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
  - $N(0, 1)$ Gaussian
  - $U(-1, 1)$ 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.
<table>
<thead>
<tr>
<th>Approach</th>
<th>Measurements $N$</th>
<th>Complexity $C^3$</th>
<th>Noise Resilience</th>
<th>RIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basis Pursuit (BP) [CRT]</td>
<td>$k \log \left(\frac{C}{k}\right)$</td>
<td></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 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

**$\alpha$:** $k$-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
- $\mathbb{E} [\|A\alpha\|^2] \approx \|\alpha\|^2$
- $\text{VAR} [\|A\alpha\|^2] \rightarrow 0$ as $N \rightarrow \infty$
Kerdock set $K_m$: $2^m$ binary symmetric $m \times m$ matrices

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

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

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

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

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

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

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

[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
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
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.

\[ \text{[CJS2009]} \]