Statistical Physics of Compressed Sensing

Song Nam Hong

Ming Hsieh Department of EE, University of Southern California, Los Angeles CA

April, 2012, Los Angeles, CA

Contents

- What's the compressed sensing and applications? (See Lectures 25 and 26)

- Focusing on asymptotic analysis of compressed sensing algorithm (e.g., L_1-minimization) using Replica Method

Problem Definition

Reconstruction Problem with linear measurements:

$$\mathbf{y} = \mathbf{F}\mathbf{x}^0$$

where

- Original vector: $\mathbf{x}^0 = (x_1, ..., x_N) \in \mathbb{R}^N$
- Measurement (or compression) matrix: $\mathbf{F} \in \mathbb{R}^{N \times P}$
- Observations: $\mathbf{y} = (y_1, ..., y_P) \in \mathbb{R}^P$

$\Rrightarrow$ For a given $\mathbf{y}$ and $\mathbf{F}$, we want to reconstruct $\mathbf{x}^0$?

Problem Definition

- If $P = N$, we can find an unique solution using matrix inversion:

$$\mathbf{x}^0 = \mathbf{F}^{-1}\mathbf{y}$$

- If $P < N$, it is under-constrained system

$$\mathcal{S} = \{\mathbf{x} : \mathbf{F}\mathbf{x} = \mathbf{y}\} \text{ with } |\mathcal{S}| >> 1$$

[Fundamental Questions]

- If we can employ a priori knowledge on $\mathbf{x}^0$ (e.g.,$P(\mathbf{x}^0)$, sparsity, etc), can we recover $\mathbf{x}^0$ with smaller number of measurements (e.g., $P < N$) ?

- If so, what is the minimum number of measurements?

Priori Knowledge

- Bernoulli-Gaussian Distribution:

$$P(x) = (1-\rho)\delta(x) + \rho\mathcal{N}(0,1)$$

- Bernoulli-Uniform Distribution:

$$P(x) = (1-\rho)\delta(x) + \rho\text{Unif}(-\sqrt{3}, \sqrt{3})$$

- Sparsity:

  Both distributions give that the number of non-zero elements of $\mathbf{x}^0$ converges to ρN as $N \to \infty$

Reconstruction (or Estimation) Methods

- MAP Estimation (Using the full-knowledge $P(x)$)

$$\begin{aligned}\hat{\mathbf{x}} &= \underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} -\ln P(\mathbf{x}|\mathbf{y}) \\ &= -\ln\delta(\mathbf{F}\mathbf{x}-\mathbf{y}) - \ln P(\mathbf{x}).\end{aligned}$$

In an optimization form, we have

$$\begin{aligned}\underset{\mathbf{x}}{\operatorname{argmin}} &\quad -\ln P(\mathbf{x}) \\ \text{subject to} &\quad \mathbf{F}\mathbf{x} = \mathbf{y}\end{aligned}$$

This problem is in general NP-Hard!

Reconstruction (or Estimation) Methods

- L_1-Minimization (Using the partial-knowledge, sparsity)

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\operatorname{argmin}} \quad ||\mathbf{x}||_1$$
$$\text{subject to} \quad \mathbf{F}\mathbf{x} = \mathbf{y}$$

  - This relaxation is based on the fact that L_1-minimization gives a sparse solution
  - Convex optimization problem

Model for Asymptotic Analysis

- Random measurement matrix and random input signal

$$\mathbf{F} \in \mathbb{R}^{P \times N} \text{ and } \mathbf{x}^0 = (x_1, ..., x_N) \text{ with } x_i \sim P(x)$$

  - $P(x)$ with $\frac{1}{N}|\mathbf{x}^0|^2 = \rho$ as $N \to \infty$ (sparsity, ρ)

- Asymptotic case

$$N, P \to \infty \text{ with } \frac{P}{N} = \alpha$$

- L_p-reconstruction $(p = 1, 2)$

Statistical Physics Model

- Hamiltonian

$$H(\mathbf{x}) = -\log(\delta(\mathbf{F}\mathbf{x} - \mathbf{y})) + ||\mathbf{x}||_p$$

