Universality in Compressed Sensing

Andrea Montanari

Stanford University

with Mohsen Bayati and Marc Lelarge

October 27, 2015

Outline

1. Compressed Sensing
2. The universality problem
3. Proof outline
4. An interesting phenomenon
5. Conclusion

Compressed Sensing

Compressed Sensing

$$y = A\,x_0 + w$$

Estimate $x_0 \in \mathbb{R}^N$ given $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times N}$.

Compressed Sensing

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}_0 + \boldsymbol{w}$$

Estimate $\boldsymbol{x}_0 \in \mathbb{R}^N$ given $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{A} \in \mathbb{R}^{n\times N}$.

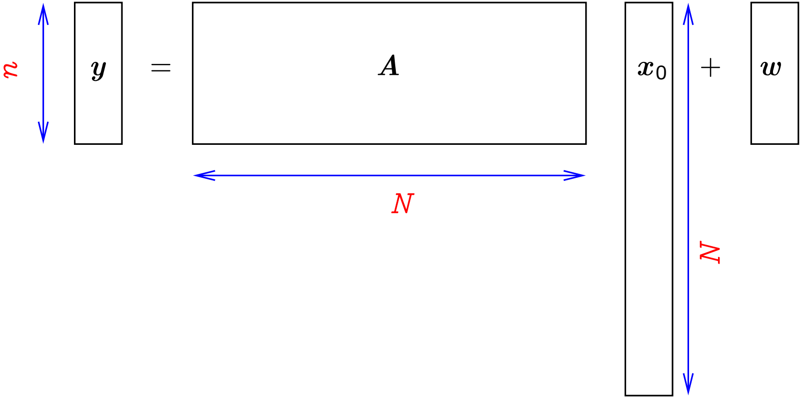
Classical case: $n > N$

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}_0 + \boldsymbol{w}$$

$$\widehat{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{A}) = \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

[Legendre, 1805; Gauss 1809]

Compressed Sensing: $n < N$



Candes, Donoho, Tao, Tropp, Indyk, Gilbert, . . . , 2006-. . . 10^4 papers

$n < N$?

Key assumption: x_0 is sparse!

Example:

$$\|x_0\|_0 = |\mathrm{supp}(x_0)| = k \ll N$$

$n < N$?

Key assumption: x_0 is sparse!

Example:

$$\|x_0\|_0 = |\text{supp}(x_0)| = k \ll N$$

Method of choice

LASSO/Basis pursuit denoising:

$$\widehat{x}(y, A) = \arg\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 \right\}$$

[Tibshirani 1996; Chen, Donoho 1994]

Noiseless case: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$

Basis pursuit $(\lambda \to 0)$

$$\begin{aligned} &\text{minimize} && \|x\|_1\,, \\ &\text{subject to} && \boldsymbol{y} = \boldsymbol{A}x\,. \end{aligned}$$

Does this work?

Noiseless case: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$

Basis pursuit $(\lambda \to 0)$

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{x}\|_1 \,, \\ &\text{subject to} && \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \,. \end{aligned}$$

Does this work?

Noiseless case: $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$

Basis pursuit ($\lambda \to 0$)

$$\begin{aligned} &\text{minimize} && \|\boldsymbol{x}\|_1 \,, \\ &\text{subject to} && \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \,. \end{aligned}$$

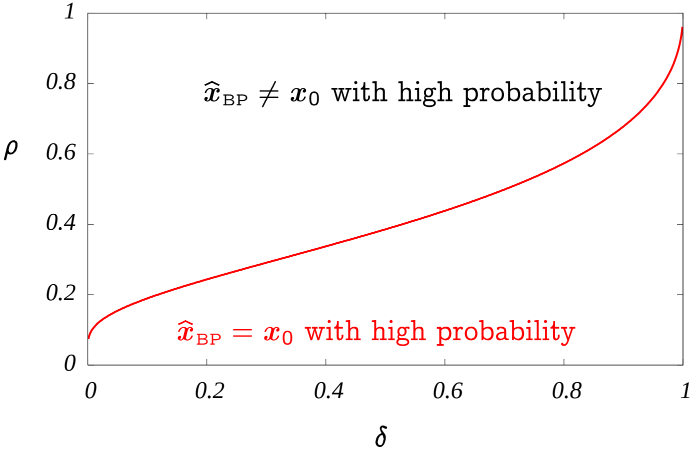
Does this work?

Does this work? Gaussian model

- $A_{ij} \sim_{i.i.d.} \mathrm{N}(0, 1/n)$
- $n/N \to \delta \in (0, 1)$
- $\|x_0\|_0/n \to \rho \in (0, 1)$

Phase diagram 1: 'ℓ_0–ℓ_1 equivalence' (noiseless)

$$N, n \to \infty, \qquad n/N = \delta, \qquad \|x_0\|_0/n = \rho$$



[Donoho 2006, Affentranger-Schneider 1992]

