Finite Sample Analysis of AMP

Cynthia Rush

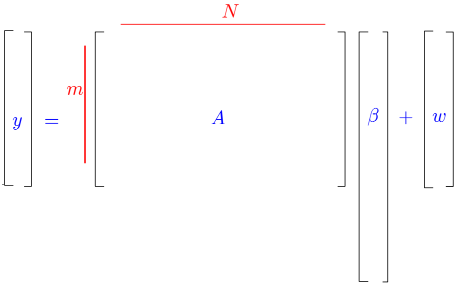
Department of Statistics
Columbia University

Statistical Physics and Machine Learning Back Together,
August 2018

Joint work with Adam Greig, Kuan Hsieh, and Ramji Venkataramanan (Cambridge)

High-Dimensional Regression

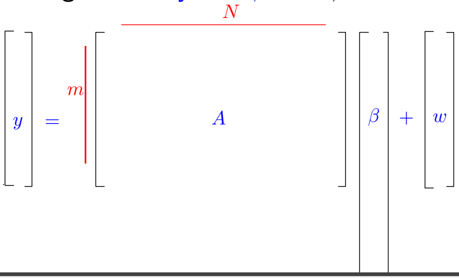
High-dimensional regression: $y = A\beta + w$,



- **Unknown vector**: $\beta \in \mathbb{R}^N$ i.i.d. with distribution p_β,
- **Measurement matrix**: $A \in \mathbb{R}^{m \times N}$,
- **Measurement noise**: $w \in \mathbb{R}^m$ i.i.d. with variance σ^2,
- **Sampling ratio**: $\frac{m}{N} \in (0, \infty)$ constant, denoted δ.

High-Dimensional Regression

High-dimensional regression: $y = A\beta + w$,



Approximate Message Passing (AMP)

- AMP is a low complexity, scalable algorithm to solve the above regression task.
- AMP derived as approximation of loopy belief propagation for dense graphs [Donoho-Maleki-Montanari '09], [Rangan '11], [Krzakala et al '12], [Schniter '11], . . .

AMP Performance Guarantees

- AMP iteratively produces estimates of β denoted $\beta^1, \beta^2, \ldots$.
- Rigorous asymptotic analysis [Bayati-Montanari '11] when A is Gaussian:

$$\text{For each } t, \quad \lim_{m \to \infty} \frac{1}{m} \|\beta - \beta^t\|^2 = \sigma_t^2.$$

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

AMP Performance Guarantees

- AMP iteratively produces estimates of β denoted $\beta^1, \beta^2, \ldots$.
- Rigorous asymptotic analysis [Bayati-Montanari '11] when A is Gaussian:

$$\text{For each } t, \quad \lim_{m \to \infty} \frac{1}{m} \|\beta - \beta^t\|^2 = \sigma_t^2.$$

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

State evolution still accurate for finite, large N

For each t,

$$P\left(\left| \frac{1}{m} \|\beta - \beta^t\|^2 - \sigma_t^2 \right| \geq \epsilon \right) \leq K_t e^{-\kappa_t N \epsilon^2}.$$

$t-$dependent constants K_t, κ_t.

AMP Algorithm

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

$$z^t = y - A\beta^t + \frac{z^{t-1}}{m} \sum_{i=1}^{N} \eta_t' \left([A^T z^{t-1} + \beta^{t-1}]_i \right),$$

$$\beta^{t+1} = \eta_t(\beta^t + A^T z^t),$$

- z^t is the 'modified residual' after step t
- η_t denoises the *effective observation* to produce β^{t+1}

AMP Algorithm

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

$$z^t = y - A\beta^t + \frac{z^{t-1}}{m} \sum_{i=1}^{N} \eta_t' \left([A^T z^{t-1} + \beta^{t-1}]_i \right),$$

$$\beta^{t+1} = \eta_t(\beta^t + A^T z^t),$$

- z^t is the 'modified residual' after step t
- η_t denoises the *effective observation* to produce β^{t+1}

The *correction* term in z^t ensures that for large enough N:

$$\beta^t + A^T z^t \approx \beta + \tau_t Z \quad \text{where } Z \text{ is } \mathcal{N}(0,1)$$

$\Rightarrow$ The *effective observation* $\beta^t + A^T z^t$ is true signal observed in independent Gaussian noise.