- Gibbs distribution with temperature β

$$P_\beta(\mathbf{x}|\mathbf{y}) = \frac{1}{Z}\exp(-\beta H(\mathbf{x})) = \frac{1}{Z}\exp(-\beta||\mathbf{x}||_p)\delta(\mathbf{F}\mathbf{x} - \mathbf{y})$$

where Z denotes a partition function

$$Z = \int d\mathbf{x}\exp(-\beta||\mathbf{x}||_p)\delta(\mathbf{F}\mathbf{x} - \mathbf{y}).$$

Statistical Physics Model

- In the low-temperature limit (e.g., $\beta \to \infty$),
  - Gibbs distribution converges to a uniform distribution of the ground states

$$\lim_{\beta \to \infty} P_\beta(\mathbf{x}|\mathbf{y}) \to \begin{cases} \frac{1}{|\mathcal{S}|}, & \text{if } \mathbf{x} \in \mathcal{S} \\ 0, & \text{otherwise.} \end{cases}$$

  where $\mathcal{S}$ denotes the set of ground states (e.g., solutions of our optimization problem).

♣ In L_p-reconstruction $(p = 1, 2)$, we have an unique solution since the objective function is convex.

Performance Measure Metric

Definition 1. *Mean Squared Error (MSE) per element:*

$$
\begin{aligned}
MSE(\mathbf{x}^0, \mathbf{F}) &= \frac{1}{N}\sum_{\mathbf{x}\in\mathcal{S}} \frac{1}{|\mathcal{S}|}|\mathbf{x}-\mathbf{x}^0|^2 \\
&= \frac{1}{N}\left\langle |\mathbf{x}-\mathbf{x}^0|^2 \right\rangle \\
&= \frac{1}{N} < |\hat{\mathbf{x}}|^2 > -\frac{2}{N} < \hat{\mathbf{x}}\cdot\mathbf{x}^0 > +\frac{1}{N}|\mathbf{x}^0|^2 \\
&= \frac{1}{N} < |\hat{\mathbf{x}}|^2 > -\frac{2}{N} < \hat{\mathbf{x}}\cdot\mathbf{x}^0 > +\rho
\end{aligned}
$$

where $< \cdot >$ *denotes the averaging with respect to Gibbs distribution.*

Order Parameters

- Parameters to determine the MSE

$$Q = \frac{1}{N}|\hat{\mathbf{x}}|^2, \; m = \frac{1}{N}\hat{\mathbf{x}} \cdot \mathbf{x}^0, \text{ and } \rho = \frac{1}{N}|\mathbf{x}^0|^2$$

- Mean squared error is computed as

$$MSE(\mathbf{x}^0, \mathbf{F}) = Q - 2m + \rho$$

- If $\hat{\mathbf{x}} = \mathbf{x}^0$, then $Q = m = \rho$ and

$$MSE(\mathbf{x}^0, \mathbf{F}) = 0$$

Reconstruction Limit $\alpha_c(\rho)$

Definition 2. *For a given ρ, a reconstruction limit (e.g., threshold) $\alpha_c(\rho)$ is defined as*

- *If* $\alpha \geq \alpha_c(\rho)$,

$$\mathbb{E}[MSE(\mathbf{F}, \mathbf{x}^0)]_{\mathbf{F},\mathbf{x}^0} = 0 \Leftrightarrow Q = m = \rho$$

- *If* $\alpha < \alpha_c(\rho)$,

$$\mathbb{E}[MSE(\mathbf{F}, \mathbf{x}^0)]_{\mathbf{F},\mathbf{x}^0} = \text{ *constant* } \Leftrightarrow Q \neq m \neq \rho$$

- We want find an $\alpha_c(\rho)$ using Replica Method
  - Replica Method $\Rightarrow Q$ and m

Replica Method

Replica Method is a standard technique to study the free energy of disordered system.

$$\mathcal{F} = -\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{\beta N}\log Z\right]$$

Essentially, it takes the following steps:

1. Using Replica Trick ($\mathbb{E}[\log Z] = \lim_{n\to 0}\frac{1}{n}\log(\mathbb{E}[Z^n])$), we can rewrite:

$$\mathcal{F} = -\lim_{N\to\infty}\frac{1}{\beta N}\lim_{n\to 0}\frac{\partial}{\partial n}\log \mathbb{E}[Z^n]$$

