arXiv:1905.05313v1 [cond-mat.dis-nn] 13 May 2019

Generalized Approximate Survey Propagation for High-Dimensional Estimation

Luca Saglietti [1 2] **Yue M. Lu** [3] **Carlo Lucibello** [4 1]

Abstract

In Generalized Linear Estimation (GLE) problems, we seek to estimate a signal that is observed through a linear transform followed by a component-wise, possibly nonlinear and noisy, channel. In the Bayesian optimal setting, Generalized Approximate Message Passing (GAMP) is known to achieve optimal performance for GLE. However, its performance can significantly degrade whenever there is a mismatch between the assumed and the true generative model, a situation frequently encountered in practice. In this paper, we propose a new algorithm, named Generalized Approximate Survey Propagation (GASP), for solving GLE in the presence of prior or model mis-specifications. As a prototypical example, we consider the phase retrieval problem, where we show that GASP outperforms the corresponding GAMP, reducing the reconstruction threshold and, for certain choices of its parameters, approaching Bayesian optimal performance. Furthermore, we present a set of State Evolution equations that exactly characterize the dynamics of GASP in the high-dimensional limit.

1. Introduction

Approximate message passing (AMP) algorithms have become a well established tool in the study of inference problems (Donoho et al., 2009; Donoho & Montanari, 2016; Advani & Ganguli, 2016) that can be represented by dense graphical models. An important feature of AMP is that its dynamical behavior in the large system limit can be exactly predicted through a dynamical system involving only scalar quantities called *State Evolution* (SE) (Bayati & Montanari, 2011). This relationship paved the way for a series of rigorous results (Rangan & Fletcher, 2012; Deshpande & Montanari, 2014; Deshpande et al., 2016). It also helps clarify the connection to several fascinating predictions obtained through the *replica analysis* in statistical physics (Mézard et al., 1987). In the optimal Bayesian setting, where one has perfect information on the process underlying data generation, AMP has been empirically shown to achieve optimal performances among polynomial algorithms for many different problems. However, in the more realistic case of mismatch between the assumed and the true generative model, i.e. when AMP is not derived on the true posterior distribution, it may become sub-optimal. A possible source of problems for the AMP class of algorithms is the outbreak of *Replica Symmetry Breaking* (Mézard et al., 1987), a scenario where an exponential number of fixed point and algorithmic barriers dominate the free energy landscape explored by AMP. This phenomena can be accentuated in case of model mismatch: a notable example is maximum likelihood estimation (as opposed to estimation by the posterior mean, which corresponds to the low temperature limit of a statistical physics model.

These considerations are well known within the physics community of disordered systems (Krzakala et al., 2016), where the problem of signal estimation is informally referred to as "crystal hunting". Estimation problems in high dimensions are characterized by a complex energy-entropy competition where the true signal is hidden in a vast and potentially rough landscape. In a wide class of problems, one observes the presence of a algorithmically "*hard*" phase for some range of values for the parameters defining the problem (e.g. signal-to-noise ration). In this regime, all known polynomial complexity algorithms fail to saturate the information theoretic bound (Ricci-Tersenghi et al., 2019). While reconstruction is possible in principle, algorithms are trapped in a region of the configuration space with low overlap with the signal and many local minima (Antenucci et al., 2019a; Ros et al., 2019).

In a recent work (Antenucci et al., 2019b), a novel message-passing algorithm, *Approximate Survey Propagation* (ASP),

[1]Microsoft Research New England, Cambridge, MA 02142, USA [2]Italian Institute for Genomic Medicine, Turin, Italy [3]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA [4]Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy. Correspondence to: Carlo Lucibello <carlo.lucibello@unibocconi.it>, Luca Saglietti <luca.saglietti@gmail.com>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

was introduced in the context of low-rank. The algorithm is based on the *1-step Replica Symmetry Breaking* (1RSB) ansatz from spin glass theory (Mézard et al., 1987), which was specifically developed to deal with landscapes populated by exponentially many local minima. It was shown that ASP on the mismatched model could reach the performance of (but not improve on) matched AMP and do far better than mismatched AMP (Antenucci et al., 2019a;b). In the present paper, we build upon these previous works and derive the ASP algorithm for *Generalized Linear Estimation* (GLE) models. Since the extension of AMP to GLE problems is commonly known as GAMP, we call *Generalized Approximate Survey Propagation* (GASP), our extension of ASP. We will show that also in this case, in presence of model mismatch, (G)ASP improves over the corresponding (G)AMP.

2. Model specification

An instance of the general class of models to which GASP can be applied is defined, for some integer M and N, by an *observed signal* $\boldsymbol{y} \in \mathbb{R}^M$ and an $M \times N$ *observation matrix* $\boldsymbol{F}$. Clearly, this scenario encompasses also GLE. We denote with $\boldsymbol{F}^\mu,\ \mu \in [M]$, the rows of $\boldsymbol{F}$ and refer to the ratio $\alpha = M/N$ as the *sampling ratio* of $\boldsymbol{F}$. We consider a probability density distribution $p(\boldsymbol{x})$ on a (possibly discrete) space χ^N, $\chi \subseteq \mathbb{R}$, defined as:

$$p(\boldsymbol{x}) = \frac{1}{Z}\, e^{-\beta \mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x})}, \tag{1}$$

where, following statistical physics jargon, β plays the role of an inverse temperature, Z is a normalization factor called partition function (both Z and p implicitly depend on $\beta, \boldsymbol{y}$ and $\boldsymbol{F}$), and $\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}$ is the Hamiltonian of the model, that in our setting takes the form:

$$\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}) = \sum_{\mu=1}^{M} \ell\big(y_\mu,\, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle\big) + \sum_{i=1}^{N} r(x_i). \tag{2}$$

Here $\langle \bullet, \bullet \rangle$ denotes the scalar product and we call ℓ and r the loss function and the regularizer of the problem respectively.

In this quite general context, the purpose of GASP is to approximately compute the marginal distribution $p(x_i)$, along with some expected quantities such as e.g. $\hat{\boldsymbol{x}} = \mathbb{E}_p \boldsymbol{x}$. The approximation entailed in GASP turns out to be exact under some assumptions in the large N limit, as we shall later see. A crucial assumption in the derivation of the GASP algorithm (and of GAMP as well), is that the entries of $\boldsymbol{F}$ are independently generated according to some zero mean and finite variance distribution.

Although the general formulation of GASP, presented in Sec. 2 of the SM, is able to deal with any model of the form (1), we will here restrict the setting to discuss Generalized Linear Estimation (Rangan, 2011).

In GLE problems, $p(x)$ is sensibly chosen in order to infer a *true signal* $\boldsymbol{x}_0 \in \mathbb{R}^N$, whose components are assumed to be independently extracted from some prior P_0, $x_{0,i} \sim P_0\ \forall i \in [N]$. The observations are independently produced by a (probabilistic) scalar channel P^{out}: $y_\mu \sim P^{\text{out}}(\bullet \,|\, \langle \boldsymbol{F}^\mu, \boldsymbol{x}_0\rangle)$.

It is then reasonable to choose $\ell\big(y, z\big) = -\log P^{\text{out}}(y|z)$, $r(x) = -\log P_0(x)$ and $\beta = 1$, so that the probability density $p(\boldsymbol{x})$ corresponds to the true posterior $P(\boldsymbol{x}|\boldsymbol{F}, \boldsymbol{y}) \propto P^{\text{out}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{F}) P_0(\boldsymbol{x})$, where $\propto$ denotes equality up to a normalization factor. We refer to this setting as to the Bayesian-optimal or matched setting (Barbier et al., 2018). Notice that in the $\beta \uparrow \infty$ limit $p(x)$ concentrates around the maximum-a-posteriori (MAP) estimate. If $\beta \neq 1$ or if the Hamiltonian doesn't correspond to the minus log posterior (e.g, when P_0 and P^{out} used in the Hamiltonian do not correspond to true ones) we talk about model mismatch.

As a testing ground for GASP, and the corresponding State Evolution, we here consider the phase retrieval problem, which has undergone intense investigation in recent years (Candes et al., 2015; Dhifallah & Lu, 2017; Chen et al., 2018; Goldstein & Studer, 2018; Mondelli & Montanari, 2018; Sun et al., 2018; Mukherjee & Seelamantula, 2018). We examine its noiseless and real-valued formulation, where observations are generated according to the process

$$\boldsymbol{x}_0 \sim \mathcal{N}(0, I_N), \tag{3}$$
$$F_i^\mu \sim \mathcal{N}(0, 1/N) \quad \forall \mu \in [M], \forall i \in [N], \tag{4}$$
$$y_\mu \sim |\langle \boldsymbol{F}^\mu, \boldsymbol{x}_0\rangle|. \tag{5}$$

for some M and N, such that $\alpha = M/N > 1$. For such generative model, we will focus on the problem of recovering $\boldsymbol{x}_0$ by minimizing the energy function $\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x})$ of Eq. (2), in the case

$$\ell(y, z) = (y - |z|)^2, \tag{6}$$
$$r(x) = \frac{1}{2}\lambda\, x^2. \tag{7}$$

Since the setting assumed for inference corresponds to MAP estimation in presence of a noisy channel, we are dealing with a case of model mismatch. The effect of the parameter λ on the estimation shall be explored in Sec. 7, but we assume $\lambda = 0$ until then. The optimization procedure will be performed using the zero-temperature (i.e. $\beta \uparrow \infty$) version of the GASP algorithm.

3. Previous work on Approximate Message Passing for Phase Retrieval

Generalized approximate message passing (GAMP) was developed and rigorously analyzed in Refs. (Rangan, 2011) and (Javanmard & Montanari, 2013). It was then applied for the first time to the (complex-valued) phase retrieval

problem in Ref. (Schniter & Rangan, 2015). In Ref. (Barbier et al., 2018) the authors report an algorithmic threshold for the perfect recovery of $\alpha_{\rm alg} \approx 1.13$, when using matched AMP on the real-valued version of the problem. This is to be compared to the information theoretic bound $\alpha_{\rm IT} = 1$.

The performance of GAMP in the MAP estimation setting, instead, was investigated in Ref. (Ma et al., 2018; 2019). A "vanilla" implementation of the zero temperature GAMP equations for the absolute value channel was reported to achieve perfect recovery for real-valued signals above $\alpha_{\rm alg} \approx 2.48$. The authors were able to show that the algorithmic threshold of GAMP in the mismatched case can however be drastically lowered by introducing regularization a regularization term ultimately continued to zero. The AMP.A algorithm proposed in (Ma et al., 2018; 2019) uses an adaptive L_2 regularization that improves the estimation threshold and also makes the algorithm more numerically robust compensating a problematic divergence that appears in the message-passing equations (see Sec. 1.3 in the SM for further details).

Another important ingredient for AMP.A's performance is initialization: in order to achieve perfect recovery one has to start from a configuration that falls within the basin of attraction of the true signal, which rapidly shrinks as the sampling ratio α decreases. A well-studied method for obtaining a configuration correlated with the signal is *spectral initialization*, introduced and studied in Refs. (Jain et al., 2013; Candes et al., 2015; Chen & Candes, 2015): in this case the starting condition is given by the principal eigenvector of a matrix obtained from the data matrix $\boldsymbol{F}$ and the labels $\boldsymbol{y}$ passed through a nonlinear processing function. The asymptotic performance of this method was analyzed in (Lu & Li, 2017), while the form of the optimal processing function was described in (Mondelli & Montanari, 2018; Luo et al., 2019). However, since the SE description is based on the assumption of the initial condition being uncorrelated with the data, in AMP.A the authors revisited the method, proposing a modification that guarantees "enough independency" while still providing high overlap between the starting point and the signal.

With the combination of these two heuristics, AMP.A is able to reconstruct the signal down $\alpha_{\rm alg} \approx 1.5$. In the present paper we will show that, with a basic continuation scheme, the 1RSB version of the zero temperature GAMP can reach the Bayes-optimal threshold $\alpha_{\rm alg} \approx 1.13$ also in the mismatched case, without the need of spectral initialization.

