Phase transitions in low-rank matrix estimation

Florent Krzakala

http://krzakala.github.io/LowRAMP/

arXiv:1503.00338 & 1507.03857

cea

Thibault Lesieur
Lenka Zdeborová

Caterina Debacco
Cris Moore
Jess Banks

SANTA FE INSTITUTE

Jiaming Xu

EPFL
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Jean Barbier
Nicolas Macris
Mohamad Dia

SIMONS INSTITUTE for the Theory of Computing

Low-rank matrix estimation

$$W = \frac{1}{\sqrt{n}} X X^T \qquad X \in \mathbb{R}^{n\times r} \qquad \text{unknown}$$

or

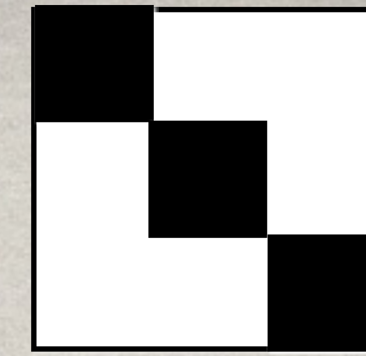
$$W = \frac{1}{\sqrt{n}} U V^T \qquad U \in \mathbb{R}^{m\times r}, V \in \mathbb{R}^{n\times r} \qquad \text{unknown}$$

Matrix W has low (finite) rank $r \ll n, m$

W is observed element-wise trough a noisy channel: $P_{\text{out}}(y_{ij}|w_{ij})$

- Goal: Estimate unknown X (or U & V) from known Y.

Some examples...

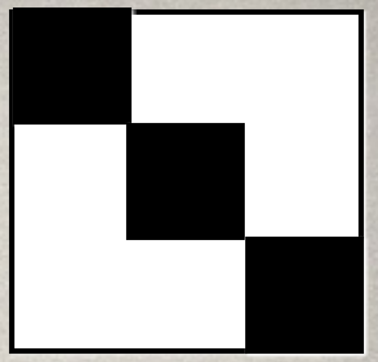


$$x_i^T = (0, \ldots, 0, 1, 0, \ldots, 0) \qquad w_{ij} = x_i^T x_j / \sqrt{n}$$

r-dimensional variable (r=rank)

(Goal: Estimate unknown X from known Y.)

Some examples...



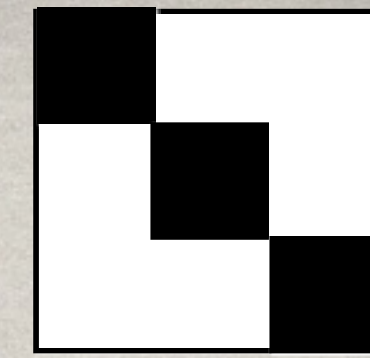
$$x_i^T = (0, \ldots, 0, 1, 0, \ldots, 0) \qquad w_{ij} = x_i^T x_j / \sqrt{n}$$

- Additive white Gaussian noise (sub-matrix localization)

$$P_{\text{out}}(y|w) = \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{(y-w)^2}{2\Delta}}$$

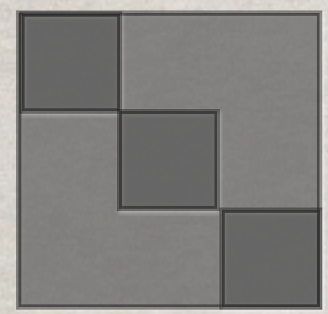
(Goal: Estimate unknown X from known Y.)

Some examples...



$$x_i^T = (0, \dots, 0, 1, 0, \dots, 0) \qquad w_{ij} = x_i^T x_j / \sqrt{n}$$

- Additive white Gaussian noise (sub-matrix localization)



$$P_{\text{out}}(y|w) = \frac{1}{\sqrt{2\pi\Delta}} e^{-\frac{(y-w)^2}{2\Delta}}$$

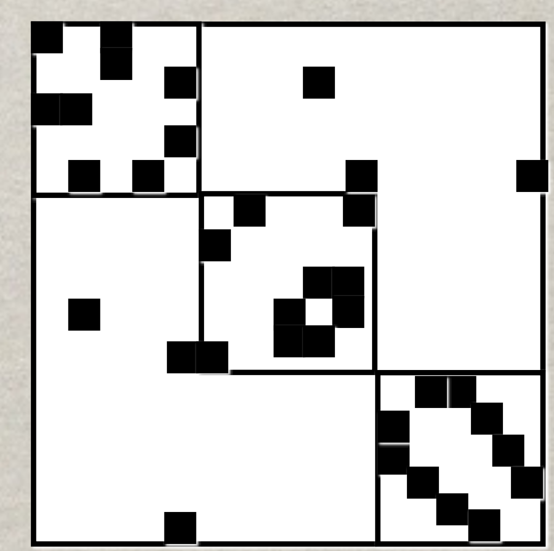
- Dense stochastic block model (community detection)

$$P_{\text{out}}(y_{ij} = 1 | w_{ij}) = p_{\text{out}} + \mu w_{ij}$$

$$P_{\text{out}}(y_{ij} = 0 | w_{ij}) = 1 - p_{\text{out}} - \mu w_{ij}$$

$$p_{\text{in}} = p_{\text{out}} + \mu / \sqrt{n}$$

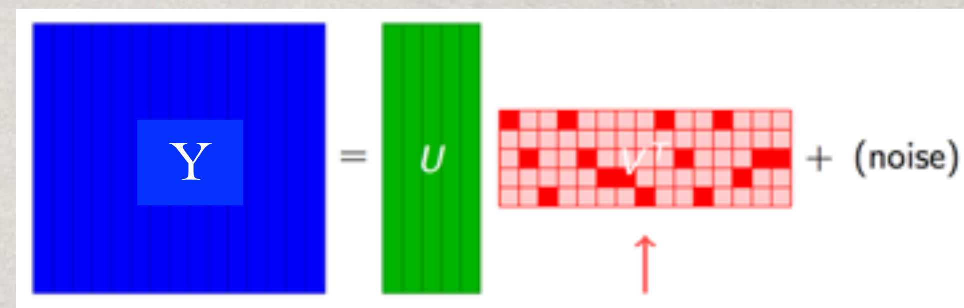
$$\Delta = \frac{p_{\text{out}}(1 - p_{\text{out}})}{\mu^2} .$$



(Goal: Estimate unknown X from known Y.)

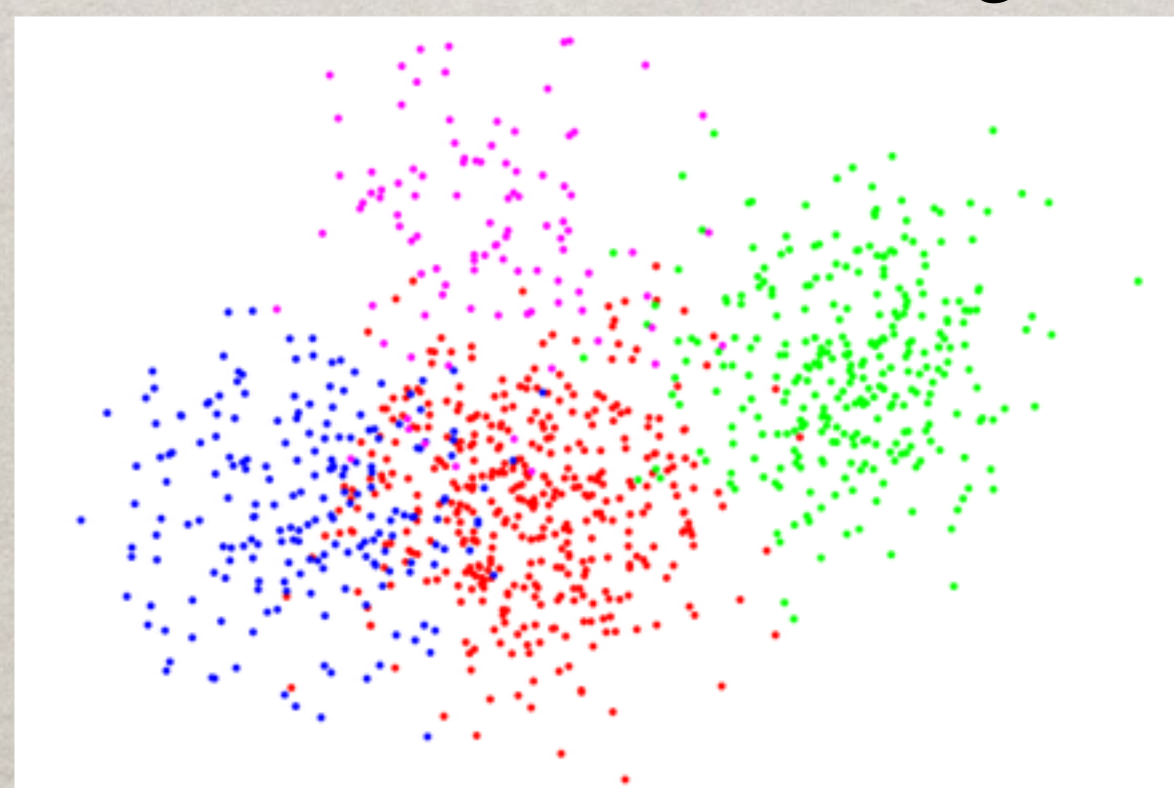
Some examples...

$$u_i^T = (v_i^1, \ldots, v_i^R) \in \mathbb{R}$$
$$v_i^T = (0, \ldots, 0, 1, 0, \ldots, 0)$$



$$W = \frac{1}{\sqrt{n}} UV^T + \mathcal{N}(0, \sigma^2)$$

- Additive white Gaussian noise (clustering mixture of Gaussians)



(Goal: Estimate unknown U and V from known Y.)

Even more examples...

- Sparse PCA, robust PCA
- Collaborative filtering (low rank matrix completion)
- 1-bit Collaborative filtering (like/unlike)
- Bi-clustering
- Planted clique (cf. Andrea Montanari yesterday)
- etc...

QUESTIONS

Many interesting problems can be formulated this way

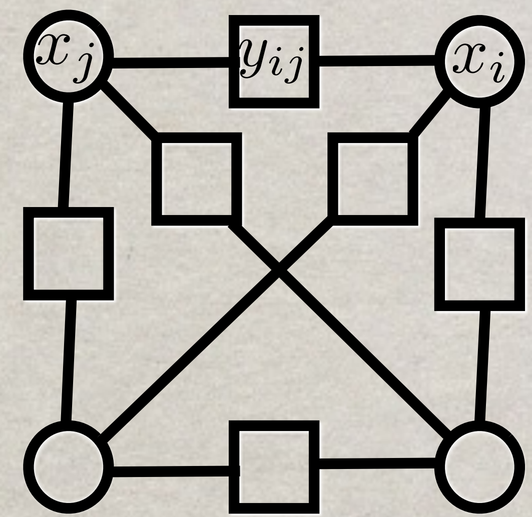
- **Q1**: When is it possible to perform a good factorization?
- **Q2:** When is it algorithmically tractable ?
- **Q3:** How good are spectral methods (main tool) ?

- ***Yesterday***: Andrea taught us how to analyze AMP for such problems with the state evolution approach.
- ***Today***: We continue in this direction and answer these questions in a probabilistic setting, with instances randomly generated from a model.

Probabilistic setting

Assume X is generated from $P(X) = \prod_{i=1}^{n} P_X(x_i)$

The posterior distribution reads $P(X|Y) = \frac{1}{Z(Y)} \prod_{i=1}^{n} P_X(x_i) \prod_{i<j} P_{\rm out}\left(y_{ij} \middle| \frac{x_i^T x_j}{\sqrt{n}}\right)$



Graphical model where y_{ij} are pair-wise observations of variables

MMSE estimator (minimal error)
Marginals probability of the posterior

Exactly Solvable ?

