Generalized Approximate Survey Propagation for High-dimensional Estimation

Microsoft Research

Luca Saglietti

IIGM Italian Institute for Genomic Medicine

Yue Lu, *Harvard University*

Carlo Lucibello, *Bocconi University*

Outline

- Generalized Linear Models (GLM)
  - Real-valued phase retrieval
  - Inference model
- Approximate message-passing
  - Effective landscapes and competition
  - Breaking the replica symmetry
  - Changing the effective landscape
- Conclusions

Generalized Linear Models

3 ingredients :

$\mathbb{R}^N$ ⟵ **TRUE SIGNAL** : $x_{0,i} \sim P_{X_0}$

$\mathbb{R}^{M\times N}$ ⟵ **OBSERVATION MATRIX** : $F_i^\mu \sim \mathcal{N}(0, 1/N)$

$\mathbb{R}^M$ ⟵ **OBSERVED SIGNAL** : $y^\mu \sim P_{out}(\cdot|F^\mu \cdot x_0)$

High-dimensional limit: $N \to \infty$

with $\alpha = M/N$ of $\mathcal{O}(1)$

Generalized Linear Models

3 ingredients :

$\mathbb{R}^N$	$\longleftarrow$	**TRUE SIGNAL** :	$x_{0,i} \sim P_{X_0}$
$\mathbb{R}^{M\times N}$	$\longleftarrow$	**OBSERVATION MATRIX** :	$F_i^\mu \sim \mathcal{N}(0, 1/N)$
$\mathbb{R}^M$	$\longleftarrow$	**OBSERVED SIGNAL** :	$y^\mu \sim P_{out}(\cdot \vert F^\mu \cdot x_0)$

High-dimensional limit: $N \to \infty$

with $\alpha = M/N$ of $\mathcal{O}(1)$

Generalized Linear Models

3 ingredients :

$\mathbb{R}^N$ ⟵ **TRUE SIGNAL** : $x_{0,i} \sim P_{X_0}$

$\mathbb{R}^{M \times N}$ ⟵ **OBSERVATION MATRIX** : $F_i^\mu \sim \mathcal{N}(0, 1/N)$

$\mathbb{R}^M$ ⟵ **OBSERVED SIGNAL** : $y^\mu \sim P_{out}(\cdot | F^\mu \cdot x_0)$

High-dimensional limit: $N \to \infty$

with $\alpha = M/N$ of $\mathcal{O}(1)$

Generalized Linear Models

3 ingredients :

$\mathbb{R}^N$	$\longleftarrow$	**TRUE SIGNAL** :	$x_{0,i} \sim P_{X_0}$
$\mathbb{R}^{M \times N}$	$\longleftarrow$	**OBSERVATION MATRIX** :	$F_i^\mu \sim \mathcal{N}(0, 1/N)$
$\mathbb{R}^M$	$\longleftarrow$	**OBSERVED SIGNAL** :	$y^\mu \sim P_{out}(\cdot \vert F^\mu \cdot x_0)$

High-dimensional limit: $N \to \infty$

with $\alpha = M/N$ of $\mathcal{O}(1)$

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad\qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad\qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad\qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

An example: Real-valued Phase Retrieval

$$P_{X_0} = \mathcal{N}(0,1) \qquad\qquad y^\mu = |F^\mu \cdot x_0| \text{ (+ noise)}$$

Fun facts about phase retrieval:

- Physically meaningful!
- $\mathbb{Z}_2$ symmetry in the signal space.
- $\alpha = 1$ should provide enough information for a perfect reconstruction.
- Gradient descent struggles to reconstruct the signal until $\alpha \sim 10$.
- Rigorous result about convexification in a $\alpha \sim \log N$ regime.

Inference Model

$$p(x) \sim e^{-\beta \mathcal{H}_{y,F}(x)}$$

GRAPHICAL MODEL

$$\mathcal{H}_{y,F}(x) = \sum_{\mu=1}^{M} \ell(y^\mu, F^\mu \cdot x) + \sum_{i=1}^{N} r(x_i)$$

Sensible choice:

$$\ell(y, z) = -\log P_{out}(y|z)$$
$$r(x) = -\log P_{X_0}(x)$$

MATCHED / MISMATCHED

Estimator $\hat{x}$:

