

Blind calibration for compressed sensing: State evolution and an online algorithm

Marylou Gabrié*¹, Jean Barbier², Florent Krzakala¹, Lenka Zdeborová³

¹Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France

²International Center for Theoretical Physics, Trieste, Italy

³Institut de Physique Théorique, CEA, CNRS, Université Paris-Saclay

March 24, 2020

Abstract

Compressed sensing allows the acquisition of compressible signals with a small number of measurements. In experimental settings, the sensing process corresponding to the hardware implementation is not always perfectly known and may require a calibration. To this end, blind calibration proposes to perform at the same time the calibration *and* the compressed sensing. Schülke and collaborators suggested an approach based on approximate message passing for blind calibration (cal-AMP) in [1, 2]. Here, their algorithm is extended from the already proposed offline case to the online case, for which the calibration is refined step by step as new measured samples are received. We show that the performance of both the offline and the online algorithms can be theoretically studied via the State Evolution (SE) formalism. Finally, the efficiency of cal-AMP and the consistency of the theoretical predictions are confirmed through numerical simulations.

1 Introduction

The efficient acquisition of sparse signals has been made possible by Compressed Sensing (CS) [3]. This technique has now many applications: in medical imaging [4, 5] for instance, where short acquisition times are preferable, or in imaging devices where measurements are costly [6]. For practical applications, correct calibration of potential unknowns coming from the measurement hardware is a central issue. If complementary pairs of measurement-signal are available a priori, a *supervised* procedure of calibration can be imagined. We assume it is not the case and focus on *blind* calibration. In blind calibration, calibration is conducted in parallel to the sparse signal recovery of the compressed sensing. This method does not require the availability of calibration samples before the reconstruction and is thus particularly appealing. Here we consider formally the case where the exact measurement process is known up to a set of variables, referred to as calibration variables.

Gribonval and co-authors discussed in a series of articles [7, 8] a decalibration example where unknown gains multiply each sensor. More precisely, they considered for each signal $\underline{x} \in \mathbb{R}^N$ an observation $\underline{y} \in \mathbb{R}^M$, produced as

$$\underline{y} = \underline{S} \underline{W} \underline{x} \tag{1}$$

with $\underline{W} \in \mathbb{R}^{M \times N}$ a known measurement matrix and \underline{S} an unknown square diagonal matrix of calibration variables. The examined problem is to reconstruct both the corresponding signals \underline{x} and the calibration matrix

*Corresponding author: marylou.gabrie@ens.fr

\underline{S} , provided several observations \underline{y} . In [7, 8], the authors showed that this question can be exactly expressed as a convex optimization problem, and thus be solved using off-the-shelf algorithms. In subsequent papers [1, 2], Schülke and collaborators used instead an Approximate Message Passing (AMP) approach, famously introduced in compressed sensing by [9]. Their calibration-AMP algorithm (cal-AMP) was proposed for Bayesian blind calibration, with both the calibration variables on the sensors and the elements of the signals reconstructed simultaneously. Remarkably, cal-AMP is not restricted to the case where the distortion of the measurements is a multiplication by a gain and is applicable in other settings. The derivation of the algorithm relies on the extension of AMP resolution technics to the Generalized Linear Model (GLM), for which linear mixing is followed by a generic probabilistic sensing process, corresponding to the Generalized-AMP (GAMP) algorithm [10, 11]. Authors of [1, 2] demonstrated their approach through empirical numerical simulations, and found that considering relatively few measured samples at a time could already allow for a good calibration.

An ensuing question is whether the calibration could be performed *online*, that is when different observations are received successively instead of being treated at once. In learning applications, it is sometimes advantageous for speed concerns to only treat a subset of training examples at a time. Sometimes also, the size of the current data sets may exceed the available memory. Methods implementing a step-by-step learning, as the data arrives, are referred to as *online*, *streaming* or *mini-batch* learning, as opposed to *offline* or *batch* learning. For instance, in deep learning, Stochastic Gradient Descent is the most popular training algorithm [12]. From the theoretical point of view, the restriction to the fully online case, where a single data point is used at a time, offers interesting possibilities of analysis, as demonstrated in different machine learning problems by [13, 14, 15]. Here we will consider the Bayesian online learning of the calibration variables.

In the present paper, we revisit and extend cal-AMP with the following contributions

- We re-derive the message passing equations from a more general formulation than [1, 2]. This strategy allows us to theoretically analyze the algorithm through a State Evolution, as first done for regular AMP in [9]. This analysis was not straightforward and remained an open problem for [2]. These contributions are presented in Section 2.
- In Section 2.4, we write the free energy, or equivalently the mutual information, of the problem thanks to the general formulation and the natural connection to committee machines [16]. We conjecture our results to be rigorously exact and discuss the missing ingredients to turn our formula into a full theorem.
- Along the lines of [17], we also consider the online version of the problem in Section 3. We propose an online cal-AMP algorithm allowing for the Bayesian adjustment of the calibration as new observations arrive. The corresponding State Evolution analysis is presented as well.
- We validate numerically the performance of the algorithms and their consistency with the theoretical analyses, on the example of gain calibration, in Section 4. In particular, the algorithm remains fast and remarkably efficient going from the offline to the online setting.

Finally, we wish to clarify notations: vector variables are underlined as \underline{x} , matrix variables are doubly underlined as $\underline{\underline{W}}$ and the symbol \propto omits the normalization factor of probability distributions.

2 Cal-AMP revisited

2.1 Model of calibrated generalized linear estimation

We consider P observations $\underline{y}_{(k)} \in \mathbb{R}^M$ gathered in a matrix $\underline{Y} \in \mathbb{R}^{M \times P}$, generated by the following *teacher* model

$$\underline{s}_0 \sim p_{s_0}(\underline{s}_0) \quad (2)$$

$$\forall k = 1, \dots, P, \quad \underline{x}_{0,(k)} \sim p_{x_0}(\underline{x}_{0,(k)}) = \prod_{i=1}^N p_{x_0}(x_{0,i,(k)}) \quad (3)$$

$$\underline{y}_{(k)} \sim p_{\text{out},0}^s(\underline{y}_{(k)} | \underline{W} \underline{x}_{0,(k)}, \underline{s}_0) = \prod_{\mu=1}^M p_{\text{out},0}^s(y_{(k),\mu} | \underline{W} \underline{x}_{0,(k)}, s_{0,\mu}). \quad (4)$$

The $\underline{x}_{0,(k)} \in \mathbb{R}^N$ are P unknown signals, gathered in a matrix $\underline{X}_0 \in \mathbb{R}^{N \times P}$, linearly mixed by a known weight (or measurement) matrix \underline{W} , and pushed through a noisy channel denoted $p_{\text{out},0}^s$. The channel is not perfectly known and depends on the realization of calibration variables $\underline{s}_0 \in \mathbb{R}^M$ (which, in the special case of a linear model (1), is the diagonal of \underline{S}). We are interested in the estimation of the unknown signals and calibration variables. The priors and the channel of the teacher model are not necessarily available and we assume the following, a priori different, *student* model,

$$p(\underline{X}, \underline{s} | \underline{Y}, \underline{W}) = \frac{1}{\mathcal{Z}(\underline{Y}, \underline{W})} p_x(\underline{X}) p_s(\underline{s}) p_{\text{out}}^s(\underline{Y} | \underline{W} \underline{X}, \underline{s}) \quad (5)$$

$$= \frac{1}{\mathcal{Z}(\underline{Y}, \underline{W})} \prod_{i=1}^N p_x(x_i) \prod_{\mu=1}^M p_s(s_\mu) p_{\text{out}}^s(y_\mu | \underline{w}_\mu^\top \underline{X}, s_\mu), \quad (6)$$

with $\underline{x}_i \in \mathbb{R}^P$, and $\underline{y}_\mu \in \mathbb{R}^P$. The corresponding factor graph is drawn on Figure 1b. Note that the distribution could be further factorized over the index k of the P observations. The corresponding message passing was derived in [1, 2]. Here we adopt instead the level of factorization above and derive the AMP algorithm on vector variables in \mathbb{R}^P . Thus cal-AMP can be seen as a special case of GAMP on vector variables as we will see in the next section. This point of view is key to obtain the State Evolution analysis of the calibration problem.

2.2 Cal-AMP as a special case of GAMP

In our derivation calibration-AMP (cal-AMP) is interpreted as a special case of Generalized Approximate Message Passing (GAMP) [11], where (a) the algorithm acts simultaneously on the vector variables and (b) the channel includes a marginalization over the calibration variables. For this first variation, the AMP algorithm on vector variables was already derived by [16], to treat committee machines (two-layers neural networks [18]) without calibration. We shall recall here the main steps of the derivation, see [19] for a recent methodological introductory review to message passing algorithms and mean-field approximations in learning problems.

2.2.1 AMP for reconstruction of multiple samples

The systematic procedure to write AMP for a given joint probability distribution consists in first writing BP on the factor graph, second project the messages on a parametrized family of functions to obtain the corresponding relaxed-BP and third close the equations on a reduced set of parameters by keeping only leading terms in the thermodynamic limit.

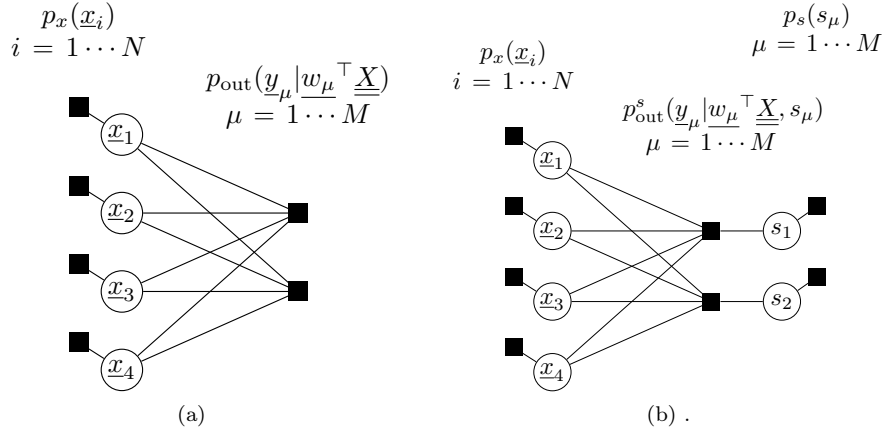


Figure 1: **(a)** Factor graph of the Generalized Linear Model (GLM) on vector variables corresponding to the joint distribution (7). **(b)** Factor graph for the GLM on vector variables with not perfectly known channel including calibration variables, corresponding to the joint distribution (6).

For the generic GLM, without the calibration variable, the posterior measure we are interested in is

$$p(\underline{X}|\underline{Y}, \underline{W}) = \frac{1}{\mathcal{Z}(\underline{Y}, \underline{W})} \prod_{i=1}^N p(\underline{x}_i) \prod_{\mu=1}^M p_{\text{out}}(\underline{y}_\mu | \underline{w}_\mu^\top \underline{X} / \sqrt{N}), \quad \underline{x}_i \in \mathbb{R}^P, \quad \underline{y}_\mu \in \mathbb{R}^P, \quad (7)$$

where the known entries of matrix \underline{W} are drawn i.i.d from a standard normal distribution (the scaling in $1/\sqrt{N}$ is from now on made explicit). The corresponding factor graph is given on Figure 1a. We are considering the simultaneous reconstruction of P signals $\underline{x}_{0,(k)} \in \mathbb{R}^N$ and therefore write the message passing on the variables $\underline{x}_i \in \mathbb{R}^P$. The major difference with the fully factorized version on scalar variables of [1, 2] is that we will consider covariance matrices between variables coming from the P observations instead of assuming complete independence.

Belief propagation (BP) We start with BP on the factor graph of Figure 1a. For all pairs of indices $i - \mu$, we define the update equations of messages

$$\tilde{m}_{\mu \rightarrow i}^{(t)}(\underline{x}_i) = \frac{1}{\mathcal{Z}_{\mu \rightarrow i}} \int \prod_{i' \neq i} d\underline{x}_{i'} p_{\text{out}} \left(\underline{y}_\mu \mid \sum_j \frac{W_{\mu j}}{\sqrt{N}} \underline{x}_j \right) \prod_{i' \neq i} m_{i' \rightarrow \mu}^{(t)}(\underline{x}_{i'}) \quad (8)$$

$$m_{i \rightarrow \mu}^{(t+1)}(\underline{x}_i) = \frac{1}{\mathcal{Z}_{i \rightarrow \mu}} p_x(\underline{x}_i) \prod_{\mu' \neq \mu} \tilde{m}_{\mu' \rightarrow i}^{(t)}(\underline{x}_i), \quad (9)$$

where $\mathcal{Z}_{\mu \rightarrow i}$ and $\mathcal{Z}_{i \rightarrow \mu}$ are normalizations that allow to interpret messages as probability distributions. To improve readability, we drop the time indices in the following derivation, and only recover them in the final algorithm.

Relaxed BP (r-BP) The second step of the derivation consists in developing messages up to order $O(1/N)$ as we take the thermodynamic limit $N \rightarrow +\infty$ (at fixed $\alpha = M/N$). At this order, we will find that it is consistent to consider the messages to be approximately Gaussian, i.e. characterized by their means and

covariances. Thus we define

$$\hat{\underline{x}}_{i \rightarrow \mu} = \int d\underline{x} \underline{x} m_{i \rightarrow \mu}(\underline{x}) \quad (10)$$

$$\underline{C}_{i \rightarrow \mu}^x = \int d\underline{x} \underline{x} \underline{x}^T m_{i \rightarrow \mu}(\underline{x}) \quad (11)$$

and

$$\underline{\omega}_{\mu \rightarrow i} = \sum_{i' \neq i} \frac{W_{\mu i'}}{\sqrt{N}} \hat{\underline{x}}_{i' \rightarrow \mu} \quad (12)$$

$$\underline{V}_{\mu \rightarrow i} = \sum_{i' \neq i} \frac{W_{\mu i'}^2}{N} \underline{C}_{i' \rightarrow \mu}^x, \quad (13)$$

where $\underline{\omega}_{\mu \rightarrow i}$ and $\underline{V}_{\mu \rightarrow i}$ are related to the intermediate variable $\underline{z}_\mu = \underline{w}_\mu^\top \underline{X}$.

