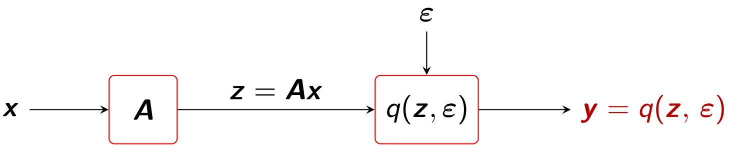
Bayes-optimal Estimation in Generalized Linear Models

Ramji Venkataramanan, University of Cambridge

(Joint work with Pablo Pascual Cobo and Kuan Hsieh)

Information Theory and Tapas Workshop, Madrid

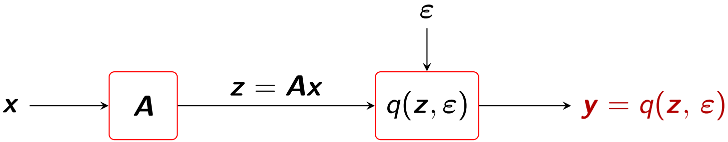
Generalized Linear Models



GOAL:

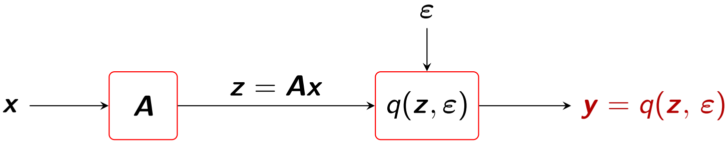
- Estimate signal $\boldsymbol{x} \in \mathbb{R}^n$ from observations $\boldsymbol{y} \equiv (y_1, \dots, y_m)$
- Known sensing matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and output function q

Examples



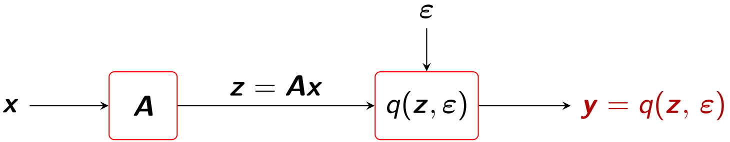
- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$

Examples

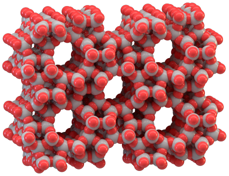


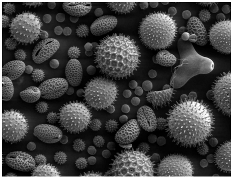
- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$
- 1-bit compressed sensing $\boldsymbol{y} = \text{sign}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon})$

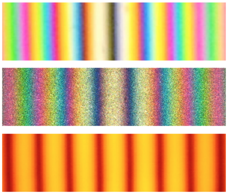
Examples

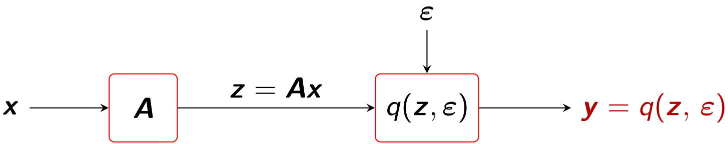


- Linear model $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \varepsilon$
- 1-bit compressed sensing $\boldsymbol{y} = \text{sign}(\boldsymbol{A}\boldsymbol{x} + \varepsilon)$
- Phase retrieval $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 + \varepsilon$


X-ray crystallography


Microscopy


Interferometry

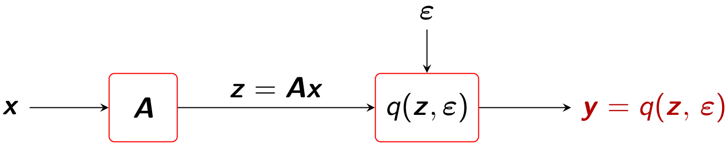


$$\boldsymbol{A} = \begin{bmatrix} \longleftarrow & \boldsymbol{a}_1 & \longrightarrow \\ & \vdots & \\ \longleftarrow & \boldsymbol{a}_m & \longrightarrow \end{bmatrix} \in \mathbb{R}^{m \times n}$$

High-dimensional regime

$$\frac{m}{n} \to \delta \quad \text{as } m, n \to \infty$$

Bayesian setting



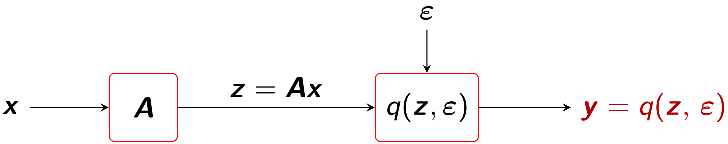
Suppose:

- $\boldsymbol{x} \sim P_X$ and $\varepsilon \sim P_\varepsilon$
- $\boldsymbol{A}$ also generated from known distribution

Bayes-optimal estimator that minimizes MSE: $\mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}$

$$\mathrm{MMSE}_n := \frac{1}{n}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}\|^2\}.$$

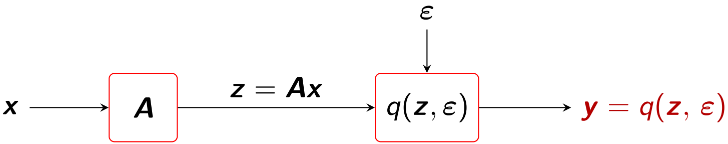
Two natural questions



$$\mathrm{MMSE}_n := \frac{1}{n}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A}, \boldsymbol{y}\}\|^2\}.$$

1. What is $\lim_{n\to\infty} \mathrm{MMSE}_n$? (for a fixed $\delta = \lim \frac{m}{n}$)

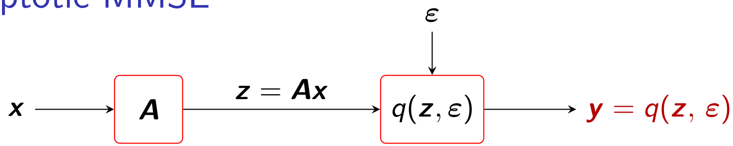
Two natural questions



$$\mathrm{MMSE}_n := \frac{1}{n}\mathbb{E}\{\|\boldsymbol{x} - \mathbb{E}\{\boldsymbol{x} \mid \boldsymbol{A},\, \boldsymbol{y}\}\|^2\}.$$

1. What is $\lim_{n\to\infty} \mathrm{MMSE}_n$? (for a fixed $\delta = \lim \frac{m}{n}$)
2. How can we design **efficient** estimators whose error approaches $\lim \mathrm{MMSE}_n$?