Bayesian optimal: $\hat{x}_{BO} = \langle x \rangle_{\beta=1}$

Maximum a posteriori: $\hat{x}_{MAP} = \langle x \rangle_{\beta=\infty}$

Inference Model

$$p(x) \sim e^{-\beta \mathcal{H}_{y,F}(x)}$$

GRAPHICAL MODEL

$$\mathcal{H}_{y,F}(x) = \sum_{\mu=1}^{M} \ell(y^\mu, F^\mu \cdot x) + \sum_{i=1}^{N} r(x_i)$$

Sensible choice:

$$\ell(y, z) = -\log P_{out}(y|z)$$
$$r(x) = -\log P_{X_0}(x)$$

MATCHED / MISMATCHED

Estimator $\hat{x}$ **:**

Bayesian optimal: $\hat{x}_{BO} = \langle x \rangle_{\beta=1}$

Maximum a posteriori: $\hat{x}_{MAP} = \langle x \rangle_{\beta=\infty}$

Inference Model

$$p(x) \sim e^{-\beta \mathcal{H}_{y,F}(x)}$$

GRAPHICAL MODEL

$$\mathcal{H}_{y,F}(x) = \sum_{\mu=1}^{M} \ell(y^\mu, F^\mu \cdot x) + \sum_{i=1}^{N} r(x_i)$$

→ Sensible choice:

$$\ell(y,z) = -\log P_{out}(y|z)$$
$$r(x) = -\log P_{X_0}(x)$$

MATCHED / MISMATCHED

Estimator $\hat{x}$ **:**

Bayesian optimal: $\hat{x}_{BO} = \langle x \rangle_{\beta=1}$

Maximum a posteriori: $\hat{x}_{MAP} = \langle x \rangle_{\beta=\infty}$

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$?

Easy (if everything is i.i.d.)

$$p(x)$$

BP

Gaussian ansatz

rBP

Close on single-site quantities

AMP (TAP)

DEFINE 2 SCALAR INFERENCE CHANNELS:

Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

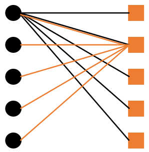
When do we get a **good estimator**?

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$?

Easy (if everything is i.i.d.)

$$p(x) \longrightarrow$$



BP

Gaussian ansatz

rBP

Close on single-site quantities

AMP (TAP)

DEFINE 2 SCALAR INFERENCE CHANNELS:

Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

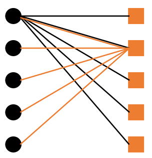
When do we get a **good estimator**?

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$?

Easy (if everything is i.i.d.)

$p(x)$ →



BP →

Gaussian ansatz

rBP →

Close on single-site quantities

AMP (TAP)

DEFINE 2 SCALAR INFERENCE CHANNELS:

Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

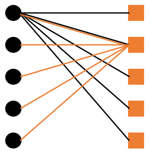
When do we get a **good estimator**?

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$?

Easy (if everything is i.i.d.)

$p(x)$ →



BP

→ Gaussian ansatz

rBP

→ Close on single-site quantities

AMP **(TAP)**

DEFINE 2 SCALAR INFERENCE CHANNELS:

Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

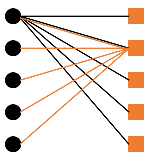
When do we get a **good estimator**?

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$?

Easy (if everything is i.i.d.)

$p(x)$ →



BP

→ Gaussian ansatz

rBP

→ Close on single-site quantities

AMP (TAP)

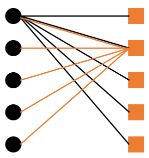
DEFINE 2 SCALAR INFERENCE CHANNELS:

Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

When do we get a **good estimator**?

Approximate Message-passing

How do we obtain $\hat{x}_{MAP}$? Easy (if everything is i.i.d.)

$p(x)$ →  → Gaussian ansatz → Close on single-site quantities

BP **rBP** **AMP (TAP)**

DEFINE 2 SCALAR INFERENCE CHANNELS:

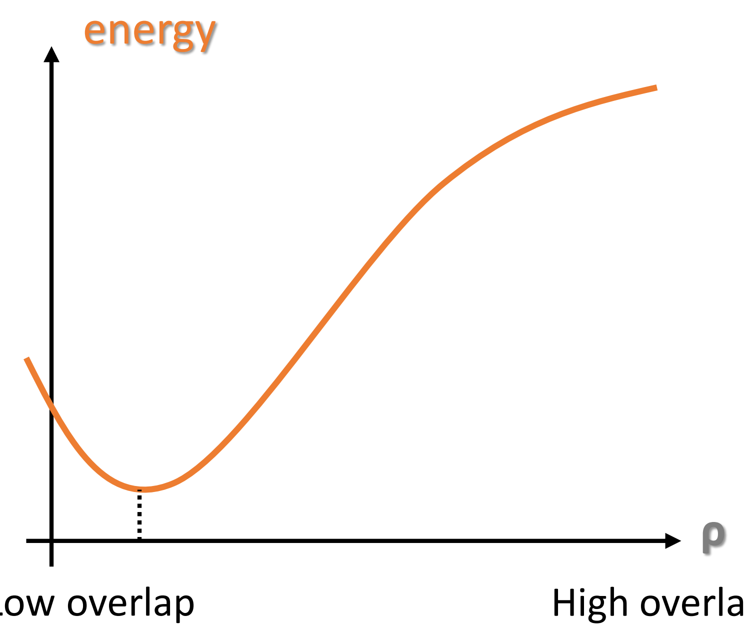
Φ_{in}	Encoding prior dependence
Φ_{out}	Encoding data dependence

When do we get a **good estimator**?

Effective Landscapes and Competition

(possible scenario)

1: $\alpha \ll 1$



overlap: $\rho = \frac{x_0 \cdot \hat{x}}{N}$

GD in this effective landscape

Stationary points ⟷ Fixed points

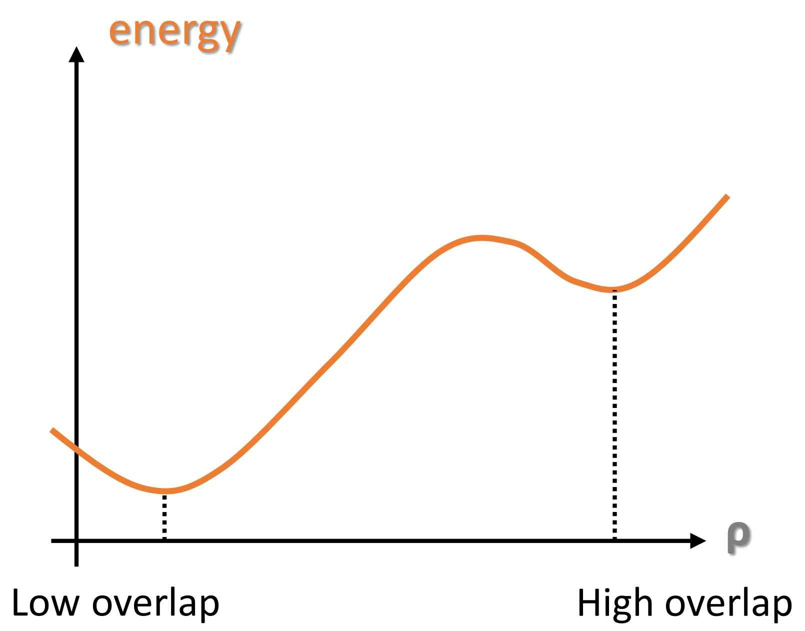
IMPOSSIBLE

α (SNR)

Effective Landscapes and Competition

(possible scenario)

