The Loss Landscape in the Phase Retrieval Problem and Generalized Approximate Survey Propagation

Yue Lu, **Carlo Lucibello** and Luca Saglietti

The Loss Landscape in the Phase Retrieval Problem

We consider the generalized linear estimation problem, where given an N-component uknown signal $x_{0,i} \sim P_{X_0}(\bullet)$, a known $M \times N$ measurement matrix $F_i^\mu \sim \mathcal{N}(0, 1/N)$, we observe M measurements $y^\mu \sim P(\bullet \,|\, \boldsymbol{F}^\mu \cdot \mathbf{x}_0)$ and want to infer $\mathbf{x}_0$.

The inference problem may be phrased as the problem of optimizing a energy/cost function of the form

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

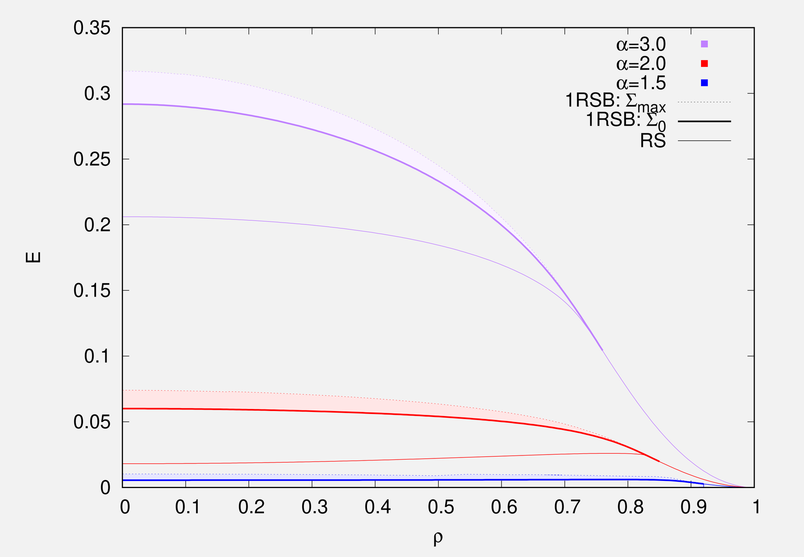
Eventually, $\mathcal{H}(\boldsymbol{x})$ can be chosen as the (minus) log-posterior. Here we focus on the phase retrieval setting, where $y^\mu = |\boldsymbol{F}^\mu \cdot \boldsymbol{x}_0|$, $\ell(y,z) = (y - |z|)^2$ and $r(x) = \lambda\, x^2/2$. We also assume $x_{0,i} \sim \mathcal{N}(0,1)$.

Finding the global minima of $\mathcal{H}(\boldsymbol{x})$ is a **non-convex** optimization problem. We perform a statistical physics analysis [1]

$$Z_{\boldsymbol{y},\boldsymbol{F}} = \int \prod_i dx_i\, e^{-\beta \mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x})}$$

$$E = \lim_{\beta\to\infty} \lim_{N\to\infty} -\frac{1}{\beta N} \mathbb{E} \log Z_{\boldsymbol{y},\boldsymbol{F}}$$

Energy at fixed "Magnetization"



We compute E using the replica method from spin glass theory, within the 1RSB ansatz:

$$\begin{aligned} E_{\rm 1RSB} = {} & \rho\hat\rho - \frac{1}{2}(q_1\,\delta\hat q - \delta q\,\hat q_1 + m q_0 \hat q_0 - m \hat q_1 q_1) \\ & - \alpha\, \mathbb{E}_{\omega,y}\, \phi_{\rm out}(\omega, q_1 - q_0, \delta q, m, y) \\ & - \mathbb{E}_B\, \phi_{\rm in}(B, \hat q_1 - \hat q_0, \delta \hat q, m) \end{aligned}$$