2. For an arbitrary positive integer n, calculate

$$\mathcal{C} = -\lim_{N\to\infty}\frac{1}{N}\log \mathbb{E}[Z^n]$$

3. Assuming the resulting expression in the above to be valid for all $n \in \mathbb{R}$ at the vicinity of $n = 0$, takes its derivative at $n = 0$ to obtain $\mathcal{F}$ as

$$\mathcal{F} = \lim_{n \to 0} \frac{1}{\beta} \frac{\partial \mathcal{C}}{\partial n}$$

[Main Observation] By Replica Method under replica symmetry,

$$\mathcal{F}_{\{\beta \to \infty\}} = \text{extr}_{Q,m,q,\hat{Q},\hat{m},\hat{q}} f(Q, q, m, \hat{Q}, \hat{q}, \hat{m})$$

where extr $f(x)$ denotes extremization of a function $f(x)$ with respect to x.

- Let Q^* and m^* be the solution of the above optimization.

$$Q^* = \frac{1}{N} < |\hat{\mathbf{x}}|^2 > \text{ and } m^* = \frac{1}{N} < \hat{\mathbf{x}} \cdot \mathbf{x}^0 >$$

Replica Method

- Step 2 in Replica Method

$$\mathcal{C} = -\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0}$$

where

$$\mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0} = \Big[\int \prod_{a=1}^{n} d\mathbf{x}^a e^{-\beta||\mathbf{x}^a||_p} \Big[\prod_{a=1}^{n} \delta(\mathbf{F}\mathbf{x}^a - \mathbf{F}\mathbf{x}^0)\Big]_{\mathbf{F}}\Big]_{\mathbf{x}^0}$$

Replica Method

◇ Replica Symmetry (RS) Assumption:

$$\mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0} = \Big[\int_{Q,q,m} \partial Q \partial q \partial m \int \prod_{a=1}^{n} d\mathbf{x}^a \mathcal{I}(\{\mathbf{x}^a\}_{a=1}^n, \mathbf{x}^0; Q, q, m\}) \\ e^{-\beta||\mathbf{x}^a||_p} \Big[\prod_{a=1}^{n} \delta(\mathbf{F}\mathbf{x}^a - \mathbf{F}\mathbf{x}^0) | Q, q, m \Big]_{\mathbf{F}} \Big]_{\mathbf{x}^0}.$$

where

$$\mathbf{x}^0 \cdot \mathbf{x}^a = Nm$$
$$\mathbf{x}^a \cdot \mathbf{x}^b = Nq \text{ for } a \neq b$$
$$\mathbf{x}^a \cdot \mathbf{x}^a = NQ$$

⇛ Integration is dominated by $\{\mathbf{x}^a : a = 1, ..., n\}$ satisfying the above equations

⇛ This assumption is not always valid, which will be checked later

Replica Method

◇ Based on the results in [1],[2]:

$$\left[\delta(\mathbf{F}\mathbf{x}^a - \mathbf{F}\mathbf{x}^0)|Q, q, m\right]_{\mathbf{F}} = \exp(N\mathcal{T}_n(Q, q, m))$$

where

$$\mathcal{T}_n(Q, q, m) = -\frac{\alpha}{2}\ln\left(\left(1 - \frac{n(Q - 2m + \rho)}{Q - q}\right)(Q - q)^{2n}(2\pi)^n\right)$$

[1] K. Takeda, A. Hatabu, and Y. Kabashima, ”Statistical mechanical analysis of the linear vector channel in digital communication”

[2] K. Takeda, S. Uda, and Y. Kabashima, ”Analysis of CDMA systems that are characterized by eigenvalue spectrum”

Replica Method

◇ We can rewrite:

$$\mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0} = \int_{Q,q,m} \partial Q \partial q \partial m e^{(N\mathcal{T}_n(Q,q,m))} \underbrace{\left[\int \prod_{a=1}^{n} d\mathbf{x}^a \mathcal{I}(\{\mathbf{x}^a\}_{a=1}^n, \mathbf{x}^0; Q, q, m\}) e^{-\beta ||\mathbf{x}^a||_p} \right]_{\mathbf{x}^0}}_{(\star)}$$

where $(\star)$ captures the volume of configuration of $\{\mathbf{x}^a\}$ to satisfy the condition given by Q, q, m.

