Information Theory and Coding for Compressed Sensing

Henry D. Pfister
(joint work with Fan Zhang)

Texas A&M University
College Station

Center for Wireless Communications Ericsson Seminar
UC San Diego
March 2nd, 2009

Outline

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 projection 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)

- Compressed sensing (CS) emerged when it was observed that random sampling could be used (with a small penalty) to achieve the sample rate reduction result without knowing the locations of important coefficients
- CS has two stages: sampling and reconstruction
  - During sampling, the signal-of-interest (SOI) is sampled by computing its projection onto a set of sampling vectors
  - 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

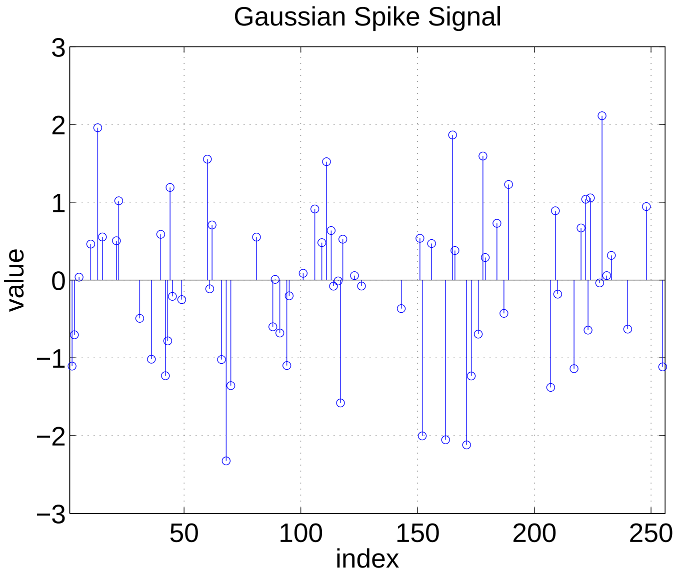
- Use m dot-product samples to reconstruct a signal
  - The signal vector is $x \in \mathbb{R}^n$
  - The $m \times n$ measurement matrix is $\Phi \in \mathbb{R}^{m \times n}$
  - The length-m sample vector is $y = \Phi x$

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

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

Exactly Sparse Signal



Reconstruction

- "Basis Pursuit" by Chen, Donoho, and Saunders (1998)
  - Constrained $\|\cdot\|_H$ minimization is NP-Hard
  - Instead, use LP to solve the relaxation: $\min_{x' \in V(y)} \|x'\|_1$

- 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: $m \geq C \|x\|_H \log n$
    - ★ Or equivalently, x is very sparse: $\|x\|_H \leq \frac{m}{C \log n}$
  - See work by Donoho, Candès, Romberg, Tao, and others

Drawbacks of Basis Pursuit

- Reconstruction Performance
  - The set $\{x \in \mathbb{R}^n | \, \|x\|_1 \leq A\}$ is bigger than $\left\{x \in \mathbb{R}^n | \, \|x\|_{0^+} \leq A\right\}$
  - Many more measurements required to reconstruct the signal
  - Solution: Instead minimize $\|x\|_p$ for $0 < p < 1$
    - Reweighted ℓ_p algorithms and coding-based techniques

- Reconstruction Complexity
  - LP (even with fast transforms) takes $\Omega\left(n^2\right)$ operations
  - Solution: Fast approximate ℓ_1 algorithms
    - Bregman iteration and fixed point continuation
    - Variations of matching pursuit

Outline

Compressed Sensing and Coding

Compressed Sensing

- Signal: $x \in \mathbb{R}^n$
  - sparse: $\|x\|_H \leq \delta n$
- Measurement matrix:
  $$\Phi \in \mathbb{R}^{m \times n}$$
  - Blind to nullspace of Φ
- Sample vector: $y = \Phi x$
- Dec: $\hat{x} = \arg \min_{x' : y = \Phi x'} \|x'\|_H$