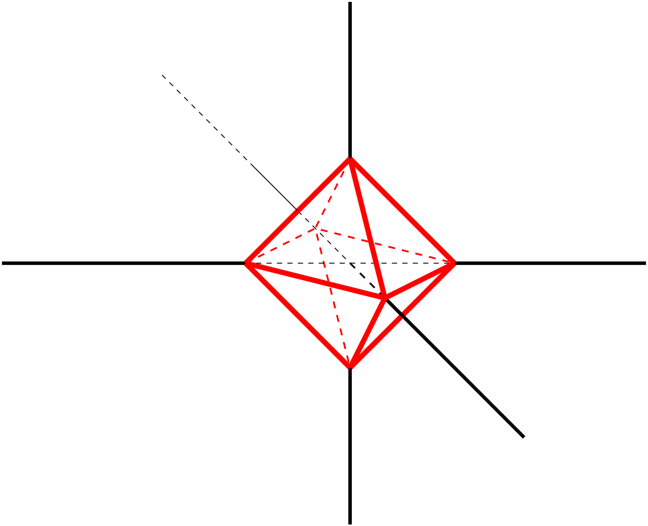
Phase transition curve

$$\delta = \frac{2\phi(\alpha)}{\alpha + 2(\phi(\alpha) - \alpha\Phi(-\alpha))}\,,$$
$$\rho = 1 - \frac{\alpha\Phi(-\alpha)}{\phi(\alpha)}\,.$$

$$\alpha \in [0,\infty), \qquad \phi(x) \equiv e^{-x^2/2}/\sqrt{2\pi}, \qquad \Phi(x) \equiv \int_{-\infty}^{x} \phi(z)\,\mathrm{d}z$$

Entr'acte: Geometric interpretation

ℓ_1 ball in $\mathbb{R}^N$



$\|x_0\|_0 = 1 \leftrightarrow$ vertices

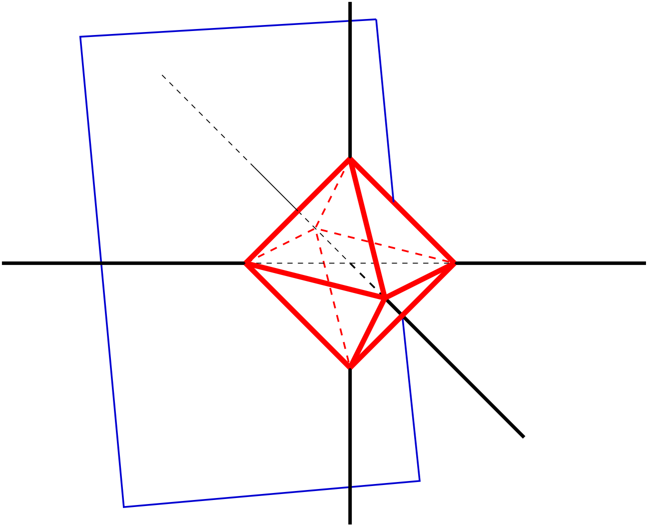
$\|x_0\|_0 = 2 \leftrightarrow$ edges

$\|x_0\|_0 = 3 \leftrightarrow$ 2-faces

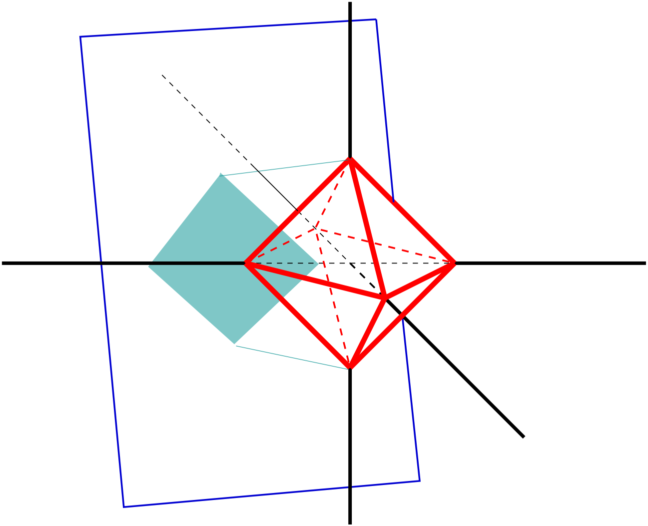
$\|x_0\|_0 = 4 \leftrightarrow$ 3-faces

...