where $\alpha = M/N$ and

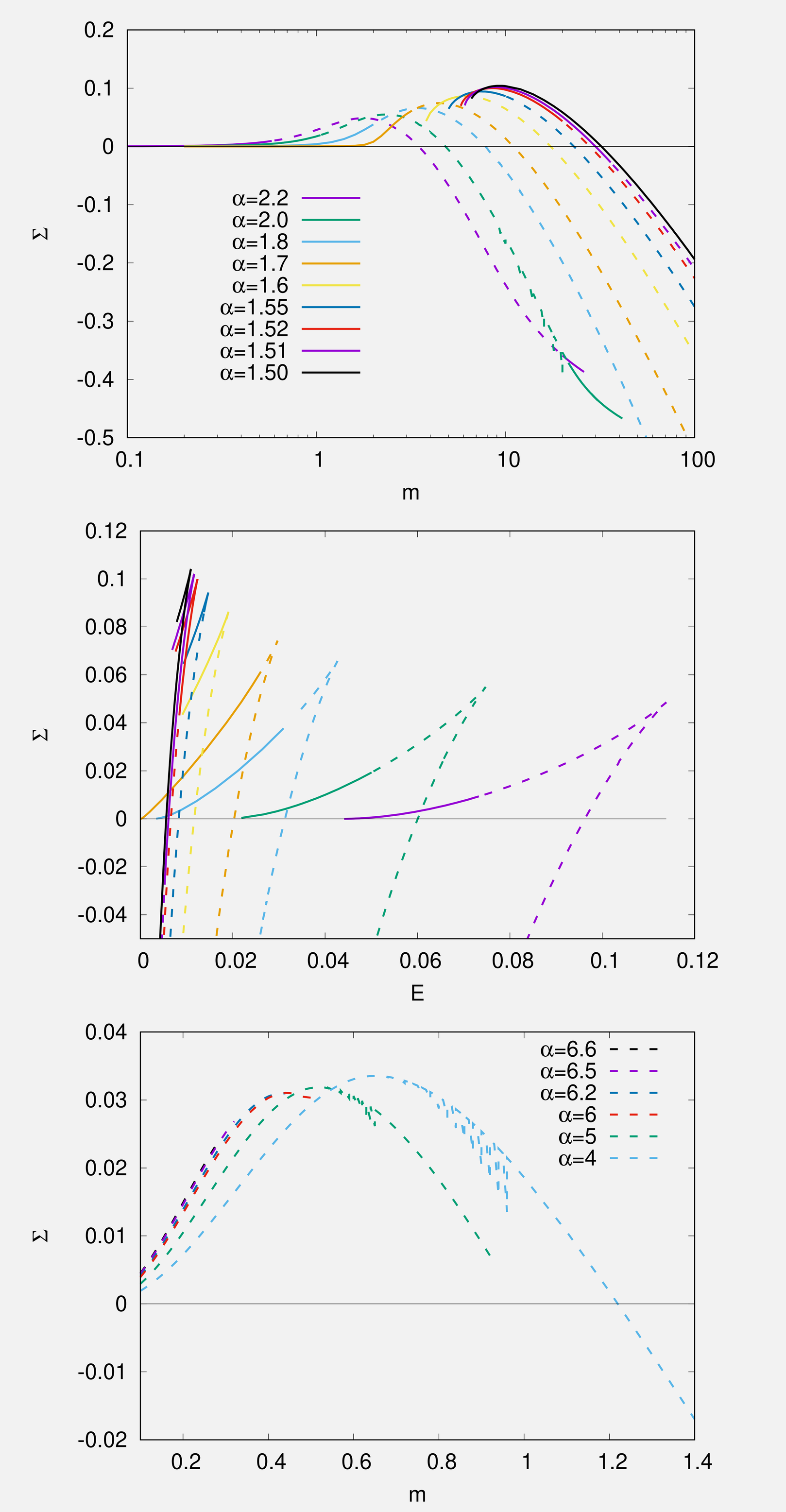
$$\begin{aligned} \omega & \sim \mathcal{N}(0, q_0) \\ y & \sim P(\bullet \,|\, \zeta) \;\text{ with }\; \zeta \sim \mathcal{N}(\rho/q_0\,\omega, \mathbb{E}[x_0^2] - \rho^2/q_0) \\ B & \sim \mathcal{N}(\hat\rho x_0, \hat q_0) \;\text{ with }\; x_0 \sim P_{X_0} \end{aligned}$$

We can explore the fixed magnetization manifold $\rho = \boldsymbol{x} \cdot \boldsymbol{x}_0 / N$ (see also [2]).

Tuning the Parisi parameter m we uncover a whole region with exponentially many local minima. Their (log-)number is accounted for by the complexity $\Sigma = m^2 \partial_m E_{\rm 1RSB}$.

