CoSaMP

Iterative signal recovery from incomplete and inaccurate samples

Joel A. Tropp

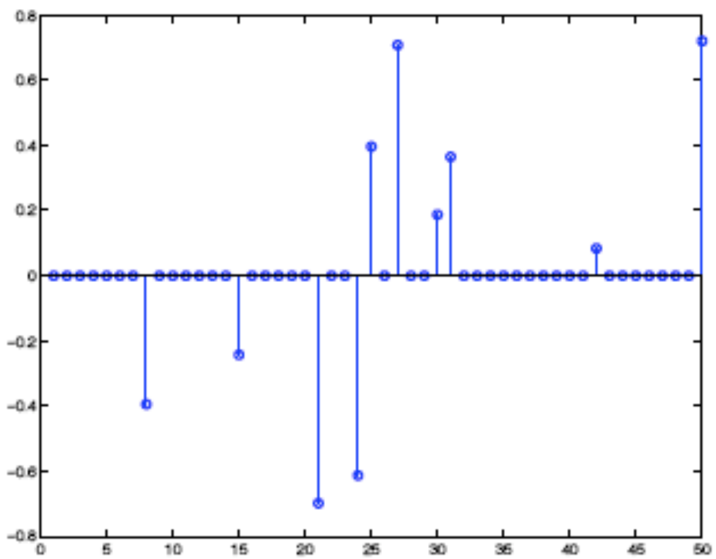
Applied & Computational Mathematics
California Institute of Technology
jtropp@acm.caltech.edu

Joint with D. Needell (UC-Davis).

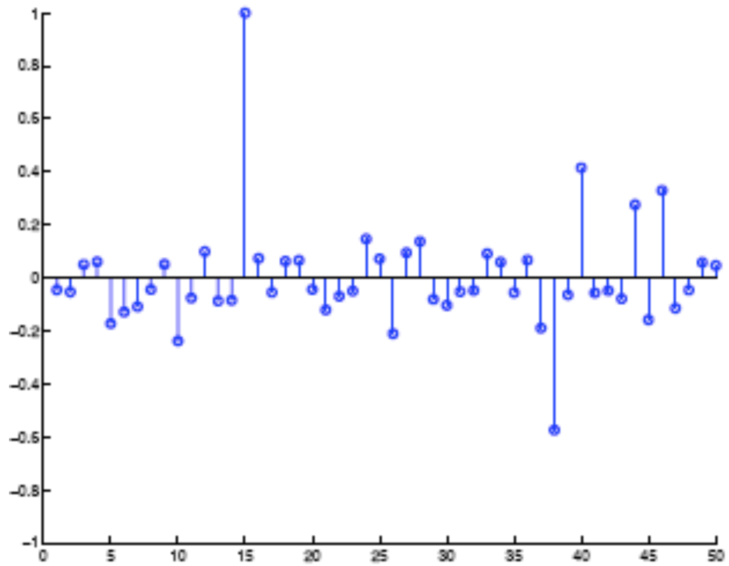
Research supported in part by NSF and DARPA