AMP Algorithm

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

$$z^t = y - A\beta^t + \frac{z^{t-1}}{m} \sum_{i=1}^{N} \eta'_t \left([A^T z^{t-1} + \beta^{t-1}]_i \right),$$

$$\beta^{t+1} = \eta_t(\beta^t + A^T z^t),$$

- z^t is the 'modified residual' after step t
- η_t denoises the *effective observation* to produce β^{t+1}

- If p_β is known, the Bayes-optimal choice for η_t which minimizes $\mathbb{E}[\|\beta - \beta^{t+1}\|^2]$ is
$$\eta_t(s) = \mathbb{E}[\beta \mid \beta + \tau_t Z = s]$$
- If p_β is unknown, partial knowledge about β can guide the choice of η_t.

The Modified Residual [Donoho-Maleki-Montanari '09]

Assume $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_i \sim \mathcal{N}(0, \sigma^2)$.

Suppose instead,

$$z^t = y - A\beta^t = w - A(\beta - \beta^t)$$

The Modified Residual [Donoho-Maleki-Montanari '09]

Assume $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_i \sim \mathcal{N}(0, \sigma^2)$.

Suppose instead,

$$z^t = y - A\beta^t = w - A(\beta - \beta^t)$$

Then effective observation:

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

The Modified Residual [Donoho-Maleki-Montanari '09]

Assume $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_i \sim \mathcal{N}(0, \sigma^2)$.

Suppose instead,

$$z^t = y - A\beta^t = w - A(\beta - \beta^t)$$

Then effective observation:

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

The Modified Residual [Donoho-Maleki-Montanari '09]

Assume $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_i \sim \mathcal{N}(0, \sigma^2)$.

Suppose instead,

$$z^t = y - A\beta^t = w - A(\beta - \beta^t)$$

Then effective observation:

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

$$\approx \beta + \sqrt{\sigma^2 + \frac{\mathbb{E}\|\beta - \beta^t\|^2}{m}}\, Z$$

The Modified Residual [Donoho-Maleki-Montanari '09]

Assume $A_{ij} \sim \mathcal{N}(0, 1/m)$ and $w_i \sim \mathcal{N}(0, \sigma^2)$.

Suppose instead,

$$z^t = y - A\beta^t = w - A(\beta - \beta^t) + \frac{z^{t-1}}{m} \sum_{i=1}^{N} \eta_t' \left([A^T z^{t-1} + \beta^{t-1}]_i \right)$$

Then effective observation:

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

$$\approx \beta + \sqrt{\sigma^2 + \frac{\mathbb{E}\|\beta - \beta^t\|^2}{m}} Z$$

- The 'correction' term asymptotically cancels out dependence between $(\mathrm{I} - A^T A)$ and $(\beta - \beta^t)$ so that

$$\beta^t + A^T z^t + \approx \beta + \tau_t Z_t, \quad \text{where } \tau_t^2 = \sigma^2 + \frac{1}{m}\mathbb{E}\|\beta - \beta^t\|^2$$

State Evolution

Define τ_t^2 as the variance of the noise in the effective observation after step t.

$$\beta^t + A^T z^t \approx \beta + \tau_t Z, \quad Z \sim \mathcal{N}(0, \mathbb{I}).$$

SE Equations:

Set $\tau_0^2 = \sigma^2 + \mathbb{E}\|\beta\|^2/m$,

$$\tau_t^2 = \sigma^2 + \frac{\mathbb{E}\|\beta - \beta^t\|^2}{m}.$$

State Evolution

Define τ_t^2 as the variance of the noise in the effective observation after step t.

SE Equations:
Set $\tau_0^2 = \sigma^2 + \mathbb{E}\|\beta\|^2/m$,

$$\tau_t^2 = \sigma^2 + \frac{\mathbb{E}\|\beta - \beta^t\|^2}{m} = \sigma^2 + \frac{\mathbb{E}\|\beta - \eta_t(\beta + \tau_{t-1} Z)\|^2}{m}.$$

$Z \sim \mathcal{N}(0,1)$ independent of $\beta \sim p_\beta$.

