Generalized Approximate Survey Propagation for High-Dimensional Estimation: Supplementary Material

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

1. A recap on Generalized Approximate Message Passing

1.1. 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:

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

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:

$$\nu^t_{i\to\mu}(x_i) \propto e^{-\beta r(x_i) + \sum_{\nu\neq\mu}\log\hat{\nu}^t_{\nu\to i}(x_i)} \tag{2}$$

$$\hat{\nu}^{t+1}_{\mu\to i}(x_i) \propto \int_{\chi^{N-1}} \prod_{j\neq i} d\nu^t_{j\to\mu}(x_j)\; e^{-\beta\ell(y_\mu, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle)}. \tag{3}$$

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.(2) 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.(3) 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:

$$\varphi^{\text{in}}(B, A) = \log \int_\chi dx\; e^{-\frac{1}{2}Ax^2 + Bx - \beta\, r(x)} \tag{4}$$

$$\varphi^{\text{out}}(\omega, V, y) = \log \int \frac{dz}{\sqrt{2\pi V}}\; e^{-\frac{1}{2V}(z-\omega)^2 - \beta\, \ell(y,z)}. \tag{5}$$

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:

$$\log \hat{\nu}^{t+1}_{\mu\to i}(x_i) = \varphi^{\text{out}}\left(\sum_j F^\mu_j \hat{x}^t_{j\to\mu} + F^\mu_i(x_i - \hat{x}^t_{i\to\mu}),\; \sum_{j\neq i}(F^\mu_j)^2 \Delta^t_{j\to\mu},\; y_\mu\right) + \text{const.} \tag{6}$$

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:

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t - \hat{x}_i^{t-1} \sum_\mu (F_i^\mu)^2 \Gamma_\mu^t \tag{7}$$

$$A_i^t = \sum_\mu (F_i^\mu)^2 \Gamma_\mu^t \tag{8}$$

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^t - g_\mu^t \sum_i (F_i^\mu)^2 \Delta_i^t \tag{9}$$

$$V_\mu^t = \sum_i (F_i^\mu)^2 \Delta_i. \tag{10}$$

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

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

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

$$\hat{x}_i^t = \partial_B \varphi_i^{\text{in},t} \tag{13}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{\text{in},t}, \tag{14}$$

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:

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t, \tag{15}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^{t-1}, \tag{16}$$

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

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1} V^{t-1} \tag{17}$$

$$g_\mu^t = \partial_\omega \varphi_\mu^{\text{out},t} \tag{18}$$

$$\Gamma_\mu^t = -\partial_\omega^2 \varphi_\mu^{\text{out},t} \tag{19}$$

$$A^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_\mu^t \tag{20}$$

$$B_i^t = \sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1} A^t \tag{21}$$

$$\hat{x}_i^t = \partial_B \varphi_i^{\text{in},t} \tag{22}$$

$$\Delta_i^t = \partial_B^2 \varphi_i^{\text{in},t} \tag{23}$$

$$V^t = c_{\boldsymbol{F}} \sum_i \Delta_i^t. \tag{24}$$

Eqs. (17-24) 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 $g_\mu^{t=0} = 0,\ \forall \mu$.

1.2. 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:

$$A \to \beta A \tag{25}$$
$$B_i \to \beta B_i \tag{26}$$
$$V \to V/\beta \tag{27}$$
$$g_\mu \to \beta g_\mu \tag{28}$$
$$\Delta_i \to \Delta_i/\beta. \tag{29}$$

With these rescalings, the GAMP equations (17-24) are left unaltered, but the expressions for the free entropies of the scalar channels become

$$\varphi^{\text{in}}(B, A) = \max_{x \in \chi} \; -r(x) - \frac{1}{2} A x^2 + B x \tag{30}$$
$$\varphi^{\text{out}}(\omega, V, y) = \max_{z} \; -\frac{(z-\omega_\mu)^2}{2V} - \ell(y, z), \tag{31}$$

as it is easy to verify.