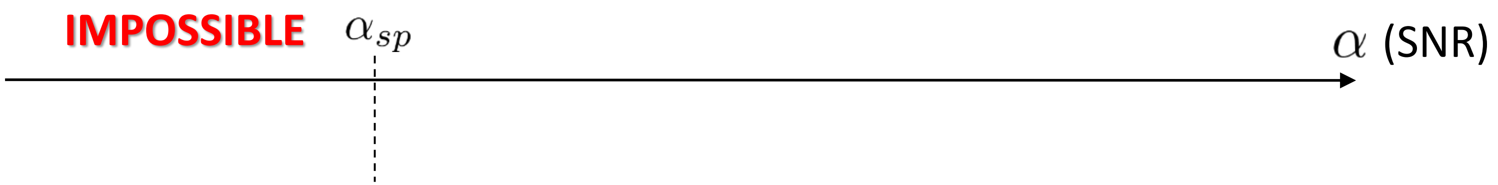
2: $\alpha_{sp} < \alpha < \alpha_{IT}$



overlap: $\rho = \frac{x_0 \cdot \hat{x}}{N}$

GD in this effective landscape

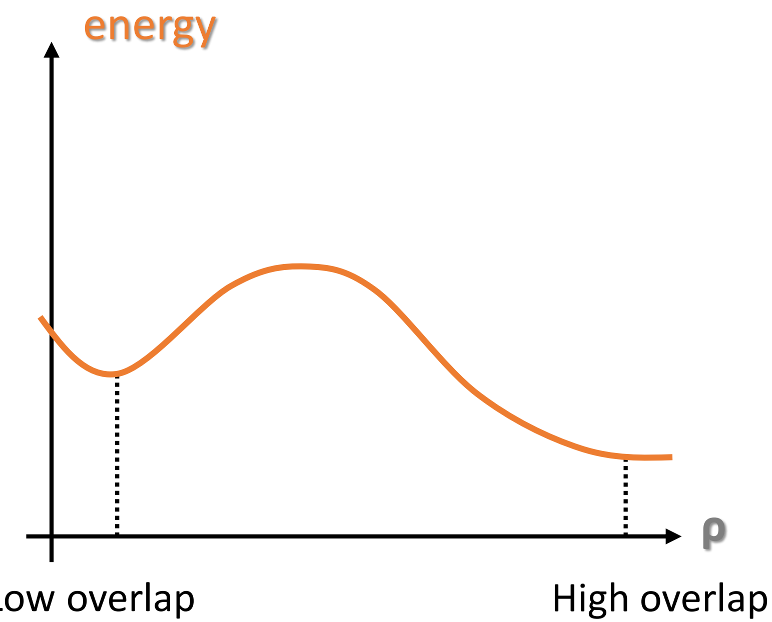
Stationary points ⟷ Fixed points



Effective Landscapes and Competition

(possible scenario)

3: $\alpha_{IT} < \alpha < \alpha_{alg}$



overlap: $\rho = \frac{x_0 \cdot \hat{x}}{N}$

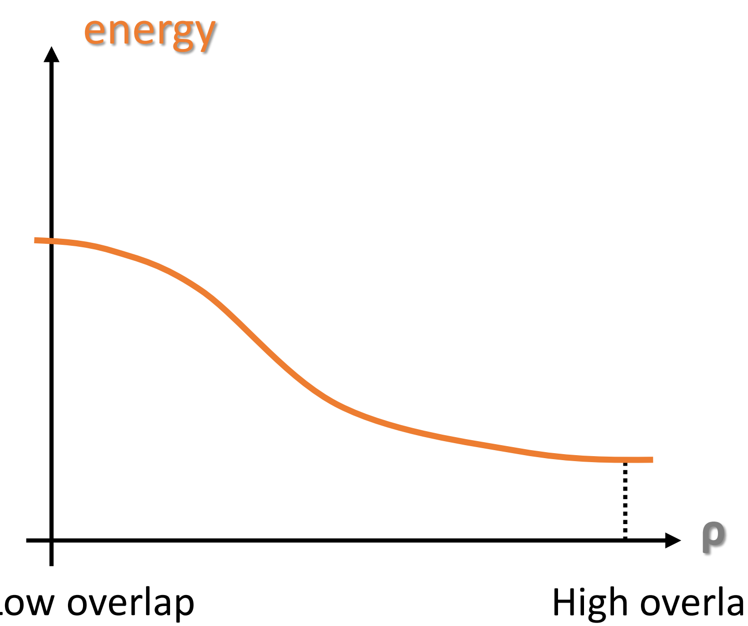
GD in this effective landscape

Stationary points ⟷ Fixed points

Effective Landscapes and Competition

(possible scenario)