Expansion of $\tilde{m}_{\mu \rightarrow i}$ - We define the Fourier transform \hat{p}_{out} of $p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu)$ with respect to its argument $\underline{z}_\mu = \underline{w}_\mu^\top \underline{X}$,

$$\hat{p}_{\text{out}}(\underline{y}_\mu | \underline{\xi}_\mu) = \int d\underline{z}_\mu p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) e^{-i \underline{\xi}_\mu^\top \underline{z}_\mu}. \quad (14)$$

Using reciprocally the Fourier representation of $p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu)$,

$$p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) = \frac{1}{(2\pi)^M} \int d\underline{\xi}_\mu \hat{p}_{\text{out}}(\underline{y}_\mu | \underline{\xi}_\mu) e^{i \underline{\xi}_\mu^\top \underline{z}_\mu}, \quad (15)$$

we decouple the integrals over the different $\underline{x}_{i'}$ in (8),

$$\tilde{m}_{\mu \rightarrow i}(\underline{x}_i) \propto \int d\underline{\xi}_\mu \hat{p}_{\text{out}}(\underline{y}_\mu | \underline{\xi}_\mu) e^{i \frac{W_{\mu i}}{\sqrt{N}} \underline{\xi}_\mu^\top \underline{x}_i} \prod_{i' \neq i} \int d\underline{x}_{i'} m_{i' \rightarrow \mu}(\underline{x}_{i'}) e^{i \frac{W_{\mu i'}}{\sqrt{N}} \underline{\xi}_\mu^\top \underline{x}_{i'}} \quad (16)$$

$$\propto \int d\underline{\xi}_\mu \hat{p}_{\text{out}}(\underline{y}_\mu | \underline{\xi}_\mu) e^{i \underline{\xi}_\mu^\top \left(\frac{W_{\mu i}}{\sqrt{N}} \underline{x}_i + \underline{\omega}_{\mu \rightarrow i} \right) - \frac{1}{2} \underline{\xi}_\mu^\top \underline{V}_{\mu \rightarrow i}^{-1} \underline{\xi}_\mu}, \quad (17)$$

where developing the exponentials of the product in (8) allows to express the integrals over the $\underline{x}_{i'}$ as a function of the definitions (12)-(13), before re-exponentiating to obtain the final result (17). Now reversing the Fourier transform and performing the integral over $\underline{\xi}$, we can further rewrite

$$\tilde{m}_{\mu \rightarrow i}(\underline{x}_i) \propto \int d\underline{z}_\mu p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) e^{-\frac{1}{2} \left(\underline{z}_\mu - \frac{W_{\mu i}}{\sqrt{N}} \underline{x}_i - \underline{\omega}_{\mu \rightarrow i} \right)^\top \underline{V}_{\mu \rightarrow i}^{-1} \left(\underline{z}_\mu - \frac{W_{\mu i}}{\sqrt{N}} \underline{x}_i - \underline{\omega}_{\mu \rightarrow i} \right)} \quad (18)$$

$$\propto \int d\underline{z}_\mu \mathbb{P}_{\text{out}}(\underline{z}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) e^{(\underline{z}_\mu - \underline{\omega}_{\mu \rightarrow i})^\top \underline{V}_{\mu \rightarrow i}^{-1} \frac{W_{\mu i}}{\sqrt{N}} \underline{x}_i - \frac{W_{\mu i}^2}{2N} \underline{x}_i^\top \underline{V}_{\mu \rightarrow i}^{-1} \underline{x}_i}, \quad (19)$$

where we are led to introduce the *output update functions*,

$$\mathbb{P}_{\text{out}}(\underline{z}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) = p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) \mathcal{N}(\underline{z}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}), \quad (20)$$

$$\mathcal{Z}_{\text{out}}(\underline{y}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) = \int d\underline{z}_\mu p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) \mathcal{N}(\underline{z}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}), \quad (21)$$

$$\underline{g}_{\text{out}}(\underline{y}_\mu; \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) = \frac{1}{\mathcal{Z}_{\text{out}}} \frac{\partial \mathcal{Z}_{\text{out}}}{\partial \underline{\omega}} \quad \text{and} \quad \partial_{\underline{\omega}} \underline{g}_{\text{out}} = \frac{\partial \underline{g}_{\text{out}}}{\partial \underline{\omega}}, \quad (22)$$

with $\mathcal{N}(\underline{z}; \underline{\omega}, \underline{V})$ the multivariate Gaussian distribution of mean $\underline{\omega}$ and covariance \underline{V} . Further expansion of the exponential in (19) to order $O(1/N)$ leads to the Gaussian parametrization

$$\tilde{m}_{\mu \rightarrow i}(\underline{x}_i) \propto 1 + \frac{W_{\mu i}}{\sqrt{N}} \underline{g}_{\text{out}} \underline{x}_i + \frac{W_{\mu i}^2}{2N} \underline{x}_i^T (\underline{g}_{\text{out}} \underline{g}_{\text{out}}^T + \partial_{\underline{\omega}} \underline{g}_{\text{out}}) \underline{x}_i \quad (23)$$

$$\propto e^{\underline{B}_{\mu \rightarrow i}^T \underline{x}_i - \frac{1}{2} \underline{x}_i^T \underline{A}_{\mu \rightarrow i} \underline{x}_i}, \quad (24)$$

with

$$\underline{B}_{\mu \rightarrow i} = \frac{W_{\mu i}}{\sqrt{N}} \underline{g}_{\text{out}}(\underline{y}_{\mu}, \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) \quad (25)$$

$$\underline{A}_{\mu \rightarrow i} = -\frac{W_{\mu i}^2}{N} \partial_{\underline{\omega}} \underline{g}_{\text{out}}(\underline{y}_{\mu}, \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}). \quad (26)$$

Consistency with $m_{i \rightarrow \mu}$ - Inserting the Gaussian approximation of $\tilde{m}_{\mu \rightarrow i}$ in the definition of $m_{i \rightarrow \mu}$, we get the parametrization

$$m_{i \rightarrow \mu}(\underline{x}_i) \propto p_x(\underline{x}_i) \prod_{\mu' \neq \mu} e^{\underline{B}_{\mu' \rightarrow i}^T \underline{x}_i - \frac{1}{2} \underline{x}_i^T \underline{A}_{\mu' \rightarrow i} \underline{x}_i} \propto p_x(\underline{x}_i) e^{-\frac{1}{2} (\underline{x}_i - \underline{\lambda}_{i \rightarrow \mu})^T \underline{\underline{\sigma}}_{i \rightarrow \mu}^{-1} (\underline{x}_i - \underline{\lambda}_{i \rightarrow \mu})} \quad (27)$$

with

$$\underline{\lambda}_{i \rightarrow \mu} = \underline{\underline{\sigma}}_{i \rightarrow \mu} \left(\sum_{\mu' \neq \mu} \underline{B}_{\mu' \rightarrow i} \right) \quad (28)$$

$$\underline{\underline{\sigma}}_{i \rightarrow \mu} = \left(\sum_{\mu' \neq \mu} \underline{A}_{\mu' \rightarrow i} \right)^{-1}. \quad (29)$$

Closing the equations - Ensuring the consistency with the definitions (10)-(11) of mean and covariance of $m_{i \rightarrow \mu}$ we finally close our set of equations by defining the *input update functions*,

$$\underline{Z}^x = \int d\underline{x} p_x(\underline{x}) e^{-\frac{1}{2} (\underline{x} - \underline{\lambda})^T \underline{\underline{\sigma}}^{-1} (\underline{x} - \underline{\lambda})} \quad (30)$$

$$\underline{f}_{\underline{1}}^x(\underline{\lambda}, \underline{\underline{\sigma}}) = \frac{1}{\underline{Z}^x} \int d\underline{x} \underline{x} p_x(\underline{x}) e^{-\frac{1}{2} (\underline{x} - \underline{\lambda})^T \underline{\underline{\sigma}}^{-1} (\underline{x} - \underline{\lambda})} \quad (31)$$

$$\underline{f}_{\underline{2}}^x(\underline{\lambda}, \underline{\underline{\sigma}}) = \frac{1}{\underline{Z}^x} \int d\underline{x} \underline{x} \underline{x}^T p_x(\underline{x}) e^{-\frac{1}{2} (\underline{x} - \underline{\lambda})^T \underline{\underline{\sigma}}^{-1} (\underline{x} - \underline{\lambda})} - \underline{f}_{\underline{1}}^x(\underline{\lambda}, \underline{\underline{\sigma}}) \underline{f}_{\underline{1}}^x(\underline{\lambda}, \underline{\underline{\sigma}})^T, \quad (32)$$

so that

$$\hat{\underline{x}}_{i \rightarrow \mu} = \underline{f}_{\underline{1}}^x(\underline{\lambda}_{i \rightarrow \mu}, \underline{\underline{\sigma}}_{i \rightarrow \mu}) \quad (33)$$

$$\underline{\underline{C}}_{i \rightarrow \mu}^x = \underline{f}_{\underline{2}}^x(\underline{\lambda}_{i \rightarrow \mu}, \underline{\underline{\sigma}}_{i \rightarrow \mu}). \quad (34)$$

The closed set of equations (12), (13), (25) (26), (28), (29), (33) and (34), with restored time indices, defines the r-BP algorithm. At convergence of the iterations, we obtain the approximated marginals

$$m_i(\underline{x}_i) = \frac{1}{\underline{Z}_i^x} p_x(\underline{x}_i) e^{-\frac{1}{2} (\underline{x}_i - \underline{\lambda}_i)^T \underline{\underline{\sigma}}_i^{-1} (\underline{x}_i - \underline{\lambda}_i)} \quad (35)$$

with

$$\underline{\lambda}_i = \underline{\sigma}_i \left(\sum_{\mu=1}^M \underline{B}_{\mu \rightarrow i} \right) \quad (36)$$

$$\underline{\sigma}_i = \left(\sum_{\mu} \underline{A}_{\mu \rightarrow i} \right)^{-1}. \quad (37)$$

While BP requires to track the iterations of $M \times N$ messages (distributions) over vectors in \mathbb{R}^P , r-BP only requires to track $O(M \times N \times P)$ variables, which is a great simplification. Nonetheless, r-BP is scarcely used as such. The computational cost can be readily further reduced with little more approximation. In the thermodynamic limit, the messages are closely related to the marginals as the contribution of the missing message between (9) and (35) is negligible to a certain extent. Careful book keeping of the order of contributions of the different terms leads to a set of closed equations on parameters of the marginals, i.e. $O(N)$ variables, corresponding to the GAMP algorithm.

Approximate message passing Given the scaling of the weights in $O(1/\sqrt{N})$, the r-BP algorithm simplifies in the thermodynamic limit. We define the pairs of marginal parameters $\underline{\omega}_\mu, \underline{V}_\mu$ and $\underline{\hat{x}}_i, \underline{C}_i^x$ similarly to the parameters $\underline{\lambda}_i$ and $\underline{\sigma}_i$ defined above. The expansion of the original message parameters $\underline{\lambda}_{i \rightarrow \mu}, \underline{\sigma}_{i \rightarrow \mu}, \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}, \underline{\hat{x}}_{i \rightarrow \mu}$ and $\underline{C}_{i \rightarrow \mu}^x$ around these new parameters leads to the final closed set of equations corresponding to the approximate message passing. In fine, we obtained the vectorized AMP for the GLM presented in Algorithm 1.

Note that, similarly to GAMP, relaxing the Gaussian assumption on the weight matrix entries to any distribution with finite second moment yields the same algorithm using the Central Limit Theorem. Furthermore, the algorithmic procedure should also generalize to a wider class of random matrices allowing for correlations between entries: the ensemble of orthogonally invariant matrices. In the singular value decomposition of such weight matrices $\underline{W} = \underline{U} \underline{S} \underline{V}^\top \in \mathbb{R}^{M \times N}$ the orthogonal basis matrices \underline{U} and \underline{V} are drawn uniformly at random from respectively $O(M)$ and $O(N)$, while the diagonal matrix of singular values \underline{S} has an arbitrary spectrum. For the GLM, without calibration variables, the signal is recovered in such cases by the (Generalized) Vector-Approximate Message Passing (G-VAMP) algorithm [20, 21], inspired from prior works in statistical physics [22, 23, 24, 25, 26] and statistical inference [27, 28].