Assumptions

We make the following assumptions:

- **Measurement matrix**: i.i.d. $\sim \mathcal{N}(0, 1/m)$.
- **Signal**: i.i.d. $\sim p_\beta$, sub-Gaussian.
- **Measurement noise**: i.i.d. $\sim p_w$, sub-Gaussian, $\mathbb{E}[w_i^2] = \sigma^2$.
- **De-noising Functions** η_t: Lipschitz continuous with weak derivative η_t' which is differentiable except possibly at a finite number of points, with bounded derivative everywhere it exists.

Assumptions

We make the following assumptions:

- **Measurement matrix**: i.i.d. $\sim \mathcal{N}(0, 1/m)$.
- **Signal**: i.i.d. $\sim p_\beta$, sub-Gaussian.
- **Measurement noise**: i.i.d. $\sim p_w$, sub-Gaussian, $\mathbb{E}[w_i^2] = \sigma^2$.
- **De-noising Functions** η_t: Lipschitz continuous with weak derivative η_t' which is differentiable except possibly at a finite number of points, with bounded derivative everywhere it exists.

Pseudo-Lipschitz (PL) Loss Functions

A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is PL if there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^n$,

$$|\phi(x) - \phi(y)| \leq L(1 + \|x\| + \|y\|)\|x - y\|.$$

E.g. $\phi(x) = \|x\|_2^2$ (ℓ_2 loss) or $\phi(x) = \|x\|_1$ (ℓ_1 loss).

Main Result

Theorem

Under the assumptions of the previous slide, for any PL function $\phi : \mathbb{R}^2 \to \mathbb{R}$, $\epsilon \in (0,1)$, and $t \geq 0$,

$$P\left(\left|\frac{1}{m}\sum_{i=1}^{N}\phi(\beta_i^{t+1}, \beta_i) - \mathbb{E}[\phi(\eta_t(\beta + \tau_t Z), \beta)]\right| \geq \epsilon\right) \leq K_t e^{-\kappa_t N \epsilon^2},$$

for $Z \sim \mathcal{N}(0,1)$, $\beta \sim p_\beta$ independent with constants K_t, κ_t.

Main Result

Theorem

Under the assumptions of the previous slide, for any PL function $\phi : \mathbb{R}^2 \to \mathbb{R}$, $\epsilon \in (0,1)$, and $t \geq 0$,

$$P\Big(\Big|\frac{1}{m}\sum_{i=1}^{N} \phi(\beta_i^{t+1}, \beta_i) - \mathbb{E}[\phi(\eta_t(\beta + \tau_t Z), \beta)]\Big| \geq \epsilon\Big) \leq K_t e^{-\kappa_t N \epsilon^2},$$

for $Z \sim \mathcal{N}(0,1)$, $\beta \sim p_\beta$ independent with constants K_t, κ_t.

Choosing PL loss function $\phi(a,b) = (a-b)^2$, the Theorem proves

$$P\left(\left|\frac{1}{N}\|\beta^{t+1} - \beta_0\|^2 - \delta(\tau_{t+1}^2 - \sigma^2)\right| \geq \epsilon\right) \leq K_t e^{-\kappa_t N \epsilon^2},$$

for τ_{t+1}^2 computed via state evolution.

Main Result

Theorem
Under the assumptions of the previous slide, for any PL function $\phi : \mathbb{R}^2 \to \mathbb{R}$, $\epsilon \in (0, 1)$, and $t \geq 0$,

$$P\left(\left|\frac{1}{m}\sum_{i=1}^{N} \phi(\beta_i^{t+1}, \beta_i) - \mathbb{E}[\phi(\eta_t(\beta + \tau_t Z), \beta)]\right| \geq \epsilon\right) \leq K_t e^{-\kappa_t N \epsilon^2},$$

for $Z \sim \mathcal{N}(0, 1)$, $\beta \sim p_\beta$ independent with constants K_t, κ_t.

- This refines an asymptotic result proved by Bayati, Montanari [Trans. IT '11]
- The finite-sample result above implies the asymptotic result (via Borel-Cantelli).

