Spatial coupling: a tool for rigorous analysis in Bayesian inference

Nicolas Macris

School of Computer and Communication Science
Ecole Polytechnique Fédérale de Lausanne

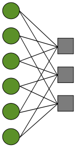
Plan

- Brief flavour of "spatial coupling" through:
  - Its origin in theory of error correcting codes for channel communication.
  - An application to lower bounds on thresholds of constraint satisfaction problems.
- Main focus of the talk: proof technique for replica formulas
  - Example of rank-one matrix factorisation.

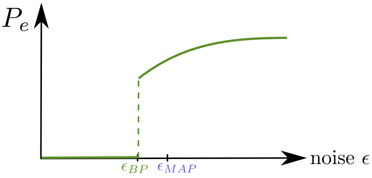
    [joint work with J. Barbier, M. Dia, F. Krzakala, T. Lesieur, and L. Zdeborova]

Spatial Coupling in coding theory: cartoon view

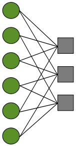
- standard LDPC code

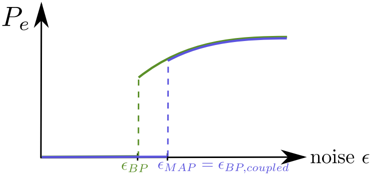


Belief Propagation decoding of transmitted bits $\rightarrow$ Prob error

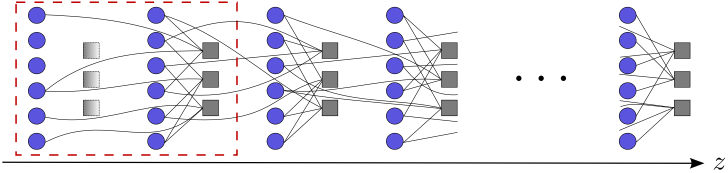


- standard LDPC code





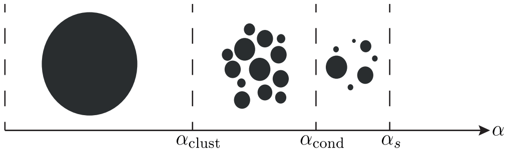
- Spatially Coupled LDPC code - bits at boundary are know and act as a seed



An engineering construction of Felstroem, Zigangirov (1999); Proofs that these codes are capacity achieving on general classes of channels [S.Kudekar, T.Richardson, R.Urbanke (2013) and H.Pfister, S.Kumar, A.Young, N.M (2013)]

Application to lower bounds for thresholds of constraint satisfaction problems

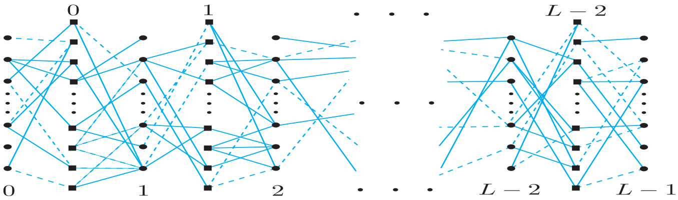
[*by the cavity method*: Parisi-Mézard-Zecchina 2002, Montanari-Ricci Tersenghi-Semerjian, Krzakala-Zdeborova 2007]



- Clustered phase is a problem for proving algorithmic lower bounds on α_s.
- Idea: remove clustered phase by spatial coupling - at the same time maintain α_s.
- Analyze simple algorithms to get lower bounds.

Coupled K-SAT model:

Chain of K-SAT models with constraints at boundaries removed.