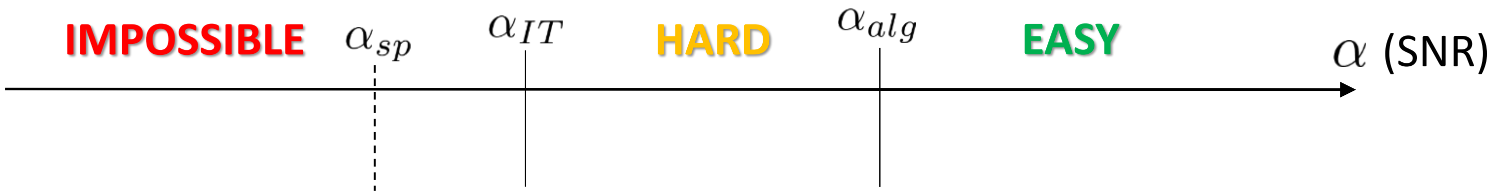
4: $\alpha > \alpha_{alg}$



overlap: $\rho = \frac{x_0 \cdot \hat{x}}{N}$

GD in this effective landscape

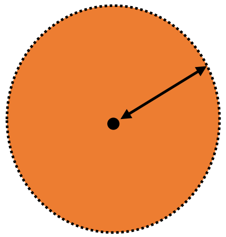
Stationary points ⟷ Fixed points



Breaking the symmetry

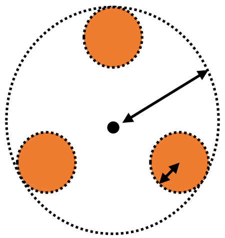
GAMP vs GASP(s)

Replica symmetry assumption



$\hat{x}$
Δ

1RSB assumption



$\hat{x}$
Δ_0
Δ_1

Input scalar channel:

$$\Phi_{in}^{RS}(B,A) = \log \int_{\mathcal{X}} dx\; e^{-\frac{1}{2}Ax^2+Bx-\beta r(x)}$$

Input scalar channel:

$$\Phi_{in}^{1RSB}(B,A_0,A_1;s) = \frac{1}{s}\log \int \mathcal{D}z\; e^{s\,\Phi_{in}^{RS}(B+\sqrt{A_0}z,A_1)}$$

SYMMETRY BREAKING PARAMETER

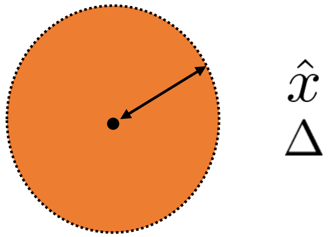
- Same computational complexity
- (Potentially) more expensive element-wise operations

- How to set the symmetry breaking parameter s ?

Breaking the symmetry

GAMP vs **GASP(s)**

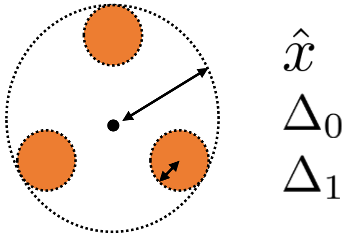
Replica symmetry assumption



Input scalar channel:

$$\Phi_{in}^{RS}(B, A) = \log \int_{\mathcal{X}} dx\ e^{-\frac{1}{2}Ax^2 + Bx - \beta r(x)}$$

1RSB assumption



Input scalar channel:

$$\Phi_{in}^{1RSB}(B, A_0, A_1; s) = \frac{1}{s} \log \int \mathcal{D}z\ e^{s\, \Phi_{in}^{RS}(B + \sqrt{A_0}z, A_1)}$$

SYMMETRY BREAKING PARAMETER

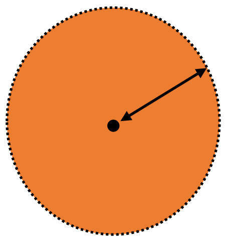
- **Same** computational complexity
- (Potentially) more expensive element-wise operations

- **How to set the symmetry breaking parameter s ?**

Breaking the symmetry

GAMP	vs	GASP(s)

Replica symmetry assumption

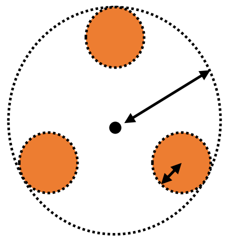


$\hat{x}$
Δ

Input scalar channel:

$$\Phi_{in}^{RS}(B, A) = \log \int_{\mathcal{X}} dx\; e^{-\frac{1}{2}Ax^2+Bx-\beta r(x)}$$

1RSB assumption



$\hat{x}$
Δ_0
Δ_1

Input scalar channel:

$$\Phi_{in}^{1RSB}(B, A_0, A_1; s) = \frac{1}{s} \log \int \mathcal{D}z\; e^{s\,\Phi_{in}^{RS}(B+\sqrt{A_0}z, A_1)}$$

SYMMETRY BREAKING PARAMETER

- **Same** computational complexity
- (Potentially) more expensive element-wise operations

- **How to set the symmetry breaking parameter s ?**

Message-passing equations

GAMP

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1} V^{t-1}$$

$$g_\mu^t = \partial_\omega \varphi_\mu^{\text{out},t}$$

$$\Gamma_\mu^t = -\partial_\omega^2 \varphi_\mu^{\text{out},t}$$

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1} A^t$$

$$\hat{x}_i^t = \partial_B \varphi_i^{\text{in},t}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{\text{in},t}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^t$$