1.3. 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.(30) and (31) become:

$$\varphi^{\text{in}}(B, A) = \frac{B^2}{2(A+\lambda)} \tag{32}$$
$$\varphi^{\text{out}}(\omega, V, y) = -\frac{(y - |\omega|)^2}{2V+1}. \tag{33}$$

Thus, Eqs. (18, 22, 23, 24) simply yield:

$$g_\mu^t = \frac{2(y_\mu - |\omega_\mu^t|)}{2V^t + 1} \operatorname{sign}(\omega_\mu^t) \tag{34}$$
$$\hat{x}_i^t = \frac{B_i^t}{A^t + \lambda} \tag{35}$$
$$\Delta_i^t = \frac{1}{A^t + \lambda} \tag{36}$$
$$V^t = N c_{\boldsymbol{F}} \frac{1}{A^t + \lambda}. \tag{37}$$

Eq. (19) is instead singular, since it involves the derivative of the sign function. Since we have

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - \frac{g_\mu^{t-1}}{A^{t-1} + \lambda} \tag{38}$$
$$g_\mu^t = \frac{2(y_\mu - |\omega_\mu^t|)}{2V^t + 1} \operatorname{sign}(\omega_\mu^t) \tag{39}$$
$$A^t = -c_{\boldsymbol{F}} \sum_\mu \partial_\omega^2 \varphi_\mu^{\text{out},t} \tag{40}$$
$$x_i^t = (A^t + \lambda) \left(\sum_\mu F_i^\mu g_\mu^t + \hat{x}_i^{t-1} \, A^t \right), \tag{41}$$

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.

2. Derivation of Generalized Approximate Survey Propagation

We will derive the GASP equation for a general GLE model specified by (1). 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.

2.1. 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:

$$\mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}) = \sum_{\mu} \ell\big(y_\mu,\, \langle \boldsymbol{F}^\mu, \boldsymbol{x}\rangle\big) + \sum_i r(x_i), \tag{42}$$

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,

$$p(\{\boldsymbol{x}^a\}_{a=1}^m) = \frac{1}{Z_{\boldsymbol{y},\boldsymbol{F}}^m} e^{-\beta \sum_{a=1}^m \mathcal{H}_{\boldsymbol{y},\boldsymbol{F}}(\boldsymbol{x}^a)}, \tag{43}$$

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:

$$\nu_{i\to\mu}(\bar{\boldsymbol{x}}_i) \propto e^{-\beta \sum_{a=1}^m r(x_i^a) + \sum_{\nu\neq\mu} \log \hat{\nu}_{\nu\to i}(\bar{\boldsymbol{x}}_i)}, \tag{44}$$

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:

$$\nu_{i\to\mu}(\bar{\boldsymbol{x}}_i) \propto \int \mathrm{d}\,h\, e^{-\frac{1}{2\Delta_{0,i\to\mu}}(h-\hat{x}_{i\to\mu})^2} \prod_a e^{-\frac{1}{2\Delta_{1,i\to\mu}}(x_i^a - h)^2}, \tag{45}$$

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

$$\langle x_i^a \rangle_{i\to\mu} = \hat{x}_{i\to\mu} \tag{46}$$
$$\langle x_i^a x_i^b \rangle_{i\to\mu} = \Delta_{0,i\to\mu} + \hat{x}_{i\to\mu}^2 \tag{47}$$
$$\langle (x_i^a)^2 \rangle_{i\to\mu} = \Delta_{1,i\to\mu} + \Delta_{0,i\to\mu} + \hat{x}_{i\to\mu}^2. \tag{48}$$

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 44. From now on the derivation is very close to that of Section 1.1 for GAMP, therefor we relax the notation and drop some indexes. Let us define the input channel free entropy:

$$\phi^{\text{in}}(B, A_0, A_1, m) = \frac{1}{m} \log \int Dz \, \left(\int_{\chi} dx \; e^{-\beta r(x) - \frac{1}{2} A_1 x^2 + (B + \sqrt{A_0} z) x} \right)^m . \tag{49}$$

