Compressed Sensing and Linear Codes over Real Numbers

Henry D. Pfister
(joint with Fan Zhang)

Texas A&M University
College Station

Information Theory and Applications Workshop
UC San Diego
January 31st, 2008

Outline

1. Introduction
   - Compressed Sensing
   - Connections with Coding
2. Benefits of the Coding Perspective
   - Large Body of Previous Work
   - Low-Density Parity-Check Codes for Compressed Sensing
   - Confessions
3. Summary and Open Problems

Outline

What is Compressed Sensing? (1)

- Compressed sensing (CS) is a relatively new area of signal processing and statistics that focuses on signal reconstruction from a small number of linear (i.e. dot product) measurements

- CS originated with the observation that many systems:
  - (1) Sample a large amount of data
  - (2) Perform a linear transform (e.g., DCT, wavelet, etc...)
  - (3) Throw away all the small coefficients

- If the locations of the important transform coefficients were known in advance, then they could be sampled directly with reduced complexity

What is Compressed Sensing? (2)

- CS is a research area that emerged when it was realized that random sampling could be used (with a small penalty) to achieve the same result without prior knowledge of the locations of important coefficients

- CS has two stages: sampling and reconstruction
  - During sampling, the signal-of-interest (SOI) is sampled by computing its dot product with a set of sampling kernels
  - During reconstruction, the SOI is estimated from the samples
  - In many cases, the number of samples required for a good estimate is much smaller than other methods (e.g., Nyquist sampling at twice the maximum frequency)

Basic Problem Statement

- Use m dot-product samples to reconstruct a signal
  - The signal vector is $x \in \mathbb{R}^n$
  - The $m \times n$ measurement matrix is $\Phi \in \mathbb{R}^{m \times n}$
  - The length-m sample vector is $y = \Phi x$

Basic Problem Statement

- Use m dot-product samples to reconstruct a signal
  - The signal vector is $x \in \mathbb{R}^n$
  - The $m \times n$ measurement matrix is $\Phi \in \mathbb{R}^{m \times n}$
  - The length-m sample vector is $y = \Phi x$
- Given y, the valid signal set is $V(y) = \{x' \in \mathbb{R}^n | \Phi x' = y\}$
  - If $m < n$, then a unique solution is not possible
  - With prior knowledge, we try to choose a "good" solutions
  - If x is i.i.d. zero-mean Gaussian, then the ML solution is

$$\hat{x} = \arg \min_{x' \in V(y)} \|x'\|_2 = \Phi^T \left(\Phi \Phi^T \right)^{-1} y$$

  - If x is sparse (w.r.t. Hamming weight $\|\cdot\|_H$), then

$$\hat{x} = \arg \min_{x' \in V(y)} \|x'\|_H$$

Reconstruction

- But constrained $\|\cdot\|_H$ minimization is NP-Hard
  - Instead use linear programming (LP) to solve $\min_{x' \in V(y)} \|x'\|_1$
  - "Basis Pursuit" by Chen, Donoho, and Saunders (1998)
- The minimizing vector is the same (w.h.p.) for both problems!
  - If Φ is chosen randomly (e.g., such that $\Phi_{ij} \sim N(0,1)$)
  - And m is large enough: $m \geq C \|x\|_H \log n$
  - See work by Donoho, Candès, Romberg, Tao, and others
- $\|\cdot\|_2$-error small even for approximately sparse signals
  - Donoho showed this for $m \geq C_p \|x\|_p \log n$ and $0 < p < 1$
- Definition: $\|x\|_p \triangleq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$ for $0 < p < \infty$

Connections to Other Research Areas

- Classical Sampling Theory
  - How many samples for reconstruction without aliasing?
  - Based on support set in the frequency domain
    - Single interval $\rightarrow$ Nyquist rate is twice the bandwidth
- Approximation Theory
  - Signal is a function drawn from a particular class of functions
  - Best k-term approximation w.r.t. to some overcomplete basis
  - Linear approximation $\rightarrow$ Kolmogorov n-width of function class
- Information Theory: Lossless and Lossy Compression
  - Signal model becomes stochastic rather than deterministic
  - Using this model, CS becomes syndrome source coding

