Sparse Regression Codes

Andrew Barron  Ramji Venkataramanan
Yale University  University of Cambridge

Joint work with Antony Joseph, Sanghee Cho, Cynthia Rush, Adam Greig, Tuhin Sarkar, Sekhar Tatikonda

ISIT 2016
Part II of the tutorial:

- Approximate message passing (AMP) decoding
- Power-allocation schemes to improve finite block-length performance

(Joint work with Cynthia Rush and Adam Greig)
SPARC Decoding

Channel output $y = A\beta + \varepsilon$

Want efficient algorithm to decode $\beta$ from $y$
AMP for Compressed Sensing

- Approximation of loopy belief propagation for dense graphs
  [Donoho-Montanari-Maleki ’09, Rangan ’11, Krzakala et al ’11…]
- Compressed sensing (CS): Want to recover $\beta$ from

$$y = A\beta + \varepsilon$$

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

In CS, we often solve LASSO: $\hat{\beta} = \arg \min_\beta \| y - A\beta \|_2^2 + \lambda \| \beta \|_1$
Min-Sum Message Passing for LASSO

Want to compute \( \hat{\beta} = \arg \min_{\beta} \sum_{i=1}^{n} (y_i - (A\beta)_i)^2 + \lambda \sum_{j=1}^{N} |\beta_j| \)
Min-Sum Message Passing for LASSO

Want to compute \( \hat{\beta} = \arg \min_\beta \sum_{i=1}^n (y_i - (A\beta)_i)^2 + \lambda \sum_{j=1}^N |\beta_j| \)

For \( j = 1, \ldots, N, \quad i = 1, \ldots, n \):

\[
M_{j \rightarrow i}^t(\beta_j) = \lambda |\beta_j| + \sum_{i' \in [n] \setminus i} \hat{M}_{i' \rightarrow j}^{t-1}(\beta_j)
\]
Min-Sum Message Passing for LASSO

Want to compute \( \hat{\beta} = \arg\min_\beta \sum_{i=1}^{n} (y_i - (A\beta)_i)^2 + \lambda \sum_{j=1}^{N} |\beta_j| \)

For \( j = 1, \ldots, N \), \( i = 1, \ldots, n \):

\[
M_{j\rightarrow i}(\beta_j) = \lambda |\beta_j| + \sum_{i' \in [n] \setminus i} \hat{M}_{i'\rightarrow j}^{t-1}(\beta_j)
\]

\[
\hat{M}_{i\rightarrow j}^{t}(\beta_j) = \min_{\beta \setminus \beta_j} (y_i - (A\beta)_i)^2 + \sum_{j' \in [N] \setminus j} M_{j'\rightarrow i}^{t}(\beta_{j'})
\]


\[
M^{t}_{j \rightarrow i}(\beta_j) = \lambda |\beta_j| + \sum_{i' \in [n]\setminus i} \hat{M}^{t-1}_{i' \rightarrow j}(\beta_j)
\]

\[
\hat{M}^{t}_{i \rightarrow j}(\beta_j) = \min_{\beta \setminus \beta_j} (y_i - (A\beta)_i)^2 + \sum_{j' \in [N]\setminus j} M^{t}_{j' \rightarrow i}(\beta_{j'})
\]

But computing these messages is infeasible:
— Each message needs to be computed for all \( \beta_j \in \mathbb{R} \)
— There are \( nN \) such messages

Further, the factor graph is not anything like a tree!
Quadratic Approximation of messages

Messages approximated by two numbers:

\[ r_{i \rightarrow j}^t = y_i - \sum_{j' \in [N] \setminus i} A_{ij'} \beta_{j' \rightarrow i}^t \]
Quadratic Approximation of messages

Messages approximated by two numbers:

\[ r_{i \rightarrow j}^t = y_i - \sum_{j' \in [N] \setminus i} A_{ij'} \beta_{j' \rightarrow i}^t \]
\[ \beta_{j \rightarrow i}^{t+1} = \eta_t \left( \sum_{i' \in [n] \setminus i} A_{i'j} r_{i' \rightarrow j}^t \right) \]

