Understanding Phase Transitions via Mutual Information and MMSE

Galen Reeves

Duke University, Durham NC, USA

ISIT Tutorial — July 7, 2019
Table of Contents

Bayes Optimal Inference

Phase Transitions in Linear Regression

Information-Theoretic Proof of RS Formula

Further Directions
Bayes Optimal Inference

Distribution $p(x, y)$ on random vectors (matrices, tensors)

unknown: $X = (X_1, \ldots, X_n)$

observed data: $Y = (Y_1, \ldots, Y_m)$

Posterior distribution of unknowns:

$$p(x | y) = \frac{p(y | x)p(x)}{p(y)}.$$

Basic questions:

1. What are the fundamental limits of inference?
2. Can we achieve these limits using efficient methods?
Marginals and Moments of Posterior Distribution

- Marginal of (random) posterior on set of indices $S \subset [n]$,

$$p(x_S \mid Y) = \int p(x \mid Y) \, dx_{S^c}$$

- Mean and covariance of (random) posterior distribution

$$\mathbb{E}[X \mid Y] = \int x \, p(x \mid Y) \, dx$$

$$\text{Cov}(X \mid Y) = \mathbb{E}[(X - \mathbb{E}[X \mid Y])(X - \mathbb{E}[X \mid Y])^T \mid Y]$$
Marginals and Moments of Posterior Distribution

- Marginal of (random) posterior on set of indices $S \subset [n],$
  \[
  p(x_S | Y) = \int p(x | Y) \, dx_{Sc}
  \]

- Mean and covariance of (random) posterior distribution
  \[
  \mathbb{E}[X | Y] = \int x \, p(x | Y) \, dx \\
  \text{Cov}(X | Y) = \mathbb{E}[(X - \mathbb{E}[X | Y])(X - \mathbb{E}[X | Y])^T | Y]
  \]

- Minimum mean-squared error (MMSE)
  \[
  \text{mmse}(X | Y) = \mathbb{E} \left[ \|X - \mathbb{E}[X | Y]\|^2 \right] = \mathbb{E}[\text{tr}(\text{Cov}(X | Y))]
  \]
  Also known as Bayes risk under squared-error loss
Approaches to Analysis

*Information Theory*
- Single-letter formulas describe fundamental limits
- Models with *structured randomness*: e.g., Markov sources (compression) and finite-state channels (communication)

*Replica Method from Statistical Physics*
- Conjectured single-letter formulas via *heuristics*
- Models with *disordered randomness*: e.g., spin glasses and modern high-dimensional inference problems

*Approximate Inference*
- Fixed-points of message passing algorithms related stationary points of potential functions.
Phase Diagram for High-Dimensional Inference

Easy – Computationally efficient methods succeed.

Hard – All known efficient methods fail but intractable methods succeed.

Impossible – All methods fail, regardless of complexity.
Example: Multiuser Detection in CDMA

\[ Y = \sum_j A_j X_j + W, \quad W \sim \mathcal{N}(0, \sigma^2 I) \]

- \( \{A_j\} \) are known signature vectors with IID entries
- \( \{X_j\} \) are IID \( \{+1, -1\} \) messages

---

\[ \frac{\mathbb{E}[\|X - \hat{X}\|^2]}{\mathbb{E}[\|X\|^2]} \]

Linear MMSE

---

Tse, Hanly’99, Verdu, Shamai’99,
Example: Multiuser Detection in CDMA

\[ Y = \sum_j A_j X_j + W, \quad W \sim \mathcal{N}(0, \sigma^2 I) \]

- \{A_j\} are known signature vectors with IID entries
- \{X_j\} are IID \{+1, -1\} messages

\begin{align*}
\mathbb{E}[\|X - \hat{X}\|^2] & \quad \mathbb{E}[\|X\|^2] \\
\frac{\mathbb{E}[\|X - \hat{X}\|^2]}{\mathbb{E}[\|X\|^2]} & \quad \text{MMSE (replica prediction)} \\
0 & \quad \text{linear MMSE}
\end{align*}