Main Result

Theorem
Under the assumptions of the previous slide, for any PL function $\phi : \mathbb{R}^2 \to \mathbb{R}$, $\epsilon \in (0,1)$, and $t \geq 0$,

$$P\left(\left|\frac{1}{m}\sum_{i=1}^{N} \phi(\beta_i^{t+1}, \beta_i) - \mathbb{E}[\phi(\eta_t(\beta + \tau_t Z), \beta)]\right| \geq \epsilon\right) \leq K_t e^{-\kappa_t N \epsilon^2},$$

for $Z \sim \mathcal{N}(0,1)$, $\beta \sim p_\beta$ independent with constants K_t, κ_t.

Constants in the Bound:

- Constants $K_t = K_1 (K_2)^t (t!)^{10}$ and $\kappa_t = \kappa_1 \kappa_2^{-t} (t!)^{-22}$ where $K_1, K_2, \kappa_1, \kappa_2 > 0$ are universal constants.
- Indicates how large t can get for deviation prob. $\to 0$: $t = o\left(\frac{\log N}{\log \log N}\right)$

Main Result: Proof Sketch

Show $\beta^t + A^T z^t \sim \beta + \tau_t Z$, with τ_t given by state evolution.

Steps

1. Characterize the conditional distribution of the effective observation and residual as sum of i.i.d. Gaussians plus deviation term.

Show:

$$(\beta^t + A^T z^t - \beta)|_{\{\text{past},\,\beta,w\}} \stackrel{d}{=} \tau_t Z_t + \Delta_t,$$

$$(z^t - w)|_{\{\text{past},\,\beta,w\}} \stackrel{d}{=} \sqrt{\tau_t^2 - \sigma^2}\,\tilde{Z}_t + \tilde{\Delta}_t,$$

where Δ_t is complicated – constructed of inner products of other output from AMP iterations up to time t.

2. Inductively obtain concentration results showing that the deviation terms are small with high probability.

Role of the Correction Term

For Δ_t to concentrate at 0, correction term takes form $\gamma_{t-1} z^{t-1}$.

Ideal Value

For all $j \leq t-1$, want *simultaneously*:

$$\gamma_{t-1} := \left(\frac{\tau_j}{\tau_{t-1}}\right)\frac{1}{\delta}\mathbb{E}\left[Z_j\, \eta_{t-1}(\beta + \tau_{t-1} Z_{t-1})\right]$$

Role of the Correction Term

For Δ_t to concentrate at 0, correction term takes form $\gamma_{t-1} z^{t-1}$.

Ideal Value

For all $j \leq t-1$, want *simultaneously*:

$$\gamma_{t-1} := \left(\frac{\tau_j}{\tau_{t-1}}\right) \frac{1}{\delta} \mathbb{E}\left[Z_j \, \eta_{t-1}(\beta + \tau_{t-1} Z_{t-1})\right]$$

$$= \quad \ldots \qquad \text{Stein's Lemma}$$

$$= \frac{1}{\delta} \mathbb{E}\left[\eta'_{t-1}(\beta + \tau_{t-1} Z_{t-1})\right] \qquad \mathbb{E}\left[Z_j Z_{t-1}\right] = \frac{\tau_{t-1}}{\tau_j}$$

Role of the Correction Term

For Δ_t to concentrate at 0, correction term takes form $\gamma_{t-1} z^{t-1}$.

Ideal Value

$$\gamma_{t-1} := \frac{1}{\delta}\mathbb{E}\left[\eta'_{t-1}(\beta + \tau_{t-1} Z_{t-1})\right].$$

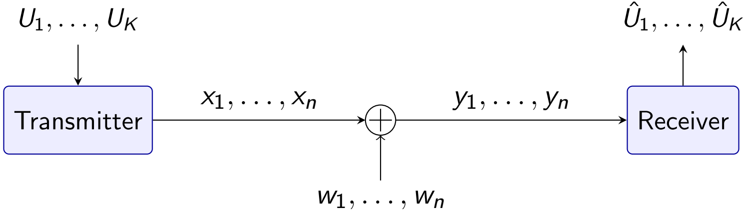
If we don't know p_β...