- When P_X and $P_{\rm out}$ known, $n \to \infty, r = O(1)$
- Approximate message passing = dense-factor-graph simplifications of belief propagation, hopefully ***asymptotically exact*** marginals of the posterior distribution.
- Exact analysis possible with ***statistical-physics*** style methods
- Many rigorous proofs possible when one works a bit ***harder***

OUTLINE

1) Message passing, State evolution, Mutual information

2) Universality property

3) Main results

4) Sketch of proof

OUTLINE

1) Message passing, State evolution, Mutual information

2) Universality property

3) Main results

4) Sketch of proof

AMP FOR GAUSSIAN ADDITIVE CHANNEL

$$w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{\textbf{AWGN}} y_{ij} = w_{ij} + \sqrt{\Delta}\xi$$

$$\mathbf{B}^t = \frac{1}{\sqrt{n}}\mathbf{S}\mathbf{a}^t - \frac{1}{\Delta}\left(\frac{1}{n}\sum_i v_i^t\right)\mathbf{a}^{t-1}$$

$$A^t = \frac{1}{n\Delta}\mathbf{a}^t \cdot \mathbf{a}^t$$

$$\mathbf{S} = \frac{\mathbf{Y}}{\Delta}$$

Mean and variance of the marginals:

$$a_i^{t+1} = f(A^t, B_i^t)$$

$$v_i^{t+1} = \left(\frac{\partial f}{\partial B}\right)(A^t, B_i^t)$$

- First written for rank r=1 in Rangan, Fletcher'12

Dependence on the prior only via a "thresholding function" f(A,B) given by the expectation of:

$$P(x) = \frac{1}{\mathcal{Z}(A,B)} P_X(x) \exp\left(B^\top x - \frac{x^\top A x}{2}\right)$$

AMP FOR GAUSSIAN ADDITIVE CHANNEL

$$w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{\textbf{AWGN}} y_{ij} = w_{ij} + \sqrt{\Delta}\xi$$

$$\mathbf{B}^t = \frac{1}{\sqrt{n}} \mathbf{S} \mathbf{a}^t - \frac{1}{\Delta} \left(\frac{1}{n} \sum_i v_i^t \right) \mathbf{a}^{t-1}$$

$$A^t = \frac{1}{n\Delta} \mathbf{a}^t \cdot \mathbf{a}^t$$

$$\mathbf{S} = \frac{\mathbf{Y}}{\Delta}$$

Mean and variance of the marginals:

$$a_i^{t+1} = f(A^t, B_i^t)$$

$$v_i^{t+1} = \left(\frac{\partial f}{\partial B} \right) (A^t, B_i^t)$$

- First written for rank r=1 in Rangan, Fletcher'12

Note: for x=±1, these are nothing but the TAP equations for the Ising Sherrington-Kirkpatrick model (on the Nishimori line)

$$a_i^{t+1} = \tanh(B_i^t) \qquad v_i^{t+1} = 1 - (a_i^{t+1})^2$$

State evolution

Single letter characterization of the AMP

$$M^t = \frac{1}{n}\hat{\mathbf{x}}_{\rm AMP}\cdot\mathbf{x}$$

$$M^{t+1} = \mathbb{E}_{x,\xi}\left[f\left(\frac{M^t}{\Delta},\frac{M^t}{\Delta}x+\sqrt{\frac{M^t}{\Delta}}\xi\right)x\right]$$

Depends on the channel only though its Fisher information.
1-parameter family of channels having the same MMSE.

cf. Yesterday's talk: rigorously proven in
Montanari, Bayati '10 - Montanari, Deshpande '14

State evolution

Single letter characterization of the AMP

$$M^t = \frac{1}{n}\hat{\mathbf{x}}_{\mathrm{AMP}} \cdot \mathbf{x}$$

$$M^{t+1} = \mathbb{E}_{x,\xi}\left[f\left(\frac{M^t}{\Delta}, \frac{M^t}{\Delta}x + \sqrt{\frac{M^t}{\Delta}}\xi\right)x\right]$$

Note: for x=±1, these are nothing but the replica symmetric equations for the Ising Sherrington-Kirkpatrick model (on the Nishimori line)

$$M^{t+1} = \int dx\mathcal{D}\xi \tanh\left(\frac{M^t}{\Delta}x + \sqrt{\frac{M^t}{\Delta}}\xi\right)$$

Replica Mutual information

Most quantities of interest can be computed from the Mutual Information
(free energy for physicist)

$$I(X;Y) = \int dx dy P(x,y) \log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

The replica methods predicts an asymptotic formula for i = I/n:

$$i_{\rm RS}(m) = \frac{m^2 + \left[\mathbb{E}_x(x^2)\right]^2}{4\Delta} - \mathbb{E}_{x,z}\left[\mathcal{J}\left(\frac{m}{\Delta}, \frac{mx}{\Delta} + \sqrt{\frac{m}{\Delta}}z\right)\right]$$

with $\mathcal{J}(A,B) = \log \int e^{Bx - Ax^2/2} p(x) de$

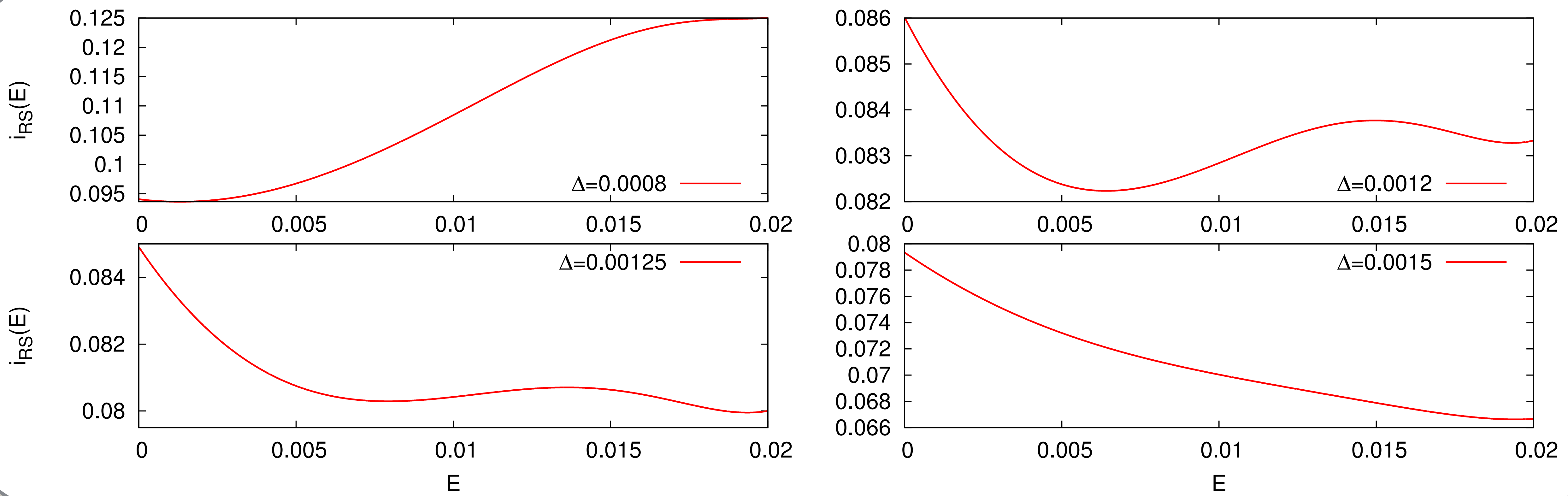
FACT: The state evolution recursion for AMP is a fixed point of $i_{\rm RS}(m)$

CONJECTURE: $\lim_{n\to\infty} \frac{I(X;Y)}{n} = \min_m i_{rs}(m)$

Replica Mutual information

$$i_{\mathrm{RS}}(m) = \frac{m^2 + \left[\mathbb{E}_x(x^2)\right]^2}{4\Delta} - \mathbb{E}_{x,z}\left[\mathcal{J}\left(\frac{m}{\Delta}, \frac{mx}{\Delta} + \sqrt{\frac{m}{\Delta}}z\right)\right] \qquad \mathcal{J}(A,B) = \log \int e^{Bx - Ax^2/2} p(x) de$$

Good for AMP | **Bad** for AMP



Good for AMP | **Good** for AMP

$$E = MSE^{vector}_{\mathrm{AMP}} = \mathbb{E}[x^2] - m$$

Replica Mutual information

Can we prove $\lim_{n\to\infty} \frac{I(X;Y)}{n} = \min_m i_{rs}(m)$? **yes!**

$$i_{\rm RS}(m) = \frac{m^2 + \left[\mathbb{E}_x(x^2)\right]^2}{4\Delta} - \mathbb{E}_{x,z}\left[\mathcal{J}\left(\frac{m}{\Delta}, \frac{mx}{\Delta} + \sqrt{\frac{m}{\Delta}}z\right)\right]$$

with $\mathcal{J}(A,B) = \log \int e^{Bx - Ax^2/2} p(x) de$

* Proven for "not-too-sparse" PCA (Montanari & Deshpande '14) and for symmetric community detection (Montanari, Abbe & Deshpande '16).
* Proven for the planted SK model by Korada & Macris '10

FK, Xu & Zdeborová '16 : Upper Bound $\frac{I}{n} \le i_{\rm rs}(m)$

(Guerra Interpolation)

Barbier, Dia, Macris, FK & Zdeborová '16: $\lim_{n\to\infty} \frac{I}{n} \ge \min_m \; i_{\rm rs}(m)$

(Spatial coupling+thermodynamic integration/I-MMSE)

FORMULAS FOR UV^T

AMP

$$B_u^{t+1} = \sqrt{\frac{1}{\Delta n}} X f_v(A_v^t, B_v^t) - \frac{1}{\Delta}\langle f_v^{t\prime}(A_v^t, B_v^t)\rangle f_u(A_u^{t-1}, B_u^{t-1})$$

$$A_u^{t+1} = \frac{1}{\Delta n} f_v(A_v^t, B_v^t) f_v(A_v^t, B_v^t)^T$$

$$B_v^{t+1} = \sqrt{\frac{1}{\Delta n}} X^T f_s(A_v^t, B_v^t) - \frac{\alpha}{\Delta}\langle f_s{}'(A_s^t, B_s^t)\rangle f_v(A_v^{t-1}, B_v^{t-1})$$

$$A_v^{t+1} = \frac{1}{\Delta n} f_u(A_u^t B_u^t) f_u(A_u^t, B_u^t)^T$$

State evolution

$$M_u^t = \frac{1}{n}\hat{\mathbf{u}}_{\mathrm{AMP}} \cdot \mathbf{u} \qquad M_u^{t+1} = \mathbb{E}_{u,\xi}\left[f_u\left(\frac{M_v^t}{\Delta}, \frac{M_v^t}{\Delta} x + \sqrt{\frac{M_v^t}{\Delta}}\xi\right) u\right]$$

$$M_v^t = \frac{1}{m}\hat{\mathbf{v}}_{\mathrm{AMP}} \cdot \mathbf{v} \qquad M_v^{t+1} = \mathbb{E}_{v,\xi}\left[f_v\left(\alpha\frac{M_u^t}{\Delta}, \alpha\frac{M_u^t}{\Delta} x + \sqrt{\alpha\frac{M_u^t}{\Delta}}\xi\right) v\right]$$

Mutual information

$$i_{\mathrm{RS}}(m_u, m_v) = \alpha\frac{m_u m_v + \left[\mathbb{E}(u^2)\right]\left[\mathbb{E}(v^2)\right]}{2\Delta} - \mathbb{E}_{u,z}\left[\mathcal{J}_u\left(\frac{m_v}{\Delta}, \frac{m_v u}{\Delta} + \sqrt{\frac{m_v}{\Delta}}z\right)\right] - \alpha\mathbb{E}_{v,z}\left[\mathcal{J}_v\left(\frac{m_u}{\Delta}, \frac{m_u v}{\Delta} + \sqrt{\frac{m_u}{\Delta}}z\right)\right]$$

OUTLINE

1) **Message passing, State evolution, Mutual information**

2) **Universality property**

3) **Main results**

4) **Sketch of proof**

WHAT ABOUT OTHER CHANNELS?

$$w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{P_{\rm out}(y_{ij}|w_{ij})} y_{ij}$$

CHANNEL UNIVERSALITY

$$w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{P_{\text{out}}(y_{ij}|w_{ij})} y_{ij}$$

Dependence on the channel **only** via:

$$S_{ij} \equiv \frac{\partial \log P_{\text{out}}(y_{ij}|w)}{\partial w}\Big|_{y_{ij},0}$$