3.1. GAMP equations at zero temperature

Here we provide a brief summary of the AMP equations for the general graphical model of Eq. (1), in the $\beta \uparrow \infty$ limit. This is both to allow an easy comparison with our novel GASP algorithm and to introduce some notation that will be useful in the following discussion. There is some degree of model dependence in the scaling of the messages when taking the zero-temperature limit: here we adopt the one appropriate for over-constrained models in continuous space. Details of the derivation can be found in Sec. 1 of the SM, along with the specialization of the equations for phase retrieval.

First, we introduce two free entropy functions associated to the input and output channels (Rangan, 2011):

$$\varphi^{\rm in}(B, A) = \max_x \; -r(x) - \frac{1}{2}Ax^2 + Bx \tag{8}$$

$$\varphi^{\rm out}(\omega, V, y) = \max_u \; -\frac{(u-\omega)^2}{2V} - \ell(y, u). \tag{9}$$

We define for convenience $\varphi_i^{{\rm in},t} = \varphi^{\rm in}(B_i^t, A^t)$ and $\varphi_\mu^{{\rm out},t} = \varphi^{\rm out}(\omega_\mu^t, V^{t-1}, y_\mu)$. In our notation the GAMP message passing equations read:

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1}\, V^{t-1} \tag{10}$$

$$g_\mu^t = \partial_\omega \varphi_\mu^{{\rm out},t} \tag{11}$$

$$\Gamma_\mu^t = -\partial_\omega^2 \varphi_\mu^{{\rm out},t} \tag{12}$$

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t \tag{13}$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1}\, A^t \tag{14}$$

$$\hat{x}_i^t = \partial_B \varphi_i^{{\rm in},t} \tag{15}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{{\rm in},t} \tag{16}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^t \tag{17}$$

where $c_{\boldsymbol{F}} = \frac{1}{MN}\sum_{\mu,i}(F_i^\mu)^2$. It is clear from the equations that the two free entropy functions are supposed to be twice differentiable. This is not the case for phase retrieval, where GAMP encounters some non-trivial numerical stability issues: during the message-passing iterations one would have to approximately evaluate an empirical average of $\partial_\omega^2 \varphi_\mu^{{\rm out},t}$, containing Dirac's δ-function. This is the problem encountered in AMP.A of Ref. (Ma et al., 2018). We will see that this problem is not present in GASP thanks to a Gaussian smoothing of the denoising function.

4. Generalized Approximate Survey Propagation

The (G)ASP algorithm builds on decades of progress within the statistical physics community in understanding and dealing with rough high-dimensional landscapes. The starting point for the derivation of the algorithm is the partition

function of m replicas (or clones) of the system $\{\boldsymbol{x}^a\}_{a=1}^m$:

$$Z^m_{\boldsymbol{y},\boldsymbol{F}} = \int \prod_{a=1}^{m}\prod_{i=1}^{N} dx_i^a \; e^{-\beta \sum_{a=1}^m \mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}^a)}. \tag{18}$$

Note that, while this probability measure factorizes trivially, setting $m \neq 1$ can introduce many important differences with respect to the standard case, both from the algorithmic and from the physics standpoints (Monasson, 1995; Antenucci et al., 2019b).

We write down the Belief Propagation (BP) equations associated to the replicated factor graph, where messages are probability distributions associated to each edge over the single-site replicated variables $\bar{\boldsymbol{x}}_i = (x_i^1, \dots, x_i^m)$. We make the assumption that the messages are symmetric under the group of replica indexes permutations. This allows for a parametrization of the message passing that can be continued analytically to any real value of m. The resulting algorithm goes under the name of 1RSB Cavity Method or, more loosely speaking, of Survey Propagation (with reference in particular to a zero temperature version of the 1RSB cavity method in discrete constraint satisfaction problems), and led to many algorithmic breakthroughs in combinatorial optimization on sparse graphical models (Mézard et al., 2002; Braunstein et al., 2005; Krzakała et al., 2007). One possible derivation of the (G)ASP algorithm is as the dense graph limit of the Survey Propagation equations, in the same way as AMP is obtained starting from BP. The derivation requires two steps. First, BP messages are projected by moment-matching onto (replica-symmetric) multivariate Gaussian distributions on the replicated variables $\bar{\boldsymbol{x}} \in \mathbb{R}^m$, which we express in the form

$$\nu(\bar{\boldsymbol{x}}) \propto \int \mathrm{d}\, h \; e^{-\frac{1}{2\Delta_0}(h-\hat{x})^2} \prod_{a=1}^{m} e^{-\frac{1}{2\Delta_1}(x^a-h)^2}; \tag{19}$$

Then, messages on the edges are conveniently expressed in term of single site quantities. We note that, some statistical independence assumptions on the entries of the measurement matrix are crucial for the derivation, as goes for AMP as well. While the starting point of the derivation assumed integer m, the resulting message passing can be analytically continued to any real m. Applying this procedure to the GLE graphical model of Eq. (1) we obtain the GASP equations. Here we consider the $\beta \uparrow \infty$ limit to deal with the MAP estimation problem. Details of the GASP derivation and the finite β GASP equations are given in Sec. 2 of the SM. Particular care has to be taken in the limit procedure, as a proper rescaling with β is needed for each parameter. For instance, as the range of sensible choices for m shrinks towards zero for increasing β, we rescale m through the substitution $m \leftarrow m/\beta$.

Relying on the definitions given Eqs. (8) and (9), we introduce the two 1RSB free entropies:

$$\phi^{\text{in}}(B, A_0, A_1, m) = \frac{1}{m} \log \int Dz \; e^{m\varphi^{\text{in}}(B+\sqrt{A_0}z,\, A_1)} \tag{20}$$

$$\phi^{\text{out}}(\omega, V_0, V_1, y, m) = \frac{1}{m} \log \int Dz \; e^{m\varphi^{\text{out}}(\omega+\sqrt{V_0}\,z,\, V_1, y)}. \tag{21}$$

Here $\int Dz$ denotes the standard Gaussian integration $\int dz \, \exp(-z^2/2)/\sqrt{2\pi}$. Using the shorthand notations $\phi_i^{\text{in},t} = \phi^{\text{in}}(B_i^t, A_0^t, A_1^t, m)$ and $\phi_\mu^{\text{out},t} = \phi^{\text{out}}(\omega_\mu^t, V_0^{t-1}, V_1^{t-1}, y_\mu, m)$ (notice the shift in the time indexes), and using again the definition $c_{\boldsymbol{F}} = \frac{1}{MN}\sum_{\mu,i}(F_i^\mu)^2$ (hence $\mathbb{E} c_{\boldsymbol{F}} = 1/N$ in our setting), the GASP equations read:

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1}\left(mV_0^{t-1} + V_1^{t-1}\right) \tag{22}$$

$$g_\mu^t = \partial_\omega \phi_\mu^{\text{out},t} \tag{23}$$

$$\Gamma_{0,\mu}^t = 2\partial_{V_1}\phi_\mu^{\text{out},t} - (g_\mu^t)^2 \tag{24}$$

$$\Gamma_{1,\mu}^t = -\partial_\omega^2 \phi_\mu^{\text{out},t} + m\Gamma_{0,\mu}^t \tag{25}$$

$$A_0^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_{0,\mu}^t \tag{26}$$

$$A_1^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_{1,\mu}^t \tag{27}$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t - \hat{x}_i^{t-1}\left(mA_0^t - A_1^t\right) \tag{28}$$

$$\hat{x}_i^t = \partial_B \phi_i^{\text{in},t} \tag{29}$$

$$\Delta_{0,i}^t = -2\partial_{A_1}\phi_i^{\text{in},t} - (\hat{x}_i^t)^2 \tag{30}$$

$$\Delta_{1,i}^t = \partial_B^2 \phi_i^{\text{in},t} - m\Delta_{0,i}^t. \tag{31}$$

$$V_0^t = c_{\boldsymbol{F}} \sum_i \Delta_{0,i}^t \tag{32}$$

$$V_1^t = c_{\boldsymbol{F}} \sum_i \Delta_{1,i}^t \tag{33}$$

Algorithm 1 GASP(m) for MAP

initialize $g_\mu = 0 \; \forall \mu$
initialize V_0, V_1, $\hat{x}_i \; \forall i$ to some values
for $t = 1$ **to** $t_{\max}$ **do**
  compute $\omega_\mu, g_\mu, \Gamma_\mu^0, \Gamma_\mu^1 \; \forall \mu$ using (22,23, 24, 25)
  compute A_0, A_1 using (26,27)
  compute $B_i, \hat{x}_i, \Delta_{0,i}, \Delta_{1,i} \; \forall i$ using (28, 29, 30, 31)
  compute V^0, V^1 using (32, 33)
end for

The computational time and memory complexity per iteration of the algorithm is the same of GAMP and is determined

by the linear operations in Eqs. (22) and (28). With respect to GAMP, we have the additional (but sub-leading) complexity due to the integrals in the input and output channels. In some special cases, the integrals in Eqs. (20) and (21) can be carried out analytically (e.g. in the phase retrieval problem).

Notice that GASP iteration reduces to standard GAMP iterations if V_0 and A_0 are initialized (or shrink) to zero, but can produce non-trivial fixed points depending on the initialization condition and on the value of m.

We remark the importance of setting the time-indices correctly in order to allow convergence (Caltagirone et al., 2014). The full algorithm is detailed in Alg. 1.

The expressions for the special case of the absolute value channel (6) and L_2 regularization (7) can be found in Sec. 2.4 of the SM. An important comment is that the divergence issue arising in AMP.A, in the same setting, does not affect GASP: the discontinuity in the expression for the minimizer of Eq. (9) is smoothed out in the 1RSB version by the Gaussian integral in Eq. (20). We also note that, in phase retrieval, a problematic initialization can be obtained by choosing configurations that are exactly orthogonal to the signal, since the message-passing will always be trapped in the uninformative fixed-point (due to the Z_2 symmetry of the problem). However, for finite size instances, a random Gaussian initial condition will have an overlap $\rho \equiv \langle \hat{\boldsymbol{x}}, \boldsymbol{x}_0 \rangle$ of order $\mathcal{O}\left(1/\sqrt{N}\right)$ with the signal, which allows to escape the uninformative fixed point whenever it is unstable (i.e. for high α).

In Fig. 1 (Top and Middle), we show the probability of a perfect recovery and convergence times of GASP for the real-valued phase retrieval problem, for different sampling ratios α and values of the symmetry-breaking parameter m, with $\lambda = 0$. The initial condition is given by $V_0^{t=0} = V_1^{t=0} = 1$ and $\hat{\boldsymbol{x}}^{t=0} \sim \mathcal{N}(0, I_N)$. Notice that standard Gaussian initialization is able to break the symmetry of the channel and, at large t, GASP matches the fixed points predicted by SE (see next Section) with a small initial overlap with the true signal ($\rho^{t=0} = 10^{-3}$). In order to achieve signal recovery at low α, the symmetry-breaking parameter has to be increased. In correspondence of values $m \approx 100$, we report an algorithmic threshold around $\alpha_{\text{alg}}^{\lambda=0} \approx 1.5$. This threshold is comparable to the one of AMP.A, without exploiting adaptive regularization and spectral initialization as AMP.A (and which could be employed also for GASP).

We report that, at fixed m, when α is increased above a certain value the message-passing will stop converging. The oscillating/diverging behavior of the messages can however be exploited for hand-tuning m, in the absence of a replica analysis to support the selection of its most appropriate value. More details can be found in Sec. 3 of the SM.