2.2.2 Treatment of calibration variables

Heuristic derivation To deal with calibration variables, we need to consider the factor graph of Figure 1b. Here belief propagation must involve a set of messages related to the variables s_μ . This new algorithm can easily be deduced from the GAMP algorithm derived above by noticing that the posterior distribution on \underline{X} in the presence of the calibration variable \underline{s} is a special case of the GLM on vector variables with the effective channel

$$p_{\text{out}}(\underline{y}_\mu | \underline{w}_\mu^\top \underline{X}) = \int d s_\mu p_{\text{out}}^s(\underline{y}_\mu | \underline{w}_\mu^\top \underline{X}, s_\mu) p_s(s_\mu). \quad (46)$$

Thus Algorithm 1 can reconstruct \underline{X} using output functions (20)-(22) which include a marginalization over \underline{s} :

$$\mathcal{Z}_{\text{out}}(\underline{y}_\mu, \underline{\omega}_\mu, \underline{V}_\mu) = \int d \underline{z}_\mu \int d s_\mu p_{\text{out}}^s(\underline{y}_\mu | \underline{z}_\mu, s_\mu) p_s(s_\mu) \mathcal{N}(\underline{z}_\mu; \underline{\omega}_\mu, \underline{V}_\mu). \quad (47)$$

The estimator of a variable computed by (sum-product) AMP is always the mean of an approximate posterior marginal distribution. For the calibration variable, the AMP posterior is already displayed in the above \mathcal{Z}_{out} ,

$$m_\mu^s(s_\mu) = \frac{1}{\mathcal{Z}_{\text{out}}} \int d \underline{z}_\mu p_{\text{out}}^s(\underline{y}_\mu | \underline{z}_\mu, s_\mu) p_s(s_\mu) \mathcal{N}(\underline{z}_\mu; \underline{\omega}_\mu, \underline{V}_\mu). \quad (48)$$

Algorithm 1 Generalized Approximate Message Passing (GAMP) for vectors

Input: matrix $\underline{Y} \in \mathbb{R}^{M \times P}$ and matrix $\underline{W} \in \mathbb{R}^{M \times N}$:

Initialize: $\hat{\underline{x}}_i, \underline{C}_{\underline{i}}^x, \underline{\sigma}_i \quad \forall i$ and $\underline{g}_{\text{out}\mu}, \partial_{\omega} \underline{g}_{\text{out}\mu} \quad \forall \mu$

repeat

1) Estimate mean and variance of z_μ given current $\hat{\underline{x}}_i$

$$\underline{V}_\mu^{(t)} = \sum_{i=1}^N \frac{W_{\mu i}^2}{N} \underline{C}_{\underline{i}}^x \quad (38)$$

$$\underline{\omega}_\mu^{(t)} = \sum_{i=1}^N \frac{W_{\mu i}}{\sqrt{N}} \hat{\underline{x}}_i^{(t)} - \sum_{i=1}^N \frac{W_{\mu i}^2}{N} (\underline{\sigma}_i^{(t)})^{-1} \underline{C}_{\underline{i}}^x \underline{\sigma}_i \underline{g}_{\text{out}\mu}^{(t-1)} \quad (39)$$

2) Estimate mean and variance of the gap between optimal z_μ and ω_μ given \underline{y}_μ

$$\partial_{\omega} \underline{g}_{\text{out}\mu}^{(t)} = \partial_{\omega} \underline{g}_{\text{out}\mu}(\underline{y}_\mu, \underline{\omega}_\mu^{(t)}, \underline{V}_\mu^{(t)}) \quad (40)$$

$$\underline{g}_{\text{out}\mu}^{(t)} = \underline{g}_{\text{out}\mu}(\underline{y}_\mu, \underline{\omega}_\mu^{(t)}, \underline{V}_\mu^{(t)}) \quad (41)$$

3) Estimate mean and variance of x_i given current optimal z_μ

$$\underline{\sigma}_i^{(t)} = \left(- \sum_{\mu=1}^M \frac{W_{\mu i}^2}{N} \partial_{\omega} \underline{g}_{\text{out}\mu}^{(t)} \right)^{-1} \quad (42)$$

$$\underline{\lambda}_i^{(t)} = \hat{\underline{x}}_i^{(t)} + \underline{\sigma}_i^{(t)} \left(\sum_{\mu=1}^M \frac{W_{\mu i}}{\sqrt{N}} \underline{g}_{\text{out}\mu}^{(t)} \right) \quad (43)$$

4) Estimate of mean and variance of \underline{x}_i augmented of the information about the prior

$$\underline{C}_{\underline{i}}^{x(t+1)} = \underline{f}_2^x(\underline{\lambda}_i^{(t)}, \underline{\sigma}_i^{(t)}) \quad (44)$$

$$\hat{\underline{x}}_i^{(t+1)} = \underline{f}_1^x(\underline{\lambda}_i^{(t)}, \underline{\sigma}_i^{(t)}) \quad (45)$$

until convergence

As a result, we can compute the estimate and uncertainty on the calibration variable, at convergence of Algorithm 1, as

$$\hat{s}_\mu = f_1^s(\underline{y}_\mu, \underline{\omega}_\mu, \underline{V}_\mu) = \frac{1}{Z_{\text{out}}} \int ds_\mu s_\mu \int dz_\mu p_{\text{out}}^s(\underline{y}_\mu | z_\mu, s_\mu) p_s(s_\mu) \mathcal{N}(z_\mu; \underline{\omega}_\mu, \underline{V}_\mu) \quad (49)$$

$$C_\mu^s = f_2^s(\underline{y}_\mu, \underline{\omega}_\mu, \underline{V}_\mu) = \frac{1}{Z_{\text{out}}} \int ds_\mu s_\mu^2 \int dz_\mu p_{\text{out}}^s(\underline{y}_\mu | z_\mu, s_\mu) p_s(s_\mu) \mathcal{N}(z_\mu; \underline{\omega}_\mu, \underline{V}_\mu) - \hat{s}_\mu^2. \quad (50)$$

Relation to original cal-AMP derivation In [1, 2], the cal-AMP algorithm was derived from the belief propagation equations on the scalar variables of the fully factorized distribution (over N , M and P). It is equivalent to Algorithm 1 if the covariance matrices $\underline{V}_\mu, \partial_{\omega} \underline{g}_{\text{out}\mu}, \underline{\sigma}_i, \underline{C}_{\underline{i}}^x$ are assumed to be diagonal. However, we recall that BP is only exact on a tree, where the incoming messages at each node are truly independent. On the dense factor graph of the GLM on scalar variables, this is approximately exact in the thermodynamic limit due to the random mixing and small scaling of the weight matrix \underline{W} . Here, by considering the reconstruction of P samples at the same time, sharing given realizations of the calibration variable \underline{s} , as well as the measurement matrix \underline{W} , additional correlations may arise. Writing the message passing on the vector variables in \mathbb{R}^P allows not to neglect them and therefore improve the final algorithm.

2.3 State Evolution for cal-AMP

In the large size limit, where $N \rightarrow +\infty$ at fixed $\alpha = M/N$, the performance of the AMP algorithm can be characterized by a set of simpler equations corresponding to a quenched average over the disorder (here the realizations of \underline{X}_0 , s_0 , \underline{Y} and \underline{W}), referred to as State Evolution (SE) [9]. Remarkably, the SE equations are equivalent to the saddle point equations of the replica free energy associated with the problem under the Replica Symmetric (RS) ansatz [29]. In [16], the teacher-student matched setting of the GLM on vectors is examined through the replica approach and the Bayes optimal State Evolution equations are obtained through this second strategy. In the following we present the alternative derivation of the State Evolution equations from the message passing without assuming a priori matched teacher and student models. Our starting point are the r-BP equations. Finally, we also introduce new State Evolution equations following the reconstruction of calibration variables.

2.3.1 State Evolution derivation for mismatched prior and channel

Definition of the overlaps The key quantities to track the dynamic of iterations and the fixed point(s) of AMP are the overlaps. Here, they are the $P \times P$ matrices

$$\underline{q} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i \hat{x}_i^T, \quad \underline{m} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i x_{0,i}^T, \quad \underline{q}_0 = \frac{1}{N} \sum_{i=1}^N x_{0,i} x_{0,i}^T. \quad (51)$$

Output parameters Under independent statistics of the entries of \underline{W} and under the assumption of independent incoming messages, the variable $\omega_{\mu \rightarrow i}$ defined in (12) is a sum of independent variables and follows a Gaussian distribution by the Central Limit Theorem. Its first and second moments are

$$\mathbb{E}_{\underline{W}} [\omega_{\mu \rightarrow i}] = \frac{1}{\sqrt{N}} \sum_{i' \neq i} \mathbb{E}_{\underline{W}} [W_{\mu i}] \hat{x}_{i' \rightarrow \mu} = 0, \quad (52)$$

$$\mathbb{E}_{\underline{W}} [\omega_{\mu \rightarrow i} \omega_{\mu \rightarrow i}^T] = \frac{1}{N} \sum_{i' \neq i} \mathbb{E}_{\underline{W}} [W_{\mu i}^2] \hat{x}_{i' \rightarrow \mu} \hat{x}_{i' \rightarrow \mu}^T \quad (53)$$

$$= \frac{1}{N} \sum_{i=1}^N \hat{x}_{i \rightarrow \mu} \hat{x}_{i \rightarrow \mu}^T + O(1/N) \quad (54)$$

$$= \frac{1}{N} \sum_{i=1}^N \hat{x}_i \hat{x}_i^T - \partial_{\underline{\lambda} f^x} \sigma_i B_{\mu \rightarrow i} \hat{x}_i^T - \left(\partial_{\underline{\lambda} f^x} \sigma_i B_{\mu \rightarrow i} \hat{x}_i^T \right)^T + O(1/N) \quad (55)$$

$$= \frac{1}{N} \sum_{i=1}^N \hat{x}_i \hat{x}_i^T + O(1/\sqrt{N}) \quad (56)$$

where we used the fact that $B_{\mu \rightarrow i}$ defined in (25) is of order $O(1/\sqrt{N})$. Similarly, the variable $z_{\mu \rightarrow i} = \sum_{i' \neq i} \frac{W_{\mu i'}}{\sqrt{N}} x_{i'}$ is Gaussian with first and second moments

$$\mathbb{E}_{\underline{W}} [z_{\mu \rightarrow i}] = \frac{1}{\sqrt{N}} \sum_{i' \neq i} \mathbb{E}_{\underline{W}} [W_{\mu i}] x_{0,i'} = 0, \quad (57)$$

$$\mathbb{E}_{\underline{W}} [z_{\mu \rightarrow i} z_{\mu \rightarrow i}^T] = \frac{1}{N} \sum_{i=1}^N x_{0,i} x_{0,i}^T + O(1/\sqrt{N}). \quad (58)$$

Furthermore, the covariance between $\underline{\omega}_{\mu \rightarrow i}$ and $\underline{z}_{\mu \rightarrow i}$ is

$$\mathbb{E}_{\underline{W}} [\underline{z}_{\mu \rightarrow i} \underline{\omega}_{\mu \rightarrow i}^T] = \frac{1}{N} \sum_{i' \neq i} \mathbb{E}_{\underline{W}} [W_{\mu i}^2] \underline{x}_{0, i'} \hat{\underline{x}}_{i' \rightarrow \mu}^T \quad (59)$$

$$= \frac{1}{N} \sum_{i=1}^N \underline{x}_{0, i'} \hat{\underline{x}}_{i \rightarrow \mu}^T + O(1/N) \quad (60)$$

$$= \frac{1}{N} \sum_{i=1}^N \underline{x}_{0, i'} \hat{\underline{x}}_i^T - \underline{x}_{0, i'} \partial_{\underline{\lambda}} f_{\underline{\lambda} \underline{\sigma}}^x \underline{\sigma}_i \underline{B}_{\mu \rightarrow i}^T + O(1/N) \quad (61)$$

$$= \frac{1}{N} \sum_{i=1}^N \underline{x}_{0, i'} \hat{\underline{x}}_i^T + O(1/\sqrt{N}). \quad (62)$$

Hence we find that for all μ -s and all i -s, $\underline{\omega}_{\mu \rightarrow i}$ and $\underline{z}_{\mu \rightarrow i}$ are approximately jointly Gaussian in the thermodynamic limit following a unique distribution $\mathcal{N}(\underline{z}_{\mu \rightarrow i}, \underline{\omega}_{\mu \rightarrow i}; \underline{0}, \underline{Q})$ with the block covariance matrix

$$\underline{Q} = \begin{bmatrix} \underline{q} & \underline{m} \\ \underline{m}^\top & \underline{q} \end{bmatrix}. \quad (63)$$

For the variance message $\underline{V}_{\mu \rightarrow i}$, defined in (13), we have

$$\mathbb{E}_{\underline{W}} [\underline{V}_{\mu \rightarrow i}] = \sum_{i' \neq i} \mathbb{E}_{\underline{W}} \left[\frac{W_{\mu i}^2}{N} \right] \underline{C}_{\mu \rightarrow i'}^x \quad (64)$$

$$= \sum_{i=1}^N \frac{1}{N} \underline{C}_{\mu \rightarrow i}^x + O(1/N) \quad (65)$$

$$= \sum_{i=1}^N \frac{1}{N} \underline{C}_{\mu \rightarrow i}^x + O(1/\sqrt{N}), \quad (66)$$

where using the developments of $\underline{\lambda}_{i \rightarrow \mu}$ and $\underline{\sigma}_{i \rightarrow \mu}$ (28)-(29) and the scaling of $\underline{B}_{\mu \rightarrow i}$ in $O(1/\sqrt{N})$ we replaced

$$\underline{C}_{\mu \rightarrow i}^x = f_{\underline{\sigma}}^x(\underline{\lambda}_{i \rightarrow \mu}, \underline{\sigma}_{i \rightarrow \mu}) = f_{\underline{\sigma}}^x(\underline{\lambda}_i, \underline{\sigma}_i) - \partial_{\underline{\lambda}} f_{\underline{\sigma}}^x \underline{\sigma}_i \underline{B}_{\mu \rightarrow i}^T = f_{\underline{\sigma}}^x(\underline{\lambda}_i, \underline{\sigma}_i) + O(1/\sqrt{N}). \quad (67)$$

Futhermore, we can check that under our assumptions

$$\lim_{N \rightarrow +\infty} \mathbb{E}_{\underline{W}} \left[\underline{V}_{\mu \rightarrow i}^2 - \mathbb{E}_{\underline{W}} [\underline{V}_{\mu \rightarrow i}]^2 \right] = 0, \quad (68)$$

where the square is taken element-wise, meaning that all $\underline{V}_{\mu \rightarrow i}$ concentrate on their mean in the thermodynamic limit, which is identical for all of them:

$$\underline{V} = \sum_{i=1}^N \frac{1}{N} \underline{C}_{\mu \rightarrow i}^x. \quad (69)$$

Input parameters Here we use the re-parametrization trick to express \underline{y}_μ as a function $g_0(\cdot)$ taking the calibration variable s_μ and a noise $\underline{\epsilon}_\mu \sim p_\epsilon(\underline{\epsilon}_\mu)$ as inputs: $\underline{y}_\mu = g_0(\underline{\omega}_\mu^\top \underline{X}_0, s_{0, \mu}, \underline{\epsilon}_\mu)$. Following (26)-(25) and