Asymptotic MMSE



- For iid Gaussian $\boldsymbol{A}$ with $A_{ij} \sim \mathrm{N}(0, \frac{1}{n})$
- Signal $\boldsymbol{x}$ iid $\sim P_X$ and noise $\boldsymbol{\varepsilon}$ iid $\sim P_\varepsilon$

[Barbier et al. '19]: Formula for asymptotic MMSE in terms of a scalar **potential function** $U(x;\, \delta)$

$$\lim_{n\to\infty} \mathrm{MMSE}_n = \underset{x\in[0,\mathrm{Var}(X)]}{\arg\min}\ U(x;\, \delta)$$

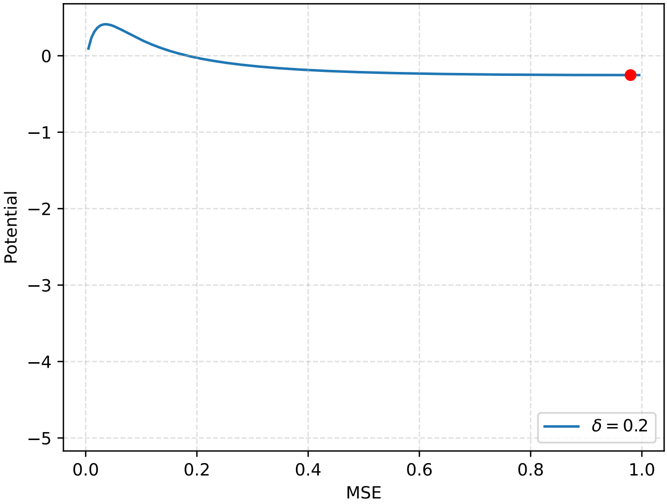
$$\lim_{n\to\infty} \frac{1}{n} I(\boldsymbol{X}\,;\ \boldsymbol{Y}) = \min_{x\in[0,\mathrm{Var}(X)]}\ a\, U(x;\, \delta) + b$$

Barbier et al. , *Optimal errors and phase transitions in high-dimensional GLMs*, PNAS, 2019

Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \text{Prior } P_X(-a) = 0.4,\ P_X(a) = 0.6$$

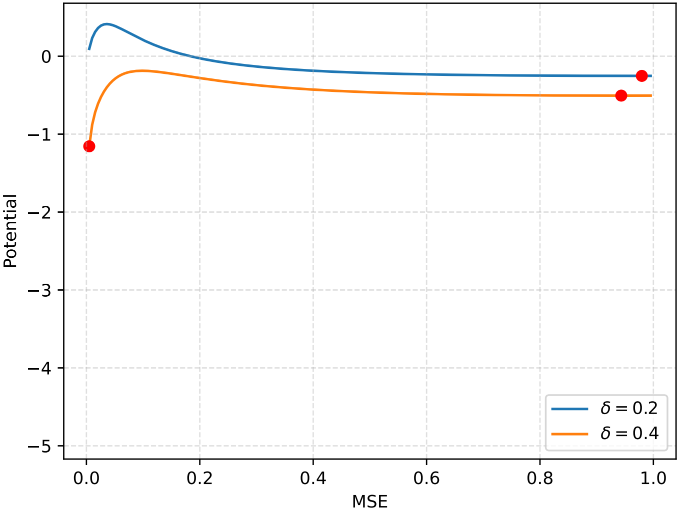
$U(x\,;\delta)$ vs x



Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior } P_X(-a) = 0.4,\ P_X(a) = 0.6$$

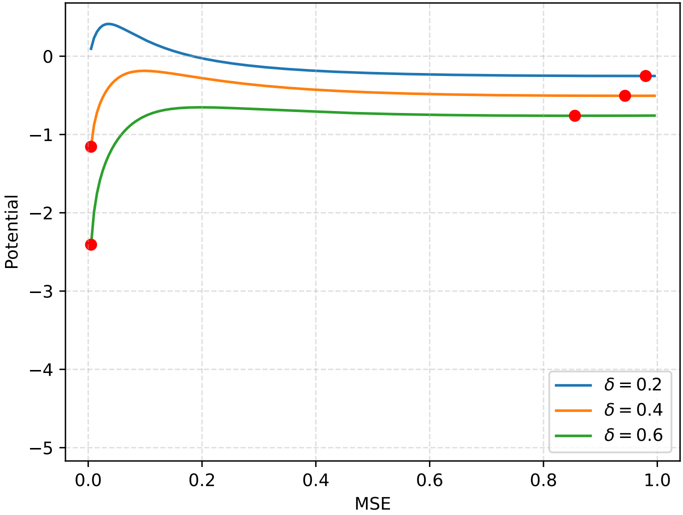
$U(x\,;\delta)$ vs x



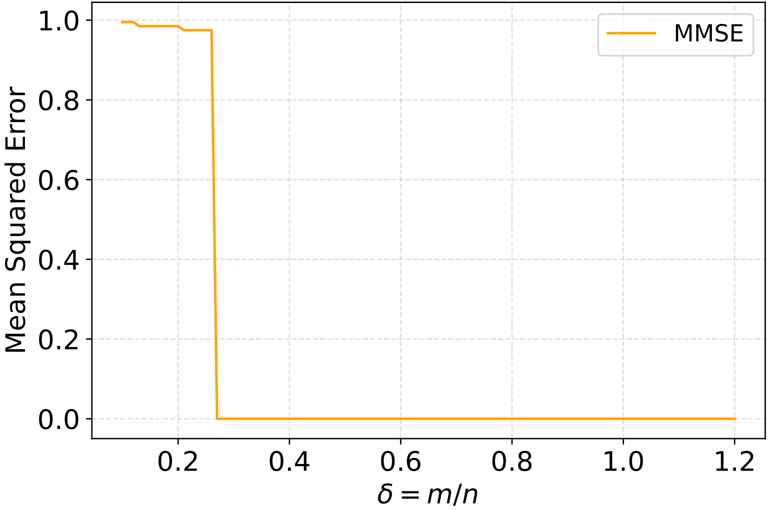
Example: Phase Retrieval

$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2$ Prior $P_X(-a) = 0.4,\ P_X(a) = 0.6$

$U(x\,;\delta)$ vs x



MMSE: Phase retrieval



Can we achieve this with efficient estimators?

Estimators

- Convex relaxations
- Iterative algorithms for non-convex objectives:
  Alternating minimization, gradient descent, . . .
- Spectral methods

Phase retrieval: [Netrapalli et al. '13], [Candes et al. '13], [Luo et al. '19], [Mondelli & Montanari '19], . . .
1-bit CS: [Plan & Vershynin '13], [Jacques et al. '13], . . .

Estimators

- Convex relaxations
- Iterative algorithms for non-convex objectives:
  Alternating minimization, gradient descent, ...
- Spectral methods

Generic techniques: can incorporate certain constraints like sparsity

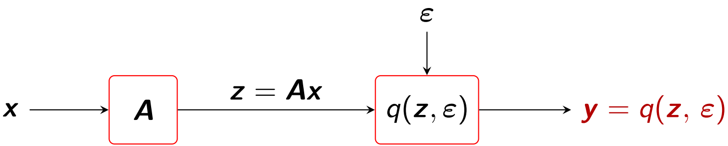
But not well-equipped to exploit specific structural info about signal, e.g., known prior

---

Phase retrieval: [Netrapalli et al. '13], [Candes et al. '13], [Luo et al. '19], [Mondelli & Montanari '19], ...