- For any value of $\alpha > 1$ there is a "simple" region aroud the true signal with possibly only one minimum.
- for $\alpha \lesssim 1.50$, the $\rho = 0$ fixed points are stable at any m;
- going up in α, the $\rho = 0$ fixed points at higher values of m become unstable towards the true signal ($\rho = 0$);
- up to $\alpha \approx 2.468$, the RS $\rho = 0$ solution is stable;
- at $\alpha \approx 6$ the physical branch of 1RSB $\rho = 0$ solutions becomes numerically unstable (fixed with damping);
- not easy to establish if the landscape simplifies at large but finite α (proved in a similar setting only for $\alpha = O(\log N)$); gradient descent possibly effective in that regime (to check).
- RSB stability analysis hindered by singularity in second derivatives

Complexity at Zero Magnetization



Generalized Approximate Survey Propagation (GASP)

The Approximate Message Passing (AMP) algorithm for statistical inference has been deeply investigated in recent years. It proved to be highly effective in the case where the generative model is known, less so in case of model mismatch.

Approximate Survey Propagation (ASP), the 1RSB correspective of AMP, was presented in [3] in the context of the planted SK model. While the hope was to overcome the algorithmic barriers of matched AMP, it was reported to perform no better.

We adapt the ASP algorithm to the generalized linear mixing setting and we name it GASP [1]. Focusing on the optimization problem (model mismatch), we report findings similar to those in [3]: for some choice of the parameters of the algorithm, performance of matched AMP are recovered.

For the phase retrieval problem, we find that GASP is numerically more robust than either matched AMP and the unmatched AMP of [4].

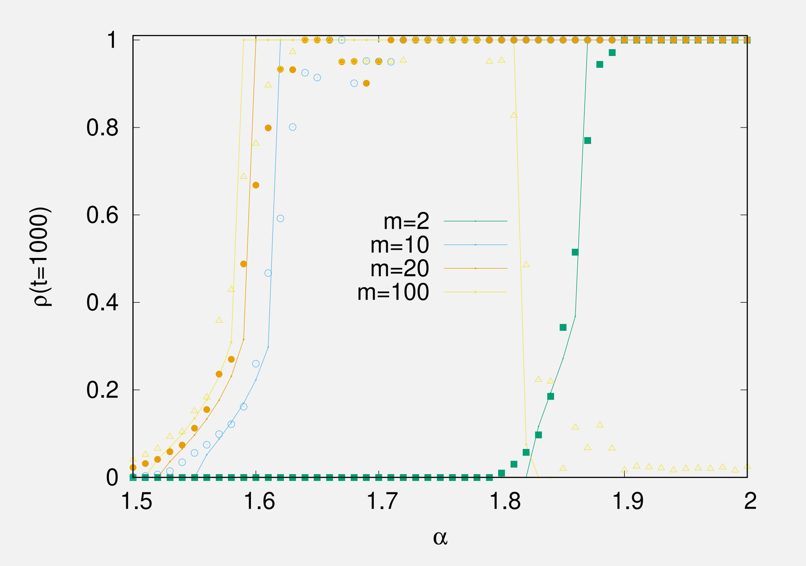
Defining

$$\begin{aligned} \phi_{\rm in}(B, A^0, A^1, m) & = \frac{1}{m}\log \int Dz\; e^{m\varphi_{\rm in}(B + \sqrt{A^0}z,\, A^1)} \\ \varphi_{\rm in}(h, A^1) & = \max_x\; -r(x) - \frac{1}{2}A^1 x^2 + hx \\ \phi_{\rm out}(\omega, V^0, V^1, m, y) & = \frac{1}{m}\log \int Dz\; e^{m\varphi_{\rm out}(\omega + \sqrt{V^0}z,\, V^1, y)} \\ \varphi_{\rm out}(h, V^1, y) & = \max_u\; -\frac{(u-h)^2}{2V^1} - \ell(y,u) \end{aligned}$$

