Sublinear Compressive Sensing (CS) and Support Weight Enumerators of Codes: A Matroid Theory Approach

Olgica Milenkovic
Joint work with Wei Dai and Hoa Vinh Pham

Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign

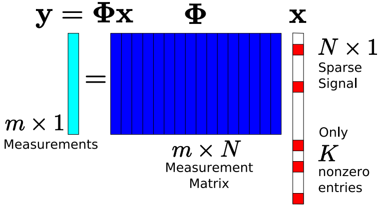
August 2009

Outline

- A brief introduction to CS;
- Why do support weight enumerators matter?
- Decoding of weighted superimposed codes: BP and OMP/SP - sublinear complexity reconstruction.
- Many open problems...

Compressive Sensing

CS: a technique that converts high dimensional signals into signals (measurements) with significantly smaller dimension ($m \ll N$).



Recovery problem: decode the signal $\mathbf{x}$ based on the measurement $\mathbf{y}$.

- Ill conditioned in general.
  - $\mathbf{\Phi}$ does not have full column rank. There are many $\mathbf{x}$ such that $\mathbf{y} = \mathbf{\Phi x}$.

When the signal is sparse, ...

When $\mathbf{x}$ is sufficiently sparse (K is small), exact reconstruction is possible.
(Kashin, 1977; Bresler et. al., 1999; Donoho et. al., 2004; Candés et. al., 2005)

Exact Reconstruction: iff $\mathbf{y}_1 - \mathbf{y}_2 = \mathbf{\Phi}(\mathbf{x}_1 - \mathbf{x}_2) \neq \mathbf{0}$,
$\forall K-$sparse $\mathbf{x}_1 \neq \mathbf{x}_2$.

$\Updownarrow$

Any $2K$-column submatrix of $\mathbf{\Phi}$ must have full rank.

Reconstruction algorithm (l_0-minimization):

$$\min \|\hat{\mathbf{x}}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}.$$

# of measurements: $m = 2K$.

Computational complexity: NP hard $\Rightarrow$ not practical for large N.

When the signal is sparse, ...

When $\mathbf{x}$ is sufficiently sparse (K is small), exact reconstruction is possible.
(Kashin, 1977; Bresler et. al., 1999; Donoho et. al., 2004; Candés et. al., 2005)

Exact Reconstruction: iff $\mathbf{y}_1 - \mathbf{y}_2 = \mathbf{\Phi}\left(\mathbf{x}_1 - \mathbf{x}_2\right) \neq \mathbf{0}$,
$\forall K-$sparse $\mathbf{x}_1 \neq \mathbf{x}_2$.

$$\Updownarrow$$

Any $2K$-column submatrix of $\mathbf{\Phi}$ must have full rank.

Reconstruction algorithm (l_0-minimization):

$$\min \|\hat{\mathbf{x}}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}.$$

# of measurements: $m = 2K$.

Computational complexity: NP hard $\Rightarrow$ not practical for large N.

When the signal is sparse, ...

When $\mathbf{x}$ is sufficiently sparse (K is small), exact reconstruction is possible.
(Kashin, 1977; Bresler et. al., 1999; Donoho et. al., 2004; Candés et. al., 2005)

Exact Reconstruction: iff $\mathbf{y}_1 - \mathbf{y}_2 = \mathbf{\Phi}\left(\mathbf{x}_1 - \mathbf{x}_2\right) \neq \mathbf{0}$,
$\forall K-$sparse $\mathbf{x}_1 \neq \mathbf{x}_2$.

$$\Updownarrow$$

Any $2K$-column submatrix of $\mathbf{\Phi}$ must have full rank.

Reconstruction algorithm (l_0-minimization):

$$\min \|\hat{\mathbf{x}}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}.$$

# of measurements: $m = 2K$.

Computational complexity: NP hard $\Rightarrow$ not practical for large N.

When the signal is sparse, ...

When $\mathbf{x}$ is sufficiently sparse (K is small), exact reconstruction is possible.
(Kashin, 1977; Bresler et. al., 1999; Donoho et. al., 2004; Candés et. al., 2005)

Exact Reconstruction: iff $\mathbf{y}_1 - \mathbf{y}_2 = \mathbf{\Phi}(\mathbf{x}_1 - \mathbf{x}_2) \neq \mathbf{0}$,
$\forall K-$sparse $\mathbf{x}_1 \neq \mathbf{x}_2$.

$\Updownarrow$

Any $2K$-column submatrix of $\mathbf{\Phi}$ must have full rank.

Reconstruction algorithm (l_0-minimization):

$$\min \|\hat{\mathbf{x}}\|_0 \text{ s.t. } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}.$$

# of measurements: $m = 2K$.

Computational complexity: NP hard $\Rightarrow$ not practical for large N.

l_1-minimization

l_1 minimization

$$\min \ \|\hat{\mathbf{x}}\|_1 \ \text{subject to } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}$$

- It is a convex optimization problem, solvable by linear programming.
- Complexity: $O\left(m^2 N^{3/2}\right)$ (Nesterov & Nemirovski, 1994)
- Performance guarantee?

