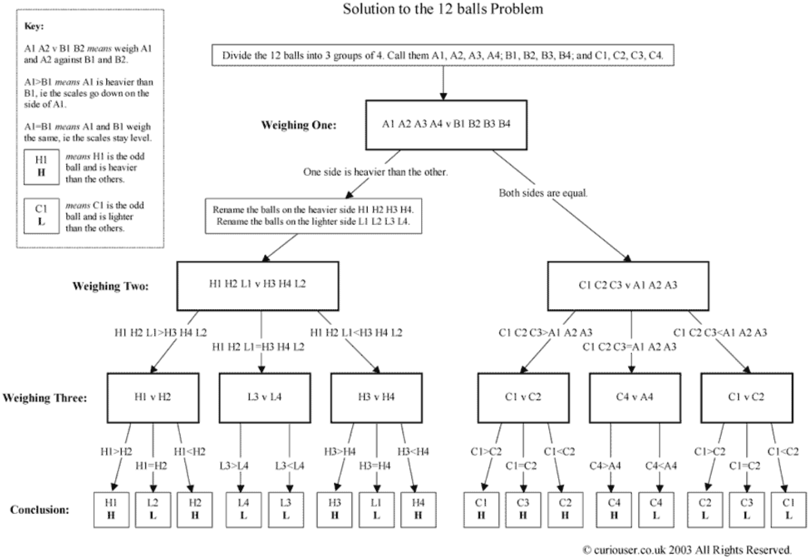


The 12 Balls Problem

- Given 12 balls of which one is heavier or lighter than the rest, find the odd ball using only three (3) weighings on a balance scale
- You will need to weigh **groups of balls** instead of individual balls $\Longrightarrow$ **indirect** measurements



An Elaborate Solution



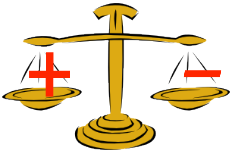
A Simpler Solution

$$
\underset{y}{\begin{bmatrix} 0.2 \\ 0 \\ 0.2 \end{bmatrix}}
=
\underset{A}{\begin{bmatrix}
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\
1 & -1 & 1 & 0 & 1 & 1 & 0 & 0 & -1 & -1 & -1 & 0 \\
1 & -1 & 0 & 1 & -1 & 0 & 1 & 0 & 0 & 1 & -1 & -1
\end{bmatrix}}
\underset{x}{\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{bmatrix}}
$$

N = 12

M = 3

S = 1



Credit: Peter Harrison
(curiouser.co.uk)

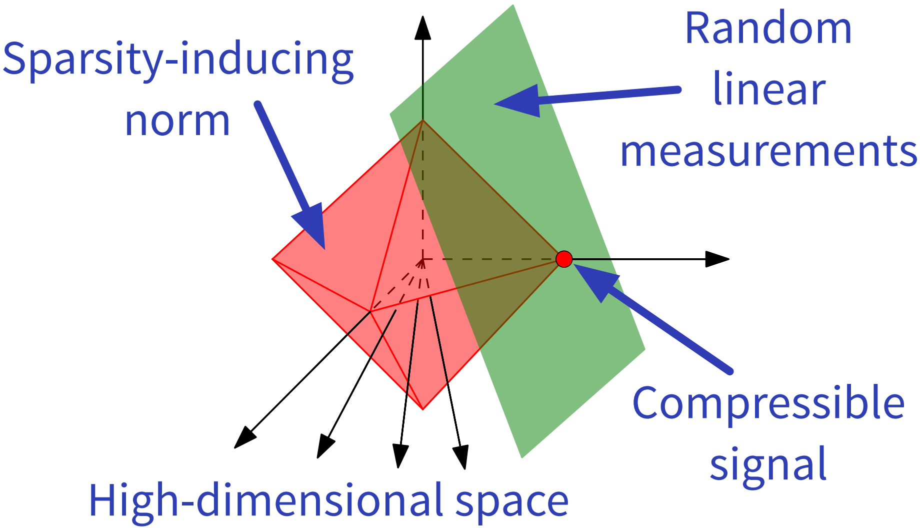
What if the odd ball is heavier?

- We can search through more balls in 3 weighings…

What if the odd ball is heavier?

- We can search through more balls in 3 weighings...
- **27** to be exact, via “ternary” search
- Non-negativity helps!