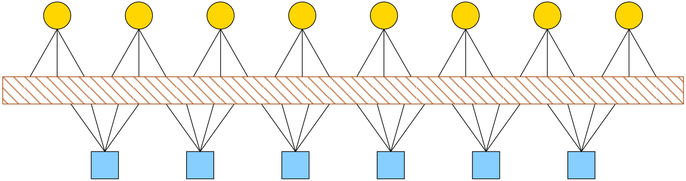
Linear Codes

- Error pattern: $e \in F^n$
  - sparse: $\Pr(e_i \neq 0) = \delta$
- Parity-check matrix:
  $$H \in F^{m \times n}$$
  - Code is nullspace of H
- Syndrome: $s = He$
- Dec: $\hat{e} = \arg \min_{e' : s = He'} \|e'\|_H$

Related Problems

- Sparse Approximation
  - Given a matrix Φ, find a sparse x such that $\|y - \Phi x\|$ is small
  - Problem studied in signal processing for quite some time
  - It is the problem that motivated basis pursuit (BP)
- Compressed Sensing
  - Choose a matrix Φ and decoder $D(y)$ such that:
    - ⋆ $D(\Phi x) \approx x$ when x is sufficently sparse
- From a coding perspective, the difference is who picks the code
  - Spase approximation: code is given
  - Compressed sensing: code can be optimized
  - Recent advances in coding favor the CS problem

Low-Density Parity-Check (LDPC) Codes



- Linear codes with sparse parity-check matrix $H \in F^{m \times n}$
  - Codewords 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)
  - A message is "verified" if correct with high probability
    - Message has value $Z \in \mathrm{GF}(q)$ and status V or U

- For the q-ary symmetric channel with large q: $(C \approx 1 - \delta)$

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

  - Verify msg if two independent observations match
  - Pr(two ind. observations give same incorrect value) = $\frac{1}{q-1}\delta^2$
  - Verify if all symbols nodes in a parity check sum to zero
  - False Verification (FV): message is *verified* but not correct

Outline

Verification-Based Decoding Algorithms

- Algorithms by Luby and Mitzenmacher (Allerton02,IT05)
  - 1st Algorithm (denoted LM1)
    - ★ Verify all symbols whose parity check sums to zero
    - ★ Use degree-1 equations to determine symbol values
  - 2nd Algorithm (denoted LM2)
    - ★ Verify all symbols whose parity check sums to zero
    - ★ Use degree-1 equations to determine symbol values
    - ★ Verify messages if any two match at symbol node
- Connection between CS and Codes over Real Numbers
  - Basic assumption: "verified symbols are correct w.h.p."
  - Holds both for large finite-fields and the real numbers

CS Reconstruction via LM1

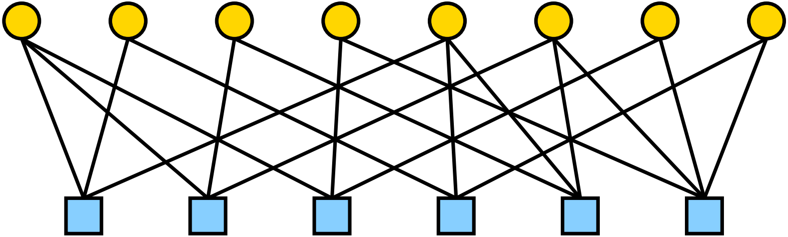
Example

Let Φ be the following parity-check matrix H

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

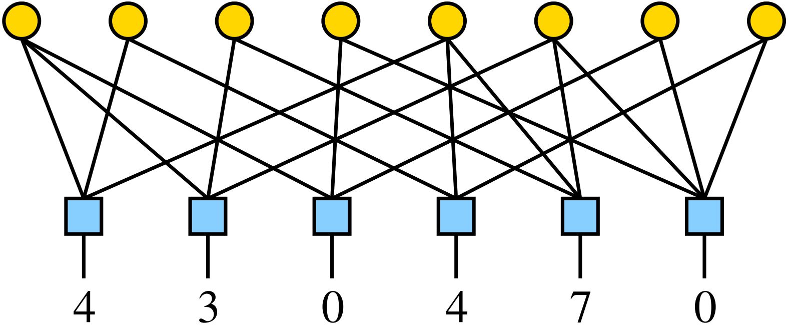
Note: We choose edge weights of one for simplicity