Use inductive hypothesis $\beta^{t-1} + A^T z^{t-1} \approx \beta + \tau_{t-1} Z_{t-1}$ to estimate

$$\hat{\gamma}_{t-1} = \frac{1}{\delta N}\sum_{i=1}^{N} \eta'_{t-1}([\beta^{t-1} + A^T z^{t-1}]_i)$$

An Application:
Sparse Regression Codes

The Additive White Gaussian Noise Channel



$$y_i = x_i + w_i, \quad i = 1, \ldots n$$

- Want to convey K bits in n channel uses
- Additive Gaussian noise: w_i iid $\sim \mathcal{N}(0, \sigma^2)$
- Average power constraint: $\frac{1}{n}\sum_i x_i^2 \leq P$
- Rate: $R = K/n$ bits/transmission
- Capacity: $\mathcal{C} = \frac{1}{2}\log(1 + snr)$

> GOAL: Codes with fast encoding & decoding with
> $Prob(\underline{\hat{U}} \neq \underline{U}) \to 0$ at rates R approaching $\mathcal{C}$

Sparse Regression Codes (SPARCs)

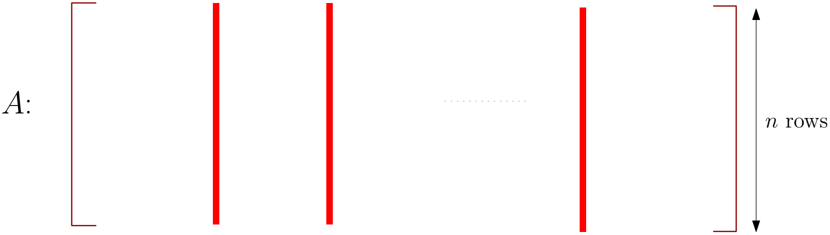
- Introduced by Barron and Joseph & shown to be capacity achieving with feasible decoding ['10, '12]

- We study an *approximate message passing* decoder:

  – Low-complexity, provably capacity-achieving with near-exponential decay of error probability

  – Good performance at practical block lengths

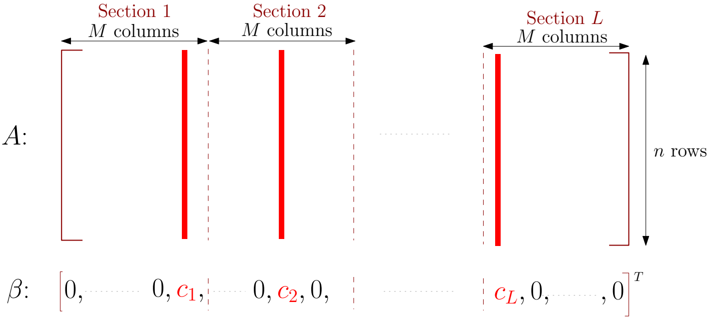
SPARC Codebook Construction



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

- A: 'Dictionary' with independent $\mathcal{N}(0, 1/n)$ entries
- Codewords $A\beta$ - *sparse* linear combinations of columns of A

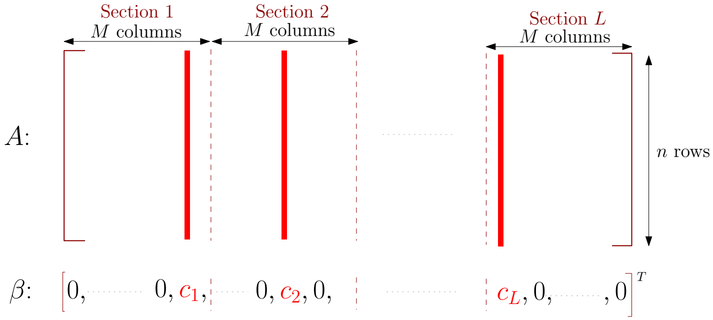
SPARC Codebook Construction



n rows, ML columns

- Codeword $A\beta$ is linear combinations of L columns of A.
- Message vector β has L entries with non-zero values $c_1, c_2, \ldots, c_L$ fixed.
- Message bits determine the location of the non-zeros.