The Ingredients of CS



The CS Sampling Process

- Sampling is described by a linear **measurement equation**

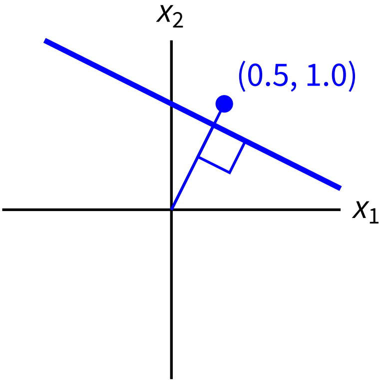
$$y = Ax,$$

  with

  - y a vector of M measurements or samples,
  - x an N-dimensional signal vector and
  - A the $M \times N$ **measurement matrix**

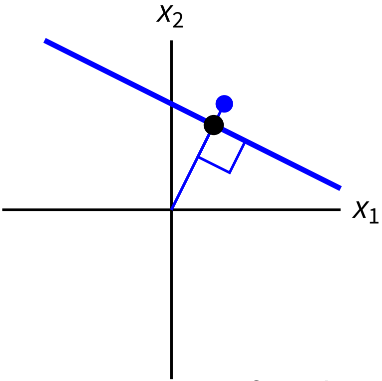
- **Question:** Can x be reconstructed from y even if $M \ll N$?
- **Surprising answer:** Yes, with high probability, as long as A satisfies certain properties and x is S**-sparse** (i.e. it has exactly S non-zero entries)

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



$$\begin{array}{cccc} \boldsymbol{y} & & \boldsymbol{A} & \boldsymbol{x} \\ \begin{bmatrix} 1.0 \end{bmatrix} & = & \begin{bmatrix} 0.5 & 1.0 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{array}$$

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



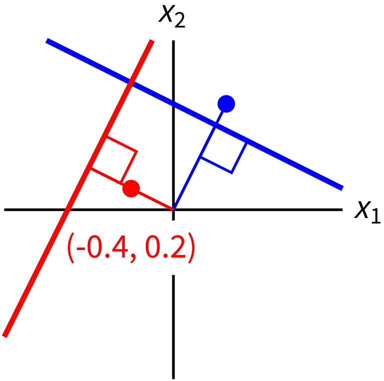
$$\begin{array}{cccc} \boldsymbol{y} & & \boldsymbol{A} & \boldsymbol{x} \\ \begin{bmatrix} 1.0 \end{bmatrix} & = & \begin{bmatrix} 0.5 & 1.0 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{array}$$

Infinitely many solutions!

Pick point closest to origin (pseudo-inverse):

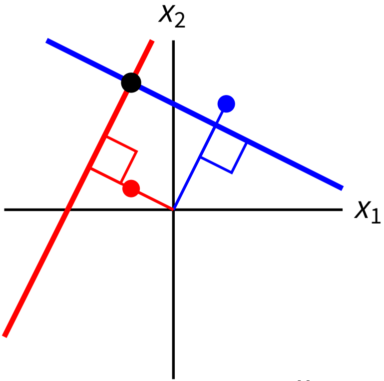
$\boldsymbol{x}^* = \boldsymbol{A}^\dagger \boldsymbol{y} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{y} = (0.4, 0.8)$

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



$$\begin{array}{ccc} \boldsymbol{y} & & \boldsymbol{A} & \boldsymbol{x} \\ \boxed{\begin{matrix} 1.0 \\ 0.4 \end{matrix}} & = & \boxed{\begin{matrix} 0.5 & 1.0 \\ -0.4 & 0.2 \end{matrix}} & \boxed{\begin{matrix} x_1 \\ x_2 \end{matrix}} \end{array}$$

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



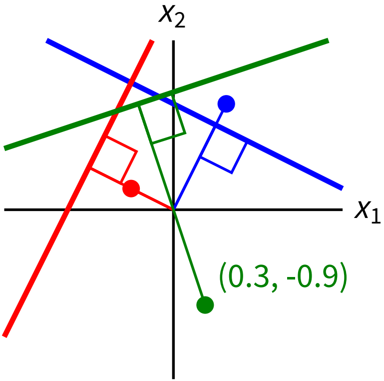
$$\begin{array}{cccc} \boldsymbol{y} & & \boldsymbol{A} & \boldsymbol{x} \\ \begin{bmatrix} 1.0 \\ 0.4 \end{bmatrix} & = & \begin{bmatrix} 0.5 & 1.0 \\ -0.4 & 0.2 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{array}$$

Usually a unique solution!

Pick the standard inverse if it exists:

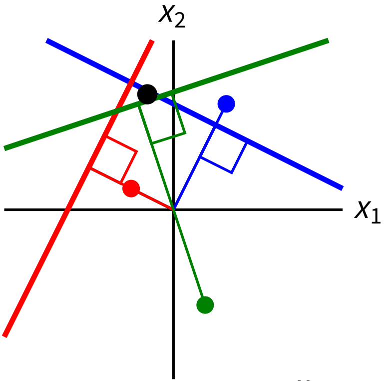
$\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{y} = (-0.4, 1.2)$

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



$$\begin{array}{ccc} \boldsymbol{y} & & \boldsymbol{A} & \boldsymbol{x} \\ \begin{bmatrix} 1.0 \\ 0.4 \\ -1.0 \end{bmatrix} & = & \begin{bmatrix} 0.5 & 1.0 \\ -0.4 & 0.2 \\ 0.3 & -0.9 \end{bmatrix} & \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \end{array}$$

Recap: What does $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ mean?