Tse, Hanly’99, Verdu, Shamai’99, Tanaka’02, Guo, Verdu’05,
Example: Multiuser Detection in CDMA

\[ Y = \sum_j A_j X_j + W, \quad W \sim \mathcal{N}(0, \sigma^2 I) \]

- \( \{A_j\} \) are known signature vectors with IID entries
- \( \{X_j\} \) are IID \( \{+1, -1\} \) messages

Tse, Hanly’99, Verdu, Shamai’99, Tanaka’02, Guo, Verdu’05, Bayati, Montanari’11
Example: Compressed Sensing

\[ Y = \sum_j A_j X_j + W, \quad W \sim \mathcal{N}(0, \sigma^2 I) \]

- \( \{ A_j \} \) are columns of known sensing matrix with IID entries
- \( \{ X_j \} \) are IID Bernoulli-Gaussian: \( p(x) = (1 - \epsilon)\delta_0(x) + \epsilon\mathcal{N}(x | 0, \epsilon^{-1}) \)

---

Donoho, Maleki, Montanari’09, Bayati, Montanari’11, R, Gastpar’12
Example: Compressed Sensing

\[ Y = \sum_j A_j X_j + W, \quad W \sim \mathcal{N}(0, \sigma^2 I) \]

- \{ A_j \} are columns of known sensing matrix with IID entries
- \{ X_j \} are IID Bernoulli-Gaussian: \( p(x) = (1 - \epsilon)\delta_0(x) + \epsilon \mathcal{N}(x \mid 0, \epsilon^{-1}) \)

---

Donoho, Maleki, Montanari’09, Bayati, Montanari’11, R, Gastpar’12
Motivating Questions

- Where do these formulas come from?
- When is AMP optimal?
- Are the replica predictions correct?
- What about other models?
Signal-Plus Noise Model (Scalar Case)

\[ Y = \sqrt{s}X + Z, \quad X \sim p(x), \quad Z \sim \mathcal{N}(0, 1) \]

Guo, Shamai, Verdu’05, R, Pfister, Dytso’18
Signal-Plus Noise Model (Scalar Case)

\[ Y = \sqrt{s}X + Z, \quad X \sim p(x), \quad Z \sim \mathcal{N}(0, 1) \]

\[ I_X(s) \triangleq I(X; \sqrt{s}X + Z) \]

\[ M_X(s) \triangleq \text{mmse}(X \mid \sqrt{s}X + Z) \]

Guo, Shamai, Verdu’05, R, Pfister, Dytso’18
**Signal-Plus Noise Model (Scalar Case)**

\[ Y = \sqrt{s}X + Z, \quad X \sim p(x), \quad Z \sim \mathcal{N}(0, 1) \]

\[ I_X(s) \triangleq I(X; \sqrt{s}X + Z) \]

\[ M_X(s) \triangleq \text{mmse}(X \mid \sqrt{s}X + Z) \]

**I-MMSE Relation:**

\[ \frac{d}{ds} I_X(s) = \frac{1}{2} M_X(s) \]

---

Guo, Shamai, Verdu’05, R, Pfister, Dytso’18
Effect of Signal Prior on MMSE Function

\[ M_X(s) \]

- Gaussian
- Bernoulli-Gaussian
- Bernoulli

Graph showing the relationship between different signal priors and the MMSE function.

\[ s \]

Values range from 0 to 5.

\[ M_X(s) \]

Values range from 0 to 1.
Effect of Signal Prior on MMSE Function (Log Scale)

\[ M_X(s) \]

- **Gaussian**
- **Bernoulli-Gaussian**
- **Bernoulli**
Table of Contents

Bayes Optimal Inference

Phase Transitions in Linear Regression

Information-Theoretic Proof of RS Formula

Further Directions
Experiment: Linear Regression with IID Gaussian matrix

- Length \( n = 10^4 \) vector \( X \) with IID Bernoulli-Gaussian entries
  \[
  X_i \sim p(x) = 0.8 \delta_0(x) + 0.2 \mathcal{N}(x \mid 0, 5)
  \]