CS Reconstruction via LM1



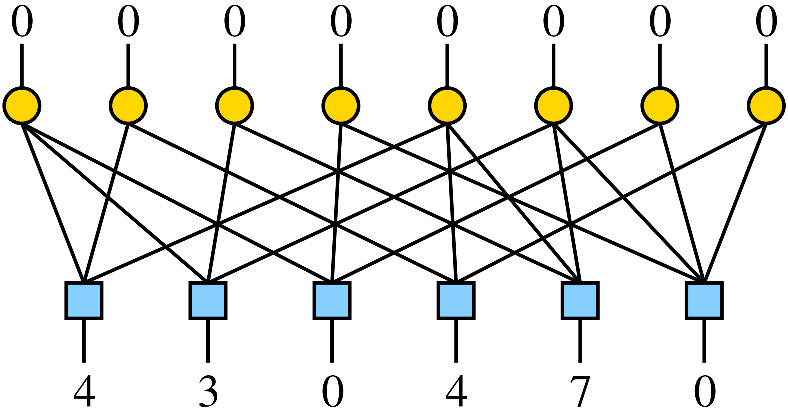
Signal (circles) measurement (squares) model

CS Reconstruction via LM1



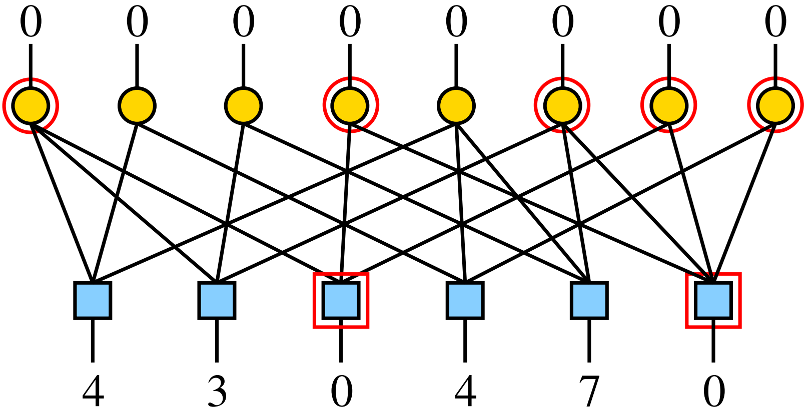
With measurements exposed

CS Reconstruction via LM1



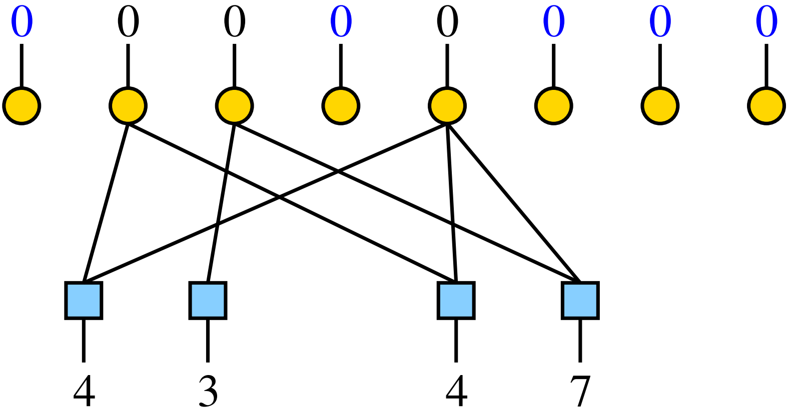
All-zero signal assumed for decoding