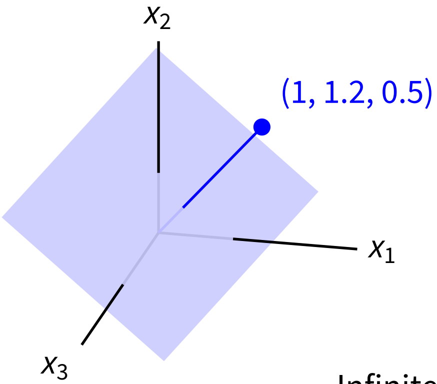
$$\overset{\boldsymbol{y}}{\begin{bmatrix} 1.0 \\ 0.4 \\ -1.0 \end{bmatrix}} = \overset{\boldsymbol{A}}{\begin{bmatrix} 0.5 & 1.0 \\ -0.4 & 0.2 \\ 0.3 & -0.9 \end{bmatrix}} \overset{\boldsymbol{x}}{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

Usually no solution!

Pick point closest to all lines (pseudo-inverse):

$\boldsymbol{x}^* = \boldsymbol{A}^\dagger \boldsymbol{y} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{y} = (-0.2471, 1.0903)$

Going up one dimension

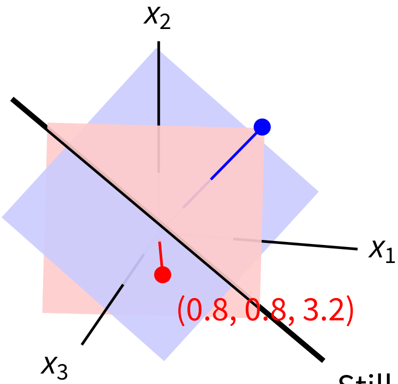


$$\underset{\boldsymbol{y}}{\boxed{0.6}} = \underset{\boldsymbol{A}}{\boxed{\begin{matrix} 1.0 & 1.2 & 0.5 \end{matrix}}} \underset{\boldsymbol{x}}{\boxed{\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}}}$$

Infinitely many solutions!

All points on blue plane...

Going up one dimension

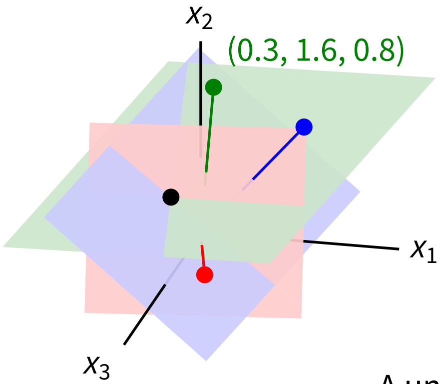


$$
\begin{array}{ccccc}
\boldsymbol{y} & & \boldsymbol{A} & & \boldsymbol{x} \\
\begin{bmatrix} 0.6 \\ 1.6 \end{bmatrix} & = & \begin{bmatrix} 1.0 & 1.2 & 0.5 \\ 0.8 & 0.8 & 3.2 \end{bmatrix} & & \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\end{array}
$$

Still infinitely many solutions!

All points on black line...

Going up one dimension



$$\begin{array}{ccc} y & A & x \\ \boxed{\begin{matrix} 0.6 \\ 1.6 \\ 1.0 \end{matrix}} = & \boxed{\begin{matrix} 1.0 & 1.2 & 0.5 \\ 0.8 & 0.8 & 3.2 \\ 0.3 & 1.6 & 0.8 \end{matrix}} & \boxed{\begin{matrix} x_1 \\ x_2 \\ x_3 \end{matrix}} \end{array}$$

A unique solution!

$$\boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{y} = (-0.14, 0.44, 0.43)$$

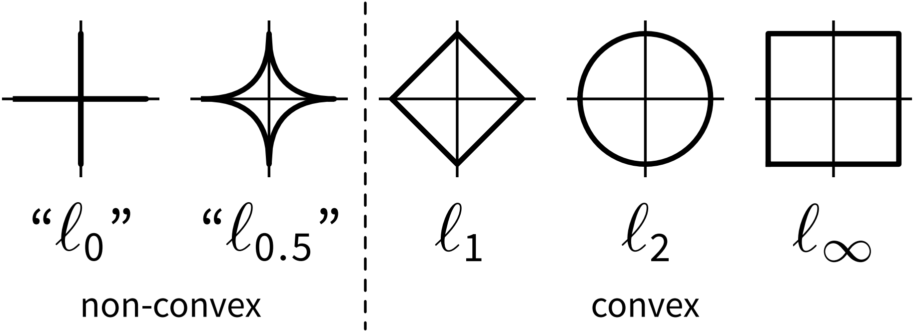
Measuring Distances: Norms

General ℓ_p-norm: $$\|\boldsymbol{x}\|_p := \left(\sum_{n=1}^{N} |x_n|^p\right)^{1/p}$$