Lower bounds by running Unit Clause:

K	3	4	$\cdots$	large K
$\alpha_{\text{spa}-\text{coup}}$	3.67	7.81	$\cdots$	$\sim 2^K \times \frac{1}{2}$
usual algor bound	3.52	5.54	$\cdots$	$\sim 2^K \times \frac{\ln K}{K}$

[D. Achlioptas, H. Hassani, R. Urbanke, N.M (SODA 2016)]

Rank one matrix factorisation: proof of the replica predictions

We are given noisy observations w_{ij} of pairwise products $s_i s_j$ for $\mathbf{s} = (s_1, \cdots, s_n)$ with i.i.d $s_i \sim P_0$ known,

$$w_{ij} = \frac{s_i s_j}{\sqrt{n}} + z_{ij}\sqrt{\Delta},$$

where $Z_{ij} \sim \mathcal{N}(0, 1)$, $1 \leq i \leq j \leq n$ (symmetric i.i.d).

Bayesian formulation:

$$p(\mathbf{X}|\mathbf{W}) = \frac{1}{P(\mathbf{W})} P_0(x_i) \exp\left(- \frac{1}{2\Delta} \sum_{i \leq j} (w_{ij} - x_i x_j / \sqrt{n})^2 \right)$$

Key quantities

- Optimal "MMSE estimators" are averages with respect to this Gibbs distribution:

$$\mathbb{E}[\mathbf{X}|\mathbf{W}] = \langle \mathbf{X} \rangle, \qquad \mathbb{E}[\mathbf{X}\mathbf{X}^T|\mathbf{W}] = \langle \mathbf{X}\mathbf{X}^T \rangle$$

- Matrix-MMSE and Mutual Information (I-MMSE formula):

$$\frac{d}{d\Delta^{-1}} \frac{1}{n} \mathrm{I}(\mathbf{S};\mathbf{W}) = \frac{1}{4n^2} \mathbb{E}_{\mathbf{S},\mathbf{W}} \|\mathbf{S}\mathbf{S}^T - \langle \mathbf{X}\mathbf{X}^T \rangle\|_F^2$$

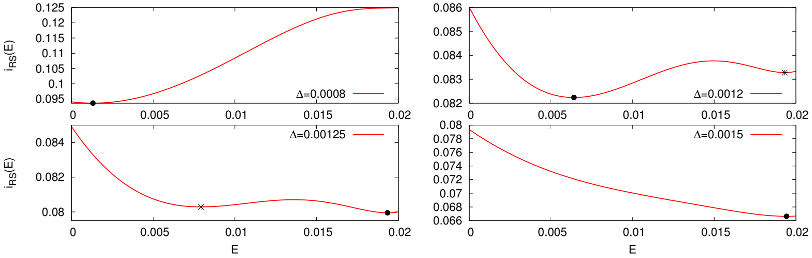
where

$$\mathrm{I}(\mathbf{S};\mathbf{W}) = \underbrace{H(\mathbf{W})}_{-\mathbb{E}_{\mathbf{W}} \log P(\mathbf{W})} - \underbrace{H(\mathbf{W}|\mathbf{S})}_{\text{trivial term}}$$

Replica symmetric potential

$$i_{\mathrm{RS}}(E;\Delta) = \frac{E^2}{4\Delta} + \mathrm{I}(S;S+\Sigma Z)$$
$$= \frac{(v-E)^2+v^2}{4\Delta} - \mathbb{E}_{S,Z}\left[\ln\left(\int dx\, P_0(x) e^{-\frac{x^2}{2\Sigma^2}+x\left(\frac{S}{\Sigma^2}+\frac{Z}{\Sigma}\right)}\right)\right]$$

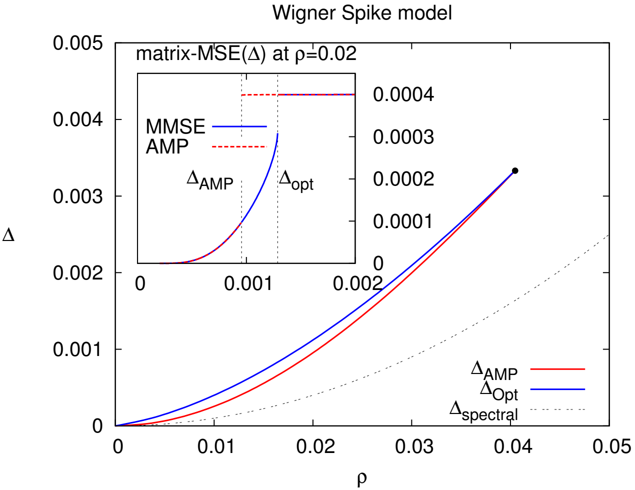
with $Z\sim\mathcal{N}(0,1)$, $S\sim P_0$, $\mathbb{E}[S^2]=v$ and $\Sigma^2 := \Delta/(v-E)$.