CS Reconstruction via LM1



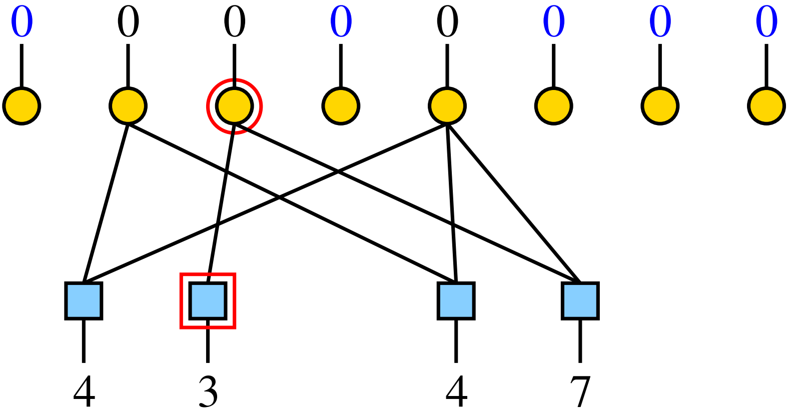
Assume all satisfied checks are correct

CS Reconstruction via LM1



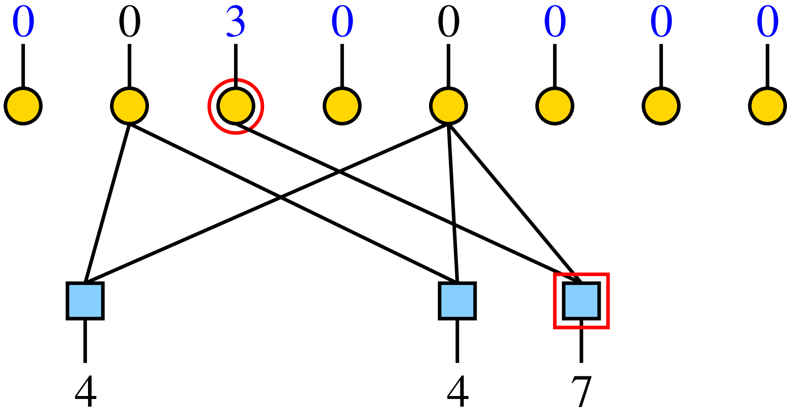
Remove edges and fix values

CS Reconstruction via LM1



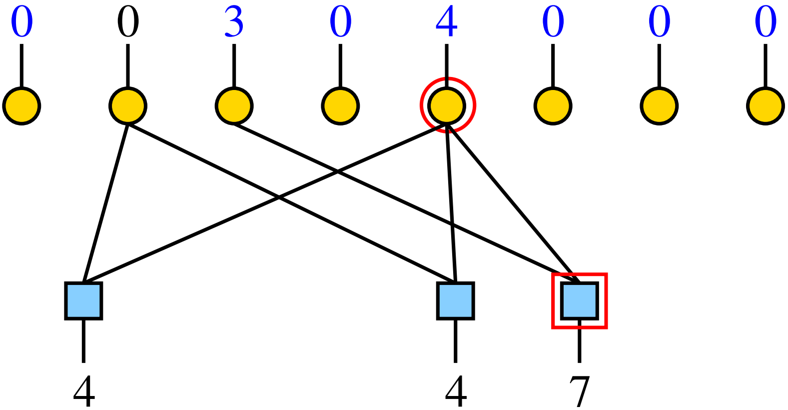
Use degree-1 check to determine a variable

