Fast Reconstruction Algorithms for Deterministic Sensing Matrices and Applications

Robert Calderbank et al.

Program in Applied and Computational Mathematics
Princeton University
NJ 08544, USA.
Introduction
What is Compressive Sensing?

- When sample by sample measurement is expensive and redundant:
  - Compressive Sensing:
    - Transform to low dimensional measurement domain
  - Machine Learning:
    - Filtering in the measurement domain
Take-Home Message

Compressed Sensing is a **Credit Card**!

We want one with no hidden charges
Geometry of Sparse Reconstruction

- **Restricted Isometry Property (RIP):** An $N \times C$ matrix $A$ satisfies $(k, \epsilon)$-RIP if for any $k$-sparse signal $x$:

\[ (1 - \epsilon)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \epsilon)\|x\|_2. \]

- **Theorem [Candes, Tao 2006]:**
  If the entries of $\sqrt{N}A$ are iid sampled from
  - $\mathcal{N}(0, 1)$ Gaussian
  - $\mathcal{U}(-1, 1)$ Bernoulli
  distribution, and $N = \Omega \left(k \log \left(\frac{C}{k}\right)\right)$, then with probability $1 - e^{-cN}$, $A$ has $(k, \epsilon)$-RIP.

- **Reconstruction Algorithm [Candes, Tao 2006 and Donoho 2006]:**
  If $A$ satisfies $(3k, \epsilon)$-RIP for $\epsilon \leq 0.4$, then given any $k$-sparse solution $x$ to $Ax = b$, the linear program

\[
\text{minimize } \|z\|_1 \text{ such that } Az = b
\]

recovers $x$ successfully, and is robust to noise.
**Expander Based Random Sensing**

$A$: Adjacency matrix of a $(2k, \epsilon)$ expander graph

- No $2k$-sparse vector in the null space of $A$

**Theorem** [Jafarpour, Xu, Hassibi, Calderbank 2008]: If $\epsilon \leq 1/4$, then for any $k$-sparse solution $x$ to $Ax = b$, the solution can be recovered successfully in at most $2k$ rounds.

**Gap**: $g_t = b - Ax_t$

RHS proxy for difference between $x_t$ and $x$.

**ALGORITHM.** Greedy reduction of gap.
Two recent results
Performance Bounds for Expander Sensing with Poisson Noise

- Let $A$: adjacency matrix of an expander graph
- $x^*$: sparse
- Noisy compressed sensing measurements $y$ in Poisson model
- $\hat{x} = \arg \min \sum_{j=1}^{N} ((Ax)_j - y_j \log(Ax)_j) + \gamma \text{pen}(x)$
- Optimization over the simplex (positive values)
- $\text{pen}$: a well chosen penalty function.
- Then $\hat{x} \approx x^*$
<table>
<thead>
<tr>
<th><strong>Approach</strong></th>
<th><strong>Measurements</strong> ( N )</th>
<th><strong>Complexity</strong> ( C^3 )</th>
<th><strong>Noise Resilience</strong></th>
<th><strong>RIP</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis Pursuit (BP) [CRT]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C^3 )</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Orthogonal Matching Pursuit (OMP) [GSTV]</td>
<td>( k \log^\alpha(C') )</td>
<td>( k^2 \log^\alpha(C') )</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Group Testing [CM]</td>
<td>( k \log^\alpha(C') )</td>
<td>( k \log^\alpha(C') )</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Greedy Expander Recovery [JXHC]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C' \log \left( \frac{C}{k} \right) )</td>
<td>No</td>
<td>RIP-1</td>
</tr>
<tr>
<td>Expanders (BP) [BGIKS]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C^3 )</td>
<td>Yes</td>
<td>RIP-1</td>
</tr>
<tr>
<td>Expander Matching Pursuit (EMP) [IR]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C' \log \left( \frac{C}{k} \right) )</td>
<td>Yes</td>
<td>RIP-1</td>
</tr>
<tr>
<td>CoSaMP [NT]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C'k \log \left( \frac{C}{k} \right) )</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SSMP [DM]</td>
<td>( k \log \left( \frac{C}{k} \right) )</td>
<td>( C'k \log \left( \frac{C}{k} \right) )</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Random Signals or Random Filters?