- Observations \( Y \in \mathbb{R}^m, A \in \mathbb{R}^{m \times n} \) with SNR = 60 dB
  \[
  Y \sim \mathcal{N}(AX, 10^{-6} I_m), \quad A_{ij} \sim \mathcal{N}(0, 1/n)
  \]
Experiment: Linear Regression with IID Gaussian matrix

- Length $n = 10^4$ vector $X$ with IID Bernoulli-Gaussian entries
  \[ X_i \sim p(x) = 0.8 \delta_0(x) + 0.2 \mathcal{N}(x \mid 0, 5) \]

- Observations $Y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ with SNR = 60 dB
  \[ Y \sim \mathcal{N}(AX, 10^{-6} I_m), \quad A_{ij} \sim \mathcal{N}(0, 1/n) \]

- Use AMP to approximate marginals of posterior: $\hat{p}(x_i \mid Y, A)$

  \[
  \text{AMP squared error:} \quad \frac{1}{n} \sum_{i=1}^{n} (X_i - \mathbb{E}_{\hat{p}}[X_i \mid Y, A])^2 \\
  \text{AMP posterior variance:} \quad \frac{1}{n} \sum_{i=1}^{n} \text{Var}_{\hat{p}}(X_{ii} \mid Y, A). 
  \]

Donoho, Maleki, Montanari’09, Matlab implementation by Schniter et. al.
Experiment: Linear Regression with IID Gaussian matrix

![Graph showing the relationship between the number of observations and the MSE. The graph plots the AMP posterior variance against the number of observations. The variance decreases as the number of observations increases.]
Experiment: Linear Regression with IID Gaussian matrix

![Graph showing the relationship between number of observations and MSE for AMP posterior variance and squared error.](image-url)
Experiment: Linear Regression with IID Gaussian matrix

![Graph showing the relationship between number of observations and MSE, with lines for AMP posterior var. and AMP squared error.](image_url)
Motivating Questions

- Why is there a big change in behavior around 3,500 observations?
- Do AMP estimates correspond to true posterior distribution?
MSE of AMP

signal-plus-noise $M_X(s)$

linear regression

AMP posterior var.
AMP squared error
Define asymptotic MSE

\[ \mathcal{E}_{\text{AMP}}(\delta) \triangleq \lim_{m,n \to \infty} \frac{1}{n} \mathbb{E} \left[ \left\| X - \hat{X}_{\text{AMP}} \right\|^2 \right], \quad m/n \to \delta \]
State Evolution for AMP

Recall from Phil's talk, the AMP per-iteration MSE $\mathcal{E}^t$ satisfies

$$\mathcal{E}^{t+1} = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}^t} \right), \quad \delta = m/n$$
State Evolution for AMP

Recall from Phil’s talk, the AMP per-iteration MSE $\mathcal{E}^t$ satisfies

$$\mathcal{E}^{t+1} = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}^t} \right), \quad \delta = m/n$$
State Evolution for AMP

Recall from Phil’s talk, the AMP per-iteration MSE $\mathcal{E}^t$ satisfies

$$\mathcal{E}^{t+1} = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}^t} \right), \quad \delta = \frac{m}{n}$$
State Evolution for AMP

Recall from Phil’s talk, the AMP per-iteration MSE $\mathcal{E}^t$ satisfies

$$\mathcal{E}^{t+1} = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}^t} \right), \quad \delta = m/n$$
State Evolution for AMP

Recall from Phil’s talk, the AMP per-iteration MSE $\mathcal{E}^t$ satisfies

$$\mathcal{E}^{t+1} = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}^t} \right), \quad \delta = m/n$$

![Phase transition graph](image)
Fixed-Point Analysis of State Evolution

$$f_{SE}(u) \triangleq M_X\left(\frac{\delta}{\sigma^2 + u}\right)$$
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
$f_{SE}(u) \triangleq MX \left( \frac{\delta}{\sigma^2 + u} \right)$