CS Reconstruction via LM1



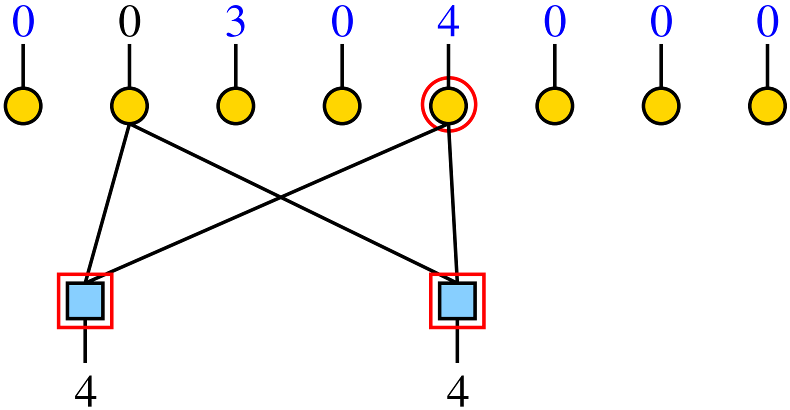
This value can be removed from all equations

CS Reconstruction via LM1



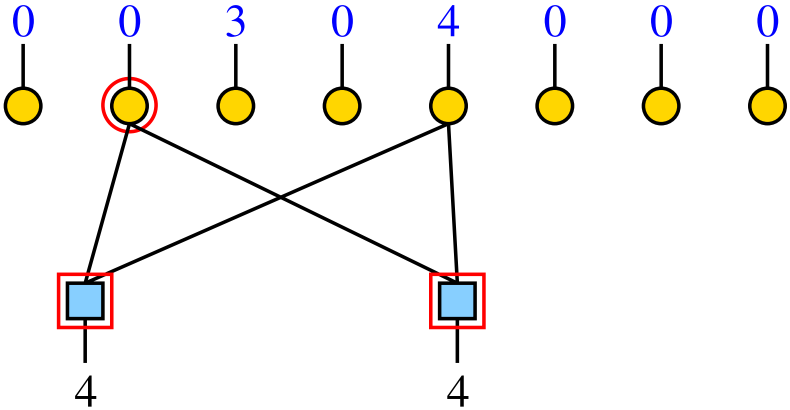
Another variable is determined

CS Reconstruction via LM1



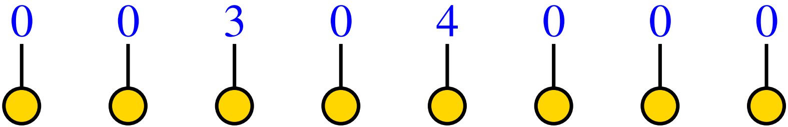
This value can be removed from all equations

CS Reconstruction via LM1



Final value is determined in two ways

CS Reconstruction via LM1



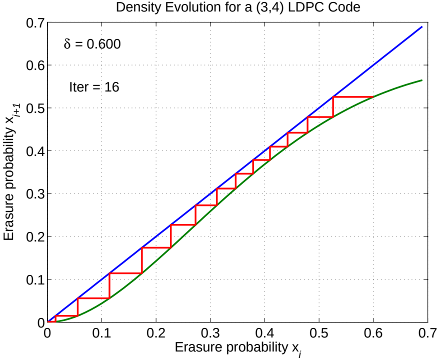
Reconstruction is successful

Sudocodes and Belief Propagation 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 CS problems
  - 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)

Outline

Warmup Example: The Binary Erasure Channel



Density evolution for a (j,k)-regular LDPC code

$$x_{i+1} = \delta \left(1 - (1 - x_i)^{k-1}\right)^{j-1}$$

Code	Rate	δ^*
(3,6)	0.5	0.4294
(3,12)	0.75	0.2105
(3,24)	0.875	0.1038

- Binary erasure channel with erasure prob. δ
- DE tracks the message erasure rate x_i after i iterations
- Gives noise threshold δ^* below which decoding succeeds w.h.p.
- Requires numerical calculations for each code ensemble

