Spin Glass Approach to Restricted Isometry Constant

Ayaka Sakata[1], Yoshiyuki Kabashima[2]

[1] Institute of Statistical Mathematics

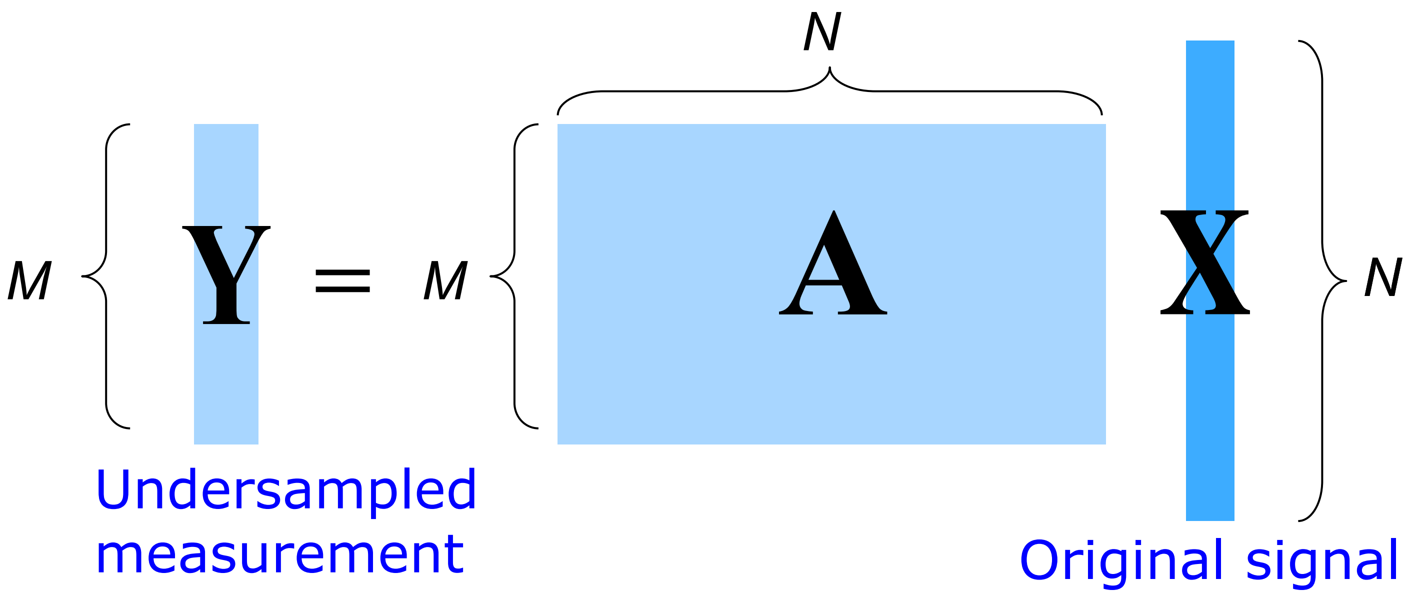
[2] Tokyo Institute of Technology

Outline

- Background: Compressed sensing
- Problem setup: Restricted isometry constant (RIC)
- Spin glass approach
  - Replica symmetric analysis
  - Improvement by replica symmetry breaking
- Summary

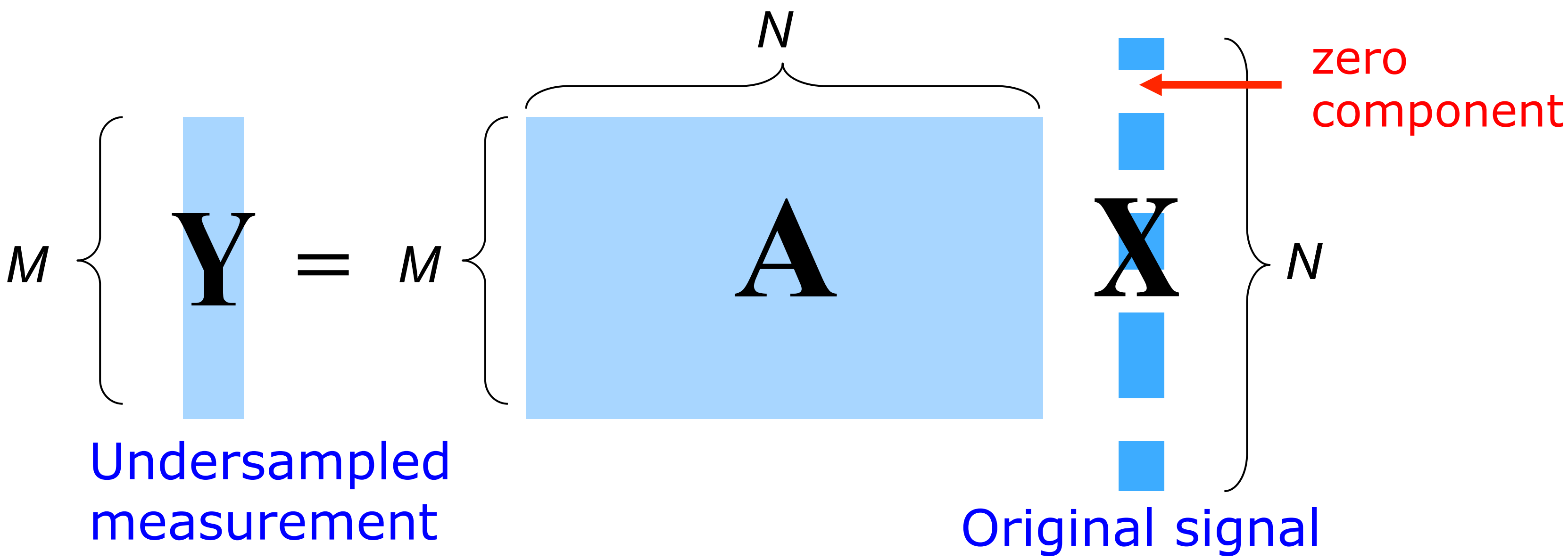
Compressed sensing

Reconstruct original signal ***X***
from its undersampled measurement ***Y***.



Compressed sensing

Reconstruct original signal ***X***
from its undersampled measurement ***Y***.