1-bit CS: [Plan & Vershynin '13], [Jacques et al. '13], ...

Approximate Message Passing



Estimator based on **AMP**

- Can be tailored to take advantage of prior info about signal
- Rigorous performance characterization via **state evolution**
  Allows us to precisely compute asymptotic MSE

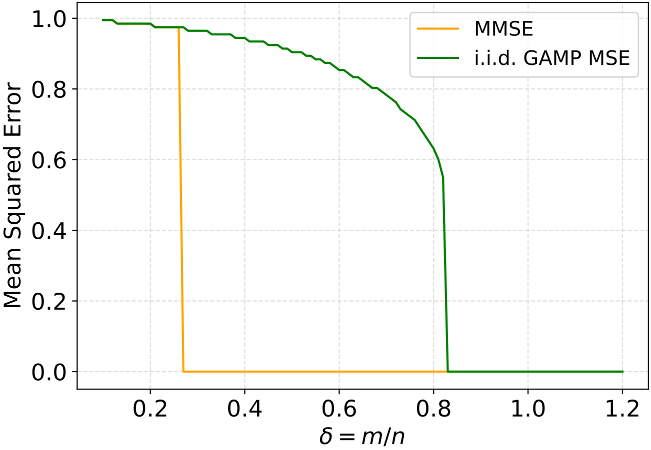
GAMP [Rangan '11]: for GLMs with i.i.d. Gaussian $\boldsymbol{A}$
– Conjectured to be optimal among poly-time estimators

[Javanmard & Montanari '13], [Schniter & Rangan '14], ...

AMP vs MMSE estimator

Phase retrieval with i.i.d. Gaussian $\boldsymbol{A}$

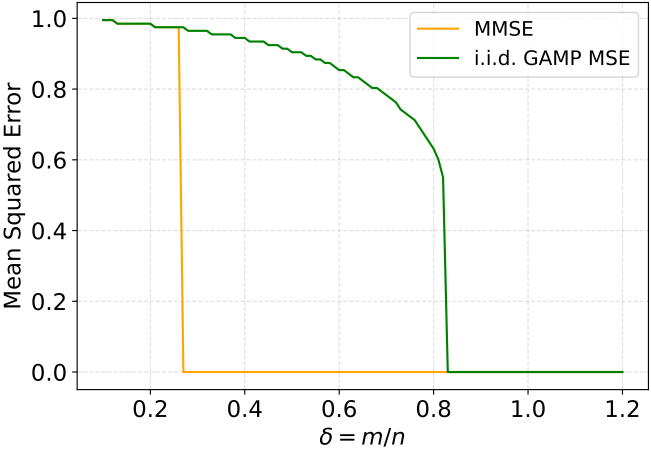
$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \text{Prior} \; : P_X(-a) = 0.4, \; P_X(a) = 0.6$$



AMP vs MMSE estimator

Phase retrieval with i.i.d. Gaussian $\boldsymbol{A}$

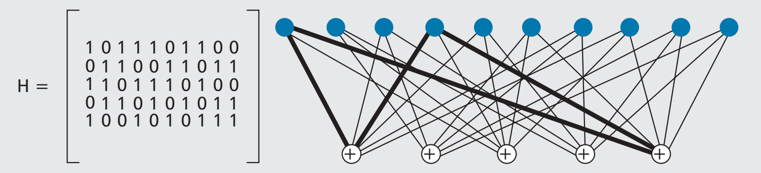
$$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior} \; : P_X(-a) = 0.4, \; P_X(a) = 0.6$$



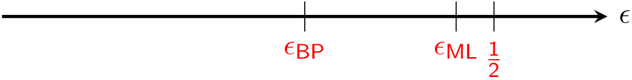
This talk: How to close this gap?

Parallel with coding theory

Consider a rate $R = \frac{1}{2}$ **regular** LDPC code. E.g.,



Used over channel with erasure probability ϵ



ϵ_{BP}: Threshold with belief propagation decoding

ϵ_{ML}: Threshold with optimal (ML) decoding

Figure from Costello et al. *Spatially coupled sparse codes on graphs: theory and practice*, 2014

Closing the gap: Can make ϵ_{BP} approach ϵ_{ML} with **spatially coupled code** [Kudekar et al. '14]

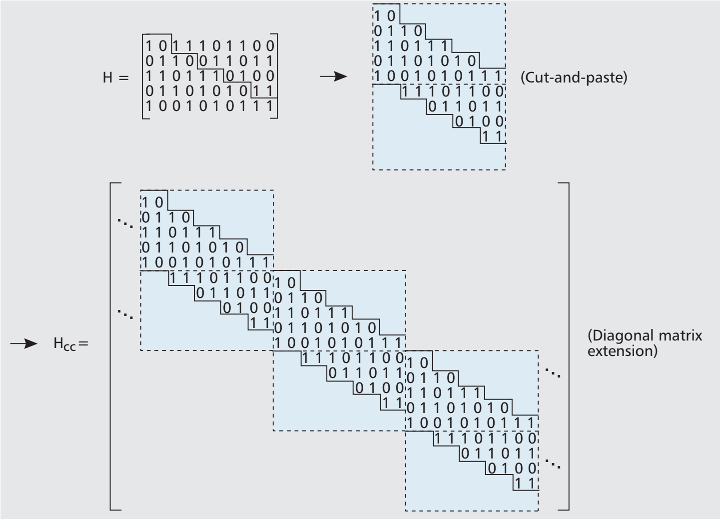


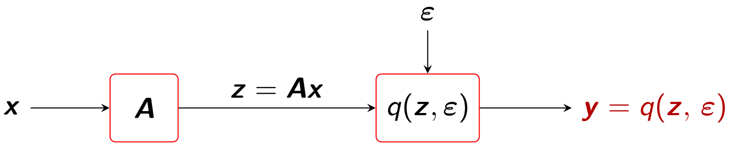
Figure from Costello et al. *Spatially coupled sparse codes on graphs: theory and practice*, 2014

LDPC codes	**GLM**
Rate R	Sampling ratio δ
Regular parity check matrix	iid Gaussian sensing matrix
BP decoder	AMP estimator
Density evolution	State evolution

LDPC codes	**GLM**
Rate R	Sampling ratio δ
Regular parity check matrix	iid Gaussian sensing matrix
BP decoder	AMP estimator
Density evolution	State evolution
ϵ_{BP}, ϵ_{ML}	δ_{AMP}, δ_{MMSE}
Spatially coupled code	Spatially coupled sensing matrix

Compressed sensing: [Kudekar, Pfister '10], [Donoho, Javanmard, Montanari '13] . . .