- **Euclidean**: $\|\boldsymbol{x}\|_2 := \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_N|^2}$
- **Manhattan**: $\|\boldsymbol{x}\|_1 := |x_1| + |x_2| + \cdots + |x_N|$
- **ℓ_0-pseudonorm**: $\|\boldsymbol{x}\|_0 := |\{n : x_n \neq 0\}|$
  number of non-zero elements of $\boldsymbol{x} \Longrightarrow$ **sparsity!**
- **Chebyshev / max-norm**: $\|\boldsymbol{x}\|_\infty := \max_n |x_n|$

Distance Contours: Unit Spheres

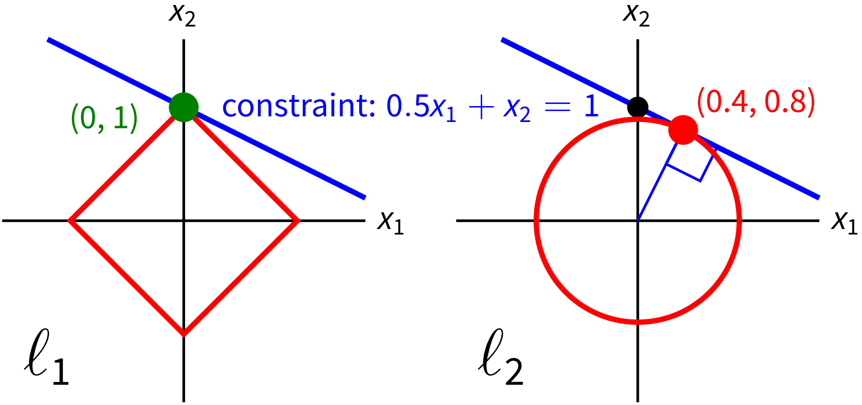
On a unit "sphere" we have $\|\boldsymbol{x}\|_p = 1$
(set of all points at the same distance from origin)
Inside of unit sphere $\Longrightarrow$ **unit ball**



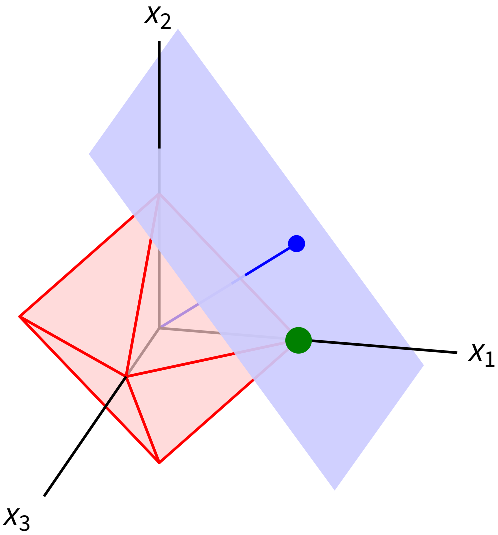
ℓ_1 Promotes Sparsity (Unlike ℓ_2)

Solve: $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_p$ subject to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$

$\Longrightarrow$ Pseudoinverse solution $\boldsymbol{x}_{\mathrm{PI}} = \boldsymbol{A}^\dagger \boldsymbol{y}$ is a **bad idea**



Going up one dimension



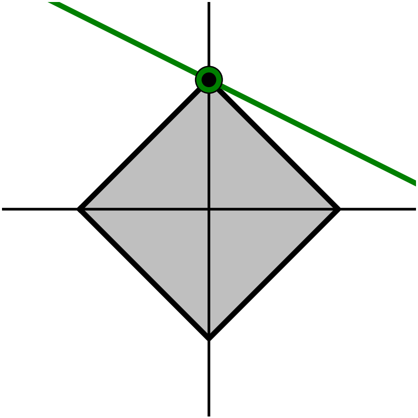
Recover Sparse Signal From Random Measurements

$N = 2$

$M = 1$

$S = 1$

SUCCESS



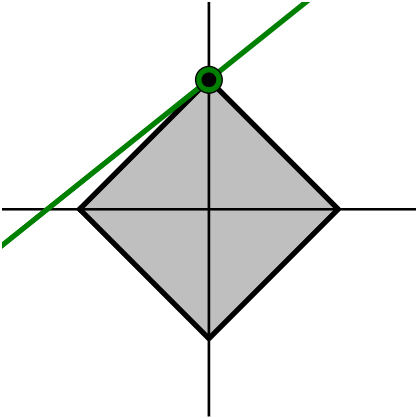
Recover Sparse Signal From Random Measurements