Fisher-score matrix

$$\frac{1}{\Delta} \equiv \mathbb{E}_{P_{\text{out}}(y|w=0)} \left[\left(\frac{\partial \log P_{\text{out}}(y|w)}{\partial w}\Big|_{y,0} \right)^2 \right]$$

Fisher information

Effective Gaussian channel with y=ΔS

$$w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{\textbf{AWGN}} y_{ij} = w_{ij} + \sqrt{\Delta}\xi$$

CHANNEL UNIVERSALITY

A physicist argument

small quantity → $W = \frac{1}{\sqrt{n}} X X^T$

$$P_{\text{out}}(Y_{ij}|W_{ij}) = e^{\log P_{\text{out}}(Y_{ij}|W_{ij})} = P_{\text{out}}(Y_{ij}|0) e^{W_{ij}S_{ij} + \frac{1}{2}W_{ij}^2 S'_{ij} + O(n^{-3/2})}$$

n^2 terms → $P_{\text{out}}(Y|W) = P_{\text{out}}(Y|0) e^{\sum_{i\leq j}(W_{ij}S_{ij} + \frac{1}{2}\underbrace{W_{ij}^2 S'_{ij}}) + O(\sqrt{n})}$

concentration

$$P_{\text{out}}(Y|W) \approx P_{\text{out}}(Y|0) e^{\sum_{i\leq j}(W_{ij}S_{ij} - \frac{1}{2\Delta}W_{ij}^2) + O(\sqrt{n})}$$

Effective Gaussian posterior probability

$$P(W|Y) \propto P(W) e^{-\frac{1}{2\Delta}\sum_{i\leq j}(\Delta S_{ij} - W_{ij})^2 + O(\sqrt{n})}.$$

CHANNEL UNIVERSALITY

Theorem I.1 (Channel Universality). *Assume model (1) with a prior $p(x)$ having a finite support, and the output channel $P_{\rm out}(y|w)$ such that at $w=0$, $\log P_{\rm out}(y|w)$ is thrice differentiable with bounded second and third derivatives and $\mathbb{E}_{P_{\rm out}(y|0)}[|\partial_w \log P_{\rm out}(y|w)|_{w=0}|^3] = O(1)$. Then the mutual information per variable satisfies*

$$I(\boldsymbol{W};\boldsymbol{Y}) = I(\boldsymbol{W};\boldsymbol{W}+\sqrt{\Delta}\,\boldsymbol{\xi}) + O(\sqrt{n}), \qquad (2)$$

where $\boldsymbol{\xi}$ is a symmetric matrix such that $\xi_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ for $i \le j$, and Δ is the inverse Fisher information (evaluated at $w=0$) of the channel $P_{\rm out}(y|w)$:

$$\frac{1}{\Delta} \equiv \mathbb{E}_{P_{\rm out}(y|0)}\left[\left(\frac{\partial \log P_{\rm out}(y|w)}{\partial w}\bigg|_{y,0}\right)^2\right]. \qquad (3)$$

arXiv:1603.08447 *FK, Xu & Zdeborova '16*

☑ Conjectured in *Lesieur, FK & Zdeborova* '15

☑ Rank 1 SBM used & proven in *Abbe, Deshpande, Montanari'16*

OUTLINE

1) Message passing, State evolution, Mutual information

2) Universality property

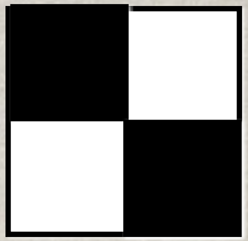
<u>3) Main results</u>

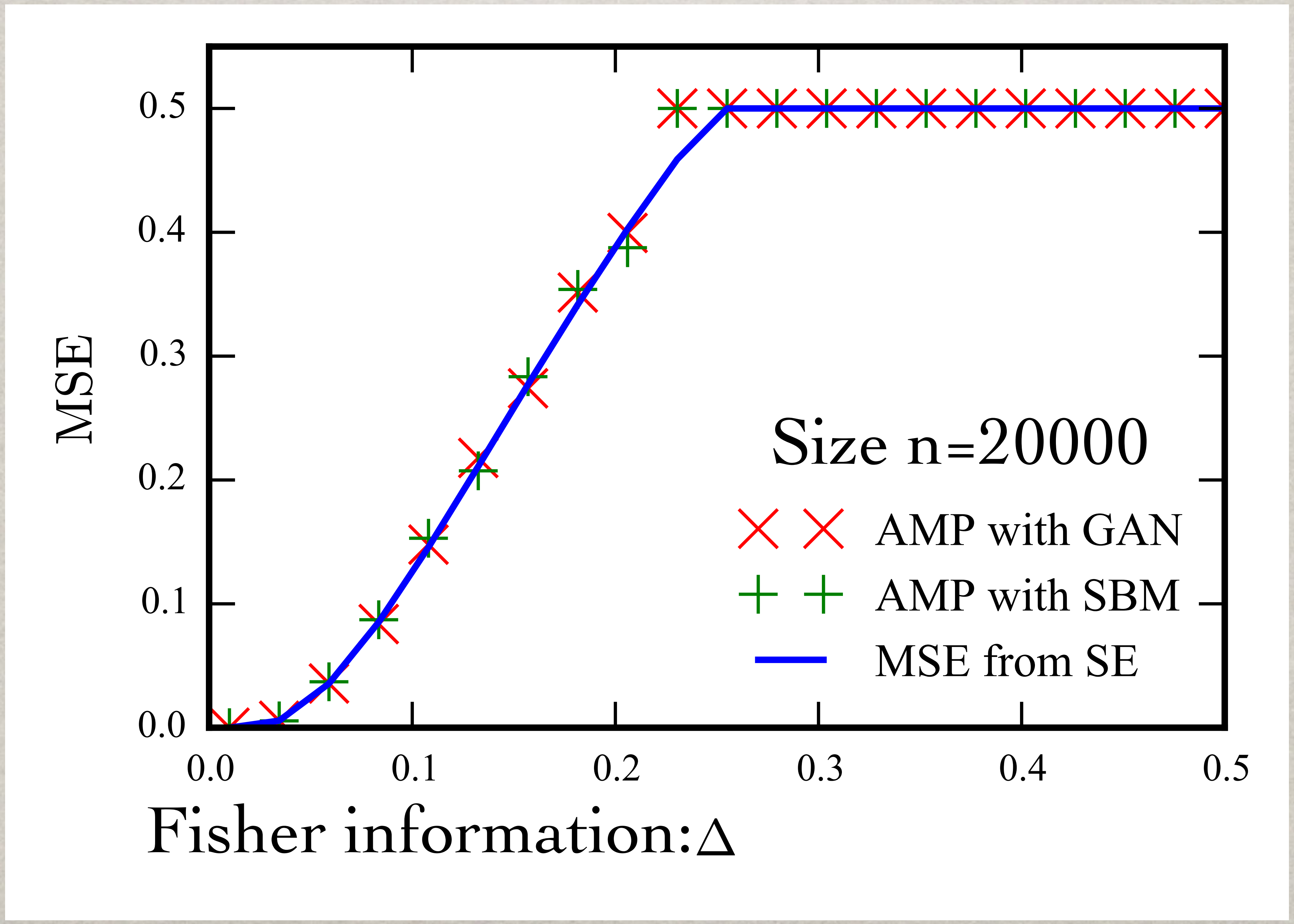
4) Sketch of proof

Minimum mean-squared error

$x_i^T = (0, \ldots, 0, 1, 0, \ldots, 0)$ $\quad W = \frac{1}{\sqrt{n}} X X^T$ $\quad Y = P_{\text{out}}(W)$

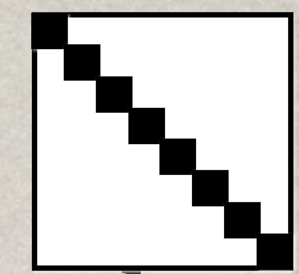
Rank r=2



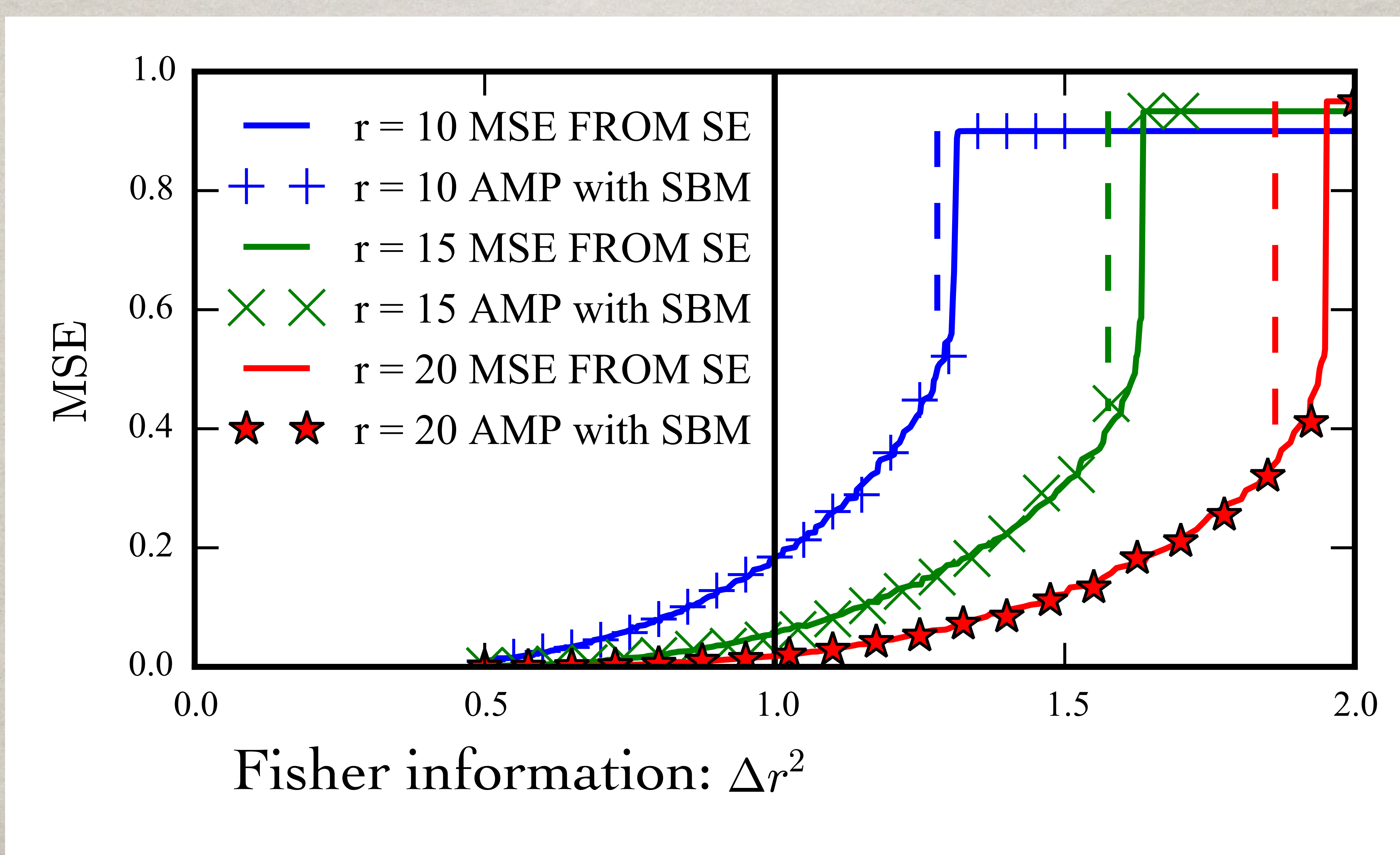


LARGER RANK ?