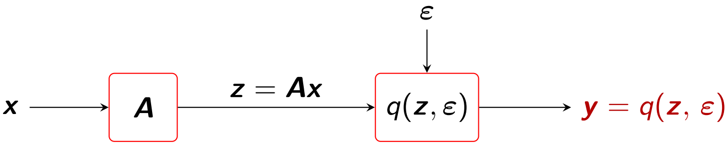
i.i.d. Gaussian GAMP



Iteratively produces estimates $\boldsymbol{x}(t)$ and $\boldsymbol{z}(t)$ for $t \geq 0$ via:

$$g_{\text{in}}(\cdot\,;\,t) : \mathbb{R} \to \mathbb{R}, \qquad g_{\text{out}}(\cdot\,, y\,;\,t) : \mathbb{R}^2 \to \mathbb{R}$$

i.i.d. Gaussian GAMP



Iteratively produces estimates $\boldsymbol{x}(t)$ and $\boldsymbol{z}(t)$ for $t \geq 0$ via:

$$g_{\text{in}}(\cdot\,;\, t) : \mathbb{R} \to \mathbb{R}, \qquad g_{\text{out}}(\cdot\,, y\,;\, t) : \mathbb{R}^2 \to \mathbb{R}$$

$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\, t) + \alpha^{x}(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\, t+1) - \alpha^{z}(t+1) g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

[Feng et al. '22]: *A unifying tutorial on Approximate Message Passing*

i.i.d. Gaussian GAMP

$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\,t) + \alpha^{x}(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t),\boldsymbol{y}\,;\,t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\,t+1) - \alpha^{z}(t+1) g_{\text{out}}(\boldsymbol{z}(t),\boldsymbol{y}\,;\,t)$$

- g_{in} and g_{out} applied row-wise
- $g_{\text{in}}, g_{\text{out}}$ Lipschitz, allow us to tailor the algorithm

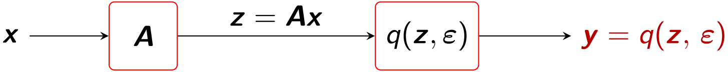
i.i.d. Gaussian GAMP

$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\, t) + \alpha^x(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\, t+1) - \alpha^z(t+1) g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

- g_{in} and g_{out} applied row-wise
- $g_{\text{in}}, g_{\text{out}}$ Lipschitz, allow us to tailor the algorithm
- Initialized with $\boldsymbol{x}^0$ and $\boldsymbol{z}(0) = \boldsymbol{A}\boldsymbol{x}^0$
- Coefficients $\alpha^x(t)$ and $\alpha^z(t)$ defined in terms of $g_{\text{in}}{}'$ and $g_{\text{out}}{}'$

Asymptotics of i.i.d Gaussian GAMP



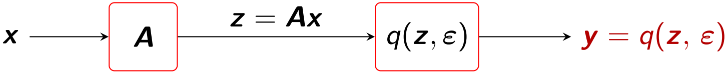
$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\, t) + \alpha^{x}(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\, t+1) - \alpha^{z}(t+1) g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

Suppose empirical distribution of $\boldsymbol{x}$ converges to law of $X \sim P_X$.
Then as $n \to \infty$:

[Rangan '11], [Javanmard, Montanari ' 13]

Asymptotics of i.i.d Gaussian GAMP



$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\,t) + \alpha^{x}(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\,t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\,t+1) - \alpha^{z}(t+1) g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\,t)$$

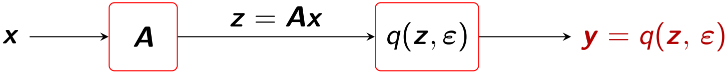
Suppose empirical distribution of $\boldsymbol{x}$ converges to law of $X \sim P_X$.
Then as $n \to \infty$:

The empirical distribution of $(\boldsymbol{x}, \boldsymbol{x}(t))$ converges to the law of

$$[X,\ \mu(t)\,X + W(t)], \quad \text{where} \quad W(t) \sim \mathrm{N}(0, \tau^{x}(t))$$

[Rangan '11], [Javanmard, Montanari ' 13]

Asymptotics of i.i.d Gaussian GAMP



$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t)\,;\, t) + \alpha^{x}(t+1)\boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1)\,;\, t+1) - \alpha^{z}(t+1) g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}\,;\, t)$$

Suppose empirical distribution of $\boldsymbol{x}$ converges to law of $X \sim P_X$.
Then as $n \to \infty$:

The empirical distribution of $(\boldsymbol{z}, \boldsymbol{z}(t))$ converges to the law of

$$[Z,\ Z(t)] \sim \mathrm{N}(0, \Lambda(t))$$

[Rangan '11], [Javanmard, Montanari ' 13]

State Evolution

The empirical distribution of $(\boldsymbol{x}, \boldsymbol{x}(t))$ converges to the law of

$$[X,\ \mu(t)\,X + W(t)], \quad \text{where} \quad W(t) \sim \mathrm{N}(0, \tau^x(t))$$

The empirical distribution of $(\boldsymbol{z}, \boldsymbol{z}(t))$ converges to the law of

$$[Z,\ Z(t)] \sim \mathrm{N}(0, \Lambda(t))$$

$\mu(t)$, $\tau^x(t)$, $\Lambda(t)$ computed via **state evolution** recursion:

$$[\mu(t),\, \tau^x(t),\, \Lambda(t)] \longrightarrow [\mu(t+1),\, \tau^x(t+1),\, \Lambda(t+1)]$$

State Evolution

The empirical distribution of $(\boldsymbol{x}, \boldsymbol{x}(t))$ converges to the law of

$$[X,\ \mu(t)\,X + W(t)], \quad \text{where} \quad W(t) \sim \mathrm{N}(0, \tau^x(t))$$

The empirical distribution of $(\boldsymbol{z}, \boldsymbol{z}(t))$ converges to the law of

$$[Z,\ Z(t)] \sim \mathrm{N}(0, \Lambda(t))$$

$\mu(t)$, $\tau^x(t)$, $\Lambda(t)$ computed via **state evolution** recursion:

$$[\mu(t),\ \tau^x(t),\ \Lambda(t)] \longrightarrow [\mu(t+1),\ \tau^x(t+1),\ \Lambda(t+1)]$$

- State evolution depends on g_{in} and g_{out}
- Analogous to density evolution for LDPC codes

Bayes GAMP

Asymptotic MSE: For $t \geq 1$,

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}(\boldsymbol{x}(t))\|^2 = \mathbb{E}\{[X - g_{\text{in}}(\mu(t)\, X + W(t))]^2\}$$

[Rangan '11], [Javanmard, Montanari ' 13]

Bayes GAMP

Asymptotic MSE: For $t \geq 1$,

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}(\boldsymbol{x}(t))\|^2 = \mathbb{E}\{[X - g_{\text{in}}(\mu(t)\, X + W(t))]^2\}$$