- For LASSO, \( \eta_t \) is the soft-thresholding operator
- We still have \( nN \) messages in each step . . .
\[ r^t_{i \rightarrow j} = y_i - \sum_{j' \in [N]} A_{ij'} \beta^t_{j' \rightarrow i} + A_{ij} \rho^t_{j \rightarrow i} \]

\[ \beta^{t+1}_{j \rightarrow i} = \eta_t \left( \sum_{i' \in [n]} A_{i'j} r^t_{i' \rightarrow j} - A_{ij} r^t_{i \rightarrow j} \right) \]

Using Taylor approximations . . .
The AMP algorithm for LASSO

[Donoho-Montanari-Maleki ’09, Rangan ’11, Krzakala et al ’11...]

\[
\begin{align*}
    r^t &= y - A\beta^t + r^{t-1} \frac{\|\beta^t\|_0}{n} \\
    \beta^{t+1} &= \eta_t(A^T r^t + \beta^t)
\end{align*}
\]

- AMP iteratively produces estimates \(\beta^0 = 0, \beta^1, \ldots, \beta^t, \ldots\)
- \(r^t\) is the ‘modified residual’ after step \(t\)
- \(\eta_t\) denoises the effective observation to produce \(\beta^{t+1}\)
The AMP algorithm for LASSO

[Donoho-Montanari-Maleki ’09, Rangan ’11, Krzakala et al ’11 . . . ]

$$r^t = y - A\beta^t + r^{t-1}/\|\beta^{t}\|_0^n$$

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

- AMP iteratively produces estimates $\beta^0 = 0, \beta^1, \ldots, \beta^t, \ldots$
- $r^t$ is the ‘modified residual’ after step $t$
- $\eta_t$ denoises the effective observation to produce $\beta^{t+1}$

The momentum term in $r^t$ ensures that asymptotically

$$A^T r^t + \beta^t \approx \beta + \tau_t Z_t \quad \text{where } Z_t \text{ is } \mathcal{N}(0, I)$$

$\Rightarrow$ The effective observation $A^T r^t + x^t$ is true signal observed in independent Gaussian noise
AMP for SPARC Decoding

\[ A: \]

\[
\begin{bmatrix}
0, & \cdots, & 0, \sqrt{nP_1}, & \cdots, & 0, \sqrt{nP_2}, 0, & \cdots, & \sqrt{nP_L}, 0, & \cdots, & 0
\end{bmatrix}^T
\]

\[ \beta: \]

\[ y = A\beta + \varepsilon, \quad \varepsilon \text{ i.i.d. } \sim \mathcal{N}(0, \sigma^2) \]

SPARC decoding is a different optimization problem from LASSO:

- Want \( \arg \min_{\beta} \| Y - A\beta \|^2 \) s.t. \( \beta \) is a SPARC message
- \( \beta \) has one non-zero per section, section size \( M \to \infty \)
- The undersampling ratio \( n/(ML) \to 0 \).

Let us revisit the (approximated) min-sum updates . . .
Approximated Min-Sum

\[ r_{i \rightarrow j}^t = y_i - \sum_{j' \in [N] \setminus i} A_{ij'} \beta_{j' \rightarrow i}^t \]

\[ \beta_{j \rightarrow i}^{t+1} = \eta_{t,j} \left( \sum_{i' \in [n] \setminus i} A_{i'j} r_{i' \rightarrow j}^t \right) \]

If for \( j \in [N] \), \( stat_{t,j} \) is approximately distributed as \( \beta_j + \tau_t Z_{t,j} \), then the Bayes optimal choice of \( \eta_{t,j} \) is ...
Approximated Min-Sum

\[ r_{i \rightarrow j}^t = y_i - \sum_{j' \in [N] \setminus i} A_{ij'} \beta_{j' \rightarrow i}^t \]

\[ \beta_{j \rightarrow i}^{t+1} = \eta_{t,j} \left( \sum_{i' \in [n] \setminus i} A_{i'j} r_{i' \rightarrow j}^t \right) \]

If for \( j \in [N] \), \( stat_{t,j} \) is approximately distributed as \( \beta_j + \tau_t Z_{t,j} \), then the Bayes optimal choice of \( \eta_{t,j} \) is ...
Bayes Optimal $\eta_t$

$$\eta_t(\text{stat}_t = s) = \mathbb{E}[\beta | \beta + \tau_t Z = s]:$$

$$\eta_{t,j}(s) = \sqrt{n P_\ell} \frac{\exp \left( s_j \sqrt{n P_\ell / \tau_t^2} \right)}{\sum_{j' \in \text{sec}_\ell} \exp \left( s_{j'} \sqrt{n P_\ell / \tau_t^2} \right)}, \quad j \in \text{section } \ell.$$ 