$\mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^{t})$
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq MX \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]
Fixed-Point Analysis of State Evolution

\[ f_{SE}(u) \triangleq M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]

\[ \mathcal{E}^{t+1} = f_{SE}(\mathcal{E}^t) \]

fixed-point \( u = f_{SE}(u) \)
Fixed-Point Analysis of State Evolution (Log Scale)
Fixed-Point Analysis of State Evolution (Log Scale)
Fixed-Point Analysis of State Evolution (Log Scale)
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.30 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.31 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.32 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.33 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.34 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.35 \]
Fixed-Point Analysis of State Evolution (Log Scale)

\[ \delta = 0.36 \]
AMP MSE is Upper Envelope of Fixed-Point Curve

\[ \text{AMP posterior var.} = \frac{E_{\text{AMP}}(\delta)}{\sigma^2} = \max \{ u : u = MX(\sigma^2 + u) \} \]
AMP MSE is Upper Envelope of Fixed-Point Curve

\[
\mathcal{E}_{\text{AMP}}(\delta) = \max \left\{ u : u = M_X \left( \frac{\delta}{\sigma^2 + u} \right) \right\}
\]
\[
\mathcal{I}(\delta) \triangleq \lim_{m,n \to \infty \atop m/n \to \delta} \frac{1}{n} I(\mathbf{X}; \mathbf{Y} \mid A)
\]

\[
\mathcal{E}(\delta) \triangleq \lim_{m,n \to \infty \atop m/n \to \delta} \frac{1}{n} \text{mmse}(\mathbf{X} \mid \mathbf{Y}, A).
\]
Recall scalar mutual information function

\[ I_X(s) = \frac{1}{2} \int_0^s M_X(s') \, ds' \]

Consider potential function:

\[ F_\delta(u) \equiv I_X(\delta u + u^2) + \delta^2 \left( \log \left( 1 + u^2 \sigma^2 \right) - u^2 + u^2 \right) \]

RS formulas defined by global minimum

\[ I_{RS}(\delta) \equiv \min_u F_\delta(u) \]

\[ E_{RS}(\delta) \equiv \arg \min_u F_\delta(u) \]
Replica-Symmetric Formulas

Recall scalar mutual information function

\[ I_X(s) = \frac{1}{2} \int_{0}^{s} M_X(s') \, ds' \]

Consider potential function:

\[ F_\delta(u) \triangleq I_X \left( \frac{\delta}{\sigma^2 + u} \right) + \frac{\delta}{2} \left( \log \left( 1 + \frac{u}{\sigma^2} \right) - \frac{u}{\sigma^2 + u} \right) \]
Replica-Symmetric Formulas

Recall scalar mutual information function

\[ I_X(s) = \frac{1}{2} \int_0^s M_X(s') \, ds' \]

Consider potential function:

\[ \mathcal{F}_\delta(u) \triangleq I_X \left( \frac{\delta}{\sigma^2 + u} \right) + \frac{\delta}{2} \left( \log \left( 1 + \frac{u}{\sigma^2} \right) - \frac{u}{\sigma^2 + u} \right) \]

RS formulas defined by global minimum

\[ \mathcal{I}_{RS}(\delta) \triangleq \min_u \mathcal{F}_\delta(u) \]
\[ \mathcal{E}_{RS}(\delta) \triangleq \arg \min_u \mathcal{F}_\delta(u) \]
RS MMSE is Global Minimizer of Potential Function

\[ \delta = 0.24 \]

\[ \mathcal{F}_\delta(u) \]

\[ \text{global min} \]
RS MMSE is Global Minimizer of Potential Function
Potential Function view of AMP MSE.

By I-MMSE relation, one finds that the stationary points of potential function satisfy fixed-point equation

\[ F'_\delta(u) = 0 \iff u = M_X \left( \frac{\delta}{\sigma^2 + u} \right) \]
Potential Function view of AMP MSE.