GASP(s)

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1}(V_1^{t-1} + sV_0^{t-1})$$

$$g_\mu^t = \partial_\omega \phi_\mu^{\text{out},t}$$

$$\Gamma_0^t = \frac{1}{s-1}\left(\partial_\omega^2 \phi_\mu^{\text{out},t} - 2\partial_{V_1} \phi_\mu^{\text{out},t} + (g_\mu^t)^2\right)$$

$$\Gamma_1^t = -\partial_\omega^2 \phi_\mu^{\text{out},t} + s\Gamma_0^t$$

$$A_0^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_0^t$$

$$A_1^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_1^t$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1}(A_1^t - sA_0^t)$$

$$\hat{x}_i^t = \partial_B \phi_i^{\text{in},t}$$

$$\Delta_{0,i}^t = \frac{1}{s-1}\left(\partial_B^2 \phi_i^{\text{in},t} + 2\partial_{A_1} \phi_i^{\text{in},t} + (\hat{x}_i^t)^2\right)$$

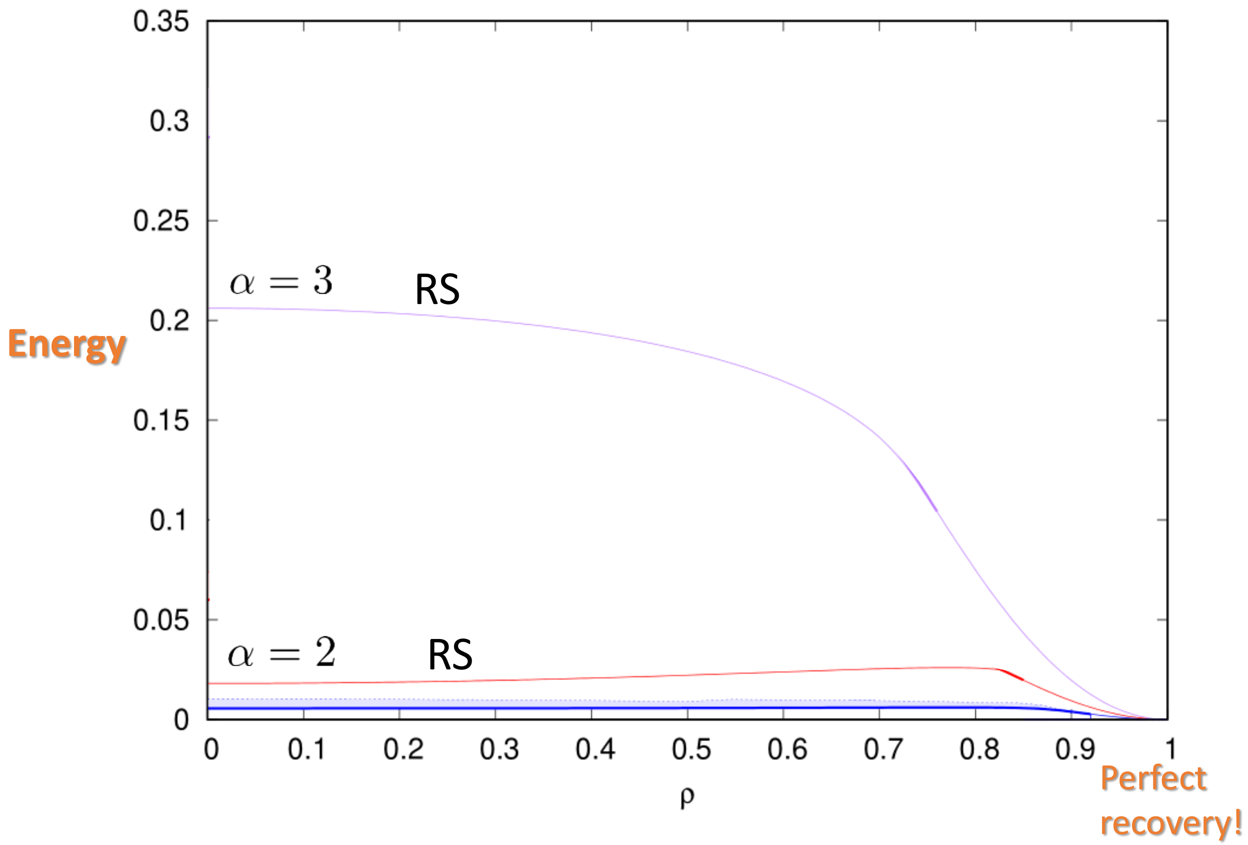
$$\Delta_{1,i}^t = \partial_B^2 \phi_i^{\text{in},t} - s\Delta_{0,i}^t$$

