Capacity-achieving Sparse Regression Codes via Aproximate Message Passing

Ramji Venkataraman

University of Cambridge

Joint work with Cynthia Rush (Yale) & Adam Greig (Cambridge)

ITA 2015

Want to construct efficient channel codes for the AWGN channel:

$$y = x + w, \qquad w \sim \mathcal{N}(0, \sigma^2)$$

Power constraint: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$

GOAL: Codes with fast encoding & decoding with probability of decoding error $\to 0$ at rates approaching

$$\mathcal{C} = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right)$$

Want to construct efficient channel codes for the AWGN channel:

$$y = x + w, \qquad w \sim \mathcal{N}(0, \sigma^2)$$

Power constraint: $\frac{1}{n}\sum_{i=1}^{n} x_i^2 \leq P$

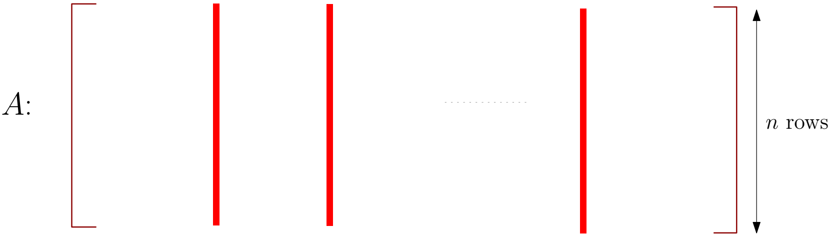
GOAL: Codes with fast encoding & decoding with probability of decoding error $\rightarrow 0$ at rates approaching

$$\mathcal{C} = \frac{1}{2}\log\left(1 + \frac{P}{\sigma^2}\right)$$

Sparse Regression Codes (SPARCs)

- Introduced by Barron and Joseph ['10, '12]
- Efficient, asymptotically capacity-achieving decoders proposed by [Barron-Joseph], [Barron-Cho]
- In this talk:
  - Fast, asymptotically capacity-achieving AMP decoder
  - Good empirical performance at practical block lengths

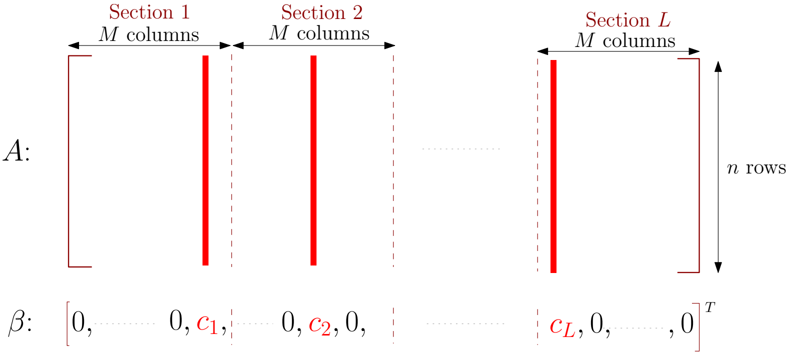
Codebook Construction



$$\beta: \quad \left[0, \cdots\cdots\, 0, c_1, \cdots\, 0, c_2, 0, \cdots\cdots\, c_L, 0, \cdots\cdots, 0\right]^T$$

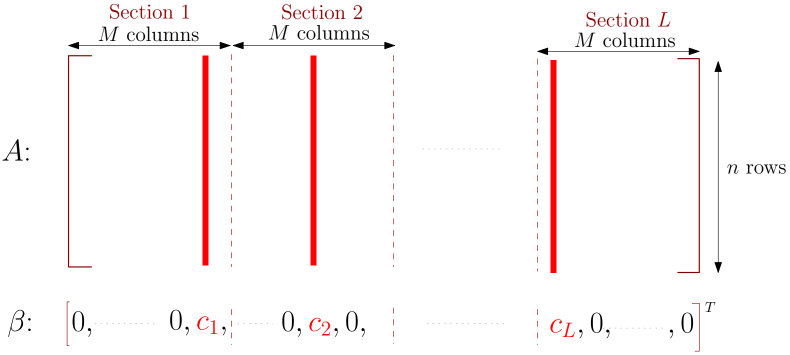
- **A**: design matrix or 'dictionary' with ind. $\mathcal{N}(0, 1/n)$ entries
- Codewords $\mathbf{A}\beta$ - *sparse* linear combinations of columns of **A**