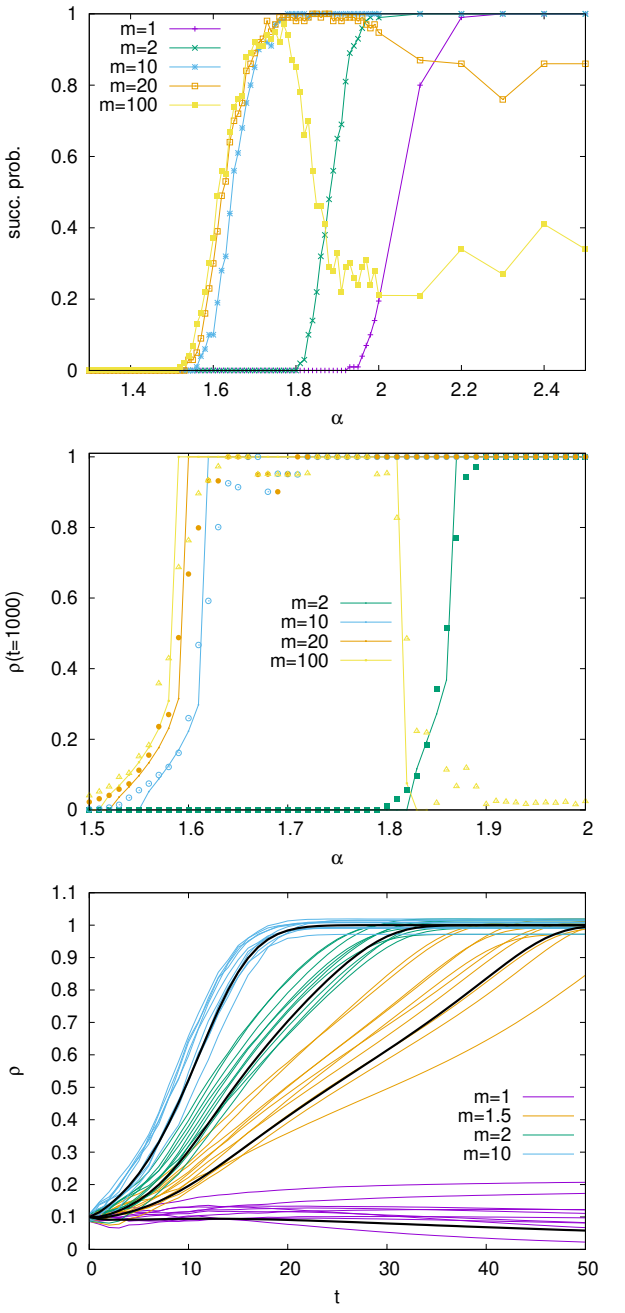

Figure 1. (Top) Probability of perfect recovery of the true signal using GASP (Alg. 1), as a function of the sampling ratio $\alpha = M/N$. (Middle) GASP and SE result after $t = 10^3$ iterations. Start at $t = 0$ with $\rho = 10^{-3}$ for SE and $\hat{\boldsymbol{x}} \sim \mathcal{N}(0, I_N)$ for GASP ($N = 10^3$, averaged over 100 samples). (Bottom) Overlap ρ^t with the true signal predicted by SE dynamics at $\alpha = 2$ and initial overlap $\rho = 0.1$ (black lines) compared to 10 GASP trajectories for each value of m. Here $N = 10^4$.

We presented here the zero-temperature limit of the GASP message-passing to solve the MAP problem. Refer to Sec. 2 of the SM for a more general formulation dealing with the class of graphical models in the form of Eq. 1.

5. State Evolution for GASP

State Evolution (SE) is a set of iterative equations involving a few scalar quantities, that were rigorously proved to track the (G)AMP dynamics, in the sense of almost sure convergence of empirical averages (Javanmard & Montanari, 2013) in the large N limit and with fixed sampling ratio $\alpha = M/N$. Following the analysis of Ref. (Rangan, 2011) for GAMP, in order to present the SE equations for GASP we assume that the observation model $y_\mu \sim P^{\text{out}}(\bullet \,|\, \langle \boldsymbol{F}^\mu, \boldsymbol{x}_0 \rangle)$ is such that can be expressed in the form $y^\mu \sim h(\langle \boldsymbol{F}^\mu, \boldsymbol{x}_0 \rangle, \xi^\mu)$ for some function $h(z, \xi)$, with ξ^μ a scalar- or vector-valued random variable modeling the noise and sampled according to some distribution P_ξ. We also set $F_i^\mu \sim \mathcal{N}(0, 1/N)$ i.i.d.. The recursion is a closed set of equations over the variables $\hat{\rho}^t, \hat{q}_0^t, A_0^t, A_1^t, \rho^t, q_0^t, V_0^t$ and V_1^t Initializing at time $t = 0$ the variables ρ, q_0, V_0 and V_1, the SE equations for $t \geq 1$:

$$\hat{\rho}^t = \alpha \mathbb{E}\big[\partial_z \partial_\omega \phi^{\text{out}}(\omega^t, V_0^{t-1}, V_1^{t-1}, h(z,\xi), m)\big] \quad (34)$$

$$\hat{q}_0^t = \alpha \mathbb{E}\big[\big(\partial_\omega \phi^{\text{out}}(\omega^t, V_0^{t-1}, V_1^{t-1}, y, m)\big)^2\big] \quad (35)$$

$$A_0^t = \alpha \mathbb{E}\big[2\partial_{V_1} \phi^{\text{out}}(\omega^t, V_0^{t-1}, V_1^{t-1}, y, m)\big] - \hat{q}_0^t \quad (36)$$

$$A_1^t = \alpha \mathbb{E}\big[-\partial_\omega^2 \phi^{\text{out}}(\omega^t, V_0^{t-1}, V_1^{t-1}, y, m)\big] + m A_0^t, \quad (37)$$

where the expectation is over the process $\omega^t \sim \mathcal{N}(0, q_0^{t-1})$, $z \sim \mathcal{N}(\rho^{t-1}/q_0^{t-1}\,\omega^t,\ \mathbb{E}[x_0^2] - (\rho^{t-1})^2/q_0^{t-1})$, $\xi \sim P_\xi$ and $y \sim h(z, \xi)$. Also, we have a second set of equations that read:

$$\rho^t = \mathbb{E}\big[x_0\, \partial_B \phi^{\text{in}}(B^t, A_0^t, A_1^t)\big] \quad (38)$$

$$q_0^t = \mathbb{E}\big[\big(\partial_B \phi^{\text{in}}(B^t, A_0^t, A_1^t, m)\big)^2\big] \quad (39)$$

$$V_0^t = \mathbb{E}\big[-2\partial_{A_1} \phi^{\text{in}}(B^t, A_0^t, A_1^t, m)\big] - q_0^t \quad (40)$$

$$V_1^t = \mathbb{E}\big[\partial_B^2 \phi^{\text{in}}(B^t, A_0^t, A_1^t)\big] - m V_0^t \quad (41)$$

where the expectation is over the Markov chain $x_0 \sim P_0$, $B^t \sim \mathcal{N}(\hat{\rho}^t x_0, \hat{q}_0^t)$.

The trajectories of V_0^t, V_1^t, A_0^t and A_1^t in GASP concentrate for large N on their expected value given by the SE dynamics. In order to frame the GASP State Evolution in the rigorous setting of Ref.(Javanmard & Montanari, 2013), we define a slightly different message-passing by replacing their GASP values for a given realization of the problem with the correspondent sample-independent SE values. Also, we replace $c_{\boldsymbol{F}}$ with the expected value $1/N$. Let us define the denoising functions:

$$\eta^{\text{out}}(\omega, y, t) = \partial_\omega \phi^{\text{out}}(\omega, V_0^{t-1}, V_1^{t-1}, y, m) \quad (42)$$

$$\eta^{\text{in}}(B, t) = \partial_B \phi^{\text{in}}(B, A_0^t, A_1^t, m) \quad (43)$$

and their vectorized extensions $\boldsymbol{\eta}^{\text{out}}(\boldsymbol{\omega}, \boldsymbol{y}, t) = (\eta^{\text{out}}(\omega_1, y_1, t), \ldots, \eta^{\text{out}}(\omega_M, y_M, t))$ and $\boldsymbol{\eta}^{\text{in}}(\boldsymbol{\omega}, \boldsymbol{y}, t) = (\eta^{\text{in}}(B_1, t), \ldots, \eta^{\text{in}}(B_N, t))$. The modified GASP message-passing then reads

$$\boldsymbol{\omega}^t = \boldsymbol{F}\, \boldsymbol{\eta}^{\text{in}}(\boldsymbol{B}^{t-1}, t-1) - d_{t-1}^{\text{in}} \boldsymbol{\eta}^{\text{out}}(\boldsymbol{\omega}^{t-1}, \boldsymbol{y}, t-1) \quad (44)$$

$$\boldsymbol{B}^t = \boldsymbol{F}^T\, \boldsymbol{\eta}^{\text{out}}(\boldsymbol{\omega}^t, \boldsymbol{y}, t) - d_t^{\text{out}} \boldsymbol{\eta}^{\text{in}}(\boldsymbol{B}^{t-1}, t-1) \quad (45)$$

where the divergence terms are given by

$$d_t^{\text{in}} = \frac{1}{N} \sum_{i=1}^{N} \partial_B \eta^{\text{in}}(B_i^t, t)$$
$$d_t^{\text{out}} = \frac{1}{N} \sum_{\mu=1}^{M} \partial_\omega \eta^{\text{out}}(\omega_\mu^t, y_\mu, t) \quad (46)$$

Message-passing (44, 45) falls within the class of AMP algorithms analyzed in Ref. (Javanmard & Montanari, 2013) (under some further technical assumptions, see Proposition 5 there). Therefore, it can be rigorously tracked by the SE Eqs. (34,41) in the sense specified in that work. In particular, denoting here $\hat{\boldsymbol{x}}^t = \boldsymbol{\eta}^{\text{in}}(\boldsymbol{B}^t, t)$, we have have almost sure converge in the large system limit of the overlap with the true signal and of the norm of $\hat{\boldsymbol{x}}^t$ to their SE estimates:

$$\lim_{N\to\infty} \frac{1}{N} \langle \hat{\boldsymbol{x}}^t, \boldsymbol{x}_0 \rangle = \rho^t \quad (47)$$

$$\lim_{N\to\infty} \frac{1}{N} \langle \hat{\boldsymbol{x}}^t, \hat{\boldsymbol{x}}^t \rangle = q_0^t \quad (48)$$

In Fig. 1(Bottom), we compare the SE dynamics to the original GASP one (Alg. 1). We compare SE prediction for the evolution of the overlap ρ to that observed in 10 sample trajectories of GASP at $N = 1000$, for a sampling ratio of $\alpha = 2$ and different values of m. The initial estimate $\hat{\boldsymbol{x}}^{t=0}$ in GASP was set to be a mixture $\hat{\boldsymbol{x}}^{t=0} \sim \mathcal{N}(0, I_N) + 0.1\boldsymbol{x}_0$. Therefore we initialize SE with $\rho^{t=0} = 0.1$, and $q^{t=0} = 1 + (\rho^{t=0})^2$. Moreover, we set $V_0^{t=0} = V_1^{t=0} = 1$ for both. As expected, we observe a good agreement between the two dynamics.

6. Effective Landscape and Message-Passing Algorithms

The posterior distribution of statistical models in the hard phase is known to be riddled with glassy states (Antenucci et al., 2019a) preventing the retrieval of the true signal, a situation which is exacerbated in the low temperature limit corresponding to MAP estimation.

Within the replica formalism, the 1RSB free energy provides a description of this complex landscape. The Parisi parameter m allows to select the contributions of different families of states. More specifically m acts as an inverse temperature

coupled to the internal free energy of the states: increasing m selects families of states with lower complexity (i.e., states that are less numerous) and lower free energy.

The fixed points of the State Evolution of GASP are in one-to-one correspondence to the stationary points of the 1RSB free energy, and while the role of m in the dynamics of SE is difficult to analyze, some insights can be gained from the static description given by the free energy.