$\mathrm{argmin}_E\, i_{\mathrm{RS}}$ is analytic in Δ except at Δ_{RS} where it has a jump discontinuity.

Phase diagram: example of the spiked Wigner model

$$P_0(s) = (1-\rho)\delta(s) + \rho\delta(s-1).$$



Theorem (One letter formula for the mutual information)

Assume $P_0(s) = \sum_{\alpha=1}^{\nu} p_\alpha \delta(s - a_\alpha)$ *and* $i_{\rm RS}(E;\Delta)$ *has at most three stationary points.*

$$\lim_{n\to+\infty} \frac{1}{n} \mathrm{I}(\mathbf{S};\mathbf{W}) = \min_{E\in[0,v]} i_{\rm RS}(E;\Delta)$$

Corollary (Exact formula for the MMSE)

Consider $\mathrm{Mmmse}_n := \frac{1}{n^2}\mathbb{E}_{\mathbf{S},\mathbf{W}} \| \mathbf{SS}^\intercal - \langle \mathbf{XX}^\intercal \rangle \|_F^2$*. For* $\Delta \neq \Delta_{\rm RS}$

$$\lim_{n\to+\infty} \mathrm{Mmmse}_n = v^2 - (v - \widetilde{E}(\Delta))^2$$

where $\widetilde{E}(\Delta) = \mathrm{argmin}_{E\in[0,v]} i_{\rm RS}(E;\Delta)$.

Remark: General bound $\leq$ [Krzakala, Zdeborova, Xu] and $=$ when there is no phase transition [Deshpande, Montanari].

Performance of Approximate Message Passing

Estimate $\langle \mathbf{X} \rangle$ by the AMP iterative algorithm $\to \hat{\mathbf{S}}^t$.

MSE of AMP $E^t := \frac{1}{n} \lim_{n\to+\infty} \mathbb{E}_{\mathbf{S},\mathbf{Z}}[\|\mathbf{S} - \hat{\mathbf{S}}^t\|_2^2]$ is tracked by State Evolution equation [M. Bayati, A. Montanari]

$$E^{t+1} = \mathrm{mmse}(\Sigma(E^t; \Delta)^{-2}), \qquad E^0 = v,$$

where $\mathrm{mmse}(\Sigma^{-2})$ is associated to the scalar AWGN channel

$$Y = S + \Sigma Z, \qquad Z \sim \mathcal{N}(0,1) \text{ and } \Sigma^2 = \Delta/(v - E).$$

Connection with RS solution:

$$E = \mathrm{mmse}(\Sigma(E; \Delta)^{-2}) \leftrightarrow \frac{d}{dE} i_{\mathrm{RS}}(E; \Delta) = 0$$

So performance of AMP is governned by the RS potential.

Definition (AMP algorithmic threshold)

$\Delta_{\rm AMP}$ is the first noise threshold s.t fixed point equation has a non-unique solution.