(35),

$$\underline{\underline{\sigma}}_i^{-1} \lambda_i = \sum_{\mu=1}^M \frac{W_{\mu i}}{\sqrt{N}} g_{\text{out}} \left(y_{\mu}, \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i} \right) \quad (70)$$

$$= \sum_{\mu=1}^M \frac{W_{\mu i}}{\sqrt{N}} g_{\text{out}} \left(g_0 \left(\sum_{i' \neq i} \frac{W_{\mu i'}}{\sqrt{N}} x_{0, i'} + \frac{W_{\mu i}}{\sqrt{N}} x_{0, i}, s_{0, \mu}, \epsilon_{\mu} \right), \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i} \right) \quad (71)$$

$$= \sum_{\mu=1}^M \frac{W_{\mu i}}{\sqrt{N}} g_{\text{out}} \left(g_0 \left(\sum_{i' \neq i} \frac{W_{\mu i'}}{\sqrt{N}} x_{0, i'}, s_{0, \mu}, \epsilon_{\mu} \right), \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i} \right) \\ + \sum_{\mu=1}^M \frac{W_{\mu i}^2}{N} \partial_z g_{\text{out}} \left(g_0 \left(z_{\mu \rightarrow i}, s_{0, \mu}, \epsilon_{\mu} \right), \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i} \right) x_{0, i}. \quad (72)$$

The first term is again a sum of independent random variables, given the $W_{\mu i}$ are i.i.d. with zero mean and the messages of type $\mu \rightarrow i$ are assumed independent. The second term has non-zero mean and can be shown to concentrate. Recalling that all $\underline{V}_{\mu \rightarrow i}$ also concentrate on \underline{V} , we obtain that $\underline{\underline{\sigma}}_i^{-1} \lambda_i$ effectively follows the Gaussian distribution (I_P is the $P \times P$ identity)

$$\underline{\underline{\sigma}}_i^{-1} \lambda_i \sim \mathcal{N} \left(\underline{\underline{\sigma}}_i^{-1} \lambda_i; \alpha \hat{\underline{m}} x_{0, i}, \sqrt{\alpha \hat{\underline{q}}} I_P \right) \quad (73)$$

with

$$\hat{\underline{q}} = \int d\underline{\epsilon} p_{\epsilon}(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) g_{\text{out}}(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V}) \times \\ g_{\text{out}}(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V})^T, \quad (74)$$

$$\hat{\underline{m}} = \int d\underline{\epsilon} p_{\epsilon}(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \partial_z g_{\text{out}}(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V}). \quad (75)$$

The inverse variance $\underline{\underline{\sigma}}_i^{-1}$ can be shown to concentrate on its mean

$$\underline{\underline{\sigma}}_i^{-1} = \sum_{\mu=1}^M \frac{W_{\mu i}^2}{N} \partial_{\omega} g_{\text{out}}(y_{\mu}, \underline{\omega}_{\mu \rightarrow i}, \underline{V}_{\mu \rightarrow i}) \simeq \alpha \hat{\underline{\chi}}, \quad (76)$$

$$\hat{\underline{\chi}} = - \int d\underline{\epsilon} p_{\epsilon}(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \partial_{\omega} g_{\text{out}}(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V}). \quad (77)$$

Closing the equations The statistics of the input parameters λ_i and $\underline{\underline{\sigma}}_i$ must be consistent with the overlap matrices given

$$\underline{V} = \frac{1}{N} \sum_{i=1}^N \underline{C}_{\underline{=i}}^x = \mathbb{E}_{\lambda, \underline{\sigma}} \left[f^x(\lambda, \underline{\sigma}) \right], \quad (78)$$

$$\underline{q} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i \hat{x}_i^{\top} = \mathbb{E}_{\lambda, \underline{\sigma}} \left[f^x(\lambda, \underline{\sigma}) f^x(\lambda, \underline{\sigma})^{\top} \right], \quad (79)$$

$$\underline{m} = \frac{1}{N} \sum_{i=1}^N \hat{x}_i x_{0, i}^{\top} = \mathbb{E}_{\lambda, \underline{\sigma}} \left[f^x(\lambda, \underline{\sigma}) x_{0, i}^{\top} \right], \quad (80)$$

which gives upon expressing the computation of the expectations (here $\mathcal{D}_{\underline{\xi}}$ is the P -dimensional multivariate standard Gaussian distribution with $\underline{0}$ mean and identity covariance)

$$\underline{V} = \int d\underline{x}_0 p_{x_0}(\underline{x}_0) \int \mathcal{D}_{\underline{\xi}} f_{\underline{2}}^x \left((\alpha \hat{\underline{\chi}})^{-1} \left(\sqrt{\alpha \hat{q}} \underline{\xi} + \alpha \hat{m} \underline{x}_0 \right); (\alpha \hat{\underline{\chi}})^{-1} \right), \quad (81)$$

$$\underline{m} = \int d\underline{x}_0 p_{x_0}(\underline{x}_0) \int \mathcal{D}_{\underline{\xi}} f_{\underline{1}}^x \left((\alpha \hat{\underline{\chi}})^{-1} \left(\sqrt{\alpha \hat{q}} \underline{\xi} + \alpha \hat{m} \underline{x}_0 \right); (\alpha \hat{\underline{\chi}})^{-1} \right) \underline{x}_0^\top, \quad (82)$$

$$\underline{q} = \int d\underline{x}_0 p_{x_0}(\underline{x}_0) \int \mathcal{D}_{\underline{\xi}} f_{\underline{1}}^x \left((\alpha \hat{\underline{\chi}})^{-1} \left(\sqrt{\alpha \hat{q}} \underline{\xi} + \alpha \hat{m} \underline{x}_0 \right); (\alpha \hat{\underline{\chi}})^{-1} \right) \times \\ f_{\underline{1}}^x \left((\alpha \hat{\underline{\chi}})^{-1} \left(\sqrt{\alpha \hat{q}} \underline{\xi} + \alpha \hat{m} \underline{x}_0 \right); (\alpha \hat{\underline{\chi}})^{-1} \right)^\top. \quad (83)$$

The State Evolution analysis of the GLM with vector variables, that allows to track the vectorial version of GAMP given by Algorithm 1, finally consists in iterating alternatively the equations (74), (75), (77), and the equations (81), (82) (83) until convergence.

Reconstruction of the calibration variable In parallel, the reconstruction of \underline{s}_0 can be followed by introducing the scalar overlaps

$$r = \frac{1}{M} \sum_{\mu=1}^M \hat{s}_{\mu}^2, \quad \nu = \frac{1}{M} \sum_{\mu=1}^M \hat{s}_{\mu} s_{0,\mu}, \quad r_0 = \frac{1}{M} \sum_{\mu=1}^M s_{0,\mu}^2. \quad (84)$$

Recalling the definition of the estimator $\hat{s}(\cdot) = f_{\underline{1}}^s(\cdot)$ given by (49) and taking the steps of the derivation above, one finds that the calibration overlaps can be computed using the SE variables introduced above,

$$r = \int d\underline{\epsilon} p_{\epsilon}(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \hat{s} \left(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V} \right)^2, \quad (85)$$

$$\nu = \int d\underline{\epsilon} p_{\epsilon}(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \hat{s} \left(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{V} \right) s_0. \quad (86)$$

Performance analysis The mean squared error (MSE) on the reconstruction of \underline{X} by the AMP algorithm is then predicted by

$$\text{MSE}(\underline{X}) = q - 2m + q_0, \quad (87)$$

where the scalar values used here correspond to the (unique) value of the diagonal elements of the corresponding overlap matrices. This MSE can be computed throughout the iterations of State Evolution. Similarly, the MSE in the reconstruction of the calibration variable can be computed as

$$\text{MSE}(\underline{s}) = r - 2\nu + r_0, \quad (88)$$

throughout the iterations. Remarkably, the State Evolution MSEs follow precisely the MSE of the cal-AMP predictors along the iterations of the algorithm provided the procedures are initialized consistently. A random initialization of \hat{x}_i and \hat{s} in cal-AMP corresponds to an initialization of zero overlap $m = 0$, $\nu = 0$, with variance of the priors $q = q_0$, $r = r_0$, in the State Evolution.

2.3.2 Bayes optimal State Evolution

The SE equations can be greatly simplified in the Bayes optimal setting where the statistical model used by the student (priors p_x and p_s , and channel p_{out}) matches the teacher model. In this case, the true unknown signal \underline{X}_0 is in some sense statistically equivalent to the estimate $\hat{\underline{X}}$ coming from the posterior. More precisely

one can prove the Nishimori identities [30, 31, 32] (or [33] for a concise demonstration and discussion) implying that

$$\underline{q} = \underline{m}, \quad \underline{V} = \underline{q}_0 - \underline{m}, \quad \underline{\hat{q}} = \underline{\hat{m}} = \underline{\hat{x}} \quad \text{and} \quad r = \nu. \quad (89)$$

As a result the State Evolution reduces to a set of three equations

$$r = \int d\underline{\epsilon} p_\epsilon(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \hat{s} \left(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{q}_0 - \underline{m} \right)^2, \quad (90)$$

$$\underline{m} = \int d\underline{x}_0 p_{x_0}(\underline{x}_0) \int \mathcal{D}\underline{\xi} f_1^x \left((\alpha \underline{\hat{m}})^{-1} \left(\sqrt{\alpha \underline{\hat{m}}} \underline{\xi} + \alpha \underline{\hat{m}} \underline{x}_0 \right); (\alpha \underline{\hat{m}})^{-1} \right) \underline{x}_0^\top, \quad (91)$$

$$\underline{\hat{m}} = \int d\underline{\epsilon} p_\epsilon(\underline{\epsilon}) ds_0 p_{s_0}(s_0) \int d\underline{\omega} d\underline{z} \mathcal{N}(\underline{z}, \underline{\omega}; \underline{0}, \underline{Q}) \underline{g}_{\text{out}} \left(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{q}_0 - \underline{m} \right) \times \underline{g}_{\text{out}} \left(g_0(\underline{z}, s_0, \underline{\epsilon}), \underline{\omega}, \underline{q}_0 - \underline{m} \right)^\top, \quad (92)$$

with the block covariance matrix

$$\underline{Q} = \begin{bmatrix} \underline{q}_0 & \underline{m} \\ \underline{m}^\top & \underline{m} \end{bmatrix}. \quad (93)$$

2.4 Replica free energy

The approximate inference via message passing presented in the previous sections is related to the replica approximation of the free energy of the problem defined as

$$F_N = -\log p(\underline{Y}|\underline{W}) = -\log \int d\underline{X} ds p_{\text{out}}^s(\underline{Y}|\underline{W}\underline{X}, s) p_x(\underline{X}) p_s(s). \quad (94)$$

More precisely, the replica approximation [34] of the asymptotic expected free energy density

$$f = -\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\underline{W}} \int d\underline{Y} p(\underline{Y}|\underline{W}) \log p(\underline{Y}|\underline{W}) \quad (95)$$

can be directly derived from GAMP on vector variables as presented in [16] (there applied to committee machines). With our notations, the Bayes optimal free energy density is

$$-f = \max \left\{ \text{extr}_{\underline{m}, \underline{\hat{m}}} \left[-\frac{1}{2} \text{Tr}(\underline{m} \underline{\hat{m}}) + \mathcal{I}_x(\underline{\hat{m}}) + \alpha \mathcal{I}_z(\underline{m}) \right] \right\} \quad (96)$$

with (again \mathcal{D} is used for a standard multivariate Gaussian density)

$$\mathcal{I}_x(\underline{\hat{m}}) = \int_{\mathbb{R}^P} \mathcal{D}\underline{\xi} \int_{\mathbb{R}^P} d\underline{x}_0 p_x(\underline{x}_0) e^{-\frac{1}{2} \underline{x}_0^\top \underline{\hat{m}} \underline{x}_0 + \underline{\xi}^\top \underline{\hat{m}}^{1/2} \underline{x}_0} \log \left(\int_{\mathbb{R}^P} d\underline{x} p_x(\underline{x}) e^{-\frac{1}{2} \underline{x}^\top \underline{\hat{m}} \underline{x} + \underline{\xi}^\top \underline{\hat{m}}^{1/2} \underline{x}} \right), \quad (97)$$

$$\mathcal{I}_z(\underline{m}) = \int_{\mathbb{R}^P} d\underline{y} \int_{\mathbb{R}^P} \mathcal{D}\underline{\xi} \int_{\mathbb{R}^P} \mathcal{D}\underline{z}_0 \int ds_0 p_s(s_0) p_{\text{out}}^s(\underline{y} | (\underline{q}_0 - \underline{m})^{1/2} \underline{z}_0 + \underline{q}^{1/2} \underline{\xi}; s_0) \times \log \left(\int_{\mathbb{R}^P} \mathcal{D}\underline{z} ds p_s(s) p_{\text{out}}^s(\underline{y} | (\underline{q}_0 - \underline{m})^{1/2} \underline{z} + \underline{q}^{1/2} \underline{\xi}; s) \right). \quad (98)$$