SPARC Codebook



n rows, ML columns

SPARC Codebook

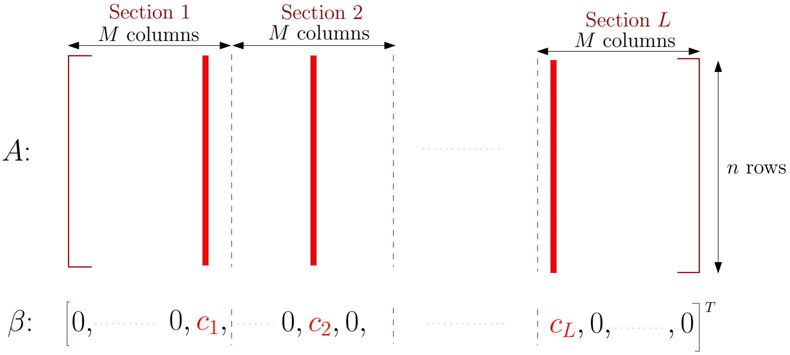


n rows, ML columns

Choosing M and L:

- For rate R codebook, need $M^L = 2^{nR}$

SPARC Codebook

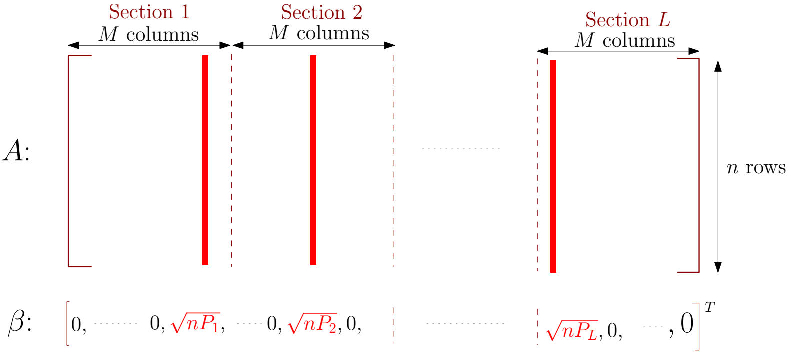


n rows, ML columns

Choosing M and L:

- For rate R codebook, need $M^L = 2^{nR}$
- Choose M polynomial in n, e.g., $\sqrt{n} \Rightarrow\ L \sim n/\log n$
- Size of A: polynomial in n

Power Allocation



Coefficients $c_1 = \sqrt{nP_1}, \ldots, c_L = \sqrt{nP_L}$ chosen such that

$$\sum_{\ell} P_\ell = P$$

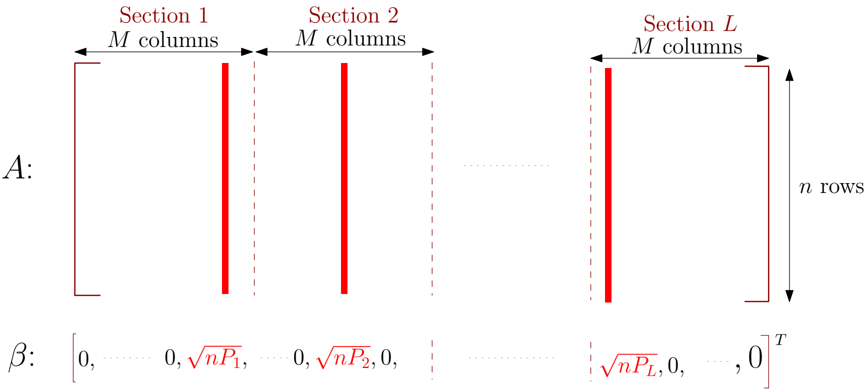
Examples:

1) Flat: $P_\ell = \frac{P}{L}$

2) Exponentially Decaying: $P_\ell \propto 2^{-\kappa\ell/L}$, with constant $\kappa > 0$

Allocation such that $P_\ell = \Theta(\frac{1}{L})$

Decoding



$$y = A\beta + w$$

Want efficient algorithm to decode $\hat{\beta}$ from y

Approximate Message Passing

Approximation of loopy belief propagation for dense graphs [Donoho-Maleki-Montanari '09], [Rangan '11], [Krzakala et al '12], [Schniter '11], ...