Replica Method

- By large-deviation principle (See [1]-[2]), we can have:

$$(\star) = \exp(N\mathcal{S}_n(Q,q,m))$$

where

- $$\mathcal{S}_n(Q,q,m) = \text{extr}_{\hat{Q},\hat{m},\hat{q}}\left\{\frac{n\hat{Q}Q}{2} - \frac{n(n-1)\hat{q}q}{2} - n\hat{m}m + \ln G(\hat{Q},\hat{q},\hat{m})\right\}$$

- $$G(\hat{Q},\hat{q},\hat{m}) = \lim_{\epsilon\to 0}\int \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz\left[\left(\int dx\exp(-(\hat{Q}+\hat{q})x^2/2 + (\sqrt{\hat{q}}z+\hat{m}x^0)x - \beta|x|^{p+\epsilon})\right)^n\right]_{x^0}$$

Replica Method

- Finally, we can obtain:

$$\mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0} = \int_{Q,q,m} \partial Q \partial q \partial m \exp\left(N\mathcal{T}_n(Q,q,m)\right) \exp\left(N\mathcal{S}_n(Q,q,m)\right)$$
$$\approx \text{extr}_{Q,q,m} \exp(N(\mathcal{T}_n(Q,q,m) + \mathcal{S}_n(Q,q,m)))$$

when $N \to \infty$.

This is based on the well-known fact that

$$e^{v_1} + \cdots e^{v_N} \approx e^{v^*} \text{ when } v^* \text{ is sufficiently large}$$

where $v^* = \max\{v_1, ..., v_N\}$.

Replica Method

$$\mathcal{C} = -\lim_{N\to\infty} \frac{1}{N} \ln \mathbb{E}[Z^n]_{\mathbf{F},\mathbf{x}^0} \approx \mathcal{T}_n(Q,q,m) + \mathcal{S}_n(Q,q,m)$$

- Free Energy Density $(\beta = \infty)$

$$\begin{aligned}
\mathcal{F}_{\{\beta=\infty\}} &= \lim_{\beta\to\infty} \lim_{n\to 0} \frac{1}{\beta}\frac{\partial \mathcal{C}}{\partial n} \\
&= extr_{Q,\chi,m,\hat{Q},\hat{\chi},\hat{m}} \Big[\frac{\alpha(Q-2m+\rho)}{2\chi} + \hat{m}m - \frac{\hat{Q}Q}{2} + \frac{\hat{\chi}\chi}{2} + \\
&\quad \int dx^0 P(x^0) \int Dz \phi_p(\hat{m}x^0 + \sqrt{\hat{\chi}}z; \hat{Q}) \Big]
\end{aligned}$$

where

$$\phi_p(h;\hat{Q}) = \lim_{\epsilon\to 0} \left\{ \min_x \left\{ \frac{\hat{Q}}{2}x^2 - hx + |x|^{p+\epsilon} \right\} \right\}$$

♢ Valid for any priori distribution with $|\mathbf{x}^0|^2 = \rho N$

Bernoulli-Gaussian Distribution

- Free Energy Density:

$$\mathcal{F}_{\{\beta=\infty\}} = \lim_{\beta\to\infty}\lim_{n\to 0}\frac{1}{\beta}\frac{\partial \mathcal{C}}{\partial n}$$
$$= extr_{Q,\chi,m,\hat{Q},\hat{\chi},\hat{m}}\Big[\frac{\alpha(Q-2m+\rho)}{2\chi}+\hat{m}m-\frac{\hat{Q}Q}{2}+\frac{\hat{\chi}\chi}{2}+$$
$$(1-\rho)\int Dz\phi_p(\sqrt{\hat{\chi}}z;\hat{Q})+\rho\int Dz\phi_p(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q})\Big]$$

where

$$\phi_p(h;\hat{Q}) = \lim_{\epsilon\to 0}\left\{\min_x\left\{\frac{\hat{Q}}{2}x^2-hx+|x|^{p+\epsilon}\right\}\right\}$$