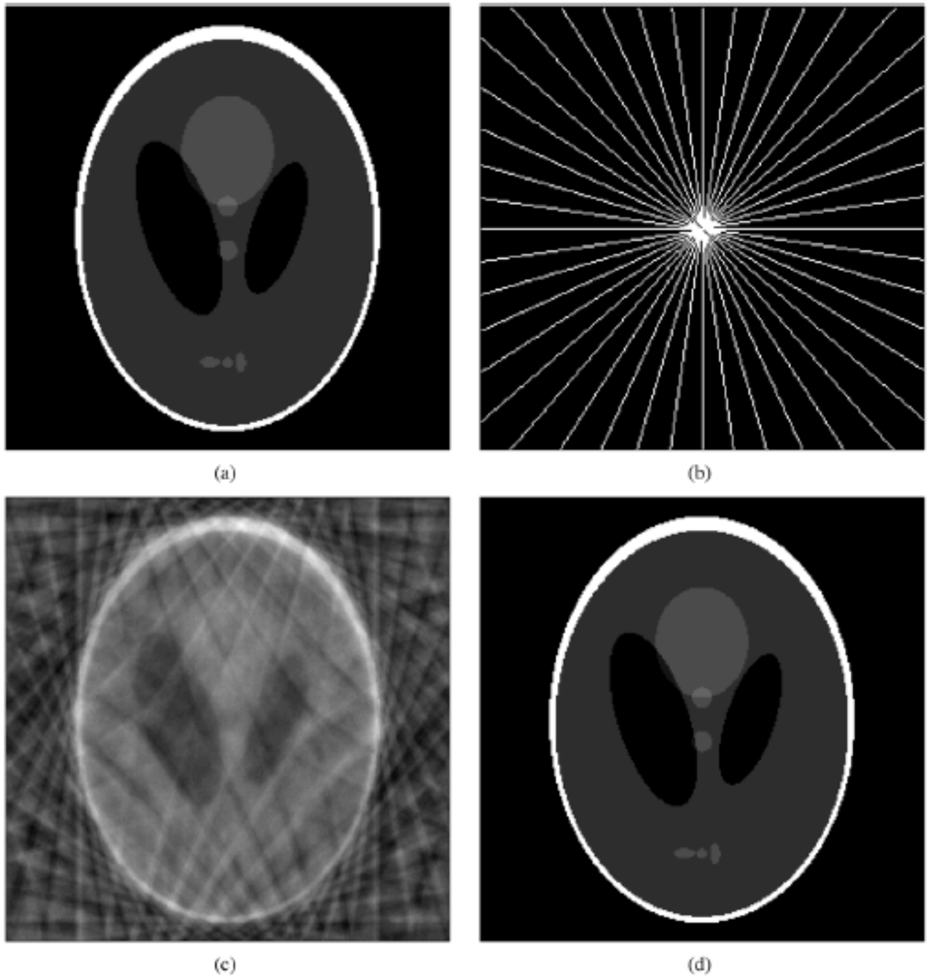


Reconstruction is possible
when ***X*** is sufficiently "***sparse***".

Practical relevance

- The problem is related to various technologies of modern signal processing
- Many application domains
  - Refraction seismic survey (mine examination)
  - Tomography（X-ray CT，MRI）
  - Single pixel camera
  - Noise removal of image
  - Data streaming computing
  - Group testing
  - etc.

Simulation of tomography



LT : Original(Logan-Shepp Phantom)
#512x512

RT : Sampling 512 points of 2D FT from 22 directions.

LB : Recovery of pseudo-inverse (standard approach)

RB : Recovery utilizing the "sparseness" of spatial variations. "Original" is perfectly recovered.

Perfect recovery is realized by only 2% samples of what Nyquist-Shannon's theory requires.
→Breaking of the conventional limit!

EJ Candes J Romberg and T. Tao, IEEE Trans. IT Vol. 52, 489—502 (2006)

Two major reconstruction methods

l_0 reconstruction

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{subject to } \mathbf{Y} = \mathbf{A}\mathbf{x}$$

l_1 reconstruction

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{x}\|_1, \text{subject to } \boldsymbol{Y} = \boldsymbol{A}\boldsymbol{x}$$

Sufficient conditions for the reconstruction

are given by

Restricted Isometriy Constant (RIC).

Restricted Isometry Constant (RIC)

Definition (Candès and Tao (2006)).

$\left(\sum_{\mu=1}^{M} A_{\mu i}^2 = 1 \text{ for } \forall i \right)$

Let $\boldsymbol{A}$ be a column-wisely normalized $M \times N$ matrix.

Let $\boldsymbol{x} \in \mathrm{R}^N$ be an arbitrary S-sparse vector, whose number of non-zero components is smaller than S.

Then, if there exists such a constant $\delta_S = \max\{\delta_S^{\min}, \delta_S^{\max}\}$ that satisfies

$$(1-\delta_S^{\min})\|\boldsymbol{x}\|_F^2 \le \|\boldsymbol{A}\boldsymbol{x}\|_F^2 \le (1+\delta_S^{\max})\|\boldsymbol{x}\|_F^2,$$

$\boldsymbol{A}$ is said to satisfy the S-restricted isometry property (RIP) with the restricted isometry constant (RIC) δ_S.

Intuitively, δ_s quantifies how $\boldsymbol{A}$ deviates from orthogonal transforms in terms of Frobenius (L_2) norm for S-sparse vectors.

Sufficient conditions for l_0, l_1 reconstruction

1. l_0 reconstruction gives the unique S-sparse solution when $d_{2S} < 1$.

2. l_1 reconstruction gives the same unique S-sparse solution as l_0 reconstruction when $d_{2S} < 2^{1/2} - 1$.

[Candes and Tao, *IEEE Trans. Inform. Theory* (2005)]

3. (Improvement of **2.**)

l_1 reconstruction gives the same unique S-sparse solution as l_0 reconstruction when $(4\sqrt{2}-3)\delta_{2S}^{\min} + \delta_{2S}^{\max} < 4(\sqrt{2}-1)$.

[Foucart and Lai, *Appl. Comput. Harmon. Anal.* (2009)]

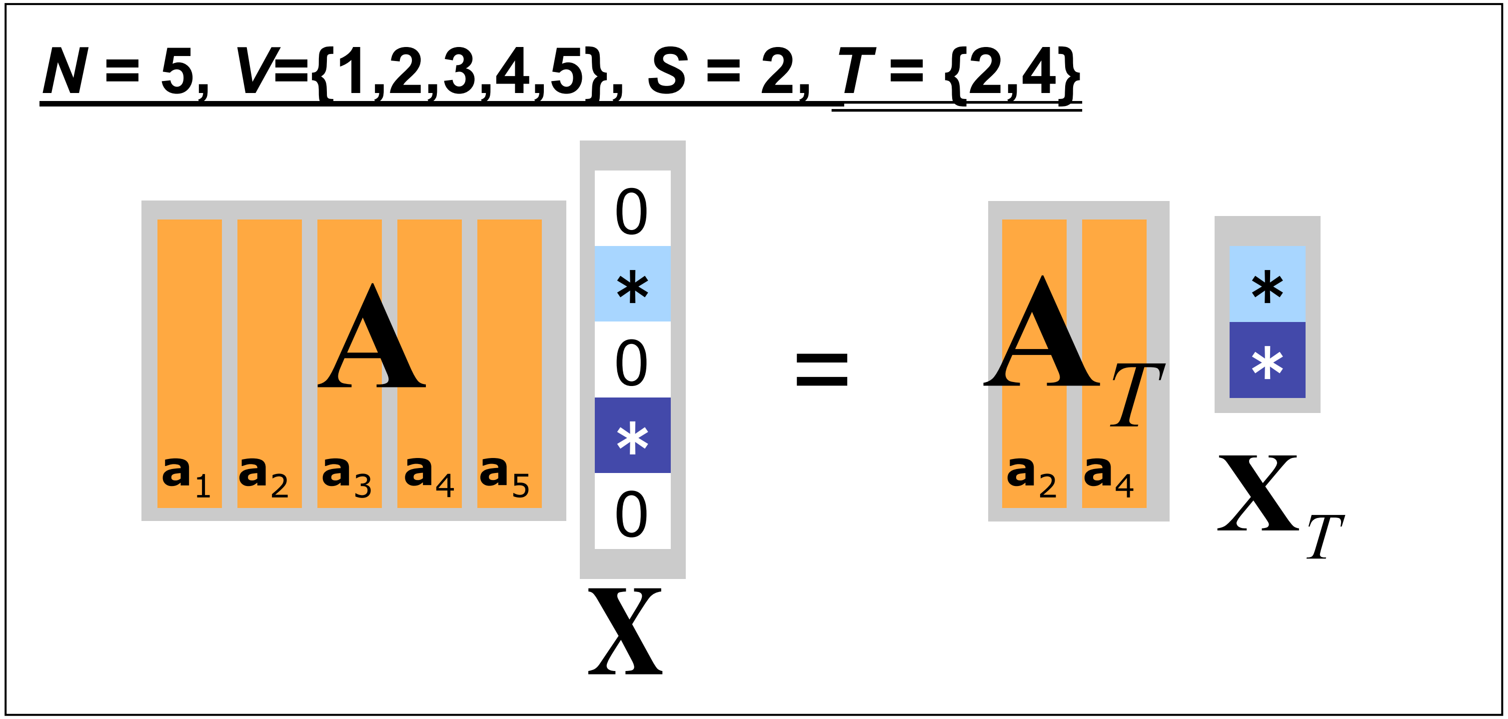
Computational difficulty

- RIC plays a key role in theoretical analysis and performance guarantee of compressed sensing.
- On top of this, it is purely of great interest as a fundamental problem of linear algebra.