By I-MMSE relation, one finds that the stationary points of potential function satisfy fixed-point equation

$$\mathcal{F}_\delta'(u) = 0 \iff u = M_X \left( \frac{\delta}{\sigma^2 + u} \right)$$

Hence, MSE of AMP is largest local minimizer:

$$\mathcal{E}_{\text{AMP}}(\delta) = \max \{ u : \mathcal{F}_\delta'(u) = 0 \}$$
AMP MSE is Largest Local Minimizer
RS Formula for MMSE

\[
\text{AMP posterior var.} \quad \text{AMP squared error} \quad \text{fixed-point curve}
\]

no. observations

\[
\begin{align*}
10^{-1} & \quad 10^{-3} & \quad 10^{-5} & \quad 10^{-7} \\
0 & \quad 1,000 & \quad 2,000 & \quad 3,000 & \quad 4,000 & \quad 5,000 & \quad 6,000 & \quad 7,000 & \quad 8,000
\end{align*}
\]
RS Formula for MMSE

AMP posterior var.
AMP squared error
fixed-point curve
RS MMSE

no. observations
Rigorous Proofs of RS Formula

- Derived using heuristic replica method from statistical physics Tanaka 2001, Guo and Verdu 2005
- First proof by R and Pfister 2016
- Alternate proof using a different method Barbier, Dia, Macris, Krzakala, 2016
Recap of Main Points

- **MSE of AMP** given by largest solution to fixed-point equation
  \[
  \mathcal{E}_{\text{AMP}}(\delta) = \max \left\{ u : u = M_X \left( \frac{\delta}{\sigma^2 + u} \right) \right\}
  \]

- **MMSE** is given by minimum energy solution to fixed-point equation
  \[
  \mathcal{E}(\delta) = \arg \min_u \left\{ I_X \left( \frac{\delta}{\sigma^2 + u} \right) + \frac{\delta}{2} \left( \log \left( 1 + \frac{u}{\sigma^2} \right) - \frac{u}{\sigma^2 + u} \right) \right\}
  \]

- Hence, AMP is optimal if and only if the largest solution is also the minimum energy solution.

- **Phase transition**: when the solution jump from one location to another.
Table of Contents

Bayes Optimal Inference

Phase Transitions in Linear Regression

Information-Theoretic Proof of RS Formula

Further Directions
Proof I: New Observations are Approx. Gaussian

Follow the approach of R. and Pfister 2016.
Proof I: New Observations are Approx. Gaussian

Follow the approach of R. and Pfister 2016.

- Let $Y^i = (Y_1, \ldots, Y_i)$ and apply chain rule

$$I(X, Y | A) = \sum_{i=0}^{m-1} I(X; Y_{i+1} | Y^i, A)$$

info. from new obs.

- If $p(x | Y^i, A)$ is weakly correlated, then $Y_{i+1} | Y^i, A$ is approx. Gaussian:

$$Y_{i+1} - E[Y_{i+1} | Y^i, A] = \sum_{j=1}^{n} A_{ij} (X_j - E[X_j | Y^i, A]) + W_i \approx N(0, \sigma^2) + \text{mmse}(X | Y^i, A)$$

- Hence,

$$I(X, Y | A) \approx m - 1 \sum_{i=0}^{m-1} I(X; Y_{i+1} | Y^i, A)$$

$$\approx m - 1 \sum_{i=0}^{m-1} \frac{1}{2} \log(1 + \frac{1}{n} \text{mmse}(X | Y^i, A) \sigma^2)$$
Proof I: New Observations are Approx. Gaussian

Follow the approach of R. and Pfister 2016.