Note that $\beta^{t+1}$ is

- the MMSE estimate of $\beta$ given the observation $\beta + \tau_t Z_t$
- $\propto$ the posterior probability of entry $j$ of $\beta_j$ being non-zero
Bayes Optimal $\eta_t$

$\eta_t(stat_t = s) = \mathbb{E}[\beta \mid \beta + \tau_t Z = s]$:

$$\eta_{t,j}(s) = \sqrt{nP_\ell} \frac{\exp \left( s_j \sqrt{nP_\ell / \tau_t^2} \right)}{\sum_{j' \in \text{sec}_\ell} \exp \left( s_{j'} \sqrt{nP_\ell / \tau_t^2} \right)}, \quad j \in \text{section } \ell.$$  

Note that $\beta^{t+1}$ is

- the MMSE estimate of $\beta$ given the observation $\beta + \tau_t Z_t$
- $\propto$ the posterior probability of entry $j$ of $\beta_j$ being non-zero

$$r_{i \rightarrow j}^t = y_i - \sum_{j' \in [N]} A_{ij'} \beta_{j' \rightarrow i}^t + A_{ij} \beta_{j \rightarrow i}^t$$

$$\beta_{j \rightarrow i}^{t+1} = \eta_t \left( \sum_{i' \in [n]} A_{i'j} r_{i' \rightarrow j}^t - A_{ij} r_{i \rightarrow j}^t \right)$$  

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

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

$$\beta_{j}^{t+1} = \eta_{t,j} (A^T r^t + \beta^t), \quad \text{for } j = 1, \ldots, ML.$$
AMP Decoder

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

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

$$\beta^{t+1}_j = \eta_{t,j}(A^T r^t + \beta^t), \quad \text{for } j = 1, \ldots, ML$$

$$\eta_{t,j}(s) = \sqrt{nP_\ell} \frac{\exp \left( s_j \sqrt{nP_\ell/\tau_t^2} \right)}{\sum_{j' \in \text{sec}_\ell} \exp \left( s_{j'} \sqrt{nP_\ell/\tau_t^2} \right)}, \quad j \in \text{section } \ell.$$ 

$\beta^{t+1}$ is the MMSE estimate of $\beta$ given that $\beta^t + A^T r^t \approx \beta + \tau_t Z_t$
The statistic $\beta^t + A^T r^t$

Suppose

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

$$\beta^t + A^T r^t = \beta + \underbrace{A^T \varepsilon}_{\mathcal{N}(0,\sigma^2)} + \underbrace{(I - A^T A)(\beta^t - \beta)}_{\approx \mathcal{N}(0,1/n)}$$
The statistic $\beta^t + A^T r^t$

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

$$\beta^t + A^T r^t = \beta + \frac{A^T w}{N(0,\sigma^2)} + (I - A^T A)(\beta^t - \beta)$$

$$+ \frac{A^T r^{t-1}}{\tau_{t-1}^2} \left( P - \frac{\|\beta^t\|^2}{n} \right)$$

- Momentum term asymptotically cancels out dependence between $(I - A^T A)$ and $(\beta - \beta^t)$ so that $\beta^t + A^T r^t \approx \beta + \tau_t Z_t$, where $\tau_t^2 = \sigma^2 + \frac{1}{n} \mathbb{E}\|\beta - \beta^t\|^2$

- Recall that the plain residual does not give statistic with the desired representation
Iteratively compute variances $\tau_t^2$

$$\tau_0^2 = \sigma^2 + P$$
$$\tau_t^2 = \sigma^2 + P(1 - x_t(\tau_{t-1})), \quad t \geq 1$$

where

$$x_t(\tau_{t-1}) = \sum_{\ell=1}^L \frac{P_\ell}{P} \mathbb{E} \left[ \exp \left( \frac{\sqrt{nP_\ell}}{\tau_{t-1}} \left( U^\ell_1 + \frac{\sqrt{nP_\ell}}{\tau_{t-1}} \right) \right) \right]$$

$$\{U^\ell_j\} \text{ are i.i.d. } \sim \mathcal{N}(0, 1)$$

With $\text{stat}_t = \beta + \tau_t Z$:

$$\frac{1}{n} \mathbb{E} \| \beta - \beta^t \|^2 = P(1 - x_t) \quad \text{and} \quad \frac{1}{n} \mathbb{E} [\beta^T \beta^t] = \frac{1}{n} \mathbb{E} \| \beta^t \|^2 = P x_t$$