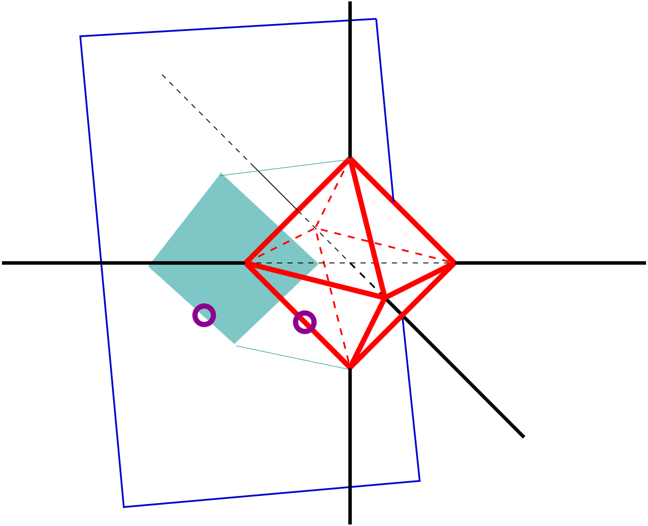
Random n-dimensional plane



Projection

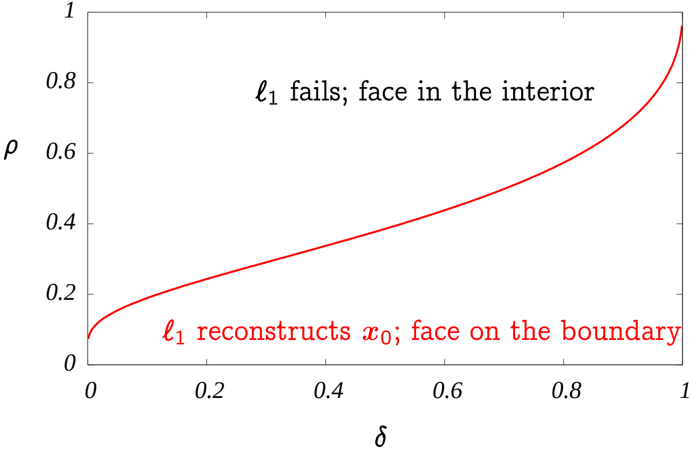


Does it fall on the boundary?



Phase diagram 1: 'ℓ_0–ℓ_1 equivalence' (noiseless)

$$N, n \to \infty, \quad n/N = \delta, \quad \|x_0\|_0/n = \rho$$



[Donoho 2006, Affentranger-Schneider 1992]

Geometric phase transition

- For $k \leq (\rho_c(\delta) - \varepsilon)\, n$, most k-faces fall on the boundary of the shadow.
- For $k \geq (\rho_c(\delta) + \varepsilon)\, n$, most k-faces fall on the boundary of the shadow.

The universality problem

Conjecture (Donoho, Tanner, 2009)

The above predictions are universal for a broad range of random matrices.

Donoho, Tanner, *Phil. Trans. R. Soc. A* (2009) **367**, 4273-4293

across a range of underlying matrix ensembles. We ran millions of linear programs using random matrices spanning several matrix ensembles and problem sizes; visually, the empirical phase transitions do not depend on the ensemble, and they agree extremely well with the asymptotic theory assuming Gaussianity. Careful statistical analysis reveals

Example: Random Fourier measurements

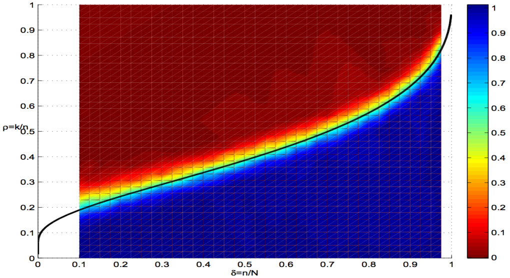
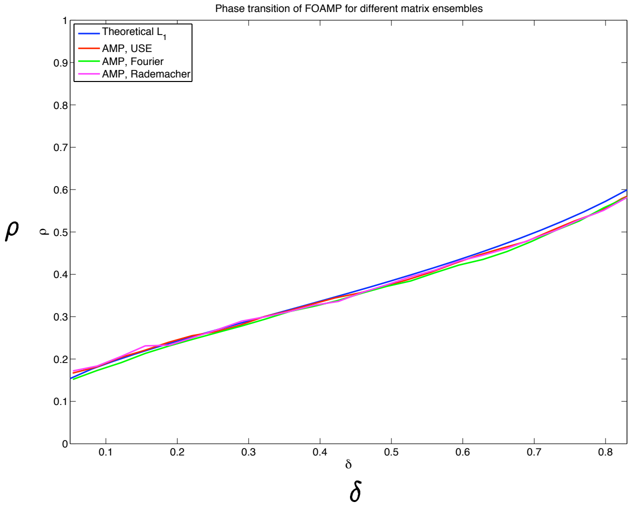


FIGURE 3. *Compressed Sensing from random Fourier measurements.* Shaded attribute: fraction of realizations in which $\ell_1$minimization (1.2) reconstructs an image accurate to within six digits. Horizontal axis: undersampling fraction $\delta = n/N$. Vertical axis: sparsity fraction $\rho = k/n$.

[Donoho, Tanner 2009]

Other examples



IID model

- A_{ij} i.i.d., $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$
- $n/N \to \delta \in (0, 1)$
- $\|x_0\|_0/n \to \rho \in (0, 1)$

Theorem (Bayati, Lelarge, Montanari, 2012)

Assume the IID model, with

- A_{ij} *subgaussian*
- $A_{ij} \stackrel{\rm d}{=} \widetilde{A}_{ij} + \varepsilon\, G_{ij}\,, \quad G_{ij} \sim {\rm N}(0,1)$

Then the $\ell_1 - -\ell_0$ phase transition is the same as for the Gaussian model. Namely

- *If $\rho < \rho_c(\delta)$, then $\mathbb{P}(\widehat{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_0) \to 1$,*
- *If $\rho > \rho_c(\delta)$, then $\mathbb{P}(\widehat{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_0) \to 0$.*

Proof outline

Two possible strategies

- **Lindeberg method:** Replace the A_{ij}'s one-by-one.

  [Korada, Montanari 2011]

  [weaker result]

- **Polynomial approximation**: Approximate function by polynomial

  [Bayati, Lelarge, Montanari 2012]

Two possible strategies

- **Lindeberg method:** Replace the A_{ij}'s one-by-one.