Restricted Isometry Property: Φ satisfies the RIP with $\delta_K \in [0, 1]$ if for all K-sparse signals $\mathbf{x}$,

$$(1 - \delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1 + \delta_K)\|\mathbf{x}\|_2^2.$$

Sufficient condition: If Φ satisfies RIP with $\delta_{2K} < \sqrt{2} - 1$, then $\hat{\mathbf{x}} = \mathbf{x}$
(Candès & Tao, 2005 and Candès 2008)

l_1-minimization

l_1 minimization

$$\min \ \|\hat{\mathbf{x}}\|_1 \ \text{subject to } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}$$

- It is a convex optimization problem, solvable by linear programming.
- Complexity: $O\left(m^2 N^{3/2}\right)$ (Nesterov & Nemirovski, 1994)
- Performance guarantee?

Restricted Isometry Property: $\mathbf{\Phi}$ satisfies the RIP with $\delta_K \in [0,1]$ if for all K-sparse signals $\mathbf{x}$,

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \le \|\mathbf{\Phi}\mathbf{x}\|_2^2 \le (1+\delta_K)\|\mathbf{x}\|_2^2.$$

Sufficient condition: If Φ satisfies RIP with $\delta_{2K} < \sqrt{2}-1$, then $\hat{\mathbf{x}} = \mathbf{x}$
(Candès & Tao, 2005 and Candès 2008)

l_1-minimization

l_1 minimization

$$\min \ \|\hat{\mathbf{x}}\|_1 \ \text{subject to } \mathbf{y} = \mathbf{\Phi}\hat{\mathbf{x}}$$

- It is a convex optimization problem, solvable by linear programming.
- Complexity: $O\left(m^2 N^{3/2}\right)$ (Nesterov & Nemirovski, 1994)
- Performance guarantee?

Restricted Isometry Property: $\mathbf{\Phi}$ satisfies the RIP with $\delta_K \in [0,1]$ if for all K-sparse signals $\mathbf{x}$,

$$(1-\delta_K)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{x}\|_2^2 \leq (1+\delta_K)\|\mathbf{x}\|_2^2.$$

Sufficient condition: If $\mathbf{\Phi}$ satisfies RIP with $\delta_{2K} < \sqrt{2}-1$, then $\hat{\mathbf{x}} = \mathbf{x}$ (Candès & Tao, 2005 and Candès 2008)

Number of Measurements

Random matrices satisfying the RIP with constant parameters
(Candès et. al., 2005; Litvak et. al., 2005; Rudelson & Vershynin 2006)

1. Random matrices with i.i.d. entries.
   1. Gaussian distribution (subGaussian distribution).
   2. Bernoulli distribution.

$$m \geq O\left(K \log N\right)$$

2. Random matrices from the Fourier ensemble.
   1. choose m rows uniformly at random.

$$m \geq O\left(K \left(\log N\right)^{c}\right)$$

This Talk

The interface between coding theory and CS

- Sublinear complexity CS: Iterative decoding (belief propagation (BP)) meets greedy algorithms;
  - Constructive methods via low-density parity-check (LDPC) coding;
  - Reconstruction via greedy matching pursuit algorithms (OMP, SP, and CoSaMP) and BP decoding with a “twist”.

Low Complexity Decoding Algorithms from CS

Recent focus on greedy algorithms:

- Orthogonal Matching Pursuit (OMP) (Tropp, 2004)
- Regularized OMP (ROMP) (Needell & Vershynin, 2007)
- Stagewise OMP (StOMP) (Donoho et. al., 2007)
- Subspace Pursuit (SP) (Dai & Milenkovic, 2008)
- Compressive Sampling Matching Pursuit (CoSaMP) (Needell & Tropp, 2008)

	Complexity	Performance
l_0 minimization	$O\left(N^K\right)$	$\delta_{2K} < 1$
l_1 minimization	$O\left(m^2 N^{3/2}\right)$	$\delta_{2K} < \sqrt{2} - 1$
OMP	$O\left(KmN\right)$	$\delta_K < \frac{1}{2K}$
SP	$O\left(KmN\right)$ or less	$\delta_{3K} < 0.16$