For phase retrieval in the MAP setting without regularization, a stable fixed-point of GAMP can be found in the space orthogonal to the signal (i.e. at overlap $\rho = 0$) for values of the sampling ratio below $\alpha^{\text{GAMP}} \approx 2.48$ (Ma et al., 2018), which is the algorithmic threshold for GAMP. For GASP instead, it is possible to see that the uninformative fixed-point is stable only below $\alpha^{\text{GASP}} \approx 1.5$, a noticeable improvement of the threshold with respect to GAMP. This is obtained by choosing the m corresponding to lowest complexity states according to the 1RSB free energy (see Sec. 3 of the SM for further details). As we will see in the following, both these thresholds can be lowered by employing a continuation strategy for the regularizer.

A thorough description of the results of the replica analysis and of the landscape properties for GLE models will be presented in a more technical future work.

7. MAP estimation with an L_2 regularizer

The objective function introduced in Eq. (2) contains a L_2 regularization term weighted by an intensity parameter λ.

Regularization plays and important role in reducing the variance of the inferred estimator, and can be crucial when the observations are noise-affected, since it lowers the sensitivity of the learned model to deviations in the training set. However, as observed in (Ma et al., 2018; 2019; Balan, 2016), regularization is also useful for its smoothing effect, and can be exploited in non-convex optimization problems even in the noiseless setting. When the regularization term is turned up, the optimization landscape gradually simplifies and it becomes easier to reach a global optimizer. However, the problem of getting stuck in bad local minima is avoided at the cost of introducing a bias. The *continuation* strategy is based on the fact that such biased estimator might be closer than the random initial configuration to the global optimizer of the unregularized objective : in a multi-stage approach, regularization is decreased (down to zero) after each warm restart.

Among the many possible continuation schedules for λ (a little decrease after each minimization, or, as in AMP.A, at the end of each iteration) in this paper we choose a simple two-stage approach: first we run GASP till convergence with a given value of $\lambda > 0$, then we set $\lambda = 0$ in the successive iterations.

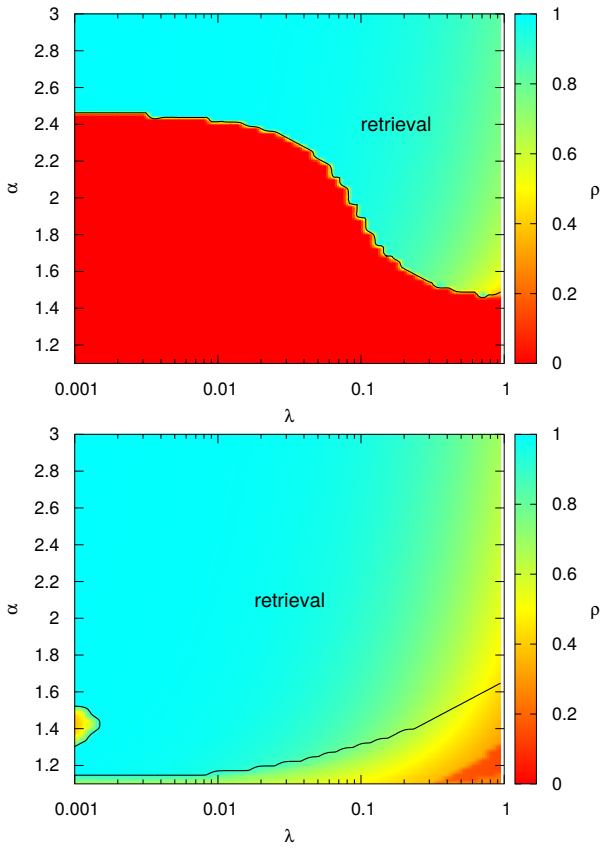

Figure 2. Phase diagrams corresponding to the SE asymptotic analysis of GAMP (top) and GASP (bottom). The color maps indicate the overlap ρ reached at convergence in the presence of an L_2 regularizer of intensity λ.

In Fig.2, we compare the asymptotic performance (tracked by SE) of GAMP and GASP for the phase retrieval problem with an L_2 regularization. The color map indicates the overlap with the signal reached at the end of the first stage of our continuation strategy (with $\lambda \neq 0$), while the black curves delimit the perfect retrieval regions, where the overlap reached at the end of stage two (with $\lambda = 0$) is $\rho = 1$.

In both cases we set the initial variances Δ to 1, and consider an initial condition with a small positive overlap with the signal, $\rho = 0.1$. An assumption of this kind is indeed needed to ensure that we avoid lingering on the fixed-point at $\rho = 0$; however, the specific value of ρ can be chosen arbitrarily (e.g., it could be taken much smaller without affecting the phase diagrams). Even in real-world applications, it is often the case that the non-orthogonality requirement is easily met, for example in many imaging applications the signal is known to be real non-negative. As explained in the previous section, we also set $q_0 = 1 + \rho^2$ in the initialization of the self-overlap parameter.

In the GASP phase diagram, for each α and λ, the value of m was set to the thermodynamic optimum value $m^\star$ (obtained at $\rho = 0$), and was kept fixed throughout the two stages of our continuation strategy. This $m^\star$ can be obtained by optimizing the 1RSB free energy over the symmetry-breaking parameter; the numerical values of m, corresponding to the points in the plot, can be found in Sec. 3 of the SM, in Fig. 1. It is not strictly necessary to fix m to this specific value, as any value in a broad range of around $m^\star$ will still be effective (see for example Fig. 2 in the SM). As expected from the numerical experiments at $\lambda = 0$, we can see from Fig. 2 that when the regularizer becomes too small an uninformative fixed-point (in $\rho = 0$) becomes attractive for the dynamics of GASP and signal recovery becomes impossible below $\alpha_{alg} \sim 1.5$ (we expect also the recovery region with $\alpha \in [1.13, 1.3]$ at $\lambda = 0.001$ to shrink and close when the regularizer is further decreased).

It is clear that the introduction of an L_2-norm is crucial for reducing the algorithmic gap of both GAMP and GASP (the information theoretic threshold is $\alpha_{\rm IT} = 1$), as previously observed in (Ma et al., 2018; 2019). In this work we find that also in GLE problems, when the mismatched setting is considered (and inference happens off the Nishimori line (Nishimori, 2001; Antenucci et al., 2019b)), the more fitting geometrical picture provided by the 1RSB ansatz can be exploited algorithmically: with a simple continuation strategy it is possible to lower the algorithmic threshold of GASP down to the Bayes-optimal value $\alpha = 1.13$.

8. Discussion

We presented Generalized Approximate Survey Propagation, a novel algorithm designed to improve over AMP in the context of GLE inference problems, when faced with a mismatch between assumed and true generative model. The algorithm, parametrized by the symmetry-breaking parameter m, allows one to go beyond some symmetry assumptions at the heart of the previous algorithms, and proves to be more suited for the MAP estimation task considered in this work.

In the prototypical case of real-valued phase retrieval, we have shown that with little tuning of m it is possible to modify the effective landscape explored during the message-passing dynamics and avoid getting stuck in otherwise attractive uninformative fixed points. Furthermore, we have seen that, even in the noiseless case, a simple continuation strategy, based on the introduction of an L_2 regularizer, can guide GASP close enough to the signal and allow its recovery, extending the region of parameters where GASP is more effective than GAMP. In some cases we observed that GASP can achieve perfect retrieval until the Bayes-optimal threshold, at the sampling ratio $\alpha \sim 1.13$. We also derived the 1RSB State Evolution equations, and showed that they can be used as a simple tool for tracking the asymptotic behaviour of GASP.

We delay a comprehensive analysis of the landscape associated to GLE models to a more technical publication, where we will also deal with the case of noisy observation channels. A straightforward follow-up of the present work could focus on the search for an adaptation scheme for the L_2 regularizer, possibly extending the work of Refs. (Ma et al., 2018; 2019), and more importantly, for a criterion to identify the best setting for the symmetry-breaking parameter. Another possible future line of work could go in the direction of relaxing some of the assumptions made in deriving the GASP algorithm over the observation matrix. This could motivate the derivation of a 1RSB version of the Vector Approximate Message Passing equations (Schniter et al., 2016). Also, the extension of GASP to deep non-linear inference model, along the lines of Ref. (Manoel et al., 2017; Fletcher et al., 2018) seems to be promising and technically feasible.

CL thanks Junjie Ma for sharing and explaining the code of their AMP.A algorithm.

References

Advani, M. and Ganguli, S. Statistical mechanics of optimal convex inference in high dimensions. *Physical Review X*, 6(3):031034, 2016.

Antenucci, F., Franz, S., Urbani, P., and Zdeborová, L. Glassy nature of the hard phase in inference problems. *Physical Review X*, 9(1):011020, 2019a.

Antenucci, F., Krzakala, F., Urbani, P., and Zdeborová, L. Approximate survey propagation for statistical inference. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(2):023401, 2019b.

Balan, R. Reconstruction of signals from magnitudes of redundant representations: The comple case. *Foundations of Computational Mathematics*, 16(3):677–721, 2016.

Barbier, J., Krzakala, F., Macris, N., Miolane, L., and Zdeborová, L. Optimal errors and phase transitions in high-dimensional generalized linear models. In *Conference On Learning Theory*, pp. 728–731, 2018.

Bayati, M. and Montanari, A. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57 (2):764–785, 2011.

Braunstein, A., Mézard, M., and Zecchina, R. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

Caltagirone, F., Zdeborová, L., and Krzakala, F. On convergence of approximate message passing. In *Information*

Theory (ISIT), 2014 IEEE International Symposium on, pp. 1812–1816. IEEE, 2014.

Candes, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Charbonneau, P., Kurchan, J., Parisi, G., Urbani, P., and Zamponi, F. Glass and jamming transitions: From exact results to finite-dimensional descriptions. *Annual Review of Condensed Matter Physics*, 8:265–288, 2017.

Chen, Y. and Candes, E. Solving random quadratic systems of equations is nearly as easy as solving linear systems. In *Advances in Neural Information Processing Systems*, pp. 739–747, 2015.

Chen, Y., Chi, Y., Fan, J., and Ma, C. Gradient descent with random initialization: Fast global convergence for nonconvex phase retrieval. *Mathematical Programming*, pp. 1–33, 2018.

Deshpande, Y. and Montanari, A. Information-theoretically optimal sparse pca. In *2014 IEEE International Symposium on Information Theory*, pp. 2197–2201. IEEE, 2014.

Deshpande, Y., Abbe, E., and Montanari, A. Asymptotic mutual information for the binary stochastic block model. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 185–189. IEEE, 2016.

Dhifallah, O. and Lu, Y. M. Fundamental limits of phasemax for phase retrieval: A replica analysis. In *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2017 IEEE 7th International Workshop on*, pp. 1–5. IEEE, 2017.

Donoho, D. and Montanari, A. High dimensional robust m-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166 (3-4):935–969, 2016.

Donoho, D. L., Maleki, A., and Montanari, A. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Fletcher, A. K., Rangan, S., and Schniter, P. Inference in deep networks in high dimensions. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1884–1888. IEEE, 2018.

Goldstein, T. and Studer, C. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, 2018.

Guo, D. and Wang, C.-C. Asymptotic mean-square optimality of belief propagation for sparse linear systems. In *Information Theory Workshop, 2006. ITW'06 Chengdu. IEEE*, pp. 194–198. IEEE, 2006.

Jain, P., Netrapalli, P., and Sanghavi, S. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 665–674. ACM, 2013.

Javanmard, A. and Montanari, A. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

Kabashima, Y., Krzakala, F., Mézard, M., Sakata, A., and Zdeborová, L. Phase transitions and sample complexity in bayes-optimal matrix factorization. *IEEE Transactions on Information Theory*, 62(7):4228–4265, 2016.