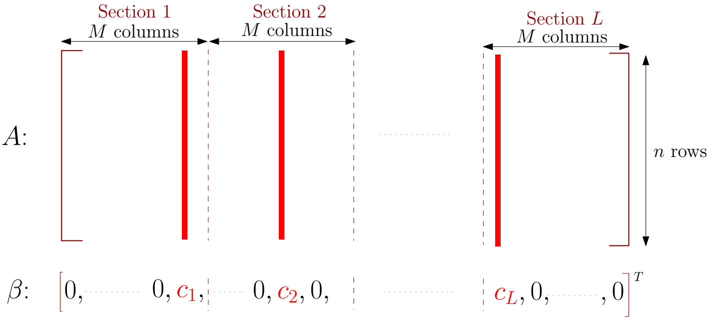
SPARC Codebook Construction



Choosing M and L:

- $M^L = 2^{nR}$ codewords for rate R, hence $L \log M = nR$
- Encoding: L chunks of input bits, each chunk with $\log M$ bits

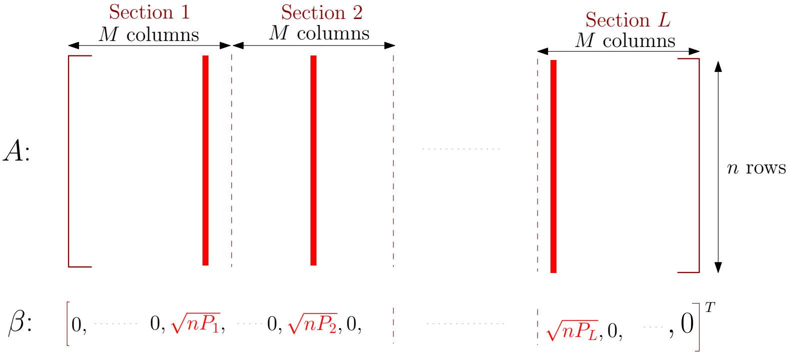
SPARC Codebook Construction



Choosing M and L:

- $M^L = 2^{nR}$ codewords for rate R, hence $L \log M = nR$
- Encoding: L chunks of input bits, each chunk with $\log M$ bits
- Choosing $M = L^a$, we get $L \sim n / \log n$
- Size of $A = n \times ML$, also polynomial in n

Power Allocation



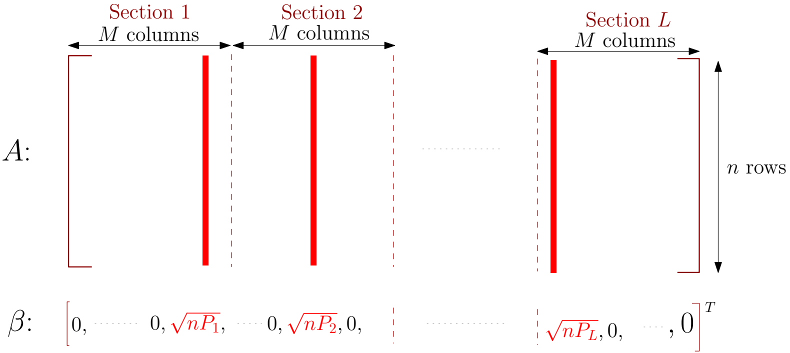
$c_1 = \sqrt{nP_1}, \ldots, c_L = \sqrt{nP_L}$ chosen such that $\sum_\ell P_\ell = P$

Examples

1. Flat: $P_\ell = \frac{P}{L}$
2. Exponentially Decaying: $P_\ell \propto e^{-\kappa \ell / L}$, constant $\kappa > 0$

For all power allocations, $P_\ell = \Theta(\frac{1}{L})$, $\sqrt{nP_\ell} = \Theta(\sqrt{\log M}) \quad \forall \ell$

Optimal Decoding



Channel output: $y = A\beta + w$

- $\hat{\beta}_{\text{opt}} = \arg\min_{\hat{\beta} \in \mathcal{B}} \|y - A\hat{\beta}\|^2$
- Probability of decoding error falls exponentially with n
  Similar performance to Shannon-style random coding [Joseph-Barron '12]
- Decoding complexity grows exponential with n

Feasible Decoders

- Adaptive successive decoders by [Joseph-Barron '12], [Barron-Cho '13] are asymptotically capacity-achieving
- **AMP decoding: [Rush, Greig, Venkataramanan '15], [Barbier, Krzakala '15]**

  – Asymptotically capacity-achieving [Rush, Greig, Venkataramanan '15]

  – Good performance at practical block lengths

  – Near-exponential decay of the error probability [Rush, Venkataramanan '17]

  – Finite sample analysis gives guidance on how to choose code parameters at finite block lengths..

Performance of AMP Decoder

Some details I'm leaving out...