Bernoulli-Gaussian Distribution

- Let $x^*(h;\hat{Q})$ be the optimal solution of $\phi_p(h;\hat{Q})$
- Performing saddle point derivative with respect to $Q, m, \chi, \hat{Q}, \hat{m}, \hat{\chi}$:

$$
\begin{aligned}
\hat{Q} &= \hat{m} = \frac{\alpha}{\chi} \\
\hat{\chi} &= \frac{\alpha(Q - 2m + \rho)}{\chi^2} \\
Q &= (1-\rho)\int x^*(\sqrt{\hat{\chi}}z;\hat{Q})^2 Dz + \rho \int x^*(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q})^2 Dz \\
m &= \rho \frac{\hat{m}}{\sqrt{\hat{\chi}+\hat{m}^2}} \int x^*(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q}) z Dz \\
\chi &= (1-\rho)\frac{1}{\sqrt{\hat{\chi}}}\int x^*(\sqrt{\hat{\chi}}z;\hat{Q}) z Dz + \rho \frac{1}{\sqrt{\hat{\chi}+\hat{m}^2}} \int x^*(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q}) z Dz
\end{aligned}
$$

L_2-minimization

$$x^*(h;\hat{Q}) = \frac{h}{\hat{Q}+2}$$

The above equations are simplified as

$$\begin{aligned}
\hat{Q} &= \hat{m} = \frac{\alpha}{\chi} \\
\hat{\chi} &= \frac{\alpha(Q-2m+\rho)}{\chi^2} \\
Q &= \frac{\hat{\chi}+\rho\hat{Q}^2}{(\hat{Q}+2)^2} \\
m &= \rho\frac{\hat{Q}}{\hat{Q}+2} \\
\chi &= \frac{1}{\hat{Q}+2}
\end{aligned}$$

L_2-minimization

- A successful solution implies the solution with $Q = m = \rho$, no restriction on other variables
- Set $\hat{Q} = \hat{m} = \infty$ and accordingly, we have $\chi = 0$
- $\hat{\chi}$ exits?

$$
\begin{aligned}
\hat{\chi} &= \alpha \frac{\hat{m}^2}{\alpha^2}\Big(\frac{\hat{\chi} + \rho \hat{m}^2}{(\hat{Q}+2)^2} - 2\rho \frac{\hat{m}}{\hat{Q}+2} + \rho\Big) \\
&= \alpha^{-1} \frac{\hat{Q}^2}{(\hat{Q}+2)^2}(\hat{\chi} + 4\rho) \\
&= \alpha^{-1}(\hat{\chi} + 4\rho)
\end{aligned}
$$

Based on this, we can see that $\alpha_c(\rho) > 1$ for any $\rho > 0$.

L_1-minimization

- Similarly, we have:

$$
\begin{aligned}
Q &= \frac{2(1-\rho)}{\hat{Q}^2}\Big(Q(\hat{\chi}^{-1/2}) - 2\sqrt{\hat{\chi}}\frac{\exp(-1/(2\hat{\chi}))}{\sqrt{2\pi}} + \int_{\hat{\chi}^{-1/2}}^{\infty} \hat{\chi} z^2 Dz\Big) \\
&\quad + \frac{\rho}{\hat{Q}^2}\int_{|z|>1/\sqrt{\hat{\chi}+\hat{m}^2}} (\sqrt{\hat{\chi}+\hat{m}^2}z - \frac{z}{|z|})^2 Dz \\
m &= \rho\frac{\hat{m}}{\sqrt{\hat{\chi}+\hat{m}^2}}\int_{|z|>1/\sqrt{\hat{\chi}+\hat{m}^2}} (\sqrt{\hat{\chi}+\hat{m}^2}z - \frac{z}{|z|}) z Dz
\end{aligned}
$$

- $\hat{\chi}$ exist?

$$
\hat{\chi} = \alpha^{-1}\Big[2(1-\rho)\Big(Q(\hat{\chi}^{-1/2}) - 2\sqrt{\hat{\chi}}\frac{\exp(-1/(2\hat{\chi}))}{\sqrt{2\pi}} + \hat{\chi}\int_{\hat{\chi}^{-1/2}}^{\infty} z^2 Dz\Big) + \rho(\hat{\chi}+1)\Big]
$$

where $Q(x)$ is a conventional *Q-function* with $Q(x) = \int_x^\infty Dz$.