$N = 2$

$M = 1$

$S = 1$

SUCCESS



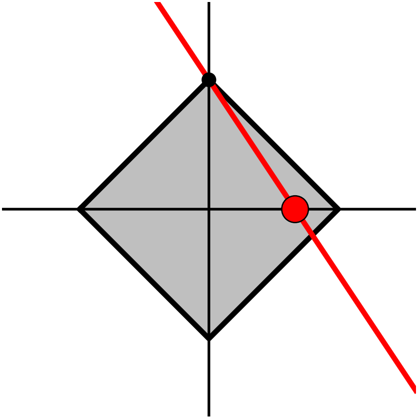
Recover Sparse Signal From Random Measurements

$N = 2$

$M = 1$

$S = 1$

FAILURE



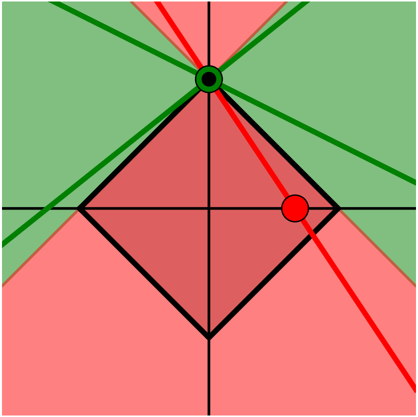
Recover Sparse Signal From Random Measurements

$N = 2$

$M = 1$

$S = 1$

SUCCESS RATE = 50%

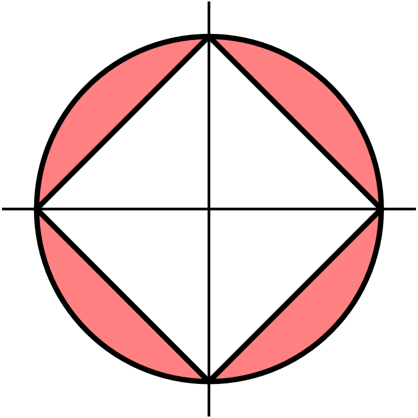


Recovery in High Dimensions

- **Much better!**
- Consider the size of ℓ_1–ball vs ℓ_2–ball
- For $N = 2, 3, 4, \ldots$:

$$\frac{V(\ell_1)}{V(\ell_2)} = \frac{2}{\pi}, \frac{1.3}{4.2}, \frac{0.7}{4.9}, \ldots$$

- We want **small spindly balls** that are hard to pierce



The Gory Details of Why

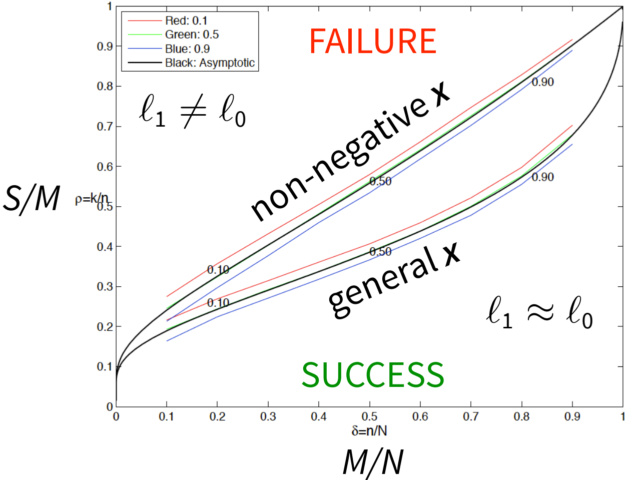
- **Candès, Romberg, Tao**, “Stable Signal Recovery from Incomplete and Inaccurate Measurements,” Comm. Pure Appl. Math., vol. 59, no. 8, pp. 1207–1223, 2006 and Candès-Tao references
  random matrix theory, Banach space geometry
- **Donoho**, “Compressed Sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006
  polytope geometry, *k*-neighborliness, Gel’fand widths

The Measurement Matrix *A*

- Compressed sensing projects the desired signal onto a **few random basis functions**, instead of many shifted impulses
- Good choices for *A* include:
    - **Gaussian matrix** with i.i.d. normal random entries
    - **Bernoulli matrix** with i.i.d. Bernoulli random entries
    - **Partial Fourier matrix** with rows drawn at random from DFT matrix (random frequencies)

Success vs Failure

Donoho-Tanner phase transition indicates where in parameter space successful recovery becomes possible



Credit: Donoho, Tanner, 2010.

But What If...