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

$$\langle \psi(\bar{\boldsymbol{x}}) \rangle = \frac{\int Dz \int_{\chi^m} \prod_{a=1}^m dx^a \; e^{-\beta r(x^a) - \frac{1}{2} A_1 (x^a)^2 + (B + \sqrt{A_0} z) x^a} \;\; \psi(\bar{\boldsymbol{x}})}{\int Dz \prod_{a=1}^m dx^a \; e^{-\beta r(x^a) - \frac{1}{2} A_1 (x^a)^2 + (B + \sqrt{A_0} z) x^a}}. \tag{50}$$

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. (44) we obtain:

$$\partial_B \phi^{\text{in}} = \langle x^a \rangle \tag{51}$$
$$\partial_B^2 \phi^{\text{in}} = (\langle (x^a)^2 \rangle - \langle x^a x^b \rangle) + m \left(\langle x^a x^b \rangle - \langle x^a \rangle^2 \right) \tag{52}$$
$$2\partial_{A_0} \phi^{\text{in}} = (\langle (x^a)^2 \rangle - \langle x^a x^b \rangle) + m \langle x^a x^b \rangle \tag{53}$$
$$2\partial_{A_1} \phi^{\text{in}} = -\langle (x^a)^2 \rangle. \tag{54}$$

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:

$$\hat{x}_{i\to\mu} = \partial_B \phi^{\text{in}}_{i\to\mu}, \tag{55}$$
$$\Delta_{0,i\to\mu} = \frac{1}{m-1} \left(\partial_B^2 \phi^{\text{in}}_{i\to\mu} + 2\partial_{A_1} \phi^{\text{in}}_{i\to\mu} + \hat{x}_{i\to\mu}^2 \right), \tag{56}$$
$$\Delta_{1,i\to\mu} = \partial_B^2 \phi^{\text{in}}_{i\to\mu} - m \Delta_{0,i\to\mu}. \tag{57}$$

Defining the messages

$$\omega_{\mu\to i} = \sum_{j\neq i} F_j^\mu \hat{x}_{j\to\mu}, \tag{58}$$
$$V_{0,\mu\to i} = \sum_{j\neq i} \left(F_j^\mu \right)^2 \Delta_{0,j\to\mu}, \tag{59}$$
$$V_{1,\mu\to i} = \sum_{j\neq i} \left(F_j^\mu \right)^2 \Delta_{1,j\to\mu}, \tag{60}$$

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

$$\hat{\nu}_{\mu\to i}(\bar{\boldsymbol{x}}_i) \propto \int_{\chi^{m(N-1)}} \prod_{j\neq i} \mathrm{d}\, \nu_{j\to\mu}(\bar{\boldsymbol{x}}_j) \; e^{-\beta \sum_a \ell(y_\mu, \langle \boldsymbol{F}^\mu, \boldsymbol{x}^a \rangle)} \tag{61}$$
$$\propto \int \mathrm{d}\, z_0 \; e^{-\frac{1}{2V_{0,\mu\to i}} (z_0 - \omega_\mu)^2} \prod_{a=1}^m \left(\int Dz_1 \; e^{-\beta \; \ell(y_\mu, F_i^\mu x_i^a + z_0 + \sqrt{V_1} z_1)} \right). \tag{62}$$

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

$$\phi^{\text{out}}(\omega, V_0, V_1, y, m) = \frac{1}{m} \log \int \frac{\mathrm{d}\, z_0}{\sqrt{2\pi V_0}}\, e^{-\frac{1}{2V_0}(z_0-\omega)^2} \left(\int Dz_1\, e^{-\beta\, \ell(y, z_0+\sqrt{V_1} z_1)} \right)^m . \tag{63}$$