  [Korada, Montanari 2011]

  [weaker result]

- **Polynomial approximation**: Approximate function by polynomial

  [Bayati, Lelarge, Montanari 2012]

Polynomial approximation

- $\boldsymbol{X} = (X_1, \dots, X_m) \sim_{i.i.d.} P_X$
- $\boldsymbol{Y} = (Y_1, \dots, Y_m) \sim_{i.i.d.} P_Y$
- $\mathbb{E}\{X_i\} = \mathbb{E}\{Y_i\}$, $\mathbb{E}\{X_i^2\} = \mathbb{E}\{Y_i^2\}$
- $F : \mathbb{R}^m \to \mathbb{R}$

Want to prove

$$\mathbb{E}\{F(\boldsymbol{X})\} \approx \mathbb{E}\{F(\boldsymbol{Y})\}$$

Prove

$$F(x) \approx \sum_{\boldsymbol{k} \in \{0,1,2\}^m} F_{\boldsymbol{k}} x^{\boldsymbol{k}} \, ,$$

$$x^{\boldsymbol{k}} \equiv \prod_{i=1}^{m} x_i^{k_i} \, .$$

Polynomial approximation

- $\boldsymbol{X} = (X_1, \dots, X_m) \sim_{i.i.d.} P_X$
- $\boldsymbol{Y} = (Y_1, \dots, Y_m) \sim_{i.i.d.} P_Y$
- $\mathbb{E}\{X_i\} = \mathbb{E}\{Y_i\}$, $\mathbb{E}\{X_i^2\} = \mathbb{E}\{Y_i^2\}$
- $F : \mathbb{R}^m \to \mathbb{R}$

Want to prove

$$\mathbb{E}\{F(\boldsymbol{X})\} \approx \mathbb{E}\{F(\boldsymbol{Y})\}$$

Prove

$$F(\boldsymbol{x}) \approx \sum_{\boldsymbol{k} \in \{0,1,2\}^m} F_{\boldsymbol{k}} \boldsymbol{x}^{\boldsymbol{k}} \,,$$

$$\boldsymbol{x}^{\boldsymbol{k}} \equiv \prod_{i=1}^{m} x_i^{k_i} \,.$$

In our case

- $\boldsymbol{A} = (A_{ij})_{i \le n, j \le N}$

$$F(\boldsymbol{A}) = \mathbb{I}\Big(\boldsymbol{x}_0 = \arg\min\big\{\|\boldsymbol{x}\|_1 \ : \ \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_0) = 0\big\}\Big)$$

First simplification: Can fix

$$x_0 = (\underbrace{1, 1, \dots, 1}_{k \text{ non-zeros}}, 0, 0, \dots, 0)$$

How to approximate it?

In our case

- $\boldsymbol{A} = (A_{ij})_{i \le n, j \le N}$

$$F(\boldsymbol{A}) = \mathbb{I}\Big(\boldsymbol{x}_0 = \arg\min\big\{\|\boldsymbol{x}\|_1 \;:\; \boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}_0) = 0\big\}\Big)$$

First simplification: Can fix

$$\boldsymbol{x}_0 = (\underbrace{1, 1, \dots, 1}_{k \text{ non}-\text{zeros}}, 0, 0, \dots, 0)$$

How to approximate it?

In our case

- $\boldsymbol{A} = (A_{ij})_{i\le n, j\le N}$

$$F(\boldsymbol{A}) = \mathbb{I}\Big(\boldsymbol{x}_0 = \arg\min\big\{\|\boldsymbol{x}\|_1 \;:\; \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_0) = 0\big\}\Big)$$

First simplification: Can fix

$$\boldsymbol{x}_0 = (\underbrace{1,1,\dots,1}_{k\ \text{non}-\text{zeros}},0,0,\dots,0)$$

How to approximate it?

Will focus on

- If $\rho < \rho_c(\delta)$, then $\mathbb{P}(\widehat{\boldsymbol{x}}(\boldsymbol{y}, \boldsymbol{A}) = \boldsymbol{x}_0) \to 1$

Optimality condition

$$x_0 \in \arg\min \{ \|x\|_1 \quad : \quad A(x - x_0) = 0 \}$$

...if and only if there exists

$$v \in \partial \|x_0\|_1 \; : \qquad v \in \mathrm{Image}(A^{\mathsf{T}})$$

Subdifferential

$$\partial \|x_0\|_1 \equiv \Big\{ v \in \mathbb{R}^N : v_i = \mathrm{sign}(x_{0,i}),\, i \in \mathrm{supp}(x_0), \\ |v_i| \le 1,\, i \notin \mathrm{supp}(x_0) \Big\}$$

Optimality condition

$$x_0 \in \arg\min\{\|x\|_1 \quad : \quad A(x - x_0) = 0\}$$

...if and only if there exists

$$v \in \partial\|x_0\|_1 \; : \qquad v \in \mathrm{Image}(A^{\mathsf{T}})$$

Subdifferential

$$\partial\|x_0\|_1 \equiv \Big\{v \in \mathbb{R}^N : v_i = \mathrm{sign}(x_{0,i}),\, i \in \mathrm{supp}(x_0),\\ |v_i| \le 1,\, i \notin \mathrm{supp}(x_0)\Big\}$$