In the case of mismatched teacher and student, the formula can be generalized and the expected free energy density

$$f = -\lim_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{\underline{W}} \int d\underline{Y} d\underline{X}_0 ds_0 p_{\text{out},0}^s(\underline{Y}|\underline{W}\underline{X}_0, s_0) p_{x_0}(\underline{X}_0) p_{s_0}(s_0) \times \log \left(\int d\underline{X} ds p_{\text{out}}^s(\underline{Y}|\underline{W}\underline{X}, s) p_x(\underline{X}) p_s(s) \right) \quad (99)$$

is now approximated as the extremum of a potential over all the overlap and auxiliary matrices

$$\underline{q} = \frac{1}{N} \sum_{i=1}^N \underline{x}_{a,i} \underline{x}_{a,i}^\top, \quad \underline{m} = \frac{1}{N} \sum_{i=1}^N \underline{x}_{0,i} \underline{x}_{a,i}^\top, \quad \underline{q}_{\underline{12}} = \frac{1}{N} \sum_{i=1}^N \underline{x}_{a,i} \underline{x}_{b,i}, \quad (100)$$

where a and b are indices of different replicas in the computation. We obtain

$$-f = \max \left\{ \text{extr}_{\underline{q}, \underline{\hat{q}}, \underline{m}, \underline{\hat{m}}, \underline{q}_{\underline{12}}, \underline{\hat{q}}_{\underline{12}}} \phi(\underline{q}, \underline{\hat{q}}, \underline{m}, \underline{\hat{m}}, \underline{q}_{\underline{12}}, \underline{\hat{q}}_{\underline{12}}) \right\}, \quad (101)$$

with

$$\phi(\underline{q}, \underline{\hat{q}}, \underline{m}, \underline{\hat{m}}, \underline{q}_{\underline{12}}, \underline{\hat{q}}_{\underline{12}}) = -\text{Tr}(\underline{m}\underline{\hat{m}}) + \text{Tr}(\underline{q}\underline{\hat{q}}) + \frac{1}{2} \text{Tr}(\underline{q}_{\underline{12}} \underline{\hat{q}}_{\underline{12}}) + \mathcal{I}_x(\underline{\hat{q}}, \underline{\hat{m}}, \underline{\hat{q}}_{\underline{12}}) + \alpha \mathcal{I}_z(\underline{q}, \underline{m}, \underline{q}_{\underline{12}}), \quad (102)$$

and

$$\begin{aligned} \mathcal{I}_x(\underline{\hat{q}}, \underline{\hat{m}}, \underline{\hat{q}}_{\underline{12}}) &= \int \mathcal{D}\underline{\xi} \int d\underline{x}_0 p_x(\underline{x}_0) \log \left(\int d\underline{x} p_x(\underline{x}) e^{\underline{x}_0^\top \underline{\hat{m}} \underline{x} - \frac{1}{2} \underline{x}^\top (\underline{\hat{q}} + \underline{\hat{q}}_{\underline{12}}) \underline{x} + \underline{\xi}^\top \underline{\hat{q}}_{\underline{12}}^{1/2} \underline{x}} \right) \\ \mathcal{I}_z(\underline{q}, \underline{m}, \underline{q}_{\underline{12}}) &= \int d\underline{y} \int \mathcal{D}\underline{\xi} \int \mathcal{D}\underline{z}_0 \int d\underline{s}_0 p_{\underline{s}_0}(\underline{s}_0) p_{\text{out},0}^s(\underline{y} | (\underline{q}_{\underline{0}} - \underline{m}\underline{q}_{\underline{12}}^{-1}\underline{m}^\top)^{1/2} \underline{z}_0 + \underline{m}\underline{q}_{\underline{12}}^{-1/2} \underline{\xi}; \underline{s}_0) \\ &\quad \times \log \left(\int \mathcal{D}\underline{z} \int d\underline{s} p_s(\underline{s}) p_{\text{out}}^s(\underline{y} | (\underline{q} - \underline{q}_{\underline{12}})^{1/2} \underline{z} + \underline{q}_{\underline{12}}^{1/2} \underline{\xi}, \underline{s}) \right). \end{aligned} \quad (104)$$

Writing the extremization conditions for the potential $\phi(\underline{q}, \underline{\hat{q}}, \underline{m}, \underline{\hat{m}}, \underline{q}_{\underline{12}}, \underline{\hat{q}}_{\underline{12}})$ yields a set of self consistency equations for the overlap matrices. This set of coupled equations can be shown to be equivalent to the SE fixed point equations, (91)-(92) for the Bayes optimal case, and (74)-(83) for the mismatched case. This equivalence reveals the connection between the replica method and message passing algorithms at first apparently unrelated.

Finally, let us comment on the validity of the formula for the asymptotic free energy in the Bayes optimal case. We conjecture that when one selects the maximum value between all the extrema in (96), the replica prediction is asymptotically correct. A rigorous justification can be provided using the approach used for the committee machine in [16], based on interpolation technics developed in [35, 36]. It is rather straightforward to repeat these steps for the present problem, which should lead to a complete proof of the conjecture (96).

3 Online algorithm and analysis

In machine learning applications, the *offline* strategy, where all the training data is exploited at once, is sometimes made difficult by its memory cost. Instead, data may be partitioned into mini-batches of a few samples, or even considered one sample at the time. This approach may be preferred even when memory is not an issue, as in deep learning with Stochastic Gradient Descent (SGD) [12]. Here we will be interested in the Bayesian online learning of the calibration variables as the observations are received. The main concept behind Bayesian online learning algorithms, such as Assumed Density Filtering (ADF) [37, 27], is to update a prior along upon receiving/processing different data points. In the statistical physics literature, Bayesian online learning algorithms were proposed and analyzed for simple neural networks [38, 39, 40, 41] and compressed sensing [42].

In the present section, we propose an online algorithm, online cal-AMP, for the calibration problem of the GLM. Our algorithm takes inspiration from the streaming version of AMP for the simple GLM of [17], which we review in the following section. We also derive the corresponding State Evolution analysis.

3.1 Streaming AMP for online reconstruction in the GLM

In [17], a mini-batch version of the GAMP algorithm is proposed. On the example of the GLM, one imagines receiving at each iteration k a subset $\underline{y}_{(k)}$ of the components of $\underline{y} \in R^M$ generated via a channel

$$\underline{y} \sim p_{\text{out}}(\underline{y} | \underline{W}\underline{x}), \quad (105)$$

Algorithm 2 Offline Gain Calibration State Evolution

Input: matrix $\underline{Y} \in \mathbb{R}^{M \times P}$ and matrix $\underline{W} \in \mathbb{R}^{M \times N}$:

Initialize:

$$t = 0, m^{(0)} = 0, V^{(0)} = q_0 = \rho,$$

$$\forall \mu = 1, \dots, N_{\text{MC}} \quad s_{0,\mu} \sim p_{s_0}(s_{0,\mu})$$

repeat

1) Draw Monte Carlo samples for update of \hat{m} (115) and r (113)

$$\forall \mu = 1, \dots, N_{\text{MC}}$$

$$\underline{z}_\mu, \underline{\omega}_\mu \sim \mathcal{N}(\underline{z}_\mu, \underline{\omega}_\mu; \underline{0}, \underline{Q}^{(t)})$$

$$\underline{\epsilon} \sim p_\epsilon(\underline{\epsilon})$$

$$\underline{y}_\mu = g_0(\underline{z}_\mu, s_{0,\mu}, \underline{\epsilon})$$

$$\text{with } \underline{Q}^{(k)} = \begin{bmatrix} q_0 I_P & m^{(t)} I_P \\ m^{(t)} I_P & m^{(t)} I_P \end{bmatrix} \in \mathbb{R}^{2P \times 2P}$$

2) Compute integrands

2.2) Compute $\hat{s}_\mu(\underline{y}_\mu, \underline{z}_\mu, \underline{\omega}_\mu, \underline{V}^{(t)})$ (123)

2.3) Compute $g_{\text{out}\mu}(\underline{y}_\mu, \underline{z}_\mu, \underline{\omega}_\mu, \underline{V}^{(t)})$ (121)

3) Update r and \hat{m}

$$r^{(t)} = \frac{1}{N_{\text{MC}}} \sum_{\mu=1}^{N_{\text{MC}}} \left(\hat{s}_\mu^{(t)} \right)^2 \quad \hat{m}^{(t)} = \frac{1}{N_{\text{MC}}} \sum_{\mu=1}^{N_{\text{MC}}} \left(g_{\text{out}\mu} \right)^2$$

3) Update $m^{(t+1)}$ by numerical integration using (114), and $V^{(t+1)} = q_0 - m^{(t+1)}$.

$$t = t + 1$$

until convergence

Output: time series $\{V^{(t)}, m^{(t)}, \hat{m}^{(t)}; t = 1 \dots t_{\text{max}}\}$

from which one wishes to reconstruct $\underline{x} \in \mathbb{R}^N$. Bayes formula gives the posterior distribution over \underline{x} after seeing k mini-batches

$$p(\underline{x} | \underline{y}_{(k)}, \{\underline{y}_{(k-1)}, \dots, \underline{y}_{(1)}\}, \underline{W}) = \frac{p(\underline{y}_{(k)} | \underline{x}, \underline{W}) p(\underline{x} | \{\underline{y}_{(k-1)}, \dots, \underline{y}_{(1)}\}, \underline{W})}{\int d\underline{x} p(\underline{y}_{(k)} | \underline{x}, \underline{W}) p(\underline{x} | \{\underline{y}_{(k-1)}, \dots, \underline{y}_{(1)}\}, \underline{W})}. \quad (106)$$

The formula suggests the iterative scheme of using as a prior on \underline{x} at iteration k the posterior obtained at iteration $k - 1$. This idea can be implemented in different approximate inference algorithms, as proposed by [43] using a variational method. In the regular version of GAMP [11] an effective factorized posterior is given at convergence by the input update functions (scalar equivalent of (31)-(32)):

$$p(\underline{x} | \underline{y}, \underline{W}) \simeq \prod_{i=1}^N \frac{1}{\mathcal{Z}_x(\lambda_i, \sigma_i)} p_x(x_i) \exp \left\{ - \frac{(\lambda_i - x_i)^2}{2\sigma_i} \right\}. \quad (107)$$

Plugging this posterior approximation in the iterative scheme yields the mini-AMP algorithm using the converged values of $\lambda_{(\ell),i}$ and $\sigma_{(\ell),i}$ at each anterior mini-batch $\ell < k$ to compute the prior

$$p_x^{(k)}(\underline{x}) = p(\underline{x} | \{\underline{y}_{(k-1)}, \dots, \underline{y}_{(1)}\}, \underline{W}) \simeq \prod_{i=1}^N \frac{1}{\mathcal{Z}_{x,i}} p_x(x_i) \exp \left\{ - \sum_{\ell=1}^{k-1} \frac{(\lambda_{(\ell),i} - x_i)^2}{2\sigma_{(\ell),i}} \right\}, \quad (108)$$

where the $\mathcal{Z}_{x,i}$ normalizes its respective marginal. Compared to a naive mean-field variational approximation of the posterior, AMP takes into account more correlations and is indeed found to perform better in experiments

reported by [17]. Another advantage of the AMP based online inference is that it is amenable to theoretical analysis by a corresponding State Evolution. In the following we apply these ideas to the online learning of the calibration variables in the model we defined in Section 2.1.

3.2 Online cal-AMP

We consider here the analysis of the online reconstruction of the calibration variable \underline{s}_0 by tracking \underline{s} as the observations $\underline{y}_{(k)} \in \mathbb{R}^M$ are treated consecutively. We readily adapt the strategy of [17], by using at step $k + 1$ the approximate posterior on \underline{s} at step k as an effective prior on s_μ in cal-AMP.

Algorithm The AMP algorithm consists here in restarting cal-AMP at each new observation $\underline{y}_{(k)}$ (that is with $P = 1$ sample in the notation of Section 2), while updating the prior used for the calibration variable. From the definition of the approximate posterior (48), we obtain the recursion on the effective prior $p_{s_\mu}^{(k+1)}$ on s_μ :

$$p_{s_\mu}^{(k+1)}(s_\mu) = m_\mu^{s,(k)}(s_\mu) \quad (109)$$

$$= \frac{1}{\mathcal{Z}_{\text{out}}^{(k)}} \int dz p_{\text{out}}^s(y_{(k),\mu}|z, s_\mu) p_{s_\mu}^{(k)}(s_\mu) \mathcal{N}(z; \omega_{(k),\mu}, V_{(k),\mu}), \quad (110)$$

where the output variables $\omega_{(k),\mu}$ and $V_{(k),\mu}$ correspond to the values at convergence (or at the last iteration t_{max}) of the cal-AMP algorithm at the previous step k . In Section 4.1, we consider the gain calibration problem and specify effective strategies to implement this recursion within the AMP algorithm.

State Evolution The streaming State Evolution analysis of the online cal-AMP is also adapted using the above recursion. Note that the effective prior at a given step/sample P (we assume the samples to be processed one at a time) depends on the output variables of the algorithm for all the previously seen samples. Expanding the recursion above we have:

$$p_{s_\mu}^{(P)}(s_\mu) = \frac{1}{\mathcal{Z}_{\text{out}}^{(P-1)}} \int \prod_{k=1}^{P-1} \left(dz_{(k)} p_{\text{out}}^s(y_{(k),\mu}|z_{(k)}, s_\mu) \mathcal{N}(z_{(k)}; \omega_{(k),\mu}, V_{(k),\mu}) \right) p_s(s_\mu) \quad (111)$$

with

$$\mathcal{Z}_{\text{out}}^{(P)} = \mathcal{Z}_{\text{out}}^{(P)}(\{y_{(k),\mu}, \omega_{(k),\mu}, V_{(k),\mu}\}_{k=1}^P), \quad (112)$$

where again for each k the output variables $\underline{\omega}_{(k)}$ and $\underline{V}_{(k)}$ are the converged values for the corresponding step k . The dependence of the normalization $\mathcal{Z}_{\text{out}}^{(P)}$ on the output variables is reflected in the definitions at step P of the output update function $g_{\text{out}}^{(P)} = \partial \log \mathcal{Z}_{\text{out}}^{(P)} / \partial \underline{\omega}_{(P)}$ and calibration update function $\hat{\underline{s}}^{(P)} = \int d\underline{s} \underline{s} m^{s,(P)}(\underline{s})$. Therefore, the State Evolution involves an averaging over all the output variables $\underline{\omega}_{(k)}$ and $\underline{V}_{(k)}$ relative to the already processed samples.