Replica Symmetry Breaking

- RS assumption is not necessarily guaranteed to be correct
- Local stability of the RS saddle point is lost against pertubations that break the RS if the following holds:

$$\frac{\alpha}{\chi^2}\Big((1-\rho)\int\Big(\frac{\partial x_p^*(\sqrt{\hat{\chi}}z;\hat{Q})}{\partial(\sqrt{\hat{\chi}}z)}\Big)^2 Dz+\rho\int\Big(\frac{\partial x_p^*(\sqrt{\hat{\chi}+\hat{m}^2}z;\hat{Q})}{\partial(\sqrt{\hat{\chi}+\hat{m}^2}z)}\Big)^2 Dz\Big)>1$$

♣ When the above condition holds for the extremum solutions, the RS is not valid

♣ Consider more general solutions taking into account the effect of replica symmetry breaking (RSB).

Numerical Results: L_1-minimization

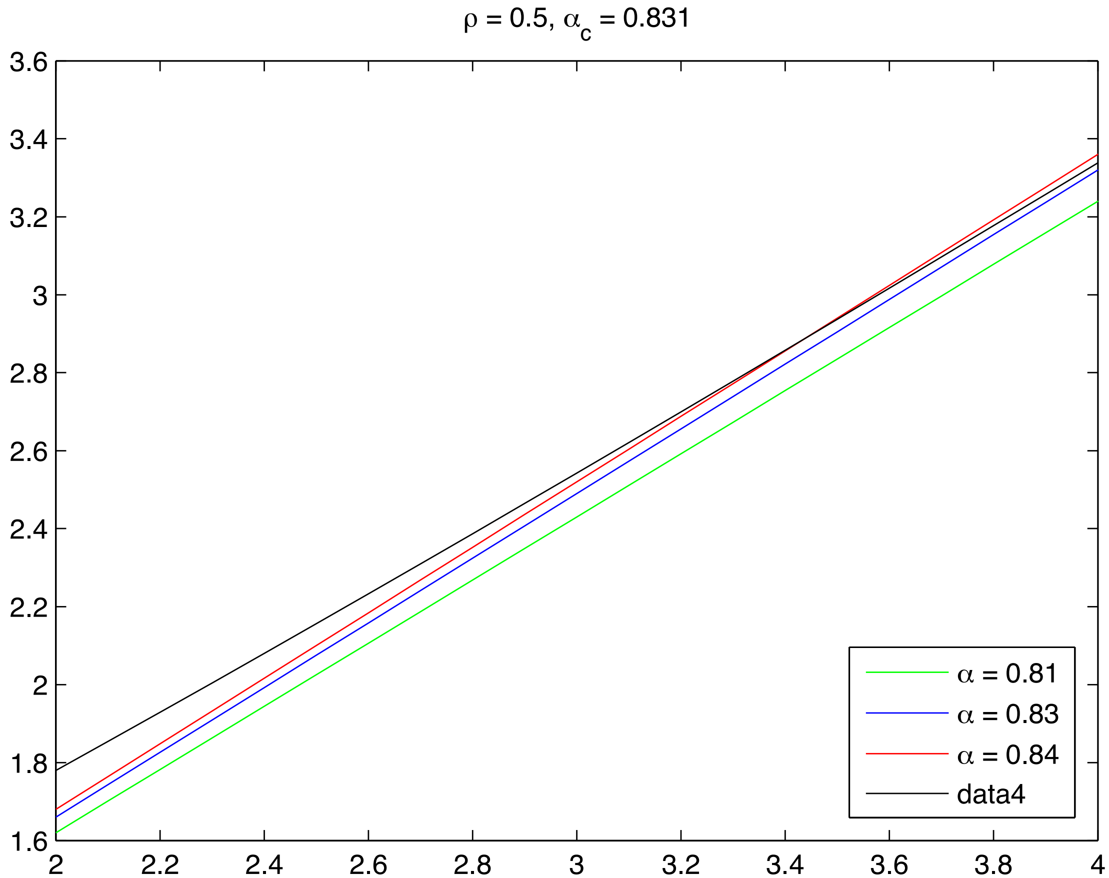

Figure 1: Fixed Point Equations: $\rho = 0.5$ and $\alpha_c(\rho) \approx 0.831$

Numerical Results: L_1-minimization

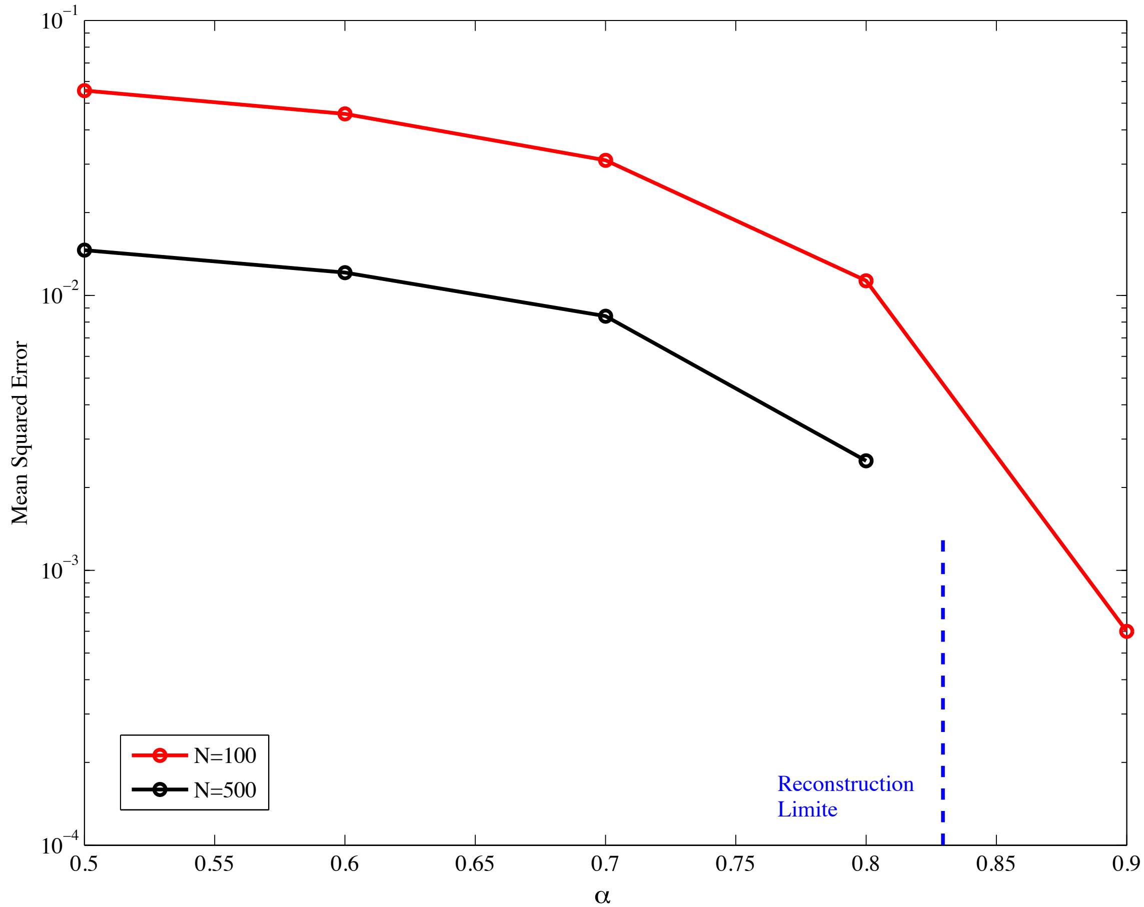

Figure 2: Finite-Length Simulation: $N = 100$ or 500 and $\rho = 0.5$

On going work...

- Working on the asymptotic analysis with Bernoulli-Uniform Distribution
- Want to see if L_1-minimization gives the same reconstruction limit $(\alpha_c(\rho))$ independent of a priori distributions if they have the same sparsity (e.g., $\frac{1}{N}|\mathbf{x}^0|^2 = \rho$)

On going work...