Krzakała, F., Montanari, A., Ricci-Tersenghi, F., Semerjian, G., and Zdeborová, L. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

Krzakala, F., Ricci-Tersenghi, F., Zdeborova, L., Zecchina, R., Tramel, E. W., and Cugliandolo, L. F. *Statistical Physics, Optimization, Inference, and Message-Passing Algorithms: Lecture Notes of the Les Houches School of Physics-Special Issue, October 2013*. Oxford University Press, 2016.

Lu, Y. M. and Li, G. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *arXiv preprint arXiv:1702.06435*, 2017.

Luo, W., Alghamdi, W., and Lu, Y. M. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Transactions on Signal Processing*, 2019.

Ma, J., Xu, J., and Maleki, A. Approximate message passing for amplitude based optimization. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3365–3374, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR. URL http://proceedings.mlr.press/v80/ma18e.html.

Ma, J., Xu, J., and Maleki, A. Optimization-based amp for phase retrieval: The impact of initialization and l2-regularization. *IEEE Transactions on Information Theory*, 2019.

Manoel, A., Krzakala, F., Mézard, M., and Zdeborová, L. Multi-layer generalized linear estimation. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pp. 2098–2102. IEEE, 2017.

Mézard, M. Mean-field message-passing equations in the hopfield model and its generalizations. *Physical Review E*, 95(2):022117, 2017.

Mezard, M. and Montanari, A. *Information, physics, and computation*. Oxford University Press, 2009.

Mézard, M., Parisi, G., and Virasoro, M. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

Mézard, M., Parisi, G., and Zecchina, R. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

Monasson, R. Structural glass transition and the entropy of the metastable states. *Physical review letters*, 75(15): 2847, 1995.

Mondelli, M. and Montanari, A. Fundamental limits of weak recovery with applications to phase retrieval. *Foundations of Computational Mathematics*, pp. 1–71, 2018.

Mukherjee, S. and Seelamantula, C. S. Phase retrieval from binary measurements. In *IEEE Signal Processing Letters*, volume 25, pp. 348–352. IEEE, 2018.

Nishimori, H. *Statistical physics of spin glasses and information processing: an introduction*, volume 111. Clarendon Press, 2001.

Rangan, S. Estimation with random linear mixing, belief propagation and compressed sensing. In *Information Sciences and Systems (CISS), 2010 44th Annual Conference on*, pp. 1–6. IEEE, 2010.

Rangan, S. Generalized approximate message passing for estimation with random linear mixing. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2168–2172. IEEE, 2011.

Rangan, S. and Fletcher, A. K. Iterative estimation of constrained rank-one matrices in noise. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 1246–1250. IEEE, 2012.

Ricci-Tersenghi, F., Semerjian, G., and Zdeborová, L. Typology of phase transitions in bayesian inference problems. *Physical Review E*, 99(4):042109, 2019.

Ros, V., Arous, G. B., Biroli, G., and Cammarota, C. Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions. *Physical Review X*, 9(1):011003, 2019.

Schniter, P. and Rangan, S. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

Schniter, P., Rangan, S., and Fletcher, A. K. Vector approximate message passing for the generalized linear model. In *Signals, Systems and Computers, 2016 50th Asilomar Conference on*, pp. 1525–1529. IEEE, 2016.

Sun, J., Qu, Q., and Wright, J. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18 (5):1131–1198, 2018.

Supplementary Material

May 15, 2019

A. A recap on Generalized Approximate Message Passing

A.1. Derivation of GAMP

For the reader's convenience and for familiarizing with the notation adopted throughout this work, we sketch the derivation of the Generalized Approximate Message Passing (GAMP) equations for Generalized Linear Estimation (GLE) models. For a longer discussion, we refer the reader to Refs. (Rangan, 2011; Ma et al., 2018; Kabashima et al., 2016). We assume the setting of Eq. (1) of the Main Text, that is a graphical model defined by the Hamiltonian:

$$\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}) = \sum_{\mu} \ell\big(y_\mu,\, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle\big) + \sum_i r(x_i), \tag{49}$$

with the further assumption that the entries of $\boldsymbol{F}$ are i.i.d. zero-mean Gaussian variables with variance $1/N$, i.e $F_i^\mu \sim \mathcal{N}(0, 1/N)$ (but the derivation also applies to non-Gaussian variables with the same mean and variance). The configuration space is assumed to be some subset χ^N of $\mathbb{R}$. For discrete spaces, integrals should be replace with summations. Also, we consider the regime of large M and N, with finite $\alpha = M/N$. The starting point for the derivation of GAMP equations is the Belief Propagation (BP) algorithm (Mezard & Montanari, 2009), characterized by the exchange of two sets of messages:

$$\nu^t_{i\to\mu}(x_i) \propto e^{-\beta r(x_i) + \sum_{\nu\neq\mu} \log \hat{\nu}^t_{\nu\to i}(x_i)} \tag{50}$$

$$\hat{\nu}^{t+1}_{\mu\to i}(x_i) \propto \int_{\chi^{N-1}} \prod_{j\neq i} d\nu^t_{j\to\mu}(x_j)\; e^{-\beta \ell(y_\mu, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle)}. \tag{51}$$

For the dense graphical model we are considering, by virtue of central limit arguments, we can relax the resulting identities among probability densities to relations among their first and second moments. The resulting approximated version of BP goes under the name of relaxed Belief Propagation (rBP) (Guo & Wang, 2006; Rangan, 2010; Mézard, 2017).

We define the expectations over the measure in Eq.(50) as $\langle \bullet \rangle^t_{i\to\mu}$, and its moments as $\langle x\rangle^t_{i\to\mu} = \hat{x}^t_{i\to\mu}$ and $\langle x^2\rangle^t_{i\to\mu} = \Delta^t_{i\to\mu} + (\hat{x}^t_{i\to\mu})^2$. In high dimensions we can see that the scalar product $\langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle$ in Eq.(51) becomes Gaussian distributed according to $\mathcal{N}(\sum_j F^\mu_j \hat{x}^t_{j\to\mu} + F^\mu_i(x_i - \hat{x}^t_{i\to\mu}), \sum_{j\neq i}(F^\mu_j)^2 \Delta^t_{j\to\mu})$.

In order to obtain the relationship between the moments of the two sets of distributions it is useful to introduce two scalar estimation functions, the input and output channels, that fully characterize the problem. The associated free entropies (Barbier et al., 2018) (i.e., log-normalization factors) can be expressed as:

$$\varphi^{\text{in}}(B, A) = \log \int_\chi dx\; e^{-\frac{1}{2}Ax^2 + Bx - \beta\, r(x)} \tag{52}$$

$$\varphi^{\text{out}}(\omega, V, y) = \log \int \frac{dz}{\sqrt{2\pi V}}\; e^{-\frac{1}{2V}(z-\omega)^2 - \beta\, \ell(y,z)}. \tag{53}$$

Then, defining $g^t_\mu = \partial_\omega \varphi^{\text{out}}(\omega', V', y)$ and $\Gamma^t_\mu = -\partial^2_\omega \varphi^{\text{out}}(\omega', V', y)$, both evaluated in $\omega' = \sum_j (F^\mu_j)^2 \Delta^t_{j\to\mu}$ and $V' = \sum_j F^\mu_j \hat{x}^t_{j\to\mu}$, we can express through them the approximate message-passing, obtained at the second order of the Taylor expansion of the messages:

$$\log \hat{\nu}^{t+1}_{\mu\to i}(x_i) = \varphi^{\text{out}}\left(\sum_j F^\mu_j \hat{x}^t_{j\to\mu} + F^\mu_i(x_i - \hat{x}^t_{i\to\mu}),\; \sum_{j\neq i}(F^\mu_j)^2 \Delta^t_{j\to\mu},\; y_\mu\right) + \text{const.} \tag{54}$$

Next, we close the equations on single site quantities, discarding terms which are sub-leading for large N and assuming zero mean and 1/N variance i.i.d entries in $\boldsymbol{F}$. Thus, we can remove the cavities and approximate the parameters of the (non-cavity) estimation channels as follows:

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t - \hat{x}_i^{t-1} \sum_\mu (F_i^\mu)^2 \Gamma_\mu^t \tag{55}$$

$$A_i^t = \sum_\mu (F_i^\mu)^2 \Gamma_\mu^t \tag{56}$$

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^t - g_\mu^t \sum_i (F_i^\mu)^2 \Delta_i^t \tag{57}$$

$$V_\mu^t = \sum_i (F_i^\mu)^2 \Delta_i. \tag{58}$$

Finally, the expectations introduced above can be obtained via the derivatives:

$$g_\mu^t = \partial_\omega \varphi_\mu^{\text{out},t} \tag{59}$$

$$\Gamma_\mu^t = -\partial_\omega^2 \varphi_\mu^{\text{out},t} \tag{60}$$

$$\hat{x}_i^t = \partial_B \varphi_i^{\text{in},t} \tag{61}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{\text{in},t}, \tag{62}$$

where we used the shorthand notation $\varphi_i^{\text{in},t} = \varphi^{\text{in}}(B_i^t, A^t)$ and $\varphi_\mu^{\text{out},t} = \varphi^{\text{out}}(\omega_\mu^t, V^{t-1}, y)$.

A slight simplification of the message passing (which involves $\mathcal{O}(N^2)$ operations per iteration), relies on the observation that due to the statistical properties of $\boldsymbol{F}$ the quantities A_i and V_μ do not depend on their indexes (Rangan, 2011), so we can define their scalar counterparts:

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t, \tag{63}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^{t-1}, \tag{64}$$

where $c_{\boldsymbol{F}} = \sum_{\mu,i} (F_i^\mu)^2/(MN) \approx 1/N$. Therefore we obtain:

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1} V^{t-1} \tag{65}$$

$$g_\mu^t = \partial_\omega \varphi_\mu^{\text{out},t} \tag{66}$$

$$\Gamma_\mu^t = -\partial_\omega^2 \varphi_\mu^{\text{out},t} \tag{67}$$

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t \tag{68}$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1} A^t \tag{69}$$

$$\hat{x}_i^t = \partial_B \varphi_i^{\text{in},t} \tag{70}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{\text{in},t} \tag{71}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^t. \tag{72}$$

Eqs. (65-72) are known as the GAMP iterations, and are valid for $t \geq 1$, given some initial condition $\hat{\boldsymbol{x}}^{t=0}$ and $V^{t=0}$, along with $g_\mu^{t=0} = 0,\ \forall \mu$.

A.2. Zero-temperature limit of GAMP

In order to apply the GAMP algorithm to MAP estimation or MAP + regularizer, we have to consider the zero-temperature limit $\beta \uparrow \infty$. The limiting form of the equations depends on the model and on the regime (e.g. low or high α). Here we

consider models defined on continuous spaces χ^N and in the high α regime (e.g. $\alpha > 1$ for phase retrieval). In this case, while taking the limit, the message have to be rescaled appropriately in order for them to stay finite. Therefore we rescale the messages through the substitutions:

$$A \to \beta A \tag{73}$$
$$B_i \to \beta B_i \tag{74}$$
$$V \to V/\beta \tag{75}$$
$$g_\mu \to \beta g_\mu \tag{76}$$
$$\Delta_i \to \Delta_i/\beta. \tag{77}$$

With these rescalings, the GAMP equations (65-72) are left unaltered, but the expressions for the free entropies of the scalar channels become

$$\varphi^{\text{in}}(B, A) = \max_{x\in\chi} \; -r(x) - \frac{1}{2}Ax^2 + Bx \tag{78}$$
$$\varphi^{\text{out}}(\omega, V, y) = \max_{z} \; -\frac{(z-\omega_\mu)^2}{2V} - \ell(y, z), \tag{79}$$

as it is easy to verify.