Very Sparse <=> Very High Rate

- CS often considers the sublinear sparsity $\delta n = o(n)$ regime
  - How does the DE threshold scale in this regime?
  - Shannon limit says that $\delta^* \leq j/k$
  - To understand this, numerical DE is not enough

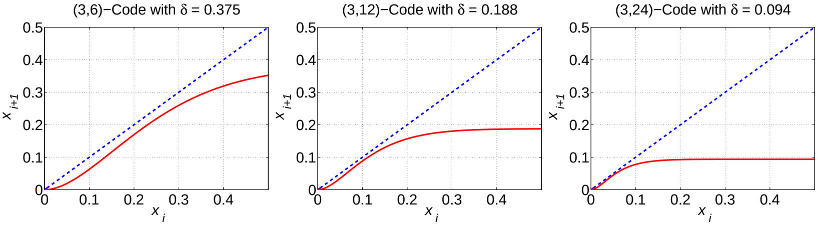
Very Sparse <=> Very High Rate

- CS often considers the sublinear sparsity $\delta n = o(n)$ regime
  - How does the DE threshold scale in this regime?
  - Shannon limit says that $\delta^* \leq j/k$
  - To understand this, numerical DE is not enough
- Let $\alpha = \dfrac{j\delta}{k-1}$ and $x_i = \dfrac{jy_i}{k-1}$ so that

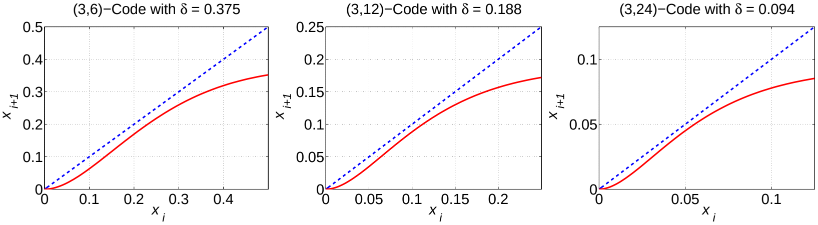
$$\begin{aligned} y_{i+1} &= \frac{j}{k-1}\delta\left(1-(1-x_i)^{k-1}\right)^{j-1} \\ &= \alpha\left(1-\left(1-\frac{jy_i}{k-1}\right)^{k-1}\right)^{j-1} \\ &\searrow \alpha\left(1-e^{-jy_i}\right)^{j-1} \end{aligned}$$

- Expression independent of k

DE Scaling Illustrated



Unscaled



Scaled

DE Scaling Analysis for the BEC

Theorem: DE Scaling for the BEC

For $\theta < 1$ and $j \geq 3$, there are constants $\overline{\alpha}_j$ and $k_0(j,\theta)$ such that

- for all $k \geq k_0(j,\theta)$ and $\delta \leq \theta \overline{\alpha}_j \left(j/(k-1)\right)$,
- iterative decoding on a BEC(δ) succeeds (w.h.p. as $n \to \infty$)
- using a randomly chosen (j,k)-regular LDPC code

- Example: $\overline{\alpha}_3 = 0.8184$ and $\theta = \frac{0.75}{0.8184}$ gives $k_0 = 9$
  - Implies $(3,k)$ codes achieve 75% of Shannon limit for all $k \geq 9$

- Proof follows from scaled DE and the concentration theorem

Reconstruction Guarantees for CS and DE

- DE for LM1: x_i = fraction of incorrect messages iteration i

$$x_{i+1} = \delta \left(1 - \left[1 - (1-\delta)\left(1-(1-x_i)^{k-1}\right)^{j-1} - x_i\right]^{k-1}\right)^{j-1}$$

  - Bad news: linear scaling $x_i \sim \frac{y_i}{k}$ fails!
  - Needs $x_i = \frac{\overline{\alpha}_j y_i}{(k-1)^{j/(j-1)}}$ and find that $\delta \sim \frac{\overline{\alpha}_j}{(k-1)^{j/(j-1)}}$