Quantitative version

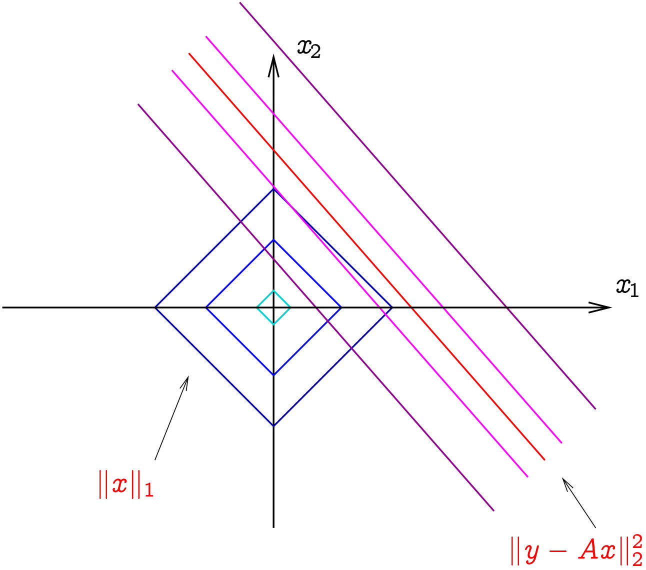
Lemma

Assume

1. *There exists* $v \in \partial\|x_0\|_1$ *and* $z \in \mathbb{R}^n$ *with* $v = A^{\mathsf{T}} z + w$ *and* $\|w\|_2 \le \sqrt{n}\,\varepsilon$, *with* $\varepsilon \le \varepsilon_0$.
2. *For* $c \in (0,1)$, *let* $S(c) \equiv \{i \in [n] : |v_i| \ge 1 - c\}$. *Then, for any* $S' \subseteq [n]$, $|S'| \le c_1 n$, $\sigma_{\min}(A_{S(c_1)\cup S'}) \ge c_2$.
3. $c_3^{-1} \le \sigma_{\max}(A)^2 \le c_3$.

Then $x_0 = \arg\min\{\|x\|_1 \;\; : \;\; A(x - x_0) = 0\}$

Where is the problem



Plan

- Construct an approximate subgradient $\boldsymbol{v}$
- Check properties $1, 2, 3$

Universality

$$\boldsymbol{v} \approx \sum_{\boldsymbol{k} \in \{0,1,2\}^{n \times N}} \boldsymbol{v}_{\boldsymbol{k}} \, \boldsymbol{A}^{\boldsymbol{k}}$$

Idea: Use an iterative algorithm to construct $\boldsymbol{v}$

‘Approximate Message Passing’ (AMP)

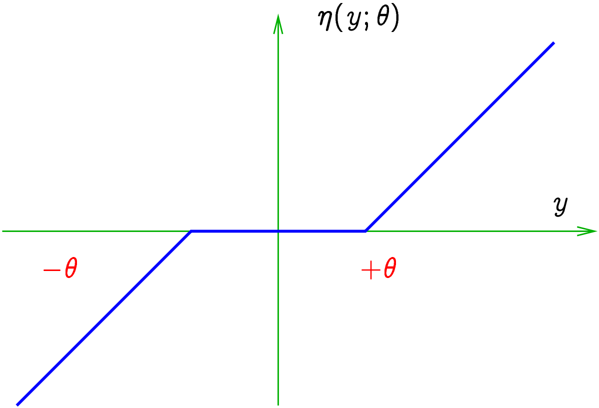
Time index $t \in \{0, 1, 2, \dots, T\}$

$$\begin{aligned} \boldsymbol{x}^{t+1} &= \eta(\boldsymbol{x}^t + \boldsymbol{A}^{\mathsf{T}} \boldsymbol{z}^t; \theta_t) \\ \boldsymbol{z}^t &= \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \mathsf{b}_t \boldsymbol{z}^{t-1} \end{aligned}$$

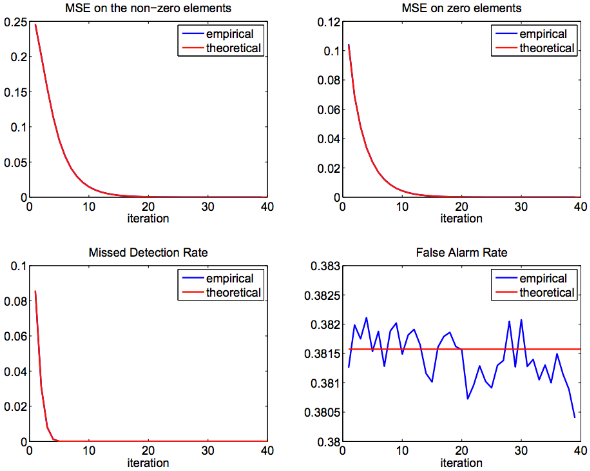
- Algorithm parameters: $\theta_t, \mathsf{b}_t \in \mathbb{R}$
- Soft thresholding function: $\eta(\,\cdot\,; \theta) : \mathbb{R} \to \mathbb{R}$

[Donoho, Maleki, Montanari, 2009]