Random Sensing

1. Outside the mainstream of signal processing: Worst Case Signal Processing
2. Less efficient recovery time
3. No explicit constructions
4. Larger storage
5. Looser recovery bounds

Deterministic Sensing

1. Aligned with the mainstream of signal processing: Average Case Signal Processing
2. More efficient recovery time
3. Explicit constructions
4. Efficient storage
5. Tighter recovery bounds
### $k$-Sparse Reconstruction with Deterministic Sensing Matrices

<table>
<thead>
<tr>
<th>Approach</th>
<th>Measurements</th>
<th>Complexity</th>
<th>Noise Resilience</th>
<th>RIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDPC Codes [BBS]</td>
<td>$k \log C$</td>
<td>$C \log C$</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Reed-Solomon codes [AT]</td>
<td>$k$</td>
<td>$k^2$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Embedding $\ell_2$ spaces into $\ell_1$ (BP) [GLR]</td>
<td>$k (\log C)^\alpha$</td>
<td>$C^3$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Extractors [Ind]</td>
<td>$kC^{o(1)}$</td>
<td>$kC^{o(1)} \log(C')$</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Discrete chirps [AHSC]</td>
<td>$\sqrt{C}$</td>
<td>$kN \log N$</td>
<td>Yes</td>
<td>StRIP</td>
</tr>
<tr>
<td>Delsarte-Goethals codes [CHS]</td>
<td>$2^{\sqrt{\log C}}$</td>
<td>$kN \log^2 N$</td>
<td>Yes</td>
<td>StRIP</td>
</tr>
</tbody>
</table>
**A:** $N \times C$ matrix satisfying
- columns form a group under pointwise multiplication
- rows are orthogonal and all row sums are zero

**α:** $k$-sparse signal where positions of the $k$ nonzero entries are equiprobable

**Theorem:** Given $\delta$ with $1 > \delta > \frac{k-1}{C-1}$, then with high probability

$$(1 - \delta)\|\alpha\|_2 \leq \|A\alpha\|_2 \leq (1 + \delta)\|\alpha\|_2$$

**Proof:** Linearity of expectation
- $\mathbb{E} [\|A\alpha\|^2] \approx \|\alpha\|^2$
- $\text{VAR} [\|A\alpha\|^2] \to 0$ as $N \to \infty$
Two recent results

Uniqueness of sparse representation and $\ell_1$ recovery

- **McDiarmid’s inequality:** Given a function $f$ for which
  $\forall \, x_1, \cdots, x_k, x_i'$:
  \[
  |f(x_i, \cdots, x_i, \cdots, x_k) - f(x_i, \cdots, x_i', \cdots, x_k)| \leq c_i,
  \]
  and given $X_1, \cdots, X_k$ independent random variables. Then
  \[
  \Pr[f(X_1, \cdots, X_k) \geq \mathbb{E}[f(X_1, \cdots, X_k)] + \eta ] \leq \exp \left( -\frac{2\eta^2}{\sum c_i^2} \right).
  \]

- **Relaxed assumption:**
  \[
  \forall \, i, j : \quad \left| \sum_x \varphi_i^i(x) \right|^2 - \left| \sum_x \varphi_j^j(x) \right|^2 \leq N^2 - \eta,
  \]
  Then:
  1. Uniqueness of sparse representation
  2. $\ell_1$ recovery of complex Steinhaus (random phase arbitrary magnitude) signals.
Kerdock set $K_m$: $2^m$ binary symmetric $m \times m$ matrices