Reconstruction Guarantees for CS and DE

- DE for LM1: x_i = fraction of incorrect messages iteration i

$$x_{i+1} = \delta \left(1 - \left[1 - (1-\delta) \left(1 - (1-x_i)^{k-1} \right)^{j-1} - x_i \right]^{k-1} \right)^{j-1}$$

  - Bad news: linear scaling $x_i \sim \frac{y_i}{k}$ fails!
  - Needs $x_i = \frac{\overline{\alpha}_j y_i}{(k-1)^{j/(j-1)}}$ and find that $\delta \sim \frac{\overline{\alpha}_j}{(k-1)^{j/(j-1)}}$

- DE Reconstruction Guarantee: "Randomized Reconstruction"
  - Signal and code chosen randomly $\implies$ Success w.h.p. as $n \to \infty$

- BP Reconstruction Guarantee: "Uniform Reconstruction"
  - There exists Φ that works for all signals $\|x\|_H \leq \frac{m}{C \log n}$

DE Scaling Analysis of LM1 Decoding 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: Randomized Reconstruction with LM1

CS with (j,k)-regular LDPC codes and LM1 decoding provides randomized reconstruction (w.h.p. as $n \to \infty$) if $\delta \leq (k-1)^{-j/(j-1)}$.

- Oversampling ratio $\leq e \left\lceil \ln \frac{1}{\delta} \right\rceil$ with linear complexity in n
  - If symbol node degree $j = \left\lceil \ln \frac{1}{\delta} \right\rceil$ increases as $\delta \to 0$

Is Linear Scaling Possible for CS?

- "Linear Scaling" = $C\|x\|_H$ measurements suffice for all $\|x\|_H$
  - Yes: Exhaustive search, random codes, and strictly sparse
  - Not possible with noise or approximately sparse signals
- Information-theoretic analysis: $(j, jn^{1-\omega})$-regular codes
  - Random signal X_1^n with measurements Y_1^m where $m = n^\omega$
    - ⋆ Non-zero entries of X_1^n drawn i.i.d. ~ Z
    - ⋆ Average number of non-zero entries per "check" is λ
  - Reconstruction requires $H(Y_1^m) \geq H(X_1^n)$ which holds only if

$$H(Z) \geq \frac{\omega}{j-1} \log \frac{n}{\lambda}$$

    - ⋆ Linear scaling essentially requires $H(Z) = \infty$ (e.g., not discrete)

LM2 Achieves Linear Scaling

- 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 $j = 4$ non-zero entries per column
  - Non-zero entries drawn from a continuous distribution

Theorem: Randomized Reconstruction with LM2

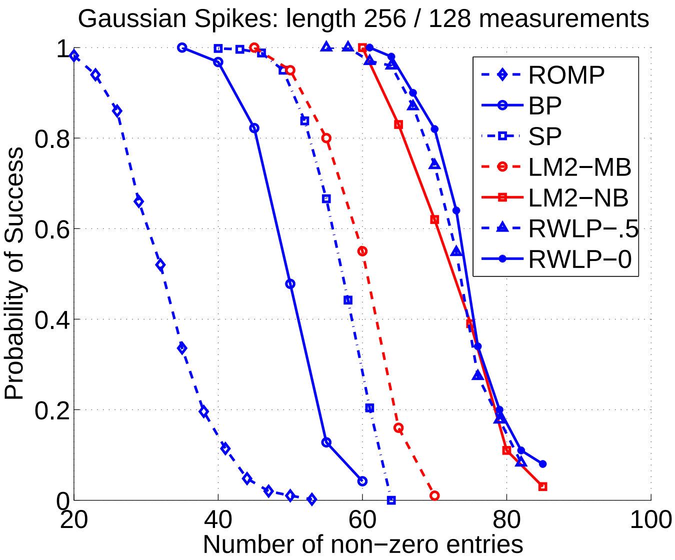
CS with $(4, k)$-regular LDPC codes and LM2 decoding (for $k \geq k_0$) provides randomized reconstruction (w.h.p. as $n \to \infty$) if $\delta \leq \frac{1}{3}(4/k)$