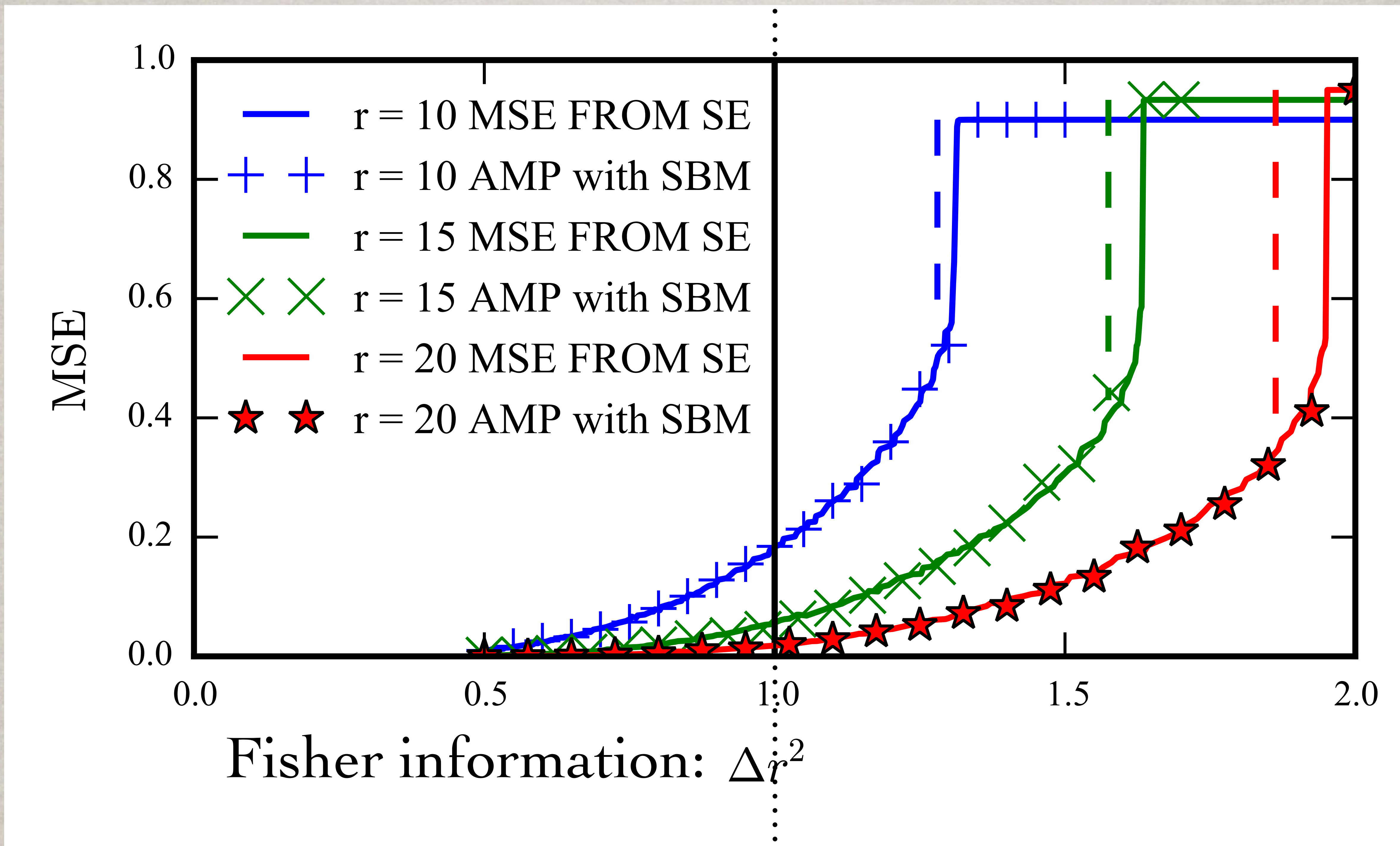
$x_i^T = (0, \ldots, 0, 1, 0, \ldots, 0)$ $\quad W = \frac{1}{\sqrt{n}} X X^T$ $\quad Y = P_{\text{out}}(W)$



Rank >=4



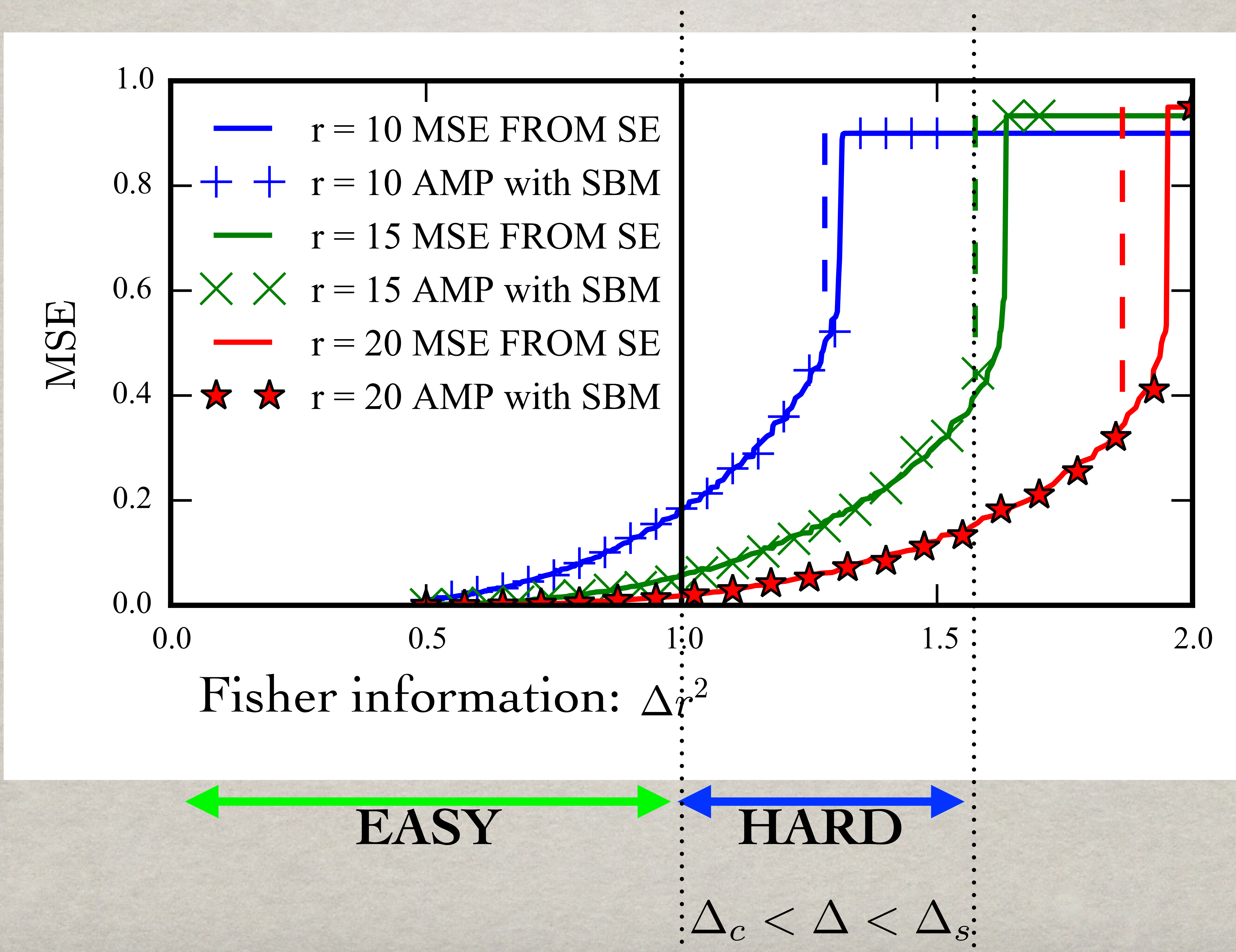
Two phase transitions for r>4



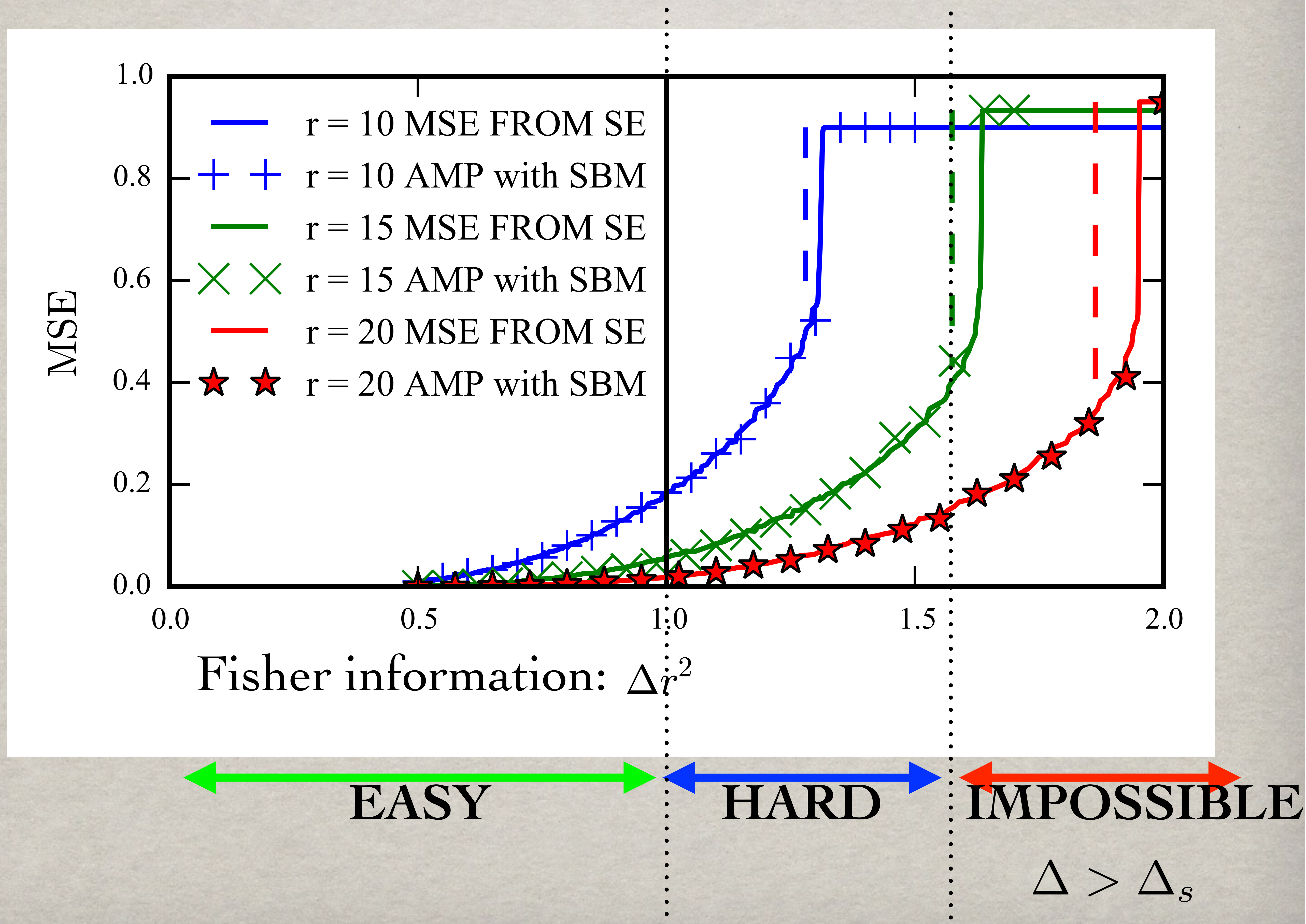
EASY

$\Delta < \Delta_c$

Two phase transitions for r>4



Two phase transitions for r>4



Two phase transitions for r>4

Same transition in spectral methods (~BBP'05)

- **Easy phase** $\Delta < \frac{1}{r^2} = \Delta_c$

$$|p_{\rm in} - p_{\rm out}| > \frac{1}{\sqrt{n}} r \sqrt{p_{\rm out}(1-p_{\rm out})}\,.$$

Decelle, Moore, FK, Zdeborova'11

- **Hard phase** $\frac{1}{r^2} < \Delta < \frac{1}{4r\ln(r)}[1+o_r(1)] = \Delta_s$