Introducing the second order expansion

$$\log \hat{\nu}_{\mu\to i}(\bar{\boldsymbol{x}}_i) = g_{\mu\to i} \sum_a x_i^a - \frac{1}{2} A_{1,\mu\to i} \sum_a (x_i^a)^2 + \frac{1}{2} A_{0,\mu\to i} \sum_{a,b} x_i^a x_i^b \tag{64}$$

we can write the last set of rSP messages as

$$g_{\mu\to i} = \partial_\omega \phi^{\text{out}}_{\mu\to i} \tag{65}$$

$$\Gamma_{0,\mu\to i} = \frac{1}{m-1} \left(\partial^2_\omega \phi^{\text{out}}_{\mu\to i} - (2\partial_{V_1} \phi^{\text{out}}_{\mu\to i} - g^2_{\mu\to i}) \right) \tag{66}$$

$$\Gamma_{1,\mu\to i} = \frac{1}{m-1} (\partial^2_\omega \phi^{\text{out}}_{\mu\to i} - m(2\partial_{V_1} \phi^{\text{out}}_{\mu\to i} - g^2_{\mu\to i})) \tag{67}$$

$$\tag{68}$$

Incoming messages on the input nodes are then given by

$$B_{i\to\mu} = \sum_{\nu\neq\mu} F_i^\nu g_{\nu\to i} \tag{69}$$

$$A_{0,i\to\mu} = \sum_{\nu\neq\mu} (F_i^\nu)^2 \Gamma_{0,\nu\to i} \tag{70}$$

$$A_{1,i\to\mu} = \sum_{\nu\neq\mu} (F_i^\nu)^2 \Gamma_{1,\nu\to i} \tag{71}$$

$$\tag{72}$$

The closed set of Equations (55-60) and (65-71), along with the free entropy definitions in Eqs. (49) and (63), define the rSP iterative message passing.

2.2. 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

$$\phi^{\text{in}}(B, A_0, A_1, m) = \frac{1}{m} \log \int Dz\, e^{m\varphi^{\text{in}}(B+\sqrt{A_0} z,\, A_1)} \tag{73}$$

$$\varphi^{\text{in}}(h, A_1) = \log \int_\chi \mathrm{d}\, x\, e^{-\beta r(x) - \frac{1}{2} A_1 x^2 + hx} \tag{74}$$

and for the output channel

$$\phi^{\text{out}}(\omega, V_0, V_1, y, m) = \frac{1}{m} \log \int Dz\, e^{m\varphi^{\text{out}}(\omega+\sqrt{V_0}\, z,\, V_1, y)}, \tag{75}$$

$$\varphi^{\text{out}}(u, V_1, y) = \log \int Dz\, e^{-\beta\, \ell(y, u+\sqrt{V_1} z)}. \tag{76}$$

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

$$\hat{x}_{i\to\mu} = \partial_B \phi^{\text{in}}(A_0, A_1, B_i - F_i^\mu g_{\mu\to i}) \tag{77}$$
$$\approx \hat{x}_i - F_i^\mu g_\mu \partial_B^2 \phi_i^{\text{in}}, \tag{78}$$

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

$$\omega_\mu = \sum_i F_i^\mu \hat{x}_i - g_\mu \sum_i \left(F_i^\mu\right)^2 \partial_B^2 \phi_i^{\text{in}}. \tag{79}$$

The other relevant equation is

$$g_{\mu\to i} = \partial_\omega \phi^{\text{out}}(V_0, V_1, \omega^\mu - F_i^\mu \hat{x}^{i\to\mu}) \tag{80}$$
$$\approx g_\mu - F_i^\mu \hat{x}^\mu \partial_\omega^2 \phi_\mu^{\text{out}}, \tag{81}$$

which analogously leads to

$$B_i = \sum_\mu F_i^\mu g_\mu - \hat{x}_i \sum_\mu \left(F_i^\mu\right)^2 \partial_\omega^2 \phi_\mu^{\text{out}}. \tag{82}$$