- Constant oversampling ratio of 3 with linear complexity in n

Comparison With Other Algorithms

- Regularized Orthgonal Matching Pursuit (ROMP)
  - Needell and Vershynin (2007)
- Basis Pursuit (BP)
  - Chen, Donoho, and Saunders (1998)
- Subspace Pursuit (SP)
  - Dai and Milenkovic (2008)
- Message-passing verification decoding (LM2-MB)
- Peeling-style verification decoding (LM2-NB)
  - Zhang and Pfister (2008)
- Iteratively Reweighted ℓ_p (RWLP-p)
  - Chartrand and Yin (2008)
- Gaussian Φ used except LM2-MB/LM2-NB use sparse $(3, 6)$

Simulation Results



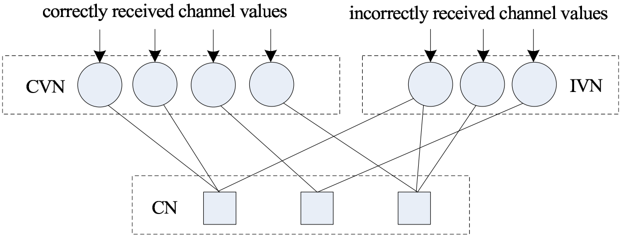
(Spikes have random locations and standard Gaussian amplitudes)

Outline

Stopping Set Analysis for the BEC

- A stopping set (s.s.) for iterative decoding on the BEC is
  - "a set of bit nodes connected even # of times to check nodes"
  - Iterative decoding progresses until it reaches a stopping set
- Uniform erasure correction
  - Let s_{min} be the size of the smallest stopping set
    - ⋆ Iterative decoding corrects all erasure patterns weight < s_{min}
  - For $j \geq 3$, random LDPC codes have $s_{min} = \gamma n$ w.h.p.
- One can also derive a scaling law for stopping sets
  - (j, k)-codes with large k have $s_{min} \sim ne(k-1)^{-j/(j-2)}$ w.h.p.

Stopping Set Analysis for LM1



- Peeling decoder removes symbols as they are verified
  - A check is removed if all edges are verified or correct
  - A symbol is verified if it is attached to a degree-1 check
- If neither condition occurs, the configuration is a stopping set

Uniform-in-Probability Reconstruction with LM1

- Signal Model
  - All length-n vectors x such that $\|x\|_H \leq \delta n$
- Sparse measurement matrix with m rows
  - (j,k)-regular code with fixed connections
  - Non-zero entries drawn from a continuous distribution
    - ★ Random for each x implies no false verification (FV) w.h.p.

Theorem: Uniform-in-Probability Reconstruction with LM1

There exists a (j,k)-regular LDPC code with $k > k_0$ such that LM1 succeeds for all x s.t. $\|x\|_H < nC_j(k-1)^{-j/(j-2)}$ assuming no FV.

- Based on combinatorial analysis of LM1 stopping sets

Summary

- Discussed connections between ECC and CS
  - Some basic connections now clear
  - Density evolution allows threshold analysis
  - Stopping-set analysis allows stronger guarantees
- Linear-time recovery algorithms for CS based on coding
  - First approach with a constant oversampling ratio

Open Problems

- Noisy observations and approximately sparse signals
  - Can one extend both the signal model and analysis?
- Signal is sparse is some basis unknown at encoder
  - Encoder cannot tranform into sparse basis before sampling
  - Can one add transform to decoding graph (e.g., ISI channels)?
- Code design for specific applications
  - Can protograph and other code designs be used for CS?
- Can one extend the DE scaling law to general channels?