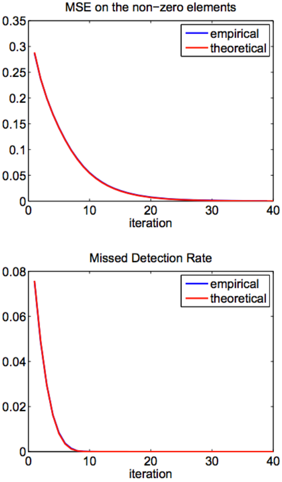
η



Algorithm evolution: $\delta = 0.2$, $\rho = 0.3$



Algorithm evolution: $\delta = 0.15$, $\rho = 0.3$



Asymptotic analysis

T fixed, $N, n \to \infty$

Dynamics + Randomness + High dimension

'State evolution': T fixed, $N, n \to \infty$

Time index $t \in \{0, 1, 2, \dots, T\}$

$$x^{t+1} = \eta(y^t; \theta_t)$$
$$y^t = x^t + A^T z^t$$
$$z^t = y - Ax^t + \mathsf{b}_t z^{t-1}$$

Theorem (Bayati, Montanari, 2010)

For A Gaussian (letting $X_0 \sim p_{X_0}$ independent of $Z \sim \mathrm{N}(0,1)$)

$$y^t \approx \mathrm{N}(x_0, \sigma_t^2 \mathrm{I}_{N\times N}),$$
$$\sigma_{t+1}^2 = \mathbb{E}\Big\{\big[\eta(X_0 + \sigma_t Z; \theta_t) - X_0\big]^2\Big\}$$

Construction of the subgradient

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{x}^t + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}^t; \theta_t)$$
$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \mathsf{b}_t \boldsymbol{z}^{t-1}$$

Note: $|\eta(x;\theta) - x| \le \theta$

$$v_i^t = \begin{cases} \text{sign}(x_{0,i}) & \text{if } i \in S\,, \\ \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^{\mathsf{T}}z^{t-1} - \eta(x^{t-1} + A^{\mathsf{T}}z^{t-1}; \theta_{t-1})\big)_i & \text{otherwise,} \end{cases}$$

Construction of the subgradient

$$x^{t+1} = \eta(x^t + A^\top z^t; \theta_t)$$
$$z^t = y - Ax^t + \mathsf{b}_t z^{t-1}$$

Note: $|\eta(x;\theta) - x| \le \theta$

$$v_i^t = \begin{cases} \text{sign}(x_{0,i}) & \text{if } i \in S\,, \\ \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^\top z^{t-1} - \eta(x^{t-1} + A^\top z^{t-1}; \theta_{t-1})\big)_i & \text{otherwise,} \end{cases}$$

Construction of the subgradient

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{x}^t + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}^t; \theta_t)$$
$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \mathsf{b}_t \boldsymbol{z}^{t-1}$$

Note: $|\eta(x;\theta) - x| \le \theta$

$$v_i^t = \begin{cases} \text{sign}(x_{0,i}) & \text{if } i \in S\,, \\ \frac{1}{\theta_{t-1}}\big(x^{t-1} + A^{\mathsf{T}}z^{t-1} - \eta(x^{t-1} + A^{\mathsf{T}}z^{t-1}; \theta_{t-1})\big)_i & \text{otherwise,} \end{cases}$$

Proof strategy

- Analyze AMP for $\boldsymbol{A}$ non-gaussian
  - Approximate η by a polynomial
  - $\Rightarrow \boldsymbol{x}^t$, $\boldsymbol{z}^t$, $\boldsymbol{v}^t$ =Polynomials $(\{A_{ij}\})$
  - Use moment method
  - $\Rightarrow$ state evolution

- Check condition for optimality.

Proof strategy

- Analyze AMP for $\boldsymbol{A}$ non-gaussian
  - Approximate η by a polynomial
  - $\Rightarrow \boldsymbol{x}^t$, $\boldsymbol{z}^t$, $\boldsymbol{v}^t$ =Polynomials $(\{A_{ij}\})$
  - Use moment method
  - $\Rightarrow$ state evolution
- Check condition for optimality.

Proof strategy

- Analyze AMP for $\boldsymbol{A}$ non-gaussian
  - Approximate η by a polynomial
  - $\Rightarrow$ $\boldsymbol{x}^t$, $\boldsymbol{z}^t$, $\boldsymbol{v}^t$ =Polynomials ($\{A_{ij}\}$)
  - Use moment method
  - $\Rightarrow$ state evolution
- Check condition for optimality.

An interesting phenomenon

Let us compare

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{x}^t + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}^t;\theta_t)$$
$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t + \mathsf{b}_t\boldsymbol{z}^{t-1}$$

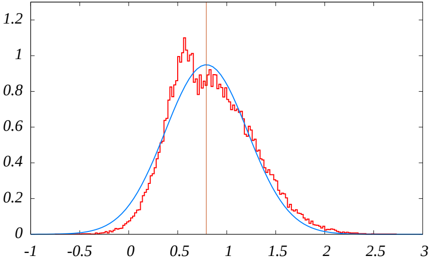
and

$$\boldsymbol{x}^{t+1} = \eta(\boldsymbol{x}^t + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}^t;\theta_t)$$
$$\boldsymbol{z}^t = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^t$$

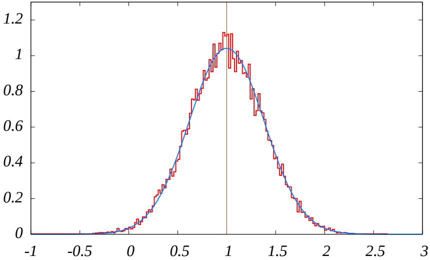
'Onsager reaction term' (spin glass theory)

What's the big deal with $-\mathrm{b}_t z^{t-1}$?