Outline

Syndrome Source Coding

- Compression via random linear projections dates to the 1970's
  - Mostly for distributed source coding (e.g., Slepian-Wolf)
  - Early history of "syndrome source coding" by Ancheta [IT-1976]
- Basic Idea → Treat random signal as an error pattern
  - Parity-check matrix computes the syndrome of the "error"
  - This syndrome is the "compressed" version of the signal
  - A syndrome decoder is used to reconstruct the "error"
  - Problem is identical to error correction coding
    - ★ For exactly sparse signals

Compressed Sensing and Coding

Compressed Sensing	Coding
• Signal: $x \in \mathbb{R}^n$	• Error pattern: $e \in F^n$
▸ sparse: $\|x\|_H \leq \delta n$	▸ sparse: $\Pr(e_i \neq 0) = \delta$
• Measurement matrix: $\Phi \in \mathbb{R}^{m \times n}$	• Parity-check matrix: $H \in F^{m \times n}$
▸ Blind to nullspace of Φ	▸ Code is nullspace of H
• Sample vector: $y = \Phi x$	• Syndrome: $s = He$
• Dec: $\hat{x} = \arg \min_{x' : y = \Phi x'} \|x'\|_H$	• Dec: $\hat{e} = \arg \min_{e' : s = He'} \|e'\|_H$

- In coding, F is usually a finite field not the real numbers[1]

[1] Except RS/BCH codes over $\mathbb{C}$ in [Wolf-TCOM83]

Minimum Distance Decoding

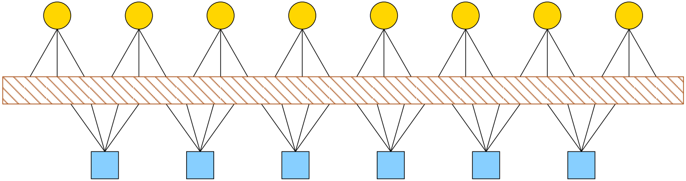
- For $H \in \mathbb{R}^{m \times n}$, we define the code $\mathcal{C} = \{x \in \mathbb{R}^n | Hx = 0\}$
  - Let the minimum distance be $d \triangleq \min_{x \in \mathcal{C} \setminus 0} \|x\|_H$
- Decode: $\hat{x} = \min_{x' : y = \Phi x'} \|x'\|_H$
  - Can fail only if $\exists\ u \in \mathcal{C} \setminus 0$ such that $\|u + x\|_H \leq \|x\|_H$
  - Failure only possible if $\|x\|_H \geq \frac{d-1}{2}$
  - This is basic coding theory!
- Easy to construct MDS codes where $d = m + 1$
  - "RS code" given by a Fourier measurement matrix
  - Random matrices give MDS codes with prob. 1
    - ★ All entries chosen i.i.d. from a continuous distribution
  - Very similar to coding over a large finite alphabet

Outline

Tools from Coding Theory

- The design and analysis of error-correcting codes has been significant research topic for roughly 60 years
  - Long random codes used to achieve capacity
  - Individual codes analyzed in terms of weight spectrum
  - Alternative error models (e.g., Lee metric decoding)
  - Capacity-achieving codes with low-complexity iterative decoding
- Challenges in applying coding to CS
  - Unrealistic error model implied by Hamming distance
  - Coding uses stochastic (rather than deterministic) signal models

Low-Density Parity-Check (LDPC) Codes



- Linear codes with sparse parity-check matrix $H \in F^{m \times n}$
  - Codes defined to be $x \in F^n$ such that $\sum_j H_{ij} x_j = 0$
- Bipartite graph representation
  - An edge connects equation node i to symbol node j if $H_{ij} \neq 0$
  - Iterative decoding passes messages along edges of graph
  - Graphs with irregular degree profiles can approach capacity

Coding Over Large Finite Alphabets

- Verification-based iterative decoding of LDPC codes
  - Introduced by Luby and Mitzenmacher (Allerton02)
  - For the q-ary symmetric channel with large q: $(C \approx 1 - \delta)$

$$\Pr(y|x) = \begin{cases} 1 - \delta & \text{if } x = y \\ \delta/(q-1) & \text{if } x \neq y \end{cases} \qquad x, y \in \mathrm{GF}(q)$$