- $x_t$: Expected power-weighted fraction of correctly decoded sections after step $t$
- $P(1 - x_t)$: interference due to undecoded sections
$x_t$ vs. $t$

SPARC: $M = 512$, $L = 1024$, $\text{snr} = 15$, $R = 7C$, $P_\ell \propto 2^{-2C\ell/L}$

\[
x_t = \frac{1}{nP} \mathbb{E}[\beta^T \beta^t], \quad x_t^* = \frac{1}{nP} \beta^T \beta^t
\]

“Power-weighted fraction of correctly decoded sections in $\beta^t$”
State Evolution

\[ \text{stat}_t = A^T r^t + \beta^t \approx \beta + \tau_t Z_t \]

with \( \tau_t^2 = \sigma^2 + P(1 - x_t) \), where \( x_t = x(\tau_{t-1}^2) \):

\[ x(\tau_{t-1}) = \sum_{\ell=1}^{L} \frac{P_\ell}{P} E \left[ \exp \left( \frac{\sqrt{nP_\ell}}{\tau_{t-1}} (U_1^{\ell} + \frac{\sqrt{nP_\ell}}{\tau_{t-1}}) \right) \right. \]

\[ \left. \exp \left( \frac{\sqrt{nP_\ell}}{\tau_{t-1}} (U_1^{\ell} + \frac{\sqrt{nP_\ell}}{\tau_{t-1}}) \right) + \sum_{j=2}^{M} \exp \left( \frac{\sqrt{nP_\ell}}{\tau_{t-1}} U_j^{\ell} \right) \right] \]

**KEY Property**

- Starting from 0, \( x_t \) increases with \( t \) for a finite number of steps \( T_n \) where \( x_{T_n} \approx 1 \)
- Starting from \( \tau_0^2 = \sigma^2 + P \), the \( \tau_t^2 \) decreases with \( t \) until \( \approx \sigma^2 \)
- AMP has effectively converted the \( A \) matrix to an identity!
State Evolution Asymptotics

Lemma [Rush, Greig, Venkataramanan ’15]:

\[ \bar{x}(\tau) := \lim x(\tau) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_{\ell}}{P} 1\{c_{\ell} > 2R_{\tau}^2\} \]

where \( c_{\ell} = LP_{\ell} \) and \( R \) is in nats.

In the large system limit:

- In step \( t + 1 \) all sections with power \( c_{\ell} > 2R_{\tau_t}^2 \) will be decodable, i.e., sent terms have weights very close to 1
- Other sections will not be decodable
- Can use this to understand decoding progression for any power allocation
Asymptotics for $P_\ell \propto e^{-2C\ell/L}$

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

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

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

\[
\xi_t = \min \left\{ \left( \frac{1}{2C} \log \left( \frac{C}{R} \right) + \xi_{t-1} \right), \ 1 \right\}.
\]
Asymptotics for $P_\ell \propto e^{-2C\ell/L}$

\[
\bar{x}_t := \lim x_t = \frac{(1 + \text{snr}) - (1 + \text{snr})^{1-\xi_t}}{\text{snr}},
\]

\[
\bar{\tau}_t^2 := \lim \tau_t^2 = \sigma^2 + P(1 - \bar{x}_t) = \sigma^2 (1 + \text{snr})^{1-\xi_t}
\]

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

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

For $R < C$, $\bar{x}_t \uparrow 1$ and $\bar{\tau}_t^2 \downarrow \sigma^2$ in exactly $T^* = \left\lceil \frac{2C}{\log(C/R)} \right\rceil$ steps.