- For problems such as compressed sensing

$$y = A\beta + w$$

  A is $n \times N$ measurement matrix, β i.i.d. with known prior
- AMP iteratively produces esitmates $\beta^1, \beta^2, \ldots$
- Faster than ℓ_1-based convex optimization, similar performance

Approximate Message Passing

Approximation of loopy belief propagation for dense graphs [Donoho-Maleki-Montanari '09], [Rangan '11], [Krzakala et al '12], [Schniter '11], . . .

- For problems such as compressed sensing

$$y = A\beta + w$$

  A is $n \times N$ measurement matrix, β i.i.d. with known prior

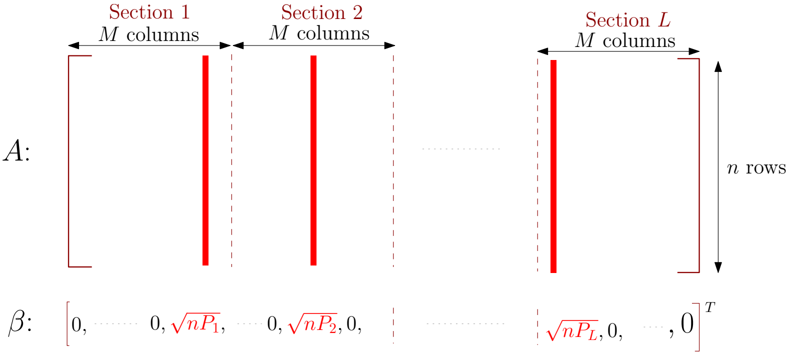
- AMP iteratively produces esitmates $\beta^1, \beta^2, \ldots$
- Faster than ℓ_1-based convex optimization, similar performance

- Rigorous asymptotic analysis [Bayati-Montanari '11] when A is Gaussian, *undersampling* ratio (n/N) is constant

$$\text{For each } t, \quad \lim_{n\to\infty} \frac{\|\beta - \beta^t\|^2}{n} = \sigma_t^2$$

- σ_t^2 can be computed via a scalar iteration — *state evolution*

AMP for SPARC



$$y = A\beta + w, \qquad w \text{ i.i.d.} \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

In SPARCs,

- β has one non-zero per section, section size $M \to \infty$
- The undersampling ratio $n/(ML) \to 0$.

AMP decoder can be derived by approximating a min-sum-like message passing algorithm for (1)

AMP Decoder

Set $\beta^0 = 0$. For $t \geq 0$:

$$z^t = y - A\beta^t + \frac{z^{t-1}}{\tau_{t-1}^2}\left(P - \frac{\|\beta^t\|^2}{n}\right),$$

$$\beta_i^{t+1} = \eta_i^t(\beta^t + A^* z^t), \qquad \text{for } i = 1, \ldots, ML$$

For $i \in$ section ℓ,

$$\eta_i^t(s) = \sqrt{nP_\ell}\,\frac{\exp\left(s_i\sqrt{nP_\ell}/\tau_t^2\right)}{\sum_{j\in \text{sec}_\ell}\exp\left(s_j\sqrt{nP_\ell}/\tau_t^2\right)}.$$

AMP Decoder

Set $\beta^0 = 0$. For $t \geq 0$:

$$z^t = y - A\beta^t + \frac{z^{t-1}}{\tau_{t-1}^2}\left(P - \frac{\|\beta^t\|^2}{n}\right),$$

$$\beta_i^{t+1} = \eta_i^t(\beta^t + A^* z^t), \qquad \text{for } i = 1, \ldots, ML$$

For $i \in$ section ℓ,

$$\eta_i^t(s) = \sqrt{nP_\ell}\,\frac{\exp\left(s_i\sqrt{nP_\ell}/\tau_t^2\right)}{\sum_{j\in \text{sec}_\ell} \exp\left(s_j\sqrt{nP_\ell}/\tau_t^2\right)}.$$

For large enough n, $\beta^t + A^* z^t$ has distribution close to $\beta + \tau_t Z$, where Z is i.i.d. $\sim \mathcal{N}(0,1)$

AMP Decoder

Set $\beta^0 = 0$. For $t \geq 0$:

$$z^t = y - A\beta^t + \frac{z^{t-1}}{\tau_{t-1}^2}\left(P - \frac{\|\beta^t\|^2}{n}\right),$$

$$\beta_i^{t+1} = \eta_i^t(\beta^t + A^* z^t), \qquad \text{for } i = 1, \ldots, ML$$