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

$$\omega_\mu^t = \sum_i F_i^\mu \hat{x}_i^{t-1} - g_\mu^{t-1}(m V_0^{t-1} + V_1^{t-1}) \tag{83}$$
$$g_\mu^t = \partial_\omega \phi_\mu^{\text{out},t} \tag{84}$$
$$\Gamma_0^t = \frac{1}{m-1}\left(\partial_\omega^2 \phi_\mu^{\text{out},t} - (2\partial_{V_1}\phi_\mu^{\text{out},t} - (g_\mu^t)^2)\right) \tag{85}$$
$$\Gamma_1^t = \frac{1}{m-1}(\partial_\omega^2 \phi_\mu^{\text{out},t} - m(2\partial_{V_1}\phi_\mu^{\text{out},t} - (g_\mu^t)^2)) \tag{86}$$
$$A_0^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_0^t \tag{87}$$
$$A_1^t = c_{\boldsymbol{F}} \sum_\mu \Gamma_1^t \tag{88}$$
$$B_i^t = \sum_\mu F_i^\mu g_\mu^t - \hat{x}_i^{t-1}(m A_0^t - A_1^t) \tag{89}$$
$$\hat{x}_i^t = \partial_B \phi_i^{\text{in},t} \tag{90}$$
$$\Delta_{0,i}^t = \frac{1}{m-1}\left(\partial_B^2 \phi_i^{\text{in},t} + 2\partial_{A_1}\phi_i^{\text{in},t} + (\hat{x}_i^t)^2\right) \tag{91}$$
$$\Delta_{1,i}^t = \partial_B^2 \phi_i^{\text{in},t} - m\Delta_{0,i}^t \tag{92}$$
$$V_0^t = c_{\boldsymbol{F}} \sum_i \Delta_{0,i}^t \tag{93}$$
$$V_1^t = c_{\boldsymbol{F}} \sum_i \Delta_{1,i}^t. \tag{94}$$

Equations (83-94), along with the free entropy definitions in Eqs. (73, 75) are the GASP iterative equations.

2.3. 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

$$A_0 \to \beta^2 A_0 \tag{95}$$
$$A_1 \to \beta A_1 \tag{96}$$
$$B \to \beta B \tag{97}$$
$$\omega \to \omega \tag{98}$$
$$V_0 \to V_0 \tag{99}$$
$$V_1 \to V_1/\beta \tag{100}$$
$$g \to \beta g \tag{101}$$
$$m \to m/\beta \tag{102}$$
$$\Delta_0 \to \Delta_0 \tag{103}$$
$$\Delta_1 \to \Delta_1/\beta, \tag{104}$$

in Equations (83-94) and Eqs. (73, 76). Taking the $\beta \to \infty$ limit we recover the GASP equations for MAP estimation presented in the Main Text.

2.4. GASP equations for real-valued phase retrieval problem

Putting together Eqs.(32) and (33), and the definitions in Eqs.(73) and (75), 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:

$$\phi^{\text{in}}(B, A_0, A_1, y, m) = -\frac{B^2}{2(A_1 + \lambda - mA_0)} - \frac{1}{2m}\log\left(1 - \frac{mA_0}{A_1 + \lambda}\right) \tag{105}$$
$$\phi^{\text{out}}(\omega, V_0, V_1, m) = \frac{1}{m}\log\left(Z_+ + Z_-\right) - \frac{1}{2m}\log\left(1 + \frac{2mV_0}{1 + 2V_1}\right), \tag{106}$$

where we defined for compactness:

$$Z_\pm = H\left(-\frac{2mV_0 y \mp \omega(1 - 2V_1)}{\sqrt{V_0(1 + 2V_1)(1 + 2V_1 + 2mV_0)}}\right)\exp\left(-\frac{m(\omega \pm y)^2}{1 + 2V_1 + 2mV_0}\right). \tag{107}$$

Moreover, the zero temperature limit of GASP Eqs. (90, 91, 92) after the rescaling discussed in previous paragraph, gives:

$$\hat{x}_i = \frac{B_i}{A_1 + \lambda - mA_0} \tag{108}$$
$$\Delta_i^0 = \frac{A_0}{(A_1 + \lambda)(A_1 + \lambda - mA_0)} \tag{109}$$
$$\Delta_i^1 = \frac{1}{A_1 + \lambda - mA_0} - m\Delta_i^0. \tag{110}$$

3. 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. 95 to 104), even the physically meaningful interval of variability of m is of course extended to $(0, \infty)$.