Run AMP decoder for $T^*$ steps to get $\beta^1, \ldots, \beta^{T^*} \to \hat{\beta}$.
Performance of AMP Decoder

The section error rate of a decoder for SPARC $S$ is

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

**Theorem [Rush, Greig, Venkataramanan ’15]:**
Fix rate $R < C$, and $b > 0$. Consider a sequence of rate $R$ SPARCs $\{S_n\}$ indexed by block length $n$, with design matrix parameters $L$ and $M = L^b$, and power allocation $\propto e^{-2C\ell/L}$. Then the section error rate of the AMP decoder converges to zero almost surely, i.e., for any $\epsilon > 0$,

$$\lim_{n_0 \to \infty} P(\mathcal{E}_{\text{sec}}(S_n) < \epsilon, \forall n \geq n_0) = 1.$$
Theorem [Rush, Venkataramanan ’16]:

For sufficiently large $n, L$, the section error rate of the AMP decoder satisfies

$$P\left( \mathcal{E}_{sec}(S_n) > \epsilon \right) \leq P\left( \frac{\| \beta^{T*} - \beta \|^2}{n} > \epsilon \frac{\sigma^2 \ln(1 + \text{snr})}{4} \right)$$

$$\leq K \exp \left( -\kappa L \epsilon^2 / (\log M)^2 T^* \right)$$
Proof Idea

Steps:
1. Characterize the conditional distribution of statistic and residual in terms of i.i.d. Gaussian plus deviation term:

Show

\[
(A^T r^t + \beta^t - \beta)|_{\text{past}, \beta, \varepsilon} \overset{d}{=} \tau_t Z_t + \Delta_t
\]

\[
(r^t - \varepsilon)|_{\text{past}, \beta, \varepsilon} \overset{d}{=} \sqrt{\tau_t^2 - \sigma^2} Z'_t + \Delta'_t
\]
Proof Idea

Steps:

1. Characterize the conditional distribution of statistic and residual in terms of i.i.d. Gaussian plus deviation term:

\[
\begin{align*}
(A^T r_t + \beta_t - \beta)_{\text{past,}\beta,\epsilon} & \overset{d}{=} \tau_t Z_t + \Delta_t \\
(r_t - \epsilon)_{\text{past,}\beta,\epsilon} & \overset{d}{=} \sqrt{\tau_t^2 - \sigma^2} Z'_t + \Delta'_t
\end{align*}
\]

\[
\Delta_t = \sum_{r=0}^{t-1} (\alpha_r^t - \hat{\alpha}_r^t) h^{r+1} + \left( \frac{m_t^\perp}{\sqrt{n}} - \tau_t^\perp \right) I - \frac{m_t^\perp}{\sqrt{n}} P_{Q_{t+1}}^\perp \right) Z_t \\
+ Q_{t+1} \left( \frac{Q_{t+1}^T Q_{t+1}}{n} \right)^{-1} \left( \frac{B_{t+1}^T m_t^\perp}{n} - \frac{Q_{t+1}^T q_t^\perp}{n} \right) \\
\Delta'_t = \ldots
\]
Proof Idea

Steps:

1. Characterize the conditional distribution of statistic and residual in terms of i.i.d. Gaussian plus deviation term.

2. Inductively obtain concentration results to show the deviation terms are small with high probability.

- Result also shows that $\frac{\|r^t\|^2}{n}$ concentrates around $\tau_t^2$
- So could use these in the decoder instead of precomputing

“Finite Sample analysis of AMP”: talk by Cynthia Rush at 17:30 on Monday
Empirical Performance

SPARC: $\text{snr} = 15$.

Power allocation plays a key role at finite block lengths!
Power Allocation

Give a picture

\[ P_\ell = \begin{cases} \kappa \cdot e^{-2aC\ell/L}, & 0 < \frac{\ell}{L} \leq f \\ \kappa \cdot e^{-2a Cf}, & f < \frac{\ell}{L} \leq 1 \end{cases} \]

Parameters \( a, f \in [0, 1] \):

- Smaller \( a \) makes the exponential decay gentler, allocating less power to initial sections
- \( f \) controls where you start flattening
- Given \( R \) and snr, can optimize over all \( a, f \) that give asymptotic \( \bar{x}_T = 1 \) for some finite \( T \)
Power Allocation

Give a picture

\[
P_\ell = \begin{cases} 
\kappa \cdot e^{-2aC\ell/L}, & 0 < \frac{\ell}{L} \leq f \\
\kappa \cdot e^{-2af}, & f < \frac{\ell}{L} \leq 1 
\end{cases}
\]