Distribution of $(x^t + A^T z^t)_i$ conditional on $x_{0,i} = 1$



without



with

Let us look at a simpler example (symmetric)

- $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- $A_{ii} = 0$, $\{A_{ij}\}_{i<j}$ iid, $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$
- $f : \mathbb{R} \to \mathbb{R}$, $f(\boldsymbol{x}) = (f(x_1), f(x_2), \dots, f(x_n))$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A} f(\boldsymbol{x}^t) - \mathsf{b}_t f(\boldsymbol{x}^{t-1})$$

$$\mathsf{b}_t = \frac{1}{n} \sum_{i=1}^{n} f'(x_i^t)$$

Special case: TAP equatios for mean field spin glasses

[Bolthausen 2011]

Let us look at a simpler example (symmetric)

- $\boldsymbol{A} \in \mathbb{R}^{n\times n}$
- $A_{ii} = 0$, $\{A_{ij}\}_{i<j}$ iid, $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$
- $f : \mathbb{R} \to \mathbb{R}$, $f(\boldsymbol{x}) = (f(x_1), f(x_2), \dots, f(x_n))$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A} f(\boldsymbol{x}^t) - \mathsf{b}_t f(\boldsymbol{x}^{t-1})$$

$$\mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{n} f'(x_i^t)$$

Special case: TAP equatios for mean field spin glasses

[Bolthausen 2011]

Let us look at a simpler example (symmetric)

- $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- $A_{ii} = 0$, $\{A_{ij}\}_{i<j}$ iid, $\mathbb{E}\{A_{ij}\} = 0$, $\mathbb{E}\{A_{ij}^2\} = 1/n$
- $f : \mathbb{R} \to \mathbb{R}$, $f(\boldsymbol{x}) = (f(x_1), f(x_2), \dots, f(x_n))$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A} f(\boldsymbol{x}^t) - \mathsf{b}_t f(\boldsymbol{x}^{t-1})$$

$$\mathsf{b}_t = \frac{1}{n} \sum_{i=1}^{n} f'(x_i^t)$$

Special case: TAP equatios for mean field spin glasses

[Bolthausen 2011]

Symmetric case

$$x^{t+1} = Af(x^t) - \mathsf{b}_t f(x^{t-1})$$

$$\mathsf{b}_t = \frac{1}{n} \sum_{i=1}^{n} f'(x_i^t)$$

Theorem (Bayati, Montanari, 2010)

If $A \sim \text{GOE}(n)$, $f \in \text{Lip}(\mathbb{R})$, *then (here* $Z \sim \text{N}(0,1)$*)*

$$x^t \approx \text{N}(0, \sigma_t^2 \text{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Theorem (Bayati, Lelarge, Montanari, 2012)

If $A \sim \text{Wigner}(n)$, $f \in \text{Poly}(\mathbb{R})$, *then (here* $Z \sim \text{N}(0,1)$*)*

$$x^t \approx \text{N}(0, \sigma_t^2 \text{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Symmetric case

$$x^{t+1} = Af(x^t) - \mathsf{b}_t f(x^{t-1})$$

$$\mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{n} f'(x_i^t)$$

Theorem (Bayati, Montanari, 2010)

If $A \sim \mathsf{GOE}(n)$, $f \in \mathsf{Lip}(\mathbb{R})$, *then (here* $Z \sim \mathrm{N}(0,1)$*)*

$$x^t \approx \mathrm{N}(0, \sigma_t^2 \mathrm{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Theorem (Bayati, Lelarge, Montanari, 2012)

If $A \sim \mathsf{Wigner}(n)$, $f \in \mathsf{Poly}(\mathbb{R})$, *then (here* $Z \sim \mathrm{N}(0,1)$*)*

$$x^t \approx \mathrm{N}(0, \sigma_t^2 \mathrm{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Symmetric case

$$x^{t+1} = Af(x^t) - \mathsf{b}_t f(x^{t-1})$$
$$\mathsf{b}_t = \frac{1}{n}\sum_{i=1}^{n} f'(x_i^t)$$

Theorem (Bayati, Montanari, 2010)

If $A \sim \mathsf{GOE}(n)$, $f \in \mathsf{Lip}(\mathbb{R})$, *then (here* $Z \sim \mathrm{N}(0,1)$*)*

$$x^t \approx \mathrm{N}(0, \sigma_t^2 \mathrm{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Theorem (Bayati, Lelarge, Montanari, 2012)

If $A \sim \mathsf{Wigner}(n)$, $f \in \mathsf{Poly}(\mathbb{R})$, *then (here* $Z \sim \mathrm{N}(0,1)$*)*

$$x^t \approx \mathrm{N}(0, \sigma_t^2 \mathrm{I}_{n\times n}), \qquad \sigma_{t+1}^2 = \mathbb{E}\{f(\sigma_t Z)^2\}.$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