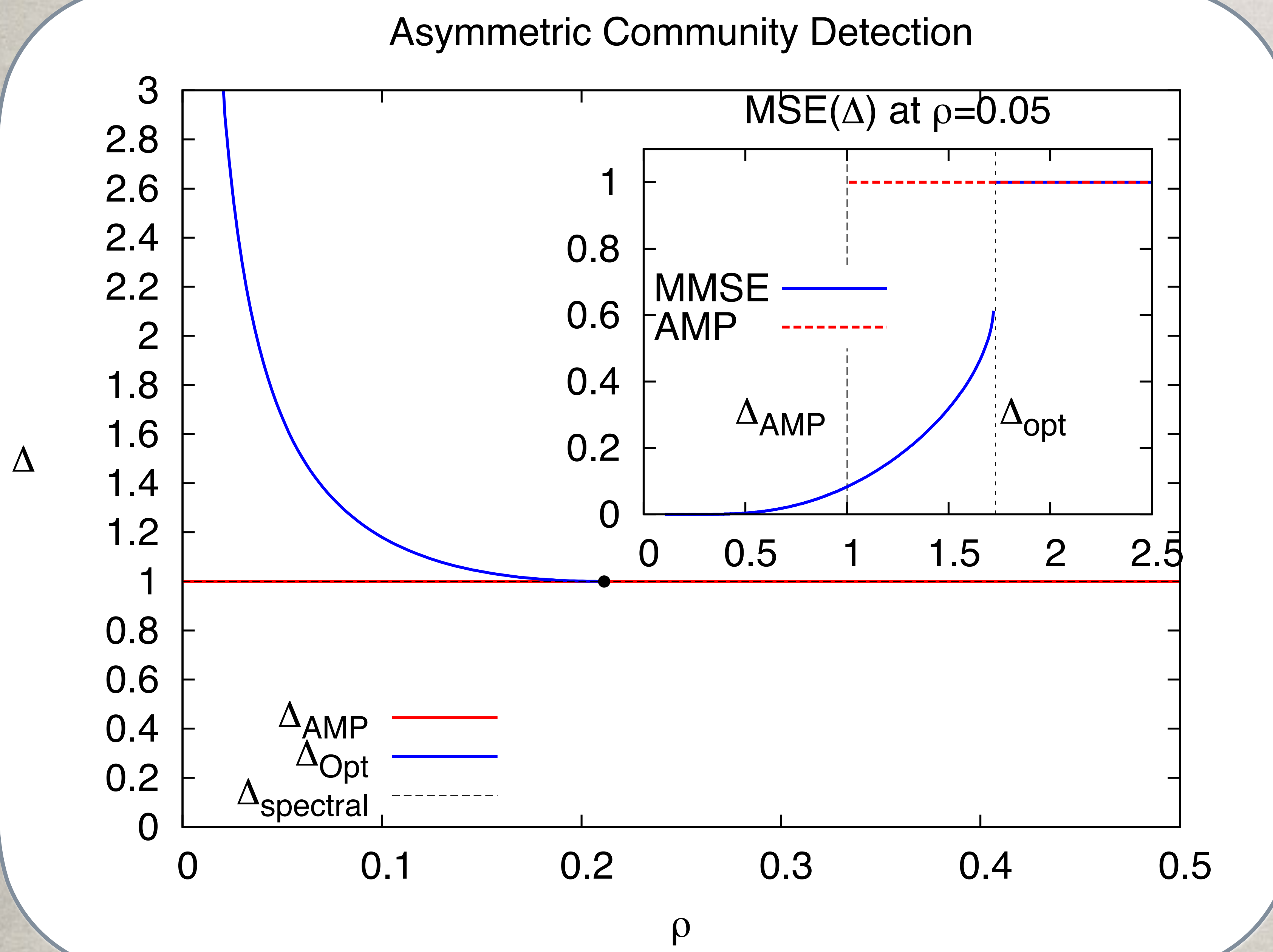
Conjectured to be hard for all polynomial algorithms

- **Impossible phase** $\Delta > \frac{1}{4r\ln(r)}[1+o_r(1)] = \Delta_s$

$$|p_{\rm in} - p_{\rm out}| < \frac{1}{\sqrt{n}} 2\sqrt{r\log r}\sqrt{p_{\rm out}(1-p_{\rm out})}\,.$$

Related to Potts glass temperature *(Kanter, Gross, Sompolinsky'85)*

Non-symmetric Community detection



2 communities +/-

size +: ρn

size -: (1-ρ)n

$$p_{++} = p + \mu \frac{1-\rho}{\rho\sqrt{n}}$$

$$p_{--} = p + \mu \frac{\rho}{(1-\rho)\sqrt{n}}$$

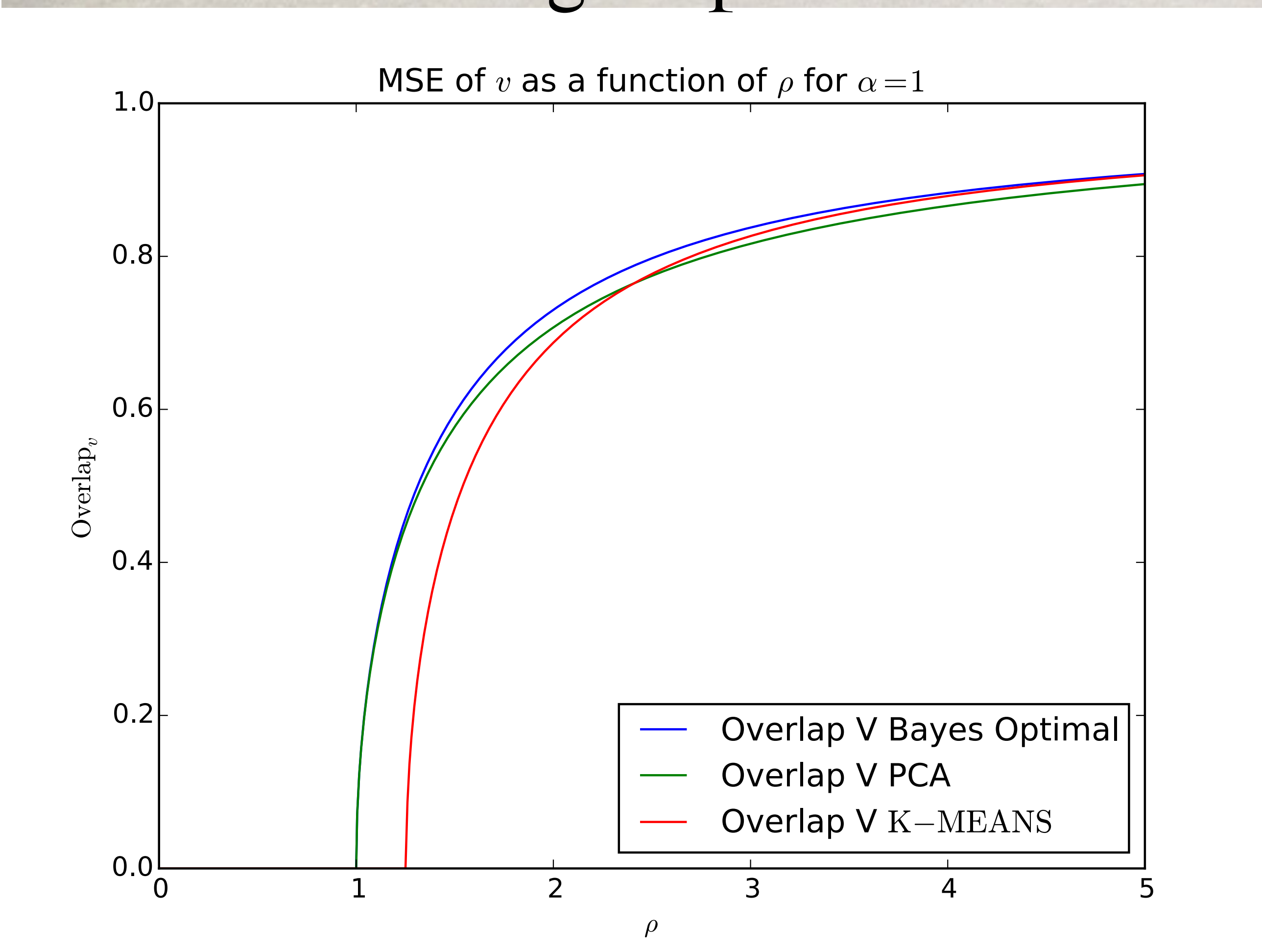
$$p_{+-} = p - \mu \frac{1}{\sqrt{n}}$$

$$P(x) = \rho\delta\left(x - \sqrt{\frac{1-\rho}{\rho}}\right) + (1-\rho)\delta\left(x + \sqrt{\frac{\rho}{1-\rho}}\right)$$