More precisely, in the Bayes optimal setting, it is easy to see from the cal-AMP SE, that the online

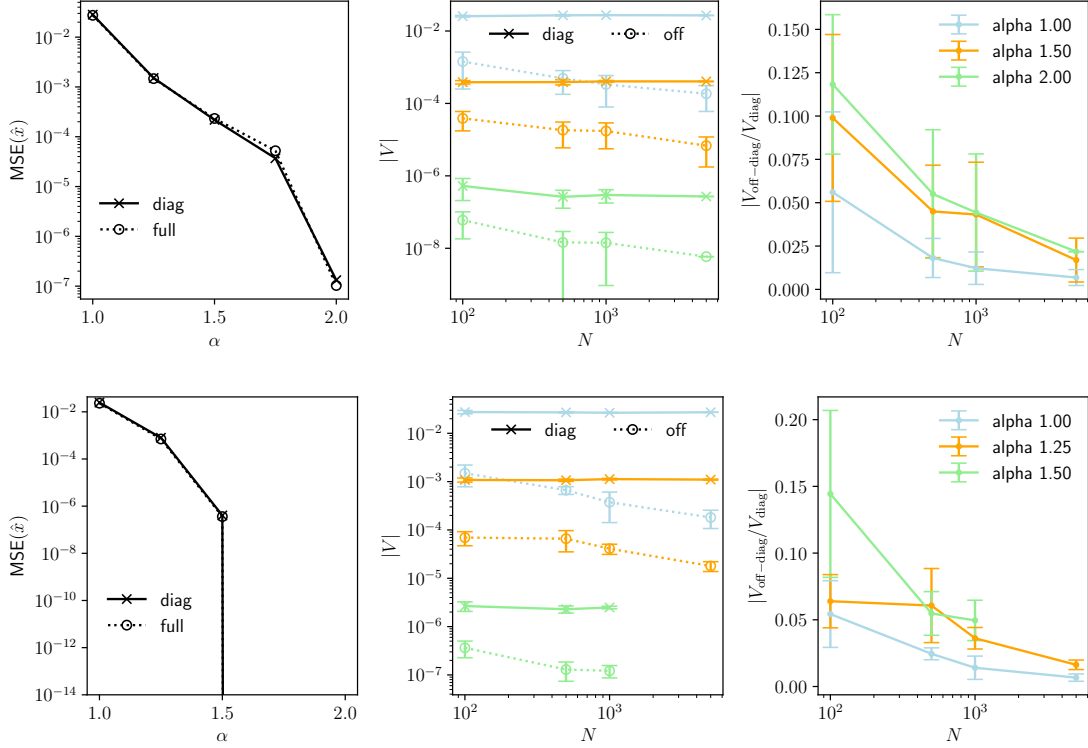


Figure 2: Full covariance matrices and departure from the diagonal ansatz for Gaussian inputs $\rho = 1$ (**top row**) $P = 2$ (**bottom row**) $P = 3$. **Left:** Comparison of the MSEs achieved by cal-AMP for full unconstrained covariance matrices and diagonal covariance matrices for $N = 1000$. **Middle:** Average value at fixed points of the diagonal elements and the off-diagonal elements of \underline{V} when unconstrained as a function of the size of the problem N . **Right:** Average absolute value of the ratio between off-diagonal and diagonal elements of \underline{V} as a function of N .

algorithm is characterized by the equations:

$$r_{(P)} = \int d\epsilon p_\epsilon(\epsilon) ds_0 p_{s_0}(s_0) \int d\omega dz \mathcal{N}(z, \omega; \underline{0}, \underline{Q}_{(P)}) \quad (113)$$

$$\prod_{k=1}^{P-1} \int d\omega_{(k)} dz_{(k)} \mathcal{N}(z_{(k)}, \omega_{(k)}; \underline{0}, \underline{Q}_{(k)}) \hat{s}^{(P)}(g_0(z, s_0, \epsilon), \omega, q_0 - m)^2,$$

$$m_{(P)} = \int dx_0 p_{x_0}(x_0) \int \mathcal{D}\xi f_1^x \left((\alpha \hat{\chi})^{-1} \left(\sqrt{\alpha \hat{m}} \xi + \alpha \hat{m}_{(P)} x_0 \right); (\alpha \hat{m}_{(P)})^{-1} \right) x_0, \quad (114)$$

$$\hat{m}_{(P)} = \int d\epsilon p_\epsilon(\epsilon) ds_0 p_{s_0}(s_0) \int d\omega dz \mathcal{N}(z, \omega; \underline{0}, \underline{Q}_{(P)}) \prod_{k=1}^{P-1} \int d\omega_{(k)} dz_{(k)} \mathcal{N}(z_{(k)}, \omega_{(k)}; \underline{0}, \underline{Q}_{(k)}) g_{\text{out}}^{(P)}(g_0(z, s_0, \epsilon), \omega, q_0 - m)^2, \quad (115)$$

where the 2×2 covariance matrices $\underline{Q}_{(k)}$, are analogous to the block covariance matrix (93) for each step k using the corresponding scalar fixed points $m_{(k)}$. In the following Section, we will discuss how to implement this State Evolution in practice, focusing on the specific problem of gain calibration.

Algorithm 3 Online Gain Calibration State Evolution

Input: matrix $\underline{Y} \in \mathbb{R}^{M \times P}$ and matrix $\underline{W} \in \mathbb{R}^{M \times N}$:

Initialize: $t = 0$,

$\forall \mu = 1, \dots, N_{\text{MC}}$

$$s_\mu^0 \sim p_{s^0}(s_\mu^0)$$

$$\Lambda_{0,\mu} = 0, \quad \Sigma_{0,\mu} = 0$$

for $k = 1, \dots, P$ **do**

Initialize: $t = 0$, $m_{(k)}^{(0)} = 0$, $V_{(k)}^{(0)} = q_0 = \rho$

repeat

1) Draw Monte Carlo samples for update of \hat{m} (115) and r (113)

$\forall \mu = 1, \dots, N_{\text{MC}}$

$$(z_{\mu,k}, \omega_{\mu,k}) \sim \mathcal{N}(z_{\mu,k}, \omega_{\mu,k}; \underline{Q}, \underline{Q}_{(k)}^{(t)})$$

$$\epsilon \sim p_\epsilon(\epsilon)$$

$$y_{\mu,k} = g^0(z_{\mu,k}, s_\mu^0, \epsilon)$$

$$\text{with } \underline{Q}_{(k)}^{(t)} = \begin{bmatrix} q_0 & m_{(k)}^{(t)} \\ m_{(k)}^{(t)} & m_{(k)}^{(t)} \end{bmatrix}$$

2) Following step (2) of Algorithm 4 with

- samples of previous steps $\{y_{\mu,l}, z_{\mu,l}, \omega_{\mu,l}\}_{l \leq k-1}$ at convergence

- current $y_{\mu,k}, z_{\mu,k}, \omega_{\mu,k}$

- current $V_{(k)}^{(t)}$

2.1) Update $\Lambda_{k,\mu}, \Sigma_{k,\mu}$

2.2) Compute \hat{s}_μ

2.3) Compute $g_{\text{out}\mu,k}$

3) Update $r_{(k)}$ and $\hat{m}_{(k)}$

$$r_{(k)}^{(t)} = \frac{1}{N_{\text{MC}}} \sum_{\mu=1}^{N_{\text{MC}}} (\hat{s}_\mu^{(t)})^2 \quad \hat{m}_{(k)}^{(t)} = \frac{1}{N_{\text{MC}}} \sum_{\mu=1}^{N_{\text{MC}}} (g_{\text{out}\mu,k})^2$$

4) Update $m_{(k)}^{(t+1)}$ by numerical integration using (114), and $V_{(k)}^{(t+1)} = q_0 - m_{(k)}^{(t+1)}$.

$t = t + 1$

until convergence

end for

Output: time series $\{V_{(k)}^{(t)}, m_{(k)}^{(t)}, \hat{m}_{(k)}^{(t)}; t = 1, \dots, t_{\text{max}}\}_{k=1}^P$

4 Numerical tests

4.1 Gain calibration setting and update functions

As a test case, we consider the following problem of gain calibration [7, 8, 1, 2]. The input signal is known to be ρ -sparse and distributed according to a Gauss-Bernoulli distribution,

$$\forall i = 1, \dots, N, \quad p_{x_0}(\underline{x}_i) = \prod_{k=1}^P (\rho \mathcal{N}(x_{(k),i}; 0, 1) + (1 - \rho) \delta(x_{(k),i})) . \quad (116)$$

Each component of the output includes a division by a calibration variable that is uniformly distributed in a positive interval $[a, b]$:

$$\forall \mu = 1, \dots, M, \quad p_{s_0}(s_\mu) = \mathbb{1}_{[a,b]}/(b-a), \quad 0 < a < b, \quad (117)$$

$$\underline{y}_\mu = \frac{1}{s_{0,\mu}}(\underline{w}_\mu^\top \underline{X}_0 + \epsilon), \quad \epsilon \sim \mathcal{N}(\epsilon, \underline{0}, \Delta I_P). \quad (118)$$

We will consider both the offline reconstruction where the P samples are exploited simultaneously and the online case.

Output functions For any value of P , the output functions have analytical expressions in this setting. The channel distribution and partition function are

$$p_{\text{out}}(\underline{y}_\mu | \underline{z}_\mu) = \int_a^b ds_\mu p_s(s_\mu)(s_\mu)^P \mathcal{N}(\underline{z}_\mu; s_\mu \underline{y}_\mu, \Delta I_P), \quad (119)$$

$$\begin{aligned} \mathcal{Z}_{\text{out}}(\underline{y}, \underline{\omega}, \underline{V}) &= \int d\underline{z} p_{\text{out}}(\underline{y} | \underline{z}) \mathcal{N}(\underline{z}; \underline{\omega}, \underline{V}) = \int d\underline{z} \int_a^b ds p_s(s)(s)^P \mathcal{N}(\underline{z}; s\underline{y}, \Delta I_P) \mathcal{N}(\underline{z}; \underline{\omega}, \underline{V}) \\ &= \int_a^b ds p_s(s)(s)^P \mathcal{N}(\underline{\omega}; s\underline{y}, \underline{V} + \Delta I_P) \end{aligned} \quad (120)$$

which give

$$\underline{g}_{\text{out}}(\underline{y}, \underline{\omega}, \underline{V}) = \frac{1}{\mathcal{Z}_{\text{out}}} \partial_{\underline{\omega}} \mathcal{Z}_{\text{out}} = (\underline{V} + \Delta I_P)^{-1} (\hat{s}(\underline{y}, \underline{\omega}, \underline{V}) \underline{y} - \underline{\omega}), \quad (121)$$

$$\partial_{\underline{\omega}} \underline{g}_{\text{out}}(\underline{y}, \underline{\omega}, \underline{V}) = C^s(\underline{y}, \underline{\omega}, \underline{V}) (\underline{V} + \Delta I_P)^{-1} \underline{y} \underline{y}^\top (\underline{V} + \Delta I_P)^{-1} - (\underline{V} + \Delta I_P)^{-1}. \quad (122)$$

These expressions involve the estimate and variance of the calibration variables under the approximate posterior which can be computed as

$$\hat{s}(\underline{y}, \underline{\omega}, \underline{V}) = f_1^s(\underline{y}, \underline{\omega}, \underline{V}) = \frac{\int_a^b ds (s)^{P+1} \mathcal{N}(\underline{z}; s\underline{y}, \underline{V} + \Delta I_P)}{\int_a^b ds (s)^P \mathcal{N}(\underline{z}; s\underline{y}, \underline{V} + \Delta I_P)} = \frac{\mathcal{I}(P+1, \nu, \delta, a, b)}{\mathcal{I}(P, \nu, \delta, a, b)}, \quad (123)$$

$$C^s(\underline{y}, \underline{\omega}, \underline{V}) = f_2^s(\underline{y}, \underline{\omega}, \underline{V}) = \frac{\mathcal{I}(P+2, \nu, \delta, a, b)}{\mathcal{I}(P, \nu, \delta, a, b)} - \hat{s}(\underline{y}, \underline{\omega}, \underline{V})^2, \quad (124)$$

where

$$\mathcal{I}(P, \nu, \delta, a, b) = \int_a^b ds s^P e^{-\frac{1}{2\delta}(s-\nu)^2}, \quad (125)$$

$$\delta = (\underline{y}^\top (\underline{V} + \Delta I_P)^{-1} \underline{y})^{-1}, \quad (126)$$

$$\delta^{-1} \nu = \underline{y}^\top (\underline{V} + \Delta I_P)^{-1} \underline{\omega} \quad (127)$$

and $\mathcal{I}(P, \nu, \delta, a, b)$ can be computed using gamma functions as explained in [1, 2].

Input functions For a Gauss-Bernoulli prior on the entries of \underline{X} , and assuming the AMP variances $\underline{\sigma}_i$ are diagonal matrices, the input update functions can be written component-wise with scalar arguments:

$$f_1^x(\lambda, \sigma) = \left(\rho \frac{\lambda}{(1+\sigma)^{3/2}} e^{-\frac{\lambda^2}{2(1+\sigma)}} \right) / \left(\rho \frac{e^{-\frac{\lambda^2}{2(1+\sigma)}}}{(1+\sigma)^{1/2}} + (1-\rho) \frac{e^{-\frac{\lambda^2}{2\sigma}}}{\sigma^{1/2}} \right), \quad (128)$$

and

$$f_2^x(\lambda, \sigma) = \left(\rho \frac{\sigma(1+\sigma) + (\lambda)^2}{(1+\sigma)^{5/2}} e^{-\frac{\lambda^2}{2(1+\sigma)}} \right) / \left(\rho \frac{e^{-\frac{\lambda^2}{2(1+\sigma)}}}{(1+\sigma)^{1/2}} + (1-\rho) \frac{e^{-\frac{\lambda^2}{2\sigma}}}{\sigma^{1/2}} \right) - f_{1,k}^x{}^2, \quad (129)$$

so that $\hat{x}_{i,k} = f_1^x(\lambda_{i,k}, \sigma_{i,kk})$ and $C_{i,kk}^x = f_2^x((\lambda_{i,k}, \sigma_{i,kk}))$.

4.2 Offline tests

The offline AMP algorithm is directly given by Algorithm 1 replacing the values of the output and input functions presented in the previous paragraph. In our numerical tests, we additionally impose that the covariance matrices \underline{V}_μ , $\partial_{\underline{\omega}} \underline{g}_{\text{out}}^\mu$ and $\underline{\sigma}_i$ are diagonal. This assumption lightens the numerics when considering a larger number of samples P .

Consequently, we first investigate how the algorithm behaves under a diagonal ansatz for the covariances. On Figure 2, we show that the performance of the algorithm with diagonal covariances is comparable to the performance of the algorithm with a full covariance matrix in terms of signal reconstruction. Furthermore, we find that the off-diagonal elements tend to vanish as we increase the size of the problem, and conjecture that they are null in the thermodynamic limit. In the following, we adopt the diagonal ansatz for covariances.