Corollary

For al $t \geq 1$ fixed, as $n \to \infty$,

$$x^t \approx \mathrm{N}(0, \mathrm{I}_{n\times n}) .$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

Corollary

For al $t \ge 1$ fixed, as $n \to \infty$,

$$\boldsymbol{x}^t \approx \mathrm{N}(0, \mathrm{I}_{n\times n})\,.$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$x_i^1 = \sum_{j \in [n] \setminus i} A_{ij} \qquad \Rightarrow \mathrm{N}(0,1)$$

$$x^2 = \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus j_1} A_{ij_1} A_{j_1,j_2} - 1$$

$$= \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1,j_2} + \sum_{j_1 \in [n] \setminus i} A_{ij_1} A_{j_1 i} - 1$$

$$\approx \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1,j_2} \qquad \Rightarrow \mathrm{N}(0,1)$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$x_i^1 = \sum_{j \in [n] \setminus i} A_{ij} \qquad \Rightarrow \mathrm{N}(0,1)$$

$$x^2 = \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus j_1} A_{ij_1} A_{j_1, j_2} - 1$$

$$= \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} + \sum_{j_1 \in [n] \setminus i} A_{ij_1} A_{j_1 i} - 1$$

$$\approx \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} \qquad \Rightarrow \mathrm{N}(0,1)$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$x_i^1 = \sum_{j\in[n]\setminus i} A_{ij} \qquad \Rightarrow \mathrm{N}(0,1)$$

$$\boldsymbol{x}^2 = \sum_{j_1\in[n]\setminus i, j_2\in[n]\setminus j_1} A_{ij_1}A_{j_1,j_2} - 1$$

$$= \sum_{j_1\in[n]\setminus i, j_2\in[n]\setminus\{j_1,i\}} A_{ij_1}A_{j_1,j_2} + \sum_{j_1\in[n]\setminus i} A_{ij_1}A_{j_1 i} - 1$$

$$\approx \sum_{j_1\in[n]\setminus i, j_2\in[n]\setminus\{j_1,i\}} A_{ij_1}A_{j_1,j_2} \qquad \Rightarrow \mathrm{N}(0,1)$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$x_i^1 = \sum_{j \in [n] \setminus i} A_{ij} \qquad \Rightarrow \mathrm{N}(0,1)$$

$$\boldsymbol{x}^2 = \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus j_1} A_{ij_1} A_{j_1, j_2} - 1$$

$$= \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} + \sum_{j_1 \in [n] \setminus i} A_{ij_1} A_{j_1 i} - 1$$

$$\approx \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} \qquad \Rightarrow \mathrm{N}(0,1)$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$\begin{aligned}
x_i^1 &= \sum_{j \in [n] \setminus i} A_{ij} && \Rightarrow \mathrm{N}(0,1) \\
\boldsymbol{x}^2 &= \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus j_1} A_{ij_1} A_{j_1, j_2} - 1 \\
&= \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} + \sum_{j_1 \in [n] \setminus i} A_{ij_1} A_{j_1 i} - 1 \\
&\approx \sum_{j_1 \in [n] \setminus i, j_2 \in [n] \setminus \{j_1, i\}} A_{ij_1} A_{j_1, j_2} && \Rightarrow \mathrm{N}(0,1)
\end{aligned}$$

Even simpler: $f(x) = x$

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}\boldsymbol{x}^t - \boldsymbol{x}^{t-1}$$

$$x_i^t \approx \sum_{\gamma \in \mathsf{NRW}(i,t)} \prod_{k=1}^{t-1} A_{\gamma(k),\gamma(k+1)} \quad \Rightarrow \mathrm{N}(0,1)$$

$$\mathsf{NRW}(i;t) \equiv \Big\{\gamma : \{0,\dots,t\} \to [n] : \quad \gamma(0) = i,\ \ \gamma(k-1) \neq \gamma(k+1)\ \forall k\Big\}$$

[$\Rightarrow$ by moment method]

General $f(x) =$poly(x)

Say $\boldsymbol{x}^0 = \mathbf{1}$, $\boldsymbol{x}^{-1} = 0$

$$\boldsymbol{x}^{t+1} = \boldsymbol{A} f(\boldsymbol{x}^t) - \mathsf{b}_t f(\boldsymbol{x}^{t-1})$$

$$x_i^t \approx \sum_{T \in \text{Trees}} c_f(T;t) \sum_{\gamma \in \mathsf{NRT}(i;t)} \prod_{e \in E(T)} A_{\gamma(e)} \quad \Rightarrow \mathrm{N}(0, \sigma_t^2)$$

$$\mathsf{NRT}(i;t) \equiv \Big\{ \gamma : V(T) \to [n] : \quad \gamma(\text{root}) = i, \\ \gamma(\text{desc}(v)) \neq \gamma(\text{parent}(v)) \; \forall v \Big\}$$

Conclusion

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,...)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,...)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,. . .)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,. . .)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,...)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!

Open problems

- Noisy compressed sensing
- Other matrix ensembles
  (i.i.d. entries, i.i.d. rows, Fourier sections,...)
- Other convex optimization problems
- Universality of AMP (non-polynomial functions)

Thanks!