- Message $\{Z\}_S$ has value $Z \in \mathrm{GF}(q)$ and status $S \in \{U, V\}$
- A message is "verified" if correct with high probability
  - Verify msg if two independent observations match
  - Pr(two ind. observations give same incorrect value) = $\frac{1}{q-1}\delta^2$
  - False Verification (FV): message is *verified* but not correct

Previous Work on Verification-Based Decoding

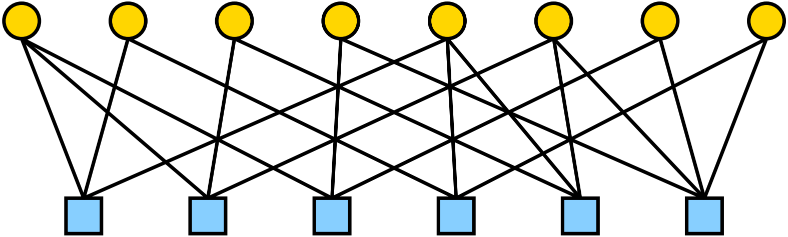
- Algorithms by Luby and Mitzenmacher (Allerton02, IT05)
  - 1st Alg. (LM1): Verify all symbols if check sums to zero
  - 2nd Alg. (LM2): LM1 + Verify if two msgs match at symbol
- Algorithms by Shokrollahi and Wang (ISIT04)
  - 1st Alg. (SW1): Same as LM2
  - 2nd Alg. (SW2): Pass list msgs with all probable values
    - Capacity-achieving but extremely high complexity
- Algorithm by Zhang and Pfister (Globecom07)
  - List-Message Passing (LMP) with Verification
  - Density evolution and optimization for list-size S
- DE gives message-passing threshold δ^* for (3,6) LDPC codes
  - $\delta^*_{LM1} = 0.169$, $\delta^*_{LM2} = 0.21$, and $\delta^*_{LMP} = 0.23$ for $S = 32$

Outline

Sudocodes and BP Approaches

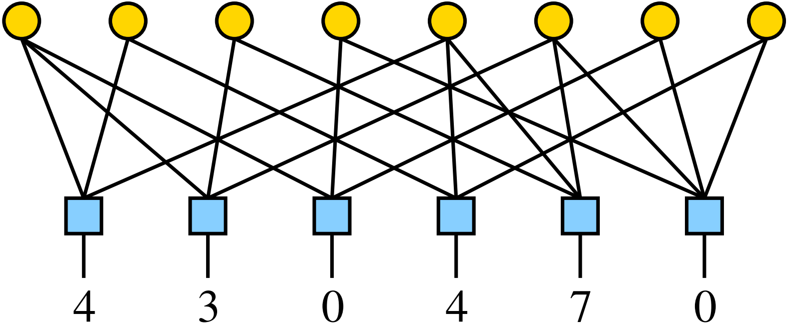
- "Sudocodes" of Sarvotham, Baron, and Baraniuk (ISIT06)
  - Measurement matrix Φ is a sparse 0/1 matrix (L ones per row)
  - Iterative decoding method based on matching coefficients
    - ⋆ Equivalent to LM2 for syndrome source coding
  - Requires $m = O(n \log n)$ for coverage (e.g., balls/bins)
  - Second phase used to handle uncovered elements
- LDPC/Belief propagation for CS by Sarvotham et al. (2006)
  - Sparse measurement matrix has -1/0/+1 entries
  - Approx. sparse signal modeled by mixture of two Gaussians
  - Belief propagation decoding used to find important coefficients
  - Algorithm evaluated primarily by simulation
- Expander LDPC codes for CS by Xu and Hassibi (ITW07)