The Sparsity Heuristic

A *sparse signal* has fewer degrees of freedom than its nominal dimension



Sparse signal

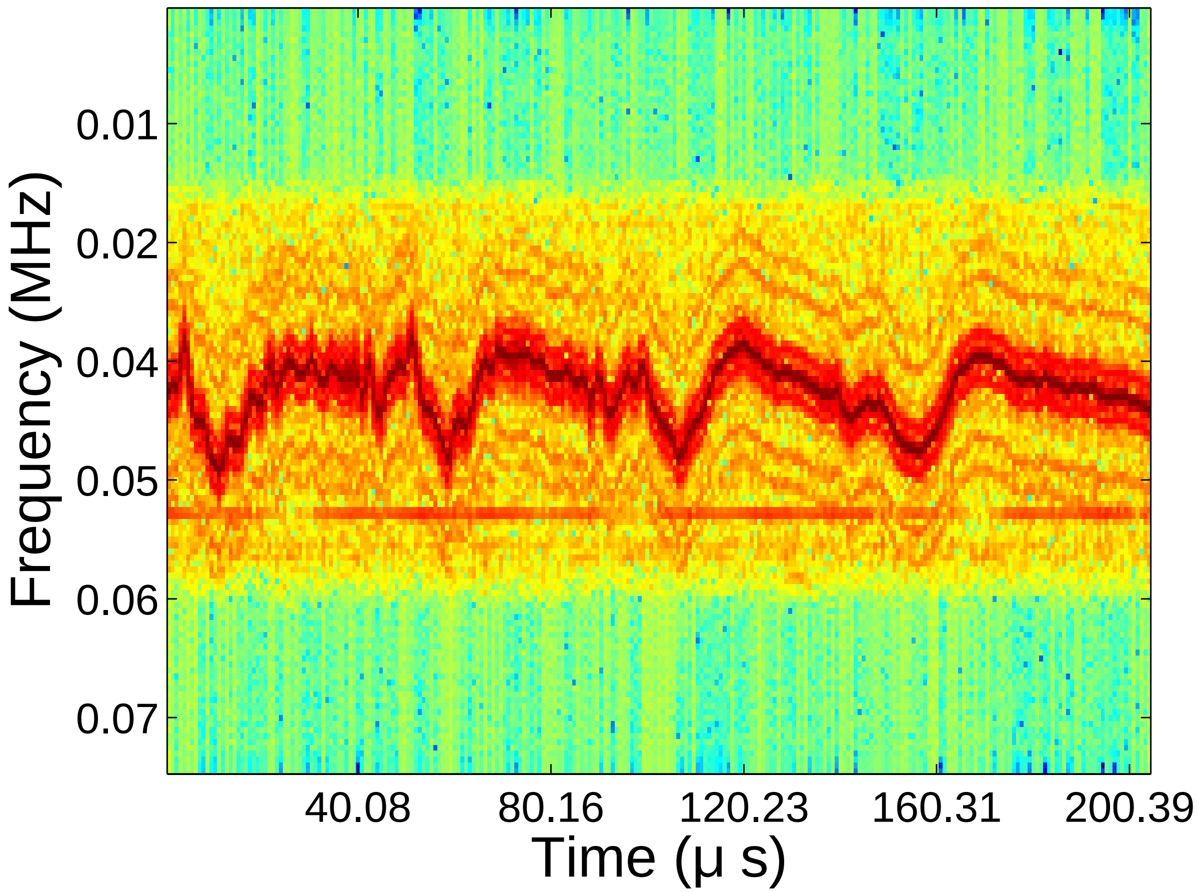


Nearly sparse signal

Example: Wavelet Sparsity

Courtesy of J. Romberg

Example: Time–Frequency Sparsity



Data provided by L3 Communications

Quantifying Sparsity

- Let $\{\boldsymbol{\psi}_k : k = 1, 2, \dots, N\}$ be an orthobasis for $\mathbb{R}^N$
- The coefficients of $\boldsymbol{x}$ with respect to the basis are

$$f_k = \langle \boldsymbol{x},\ \boldsymbol{\psi}_k \rangle \qquad \text{for } k = 1, 2, \dots, N$$

- The signal is *s-sparse* when $\#\{k : f_k \neq 0\} \leq s$
- Generalization: the signal is *p-compressible* with magnitude R if

$$|f|_{(k)} \leq R \cdot k^{-1/p} \qquad \text{for } k = 1, 2, \dots, N$$

- p-compressible is slightly weaker than "in ℓ_p" for each $p > 0$

Approximating Compressible Signals

- Consider a signal p-compressible w.r.t. the standard basis

$$|x|_{(k)} \le R \cdot k^{-1/p} \qquad \text{for } k = 1, 2, 3, \dots$$

- Approximating $\boldsymbol{x}$ by its s largest terms gives error

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{x}_s\|_2 &\le R \cdot \left[\sum\nolimits_{k>s} k^{-2/p} \right]^{1/2} \\ &\approx R \cdot \left[\int_s^\infty u^{-2/p} \, \mathrm{d}u \right]^{1/2} \approx R \cdot s^{1/2-1/p} \end{aligned}$$

- Compressible signals are well approximated by sparse signals

- Fundamental idea behind transform coding

Counting Bits

- Consider the class of 0–1 signals in $\mathbb{R}^N$ with exactly s ones
- Clearly need *at least* $\log_2 \binom{N}{s}$ bits to distinguish signals
- By Stirling's approximation, about $s \log(N/s)$ bits
- When $s \ll N$, signals contain much less information than the ambient dimension suggests
- A simple *adaptive* coding scheme can achieve this rate

What is a Sample?