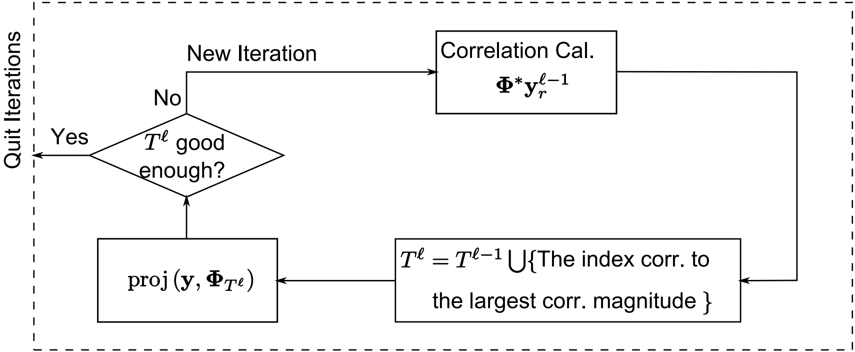
Orthogonal Matching Pursuit (OMP) Algorithm

Input: $\mathbf{\Phi}$, $\mathbf{y}$, K

Initialization:

$$T^0 = \phi, \ \mathbf{y}_r^0 = \mathbf{y}.$$

Iteration:



Output: solution obtained after K iterations

Subspace Pursuit (SP) algorithm

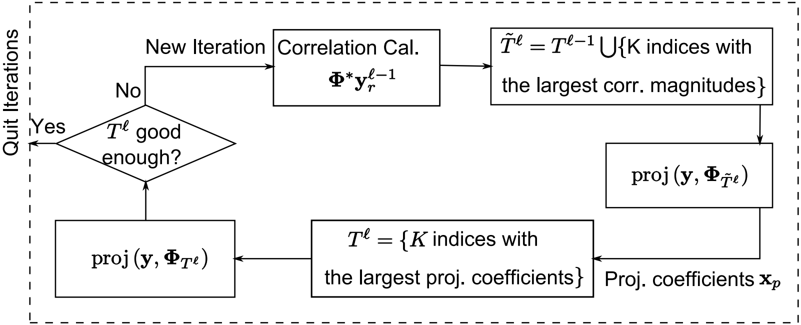
Input: $\mathbf{\Phi}$, $\mathbf{y}$, K

Initialization:

$T^0 = \{K$ indices corresponding to the largest magnitudes of $\mathbf{\Phi}^*\mathbf{y}\}$.

$\mathbf{y}_r^0 = \mathrm{resid}\left(\mathbf{y}, \mathbf{\Phi}_{T^0}\right)$.

Iteration:



Output:

$\hat{\mathbf{x}}$: $\hat{\mathbf{x}}_{T^\ell} = \mathbf{\Phi}_{T^\ell}^\dagger \mathbf{y}$ and $\hat{\mathbf{x}}_{(T^\ell)^c} = \mathbf{0}$.

LDPC Applications in CS

- Complexity of greedy strategies is dominated by correlation computation
  - Complexity is $O(mN)$.

- Use LDPC codebook for sensing matrix design
  - Mimics the Bernoulli matrix;
  - Introduce structure for storage saving.

- Correlation computation via BP
  - ML decoding = finding the largest correlation.
  - Decoding complexity: from $O(mN)$ to $O(m)$.

LDPC Applications in CS

- Complexity of greedy strategies is dominated by correlation computation
  - Complexity is $O(mN)$.

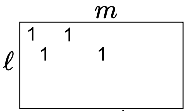
$$\underset{\mathbf{H} \in \mathbb{F}_2^{\ell \times m}}{\ell \overset{m}{\boxed{\begin{array}{cccc} 1 & 1 & & \\ & 1 & & 1 \\ & & & \end{array}}}} \Longrightarrow \mathbf{C} \in \mathbb{F}_2^{m \times 2^{m-\ell}} \underset{\text{BPSK Mapping}}{\Longrightarrow} \mathbf{\Phi} \in \{-1, +1\}^{m \times N}$$

- Use LDPC codebook for sensing matrix design
  - Mimics the Bernoulli matrix;
  - Introduce structure for storage saving.

- Correlation computation via BP
  - ML decoding = finding the largest correlation.
  - Decoding complexity: from $O(mN)$ to $O(m)$.

LDPC Applications in CS

- Complexity of greedy strategies is dominated by correlation computation
  - Complexity is $O(mN)$.



$\mathbf{H} \in \mathbb{F}_2^{\ell \times m}$

$$\Longrightarrow \mathbf{C} \in \mathbb{F}_2^{m \times 2^{m-\ell}} \underset{\text{BPSK Mapping}}{\Longrightarrow} \boldsymbol{\Phi} \in \{-1, +1\}^{m \times N}$$

- Use LDPC codebook for sensing matrix design
  - Mimics the Bernoulli matrix;
  - Introduce structure for storage saving.

- Correlation computation via BP
  - ML decoding = finding the largest correlation.
  - Decoding complexity: from $O\left(mN\right)$ to $O\left(m\right)$.