- Bayes-optimal choice of g_{in}:

$$g^*_{\text{in}}(X(t)) = \mathbb{E}[X \mid \mu(t)\, X + W(t) = X(t)]$$

  $g^*_{\text{in}}(\boldsymbol{x}(t))$ is the MMSE estimate of $\boldsymbol{x}$ given $\boldsymbol{x}(t)$

- Can also determine Bayes-optimal g^*_{out}

[Rangan '11], [Javanmard, Montanari ' 13]

Fixed points of Bayes GAMP

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x}-g_{\text{in}}^{*}(\boldsymbol{x}(t))\|^2 = \mathbb{E}\{ [X-g_{\text{in}}^{*}(X+W(t))]^2\}, \quad W(t) \sim \mathrm{N}(0, \tau^{x}(t)$$

Run to "convergence" $\Rightarrow$ MSE determined by $\lim_{t\to\infty} \tau^{x}(t)$

Fixed points of Bayes GAMP

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}^*(\boldsymbol{x}(t))\|^2 = \mathbb{E}\{ [X - g_{\text{in}}^*(X + W(t))]^2\}, \quad W(t) \sim \mathrm{N}(0, \tau^x(t)$$

Run to "convergence" $\Rightarrow$ MSE determined by $\lim_{t\to\infty} \tau^x(t)$

State evolution

Given $\tau^x(t)$, compute:

$$\tau^z(t) = \frac{1}{\delta}\text{mmse}\big(\tau^x(t)\big)$$

$$\tau^x(t+1) = \tau^z(t)\Big[1 - \frac{1}{\tau^x(t)}\mathbb{E}\{\text{Var}(Z \mid Z(t),\ Y)\}\Big]^{-1}$$

Fixed points of Bayes GAMP

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x}-g_{\text{in}}^*(\boldsymbol{x}(t))\|^2 = \mathbb{E}\{ [X-g_{\text{in}}^*(X+W(t))]^2\}, \quad W(t) \sim \mathrm{N}(0, \tau^x(t)$$

Run to "convergence" $\Rightarrow$ MSE determined by $\lim_{t\to\infty} \tau^x(t)$

State evolution

Given $\tau^x(t)$, compute:

$$\tau^z(t) = \frac{1}{\delta}\text{mmse}\big(\tau^x(t)\big)$$

$$\tau^x(t+1) = \tau^z(t)\left[1 - \frac{1}{\tau^x(t)}\mathbb{E}\{\text{Var}(Z \mid Z(t),\ Y)\}\right]^{-1}$$

Can determine $\lim_{t\to\infty} \tau^x(t)$ via potential function $U(x;\delta)$

Fixed points of Bayes GAMP

$$\tau^z(t) = \frac{1}{\delta}\text{mmse}\big(\tau^x(t)\big)$$
$$\tau^x(t+1) = \tau^z(t)\left[1 - \frac{1}{\tau^x(t)}\mathbb{E}\{\text{Var}(Z \mid Z(t),\ Y)\}\right]^{-1}$$

Proposition

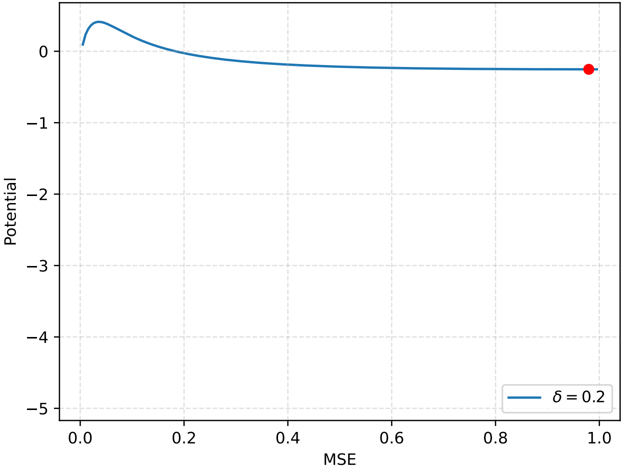
$$\lim_{t\to\infty}\lim_{n\to\infty}\frac{1}{n}\|\boldsymbol{x} - \bar{g}^*_{\text{in}}(\boldsymbol{x}(t);\ t)\|^2$$
$$= \max\left\{x \in [0, \text{Var}(X)]\ :\ \frac{\partial U(x;\delta)}{\partial x} = 0\right\}.$$

MSE of Bayes GAMP given by **largest** stationary point of $U(x;\delta)$

Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior } P_X(-a) = 0.4,\ P_X(a) = 0.6$$

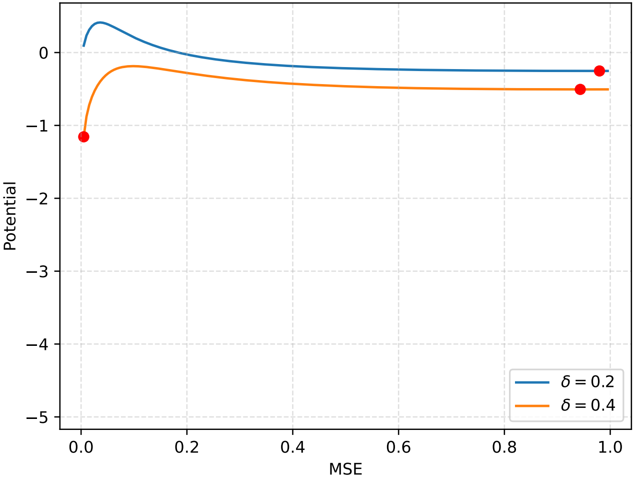
$U(x\,;\delta)$ vs x



Example: Phase Retrieval

$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior } P_X(-a) = 0.4, \;\; P_X(a) = 0.6$

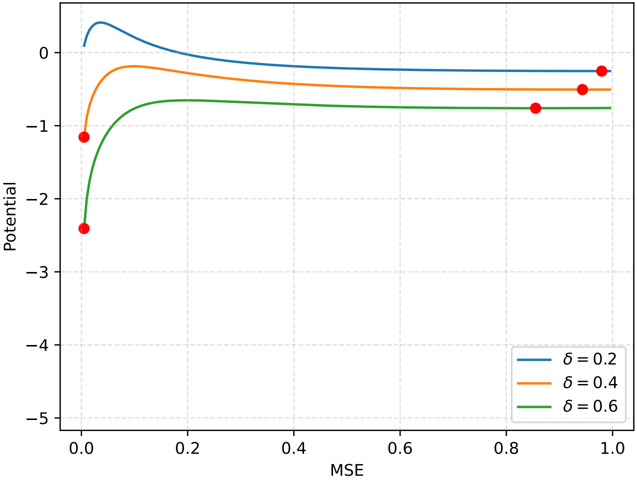
$U(x\,;\delta)$ vs x



Example: Phase Retrieval

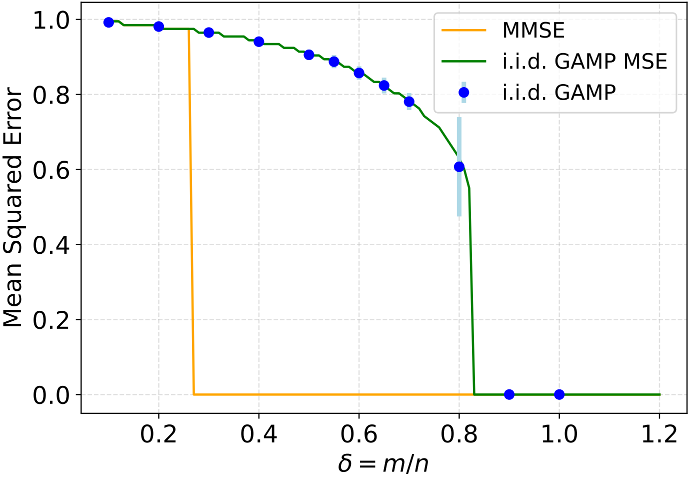
$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \text{Prior } P_X(-a) = 0.4, \; P_X(a) = 0.6$$