Tensor $C^0(x, y, a): \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \to \mathbb{F}_2$ given by

$$\text{Tr}[xya] = (x_0, \ldots, x_{m-1})P^0(a)(y_0, \ldots, y_{m-1})^T$$

**Theorem:** The difference of any two matrices $P^0(a)$ in $K_m$ is nonsingular

**Proof:** Non-degeneracy of the trace

**Example:** $m = 3$, primitive irreducible polynomial $g(x) = x^3 + x + 1$

$$P^0(100) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \; P^0(010) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \; P^0(001) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$
Tensor $C^t(x, y, a): \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \times \mathbb{F}_{2^m} \to \mathbb{F}_2$ given by

$$C^t(x, y, a) = \text{Tr}[(xy^{2^t} + x^{2^t}y)a]$$

$$= (x_0, \ldots, x_{m-1})P^t(a)(y_0, \ldots, y_{m-1})^T$$

**Delsarte-Goethals Set** $DG(m, r)$: $2^{(r+1)m}$ binary symmetric $m \times m$ matrices

$$DG(m, r) = \left\{ \sum_{t=0}^{r} P^t(a_t)|a_0, \ldots, a_r \in \mathbb{F}_{2^m} \right\}$$

**Framework for exploiting prior information about the signal**

**Theorem:** The difference of any two matrices in $DG(m, r)$ has rank at least $m - 2r$

**Proof:** Non-degeneracy of the trace
The Delsarte-Goethals structure imparts an order of preference on the columns of a Reed-Muller sensing matrix

\[ K_m = DG(m, 0) \subset DG(m, 1) \subset \cdots \subset DG \left( m, \frac{m-1}{2} \right) \]

Better inner products $\leftrightarrow$ Worse inner products

If a prior distribution on the positions of the sparse components is known, the DG structure provides a means to assign the best columns to the components most likely present.
Reed-Muller Sensing Matrices

\[ A = [\phi^P,b(x)] : P \in DG(m, r), \ b \in \mathbb{Z}_2^m \]

\( A \) has \( N = 2^m \) rows and \( C = 2^{(r+2)m} \) columns

\[ \phi^P,b(x) = i^{\text{wt}(d_p)+2\text{wt}(b)} i^x P x^T + 2b x^T \]

- Union of \( 2^{(r+1)m} \) orthonormal basis \( \Gamma_P \)
- Coherence between bases \( \Gamma_P \) and \( \Gamma_Q \) determined by \( R = \text{rank}(P + Q) \)

**Theorem:** Any vector in \( \Gamma_P \) has inner product \( 2^{-R/2} \) with \( 2^R \) vectors in \( \Gamma_Q \) and is orthogonal to the remaining vectors

**Proof:** Exponential sums or properties of the symplectic group \( Sp(2m, 2) \)
Quadratic Reconstruction Algorithm

\[ f(x+a)f(x) = \frac{1}{N} \sum_{j=1}^{k} |\alpha_j|^2 (-1)^{aP_j}x^T + \frac{1}{N} \sum_{j \neq t} \alpha_j \overline{\alpha_t} \phi_{P_j,b_j}(x+a)\overline{\phi_{P_t,b_t}(x)} \]

\[ \frac{1}{N} \sum_{j=1}^{k} |\alpha_j|^2 (-1)^{aP_j}x^T : \text{Concentrates energy at } k \text{ Walsh-Hadamard tones.} \]

\[ \frac{1}{N} \sum_{j=1}^{k} |\alpha_j|^4 : \text{Signal energy in the Walsh-Hadamard tones} \]

The second term distributes energy uniformly across all \( N \) tones – the \( l^{th} \) Fourier coefficient is

\[ \Gamma^l_a = \frac{1}{N^{3/2}} \sum_{j \neq t} \alpha_j \overline{\alpha_t} \sum_x (-1)^{lx^T} \phi_{P_j,b_j}(x+a)\overline{\phi_{P_t,b_t}(x)} \]