- A *sample* is the value of a linear functional applied to the signal
- Examples:
  - CCD: Point intensity of an image
  - ADC: Voltage of an electrical signal at a point in time
  - MRI: Frequency in the 2D Fourier transform of an image
  - CAT: Line integral of density in one direction
- Some of these technologies acquire samples in batches
- We wish to acquire signals with as few samples as possible

Compressive Sampling and Signal Recovery

- Design linear sampling operator $\mathbf{\Phi} : \mathbb{C}^N \to \mathbb{C}^m$
- Suppose $\boldsymbol{x}$ is an unknown (compressible) signal in $\mathbb{C}^N$
- Collect noisy samples $\boldsymbol{u} = \mathbf{\Phi}\boldsymbol{x} + \boldsymbol{e}$
- Problem: Given samples $\boldsymbol{u}$, approximate $\boldsymbol{x}$

Restricted Isometries

- Abstract property of sampling operator supports efficient sampling
- $\mathbf{\Phi}$ has the *restricted isometry property* of order $2s$ when

$$(1-\mathrm{c})\,\|\boldsymbol{x}\|_2^2 \le \|\mathbf{\Phi}\boldsymbol{x}\|_2^2 \le (1+\mathrm{c})\,\|\boldsymbol{x}\|_2^2 \qquad \text{whenever } \|\boldsymbol{x}\|_0 \le 2s$$

- $\mathbf{\Phi}$ preserves geometry of s-sparse signals (take $\boldsymbol{x} = \boldsymbol{y} - \boldsymbol{z}$)
- W.h.p., a Gaussian sampling operator has RIP$(2s)$ when

$$m \ge \mathrm{C}s\log(N/s)$$

- Gaussian matrices are practically useless

References: [Candès–Tao 2006, Rudelson–Vershynin 2006]

Practical Sampling Operators

- Partial Fourier matrices [CRT 2006]
  - Each row of $\mathbf{\Phi}$ is chosen at random from rows of unitary DFT $\mathscr{F}_N$
- Random demodulator [Rice DSP 2006]

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \dots & 1 & & & & \\ & & & 1 & \dots & 1 & \\ & & & & & & \ddots \end{bmatrix}_{m \times N} \begin{bmatrix} \pm 1 & & \\ & \pm 1 & \\ & & \ddots \end{bmatrix}_{N \times N} \mathscr{F}_N$$

- W.h.p., both have RIP$(2s)$ when $m \geq \mathrm{C} s \log^{\alpha} N$
- Certain technologies can acquire these samples efficiently
- Fast matrix–vector multiplies!

Desiderata for Recovery Algorithm

- Works for general sampling schemes
- Succeeds with minimal number of samples
- Tolerates noise in samples
- Produces approximations with optimal error bound
- Yields rigorous guarantees on resource requirements
- Exploits structured sampling matrices

$\textsc{CoSaMP}(\boldsymbol{\Phi},\ \boldsymbol{u},\ s)$

Input: Sampling operator $\boldsymbol{\Phi}$, noisy sample vector $\boldsymbol{u}$, sparsity level s