- Unfortunately, *"there are no known large (systematic) matrices with bounded RICs (and computing RICs is strongly NP-hard),…".*
  - Wikipedia
- On the other hand, many random matrices have been shown to remain bounded. Therefore, much effort has been paid for improving the bounds.
  - *Our study is along this line.*

Why computationally difficult?

Suppose a situation where non-zero components are fixed.



$$\lambda_{\min}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T)\|\mathbf{x}_T\|_F^2 \le \|\mathbf{A}_T\mathbf{x}_T\|_F^2 \le \lambda_{\max}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T)\|\mathbf{x}_T\|_F^2$$

The change of norm can be easily characterized by eigenvalues of “sub-matrix” A_T.

Why computationally difficult?

- In evaluation of RIC, the inequalities must hold for <u>all possible combinations of the positions of non-zeros.</u>

- This causes a combinatorial difficulty, yielding an exact expression of RIC as

$$\delta_S = \max\{1 - \min_{T:|T|=S, T\subseteq V} \lambda_{\min}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T), \max_{T:|T|=S, T\subseteq V} \lambda_{\max}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T) - 1\}.$$

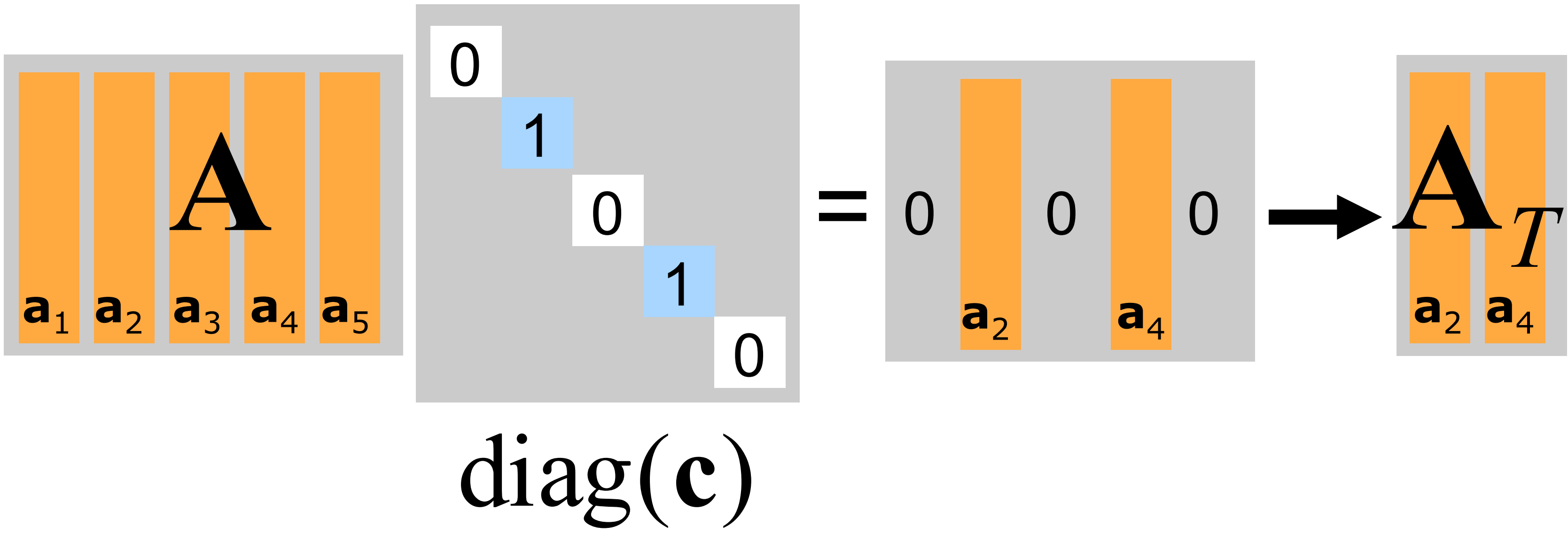
All possible column choices
=> Combinatorial difficulty

Structures of the problem and the difficulty are <u>analogous to those of spin glass problems.</u>

Analogy to spin glasses

The choice of T are represented by binary vector $\{c_i\} \in \{0, 1\}^N$.

***N* = 5, *S* = 2, *T* = {2,4} → *c* = {0, 1, 0, 1, 0}**



$\lambda_{\min}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T)$ and $\lambda_{\max}(\mathbf{A}_T^{\mathrm{T}}\mathbf{A}_T)$ can be regarded as \`energy functions' of $\boldsymbol{c}$ given $\mathbf{A}$.

Analogy to spin glasses

$$\delta_S = \max\{1 - \min_{\mathbf{c}:\sum_{i=1}^{N} c_i = S} \lambda_{\min}(\mathbf{c}\,|\,\mathbf{A}),\ \max_{\mathbf{c}:\sum_{i=1}^{N} c_i = S} \lambda_{\max}(\mathbf{c}\,|\,\mathbf{A}) - 1\}$$

RIC is given by minimum and maximum 'energy' of ***c***.

- Energy of '0-1 spin' $\boldsymbol{c}$

$$\Lambda_+(\boldsymbol{c}\,|\,\boldsymbol{A}) = \lambda_{\min}(\boldsymbol{c},\boldsymbol{A})$$
$$\Lambda_-(\boldsymbol{c}\,|\,\boldsymbol{A}) = \lambda_{\max}(\boldsymbol{c},\boldsymbol{A})$$

- 'Canonical distribution' of $\boldsymbol{c}$

$$P(\boldsymbol{c}\,|\,\boldsymbol{A};\mu) \propto \exp\left(-\mu N \Lambda_{\mathrm{sgn}(\mu)}(\boldsymbol{c}\,|\,\boldsymbol{A})\right)\delta\left(\sum_{i=1}^{N} c_i - S\right)$$

- 'Quenched randomness'

$$P(A) = \left(2\pi M^{-1}\right)^{-MN/2} \exp\left(-\frac{M}{2}\sum_{\mu,i} A_{\mu i}^2\right)$$

Spin glass approach

- Free entropy (free energy)

$Z(\mu \mid \boldsymbol{A})$: Partition function

$$\phi(\mu \mid \boldsymbol{A}) = \frac{1}{N}\log\left[\sum_{\boldsymbol{c}\in\{0,1\}^N} \exp\left(-\mu N \Lambda_{\mathrm{sgn}(\mu)}(\boldsymbol{c} \mid \boldsymbol{A})\right)\delta\left(\sum_{i=1}^{N} c_i - S\right)\right]$$