**Theorem:** \( \lim_{N \to \infty} \mathbb{E}[N^2|\Gamma^l_a|^2] = \sum_{j \neq t} |\alpha_j|^2|\alpha_t|^2 \)

[Note: \( ||f||^4 = \left( \sum_{x,a} |f(x+a)f(x)|^2 \right)^2 \)]
Example: $N = 2^{10}$ and $C = 2^{55}$
**Information Theoretic Rule of Thumb:** Number of measurements $N$ required by Basis Pursuit satisfies

$$N > k \log_2 \left( 1 + \frac{C}{k} \right)$$

**RM(2, m):** $C = 2^{55}$, $k = 20$

$N = 1024$ versus $1014$

**Kerdock Sensing:** $C = 2^{20}$, $k = 70$

$N = 1024$ versus $971$
Still images with controlled sparsity are reconstructed with good fidelity using compressive sensing with chirp matrices:

- Original $128 \times 128$ image

- Sparsified image
  - Daubechies-4 wavelet expansion
  - 10% of coefficients (1636) retained

- Image reconstructed with deterministic algorithm from 4099 chirp measurements
  - About 4:1 compression
  - Essentially lossless reconstruction of sparsified image
**Deutsche Telekom:** Energy cost of operations greater than people cost

**Orthogonal CDMA:** RM(1,m)

1 Walsh function ↔ 1 bit to 1 user

**Compressive CDMA:** RM(2,m)

1 column ↔ 1 bit to \( \binom{m}{2} \) users

↔ \( \binom{m}{2} \) bits to 1 user

\[ \Phi = \begin{bmatrix} \phi_{P_i, b_i} \end{bmatrix}^{2m \times 2\left(\binom{m+1}{2}\right)} \]

\( m=10: \)

\[ X = \sum_{i=1}^{5} \sqrt{W_i} \phi_{P_i, b_i} \]

\( W_i = \text{power of signal intended for } i^{th} \text{ annulus} \)
Preliminary Assessment of Energy Savings

- **Orthogonal CDMA**: Many Sequences, No Interference, Low signal processing complexity
- **Compressive CDMA**: Few Sequences, Interference, More complex signal processing

![Graph showing Bit Error Rate for Worst User vs Normalized power](image)

- Orthogonal CDMA: $m=10, k=5$: MATLAB implementation of $k$-sparse signal reconstruction $\approx 1/20$ seconds
Using Chirps for A/D Conversion

- **Application:** Sparse signals of tones over large band

- **Idea:** Non-uniform sampling to convert pure tones to chirps

- **Leverage RIP results from compressed sensing measurements**
Motivation: A/D Metrics and Progress

- Standard Performance metric
  \[ P = 2^{SNR \text{ Bits}} f_{sampling} \]

- Captures bandwidth/resolution trade-off. Eg: \( \Delta\Sigma \) modulation

- [Walden (1999)] Slow rate of progress: 1.5 bit increase / 8 Years
Motivation: Nyquist Folding Analog-to-Information Receiver [G. Fudge et al.]

- Sampling at zero-crossing of a phase-modulated signal
- Undersampling aliases “Nyquist zones” together
- Stretching/reflection of phase-modulation resolves “Nyquist zone”
- Recovery visualized by spectrogram
Chirp Sampling

- Chirp Sample times: \( t_n = nT + n^2T\epsilon \)
  - Converts pure tones to linear chirps
    
    \[ e^{j\omega t} \rightarrow e^{j\omega Tn + j\omega T\epsilon n^2} \]

- Can pick \( T > \frac{1}{\text{bandwidth}} \) (under sampling)

Discretized Model:

- Pick \( P, Q \) with \( Q \) prime, \( P \not\approx Q \)
- Sample times \( t_n = AQn + BPn^2 \) for \( A, B \in \mathbb{Z} \)
  \( n = 0, 1, \ldots, P-1 \)

  \[ e^{2\pi j \frac{k}{PQ} t} \rightarrow e^{2\pi j \frac{Ak}{P} n} e^{2\pi j \frac{Bk}{Q} n^2} \]
Recovery Conditions via Compressed Sensing