- Let \( Y^i = (Y_1, \ldots, Y_i) \) and apply chain rule

\[
I(X, Y | A) = \sum_{i=0}^{m-1} I(X; Y_{i+1} | Y^i, A) \]

- If \( p(x | Y^i, A) \) is weakly correlated, then \( Y_{i+1} | Y^i, A \) is approx. Gaussian:

\[
Y_{i+1} - \mathbb{E}[Y_{i+1} | Y^i, A] = \sum_{j=1}^{n} A_{ij} (X_j - \mathbb{E}[X_j | Y^i, A]) + W_i \\
\approx \mathcal{N}\left(0, \frac{1}{n} \text{mmse}(X | Y^i, A) + \sigma^2\right)
\]
Proof I: New Observations are Approx. Gaussian

Follow the approach of R. and Pfister 2016.

Let $Y^i = (Y_1, \ldots, Y_i)$ and apply chain rule

$$I(X, Y \mid A) = \sum_{i=0}^{m-1} I(X; Y_{i+1} \mid Y^i, A)$$

info. from new obs.

If $p(x \mid Y^i, A)$ is weakly correlated, then $Y_{i+1} \mid Y^i, A$ is approx. Gaussian:

$$Y_{i+1} - \mathbb{E}[Y_{i+1} \mid Y^i, A] = \sum_{j=1}^{n} A_{ij} (X_j - \mathbb{E}[X_j \mid Y^i, A]) + W_i$$

$$\approx \mathcal{N}\left(0, \frac{1}{n} \text{mmse}(X \mid Y^i, A) + \sigma^2\right)$$

Hence,

$$I(X, Y \mid A) \approx \sum_{i=0}^{m-1} \frac{1}{2} \log \left(1 + \frac{1}{n} \frac{\text{mmse}(X \mid Y^i, A)}{\sigma^2}\right)$$
Proof II: MMSE Satisfies Fixed-Point Equation

- By symmetry, MMSE is the same for all entries:

\[
\frac{1}{n} \text{mmse}(X \mid Y, A) = \text{mmse}(X_n \mid Y, A)
\]
Proof II: MMSE Satisfies Fixed-Point Equation

- By symmetry, MMSE is the same for all entries:

\[
\frac{1}{n} \text{mmse}(X | Y, A) = \text{mmse}(X_n | Y, A)
\]

- Let \( Q \) be an orthogonal matrix such that the first \((m - 1)\) transformed observations are independent of \( X_n \)

\[
(QY)_i = \tilde{A}_i X^{n-1} + \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, m - 1
\]

\[
(QY)_m \approx \sqrt{m/n} X_n + \tilde{A}_m X^{n-1} + \mathcal{N}(0, \sigma^2)
\]

effective Gaussian noise
Proof II: MMSE Satisfies Fixed-Point Equation

- By symmetry, MMSE is the same for all entries:
  \[
  \frac{1}{n} \text{mmse}(X \mid Y, A) = \text{mmse}(X_n \mid Y, A)
  \]

- Let \( Q \) be an orthogonal matrix such that the first \((m - 1)\) transformed observations are independent of \( X_n \)
  \[
  (QY)_i = \tilde{A}_iX^{n-1} + \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, m - 1
  \]
  \[
  (QY)_m \approx \sqrt{m/n}X_n + \tilde{A}_mX^{n-1} + \mathcal{N}(0, \sigma^2)\quad \text{effective Gaussian noise}
  \]

- If \( p(x \mid (QY)^{m-1}, A) \) is weakly correlated, then effective noise is approx. Gaussian and this leads to the fixed-point equation:
  \[
  \text{mmse}(X_n \mid Y, A) \approx M_X \left( \frac{m/n}{\sigma^2 + \frac{1}{n} \text{mmse}(X^{n-1} \mid (QY)^m, A)} \right)
  \]
Proof III: Weak Correlation Almost Everywhere

By chain rule and problem symmetry, we obtain:

\[ I(X; Y_{m+1} | Y^m, A) = I(X; Y_1 | A) - \sum_{i=0}^{m-1} I(Y_{i+1}; Y_{i+2} | Y^i, A) \]
Proof III: Weak Correlation Almost Everywhere