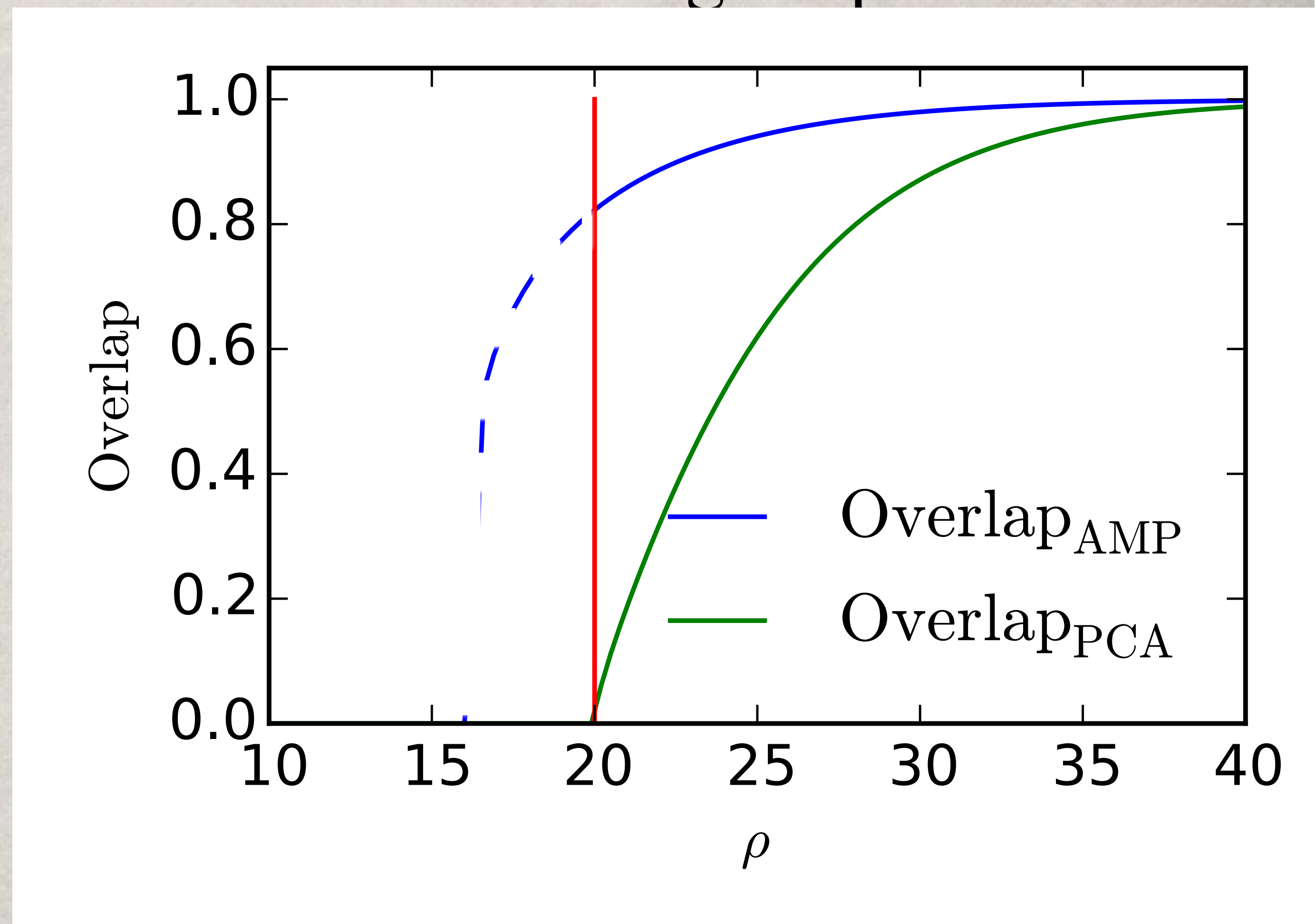
$$\Delta = \frac{p(1-p)}{\mu^2}$$

CLUSTERING MIXTURES OF GAUSSIANS IN HIGH DIMENSIONS

2 groups



20 groups

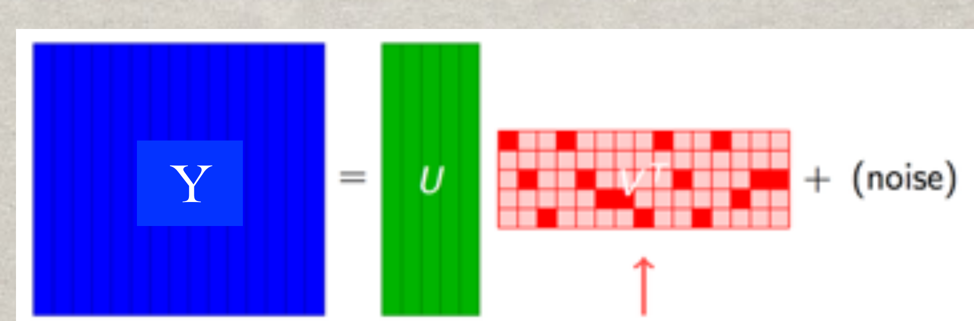


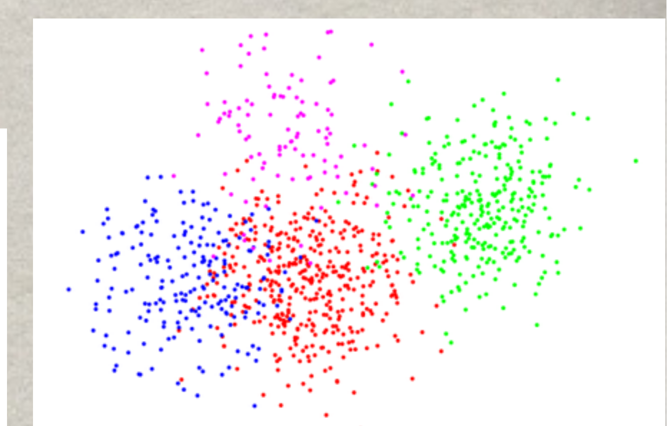
$$u_i^T = (v_i^1, \ldots, v_i^R) \in \mathbb{R}$$

α=m/n=1

$$v_i^T = (0, \ldots, 0, 1, 0, \ldots, 0)$$

$$W = \frac{1}{\sqrt{n}} U V^T \quad + \mathcal{N}(0, 1/\rho)$$

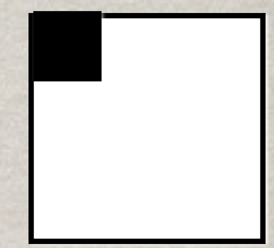




- Algorithm first proposed by Matsushita, Tanaka '13

What about RANK 1 sparse PCA?

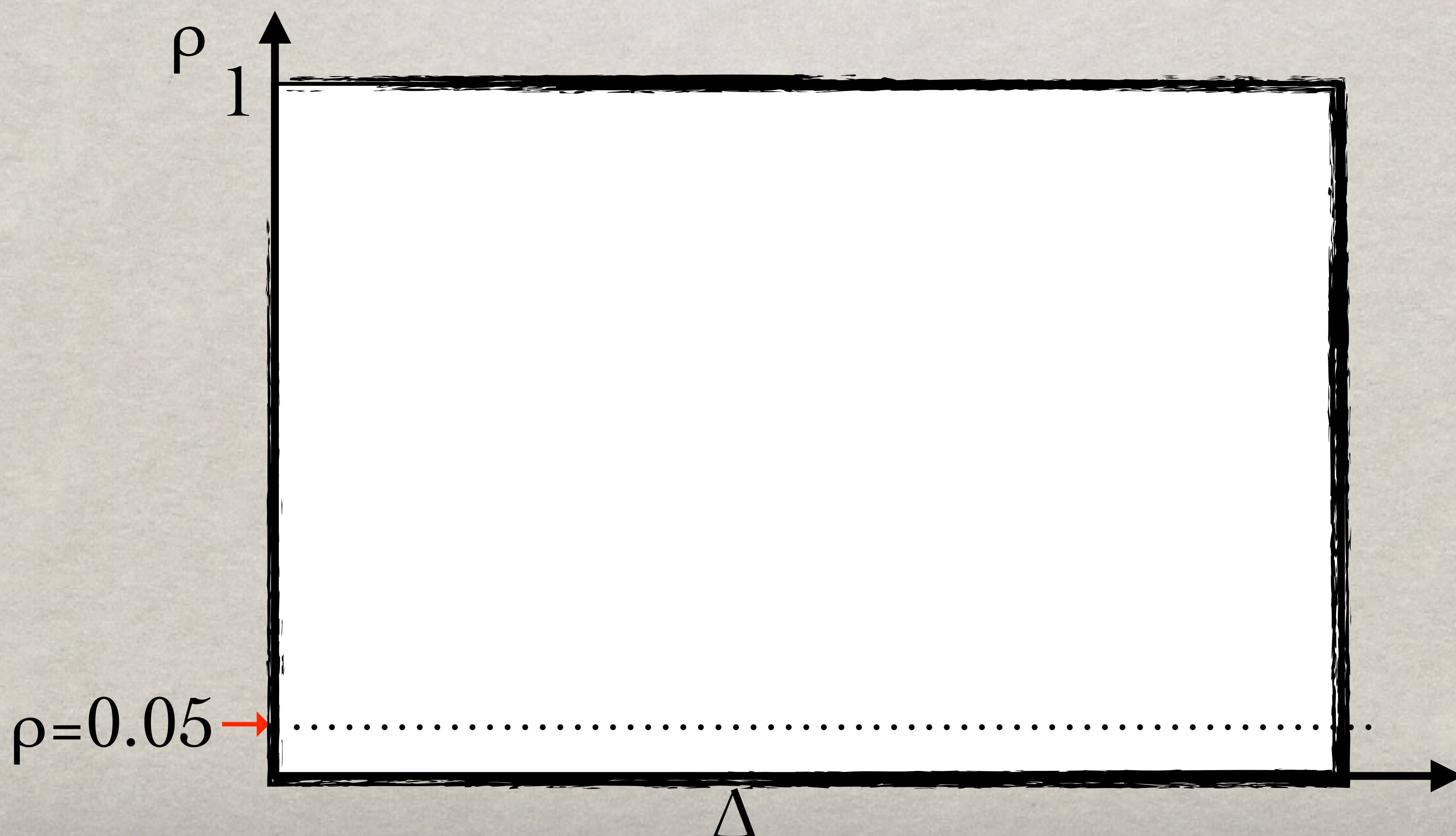
$$X_i = \{0, 1\} \qquad W = \frac{1}{\sqrt{n}} X X^T \qquad Y = W + \mathcal{N}(0, \Delta)$$



$$P(x) = \rho \delta_{x,1} + (1 - \rho) \delta_{x,0}$$

Rank r=1

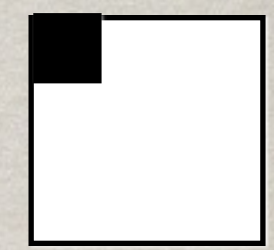
For ρ > 0.05, Montanari and Deshpande showed that AMP achieved the optimal MMSE



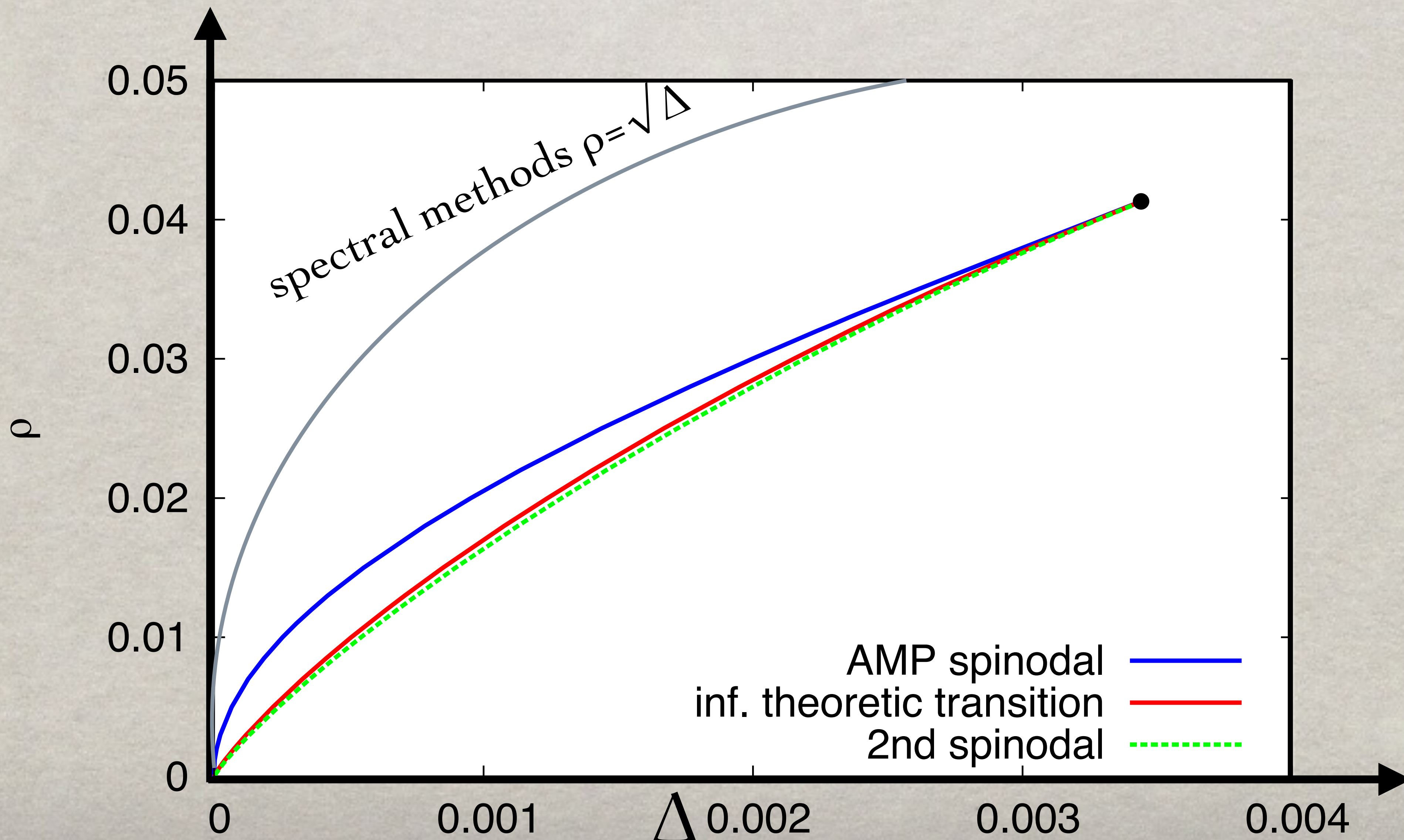
What about RANK 1 sparse PCA?