Output: An s-sparse approximation $\boldsymbol{a}$ of the target signal

---

$k = 0$	{ Initialization }
$\boldsymbol{a}^k = \boldsymbol{0}$	
while halting criterion *false*	
$\quad \boldsymbol{v} \leftarrow \boldsymbol{u} - \boldsymbol{\Phi}\boldsymbol{a}^k$	{ Update samples }
$\quad \boldsymbol{y} \leftarrow \boldsymbol{\Phi}^* \boldsymbol{v}$	{ Form signal proxy }
$\quad \Omega \leftarrow \mathrm{supp}(\boldsymbol{y}_{2s})$	{ Identification }
$\quad T \leftarrow \Omega \cup \mathrm{supp}(\boldsymbol{a}^k)$	{ Merge supports }
$\quad \boldsymbol{b}\vert_T \leftarrow \boldsymbol{\Phi}_T^\dagger \boldsymbol{u}$	{ Signal estimation by least squares }
$\quad \boldsymbol{b}\vert_{T^c} \leftarrow \boldsymbol{0}$	
$\quad \boldsymbol{a}^{k+1} \leftarrow \boldsymbol{b}_s$	{ Prune to obtain next approximation }
$\quad k \leftarrow k + 1$	
end *while*	
$\boldsymbol{a} \leftarrow \boldsymbol{a}^k$	{ Return final approximation }

Cost per Iteration

- Update samples and form signal proxy:

$$\boldsymbol{y} \leftarrow \boldsymbol{u} - \boldsymbol{\Phi}\boldsymbol{a}^k \qquad \text{and} \qquad \boldsymbol{y} \leftarrow \boldsymbol{\Phi}^*\boldsymbol{v}$$

- One matrix–vector multiplication each
- Signal approximation by least squares:

$$\boldsymbol{b}_T \leftarrow \boldsymbol{\Phi}_T^\dagger \boldsymbol{u}$$

- Use conjugate gradient to apply pseudoinverse
- Each iteration requires two matrix–vector mulitplies
- Assuming RIP($2s$), constant number of iterations for fixed accuracy

- Constant number of matrix–vector multiplies per CoSaMP iteration!

Performance Guarantee

Theorem 1. [CoSaMP] *Suppose that*

- *the sampling matrix* $\boldsymbol{\Phi}$ *has* $\mathrm{RIP}(2s)$,
- *the sample vector* $\boldsymbol{u} = \boldsymbol{\Phi x} + \boldsymbol{e}$,
- η *is a precision parameter,*
- $\mathscr{L}$ *bounds cost of a matrix–vector multiply with* $\boldsymbol{\Phi}$ *or* $\boldsymbol{\Phi}^*$.

Then CoSaMP produces a $2s$*-sparse approximation* $\boldsymbol{a}$ *such that*

$$\|\boldsymbol{x} - \boldsymbol{a}\|_2 \le \mathrm{C} \max \left\{ \eta, \frac{1}{\sqrt{s}} \|\boldsymbol{x} - \boldsymbol{x}_s\|_1 + \|\boldsymbol{e}\|_2 \right\}$$

with execution time $\mathrm{O}(\mathscr{L} \cdot \log(\|\boldsymbol{x}\|_2 / \eta))$.

- Need $m \ge \mathrm{C} s \log^{\alpha} N$ samples for restricted isometry hypothesis

Error Bound for Compressible Signals

Corollary 2. [Compressible signals] *Suppose*

- *the sampling matrix* $\mathbf{\Phi}$ *has* $\mathrm{RIP}(2s)$,
- *the signal* $\boldsymbol{x}$ *is* p*-compressible with magnitude* R,
- *the sample vector* $\boldsymbol{u} = \mathbf{\Phi}\boldsymbol{x} + \boldsymbol{e}$,
- $\mathscr{L}$ *bounds cost of a matrix–vector multiply with* $\mathbf{\Phi}$ *or* $\mathbf{\Phi}^*$.

Then CoSaMP produces a $2s$*-sparse approximation* $\boldsymbol{a}$ *such that*

$$\|\boldsymbol{x} - \boldsymbol{a}\|_2 \leq \mathrm{C}\left[Rp^{-1} \cdot s^{1/2-1/p} + \|\boldsymbol{e}\|_2\right]$$

with execution time $\mathrm{O}(\mathscr{L} \cdot p^{-1} \log s)$.

To learn more...

E-mail:

- `jtropp@acm.caltech.edu`
- `dneedell@math.ucdavis.edu`

Web: `http://www.acm.caltech.edu/~jtropp`

Relevant Papers:

- NTV, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," accepted to ACHA
- T and Rice DSP, "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals," in preparation
- N and Vershynin, "Stable signal recovery from incomplete and inaccurate samples," submitted
- T and Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit," *Trans. IT*, Dec. 2007.