For $i \in$ section ℓ,

$$\eta_i^t(s) = \sqrt{nP_\ell}\, \frac{\exp\left(s_i \sqrt{nP_\ell}/\tau_t^2\right)}{\sum_{j \in \mathrm{sec}_\ell} \exp\left(s_j \sqrt{nP_\ell}/\tau_t^2\right)}.$$

Then $\beta^{t+1} = \eta^t(s) = \mathbb{E}[\beta \mid \beta + \tau_t Z = s]$. It is

- the MMSE estimate of β given the observation $\beta + \tau_t Z$
- $\propto$ the posterior probability of entry i of β being non-zero

The test statistics

Suppose

$$z^t = y - A\beta^t$$

The test statistics

Suppose

$$z^t = y - A\beta^t$$

Then

$$\beta^t + A^* z^t \;=\; \beta \;+\; A^* w \;+\; (\mathrm{I} - A^* A)(\beta - \beta^t)$$

The test statistics

Suppose

$$z^t = y - A\beta^t$$

$$\beta^t + A^* z^t \;=\; \beta \;+\; \underbrace{A^* w}_{\mathcal{N}(0,\sigma^2)} \;+\; (\mathrm{I} - A^* A)(\beta - \beta^t)$$

The test statistics

Suppose

$$z^t = y - A\beta^t$$

$$\beta^t + A^* z^t \; = \; \beta \; + \; \underbrace{A^* w}_{\mathcal{N}(0,\sigma^2)} \; + \; \underbrace{(\mathrm{I} - A^* A)}_{\approx \mathcal{N}(0,1/n)}(\beta - \beta^t)$$

The test statistics

Suppose

$$z^t = y - A\beta^t \; + \; \frac{z^{t-1}}{\tau_{t-1}^2}\left(P - \frac{\|\beta^t\|^2}{n}\right)$$

$$\beta^t + A^* z^t \; = \; \beta \; + \; \underbrace{A^* w}_{\mathcal{N}(0,\sigma^2)} \; + \; \underbrace{(\mathrm{I} - A^* A)}_{\approx \mathcal{N}(0,1/n)}(\beta - \beta^t)$$

The constants τ_t

$$\beta^t + A^* z^t \;\sim\; \beta + \tau_t Z$$

τ_t^2 is the variance of the noise in the test statistic after step t

$$\tau_0^2 = \sigma^2 + P,$$
$$\tau_{t+1}^2 = \sigma^2 + \frac{\mathbb{E}[\|\beta - \beta^{t+1}\|^2]}{n} = \sigma^2 + P(1 - x_{t+1}),$$

where

$$x_{t+1} = \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbb{E}\left[\frac{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right)}{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right) + \sum_{j=2}^{M} \exp\left(\frac{\sqrt{nP_\ell}}{\tau_t} U_j^\ell\right)} \right]$$

$\{U_j^\ell\}$ are i.i.d. $\sim \mathcal{N}(0,1)$

State Evolution

Thus, when $\beta^t + A^* z^t \sim \beta + \tau_t Z$:

$$\beta^{t+1} \sim \eta^t(\beta + \tau_t Z)$$

$$\beta^{t+2} \sim \eta^t(\beta + \tau_{t+1} Z)$$

$$\vdots$$

State Evolution

Thus, when $\beta^t + A^* z^t \sim \beta + \tau_t Z$:

$$\beta^{t+1} \sim \eta^t(\beta + \tau_t Z)$$

$$\beta^{t+2} \sim \eta^t(\beta + \tau_{t+1} Z)$$

$$\vdots$$

$\tau_{t+1}^2 = \sigma^2 + P(1 - x_{t+1})$, where

$$x_{t+1} = \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbb{E}\left[\frac{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right)}{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right) + \sum_{j=2}^{M} \exp\left(\frac{\sqrt{nP_\ell}}{\tau_t} U_j^\ell\right)} \right]$$