$X_i = \{0,1\}$ $\quad W = \frac{1}{\sqrt{n}} X X^T \quad Y = W + \mathcal{N}(0,\Delta)$

$P(x) = \rho\delta_{x,1} + (1-\rho)\delta_{x,0}$ Rank r=1

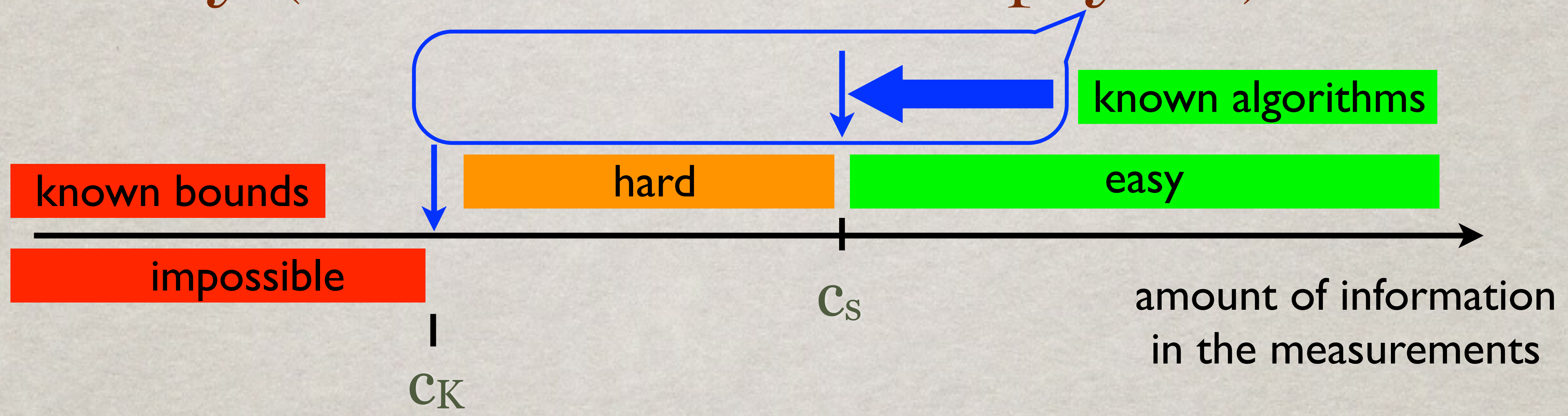


AMP achieved the optimal MMSE everywhere EXCEPT between the blue and red curves



Easy, Hard and impossible inference

- A very common phenomena: Information is hidden by metastability (1st order transition in physics)



- The hard phase quantified also in: planted constraint satisfaction, compressed sensing, stochastic block model, dictionary learning, blind source separation, sparse PCA, error correcting codes, hidden clique problem, others ….

- Conjecture: hard for **<u>all</u>** polynomial algorithms

OUTLINE

1) Message passing, State evolution, Mutual information

2) Universality property

3) Main results

<u>4) Sketch of proof</u>

SKETCH OF THE PROOF

Result 1 [FK, Xu & Zdeborová] : Upper Bound

$$\frac{I}{n} \leq i_{\mathrm{Bethe}}(m) = \frac{m^2 + \left[\mathbb{E}_x(x^2)\right]^2}{4\Delta} - \mathbb{E}_{x,z}\left[\mathcal{J}\left(\frac{m}{\Delta}, \frac{mx}{\Delta} + \sqrt{\frac{m}{\Delta}}z\right)\right]$$

(note that this is true at any value of n, not only asymptotically)

Method: Guerra's interpolation+Nishimori identities

SKETCH OF THE PROOF

Result 1 [FK, Xu & Zdeborová] : Upper Bound

Interpolate the factorization problem at t=1 from a denoising problem at t=0

$$P_t(x|Y,D) = \frac{1}{Z_t} p_0(x) e^{t \sum_{i \leq j} \left[- \frac{\left(\frac{x_i x_j}{\sqrt{n}} - Y_{ij} \right)^2}{2\Delta} \right] + (1-t) \sum_i \left[\frac{(D_i - x_i)^2}{2\Delta_D} \right]}$$

Factorization: $w_{ij} = x_i^T x_j / \sqrt{n} \xrightarrow{\mathcal{N}(0, \Delta/t)} y_{ij}$

Denoising: $x_i \xrightarrow{\mathcal{N}(0, \Delta_D/(1-t))} D_i$ with $\Delta_D = \frac{\Delta}{m}$

$$I^{t=1}(x;Y) = I^{t=0}(x;y) + \int_0^1 \frac{d}{dt} I^t(x;Y,y)$$

SKETCH OF THE PROOF

Result 1 [FK, Xu & Zdeborová] : Upper Bound

Interpolate the factorization problem at t=1 from a denoising problem at t=0

$$P_t(x|Y,D) = \frac{1}{Z_t} p_0(x) e^{t \sum_{i \le j} \left[- \frac{\left(\frac{x_i x_j}{\sqrt{n}} - Y_{ij} \right)^2}{2\Delta} \right] + (1-t) \sum_i \left[\frac{(D_i - x_i)^2}{2\Delta_D} \right]}$$

$$I^{t=1}(x;Y) = I^{t=0}(x;y) + \int_0^1 \frac{d}{dt} I^t(x;Y,y)$$

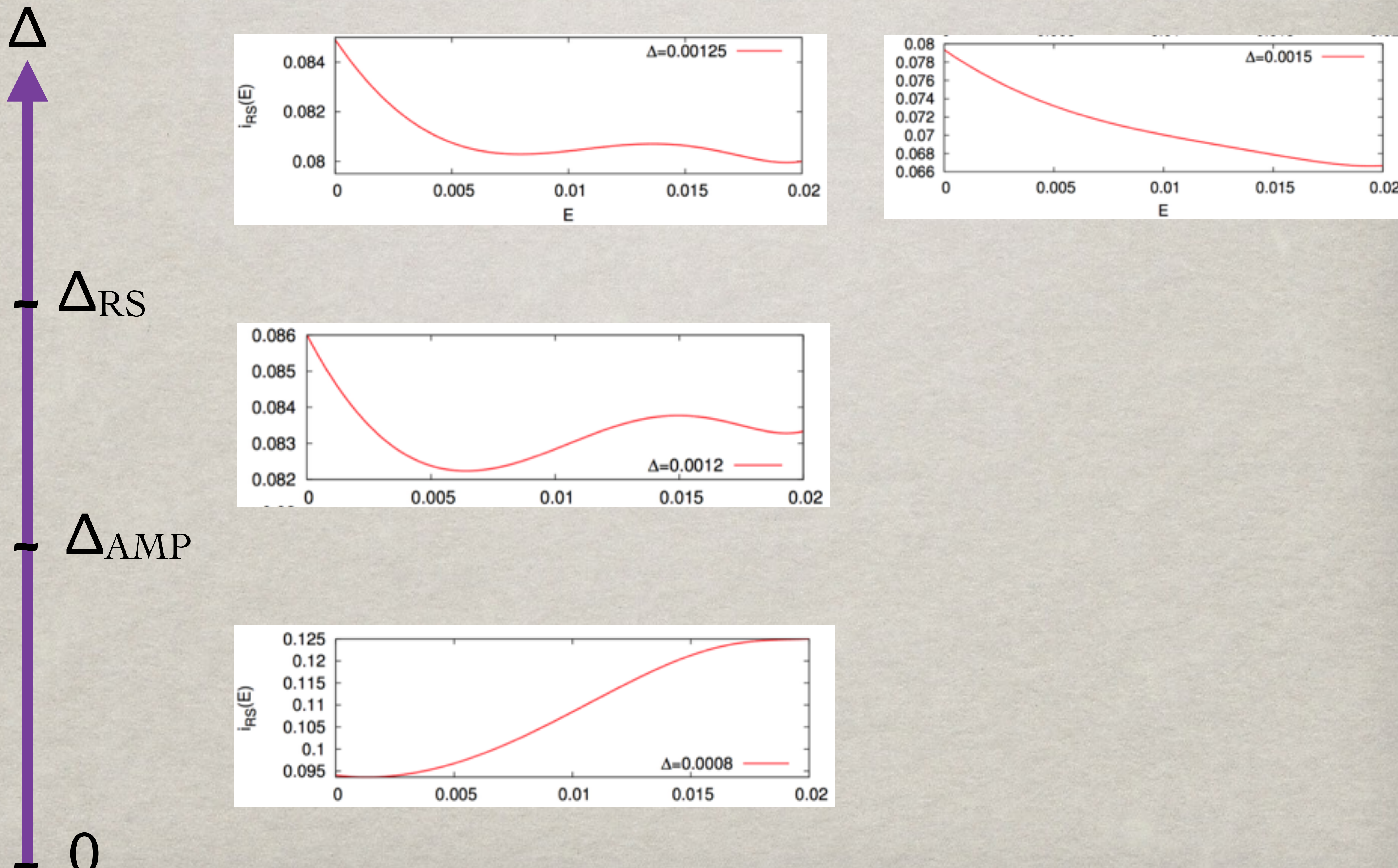
Use Stein lemma and Nishimori's identities, one can show that

$$\frac{I(x;Y)}{n} = i_{rs}(m) - \int_0^1 dt \frac{(m - \mathbb{E}_t[x x_0])^2}{4\Delta} < i_{rs}(m)$$

Remark: for estimation problems, Guerra's interpolation yields an upper bound while in the usual case (i.e. CSP) it yields a lower bound

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound

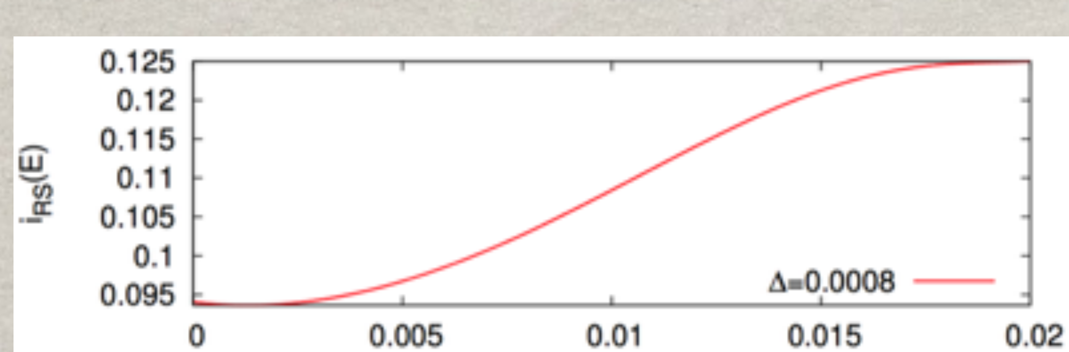


SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound

Step 1: For $\Delta < \Delta_{\mathrm{AMP}}$

$$\mathrm{MSE}_{\mathrm{AMP}}(\Delta) \geq \mathrm{MMSE}(\Delta)$$



SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound

Step 1: For $\Delta < \Delta_{\text{AMP}}$

$$\text{MSE}_{\text{AMP}}(\Delta) \geq \text{MMSE}(\Delta)$$

I-mmse theorem (Guo-Verdu)

(For physicists: this is just thermodynamics)

$$\frac{d}{d\Delta^{-1}} i(\Delta) = \frac{1}{4}\text{MMSE}(\Delta)$$

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound

Step 1: For $\Delta < \Delta_{\mathrm{AMP}}$

$$\mathrm{MSE}_{\mathrm{AMP}}(\Delta) \geq \mathrm{MMSE}(\Delta)$$

$$\frac{d}{d\Delta^{-1}}\left[\min_E i_{RS}(E;\Delta)\right] \geq \frac{d}{d\Delta^{-1}} i(\Delta)$$

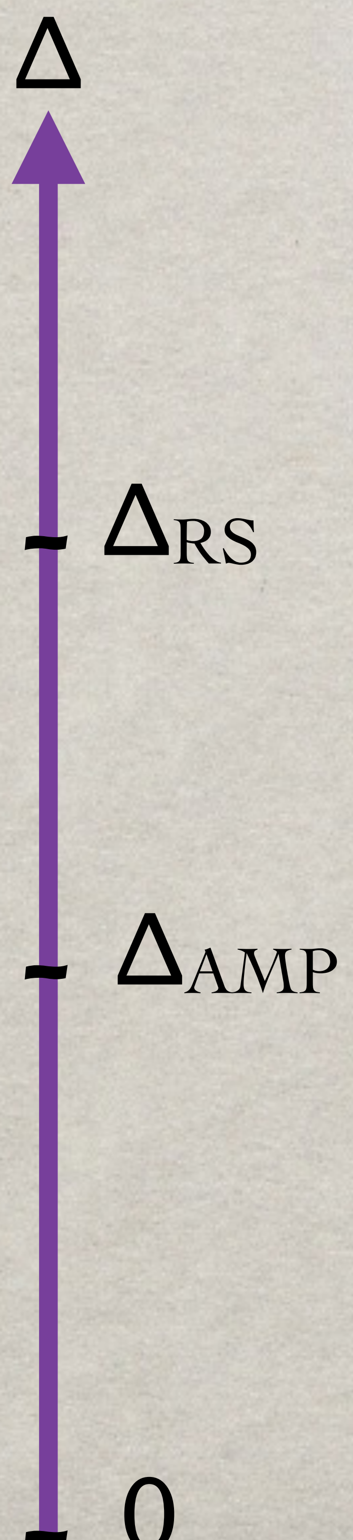
I-mmse theorem (Guo-Verdu)