A.3. GAMP equations for real-valued phase retrieval and AMP.A equations

In the special case of the phase retrieval problem, with a loss $\ell(y, \omega) = (y - |\omega|)^2$ and L_2-norm $r(x) = \lambda x^2/2$ and at zero temperature, the two scalar estimation channels of Eqs.(78) and (79) become:

$$\varphi^{\text{in}}(B, A) = \frac{B^2}{2(A+\lambda)} \tag{80}$$
$$\varphi^{\text{out}}(\omega, V, y) = -\frac{(y-|\omega|)^2}{2V+1}. \tag{81}$$

Thus, Eqs. (66, 70, 71, 72) simply yield:

$$g_\mu^t = \frac{2(y_\mu - |\omega_\mu^t|)}{2V^t+1}\,\text{sign}(\omega_\mu^t) \tag{82}$$
$$\hat{x}_i^t = \frac{B_i^t}{A^t+\lambda} \tag{83}$$
$$\Delta_i^t = \frac{1}{A^t+\lambda} \tag{84}$$
$$V^t = N c_{\boldsymbol{F}} \frac{1}{A^t+\lambda}. \tag{85}$$

Eq. (67) is instead singular, since it involves the derivative of the sign function. Since we have

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - \frac{g_\mu^{t-1}}{A^{t-1}+\lambda} \tag{86}$$
$$g_\mu^t = \frac{2(y_\mu - |\omega_\mu^t|)}{2V^t+1}\,\text{sign}(\omega_\mu^t) \tag{87}$$
$$A^t = -c_{\boldsymbol{F}} \sum_\mu \partial_\omega^2 \varphi_\mu^{\text{out},t} \tag{88}$$
$$x_i^t = (A^t+\lambda)\left(\sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1} A^t\right), \tag{89}$$

because of the singularity, the value of A^t cannot be simply evaluated on a given finite sample. A possible way of dealing with this issue is to use a smoothing strategy in the first iterations of the message passing, replacing the sign function with a

continuous version of it. Alternatively, in Ref. (Ma et al., 2018; 2019), the author propose to self-consistently adapt the regularizer λ at each time step in order to absorb the divergent contribution. Also, the dynamics A^t can be replaced by the corresponding and non-singular SE estimate. We find that all these solutions are difficult to implement in a robust way and lead to some numerical instabilities that have to be dealt with great care. As we commented in the Main Text, this problem does not affect the GASP version of the algorithm, because of the additional Gaussian kernel that smoothens the output scalar estimation channel.

B. Derivation of Generalized Approximate Survey Propagation

We will derive the GASP equation for a general GLE model specified by (49). As already explained in the Main Text, we will follow (Antenucci et al., 2019b) and work within the (real) replicas formalism. The derivation is similar to the one outlined for the GAMP algorithm, which goes from Belief Propagation (BP) to relaxed Belief Propagation (rBP) to Approximate Message Passing (AMP). In fact, GASP is obtained by applying the very same procedure that leads to GAMP to an auxiliary graphical model that corresponds to considering multiple copies of the system.

B.1. Relaxed Survey Propagation

As an intermediate step toward the derivation of GASP equations, we derive the relaxed Survey Propagation (rSP) equations for out GLE problem. This corresponds to a Gaussian closure of the standard BP equations on the replicated factor graph of the problem, under replica symmetric assumptions. We assume the setting of Eq. (1) of the Main Text, that is a graphical model defined by the Hamiltonian:

$$\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}) = \sum_{\mu} \ell\big(y_\mu, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle\big) + \sum_i r(x_i), \tag{90}$$

with the further assumption that the entries of $\boldsymbol{F}$ are i.i.d. zero-mean Gaussian variables with variance $1/N$, i.e $F_i^\mu \sim \mathcal{N}(0, 1/N)$ (but the derivation also applies to non-Gaussian variables with the same mean and variance). The configuration space is assumed to be some subset χ^N of $\mathbb{R}$. For discrete spaces, integrals should be replaced with summations. Also, we consider the regime of large M and N, with finite $\alpha = M/N$.

Quite peculiarly, the family of message passing algorithm corresponding to the 1RSB framework (i.e. SP, rSP, ASP), are simply obtained as the BP, rBP and AMP equations for a *replicated* graphical model,

$$p(\{\boldsymbol{x}^a\}_{a=1}^m) = \frac{1}{Z^m_{\boldsymbol{y},\boldsymbol{F}}} e^{-\beta \sum_{a=1}^m \mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}^a)}, \tag{91}$$

where m is the number of replicas. The parameter m is not to be confused with the number of replica n that it is usually sent to zero in the replica trick, but it has to be interpreted as the Parisi symmetry breaking parameter in the 1RSB scheme or as the number of real clones within Monasson's method (Monasson, 1995)). While the replicated model is trivially factorized over the replicas, a highly non-trivial picture emerges when p is considered as the limit distribution obtained by inserting a coupling term among the replicas and then letting it go to zero. Since the discussion about this technique (pioneered by Monasson in Ref. (Monasson, 1995)) is quite articulated and has its root in a few decades of development in spin-glass theory, we refer the interested reader to (Mézard et al., 1987; Mezard & Montanari, 2009; Antenucci et al., 2019a;b) and reference therein for an overview of the theoretical aspects behind this approach. From here on we present the innovative aspects of our contribution, which extends the work of Ref. (Antenucci et al., 2019b) to GLE models.

We denote with $\boldsymbol{x}_i \in \xi^m$ the replicated variable on site i, and write a first set of BP equations in the form:

$$\nu_{i\to\mu}(\bar{\boldsymbol{x}}_i) \propto e^{-\beta \sum_{a=1}^m r(x_i^a) + \sum_{\nu\neq\mu} \log \hat{\nu}_{\nu\to i}(\bar{\boldsymbol{x}}_i)}, \tag{92}$$

where we omit time indexes. In the large N limit, we can exploit the statistical assumptions on $\boldsymbol{F}$ and the central limit theorem to perform a Gaussian approximation of the messages. Also, we assume symmetry of the messages $\nu_{i\to\mu}(\bar{\boldsymbol{x}}_i)$ under permutation of replica indexes, which holds self-consistently if one makes a similar assumption also on the messages $\hat{\nu}_{\nu\to i}(\bar{\boldsymbol{x}}_i)$. Messages are then multivariate Gaussian distribution conveniently parametrized by the mean $\hat{x}_{i\to\mu}$ and two parameters $\Delta_{0,i\to\mu}$ and $\Delta_{1,i\to\mu}$ in the form:

$$\nu_{i\to\mu}(\bar{\boldsymbol{x}}_i) \propto \int \mathrm{d}\, h\, e^{-\frac{1}{2\Delta_{0,i\to\mu}}(h-\hat{x}_{i\to\mu})^2} \prod_a e^{-\frac{1}{2\Delta_{1,i\to\mu}}(x_i^a - h)^2}, \tag{93}$$

also known as caging ansatz in the glass and spin-glass community (Charbonneau et al., 2017). According to this Gaussian projection, the first and second moments of messages are given by

$$\langle x_i^a \rangle_{i\to\mu} = \hat{x}_{i\to\mu} \tag{94}$$
$$\langle x_i^a x_i^b \rangle_{i\to\mu} = \Delta_{0,i\to\mu} + \hat{x}_{i\to\mu}^2 \tag{95}$$
$$\langle (x_i^a)^2 \rangle_{i\to\mu} = \Delta_{1,i\to\mu} + \Delta_{0,i\to\mu} + \hat{x}_{i\to\mu}^2. \tag{96}$$

The values of $\hat{x}_{i\to\mu}$, $\Delta_{0,i\to\mu}$ and $\Delta_{1,i\to\mu}$ can be obtained by matching the moments of the r.h.s. of 92. From now on the derivation is very close to that of Section A.1 for GAMP, therefor we relax the notation and drop some indexes. Let us define the input channel free entropy:

$$\phi^{\text{in}}(B, A_0, A_1, m) = \frac{1}{m} \log \int Dz \left(\int_\chi dx \; e^{-\beta r(x) - \frac{1}{2} A_1 x^2 + (B + \sqrt{A_0} z) x} \right)^m. \tag{97}$$

Let us also denote with $\langle \psi(\bar{\boldsymbol{x}}) \rangle$ the expectation over the corresponding measure, in the m-replicated space, of a test function ψ, that is

$$\langle \psi(\bar{\boldsymbol{x}}) \rangle = \frac{\int Dz \int_{\chi^m} \prod_{a=1}^m dx^a \; e^{-\beta r(x^a) - \frac{1}{2} A_1 (x^a)^2 + (B + \sqrt{A_0} z) x^a} \; \psi(\bar{\boldsymbol{x}})}{\int Dz \prod_{a=1}^m dx^a \; e^{-\beta r(x^a) - \frac{1}{2} A_1 (x^a)^2 + (B + \sqrt{A_0} z) x^a}}. \tag{98}$$

For appropriate values of $B_{\nu\to i}$, $A_{0,\nu\to i}$ and $A_{1,\nu\to i}$ to be determined by second order expansion of $\log \hat{\nu}_{\nu\to i}(\bar{x}_i)$, and for replica indexes a and b,$a \neq b$, from Eq. (92) we obtain:

$$\partial_B \phi^{\text{in}} = \langle x^a \rangle \tag{99}$$
$$\partial_B^2 \phi^{\text{in}} = (\langle (x^a)^2 \rangle - \langle x^a x^b \rangle) + m \left(\langle x^a x^b \rangle - \langle x^a \rangle^2 \right) \tag{100}$$
$$2\partial_{A_0} \phi^{\text{in}} = (\langle (x^a)^2 \rangle - \langle x^a x^b \rangle) + m \langle x^a x^b \rangle \tag{101}$$
$$2\partial_{A_1} \phi^{\text{in}} = -\langle (x^a)^2 \rangle. \tag{102}$$

Using the above formulas, we can project the measure on $\mathbb{R}^m$ corresponding to ϕ^{in} onto the space of replica-symmetric Gaussian distributions, parametrized by $\hat{x}$, Δ_0 and Δ_1. Defining for convenience $\phi^{\text{in}}_{i\to\mu} = \phi^{\text{in}}(A_{0,i\to\mu}, A_{1,i\to\mu}, B_{i\to\mu})$, with the quantities $A_{0,i\to\mu}$, $A_{1,i\to\mu}$ and $B_{i\to\mu}$ to be defined later, by moment matching we obtain:

$$\hat{x}_{i\to\mu} = \partial_B \phi^{\text{in}}_{i\to\mu}, \tag{103}$$
$$\Delta_{0,i\to\mu} = \frac{1}{m-1} \left(\partial_B^2 \phi^{\text{in}}_{i\to\mu} + 2\partial_{A_1} \phi^{\text{in}}_{i\to\mu} + \hat{x}^2_{i\to\mu} \right), \tag{104}$$
$$\Delta_{1,i\to\mu} = \partial_B^2 \phi^{\text{in}}_{i\to\mu} - m\Delta_{0,i\to\mu}. \tag{105}$$

Defining the messages

$$\omega_{\mu\to i} = \sum_{j\neq i} F_j^\mu \hat{x}_{j\to\mu}, \tag{106}$$
$$V_{0,\mu\to i} = \sum_{j\neq i} \left(F_j^\mu\right)^2 \Delta_{0,j\to\mu}, \tag{107}$$
$$V_{1,\mu\to i} = \sum_{j\neq i} \left(F_j^\mu\right)^2 \Delta_{1,j\to\mu}, \tag{108}$$

we can express the central limit approximation for the BP equations at factor node μ as

$$\hat{\nu}_{\mu\to i}(\bar{\boldsymbol{x}}_i) \propto \int_{\chi^{m(N-1)}} \prod_{j\neq i} \mathrm{d}\, \nu_{j\to\mu}(\bar{\boldsymbol{x}}_j) \; e^{-\beta \sum_a \ell(y_\mu, \langle \boldsymbol{F}^\mu, \boldsymbol{x}^a \rangle)} \tag{109}$$
$$\propto \int \mathrm{d}\, z_0 \; e^{-\frac{1}{2V_{0,\mu\to i}} (z_0 - \omega_\mu)^2} \prod_{a=1}^m \left(\int Dz_1 \; e^{-\beta \; \ell(y_\mu, F_i^\mu x_i^a + z_0 + \sqrt{V_1} z_1)} \right). \tag{110}$$