$U(x\,;\delta)$ vs x



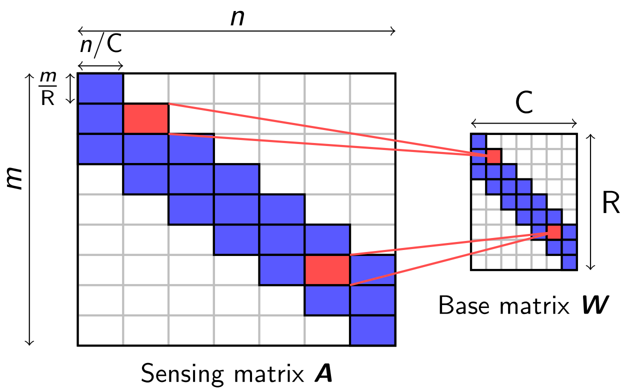
Example: Phase Retrieval

$\boldsymbol{y} = |\boldsymbol{Ax}|^2$ Prior $P_X(-a) = 0.4,\ P_X(a) = 0.6$



Can we get the MSE of GAMP to approach global minimum?

Spatially coupled sensing matrix

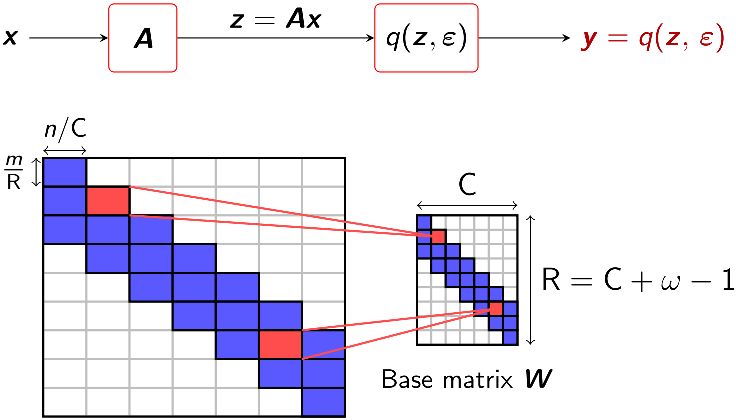


$$A_{jk} \sim \mathrm{N}(0, W_{\mathrm{rc}}) \quad \text{for } j \in \text{block r and } k \in \text{block c}$$

W_{rc} chosen so that each column of $\boldsymbol{A}$ has $\mathbb{E}[\text{squared-norm}] = 1$

[Donoho, Javanmard, Montanari '13] [Barbier and Krzakala '17] [Liang, Ma and Ping '17] [Hsieh, Rush, V '21] . . .

High-level idea

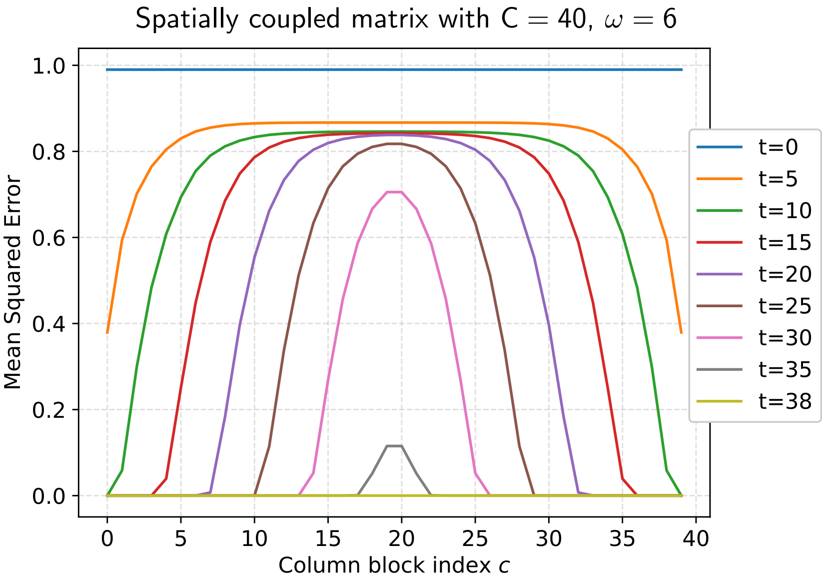


Each little block an iid sensing matrix that multiplies a section of $\boldsymbol{x}$

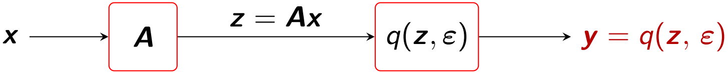
First and last sections have observations with less interference $\Rightarrow$

Can be recovered more easily $\Rightarrow$ helps recover adjacent sections

Decoding wave



Spatially coupled GAMP

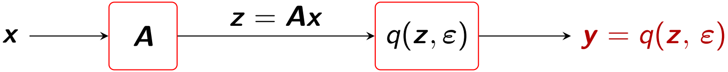


$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t), \mathrm{c}\,;\ t) + \boldsymbol{\alpha}^{x}(t+1) \odot \boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}, \mathrm{r}\,;\ t)$$

$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1), \mathrm{c}\,;\ t+1) - \boldsymbol{\alpha}^{z}(t+1) \odot g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}, \mathrm{r}\,;\ t)$$

- g_{in} and g_{out} now depend on the column and row section

Spatially coupled GAMP

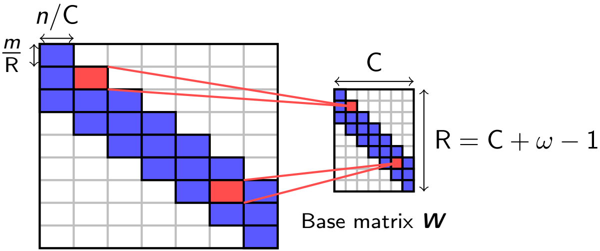


$$\boldsymbol{x}(t+1) = g_{\text{in}}(\boldsymbol{x}(t), \mathsf{c}\,;\; t) + \boldsymbol{\alpha}^{x}(t+1) \odot \boldsymbol{A}^{\mathsf{T}} g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}, \mathsf{r}\,;\; t)$$