This is also a spinodal point or solution of

$$\frac{d}{dE} i_{\rm RS}(E;\Delta) = \frac{d^2}{dE^2} i_{\rm RS}(E;\Delta) = 0.$$

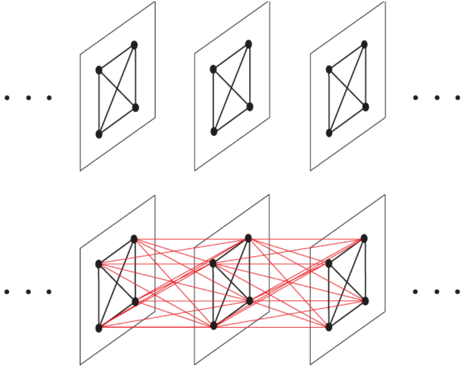
Corollary (Performance of AMP)

AMP yields upon convergence the matrix-MMSE as well as the vector-MMSE iff $\Delta < \Delta_{\rm AMP}$ *or* $\Delta > \Delta_{\rm RS}$*:*

$$E^{\infty} = argmin_{E\in[0,v]} i_{\rm RS}(E;\Delta).$$

Spatially coupled model I: periodic construction

Ring of length $L+1$ with *blocks* positioned at $\mu \in \{0, \ldots, L\}$ and coupled to neighboring blocks $\{\mu - w, \ldots, \mu + w\}$.



$$w_{i_\mu j_\nu} = s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda_{\mu\nu}}{n}} + z_{i_\mu j_\nu} \sqrt{\Delta},$$

Spatially coupled model II: seeded construction

The construction is completed by introducing a seed in the ring:

Assume perfect knowledge of the signal components

$$s_{i_\mu} \text{ for } \mu \in \{-w-1, \ldots, +w-1\} \mod L+1$$

This seed allows to close the gap between the algorithmic and information theoretical limits for the chain.

Two fundamental claims

Lemma (Equality of mutual informations)

For any fixed w the following limits exist and are equal:

$$\lim_{n\to+\infty} I^{\text{periodic}}_{w,L}(\boldsymbol{S};\boldsymbol{W})/nL = \lim_{n\to+\infty} \mathrm{I}(\boldsymbol{S};\boldsymbol{W})/n$$

$$\lim_{L\to+\infty}\lim_{n\to+\infty} I^{\text{seeded}}_{w,L}(\boldsymbol{S};\boldsymbol{W})/nL = \lim_{n\to+\infty} \mathrm{I}(\boldsymbol{S};\boldsymbol{W})/n$$

Lemma (Threshold saturation)

Let $\Delta_{\text{AMP,coup}} := \liminf_{w\to+\infty} \liminf_{L\to+\infty} \Delta_{\text{AMP},w,L}$. *We have*

$$\Delta_{\text{AMP,coup}} \geq \Delta_{\text{RS}}$$

First lemma: A Guerra style interpolation method.

MI is the free energy associated to

$$\mathcal{H}(\mathbf{x},\mathbf{z},\Lambda) = \frac{1}{\Delta}\sum_{\mu=0}^{L}\Lambda_{\mu\mu}\sum_{i_\mu\le j_\mu}\Big(\frac{x_{i_\mu}^2 x_{j_\mu}^2}{2n} - \frac{s_{i_\mu}s_{j_\mu}x_{i_\mu}x_{j_\mu}}{n} - \frac{x_{i_\mu}x_{j_\mu}z_{i_\mu j_\mu}\sqrt{\Delta}}{\sqrt{n\Lambda_{\mu\mu}}}\Big)$$

$$+ \frac{1}{\Delta}\sum_{\mu=0}^{L}\sum_{\nu=\mu+1}^{\mu+w}\Lambda_{\mu\nu}\sum_{i_\mu,j_\nu}\Big(\frac{x_{i_\mu}^2 x_{j_\nu}^2}{2n} - \frac{s_{i_\mu}s_{j_\nu}x_{i_\mu}x_{j_\nu}}{n} - \frac{x_{i_\mu}x_{j_\nu}z_{i_\mu j_\nu}\sqrt{\Delta}}{\sqrt{n\Lambda_{\mu\nu}}}\Big).$$

Take a pair of systems Λ and Λ' and an *interpolated Hamiltonian*, $t\in[0,1]$, $\mathcal{H}(\mathbf{x},\mathbf{z},t\Lambda)+\mathcal{H}(\mathbf{x},\mathbf{z}',(1-t)\Lambda')$.