the GASP message passing reads

$$\begin{aligned} V_t^0 & = \overline{\Delta^0_{i,t-1}}^{\,i} \\ V_t^1 & = \overline{\Delta^1_{i,t-1}}^{\,i} \\ \omega_{\mu,t} & = \sum_i F_i^\mu \hat x_{i,t-1} - g_{\mu,t-1}\,(m V_t^0 + V_t^1) \\ g_{\mu,t} & = \partial_\omega \phi^\mu_{{\rm out},t} \\ \Gamma^0_{\mu,t} & = 2\partial_{V^1}\phi^\mu_{{\rm out},t} - g^2_{\mu,t} \\ \Gamma^1_{\mu,t} & = -\partial^2_\omega \phi^\mu_{{\rm out},t} + m\Gamma^0_{\mu,t} \\ A_t^0 & = \alpha\overline{\Gamma^0_{\mu,t}}^{\,\mu} \\ A_t^1 & = \alpha\overline{\Gamma^1_{\mu,t}}^{\,\mu} \\ B_{i,t} & = \sum_\mu F_i^\mu g_{\mu,t} - \hat x_{i,t-1}\,(m A_t^0 - A_t^1) \\ \hat x_{i,t} & = \partial_B \phi^i_{{\rm in},t} \\ \Delta^0_{i,t} & = -2\partial_{A^1}\phi^i_{{\rm in},t} - \hat x^2_{i,t} \\ \Delta^1_{i,t} & = \partial^2_B \phi^i_{{\rm in},t} - m\Delta^0_{i,t} \end{aligned}$$

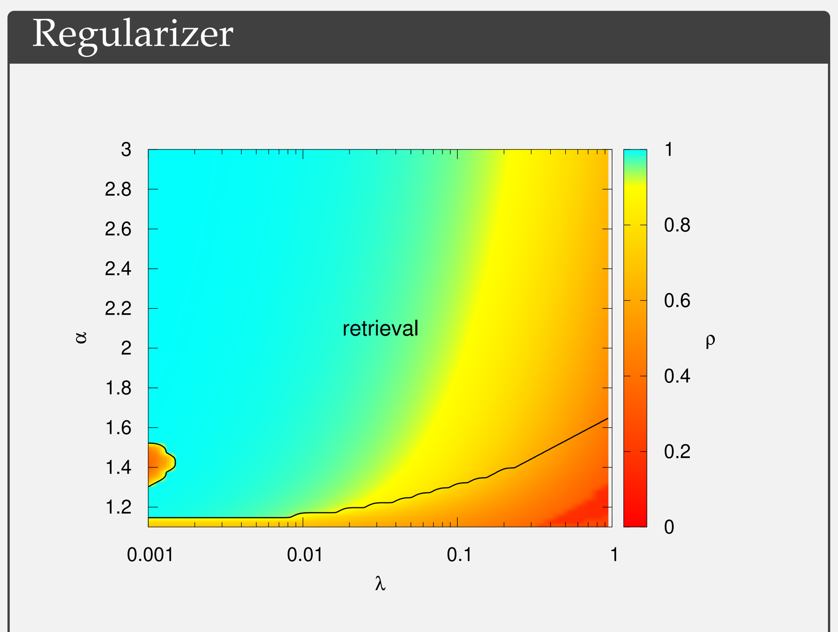
GASP and SE



GASP with no regularization retrieves the signals above $\alpha \approx 1.5$, while its RS counterpart works only above $\alpha \approx 2.5$ (and contains some hard to estimate delta functions). Optimal Bayes AMP's algorithmic transition is at $\alpha \approx 1.13$ and information theoretic bound for recovery is at $\alpha = 1$.

It is easy to prove, building on previous works, that for large system size the time evolution of GASP can be rigorously tracked by the State Evolution scalar equations.

Regularizer



Adding a L_2 regularizer can improve performances. A simple protocol where 1) λ is fixed to some value until GASP convergence; 2) it is dropped to zero; 3) GASP is run again till convergence; results effective in perfectly recovering the true signal in a vast region of the phase space.

A future challenge is to devise an adaptive scheme for m_t and λ_t to facilitate the signal retrieval and eliminate the need to hand-tune extra-parameters.

Hot-starting GASP with spectral initialization should also be possible [4].

References

[1] Lu, Lucibello, Saglietti (in preparation).
[2] Ros, Ben Arous, Biroli, Cammarota. Complex energy landscapes in spiked-tensor and simple glassy models (2018).
[3] Antenucci, Krzakala, Urbani, Zdeborová. Approximate Survey Propagation for Statistical Inference (2018).
[4] Ma, Xu, Maleki. Optimization-based AMP for Phase Retrieval (2018).