CS Reconstruction via LM1



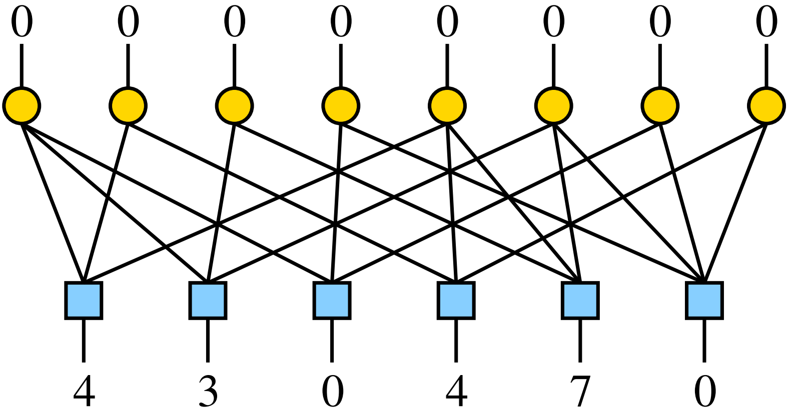
Signal (circles) measurement (squares) model

CS Reconstruction via LM1



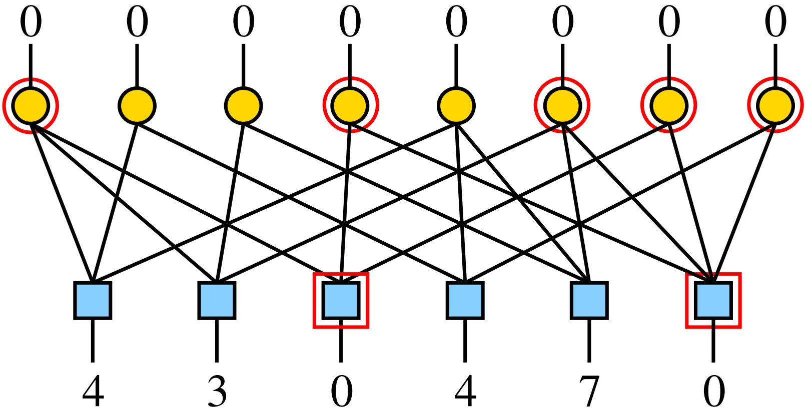
With measurements exposed

CS Reconstruction via LM1



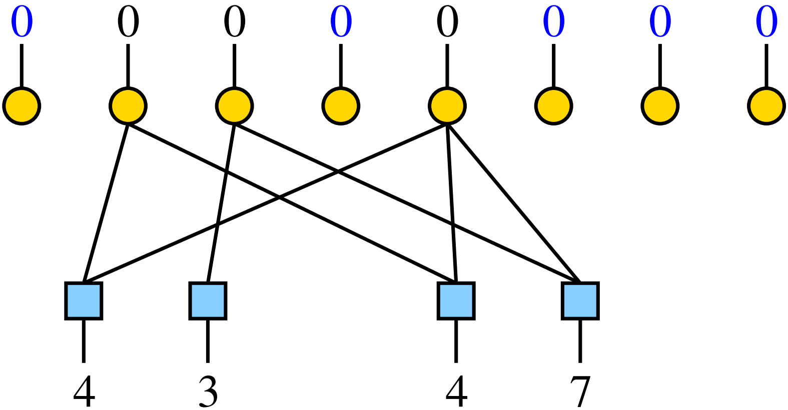
All-zero signal assumed for decoding

CS Reconstruction via LM1



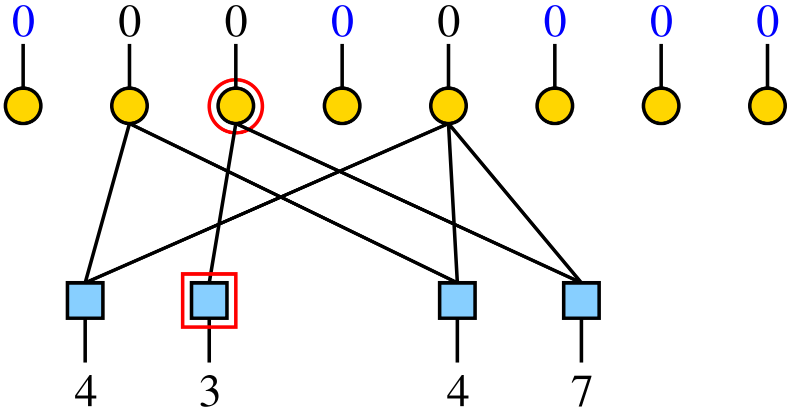