Parameters \(a, f \in [0, 1]\):

- Smaller \(a\) makes the exponential decay gentler, allocating less power to initial sections
- \(f\) controls where you start flattening
- Given \(R\) and snr, can optimize over all \(a, f\) that give asymptotic \(\bar{x}_T = 1\) for some finite \(T\)

Can we have a non-parametric algorithm to get optimal power allocation for a given \(R\) and snr?
Algorithmic Power Allocation

\[ \bar{x}(\tau_t^2) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbf{1}\{c_\ell > 2R\tau_t^2\} \text{ where } c_\ell = LP_\ell \]

- Fix target number of decoding steps \( T^* \)
- Asymptotically want to decode \( L/T^* \) sections in each step

With \( \tau_0^2 = \sigma^2 + P \), set \( c_\ell = 2R\tau_0^2 + \delta \) for \( \ell \leq L/T^* \)
Algorithmic Power Allocation

\[ \bar{x}(\tau_t^2) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbb{1}\{c_\ell > 2R\tau_t^2\} \text{ where } c_\ell = LP_\ell \]

- Fix target number of decoding steps \( T^* \)
- Asymptotically want to decode \( \frac{L}{T^*} \) sections in each step

\[ \tau_1^2 = \sigma^2 + P(1 - \bar{x}(\tau_0^2)) \]

Compare \( 2R\tau_1^2 + \delta \) with allocating remaining power equally, choose the greater for the next \( \frac{L}{T^*} \) sections
Algorithmic Power Allocation

\[ \tilde{x}(\tau_t^2) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbf{1}\{c_\ell > 2R\tau_t^2\} \text{ where } c_\ell = LP_\ell \]

- Fix target number of decoding steps \( T^* \)
- Asymptotically want to decode \( L/T^* \) sections in each step

\[ \tau_2^2 = \sigma^2 + P(1 - \tilde{x}(\tau_1^2)) \]

Compare \( 2R\tau_2^2 + \delta \) with allocating remaining power equally, choose the greater for the next \( L/T^* \) sections
Algorithmic Power Allocation

\[ \bar{x}(\tau_t^2) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_\ell}{P} \mathbf{1}\{c_\ell > 2R \tau_t^2\} \text{ where } c_\ell = LP_\ell \]

- Fix target number of decoding steps \( T^* \)
- Asymptotically want to decode \( \frac{L}{T^*} \) sections in each step

\[ \tau_3^2 = \sigma^2 + P(1 - \bar{x}(\tau_3^2)) \]

Compare \( 2R \tau_3^2 + \delta \) with allocating remaining power equally, choose the greater for the next \( \frac{L}{T^*} \) sections.
Algorithmic Power Allocation

\[ \tilde{x}(\tau_t^2) = \lim_{L \to \infty} \sum_{\ell=1}^{L} \frac{P_\ell}{P} 1\{c_\ell > 2R\tau_t^2\} \text{ where } c_\ell = LP_\ell \]

- Fix target number of decoding steps \( T^* \)
- Asymptotically want to decode \( L/T^* \) sections in each step

\[ \tau_3^2 = \sigma^2 + P(1 - \tilde{x}(\tau_3^2)) \]

Compare \( 2R\tau_3^2 + \delta \) with allocating remaining power equally, choose the greater for the next \( L/T^* \) sections.
- Algorithmic power alloc. performs close to (or even beats!) numerically optimized for wide range of snr and $R$ values
- Optimal value of tolerance $\delta$ varies (slightly) with snr
Spatially-coupled design matrices [Barbier, Krzakala ’15]: Another way to improve performance at finite block-lengths

- Band-diagonal structure for $A$; each block consists of random rows of a Hadamard matrix
- First few sections of $\beta$ are oversampled to kick-start the decoding progression
- Good empirical results at finite block lengths
- (Non-rigorous) Replica analysis indicates that this is asymptotically capacity-achieving
Complexity of AMP Decoder

Complexity determined by matrix-vector mults. $A\beta^t$ and $A^Tr^t$

For Gaussian $A$:
- Complexity and memory both $O(nN)$
- Hard to implement for very large $M, L$, e.g., $M = L = 1024$
Complexity of AMP Decoder