$$V_0^t = c_{\boldsymbol{F}} \sum_i \Delta_{0,i}^t$$

$$V_1^t = c_{\boldsymbol{F}} \sum_i \Delta_{1,i}^t.$$

Changing the Effective Landscape

Phase retrieval, noiseless case
No regularizer



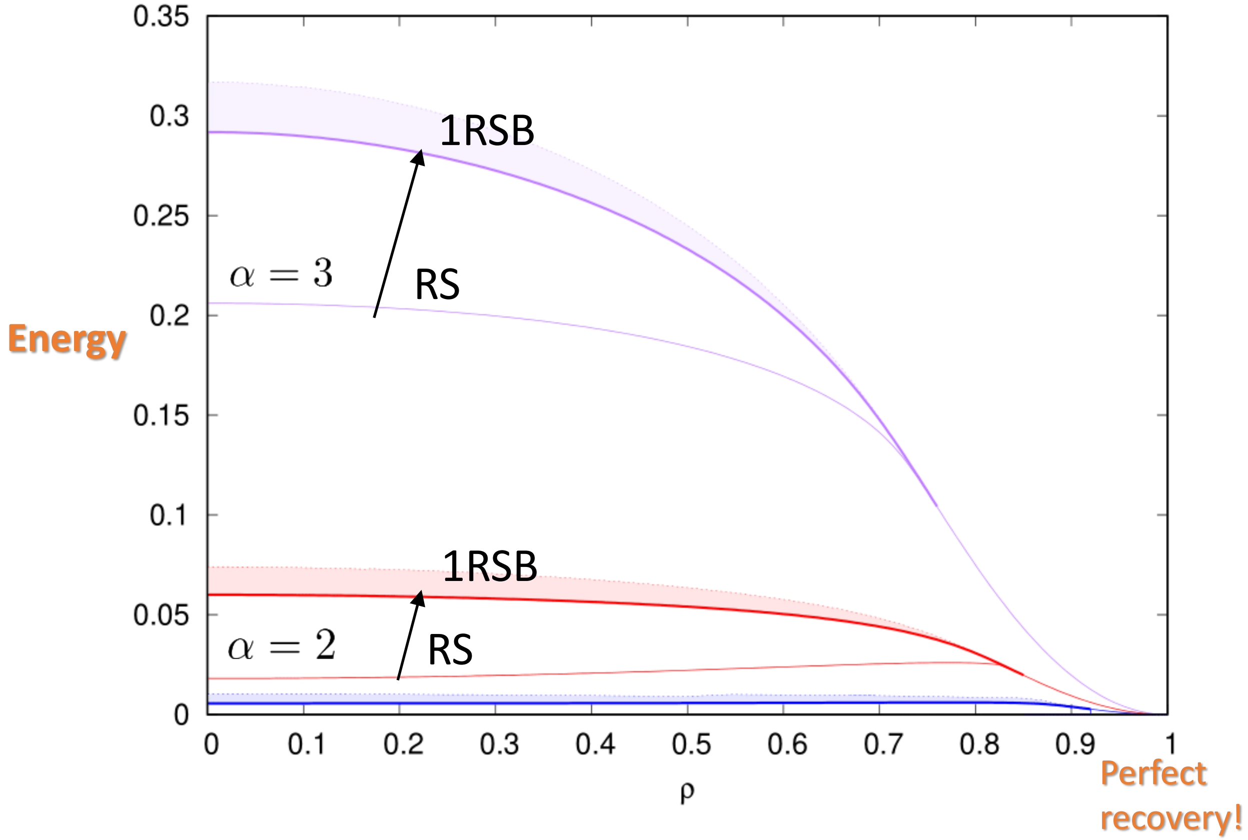
s : explore minima at different energy levels

$s^\star \longrightarrow$ **GROUND STATE**

RS :	Hard below	$\alpha_{alg} \sim 2.5$
1RSB :	Hard below	$\alpha_{alg} \sim 1.5$