- Derivation of AMP algorithm for SPARCs problem.
- Prove state evolution accurately predicts performance.

State evolution tells us:

- Run AMP decoder for T^* iterations, where

$$T^* := 1 + \left\lceil \left(\frac{1}{2\mathcal{C}} \log \left(\frac{\mathcal{C}}{R(1+\alpha/2)} \right) \right)^{-1} \right\rceil,$$

  $\alpha \in [0, 1/2]$ constant.

- After $T^* - 1$ iterations, we are guaranteed that

$$\sigma^2 \leq \tau_{T^*}^2 \leq \sigma^2 + \kappa M^{-c\alpha^2}.$$

- At final iteration T^*, generate

$$\beta^{T^*} = \eta_{T^*}\Big(\underbrace{A^T z^{T^*-1} + \beta^{T^*-1}}_{\approx \beta + \sigma Z} \Big).$$

Performance of AMP Decoder

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: [Rush, Greig, Venkataramanan '15]
Fix any rate $R < \mathcal{C}$ and $b > 0$. Consider a SPARC $\mathcal{S}_n$ with rate R, block length n, design matrix parameters L and $M = L^b$ determined according to

$$L \log M = nR,$$

and exponentially decaying power allocation, $P_\ell \propto 2^{-2\mathcal{C}\ell/L}$.
Then the section error rate of the AMP decoder converges to zero almost surely, i.e., for any $\epsilon > 0$,

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

Performance of AMP Decoder

Theorem: [Rush, Venkataramanan '17]
Fix any rate $R < \mathcal{C}$. Consider a SPARC $\mathcal{S}_n$ with rate R, block length n, design matrix parameters L, M determined according to

$$L \log M = nR,$$

and an exponentially decaying power allocation with $P_\ell \propto 2^{-2\mathcal{C}\ell/L}$. Let $\epsilon > \kappa M^{-c\alpha^2}$. Then, the section error rate of the AMP decoder satisfies

$$\mathbb{P}\left(\mathcal{E}_{sec}(\mathcal{S}_n) > \epsilon\right) \leq \; K_{T^*} \exp\left\{ \frac{-\kappa_{T^*} L \epsilon^2}{(\log M)^{2T^*-1}} \right\},$$

- Constants K_{T^*}, κ_{T^*} depend only on T^*, no further dependence on n, M, L, ϵ.
- The dependence on the rate R is only via T^*:
  Recall $T^* \approx 1 + 2\mathcal{C} / \log\left(\frac{\mathcal{C}}{R}\right)$.

Error Exponent

Consider the choice $M = L^a$ for constant $a > 0 \;\Rightarrow\; L \sim \frac{n}{\log n}$

For this choice:

$$\mathbb{P}\left(\mathcal{E}_{sec}(\mathcal{S}_n) > \epsilon\right) \leq \; K_{T^*} \exp\left\{\frac{-n\kappa_{T^*}\epsilon^2}{(\log n)^{2T^*}}\right\}$$

- Can also consider other choices such as $L \sim \frac{n}{\log\log n}, \quad M \sim \log n$
- Leads to different trade-offs with respect to error exponent, complexity, and gap-to-capacity

Performance of AMP Decoder

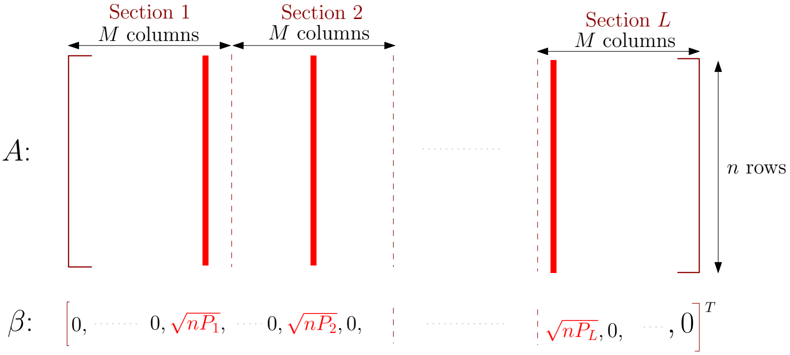
Finite sample analysis allows us to analyze the gap to capacity Δ_R:

- For fixed section error rate ϵ, how fast can R approach $\mathcal{C}$ with growing n?
- For the choice $M = L^a$,

$$\Delta_R \text{ is of order } \sqrt{\frac{\log \log n}{\log n}}.$$

- For AWGN channels, no known coding scheme that provably achieves a polynomial gap to capacity with efficient decoding.