After differentiation, Gaussian integration by parts, and Nishimori identities:

$$-\frac{1}{nL}\frac{d}{dt}\mathbb{E}_{\mathbf{s},\mathbf{z},\mathbf{z}'}[\ln\mathcal{Z}_t] = \frac{1}{4\Delta L}\mathbb{E}_{\mathbf{s},\mathbf{z},\mathbf{z}'}[\langle\mathbf{q}^\intercal\Lambda\mathbf{q} - \mathbf{q}^\intercal\Lambda'\mathbf{q}\rangle_t] + \mathcal{O}(\frac{1}{nL})$$

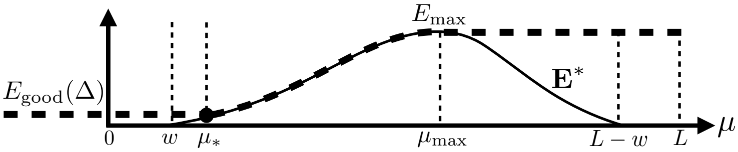
where $q_\mu = \frac{1}{n}\sum_{i_\mu=1}^{n} s_{i_\mu}x_{i_\mu}$. Choose suitable Λ and Λ' to compare quadratic forms and get

$$\lim_n \frac{1}{nL} I_{w=0,L} \le \lim_n \frac{1}{nL} I_{w,L} \le \lim_n \frac{1}{nL} I_{w=L/2,L}$$

Threshold saturation: rough sketch

Coupled AMP is controlled by the coupled version of RS potential: $i_{w,L}(\{E_\mu\}_0^L; \Delta)$.

<u>Main intuition</u>: initial MSE profile is driven to flat $E_{\rm good}$ for all $\mu \in \{0, \cdots, L\}$, because it is energetically favourable for the seed grows and a reconstruction wave develops.



<u>Proof idea</u>: suppose there is an increasing fixed point of state evolution and show that a small shift lowers the free energy.

$$i_{w,L}(\mathcal{S}(\mathbf{E}^{\rm s}); \Delta) - i_{w,L}(\mathbf{E}^{\rm s}; \Delta) \leq -|i_{\rm RS}(E_{\rm bad}; \Delta) - i_{\rm RS}(E_{\rm good}; \Delta)|$$

Back to uncoupled system: proof of RS formula

Consequence of the two lemmas:

$$\Delta_{\rm RS} \underbrace{\leq}_{\rm threshold\ saturation} \Delta_{\rm AMP,coup} \underbrace{\leq}_{\rm sub-optimal\ alg} \Delta_{\rm Opt,coup} \underbrace{=}_{\rm equality\ MI} \Delta_{\rm Opt}$$

We can also argue that:

$$\Delta_{\rm Opt} \leq \Delta_{\rm RS}$$

- This argument combines sub optimality of AMP and a "Guerra-Toninelli interpolation bound" for original (uncoupled) system.
- In summary we get $\Delta_{\rm Opt} = \Delta_{\rm RS}$. In the process we get the whole replica formula as a bonus.

Proof of $\Delta_{\rm Opt} \leq \Delta_{\rm RS}$

- We use $\quad \text{Matrix MSE}_{\rm AMP} \geq \text{Matrix MMSE}$:

$$\frac{\partial i_{\rm RS}(E^{\infty};\Delta)}{\partial \Delta} \geq \limsup_{n\to+\infty} \frac{1}{n}\frac{d\mathrm{I}(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}$$

- $\Delta \leq \Delta_{\rm AMP} \to E^{\infty} = \text{argmin}\, i_{RS}(E,\Delta)$ thus

$$\frac{d}{d\Delta^{-1}} \min_{E} i_{\rm RS}(E;\Delta) \geq \limsup_{n\to+\infty} \frac{1}{n}\frac{d\mathrm{I}(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}.$$

- Integrate on $[0,\Delta] \subset [0,\Delta_{\rm AMP}]$ to obtain

$$\min_{E\in[0,v]} i_{\rm RS}(E;\Delta) \leq \liminf_{n\to+\infty} \mathrm{I}(\mathbf{S};\mathbf{W})/n$$

- But we know "$\geq$" by Guerra-Toninelli interpolation.