Complexity determined by matrix-vector mults. \(A \beta^t\) and \(A^T r^t\)

For Gaussian \(A\):

- Complexity and memory both \(O(nN)\)
- Hard to implement for very large \(M, L\), e.g., \(M = L = 1024\)

For practical implementation, use Hadamard design:

- For \(N = 2^m\), Hadamard matrix \(\mathcal{H}_m \in \mathbb{R}^{N \times N}\):

\[
\mathcal{H}_m = \begin{bmatrix}
\mathcal{H}_{m-1} & \mathcal{H}_{m-1} \\
\mathcal{H}_{m-1} & -\mathcal{H}_{m-1}
\end{bmatrix}, \quad \mathcal{H}_0 = 1.
\]

- Design matrix \(A\): Pick \(n\) random rows of \(\mathcal{H}_m\)
- Multiplications via fast Walsh-Hadamard Transform \(\Rightarrow\)
  Complexity: \(O(N \log N) \sim n^{1+\epsilon}\)
- Don’t need to store \(A\) in memory!
Comparison of SPARC decoders

In both decoders, \( \beta_{j}^{t+1} = \eta_{t,j}(stat_t) \), for \( j = 1, \ldots, ML \)

\textbf{AMP:}

\[
\beta_{j}^{t} = y - A\beta^{t} + \frac{r^{t-1}}{\tau_{t-1}^2} \left( P - \frac{\|\beta^{t}\|^2}{n} \right),
\]

\[
stat_t = \beta^{t} + A^T r_t,
\]

\textbf{Adaptive Soft-Decision Decoding:}

\[
G_t = \text{part of } A\beta^{t} \text{ orthogonal to } A\beta^{t-1}, \ldots, Y
\]

\[
Z_t = \sqrt{n} A^T G_t / \| G_t \|
\]

\[
Z_t^{comb} = \lambda_0 Z_0 + \lambda_1 Z_1 + \ldots + \lambda_t Z_t, \quad \sum \lambda_k^2 = 1
\]

\[
stat_t = \tau_t Z_t^{comb} + \beta^{t}
\]
Comparison of SPARC decoders

In both decoders, \( \beta_j^{t+1} = \eta_{t,j}(\text{stat}_t) \), for \( j = 1, \ldots, ML \)

AMP:

\[
\begin{align*}
\dot{r}^t &= y - A\beta^t + \frac{r^{t-1}}{\tau_{t-1}^2} \left( P - \frac{\|\beta_t\|^2}{n} \right), \\
\text{stat}_t &= \beta^t + A^T \dot{r}^t,
\end{align*}
\]

Adaptive Soft-Decision Decoding:

\[
G_t = \text{part of } A\beta^t \text{ orthogonal to } A\beta^{t-1}, \ldots, Y
\]

\[
\mathcal{Z}_t = \sqrt{n} A^T G_t / \| G_t \|
\]

\[
\mathcal{Z}_{t}^{\text{comb}} = \lambda_0 \mathcal{Z}_0 + \lambda_1 \mathcal{Z}_1 + \ldots + \lambda_t \mathcal{Z}_t, \quad \sum \lambda_k^2 = 1
\]

\[
\text{stat}_t = \tau_t \mathcal{Z}_{t}^{\text{comb}} + \beta^t
\]

- For both, \( \Pr \left( |\frac{1}{n} \beta^T \beta^t - x_t P| > \delta \right) \leq K_t \exp(-\kappa_n \delta^2 / (\log M)^2t) \)
- Constants, analysis techniques are different
Summary + Future Directions

- AMP for low-complexity, parameter-free SPARC decoding
- For any rate $R < C$, the probability of section error rate $> \epsilon$ decays exponentially as $L\epsilon^2 / (\log M)^t$
- With Hadamard-based matrices, can implement the decoder for large block lengths, making SPARCs an attractive alternative to coded modulation

Open Questions:

1. Theoretical Guarantees for the Hadamard-based SPARC

2. Can we combine power allocation and spatial coupling to get good empirical performance close to $C$ at reasonable block lengths?

3. The BIG question: Can we design feasible decoders whose gap to capacity is $O\left(\frac{1}{n^a}\right)$ for some $a \in (0, \frac{1}{2})$?
References