Assume all satisfied checks are correct

CS Reconstruction via LM1



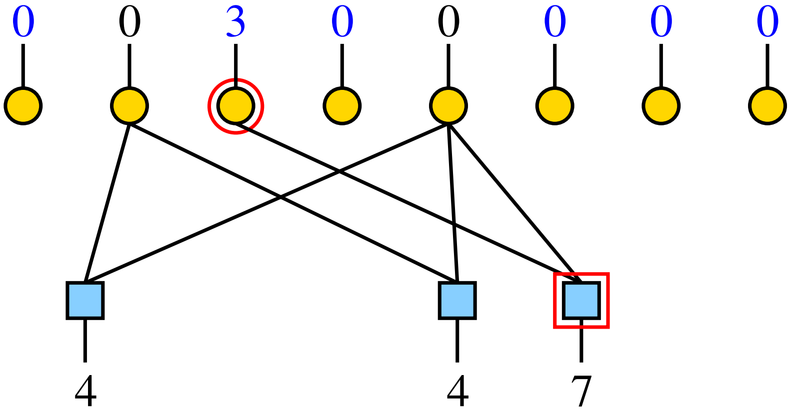
Remove edges and fix values

CS Reconstruction via LM1



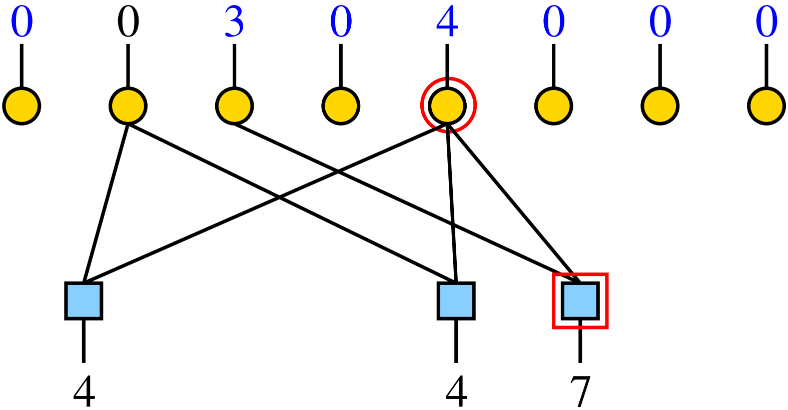
Use degree-1 check to determine a variable

CS Reconstruction via LM1



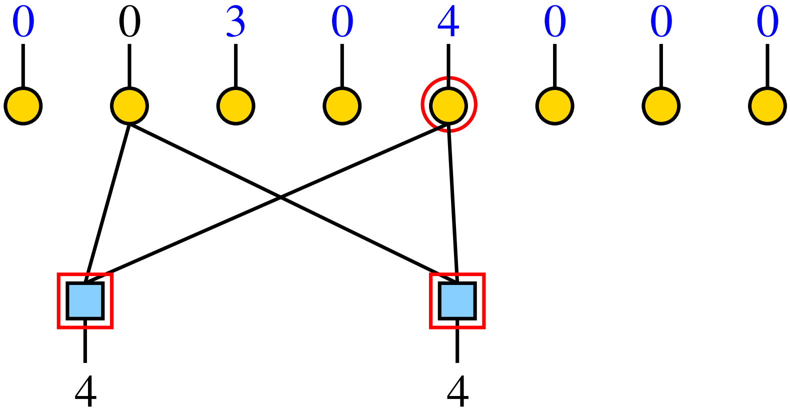
This value can be removed from all equations