- Typical properties can be assessed by the replica method.

$$\phi(\mu) = \left[\phi(\mu \mid \boldsymbol{A})\right]_{\boldsymbol{A}} = \lim_{n\to 0}\frac{\partial}{\partial n}\log\left[Z^n(\mu \mid \boldsymbol{A})\right]_{\boldsymbol{A}}$$

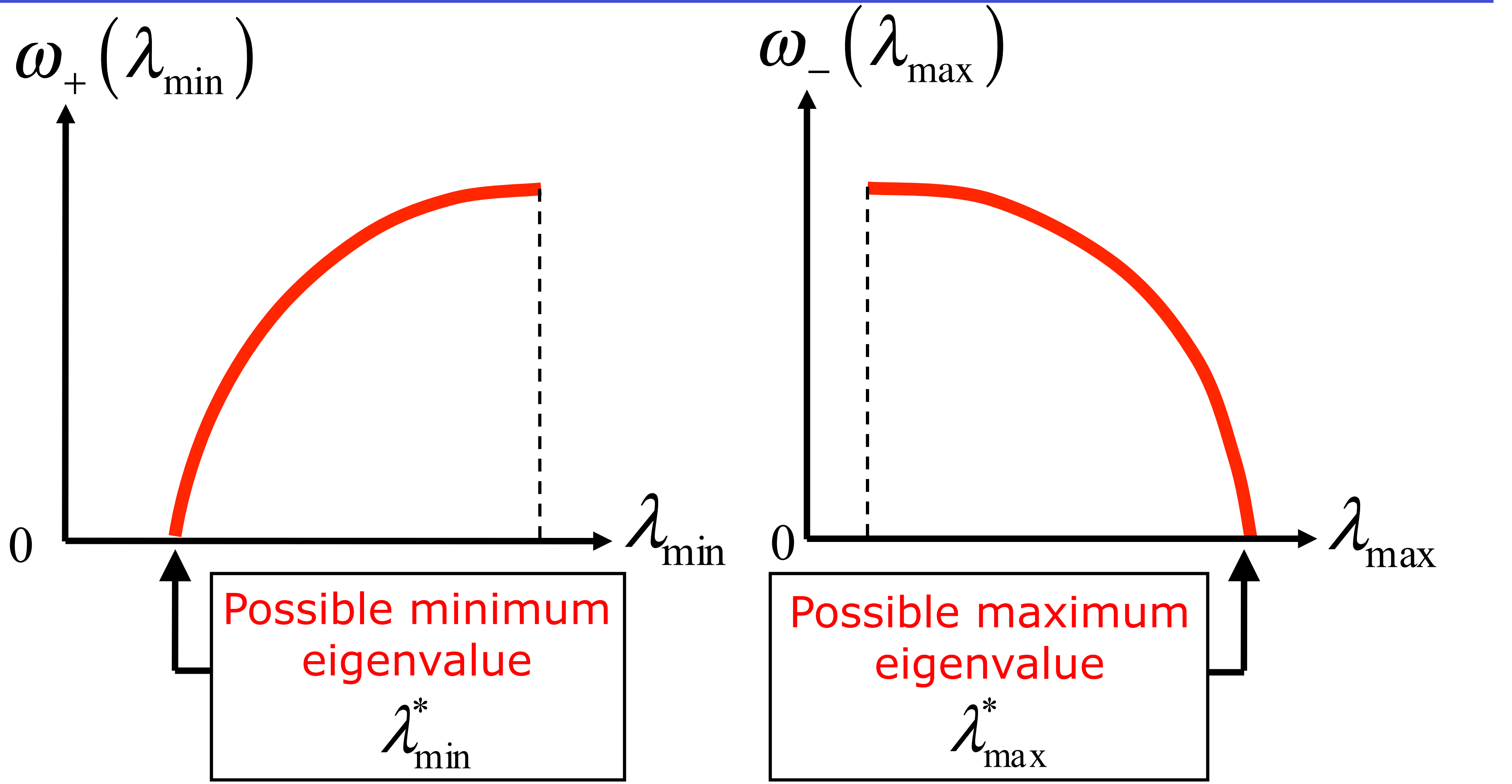
Energy (min/max eigenvalues)

$$\lambda(\mu) = -\frac{\partial \phi(\mu)}{\partial \mu}$$

Entropy (# of *T* producing λ)

$$\omega(\mu) = \phi(\mu) - \mu\frac{\partial \phi(\mu)}{\partial \mu}$$

Possible minimum and maximum eigenvalues



$$\delta_S = \max\{1 - \lambda^*_{\min}, \lambda^*_{\max} - 1\}$$

Replica symmetric analysis

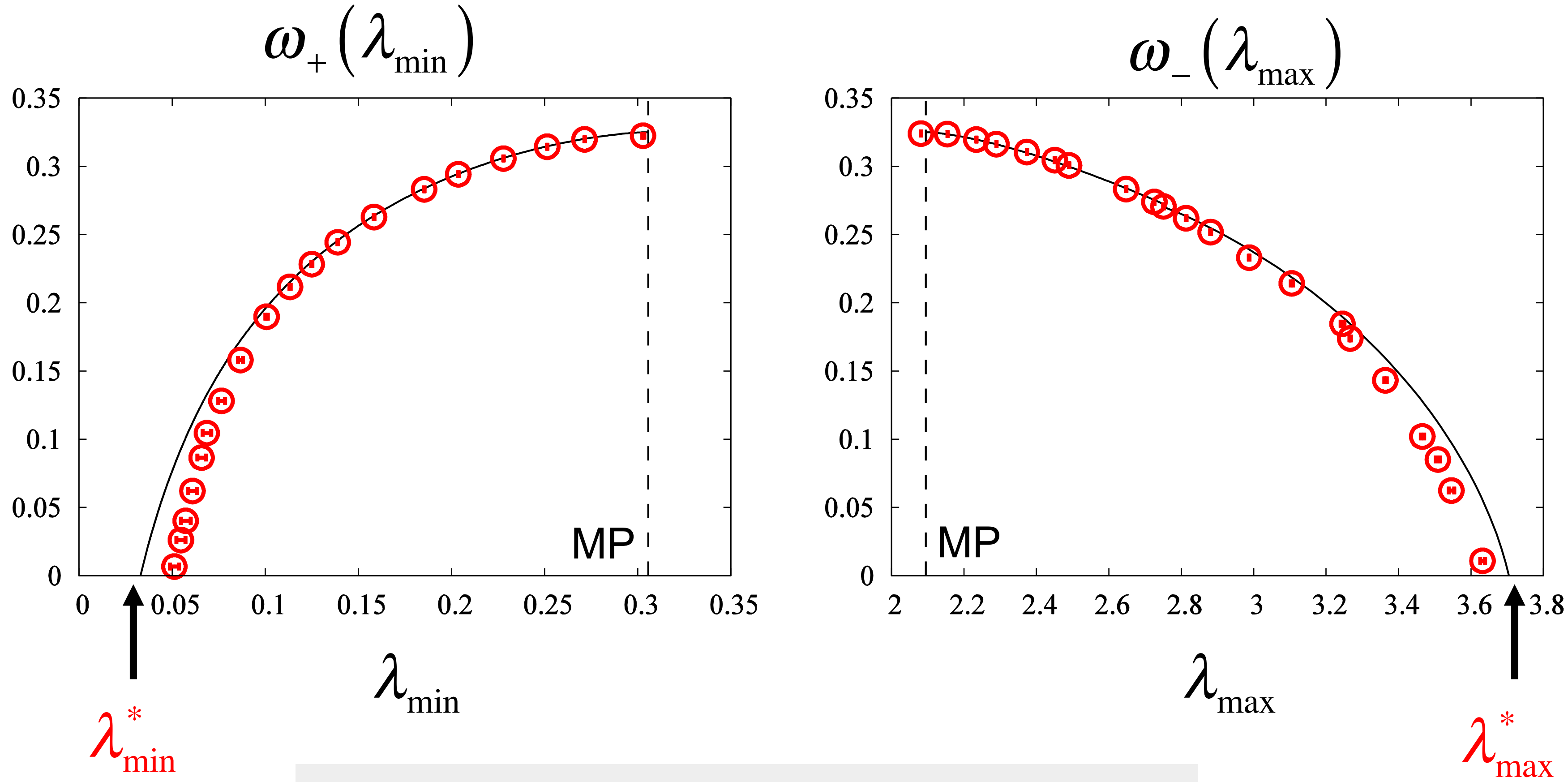
$$\phi(\mu) = -\frac{\alpha}{2}\log\{\alpha+\chi+\mu(1-q)\} + \frac{\alpha}{2}\log(\alpha+\chi)$$
$$-\frac{\alpha\mu q}{2\{\alpha+\chi+\mu(1-q)\}} + \frac{\hat{Q}}{2} - \frac{\hat{q}_1}{2}\left(1+\frac{\chi}{\mu}\right) + \frac{\hat{q}_0 q}{2} + K\rho$$
$$+\int Dz \log\left(1+e^{-K}\int Dy \exp\left(\frac{(\sqrt{\hat{q}_1-\hat{q}_0}\,y+\sqrt{\hat{q}_0}\,z)^2}{2\hat{Q}}\right)\right)$$