Thus we already get equality for $\Delta \leq \Delta_{\rm AMP}$.

Proof of $\Delta_{\mathrm{Opt}} \leq \Delta_{\mathrm{RS}}$ continued

- By definition of thresholds we know already $\Delta_{\mathrm{AMP}} \leq \Delta_{\mathrm{RS}}$.
- Suppose $\Delta_{\mathrm{AMP}} \leq \Delta_{\mathrm{RS}} < \Delta_{\mathrm{Opt}}$. Then, by analyticity, both $\lim_{n\to\infty} \mathrm{I}(\mathbf{S};\mathbf{W})/n$ and $\min_E i_{\mathrm{RS}}(E;\Delta)$ are equal up to Δ_{RS}.
- Repeat the integration argument on $[\Delta_{\mathrm{RS}}, \Delta]$:

$$\int_{\Delta_{\mathrm{RS}}}^{\Delta} d\Delta \frac{d}{d\Delta} \min_E i_{\mathrm{RS}}(E;\Delta) \leq \liminf_{n\to+\infty} \frac{1}{n} \int_{\Delta_{\mathrm{RS}}}^{\Delta} d\Delta \frac{d\mathrm{I}(\mathbf{S};\mathbf{W})}{d\Delta}.$$

which leads again to

$$\min_{E\in[0,v]} i_{\mathrm{RS}}(E;\Delta) \leq \liminf_{n\to+\infty} \mathrm{I}(\mathbf{S};\mathbf{W})/n$$

- and to equality with the help of Guerra-Toninelli: $\geq$
- But equality for all $\Delta \in [0, \Delta_{\mathrm{Opt}}[$ with $\Delta_{\mathrm{RS}} < \Delta_{\mathrm{Opt}}$ means $\min_E i_{\mathrm{RS}}(E;\Delta)$ is analytic at Δ_{RS} which is a contradiction.

To summarize:

- The integration argument gave RS formula for $\Delta < \Delta_{\mathrm{AMP}}$.
- We managed to show $\Delta_{\mathrm{RS}} = \Delta_{\mathrm{Opt}}$.
- By analyticity the RS formula extends up to Δ_{Opt}.
- A final integration argument on $[\Delta_{\mathrm{Opt}}, \Delta]$ finishes the proof.

Remark:

- the integration argument does not work for $\Delta \in [\Delta_{\mathrm{AMP}}, \Delta_{\mathrm{RS}}]$.

$$\underbrace{\frac{\partial i_{\mathrm{RS}}(E^{\infty};\Delta)}{\partial \Delta^{-1}}}_{\neq \frac{d}{d\Delta^{-1}} \min_E i_{\mathrm{RS}}(E;\Delta)} \geq \limsup_{n\to+\infty} \frac{1}{n} \frac{d\mathrm{I}(\mathbf{S};\mathbf{W})}{d\Delta^{-1}}$$

Conclusion

- The ideas are more or less generic.
- Proof of RS formulas in coding LDPC, LDGM (spin systems on sparse graph on their Nishimori line). [A. Giurgiu, R. Urbanke, H. Pfister, S. Kumar, A.Young, N.M]
- In linear estimation/ compressive sensing [J. Barbier, M. Dia, F. Krzakala, N.M]. [case without phase transition: Montanari-Tse 2006; other recent proof G. Reeves, H. Pfister]
- New provable lower bounds on K-SAT, Q-COL thresholds [D. Achlioptas, H. Hassani, R. Urbanke, N.M].
- Work in progress: non-linear estimation, low rank matrix factorisation etc.