Changing the Effective Landscape

Phase retrieval, noiseless case
No regularizer



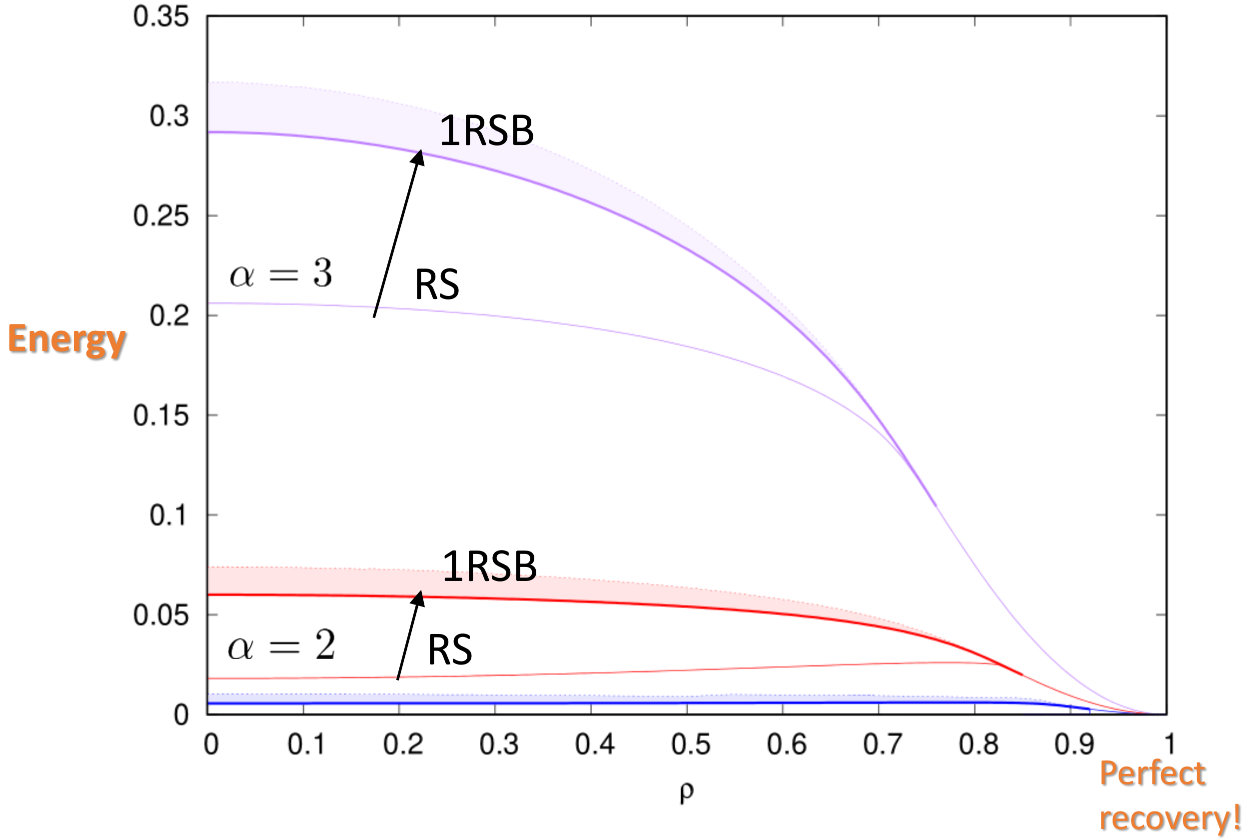
s : explore minima at different energy/complexity levels

$s^\star \longrightarrow$ **GROUND STATE**

RS : Hard below $\alpha_{alg} \sim 2.5$
1RSB : Hard below $\alpha_{alg} \sim 1.5$

Changing the Effective Landscape

Phase retrieval, noiseless case
No regularizer



s : explore minima at different energy/complexity levels

$s^\star \longrightarrow$ **GROUND STATE**

RS :	Hard below	$\alpha_{alg} \sim 2.5$
1RSB :	Hard below	$\alpha_{alg} \sim 1.5$

Conclusions

- In **mismatched** inference settings the **RS assumption** can be wrong.
- **GASP** can improve over **GAMP**. Same O(N^2) complexity.
- Simple **continuation strategy** can push GASP down to the BO algorithmic threshold.
- For more details please **check my poster** this evening!

THANK YOU FOR YOUR ATTENTION!