- **there is measurement noise?**
  - CS techniques are stable $\Longrightarrow$ reconstruction errors bounded
- **signal is smooth instead of sparse?**
  - Maybe the gradient is sparse $\Longrightarrow$ use different **TV-norm**
  - Represent signal in different basis where it will be sparse (e.g. wavelets) $\Longrightarrow$ problem changes to $y = AWx$
- **signal is only approximately sparse?**
  - CS works if signal representation is **compressible** with amplitudes decaying according to power law
  - Most natural signals are compressible!

CS for Radio Astronomy

- Consider simplified imaging equation expressing visibilities V in terms of image brightness I,

$$V(u_j, v_j) = \sum_{k=1}^{N} I(l_k, m_k) e^{-i2\pi(u_j l_k + v_j m_k)}$$

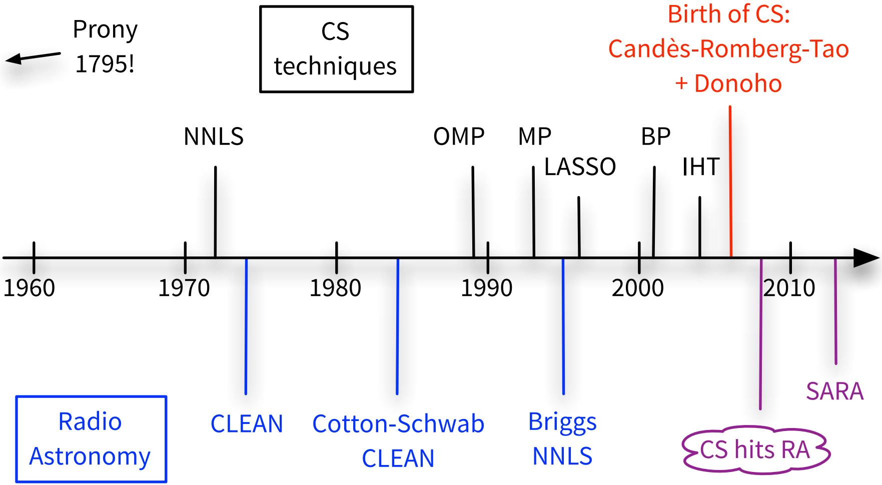
- In matrix form it becomes $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, with M visibilities $y_j = V(u_j, v_j)$, N image pixels $x_k = I(l_k, m_k)$ and matrix entries $a_{jk} = \exp\{-i2\pi(u_j l_k + v_j m_k)\}$
- **Natural fit to CS**: the interferometer does random projections for you! (similar situation in MRI)

Reconstruction Algorithms

Various classes of CS algorithms exist, of which the most popular are:

- **Convex relaxation** (BP, NESTA, SARA, ...)
- **Greedy methods** (CLEAN, MP, OMP, CoSaMP, ...)
- Iterative thresholding (IHT, AMP, FISTA, ...)
- Combinatorial algorithms (chaining pursuit, Heavy-Hitters on Steroids (HHS), ...)
- Bayesian methods (MAP with Laplacian prior...)

Practice Precedes Theory



Just Relax: Basis Pursuit (BP)

- Ideal sparse reconstruction minimises $\|\boldsymbol{x}\|_0$ while being consistent with the measurements $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$
- This is intractable, so use next best norm instead, which is the ℓ_1 **norm** $\Longrightarrow$ **convex relaxation** of ℓ_0
- **Basis Pursuit** solves the convex optimisation problem

$$\text{(BP)} \qquad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$

Handling Noise in Basis Pursuit

- For noisy measurements, change to one of

$$(\mathrm{BP}_\epsilon) \qquad \min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \epsilon$$

$$(\mathrm{QP}_\lambda) \qquad \min_{\boldsymbol{x}} \left(\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1\right)$$

- **Quadratic Program** QP_λ is least-squares with ℓ_1 regularisation
- Tune parameters ϵ and λ based on SNR
- Easy to add constraints such as non-negativity of $\boldsymbol{x}$, e.g. **BP**$_\epsilon+$ and **QP**$_\lambda+$

Greedy: Matching Pursuit (MP)

- Views recovery problem as finding a sparse representation for the $M \times 1$ **measurement vector** $\boldsymbol{y} = \sum_{j=1}^{N} x_j \boldsymbol{a}_j$, based on the **columns** $\boldsymbol{a}_j$ of $\boldsymbol{A}$ (i.e. only a few x_j terms are non-zero)
- MP terminology: $\boldsymbol{A}$ is **dictionary** of **atoms** $\boldsymbol{a}_j$
- MP approximately solves the problem

$$\text{(MP)} \qquad \min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_0 \leq S$$