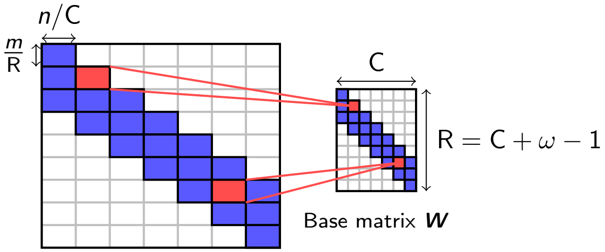
$$\boldsymbol{z}(t+1) = \boldsymbol{A} g_{\text{in}}(\boldsymbol{x}(t+1), \mathsf{c}\,;\; t+1) - \boldsymbol{\alpha}^{z}(t+1) \odot g_{\text{out}}(\boldsymbol{z}(t), \boldsymbol{y}, \mathsf{r}\,;\; t)$$

- g_{in} and g_{out} now depend on the column and row section
- $\boldsymbol{\alpha}^{x}(t+1) = [\ \alpha_1^{x}(t+1), \ldots, \alpha_{\mathsf{C}}^{x}(t+1)\ ]$
- $\boldsymbol{\alpha}^{z}(t+1) = [\ \alpha_1^{z}(t+1), \ldots, \alpha_{\mathsf{R}}^{z}(t+1)\ ]$

Asymptotics of SC-GAMP



Asymptotics of SC-GAMP



The empirical distribution of $(\boldsymbol{x}_{\mathrm{c}}, \boldsymbol{x}_{\mathrm{c}}(t))$ converges to the law of

$$[X,\ X + W_{\mathrm{c}}(t)], \quad \text{where} \quad W(t) \sim \mathrm{N}(0, \tau_{\mathrm{c}}^{x}(t))$$

for $\mathrm{c} = 1, \ldots, \mathrm{C}$

The empirical distribution of $(\boldsymbol{z}_{\mathrm{r}}, \boldsymbol{z}_{\mathrm{r}}(t))$ converges to the law of

$$[Z_{\mathrm{r}},\ Z_{\mathrm{r}}(t)] \sim \mathrm{N}(0, \Lambda_r(t))$$

for $\mathrm{r} = 1, \ldots, \mathrm{R}$

SC-GAMP Performance

State evolution has $\mathsf{C} + \mathsf{R}$ parameters:

$$\{\tau_1^x(t), \ldots, \tau_{\mathsf{C}}^x(t),\, \Lambda_1(t), \ldots, \Lambda_{\mathsf{R}}(t)\} \;\longrightarrow$$
$$\{\tau_1^x(t+1), \ldots, \tau_{\mathsf{C}}^x(t+1),\, \Lambda_1(t+1), \ldots, \Lambda_{\mathsf{R}}(t+1)\}$$

SC-GAMP Performance

State evolution has $\mathsf{C}+\mathsf{R}$ parameters:

$$\{\tau_1^x(t), \ldots, \tau_{\mathsf{C}}^x(t), \Lambda_1(t), \ldots, \Lambda_{\mathsf{R}}(t)\} \longrightarrow$$
$$\{\tau_1^x(t+1), \ldots, \tau_{\mathsf{C}}^x(t+1), \Lambda_1(t+1), \ldots, \Lambda_{\mathsf{R}}(t+1)\}$$

Theorem (Asymptotic MSE): For $t \geq 1$

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}^*(\boldsymbol{x}(t))\|^2 = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} \mathbb{E}\{\, [X - g_{\text{in}}^*(X + W_{\mathsf{c}}(t), \mathsf{c})]^2 \,\}$$

where $W_{\mathsf{c}}(t) \sim \mathsf{N}(0, \tau_{\mathsf{c}}^x(t))$

Fixed points of Bayes SC-GAMP

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g^*_{\text{in}}(\boldsymbol{x}(t))\|^2 = \frac{1}{\mathsf{C}} \sum_{\mathsf{c}=1}^{\mathsf{C}} \mathbb{E}\{ [X - g^*_{\text{in}}(X + W_{\mathsf{c}}(t), \mathsf{c})]^2 \}$$

where $W_{\mathsf{c}}(t) \sim \mathrm{N}(0, \tau^x_{\mathsf{c}}(t))$

Run SC-GAMP to convergence $\Rightarrow$ MSE determined by $\lim_{t\to\infty} \{\tau^x_1(t), \dots, \tau^x_{\mathsf{C}}(t)\}$

How to determine fixed points of this **coupled** recursion?

Fixed points of Bayes SC-GAMP

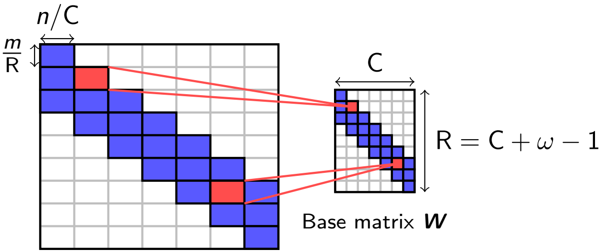
$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g^*_{\text{in}}(\boldsymbol{x}(t))\|^2 = \frac{1}{\mathsf{C}}\sum_{\mathsf{c}=1}^{\mathsf{C}} \mathbb{E}\{\,[X - g^*_{\text{in}}(X + W_{\mathsf{c}}(t), \mathsf{c})]^2\,\}$$

where $W_{\mathsf{c}}(t) \sim \mathsf{N}(0, \tau^x_{\mathsf{c}}(t))$

Run SC-GAMP to convergence $\Rightarrow$ MSE determined by $\lim_{t\to\infty}\{\tau^x_1(t), \ldots, \tau^x_{\mathsf{C}}(t)\}$

How to determine fixed points of this **coupled** recursion?

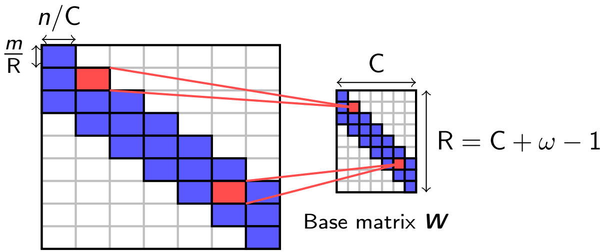
[Yedla et al. '14]: *A simple proof of Maxwell saturation for coupled scalar recursions*



Theorem (Fixed point of SC-GAMP): Fix $\gamma > 0$. Then for $\omega > \omega_0$ and $t > t_0$:

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}^{*}(\boldsymbol{x}(t);\, t)\|^2 \leq \left(\operatorname*{arg\,min}_{x\in[0,\text{Var}(X)]} U(x;\delta_{\text{in}}) + \gamma \right) \frac{\mathsf{C}+\omega}{\mathsf{C}}.$$

Here $\delta_{\text{in}} = \delta\frac{\mathsf{C}}{\mathsf{R}}$ is the inner sampling ratio.