- $\alpha = M/N$ (compression rate)
- $\rho = S/N$ (fraction of non-zero components)
- $\{q,\chi,\hat{Q},\hat{q}_1,\hat{q}_0,K\}$ are determined by saddle point equations.

Replica symmetric entropy

- $\alpha = 0.5, \rho = 0.1$

※ $\alpha = \dfrac{M}{N}, \rho = \dfrac{S}{N}$

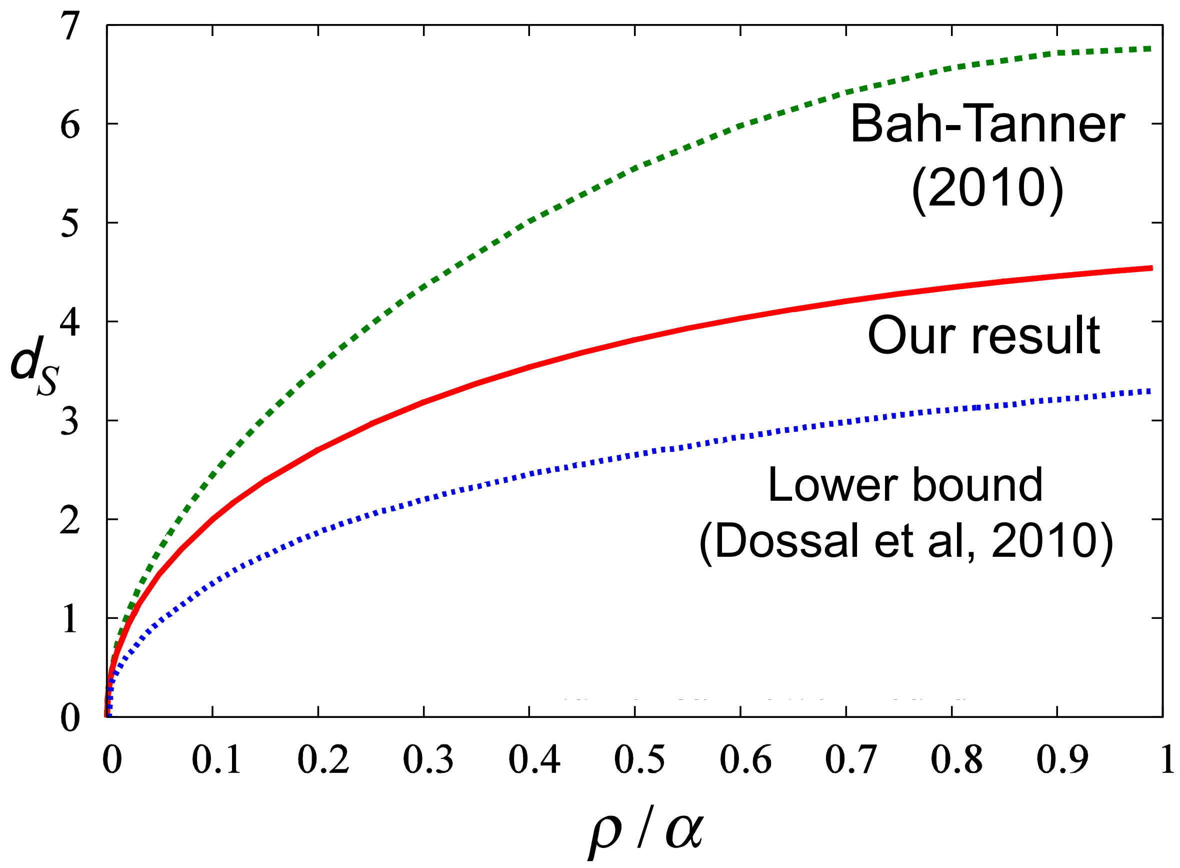


$$\delta_S = \max\{1 - \lambda^*_{\min}, \lambda^*_{\max} - 1\}$$

Replica symmetric RIC

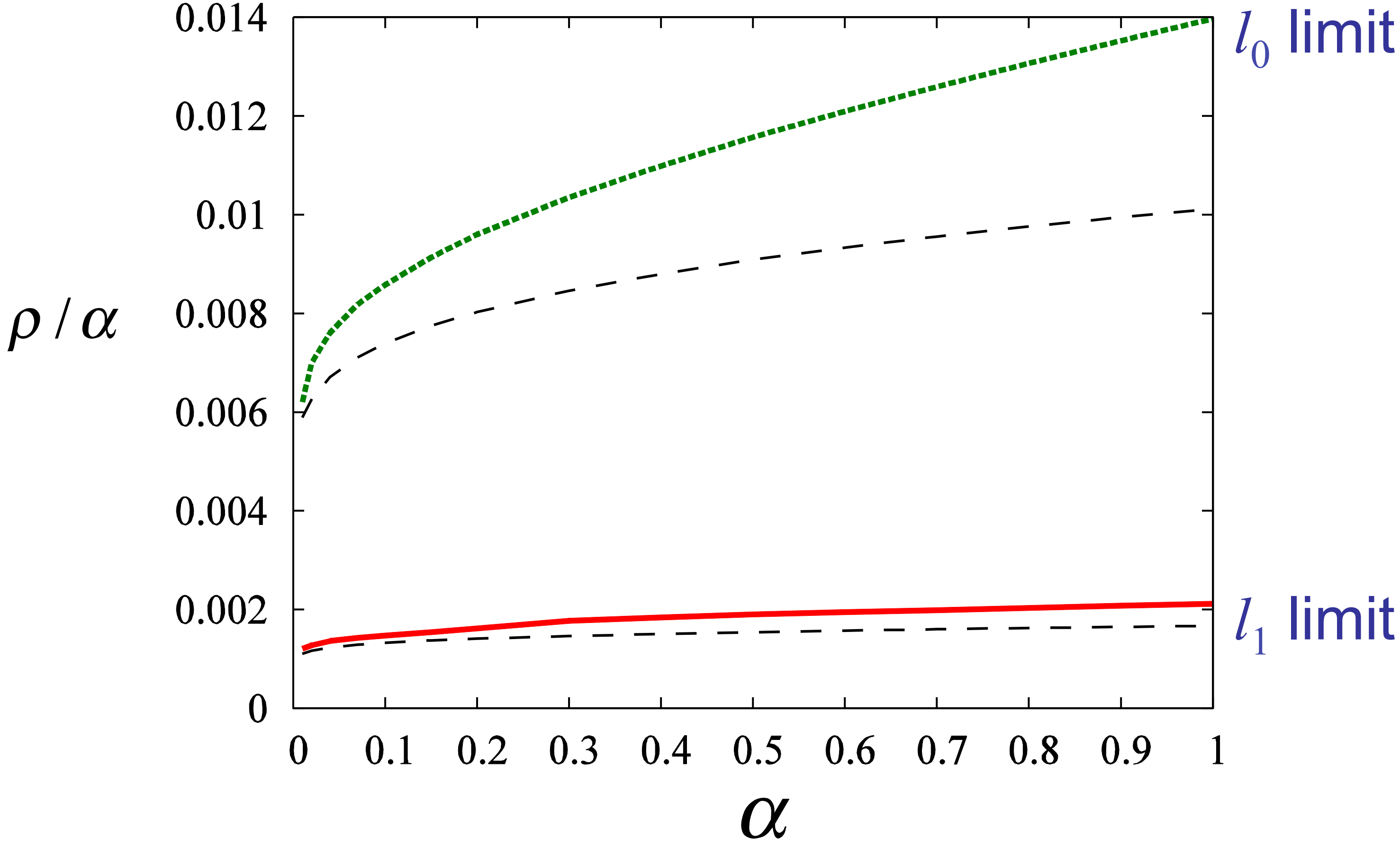
- d_S at $\alpha = 0.5$

※ $\alpha = \frac{M}{N}, \rho = \frac{S}{N}$



l_0, l_1 reconstruction limit

※ $\alpha = \frac{M}{N}, \rho = \frac{S}{N}$