By chain rule and problem symmetry, we obtain:

\[ I(X; Y_{m+1} \mid Y^m, A) = I(X; Y_1 \mid A) - \sum_{i=0}^{m-1} I(Y_{i+1}; Y_{i+2} \mid Y^i, A) \]

Correlation in \( p(x \mid Y^i, A) \) is upper bounded by

\[ \frac{1}{n^2} \sum_{k, \ell} \mathbb{E}[\text{Cov}(X_k, X_\ell \mid Y^i, A)] \leq \Psi(I(Y_{i+1}; Y_{i+2} \mid Y^i, A)) \]

where \( \Psi(\cdot) \) is concave and non-decreasing.
Proof III: Weak Correlation Almost Everywhere

- By chain rule and problem symmetry, we obtain:

\[
I(X; Y_{m+1} | Y^m, A) = I(X; Y_1 | A) - \sum_{i=0}^{m-1} I(Y_{i+1}; Y_{i+2} | Y^i, A)
\]

- Correlation in \( p(x | Y^i, A) \) is upper bounded by

\[
\frac{1}{n^2} \sum_{k,\ell} \mathbb{E} \left[ \text{Cov}(X_k, X_\ell | Y^i, A) \right] \leq \Psi \left( I(Y_{i+1}; Y_{i+2} | Y^i, A) \right)
\]

where \( \Psi(\cdot) \) is concave and non-decreasing.

- But \( \sum_{i=0}^{\infty} I(Y_{i+1}; Y_{i+2} | Y^i, A) \leq I(Y_1; X | A) \) is finite, and thus

\[
\# \{ i : p(x | Y^i, A) \text{ has significant correlation} \} = o(n)
\]
Proof IV: Resolve Limits via Area Theorem

Show that

\[
\frac{1}{n} I(X; Y \mid A) \approx \frac{1}{n} \sum_{i=1}^{m} \frac{1}{2} \log \left(1 + \frac{1}{n} \frac{\text{mmse}(X \mid Y_i, A)}{\sigma^2}\right)
\]

\[
\frac{1}{n} \text{mmse}(X \mid Y, A) \approx M_X \left(\frac{m/n}{\sigma^2 + \frac{1}{n} \text{mmse}(X \mid Y, A)}\right)
\]
Proof IV: Resolve Limits via Area Theorem

Show that

\[
\frac{1}{n} I(X; Y | A) \approx \frac{1}{n} \sum_{i=1}^{m} \frac{1}{2} \log \left( 1 + \frac{\frac{1}{n} \text{mmse}(X | Y^i, A)}{\sigma^2} \right)
\]

\[
\frac{1}{n} \text{mmse}(X | Y, A) \approx M_X \left( \frac{m/n}{\sigma^2 + \frac{1}{n} \text{mmse}(X | Y, A)} \right)
\]

This implies asymptotic constraints

\[
I(\delta) = \int_0^\delta \frac{1}{2} \log \left( 1 + \frac{\mathcal{E}(\delta')}{\sigma^2} \right) \, d\delta' \quad \text{integral constraint}
\]

\[
\mathcal{E}(\delta) = M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}(\delta)} \right), \quad \text{a.e.} \quad \text{fixed-point constraint}
\]
Proof IV: Resolve Limits via Area Theorem

\[ I(\delta) \approx \int_{0}^{\delta} \frac{1}{2} \log \left( 1 + \frac{\mathcal{E}(\delta')}{\sigma^2} \right) \, d\delta' \quad \text{integral constraint} \]

\[ \mathcal{E}(\delta) \approx M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}(\delta)} \right), \quad \text{a.e. fixed-point constraint} \]

\[ \frac{1}{2} \log \left( 1 + \frac{\mathcal{E}(\delta)}{\sigma^2} \right) \]

\[ \delta = \frac{m}{n} \]

find jump such that area under curve matches mutual information in large-\( \delta \) limit
Proof IV: Resolve Limits via Area Theorem

\[ \mathcal{I}(\delta) \approx \int_0^\delta \frac{1}{2} \log \left( 1 + \frac{\mathcal{E}(\delta')}{\sigma^2} \right) d\delta' \]  
integral constraint

\[ \mathcal{E}(\delta) \approx M_X \left( \frac{\delta}{\sigma^2 + \mathcal{E}(\delta)} \right), \quad \text{a.e.} \]  
fixed-point constraint

\[ \frac{1}{2} \log \left( 1 + \frac{\mathcal{E}(\delta)}{\sigma^2} \right) \]

find jump such that area under curve matches mutual information in large-\( \delta \) limit
Table of Contents