Corollary (Bayes optimality of SC-GAMP): Fix $\epsilon > 0$. Then for $\omega > \omega_0$, sufficiently large C and $t > t_0$ we have:

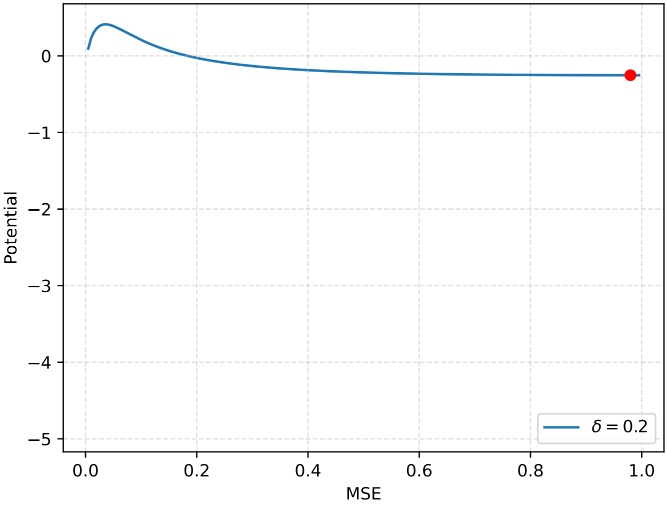
$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x} - g_{\text{in}}^*(\boldsymbol{x}(t);\, t)\|^2 \leq \operatorname*{arg\,min}_{x\in[0,\text{Var}(X)]} U(x;\delta) + \epsilon.$$

Analogous to threshold saturation in SC-LDPC codes

Example: Phase Retrieval

$$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior } P_X(-a) = 0.4,\ \ P_X(a) = 0.6$$

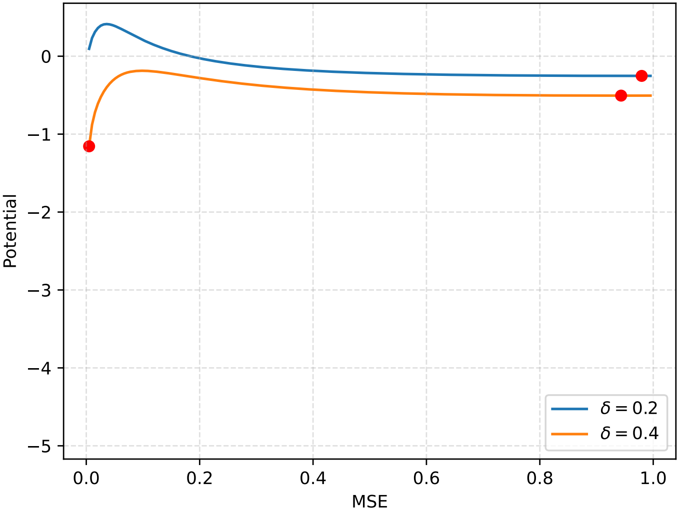
$U(x\,;\delta)$ vs x



Example: Phase Retrieval

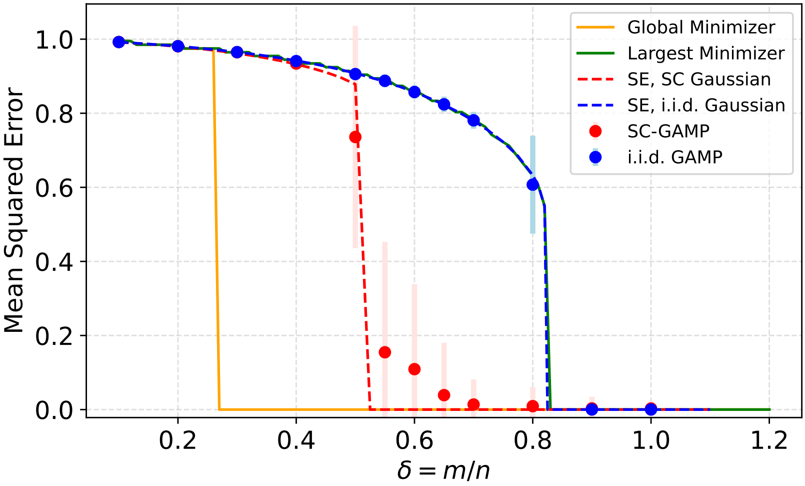
$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|^2 \qquad \text{Prior } P_X(-a) = 0.4,\ \ P_X(a) = 0.6$$

$U(x\,;\delta)$ vs x



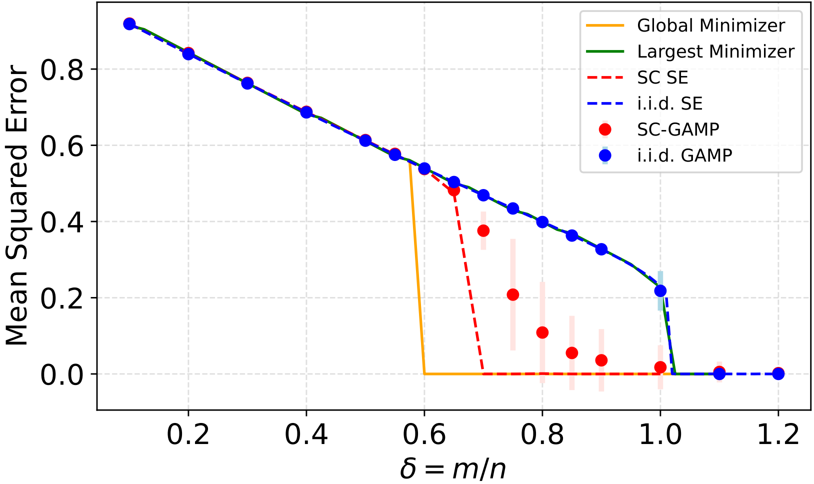
Phase retrieval

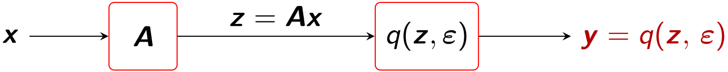
$$\boldsymbol{y} = |\boldsymbol{Ax}|^2 \qquad \text{Prior } P_X(-a) = 0.4, \ P_X(a) = 0.6$$



ReLU model

$$\boldsymbol{y} = \max\left(\boldsymbol{A}\boldsymbol{x},\, 0\right) \qquad \text{Prior } P_X(-b) = P_X(b) = 0.25,\ P_X(0) = 0.5$$





Performance of optimal estimator with iid Gaussian design achieved by *spatially coupled design with message passing estimator*

Future directions

Spatial coupling with *structured* random matrices

– E.g., Fourier, DCT, Hadamard based matrices

– Enables faster AMP-like algorithms