CS Reconstruction via LM1



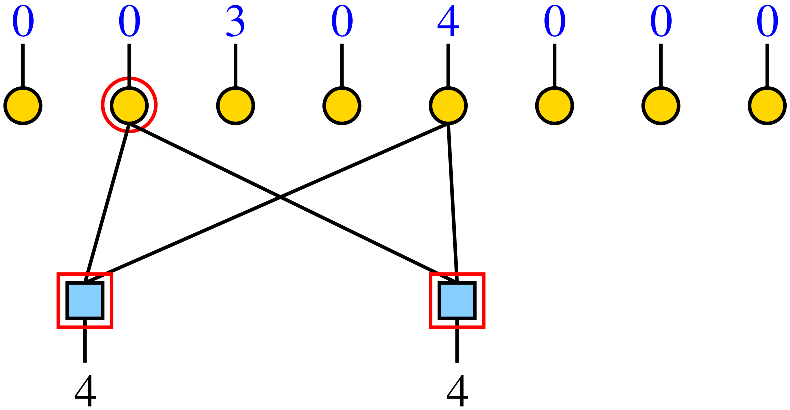
Another variable is determined

CS Reconstruction via LM1



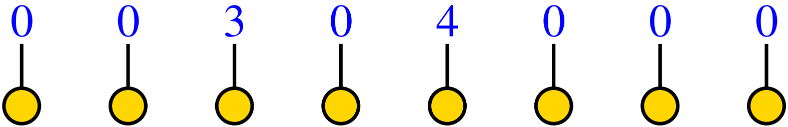
This value can be removed from all equations

CS Reconstruction via LM1



Final value is determined in two ways

CS Reconstruction via LM1



Reconstruction is successful

New Result: Analysis of LM1 for CS

- Probabilistic signal model
  - Length-n vector x with i.i.d. elements: $\Pr(x_i \neq 0) = \delta$
  - $\|x\|_H \leq (1+\epsilon)\delta n$ with high probability as $n \to \infty$
- Sparse measurement matrix with m rows
  - Regular degree profile with j non-zero entries per column
  - Non-zero entries drawn from a continuous distribution

Theorem: LM1 Recovery for CS

LM1 recovery for CS with (j,k)-regular LDPC codes reconstructs x (with high probability as $n \to \infty$) if $\delta < (k-1)^{-j/(j-1)}$.

- Recovery time is linear in n with $m = j\delta^{-1/j}\, n$ measurements
  - Choosing $j = \left\lceil \ln \frac{1}{\delta} \right\rceil$ reduces this to $m \approx e \left\lceil \ln \frac{1}{\delta} \right\rceil n$ measurements

New Result: Near Optimal Recovery

- Probabilistic signal model
  - Length-n vector x with i.i.d. elements: $\Pr(x_i \neq 0) = \delta$
  - $\|x\|_H \leq (1+\epsilon)\delta n$ with high probability as $n \to \infty$
- Sparse measurement matrix with m rows
  - Irregular degree profile for capacity-achieving erasure codes
  - Non-zero entries drawn from a continuous distribution

Theorem: Near Optimum Recovery with LMP

List-message passing decoding with verification reconstructs x (with high probability as $n \to \infty$), for any $\epsilon > 0$, if $m \geq (1+\epsilon)\delta n$.

- Complexity linear in n (but constant grows as $\delta \to 0$ or $\epsilon \to 0$)

Conjecture: Uniform Recovery

- Reconstruct all x with some maximum $\|x\|_H$
  - Assume length-n vector x such that $\|x\|_H \leq \delta n$
- Sparse measurement matrix with m rows
  - (j,k)-Regular degree profile used for row and column weights
  - Non-zero entries drawn from a continuous distribution

Conjecture: Uniform Recovery with LM1

LM1 recovery with randomly chosen (j,k)-regular LDPC codes (w.h.p. as $n \to \infty$) recovers all x if $\delta < \frac{1}{2}(k-1)^{-2/j}$.

- Based on a preliminary stopping set analysis of LM1

Outline

Exact Versus Approximate Sparsity

- Connection to coding theory is strongest for exact sparsity
  - Current analysis works only in that case (unless we cheat)
    - Use ϵ-neighborhood for verification, assume no FV
  - DE analysis of [SBB06] decoder in progress
    - This should extend our results to approximate sparsity

Transform Domain Issues

- Encoding and decoding interacts with the domain of sparsity
  - Sparsifying transform can be built into encoding in two ways
    - Transform then encode
    - Add transform to decoding graph (think ISI channels)
- In contrast
  - Basis pursuit allows sampling before transform is known

Summary and Open Problems

- Discussed connections between ECC and CS
  - Connection is not new, but coding brings many new tools
- Linear-time recovery algorithms for CS based on coding
  - Capacity achieving codes can provide near optimal recovery
  - Uniform recovery conjecture based on stopping set analysis
- Open Questions
  - Extension of analysis to approximately sparse signals
    - Must extend both signal model and DE analysis
  - Overcoming transform domain issues with linear complexity