The expansion of the message $\hat{\nu}_{\mu\to i}(\bar{\boldsymbol{x}}_i)$ that we use for our Gaussian closure of the BP messages are conveniently expressed in terms of the derivatives of the output channel free entropy

$$\phi^{\text{out}}(\omega, V_0, V_1, y, m) = \frac{1}{m}\log\int \frac{\mathrm{d}\, z_0}{\sqrt{2\pi V_0}}\, e^{-\frac{1}{2V_0}(z_0-\omega)^2}\, \left(\int Dz_1\, e^{-\beta\, \ell(y, z_0+\sqrt{V_1}z_1)}\right)^m. \tag{111}$$

Introducing the second order expansion

$$\log\hat{\nu}_{\mu\to i}(\bar{\boldsymbol{x}}_i) = g_{\mu\to i}\sum_a x_i^a - \frac{1}{2}A_{1,\mu\to i}\sum_a (x_i^a)^2 + \frac{1}{2}A_{0,\mu\to i}\sum_{a,b} x_i^a x_i^b \tag{112}$$

we can write the last set of rSP messages as

$$g_{\mu\to i} = \partial_\omega \phi^{\text{out}}_{\mu\to i} \tag{113}$$

$$\Gamma_{0,\mu\to i} = \frac{1}{m-1}\left(\partial^2_\omega \phi^{\text{out}}_{\mu\to i} - (2\partial_{V_1}\phi^{\text{out}}_{\mu\to i} - g^2_{\mu\to i})\right) \tag{114}$$

$$\Gamma_{1,\mu\to i} = \frac{1}{m-1}(\partial^2_\omega \phi^{\text{out}}_{\mu\to i} - m(2\partial_{V_1}\phi^{\text{out}}_{\mu\to i} - g^2_{\mu\to i})) \tag{115}$$

$$\tag{116}$$

Incoming messages on the input nodes are then given by

$$B_{i\to\mu} = \sum_{\nu\neq\mu} F_i^\nu g_{\nu\to i} \tag{117}$$

$$A_{0,i\to\mu} = \sum_{\nu\neq\mu} (F_i^\nu)^2 \Gamma_{0,\nu\to i} \tag{118}$$

$$A_{1,i\to\mu} = \sum_{\nu\neq\mu} (F_i^\nu)^2 \Gamma_{1,\nu\to i} \tag{119}$$

$$\tag{120}$$

The closed set of Equations (103-108) and (113-119), along with the free entropy definitions in Eqs. (97) and (111), define the rSP iterative message passing.

B.2. The GASP Equations

Under our statistical assumptions on the sensing matrix $\boldsymbol{F}$, in order to reduce the computational complexity of rSP, it is possible to close the equations the rSP message passing in terms of single site or scalar quantities $\omega_\mu, g_\mu, \Gamma_{0,\mu}, \Gamma_{1,\mu}, A_0, A_1, B_i, \Delta_{0,i}, \Delta_{1,i}, V_0$ and V_1, therefore obtaining the GASP equation. In fact, the values $A_{0,i\to\mu}, A_{1,i\to\mu}$ and $V_{0,\mu\to i}, V_{1,\mu\to i}$ concentrate and can be straightforwardly replaced by their scalar counterparts. In order to present in this section all of the necessary ingredients of the GASP algorithm, we rewrite here the two scalar channel free entropies from previous section. Adopting a form that makes clear the nested structure of the 1RSB free-entropy and it's relation to the corresponding RS free entropy used in GAMP, we write fro the input channel

$$\phi^{\text{in}}(B, A_0, A_1, m) = \frac{1}{m}\log\int Dz\, e^{m\varphi^{\text{in}}(B+\sqrt{A_0}z,\, A_1)} \tag{121}$$

$$\varphi^{\text{in}}(h, A_1) = \log\int_\chi \mathrm{d}\, x\, e^{-\beta r(x) - \frac{1}{2}A_1 x^2 + hx} \tag{122}$$

and for the output channel

$$\phi^{\text{out}}(\omega, V_0, V_1, y, m) = \frac{1}{m}\log\int Dz\, e^{m\varphi^{\text{out}}(\omega+\sqrt{V_0}\, z,\, V_1, y)}, \tag{123}$$

$$\varphi^{\text{out}}(u, V_1, y) = \log\int Dz\, e^{-\beta\, \ell(y, u+\sqrt{V_1}z)}. \tag{124}$$

As usual, $\int Dz$ denotes standard Gaussian integration $\int dz\, \exp(-z^2/2)/\sqrt{2\pi}$. We will use the notation $\phi_i^{\text{in}} = \phi^{\text{in}}(B_i, A_0, A_1, m)$ and $\phi_\mu^{\text{out}} = \phi^{\text{out}}(\omega_\mu, V_0, V_1, y_\mu, m)$ and drop time indexes for the time being. Given the definition $B_i = \sum_\mu F_i^\mu g_{\mu\to i}$, we can write

$$\hat{x}_{i\to\mu} = \partial_B \phi^{\text{in}}(A_0, A_1, B_i - F_i^\mu g_{\mu\to i}) \tag{125}$$
$$\approx \hat{x}_i - F_i^\mu g_\mu \partial_B^2 \phi_i^{\text{in}}, \tag{126}$$

which can be then inserted in the definition $\omega_\mu = \sum_i F_i^\mu \hat{x}_{i\to\mu}$ resulting in

$$\omega_\mu = \sum_i F_i^\mu \hat{x}_i - g_\mu \sum_i (F_i^\mu)^2 \partial_B^2 \phi_i^{\text{in}}. \tag{127}$$

The other relevant equation is

$$g_{\mu\to i} = \partial_\omega \phi^{\text{out}}(V_0, V_1, \omega^\mu - F_i^\mu \hat{x}^{i\to\mu}) \tag{128}$$
$$\approx g_\mu - F_i^\mu \hat{x}^\mu \partial_\omega^2 \phi_\mu^{\text{out}}, \tag{129}$$

which analogously leads to

$$B_i = \sum_\mu F_i^\mu g_\mu - \hat{x}_i \sum_\mu (F_i^\mu)^2 \partial_\omega^2 \phi_\mu^{\text{out}}. \tag{130}$$

We now introduce back the time indexes, and use the shorthand notations $\phi_i^{\text{in},t} = \phi^{\text{in}}(B_i^t, A_0^t, A_1^t, m)$ and $\phi_\mu^{\text{out},t} = \phi^{\text{out}}(\omega_\mu^t, V_0^{t-1}, V_1^{t-1}, y_\mu, m)$. Using again the definition $c_{\boldsymbol{F}} = \frac{1}{MN}\sum_{\mu,i}(F_i^\mu)^2$ (hence $\mathbb{E}\, c_{\boldsymbol{F}} = 1/N$ in our setting), with some initialization for $\hat{x}_i^{t=0}$,$V_0^{t=0}$, $V_1^{t=0}$ and setting $g_\mu^{t=0} = 0$, we finally obtain

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1}(mV_0^{t-1} + V_1^{t-1}) \tag{131}$$
$$g_\mu^t = \partial_\omega \phi_\mu^{\text{out},t} \tag{132}$$
$$\Gamma_0^t = \frac{1}{m-1}\left(\partial_\omega^2 \phi_\mu^{\text{out},t} - (2\partial_{V_1}\phi_\mu^{\text{out},t} - (g_\mu^t)^2)\right) \tag{133}$$
$$\Gamma_1^t = \frac{1}{m-1}(\partial_\omega^2 \phi_\mu^{\text{out},t} - m(2\partial_{V_1}\phi_\mu^{\text{out},t} - (g_\mu^t)^2)) \tag{134}$$
$$A_0^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_0^t \tag{135}$$
$$A_1^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_1^t \tag{136}$$
$$B_i^t = \sum_\mu F_i^\mu g_\mu^t - \hat{x}_i^{t-1}(mA_0^t - A_1^t) \tag{137}$$
$$\hat{x}_i^t = \partial_B \phi_i^{\text{in},t} \tag{138}$$
$$\Delta_{0,i}^t = \frac{1}{m-1}\left(\partial_B^2 \phi_i^{\text{in},t} + 2\partial_{A_1}\phi_i^{\text{in},t} + (\hat{x}_i^t)^2\right) \tag{139}$$
$$\Delta_{1,i}^t = \partial_B^2 \phi_i^{\text{in},t} - m\Delta_{0,i}^t \tag{140}$$
$$V_0^t = c_{\boldsymbol{F}} \sum_i \Delta_{0,i}^t \tag{141}$$
$$V_1^t = c_{\boldsymbol{F}} \sum_i \Delta_{1,i}^t. \tag{142}$$

Equations (131-142), along with the free entropy definitions in Eqs. (121, 123) are the GASP iterative equations.

B.3. Zero Temperature Limit

In order to apply the GASP algorithm to MAP estimation, we have to consider the zero-temperature limit $\beta \uparrow \infty$ of the message passing. The limiting form of the equations depends on the model and on the regime (e.g. low or high α). Here we

consider models defined on continuous spaces χ^N and in the high α regime (e.g. $\alpha > 1$ for phase retrieval). In this case, while taking the limit, the messages have to be rescaled appropriately in order to keep them finite. Therefore, we rescale the messages through the substitutions

$$A_0 \to \beta^2 A_0 \tag{143}$$
$$A_1 \to \beta A_1 \tag{144}$$
$$B \to \beta B \tag{145}$$
$$\omega \to \omega \tag{146}$$
$$V_0 \to V_0 \tag{147}$$
$$V_1 \to V_1/\beta \tag{148}$$
$$g \to \beta g \tag{149}$$
$$m \to m/\beta \tag{150}$$
$$\Delta_0 \to \Delta_0 \tag{151}$$
$$\Delta_1 \to \Delta_1/\beta, \tag{152}$$

in Equations (131-142) and Eqs. (121, 124). Taking the $\beta \to \infty$ limit we recover the GASP equations for MAP estimation presented in the Main Text.

B.4. GASP equations for real-valued phase retrieval problem

Putting together Eqs.(80) and (81), and the definitions in Eqs.(121) and (123), we can obtain the zero temperature limit of the two GASP scalar estimation channels, in the special case of the phase retrieval loss $\ell(y,u) = (y - |u|)^2$ and an L_2-norm $r(x) = \lambda x^2/2$. The expressions simply become:

$$\phi^{\text{in}}(B, A_0, A_1, y, m) = -\frac{B^2}{2(A_1 + \lambda - mA_0)} - \frac{1}{2m}\log\left(1 - \frac{mA_0}{A_1 + \lambda}\right) \tag{153}$$
$$\phi^{\text{out}}(\omega, V_0, V_1, m) = \frac{1}{m}\log\left(Z_+ + Z_-\right) - \frac{1}{2m}\log\left(1 + \frac{2mV_0}{1 + 2V_1}\right), \tag{154}$$

where we defined for compactness:

$$Z_\pm = H\left(-\frac{2mV_0 y \mp \omega(1 - 2V_1)}{\sqrt{V_0(1 + 2V_1)(1 + 2V_1 + 2mV_0)}}\right)\exp\left(-\frac{m(\omega \pm y)^2}{1 + 2V_1 + 2mV_0}\right). \tag{155}$$

Moreover, the zero temperature limit of GASP Eqs. (138, 139, 140) after the rescaling discussed in previous paragraph, gives:

$$\hat{x}_i = \frac{B_i}{A_1 + \lambda - mA_0} \tag{156}$$
$$\Delta_i^0 = \frac{A_0}{(A_1 + \lambda)(A_1 + \lambda - mA_0)} \tag{157}$$
$$\Delta_i^1 = \frac{1}{A_1 + \lambda - mA_0} - m\Delta_i^0. \tag{158}$$

C. Setting the symmetry-breaking parameter

The 1RSB formalism, from which the (G)ASP equations are derived, is based on the introduction of a symmetry-breaking parameter, the so-called Parisi parameter m (Mézard et al., 1987), that allows the description of the fine structure of highly non-convex ("glassy") landscapes.