Bayes Optimal Inference

Phase Transitions in Linear Regression

Information-Theoretic Proof of RS Formula

Further Directions
Inference with Mismatched Prior

- Postulated distribution $q(x, y)$ differs from true distribution $p(x, y)$

- Example: MAP estimation via expectation with modified likelihood:
  \[
  q_\beta(y, x) \propto [p(y \mid x)]^\beta p(x)
  \]
  \[
  \int x q_\beta(x \mid y) \, dx \xrightarrow{\beta \to \infty} \arg \max_x p(x \mid y)
  \]
  \[
  \text{postulated mean} \quad \text{MAP estimate}
  \]

- Conjectured formulas for mismatched MMSE and M-estimators (e.g., MAP, MLE, LASSO)

- But replica symmetry does not hold in general!

---

Guo, Verdu’05, Rangan, Fletcher, Goyal’09, Bereyhi, Müller, Schulz-Baldes’17
Information in Multilayer Networks

Feed forward network with random weight matrices \( \{W_\ell\} \)

- Component-wise activation functions \( \{\sigma_\ell\} \)

\[
x \xrightarrow{\sigma_1(W_1x)} h_1 \xrightarrow{\ldots} h_{L-1} \xrightarrow{\sigma_L(W_Lh_{L-1})} y
\]

- Separable prior \( p(x) \) and separable distributions \( \{p_\ell(\cdot | \cdot)\} \)

\[
p(x) \xrightarrow{X} p_1(\cdot | W_1x) \xrightarrow{H_1} \ldots \xrightarrow{H_{L-1}} p_L(\cdot | W_Lh_{L-1}) \xrightarrow{Y}
\]

Manoel, Krzakala, Mézard, Zdeborová’17, Fletcher, Rangan, Schniter’18, R.’17, Gabrié, et al.’18
Information in Multilayer Networks

Feed forward network with random weight matrices \( \{W_\ell\} \)

- Component-wise activation functions \( \{\sigma_\ell\} \)

\[
x \xrightarrow{\sigma_1(W_1x)} h_1 \xrightarrow{\ldots} h_{L-1} \xrightarrow{\sigma_L(W_Lh_{L-1})} y
\]

- Separable prior \( p(x) \) and separable distributions \( \{p_\ell(\cdot | \cdot)\} \)

\[
p(x) \xrightarrow{X} p_1(\cdot | W_1x) \xrightarrow{H_1} \ldots \xrightarrow{H_{L-1}} p_L(\cdot | W_Lh_{L-1})
\]

What is mutual information between layers?

- State evolution for multilayer AMP / VAMP
- Conjectured formulas via heuristics
- Observed behavior under restricted learning of weights

Manoel,Krzakala,Mézard,Zdeborová’17, Fletcher,Rangan,Schniter’18, R.’17, Gabrié,et.al’18
Phase Diagrams in Other Problem

- Recent progress
  - Multilayer networks (random initialization)
  - Bilinear and multi-linear models (low rank) [See Jean’s Talk]
  - Community detection (dense case)
Phase Diagrams in Other Problem

- Recent progress
  - Multilayer networks (random initialization)
  - Bilinear and multi-linear models (low rank) [See Jean’s Talk]
  - Community detection (dense case)

- Future challenges
  - Multilayer networks (learned from data)
  - Bilinear and multi-linear models (extrinsic rank)
  - Community detection (sparse case)
References I