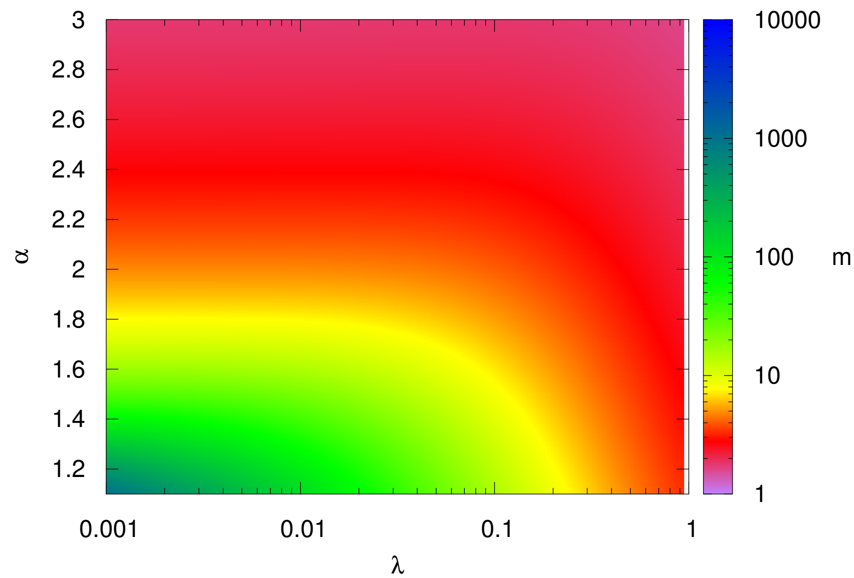

Figure 1. Optimal value of the symmetry-breaking parameter $m = m^{\star}$ (as employed in the GASP phase diagram in Fig. 2 in the Main Text, bottom plot), for different values of the regularizer λ.

In Fig. 1, 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. 1 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. 2 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, 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

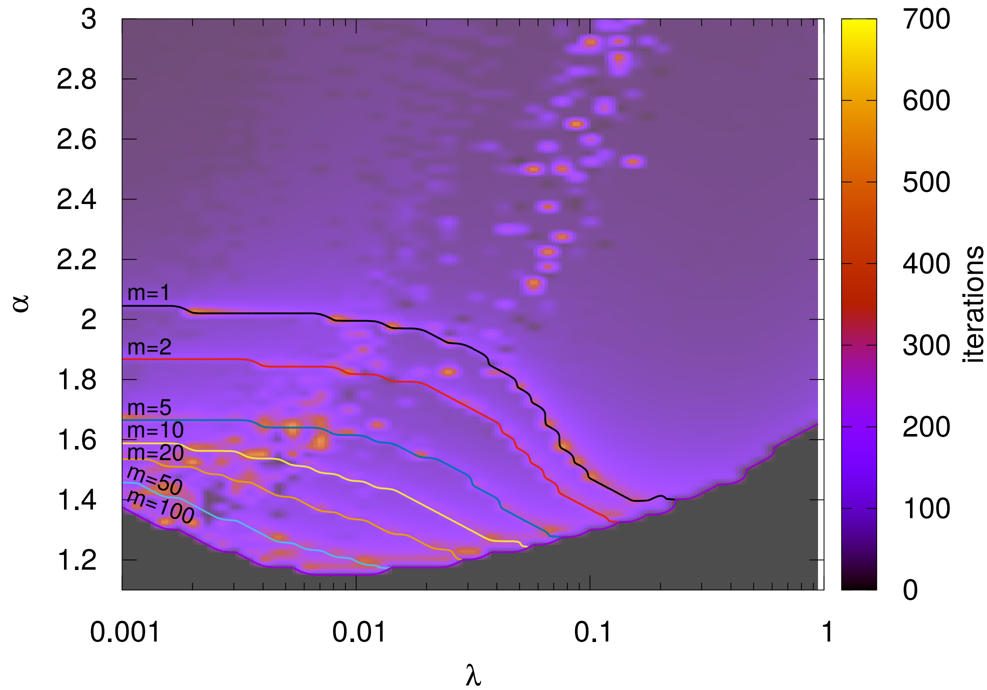

Figure 2. Total number of iterations to convergence for GASP(m). The colored curves delimit (from below) the perfect recovery regions of GASP with the indicated value for m.

obtained at convergence of the message-passing.