- Properties of $\Theta$
  - Columns form a group
  - Rows are a tight frame

- Bound on column sum / inner-product

$$\left| \sum_{n=0}^{P-1} e^{2\pi j \frac{\alpha}{P} n} e^{2\pi j \frac{\beta}{Q} n^2} \right| \leq CP^{1-\frac{\delta}{2}} \quad \delta \approx 1$$

for constant $C$ independent of $P, Q$. 
For $P$ signal samples in noise given by

$$Z_n = be^{j\omega t_n} + W_n \quad W_n \sim \mathcal{N}_C(0, 2\sigma^2)$$

Cramér-Rao lower bound on $\hat{\omega}$ with unknown $b$

$$\text{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2(S_2 - S_1^2/P)}$$

where

$$S_1 = \sum_{n=0}^{P-1} t_n, \quad S_2 = \sum_{n=0}^{P-1} t_n^2$$

Uniform sampling: $t_n = nT$

Chirp sampling: $t_n = nT + n^2T\epsilon$

$$\text{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2}\frac{12}{T^2P(P^2 - 1)} \quad \text{CRLB is Achievable with good SNR}$$

$$\text{var}\{\hat{\omega}\} \geq \frac{\sigma^2}{|b|^2}\frac{1}{T^2O(\epsilon^2P^5)}$$
Recovery from Chirp Sampling

- Can leverage efficiency of FFT after simple transform, converting chirps to tones

\[ f[n] = \bar{y}[n]y[n + D] \]
\[ = |b_1|e^{j\omega_1 D^2 T + j\omega_1 DT}e^{j2\omega_1 DT n} + |b_2|e^{j\omega_2 D^2 T + j\omega_2 DT}e^{j2\omega_2 DT n} + \cdots + \text{cross terms} \]

- FFT upon \( f[n] \) gives initial estimates of \( \omega_i \) from which we can narrow the search

![Initial FFT on f[n]](image1)
![Rife Boorstyn refinement on \( \omega_i \) from f[n]](image2)
![Final refinement on original samples y[n]](image3)
Simulation Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P ) Samples</td>
<td>256</td>
</tr>
<tr>
<td>( T ) Sample Rate</td>
<td>1/200 s</td>
</tr>
<tr>
<td>( \epsilon ) Relative Chirp rate</td>
<td>1/10</td>
</tr>
<tr>
<td>( f_{\text{max}} ) Observed BW</td>
<td>1000 Hz</td>
</tr>
</tbody>
</table>

\( 5 \times \) undersampling (complex tones)
Passive Network Monitoring

Current methods
- fine grained analysis at a single node or flow
- collection of coarse statistics network wide

Limitations
- fail to leverage diverse detailed data from multiple vantage points
- too complex to extract knowledge from massive high-dimensional datasets
Recovery of low rank matrices: Keshavan, Montanari & Oh’09

$M - n \times m$ random matrix with rank $r$

$M = UV$ where $U$, $V$ are independent random matrices with i.i.d. entries

$M$ can be recovered up to precision $\delta$ from a random subset of $C(r, \delta)n$ observations. This can be accomplished efficiently via stochastic local search.

Verification of low rank: Rigidity Theory of matrices A. Singer

$M$ is rank $r$, for example $M_{ij} = \langle c_i, c_j \rangle$ $c_i \in \mathbb{R}^r$

Given $m$ entries, realization is rigid in $r$ dimensions (completable) if for all observed $(i, j)$

$\langle c_i, v_j \rangle + \langle c_j, v_i \rangle = 0$ $\equiv$ $\dim(\text{null}(C')) \leq r(r - 1)/2$
Content Distribution Network running on Planetlab

Monitors: Subset of Planetlab P nodes
End Hosts: Clients c, Sources S
Remaining Planetlab P nodes
Compressive Learning

Is it possible to find needles in compressively sampled haystack?

If features can be learned in data domain, can they also be learned in the measurement domain?

Curse of Dimensionality
The error of SVM in the measurement domain is with high probability close to the error of the best linear classifier in the data domain.