Matching Pursuit Algorithm

- Initialise **residual** $\boldsymbol{r}^{(0)} = \boldsymbol{y}$
- At kth iteration, select atom which fits residual best, as $\boldsymbol{a}^{(k)} = \arg\max_{\boldsymbol{a}} \left|\langle \boldsymbol{r}^{(k)}, \boldsymbol{a}\rangle\right|$, which amounts to picking the peak of $\left|\boldsymbol{A}^H \boldsymbol{r}^{(k)}\right|$
- Update residual to $\boldsymbol{r}^{(k+1)} = \boldsymbol{r}^{(k)} - a_k \boldsymbol{a}^{(k)}$, with $a_k = \langle \boldsymbol{r}^{(k)}, \boldsymbol{a}\rangle$
- Stop when residual becomes small enough
- Recovered signal has non-zero entries a_k at locations of selected atoms

Orthogonal Matching Pursuit

- This is identical to MP, but adds a least-squares fit step after selecting a new atom, which readjusts the amplitudes of all atoms to best fit the data
- Easy to add non-negativity constraint **(OMP+)**
- In practice, OMP is preferred to plain MP, as it converges faster
- OMP is typically faster than BP and simpler to code
- BP problem is convex $\Longrightarrow$ single global optimum

Relating CLEAN to CS

- Högbom CLEAN is identical to MP, but forms residual in image space instead of in measurement (uv) space
- Clark CLEAN subtracts multiple components in one iteration $\Longrightarrow$ many MP variants such as ROMP and StOMP do too
- Cotton-Schwab CLEAN actually operates in measurement (uv) space like standard MP
- CLEAN loop gain idea not prevalent in MP literature $\Longrightarrow$ rather rebalances components as in OMP

Relating NNLS to CS

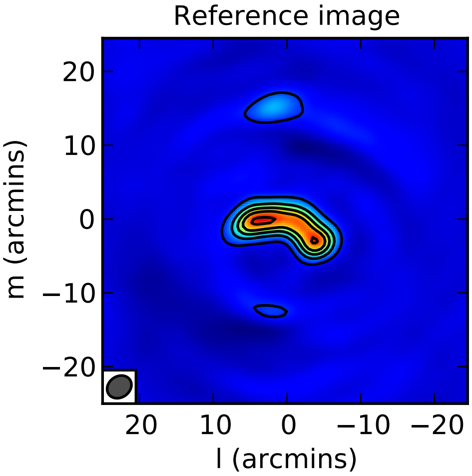
- Consider Non-Negative Least Squares (Briggs, 1995)
- NNLS is identical to OMP with non-negativity constraint, but operates in the image domain instead of uv domain, solving

$$\boldsymbol{A}^H \boldsymbol{y} = \boldsymbol{A}^H \boldsymbol{A} \boldsymbol{x} \quad \text{subject to} \quad \|\boldsymbol{x}\|_0 \leq S \quad \text{and} \quad \boldsymbol{x} \geq 0$$

- This explains the tendency of NNLS to compact flux
- The CS version improves on standard NNLS by operating directly in uv domain: improved accuracy and reduced memory usage ($M \times S$ instead of $N \times N$)
- Standard OMP fits in between CLEAN and NNLS

Observation

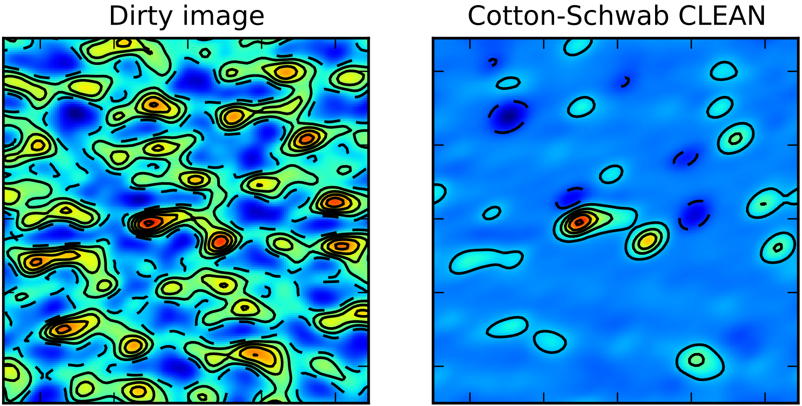
- PKS 1610-60 galaxy
- 12.8 hours at 1822 MHz
- Flagged, calibrated and averaged in MIRIAD (Laura Richter)
- $M = 94390$ visibilities
- Made 100×100 image in CASA with Cotton-Schwab CLEAN (3' restoring beam)
- CLEAN and CS methods very similar since $M > N$