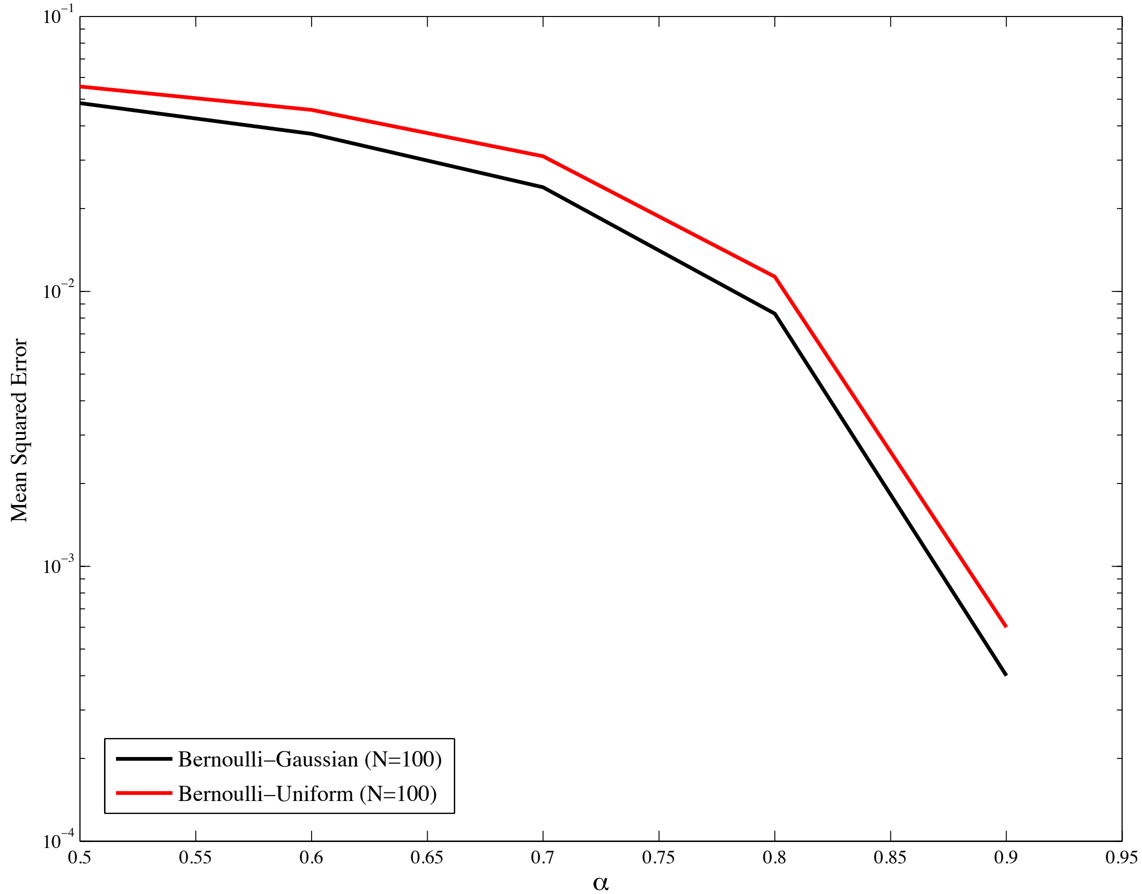

Figure 3: Finite-Length Simulation: $N = 100$ and $\rho = 0.5$

References

[1] K. Takeda, A. Hatabu, and Y. Kabashima, ”Statistical mechanical analysis of the linear vector channel in digital communication”

[2] K. Takeda, S. Uda, and Y. Kabashima, ”Analysis of CDMA systems that are characterized by eigenvalue spectrum”

[3] D. Guo and S. Verdye, ”Randomly spread CDMA: Asymptotics via statistical physics”

[4] S. Rangan, ”Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing”

[5] Y. Kabashima, T. Wadayama, and T. Tanaka, “Statistical Mechanical Analysis of a Typical Reconstruction Limit of Compressed Sensing”