※ Dashed lines are Bah – Tanner (2010)

Improvement by RSB

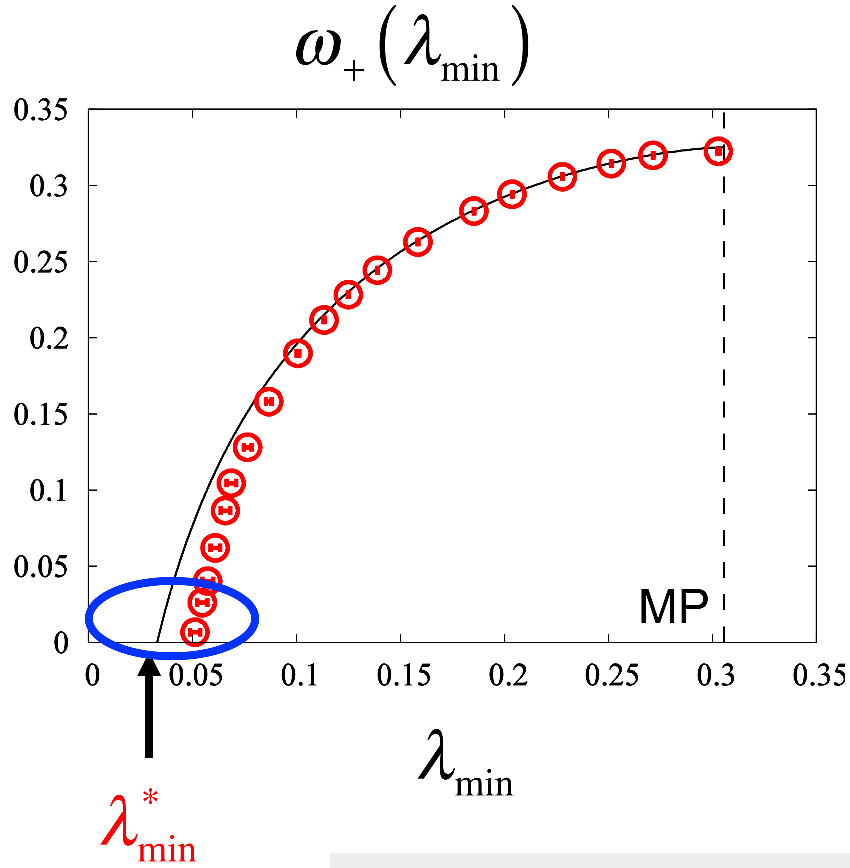
- The RS RIC estimate is lower than *any existing upper bounds,* being consistent with a known lower bound.
- On the other hand, there are non-negligible deviations from the experimental data.

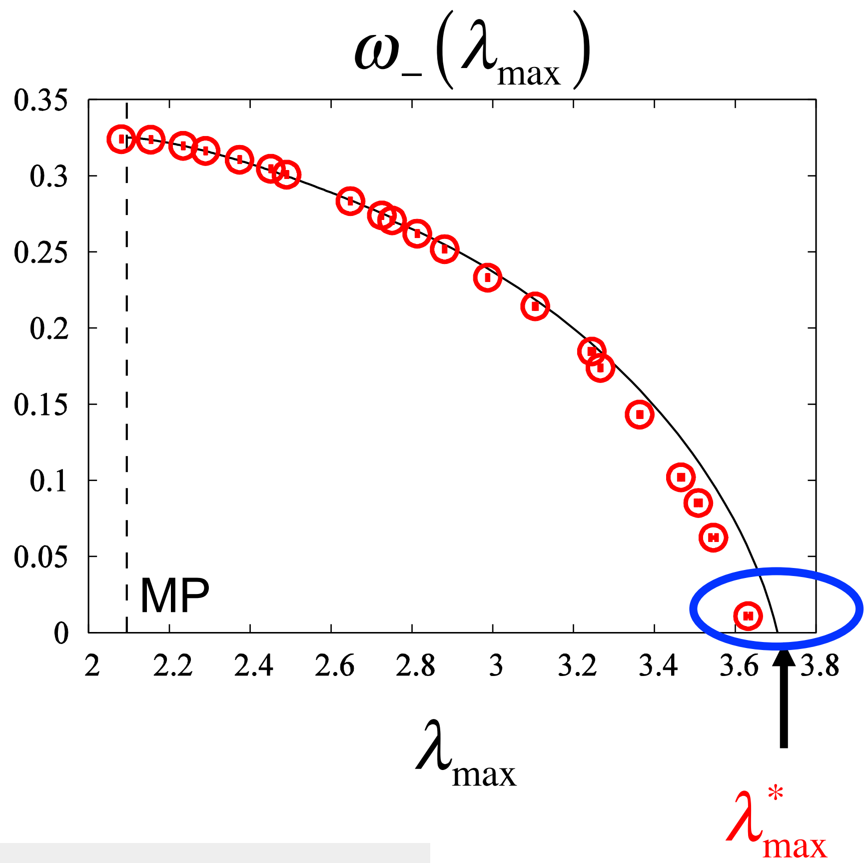
- In fact, detailed analysis shows that the replica symmetry is broken for the left and right edges of the entropy curves.
- However, physical interpretation of RSB indicates that the RS estimates still serve as *upper bounds* of RIC, and the bounds are improved as the higher RSBs are taken into account.