Under the assumption of diagonal covariance matrices, the State Evolution equations (91), (92) involve $P \times P$ matrices proportional to the identity (given the P examples are statistically equivalent). Therefore they can be parametrized with one scalar value, the unique diagonal element, noted respectively m and \hat{m} . While the update of m only requires a two dimensional integral that can be performed numerically, we resort to a Monte Carlo sampling for the update of \hat{m} . The procedure is described in Algorithm 2.

On the top row of Figure 4 and respectively of Figure 5, we report the reconstruction performance by the cal-AMP algorithm of the signal \underline{X}_0 and the calibration variable \underline{s}_0 , in the plane $\rho - \alpha$, for different values of P . These phase diagrams are similar to the ones reported in [1, 2] and will be compared with the online case described in the next Section. The performances are compared with the information theoretic threshold α_{min} corresponding to the minimal number of observations necessary to reconstruct the unknown signal and calibration variables,

$$M_{\text{min}} \times P = \rho N \times P + M_{\text{min}} \Rightarrow \alpha_{\text{min}} = \frac{M_{\text{min}}}{N} = \rho \frac{P}{P-1}, \quad (130)$$

and the sharp threshold of reconstruction of the fully calibrated Bayesian AMP α_{CS} [44].

On Figure 3a, we check numerically that the derived State Evolution predicts the behavior of the AMP algorithm for gain calibration. The MSEs of the two procedures are indeed consistent along the iterations of the algorithm. On Figure 3b, we also report an almost perfect agreement of the fixed points of SE and cal-AMP in terms of the MSE of the cal-AMP estimate of \underline{X}_0 as we vary the number of samples P .

4.3 Online tests

The online algorithms include supplementary operations to update the effective prior on the calibration variable from one step to the next. In this setting the recursion is

$$p_{s_\mu}^{(k+1)}(s_\mu) = \frac{1}{\mathcal{Z}_{\text{out}}^{(k+1)}} \int dz \mathcal{N}(\underline{y}_{(k),\mu}; z/s_\mu, \Delta) \mathcal{N}(z; w_{(k),\mu}, V_{(k),\mu}) p_{s_\mu}^{(k)}(s_\mu), \quad (131)$$

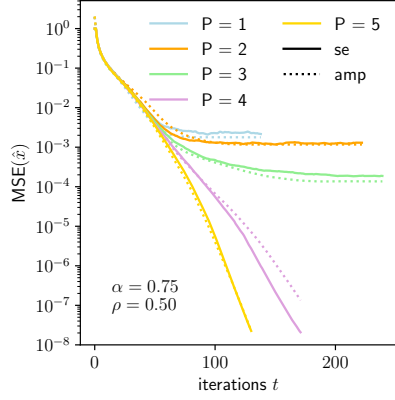
$$\propto s_\mu \exp \left\{ -\frac{(s_\mu - \nu_{(k),\mu})^2}{2\delta_{(k),\mu}} \right\} p_{s_\mu}^{(k)}(s_\mu) \quad (132)$$

so that

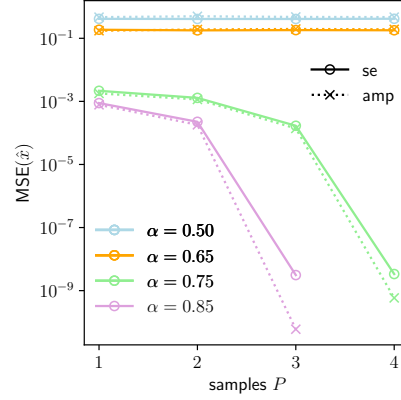
$$p_{s_\mu}^{(P)}(s_\mu) \propto s_\mu^P \exp \left\{ -\frac{(s - \Lambda_{(P),\mu})^2}{2\Sigma_{(P),\mu}} \right\} \quad \text{with} \quad \begin{cases} \Sigma_{(P),\mu}^{-1} = \sum_{k=1}^P \delta_{(k),\mu}^{-1} \\ \Lambda_{(P),\mu} = \Sigma_{(P),\mu} \left(\sum_{k=1}^P \delta_{(k),\mu}^{-1} \nu_{(k),\mu} \right), \end{cases} \quad (133)$$

yielding a posterior on the calibration variable at step P with an identical form to the posterior of the offline algorithm (albeit with parameters computed differently). Therefore \hat{s}_μ and C_μ^s can still be computed analytically following (123) and (124). We provide a pseudo-code for the online cal-AMP in Algorithm 4 and a pseudo-code for online SE in Algorithm 3.

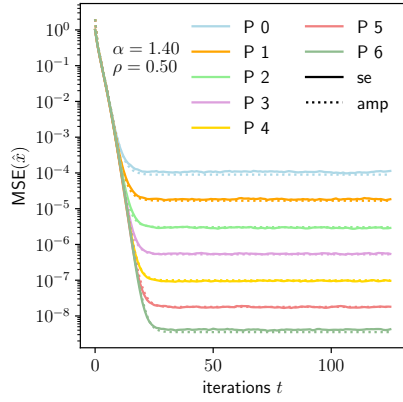
On the bottom rows of Figure 4 and 5 we plot phase diagrams obtained with online cal-AMP. Compared to the offline diagrams, we find that the reconstruction requires more samples to achieve comparable levels of accuracy. On Figure 3c and 3d we check the consistency of the cal-AMP and SE fixed points in terms of the MSE of cal-AMP in reconstructing \underline{X}_0 .



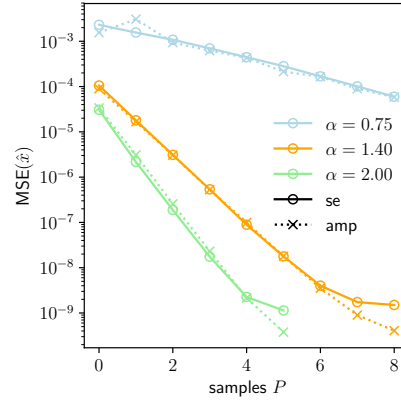
(a) Offline



(b) Offline



(c) Online



(d) Online

Figure 3: Comparison of mean squared errors in reconstructing the signal \underline{X}_0 using cal-AMP and predicted by SE for the gain calibration problem with calibration variables uniformly distributed in $[0.95, 1.05]$ and $N = 10^4$. Along the iterations of the algorithm and comparing fixed points we find a very good agreement of the two procedures. Nevertheless, the Monte Carlo integration in the SE does not allow for great numerical precision for low errors. We start seeing discrepancies between SE and AMP for errors below 10^{-8} .

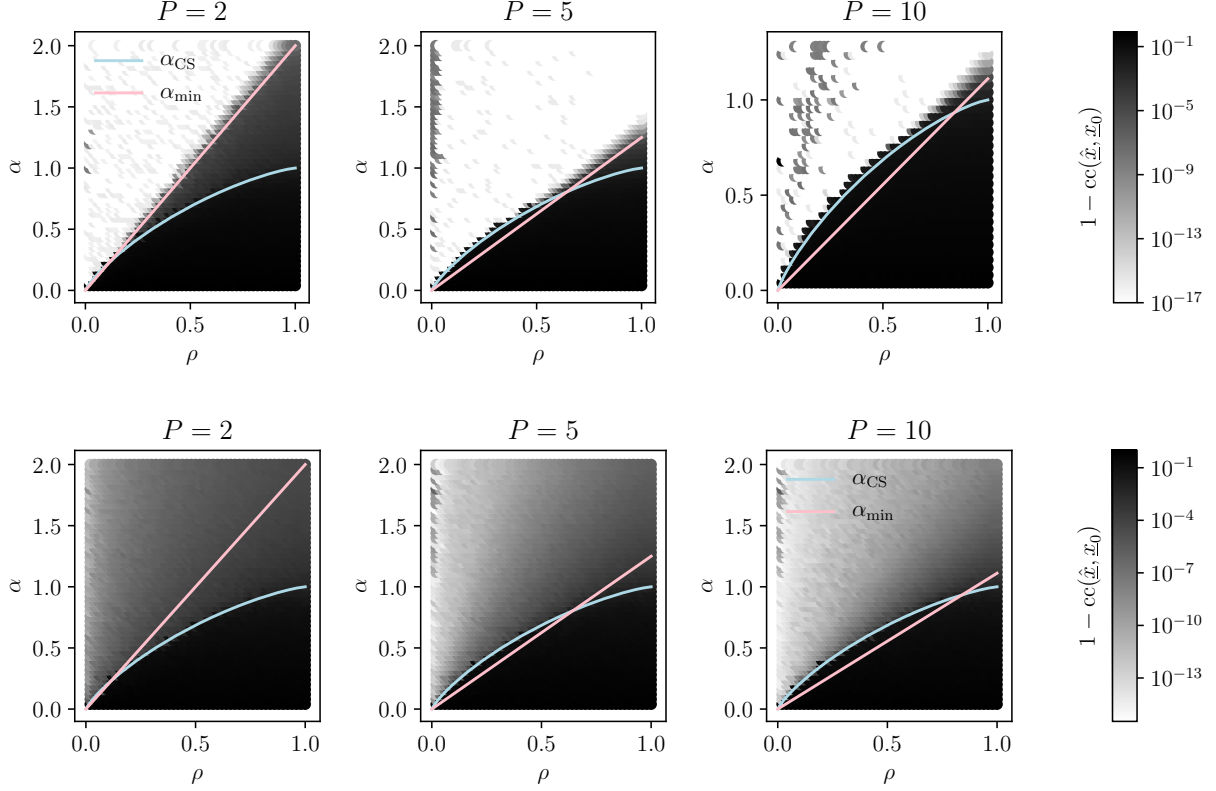


Figure 4: Normalized cross correlation between cal-AMP estimate $\hat{\underline{X}}$ and teacher signal \underline{X}_0 for the gain calibration problem with calibration variables uniformly distributed in $[0.95, 1.05]$ and $N = 10^3$. Diagrams are plotted as a function of the measurement rate $\alpha = M/N$ and the sparsity level ρ for the offline (**top row**) and online (**bottom row**) algorithms, for increasing number of available samples P . The blue line α_{CS} is the phase transition threshold for a perfectly calibrated channel. The pink line α_{min} marks the strict lower bound on the number of measurements necessary for reconstruction. The online cal-AMP requires more samples than the offline version to achieve comparable errors, nevertheless above the transition relatively low errors are already reached at $P = 10$. We note that the algorithms are sometimes unstable at low ρ , leading to unexpectedly high MSEs (top left corner of top right diagram).

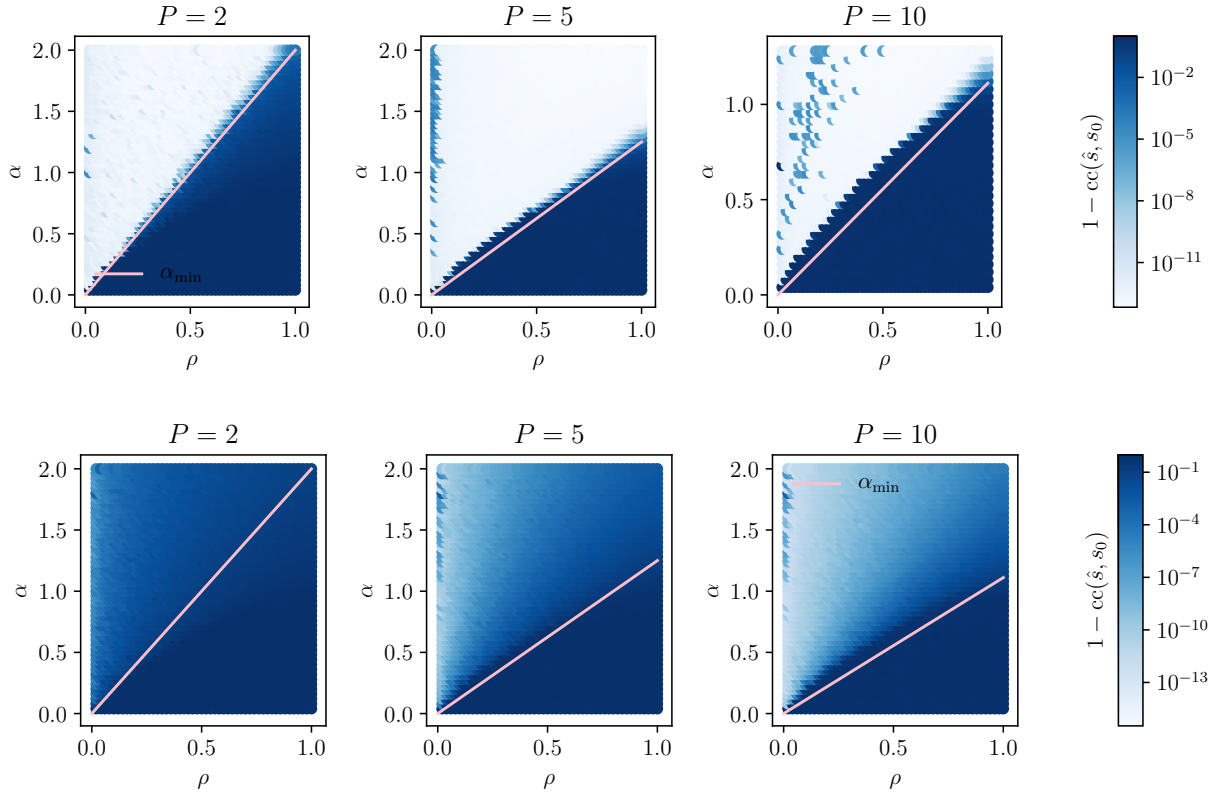


Figure 5: Normalized cross correlation between cal-AMP estimate \hat{s} and teacher signal s_0 for the gain calibration problem with calibration variables uniformly distributed in $[0.95, 1.05]$ and $N = 10^3$. Diagrams are plotted as a function of the measurement rate $\alpha = M/N$ and the sparsity level ρ for the offline (**top row**) and online (**bottom row**) algorithms, for increasing number of available samples P . The pink line α_{\min} marks the strict lower bound on the number of measurements necessary for reconstruction. Results are similar to the diagrams in terms of errors on the signal \hat{X} of Figure 4.