As a last data point, we report in Fig. 3 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.

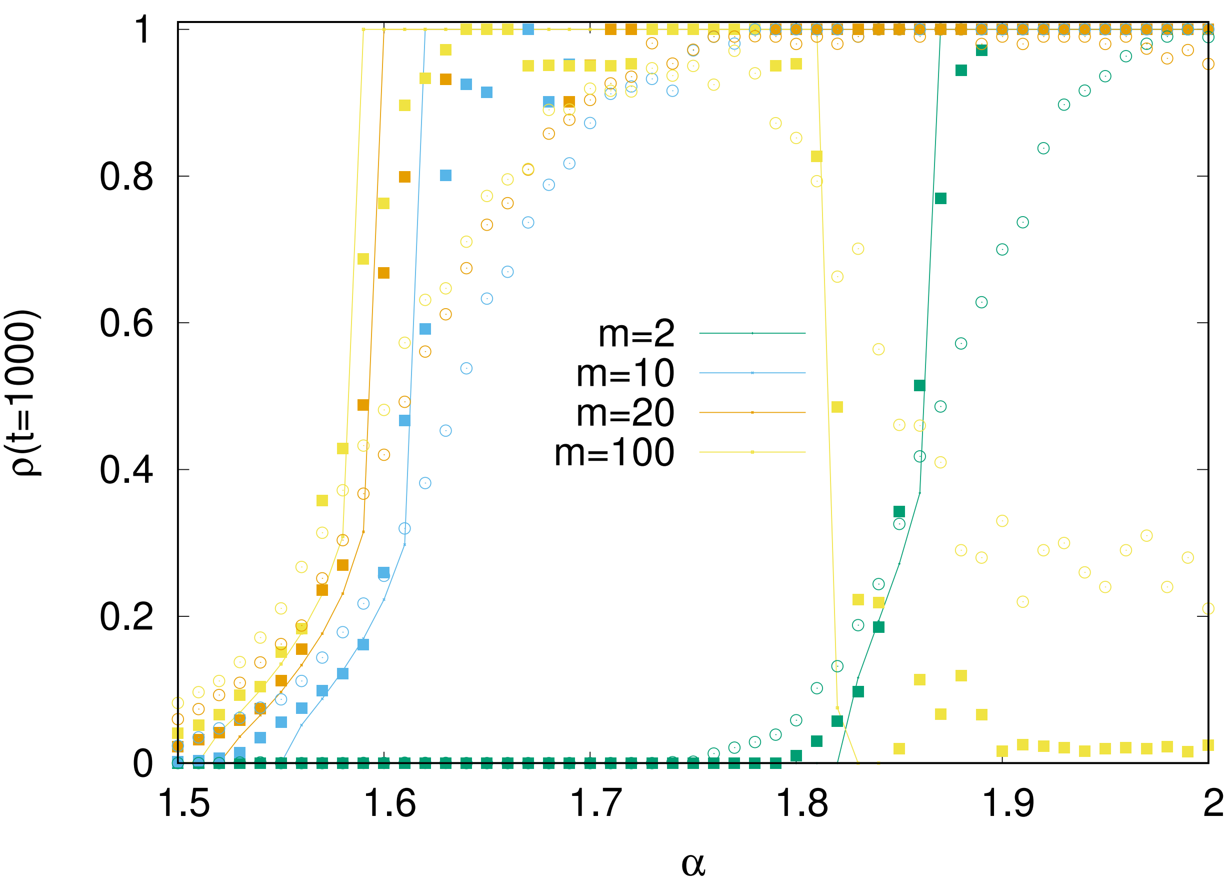

Figure 3. GASP and SE result after $t = 10^3$ iterations. Start at $t = 0$ with $\rho = 10^{-3}$ for SE and $\hat{\boldsymbol{x}} \sim \mathcal{N}(0, I_N)$ for GASP. Circles are for $N = 10^3$, squares for $N = 10^4$, . results are averaged over 100 samples. Lines are predictions from SE.

References

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.

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.

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.

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.

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.

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.

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.

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

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.