State Evolution

Thus, when $\beta^t + A^* z^t \;\sim\; \beta + \tau_t Z$:

$$\beta^{t+1} \;\sim\; \eta^t(\beta + \tau_t Z)$$

$$\beta^{t+2} \;\sim\; \eta^t(\beta + \tau_{t+1} Z)$$

$$\vdots$$

$\tau_{t+1}^2 = \sigma^2 + P(1 - x_{t+1})$, where

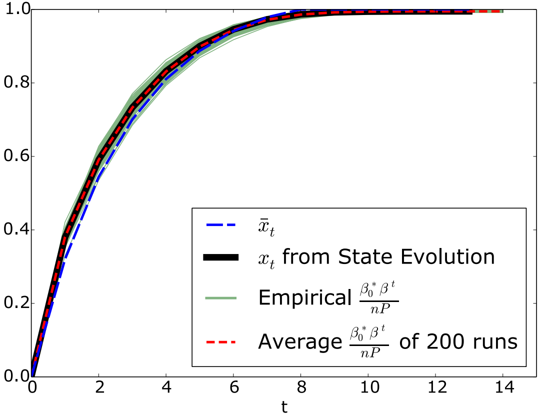
$$x_{t+1} = \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbb{E}\left[\frac{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right)}{\exp\left(\frac{\sqrt{nP_\ell}}{\tau_t}\left(U_1^\ell + \frac{\sqrt{nP_\ell}}{\tau_t}\right)\right) + \sum_{j=2}^{M} \exp\left(\frac{\sqrt{nP_\ell}}{\tau_t} U_j^\ell\right)} \right]$$

KEY property

- x_t *increases* with t for a finite number of steps T_n
- $x_t \approx 1 \;\Rightarrow\; \tau_t^2 \approx \sigma^2$, i.e., the test statistic is $\sim \beta + \sigma Z$, i.e.,
- AMP has effectively converted the A matrix to an identity!

x_t vs. t

SPARC: $M = 512, L = 1024, \text{snr} = 15, R = 0.7\mathcal{C}, P_\ell \propto 2^{-2\mathcal{C}\ell/L}$



$$x_t = \frac{1}{n}\mathbb{E}[\beta^* \beta^t]$$

"Power-weighted fraction of correctly decoded sections in β^t"

Asymptotics

Nice closed-form expression can be obtained for $\bar{x}_t := \lim_{n\to\infty} x_t$

Example: With $P_\ell \propto 2^{-2\mathcal{C}\ell/L}$

$$\bar{x}_t := \lim x_t = \frac{(1+\mathsf{snr}) - (1+\mathsf{snr})^{1-\xi_{t-1}}}{\mathsf{snr}},$$
$$\bar{\tau}_t^2 := \lim \tau_t^2 = \sigma^2 + P(1-\bar{x}_t) = \sigma^2\,(1+\mathsf{snr})^{1-\xi_{t-1}}$$

where $\xi_{-1} = 0$ and for $t \geq 0$,

$$\xi_t = \min\left\{\left(\frac{1}{2\mathcal{C}}\log\left(\frac{\mathcal{C}}{R}\right) + \xi_{t-1}\right),\ 1\right\}.$$

Asymptotics

Nice closed-form expression can be obtained for $\bar{x}_t := \lim_{n\to\infty} x_t$

Example: With $P_\ell \propto 2^{-2\mathcal{C}\ell/L}$

$$\bar{x}_t := \lim x_t = \frac{(1+\text{snr}) - (1+\text{snr})^{1-\xi_{t-1}}}{\text{snr}},$$

$$\bar{\tau}_t^2 := \lim \tau_t^2 = \sigma^2 + P(1-\bar{x}_t) = \sigma^2\,(1+\text{snr})^{1-\xi_{t-1}}$$

where $\xi_{-1} = 0$ and for $t \geq 0$,

$$\xi_t = \min\left\{ \left(\frac{1}{2\mathcal{C}} \log\left(\frac{\mathcal{C}}{R}\right) + \xi_{t-1} \right),\ 1 \right\}.$$

For $R < \mathcal{C}$, $\bar{x}_t \nearrow 1$ and $\bar{\tau}_t^2 \searrow 0$ in exactly $T^* = \lceil \frac{2\mathcal{C}}{\log(\mathcal{C}/R)} \rceil$ steps

Run AMP decoder for T^* steps to get $\beta^1, \ldots, \beta^{T*} \ \rightarrow\ \hat{\beta}$

Main Result

The *section error rate* of a decoder for SPARC $\mathcal{S}$ is

$$\mathcal{E}_{sec}(\mathcal{S}) := \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{1}\{\hat{\beta}_\ell \neq \beta_\ell\}.$$

Theorem

Fix any rate $R < \mathcal{C}$, and $b > 0$. Consider a sequence of rate R SPARCs $\{\mathcal{S}_n\}$ indexed by block length n, with design matrix parameters L and $M = L^b$, and power allocation $\propto 2^{-2\mathcal{C}\ell/L}$.