In replica analyses, the physical meaning of m is the following: when the studied model develops a 1RSB structure, by tuning m in its natural range of variability $(0, 1]$, it is possible to focus the Gibbs measure on the different families of exponentially numerous "states" (i.e., basins of solutions of the inference problem) that populate the loss landscape (Mézard

et al., 1987). The dominant states, i.e. those where a perfect sampling algorithm would land with high probability, are described at the thermodinamically optimal value $m^\star$, that extremizes the free-energy of the model.

In the real-replica formalism employed to derive the ASP equations (Monasson, 1995; Antenucci et al., 2019b), however, m is an external parameter that can be analytically continued to take any real value, and is no-longer strictly bound to the interval $(0, 1]$. In fact, both the algorithm and its SE characterization are valid even if the model has not developed a proper 1RSB structure, and m can be simply thought as a parametrization the family of algorithms ASP(m) (Antenucci et al., 2019b). We note that, in the zero-temperature limit, when the proper scaling of m with $\beta \to \infty$ is chosen (Eqs. 143 to 152), even the physically meaningful interval of variability of m is of course extended to $(0, \infty)$.

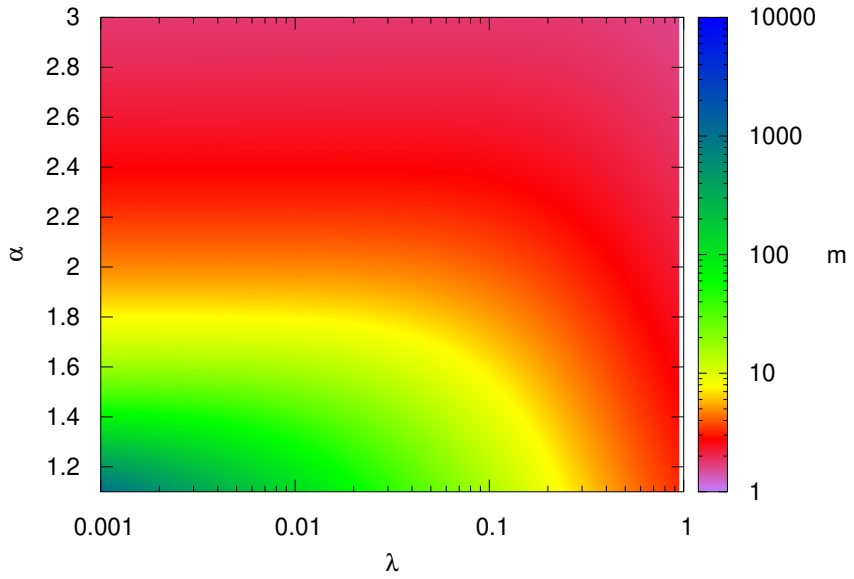

Figure 3. Optimal value of the symmetry-breaking parameter $m = m^\star$ (as employed in the GASP phase diagram in Fig. 2 in the Main Text, bottom plot), for different values of the regularizer λ.

In Fig. 3, we show the numerical values of the thermodynamic optima $m = m^\star$ in the zero-temperature phase retrieval problem (obtained analytically in correspondence of $\rho = 0$, at varying values of α and λ, from a replica computation that will be presented in a more technical future work). These are the values that were employed in the corresponding GASP phase diagram, presented in the Main Text in Fig. 2.

We remark, however, that this particular choice was mostly due to the need of consistency in the criterion for fixing m throughout the various regions of the phase diagram. In fact, as it was already noted in the Bayesian case (Antenucci et al., 2019b), the thermodynamical optimum might not be the best choice for m, since other values seem to allow better inference (e.g., a decreased final MSE). Since we are here interested in the MAP estimation task, our performance evaluation is based solely on the possibility of achieving retrieval of the signal. This condition is definitely less demanding than that of obtaining the best MSE, and in fact we find that wide ranges of values for m are effective in correspondence of each α and λ. Fig. 3 is nevertheless indicative of how m should be incremented when the observation matrix gets smaller or when weaker regularizers are employed.

In order to show the robustness of GASP(m) with respect to the choice of different values for the symmetry-breaking parameter, in Fig. 4 we plot the total number of iterations required to converge to the signal (indicated by the color map), for fixed values of m. The plotted number of iterations include both stages in our simple *continuation* strategy. As it can be seen in the plot, this total number tends to increase as α is lowered, since the inference problem becomes harder.

The colored curves mark the lower border of the regions of effectiveness of GASP(m), with m fixed in each region, at which the number of iterations required by the algorithm diverge. It is clear, indeed, that a careful fine-tuning of m is unnecessary,

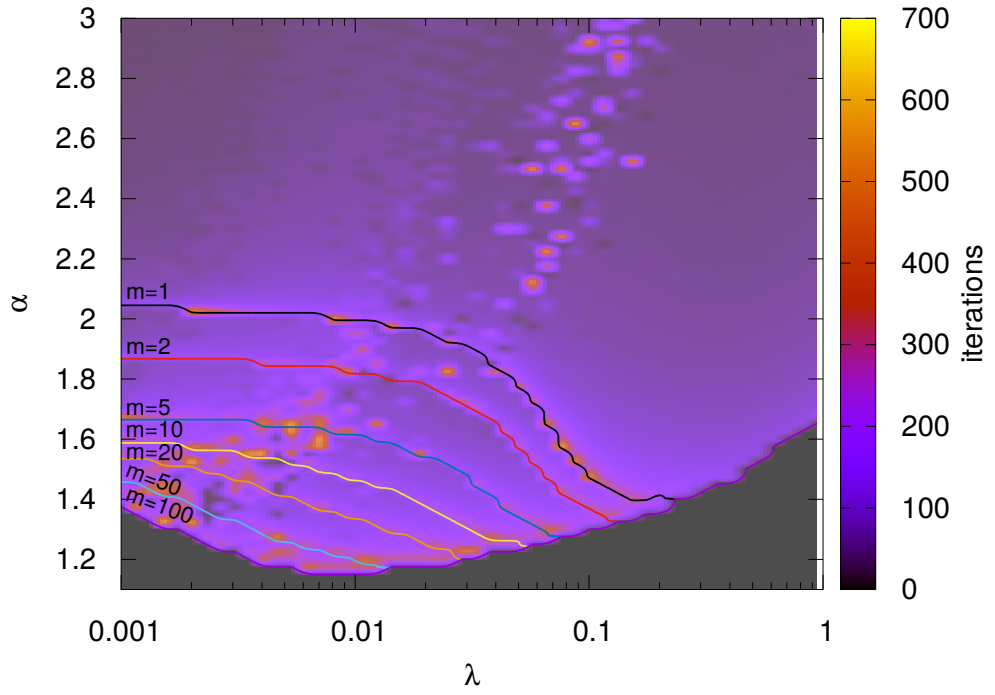

Figure 4. Total number of iterations to convergence for GASP(m). The colored curves delimit (from below) the perfect recovery regions of GASP with the indicated value for m.

and that it is quite intuitive how to adapt it when a different instance of the problem is given. For example, in the noiseless case, a basic strategy is to fix λ in the range $[0.001 : 0.01]$ and then test $\mathcal{O}(1)$ different values for m, until $MSE = 0$ is obtained at convergence of the message-passing.

As a last data point, we report in Fig. 5 the behaviour of the overlap with the true signal of the estimator given by GASP, for two different system sizes, large times and as a function of α. We observe that for large N transitions become sharper and experimental points approach the asymptotic prediction from SE.

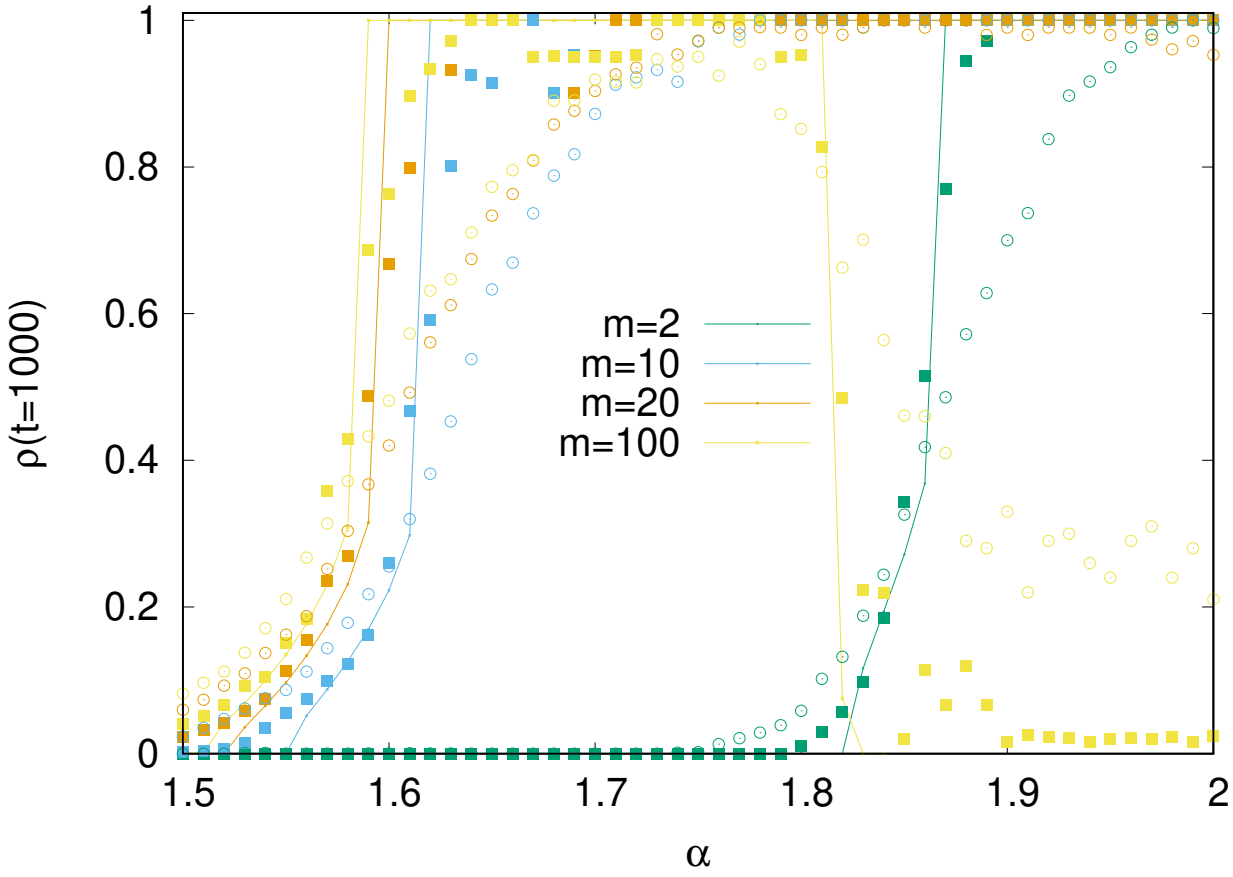

Figure 5. GASP and SE result after $t = 10^3$ iterations. Start at $t = 0$ with $\rho = 10^{-3}$ for SE and $\hat{\boldsymbol{x}} \sim \mathcal{N}(0, I_N)$ for GASP. Circles are for $N = 10^3$, squares for $N = 10^4$, . results are averaged over 100 samples. Lines are predictions from SE.