Incoherence parameter μ

$$\mu \triangleq \max_{i \neq j} |\langle \boldsymbol{\varphi}_i,\ \boldsymbol{\varphi}_j \rangle|,$$

- Sufficient condition of exact reconstruction for OMP (Tropp 2003):

$$\mu \leq \frac{1}{2K}$$

- Equivalent to Hamming distance requirement for LDPC codes

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j.$$

Proposition: A random LDPC code with row sums $d_c \geq 3$ and $m = O(K^2 \log N)$ satisfies

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j$$

with high probability.

Incoherence parameter μ

$$\mu \triangleq \max_{i \neq j} |\langle \boldsymbol{\varphi}_i,\ \boldsymbol{\varphi}_j \rangle|,$$

- Sufficient condition of exact reconstruction for OMP (Tropp 2003):

$$\mu \leq \frac{1}{2K}$$

- Equivalent to Hamming distance requirement for LDPC codes

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j.$$

Proposition: A random LDPC code with row sums $d_c \geq 3$ and $m = O(K^2 \log N)$ satisfies

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j$$

with high probability.

Incoherence parameter μ

$$\mu \triangleq \max_{i \neq j} |\langle \boldsymbol{\varphi}_i,\ \boldsymbol{\varphi}_j \rangle|,$$

- Sufficient condition of exact reconstruction for OMP (Tropp 2003):

$$\mu \leq \frac{1}{2K}$$

- Equivalent to Hamming distance requirement for LDPC codes

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j.$$

Proposition: A random LDPC code with row sums $d_c \geq 3$ and $m = O(K^2 \log N)$ satisfies

$$\frac{1}{2} - \frac{1}{4K} < \frac{d_H(\mathbf{c}_i, \mathbf{c}_j)}{m} < \frac{1}{2} + \frac{1}{4K}, \quad \forall i \neq j$$

with high probability.

RIP property

- Gershgorin Circle Theorem: For all $\mathbf{A} \in \mathbb{C}^{n \times n}$,

$$\{\lambda_i\} \subset \bigcup_{i=1}^{n} D\left(a_{i,i}, \sum_{j \neq i} |a_{i,j}|\right).$$

- RIP holds!
  For all eigenvalues of $\mathbf{\Phi}_T^* \mathbf{\Phi}_T$,

$$\begin{aligned} \left|\lambda\left(\mathbf{\Phi}_T^* \mathbf{\Phi}_T\right) - 1\right| &\leq \max_j \sum_{l \neq j} \left|\langle \boldsymbol{\varphi}_j, \boldsymbol{\varphi}_l \rangle\right| \\ &\leq K\mu \leq \frac{1}{2}, \end{aligned}$$

  which implies

$$\delta_K \leq 1/2.$$

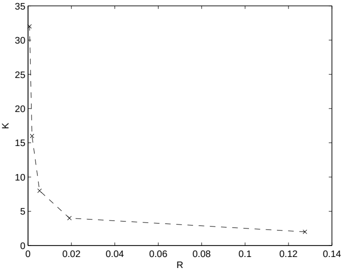
LDPC Code Rate for CS

A necessary condition: Unless the LDPC code family satisfies

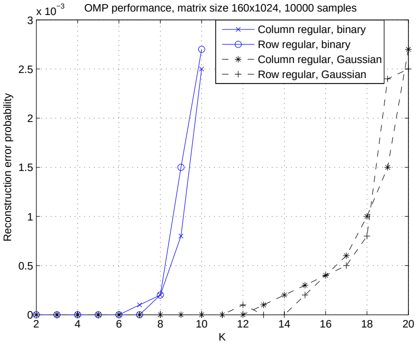
$$R < 1 - (1 - \frac{\sqrt{2}}{K}) \frac{\log_2(K-1)}{\log_2(K)} - \frac{H(\sqrt{2}/K)}{K},$$

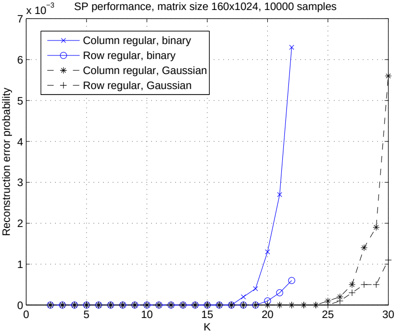
the RIP constant cannot satisfy $\delta_K < \sqrt{2} - 1$.

Proof is based on connection between the RIP and generalized Hamming weights of a code.



Performance of standard OMP and SP algorithms





Extensions

- List-based BP decoding algorithm.
  - Motivated by the significant performance improvement of SP compared with OMP.
  - Instead of outputing the ML codeword, we output a list of K codewords that have large likelihood.
- Multiple basis belief propagation (MBBP) Algorithm
  - An LDPC code can have different parity check matrices (bases).
  - The performance of BP algorithm highly depends on the chosen basis.
  - We propose to run BP algorithm on multiple bases and choose the best output codeword.

Thank you!