Replica symmetric entropy (again)

- $\alpha = 0.5, \rho = 0.1$

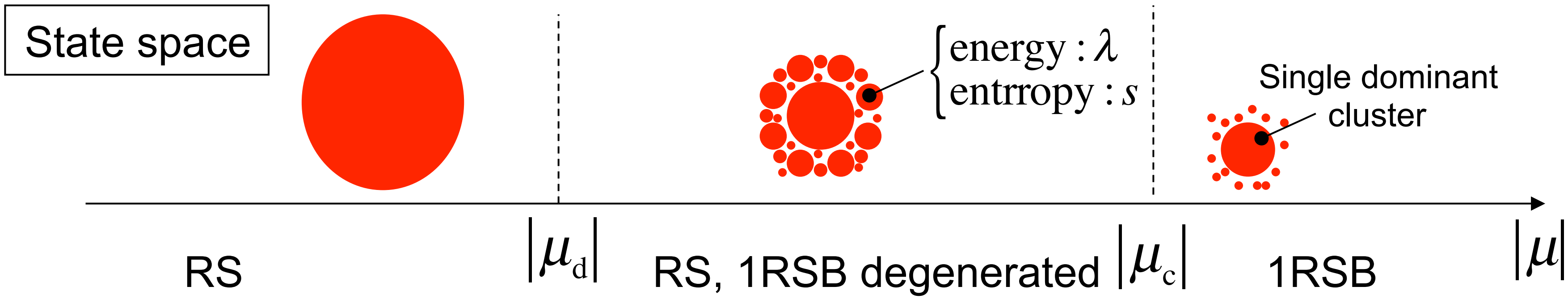
※ $\alpha = \frac{M}{N}, \rho = \frac{S}{N}$





$$\delta_S = \max\{1 - \lambda^*_{\min}, \lambda^*_{\max} - 1\}$$

Physical interpretation of RSB



Complexity = entropy of pure states

$$\Sigma(\lambda,s)=\frac{1}{N}\log\left(\#\text{pure states specified by } (\lambda,s)\right)$$

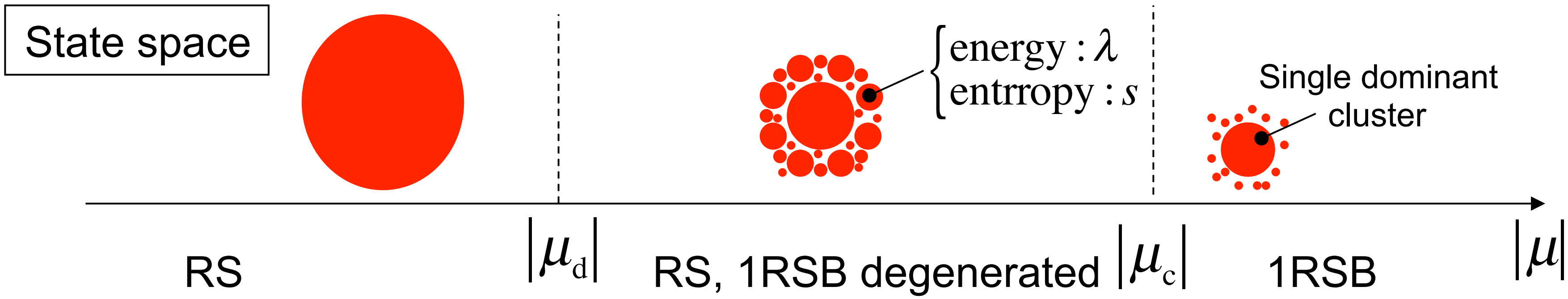
Non-negativity constraint: Must be non-negative.

Total entropy

$$\omega(\lambda)=\frac{1}{N}\log\left(\int ds\exp\left(N\left(s+\Sigma(\lambda,s)\right)\right)\right)\simeq\max_{s}\left\{s+\Sigma(\lambda,s)\right\}$$

Cf) Montanari and Ricci-Tersenghi (2003), Krzakala et al (2007), …

Physical interpretation of RSB



RS evaluation: the complexity constraint is <u>ignored.</u>

$$\omega_{\mathrm{RS}}(\lambda) = \max_{s}\left\{ s + \Sigma(\lambda, s) \right\}$$

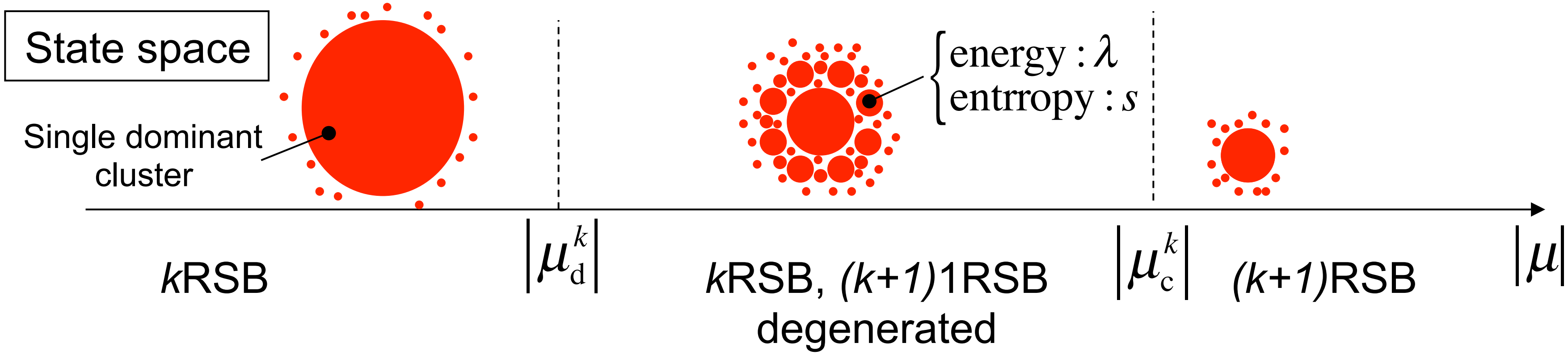
1RSB evaluation: the complexity constraint is <u>taken into account</u>.

$$\omega_{\mathrm{1RSB}}(\lambda) = \max_{s}\left\{ s + \Sigma(\lambda, s) \middle| \Sigma(\lambda, s) \geq 0 \right\}$$

The non-negativity constraint of the 1RSB evaluation means

$$\omega_{\mathrm{RS}}(\lambda) \geq \omega_{\mathrm{1RSB}}(\lambda).$$

Physical interpretation of RSB



If necessary, a similar argument may be applied for the dominant cluster of 1RSB, which yields

$$\omega_{\mathrm{RS}}(\lambda) \geq \omega_{\mathrm{1RSB}}(\lambda) \geq \omega_{\mathrm{2RSB}}(\lambda).$$