Algorithm 4 Online Gain Calibration Approximate Message Passing

Input: matrix $\underline{Y} \in \mathbb{R}^{M \times P}$ and matrix $\underline{W} \in \mathbb{R}^{M \times N}$:

Initialize: $\Lambda_{0,\mu} = 0, \Sigma_{0,\mu} = 0,$

for $k = 1, \dots, P$: **do**

Initialize: $\hat{x}_{i,k}, C_{i,k}^x \quad \forall i$ and $g_{\text{out}\mu,k}, \partial_\omega g_{\text{out}\mu,k} \quad \forall \mu$

repeat

1) Estimate mean and variance of $\underline{z}_{(k)}$ given current $\hat{\underline{x}}_{(k)}$

$$V_{\mu,k}^{(t)} = \sum_{i=1}^N W_{\mu i}^2 C_{i,k}^{x(t)} \quad (134)$$

$$\omega_{\mu,k}^{(t)} = \sum_{i=1}^N W_{\mu i} \hat{x}_{i,k}^{(t)} - \sum_{i=1}^N W_{\mu i}^2 C_{i,k}^{x(t)} g_{\text{out}\mu,k}^{(t-1)} \quad (135)$$

2) Exploit current $\underline{y}_{(k)}$ and inherited $\Lambda_{k-1,\mu}, \Sigma_{k-1,\mu}$

2.1) Update recursion parameters given $\underline{y}_{(k)}$

$$\delta_{\mu,k}^{(t)-1} = y_{\mu,k}^2 / (V_{\mu,k}^{(t)} + \Delta) \quad (136)$$

$$\nu_{\mu,k}^{(t)} = \delta_{\mu,k} y_{\mu,k} \omega_{\mu,k}^{(t)} / (V_{\mu,k}^{(t)} + \Delta) \quad (137)$$

$$\Sigma_{k,\mu}^{(t)} = \Sigma_{k-1,\mu} + \delta_{\mu,k}^{(t)} \quad (138)$$

$$\Lambda_{k,\mu}^{(t)} = \Sigma_{k,\mu}^{(t)} \left(\Sigma_{k-1,\mu} \Lambda_{k-1,\mu} + \delta_{\mu,k}^{(t)} \nu_{\mu,k}^{(t)} \right) \quad (139)$$

2.2) Update estimates \hat{s} and C^s

$$\hat{s}_\mu^{(t)} = \frac{\mathcal{I}(k+1, \Lambda_{k,\mu}, \Sigma_{k,\mu}, a, b)}{\mathcal{I}(k, \Lambda_{k,\mu}, \Sigma_{k,\mu}, a, b)} \quad (140)$$

$$C_\mu^{s(t)} = \frac{\mathcal{I}(k+2, \Lambda_{k,\mu}, \Sigma_{k,\mu}, a, b)}{\mathcal{I}(k, \Lambda_{k,\mu}, \Sigma_{k,\mu}, a, b)} \quad (141)$$

2.3) Update $\underline{g}_{\text{out}}$ and $\partial_\omega \underline{g}_{\text{out}}$

$$\partial_\omega g_{\text{out}\mu,k}^{(t)} = C_\mu^{s(t)} y_{\mu,k}^2 / (V_{\mu,k}^{(t)} + \Delta)^2 - (V_{\mu,k}^{(t)} + \Delta)^{-1} \quad (142)$$

$$g_{\text{out}\mu,k}^{(t)} = (\hat{s}_\mu^{(t)} y_{\mu,k} - \omega_{\mu,k}^{(t)}) / (V_{\mu,k}^{(t)} + \Delta), \quad (143)$$

3) Estimate mean and variance of \underline{x} given current optimal \underline{z}

$$\sigma_{i,k}^{(t)} = \left(- \sum_{\mu=1}^M W_{\mu i}^2 \partial_\omega g_{\text{out}\mu,k}^{(t)} \right)^{-1} \quad (144)$$

$$\lambda_{i,k}^{(t)} = \hat{x}_{i,k}^{(t)} + \sigma_{i,k}^{(t)} \left(\sum_{\mu=1}^M W_{\mu i} g_{\text{out}\mu,k}^{(t)} \right) \quad (145)$$

4) Estimate of mean and variance of \underline{x} augmented of the information about the prior

$$C_{i,k}^{x(t+1)} = f_2^x(\lambda_{i,k}^{(t)}, \sigma_{i,k}^{(t)}) \quad (146)$$

$$\hat{x}_{i,k}^{(t+1)} = f_1^x(\lambda_{i,k}^{(t)}, \sigma_{i,k}^{(t)}) \quad (147)$$

$t = t + 1$

until convergence

end for

Output: Estimates and variances $\hat{\underline{x}}_i, C_i^x, \hat{s}_\mu, C_\mu^s$

5 Conclusion

In this paper, we presented a theoretical characterization of the cal-AMP algorithm for blind calibration. The derivation of the corresponding State Evolution was presented from the message passing equations of a vectorial version of the GAMP algorithm. Furthermore, we extended the algorithm and its analysis to the online scenario, which allows to adapt the calibration upon receiving new observations. The validity of the newly proposed algorithms and analysis were demonstrated through numerical experiments on the case of gain calibration. In the offline setting, cal-AMP achieves near-perfect reconstruction close to the theoretical lower bound. In the online setting, cal-AMP is also able to calibrate and reconstruct from a small number of samples.

For practical applications, the extension to the online case is of particular interest. It enables continuous improvement of hardware sensing channel, without requiring to keep past observations in memory. A natural extension will also be to consider the G-VAMP [20, 21] version of cal-AMP, which will enable a generalization of the algorithm to measurement matrices with non-i.i.d. entries. From the theoretical point of view, we note that the introduced SE analysis in the non-Bayes optimal setting should make it possible to characterize other reconstruction algorithms, as algorithms proposed in [7, 8]. Also, it will be interesting to confirm whether cal-AMP can also succeed in a multi-layer setting, likewise the recent extension of GAMP to ML-AMP [45]. This generalization will open the way to message passing algorithms solutions in more complex calibration problems.

Acknowledgements We wish to thank Christophe Schülke for his dedication to the calibration problem. We would like to thank the Kavli Institute For Theoretical Physics for its hospitality for an extended stay, during which parts of this work were conceived and carried out. This work is supported by the French Agence Nationale de la Recherche under grant ANR17-CE23-0023-01 PAIL and the European Union’s Horizon 2020 Research and Innovation Program 714608-SMiLe. This material is based upon work supported by Google Cloud. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. Additional funding is acknowledged by MG from ‘Chaire de recherche sur les modèles et sciences des données’, Fondation CFM pour la Recherche-ENS.

References

- [1] Christophe Schülke, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Blind Calibration in Compressed Sensing using Message Passing Algorithms. *Advances in Neural Information Processing Systems 26*, pages 1–9, 2013.
- [2] Christophe Schülke, Francesco Caltagirone, and Lenka Zdeborová. Blind sensor calibration using approximate message passing. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(11):P11013, nov 2015.
- [3] Emmanuel J. Candès and Terence Tao. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, dec 2006.
- [4] Michael Lustig, David L. Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, dec 2007.
- [5] Ricardo Otazo, Daniel Kim, Leon Axel, and Daniel K. Sodickson. Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI. *Magnetic Resonance in Medicine*, 64(3):767–776, 2010.
- [6] J. Romberg, M.F. F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, K.F. Kelly, R.G. Baraniuk, Ting Sun, K.F. Kelly, and R.G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, 2008.

- [7] Rémi Gribonval, Gilles Chardon, and Laurent Daudet. Blind calibration for compressed sensing by convex optimization. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2713–2716. IEEE, mar 2012.
- [8] Hao Shen, Martin Kleinstueber, Cagdas Bilen, and Rémi Gribonval. A conjugate gradient algorithm for blind sensor calibration in sparse recovery. In *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–5. IEEE, sep 2013.
- [9] David L. Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, nov 2009.
- [10] Yoshiyuki Kabashima and Shinsuke Uda. A BP-based algorithm for performing Bayesian inference in large perceptron-type networks, 2004.
- [11] Sundeep Rangan. Generalized Approximate Message Passing for Estimation with Random Linear Mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, jul 2011.
- [12] Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization Methods for Large-Scale Machine Learning. *SIAM Review*, 60(2):223–311, 2018.
- [13] Chuang Wang and Yue M. Lu. Online learning for sparse PCA in high dimensions: Exact dynamics and phase transitions. *2016 IEEE Information Theory Workshop, ITW 2016*, (3):186–190, 2016.
- [14] Ioannis Mitliagkas, Constantine Caramanis, and Prateek Jain. Memory Limited, Streaming PCA. In *Neural Information Processing Systems 2013*, 2013.
- [15] Chuang Wang, Jonathan Mattingly, and Yue M. Lu. Scaling Limit: Exact and Tractable Analysis of Online Learning Algorithms with Applications to Regularized Regression and PCA. *arXiv preprint*, 1712.04332, 2017.
- [16] Benjamin Aubin, Antoine Maillard, Jean Barbier, Florent Krzakala, Nicolas Macris, and Lenka Zdeborová. The committee machine: Computational to statistical gaps in learning a two-layers neural network. In *Neural Information Processing Systems 2018*, number NeurIPS, pages 1–44, 2018.
- [17] Andre Manoel, Florent Krzakala, Eric W. Tramel, and Lenka Zdeborová. Streaming Bayesian inference: Theoretical limits and mini-batch approximate message-passing. *55th Annual Allerton Conference on Communication, Control, and Computing, Allerton 2017*, 2018-Janua(i):1048–1055, 2018.
- [18] Timothy L.H. Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65(2):499–556, 1993.
- [19] Marylou Gabrié. Mean-field inference methods for neural networks. *Journal of Physics A: Mathematical and Theoretical*, 1911.00890, mar 2020.
- [20] Sundeep Rangan, Philip Schniter, and Alyson K. Fletcher. Vector approximate message passing. In *2017 IEEE International Symposium on Information Theory (ISIT)*, volume 1, pages 1588–1592. IEEE, jun 2017.
- [21] Philip Schniter, Sundeep Rangan, and Alyson K. Fletcher. Vector approximate message passing for the generalized linear model. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1525–1529. IEEE, nov 2016.
- [22] Manfred Opper and Ole Winther. Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling. *Physical Review E*, 64(5):056131, oct 2001.
- [23] Manfred Opper and Ole Winther. Tractable approximations for probabilistic models: The adaptive Thouless-Anderson-Palmer mean field approach. *Physical Review Letters*, 86(17):3695–3699, 2001.

- [24] Yoshiyuki Kabashima. Inference from correlated patterns: A unified theory for perceptron learning and linear vector channels. *Journal of Physics: Conference Series*, 95(1):012001, jan 2008.
- [25] Takashi Shinzato and Yoshiyuki Kabashima. Learning from correlated patterns by simple perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 42(1):015005, jan 2009.
- [26] Yoshiyuki Kabashima and Mikko Vehkaperä. Signal recovery using expectation consistent approximation for linear observations. *IEEE International Symposium on Information Theory - Proceedings*, pages 226–230, 2014.
- [27] Thomas P Minka. A family of algorithms for approximate Bayesian inference. *PhD Thesis*, 2001.
- [28] Manfred Opper and Ole Winther. Expectation consistent free energies for approximate inference. *Advances in Neural Information Processing Systems*, 17:1001–1008, 2005.
- [29] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, sep 2016.
- [30] Manfred Opper and David Haussler. Calculation of the learning curve of Bayes optimal classification algorithm for learning a perceptron with noise. In *COLT '91 Proceedings of the fourth annual workshop on Computational learning theory*, pages Pages 75–87, 1991.
- [31] Yukito Iba. The Nishimori line and Bayesian statistics. *Journal of Physics A: Mathematical and General*, 32(21):3875–3888, may 1999.
- [32] Hidetoshi Nishimori. *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Clarendon Press, 2001.
- [33] Yoshiyuki Kabashima, Florent Krzakala, Marc Mézard, Ayaka Sakata, and Lenka Zdeborová. Phase Transitions and Sample Complexity in Bayes-Optimal Matrix Factorization. *IEEE Transactions on Information Theory*, 62(7):4228–4265, 2016.
- [34] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin Glass Theory and Beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. WORLD SCIENTIFIC, nov 1986.
- [35] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.
- [36] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference. *Probability Theory and Related Fields*, 174(3-4):1133–1185, aug 2019.
- [37] Manfred Opper and Ole Winther. A Bayesian approach to on-line learning. In *On-line learning in neural networks*, pages 363–378. Cambridge University Press, 1999.
- [38] O Kinouchi and N Caticha. Optimal generalization in perceptions. *Journal of Physics A: Mathematical and General*, 25(23):6243–6250, dec 1992.
- [39] Michael Biehl and Peter Riegler. On-Line Learning with a Perceptron. *Europhysics Letters (EPL)*, 28(7):525–530, dec 1994.
- [40] Sara Solla and Ole Winther. Optimal Perceptron Learning: an On-line Bayesian Approach. In *On-Line Learning in Neural Networks*, pages 379–398. Cambridge, 1999.
- [41] David Saad. *On-Line Learning in Neural Networks*. Cambridge University Press, jan 1999.
- [42] Paulo V. Rossi, Yoshiyuki Kabashima, and Jun Ichi Inoue. Bayesian online compressed sensing. *Physical Review E*, 94(2):022137, aug 2016.

- [43] Tamara Broderick, Nicholas Boyd, Andre Wibisono, Ashia C. Wilson, and Michael I. Jordan. Streaming Variational Bayes. *Neural Information Processing Systems 2013*, pages 1–9, 2013.
- [44] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *J. Stat. Mech*, 2012.
- [45] Andre Manoel, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Multi-layer generalized linear estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2098–2102. IEEE, jun 2017.