Experimental Setup

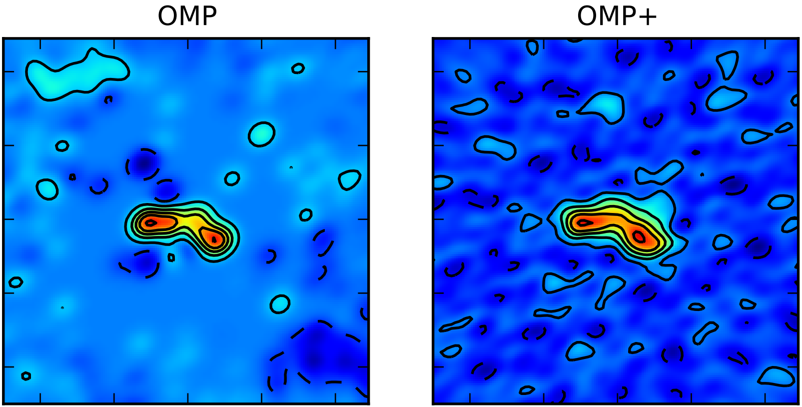
- Up the challenge: selected 10-minute segment to produce **snapshot image**
- $N = 10000$ pixels, $M = 1140$ measurements, about $S = 200$ components
- Methods tested via CASA and compsense:
    - Cotton-Schwab CLEAN (loop gain 0.1, max 2000 iters)
    - OMP, OMP$+$ (max 200 iterations)
    - QP$_\lambda$, QP$_\lambda +$ (λ automatically tuned to reflect SNR)
    - BP$_\epsilon$, BP$_\epsilon +$ (ϵ automatically tuned to reflect SNR)

Results: Standard CLEAN



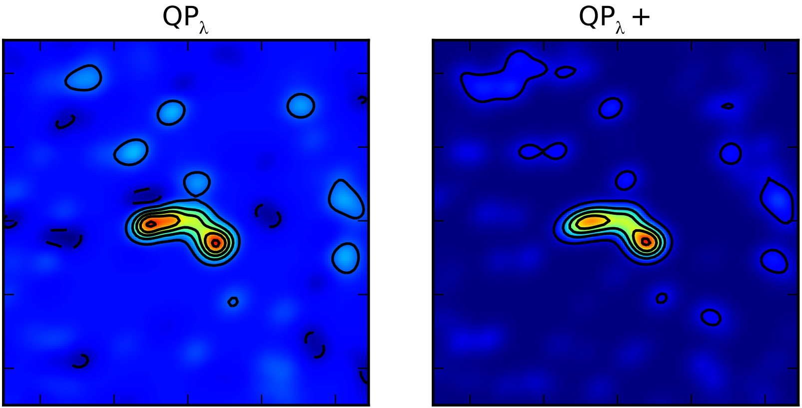
CLEAN does not pick up central part of galaxy

Results: OMP and OMP+



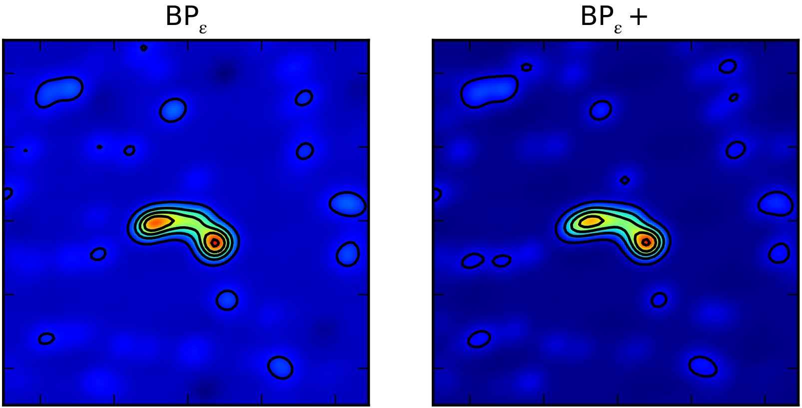
OMP+ has small, lumpy residual (very few components)

Results: QP_λ and $QP_\lambda+$



Good correspondence with reference image

Results: BP_ϵ and $BP_\epsilon+$



Good correspondence with reference image

Sparsity, Dynamic Range, CPU

Method	# Comps	DR	CPU time (s)
Cotton-Schwab	188	16.5	3.6
OMP	105	11.4	6.4
OMP+	39	16.4	15.1
QP_λ	98	24.5	47.9
$QP_\lambda+$	212	42.4	46.8
$(BP_\epsilon$	119	37.4	235.6)
$BP_\epsilon+$	145	36.8	76.2

Conclusions

CLEAN works because:

- Astronomical images consist of point sources and blobs with amplitudes that decay according to a power law
- Telescopes produce indirect measurements that are semi-random in Fourier plane
- CLEAN is a version of matching pursuit that approximately solves the CS reconstruction problem

Exciting time for deconvolution - new algorithms, performance guarantees