Applying the similar argument repeatedly concludes

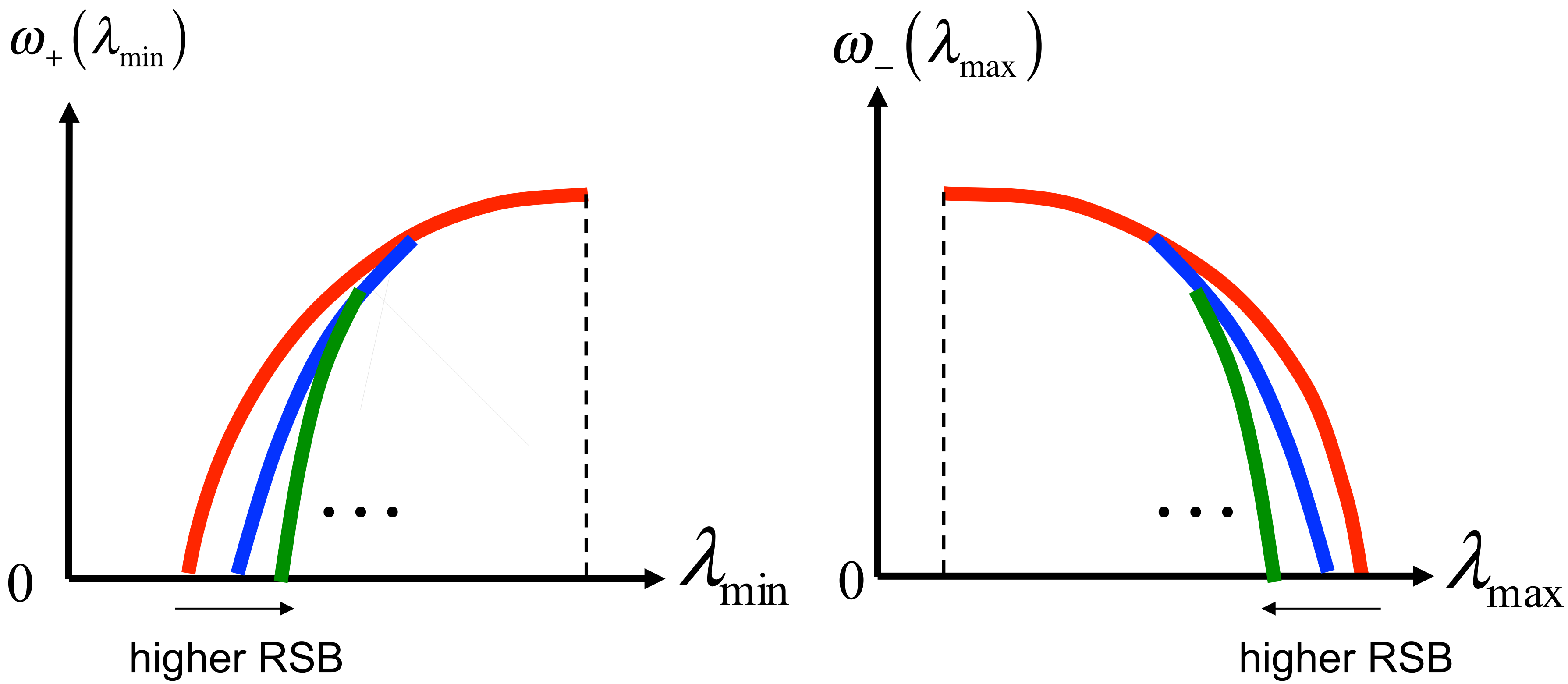
$$\omega_{\mathrm{RS}}(\lambda) \geq \omega_{\mathrm{1RSB}}(\lambda) \geq \omega_{\mathrm{2RSB}}(\lambda) \geq \ldots.$$

Monotonic improvement by RSB

The series of inequalities

$$\omega_{\mathrm{RS}}(\lambda) \geq \omega_{\mathrm{1RSB}}(\lambda) \geq \omega_{\mathrm{2RSB}}(\lambda) \geq \ldots.$$

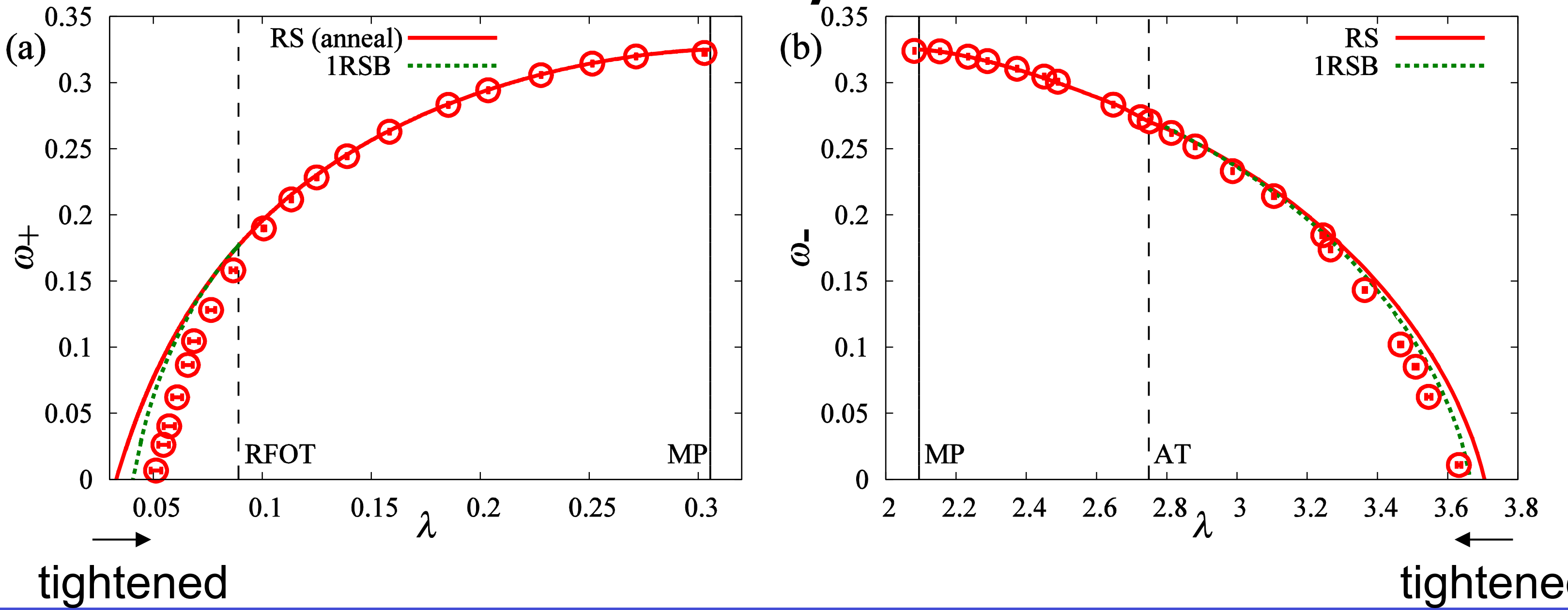
indicates that bounds are monotonically improved by incorporating the higher RSB.



1RSB results

- The estimates are actually improved by the 1RSB solution!
- Two scenarios for the RSB transition
  - λ_{min} : Random first order transition (RFOT)
  - λ_{max} : de Almeida-Thouless instability (full RSB)

$$\alpha = 0.5, \rho = 0.1$$



Summary

- Evaluation of the restricted isometry constant (RIC) can be formulated as a spin glass problem.
- We provided a replica based-framework for accurate evaluation of RICs.
  - Replica evaluation provides the current best accuracy.
  - Although the RS solution is not thermodynamically stable, the physical interpretation of RSB implies that the RS estimates still have the meaning of “upper-bounds”.
  - The bounds are monotonically improved by taking the higher RSB into account.

- Future work
  - Application to other matrix ensembles
  - Mathematical justification

Thank you for your attention.

- Acknowledgments
  - RIKEN special postdoctoral researcher (SPDR) fellowship, KAKENHI No. 26880028 (AS)
  - KAKENHI No. 25120013 (YK)