Then the section error rate of the AMP decoder converges to zero almost surely, i.e., for any $\epsilon > 0$,

$$\lim_{n_0\to\infty} P\left(\mathcal{E}_{sec}(\mathcal{S}_n) < \epsilon, \; \forall n \geq n_0\right) = 1.$$

Main Result

The *section error rate* of a decoder for SPARC $\mathcal{S}$ is

$$\mathcal{E}_{sec}(\mathcal{S}) := \frac{1}{L}\sum_{\ell=1}^{L} \mathbf{1}\{\hat{\beta}_\ell \neq \beta_\ell\}.$$

Theorem

Fix any rate $R < \mathcal{C}$, and $b > 0$. Consider a sequence of rate R SPARCs $\{\mathcal{S}_n\}$ indexed by block length n, with design matrix parameters L and $M = L^b$, and power allocation $\propto 2^{-2\mathcal{C}\ell/L}$.

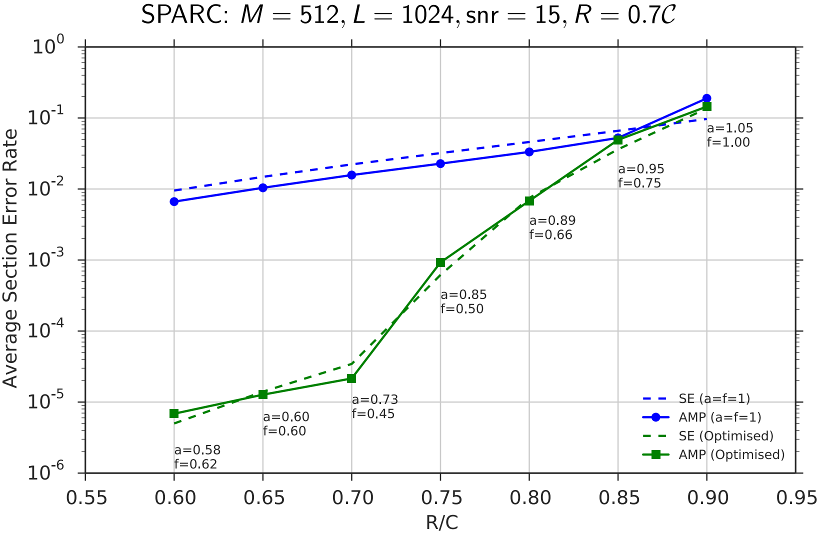
Then the section error rate of the AMP decoder converges to zero almost surely, i.e., for any $\epsilon > 0$,

$$\lim_{n_0 \to \infty} P\left(\mathcal{E}_{sec}(\mathcal{S}_n) < \epsilon, \ \forall n \geq n_0\right) = 1.$$

Proof: Show that asymptotically

- $s^t = (A^*\beta^t + z^t) \ \sim \ \beta + \bar{\tau}_t Z$, with $\bar{\tau}_t$ given by state evolution,
- $\|\beta - \beta^t\|^2 \ \xrightarrow{a.s.} \ P(1 - \bar{x}_t), \ t \leq T^*$

Empirical Performance



Power allocation plays a key role at finite block lengths!

Summary

- AMP provably achieves $R < \mathcal{C}$, complexity linear in matrix size
- [Barbier-Krzakala'14]:AMP decoder with different update rules

Summary

- AMP provably achieves $R < \mathcal{C}$, complexity linear in matrix size
- [Barbier-Krzakala'14]:AMP decoder with different update rules

Future directions:

- Rules of thumb for power allocation
- Use *Hadamard* instead of Gaussian design matrix:
  AMP complexity reduced to $\sim n^{1+\epsilon}$; much less memory
- Scaling laws à la polar codes, spatially coupled codes
- AMP encoder for SPARC lossy compression
- Implement binning, superposition by nesting SPARC channel & source codes: fast codes for Wyner-Ziv, Gelfand-Pinsker ...

`http://arxiv.org/abs/1501.05892`