(For physicists: this is just thermodynamics)

$$\frac{d}{d\Delta^{-1}} i(\Delta) = \frac{1}{4}\mathrm{MMSE}(\Delta)$$

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound



Step 1: For $\Delta < \Delta_{\rm AMP}$

$$\mathrm{MSE}_{\rm AMP}(\Delta) \geq \mathrm{MMSE}(\Delta)$$

$$\frac{d}{d\Delta^{-1}}\left[\min_E i_{RS}(E;\Delta)\right] \geq \frac{d}{d\Delta^{-1}} i(\Delta)$$

$$\frac{d}{d\Delta}\left[\min_E i_{RS}(E;\Delta)\right] \leq \frac{d}{d\Delta} i(\Delta)$$

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound

Step 1: For $\Delta < \Delta_{AMP}$

$$\mathrm{MSE}_{\mathrm{AMP}}(\Delta) \geq \mathrm{MMSE}(\Delta)$$

$$\frac{d}{d\Delta^{-1}}\left[\min{}_E i_{RS}(E;\Delta)\right] \geq \frac{d}{d\Delta^{-1}} i(\Delta)$$

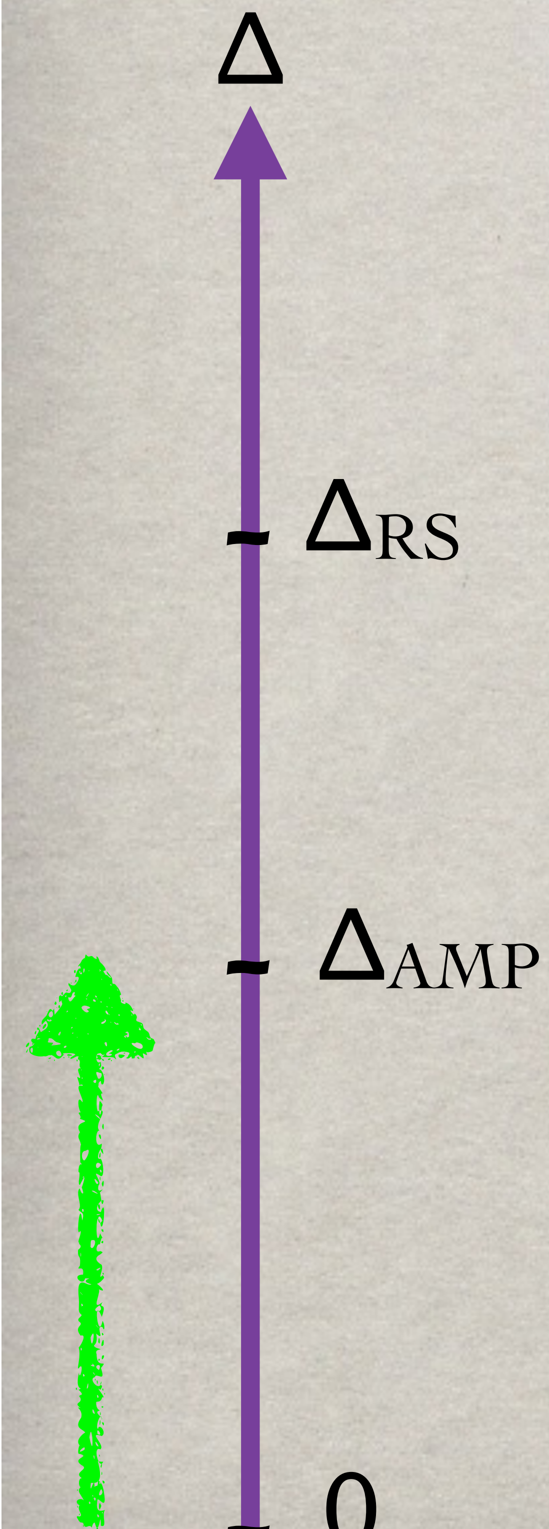
$$\frac{d}{d\Delta}\left[\min{}_E i_{RS}(E;\Delta)\right] \leq \frac{d}{d\Delta} i(\Delta)$$

$$\min{}_E i_{RS}(E;\Delta) - \min{}_E i_{RS}(E;0) \leq i(\Delta) - i(0)$$

Both are equal to H(x)

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound



Step 1: For $\Delta < \Delta_{\rm AMP}$

$$\mathrm{MSE}_{\rm AMP}(\Delta) \geq \mathrm{MMSE}(\Delta)$$

$$\frac{d}{d\Delta^{-1}}\left[\min_E i_{RS}(E;\Delta)\right] \geq \frac{d}{d\Delta^{-1}} i(\Delta)$$

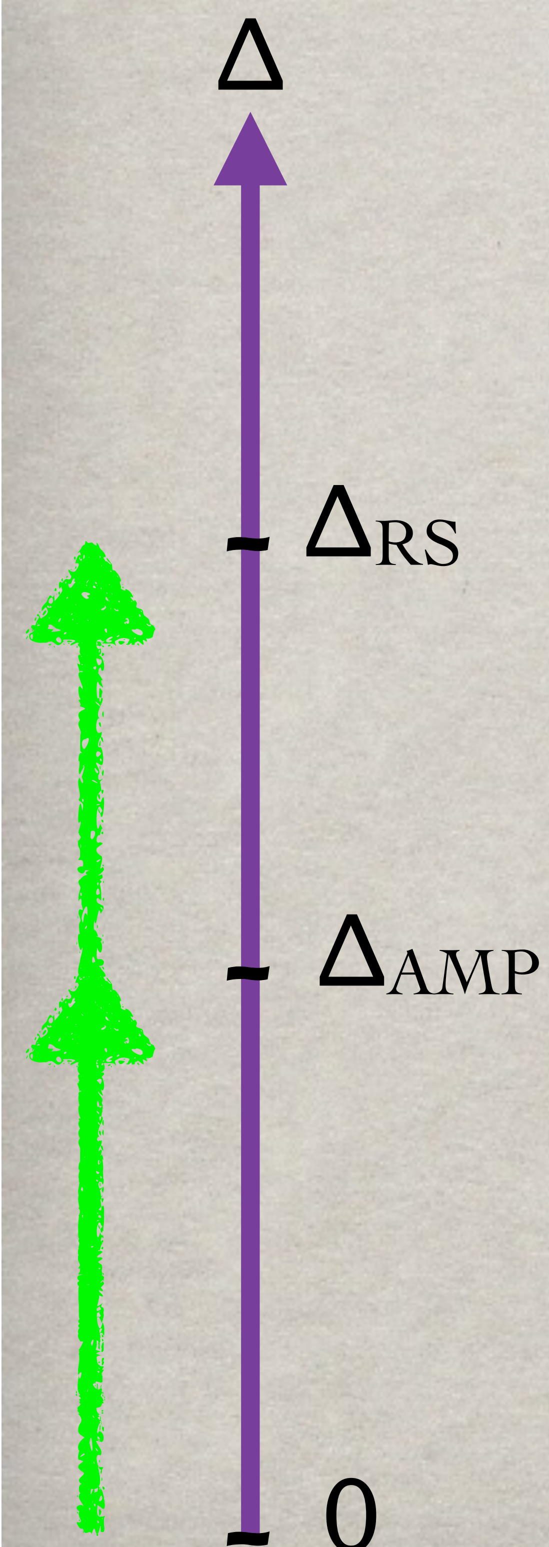
$$\frac{d}{d\Delta}\left[\min_E i_{RS}(E;\Delta)\right] \leq \frac{d}{d\Delta} i(\Delta)$$

$$\min_E i_{RS}(E;\Delta) - \min_E i_{RS}(E;0) \leq i(\Delta) - i(0)$$

$$\min_E i_{\rm RS}(\Delta, E) \leq i(\Delta)$$

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, Macris, FK & Zdeborová] : Converse Lower bound



Step 2: Prove that the true thermodynamic potential is analytic until Δ_{RS}

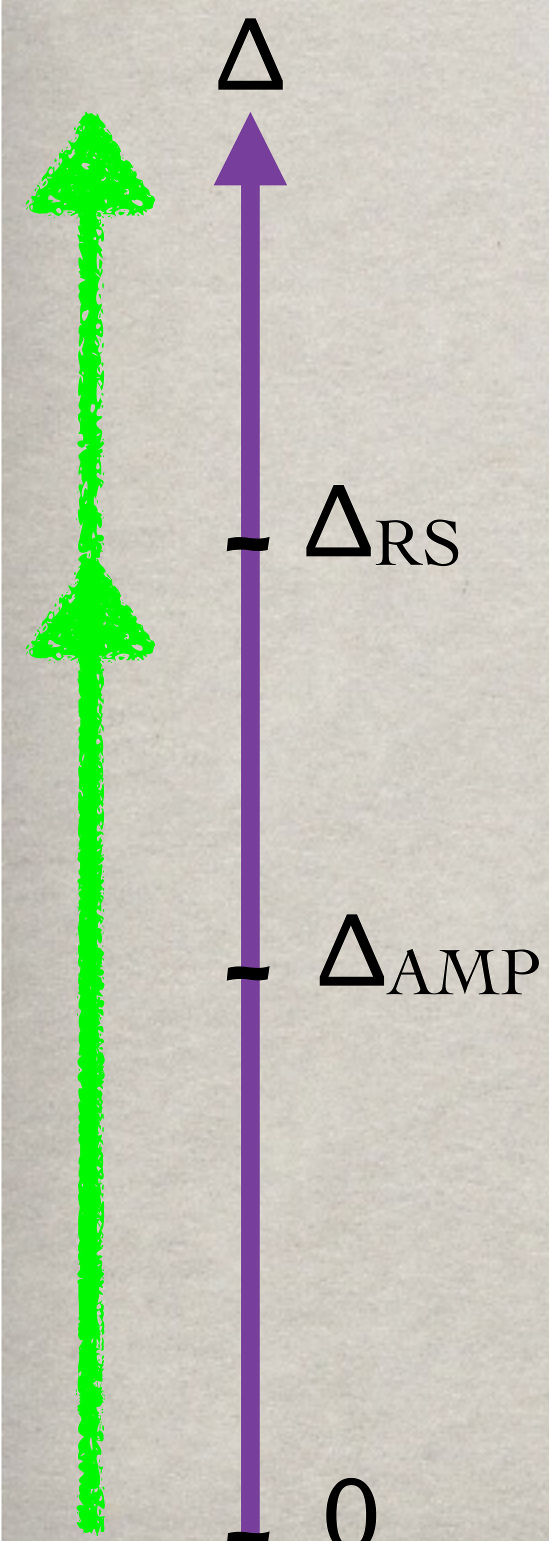
This is the hardest part. Sketch:

1) Map the problem to its "spatially coupled" version
2) Prove both models have the same mutual information *(with Guerra's type interpolation)*
3) Prove the spatially coupled version is analytic until Δ_{RS} *(Threshold saturation)*

More in Nicolas Macris's talk this afternoon

SKETCH OF THE PROOF

Result 2 [Barbier, Dia, *Macris*, FK & Zdeborová] : Converse Lower bound



Step 3:

For $\Delta > \Delta_{RS}$,
AMP provides back the MMSE
Use again the I-MMSE theorem
from Δ_{RS} to $\Delta > \Delta_{RS}$

Conclusions

- MMSE in a random setting of the low-rank matrix estimation evaluated *(Proven rigorously for symmetric problems, open for generic UV^T factorization)*.
- Promising proof strategy for the replica formula in estimation problems (possible generalization for larger ranks, tensors, etc…)
- Dependence on the output channel through its Fisher information: *universality*
- Two phase transitions: an algorithmic one and an information theoretic one. Sometimes equal, but not always; AMP reach the MMSE in a large region.
- When the problem is not balanced (i.e. if the prior has a non-zero mean) AMP has a *better detectability transition* than spectral methods.
- Promising algorithm with good convergence properties for applications beyond random setting (example: Restricted Boltzmann Machine).
- **Open-source implementation available**
  http://krzakala.github.io/LowRAMP/