Choosing L, M



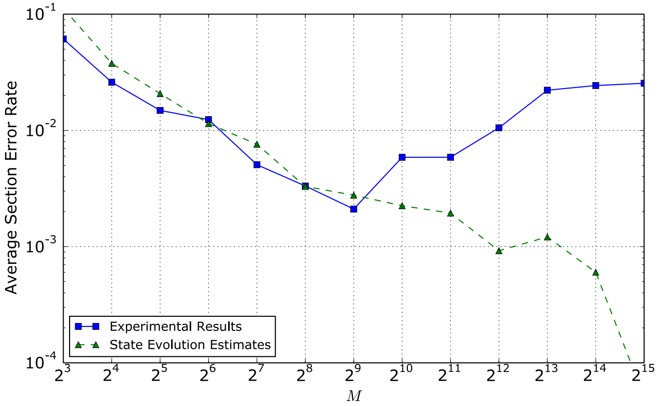
$$L \log M = nR$$

Larger $L, M \Rightarrow$ better performance?

Let's try increasing M with L fixed

Choosing L, M

- State evolution prediction: $0 \leq \frac{1}{n}\mathbb{E}\|\beta - \beta^T\|^2 \leq \kappa M^{-c\alpha^2}$
- So we might expect performance to improve with $\uparrow M$

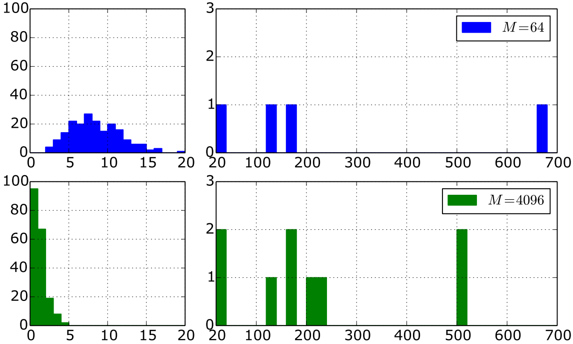


AMP performance with increasing M, for $L = 1024$, $R = 1.5$, $\frac{E_b}{N_0} = 5.7dB$

[Greig and Venkataramanan '17]

$$\mathbb{P}\left(\mathcal{E}_{sec}(\mathcal{S}_n) > \epsilon\right) \leq \; K_{T^*} \exp\left\{ \frac{-\kappa_{T^*} L \epsilon^2}{(\log M)^{2T^*-1}} \right\},$$

- Larger $M \Rightarrow$ worse concentration (with $L, R,$ snr fixed)
- T^* is of the order of tens, so this effect is significant!



$M = 64$ vs. $M = 4096$: Histogram of section errors

Increasing L improves concentration
Increasing M useful up to a point

Summary

Finite Sample AMP

- SE accurate in predicting AMP performance for large, finite N.
- Prob. of deviation from SE predictions falls exponentially like $N\epsilon^2$.
- Theoretical support for empirical findings of such.

Future directions:

- Extensions to LASSO risk, high-dimensional M-estimation, low-rank matrix estimation, AMP with spatially coupled matrices, and GAMP?
- Theoretical results for general A matrices (iid uniform Bernoulli, partial DFT, . . .)

Summary

AMP for SPARCs

- AMP for low-complexity SPARC decoding.
- Any rate $R < \mathcal{C}$, probability of section error rate $> \epsilon$ decays exponentially as

$$L\epsilon^2/(\log M)^{T^*} \approx n\epsilon^2/(\log n)^{T^*}.$$

Future directions:

- Theoretical Guarantees for the Hadamard-based SPARC.
- Theoretical analysis of spatially-coupled design matrices to improve performance at finite block-lengths.
- Will combining power allocation & spatial coupling give better empirical performance at reasonable block lengths near $\mathcal{C}$?
- The BIG question: Can we design feasible decoders with $O(1/n^\alpha)$ gap to capacity for some $\alpha \in (0, 1/2)$?