Tutorial: Approximate Message Passing & Replicas for committee machines

Benjamin Aubin[†]

† *Institut de Physique Théorique*
CNRS & CEA & Université Paris-Saclay, Saclay, France

Contents

1 Introduction

The purpose of these notes is to provide a short introduction to **Approximate Message Passing (AMP) algorithms** and **replicas** computation, that we illustrate for **committee machines**, that are a simple generalization of **generalized linear models**. For more details, see in particular the references [1, 2, 3, 4, 5, 6, 7].

1.1 Model

We revisit the so-called *teacher-student* scenario from statistical physics [8, 9, 10, 11, 12, 13, 14]. We assume that the synthetic dataset is generated by a *teacher* model.

1.1.1 Data

We assume that the matrix of data inputs $\mathbf{X} \in \mathbb{R}^{n\times d}$, that contains n d-dimensional samples, is drawn i.i.d with distribution $\mathrm{P_x}(\mathbf{X}) = \prod_{i,\mu=1}^{d,n} \mathrm{P_x}(x_{\mu i})$.

For simplicity, we will consider them to be i.i.d Gaussian with zero mean and unit variance: $\forall \mu \in [\![n]\!], \mathbf{x}_\mu \sim \mathcal{N}_{\mathbf{x}}(\mathbf{0}, \mathbf{I}_d)$.

1.1.2 Teacher - Planted solution

Moreover, we draw a **planted solution** $\mathbf{W}^\star$ from a separable distribution (across the input dimension) $\mathrm{P_{w^\star}}(\mathbf{W}^\star) = \prod_{i=1}^{d} \mathrm{P_{w^\star}}(\mathbf{w}_i^\star)$. Therefore the weights could possibly be correlated across the second dimension.

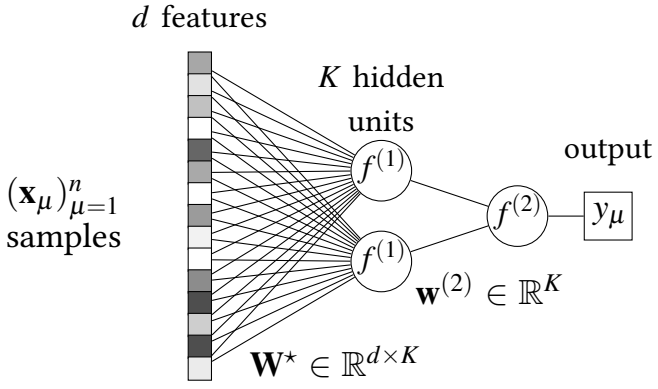

Figure 1 – Illustration of the *committee machine*: it is one of the simplest models belonging to the considered model class (1), and on which we focus to illustrate our results. It is a two-layers neural network with sign activation functions $f^{(1)}, f^{(2)} = \mathrm{sign}$ and weights $\mathbf{w}^{(2)}$ fixed to unity. It is represented for $K = 2$.

1.1.3 Channel

The labels are then generated by multiplying the data matrix $\mathbf{X}$ and the planted weights $\mathbf{W}^\star$ and passing their product to a **channel** φ_{out}, acting component-wise, according to:

$$\mathbf{y} = \varphi_{\text{out}}\left(\left\{\frac{1}{\sqrt{d}}\mathbf{X}\mathbf{w}_k^\star\right\}_{k=1}^{K}\right) \quad \text{or} \quad \mathbf{y} \sim \mathrm{P}_{\text{out}}\left(\cdot\,\middle|\,\left\{\frac{1}{\sqrt{d}}\mathbf{X}\mathbf{w}_k^\star\right\}_{k=1}^{K}\right), \tag{1}$$

and illustrated in Fig. 1 in the case of a committee machine with fixed weights $\mathbf{w}^{(2)}$ in the second layer.

The function $\varphi_{\text{out}} : \mathbb{R}^K \mapsto \mathbb{R}$ represents a deterministic, or stochastic function associated to a probability distribution P_{out}, applied component-wise to each sample.

Remark 1.1. *Notice that the factor $\frac{1}{\sqrt{d}}$ is present to insure that the variance of the input data is normalized to the unit.*

1.1.4 Student

The committee machine estimation problem consists of trying to fit the n input-output observations $\{\mathbf{X}, \mathbf{y}\} \in \mathbb{R}^{n\times d} \times \mathbb{R}^n$.

In the case the **student** has exactly the same architecture, namely $\mathbf{W}^\star \in \mathbb{R}^{d\times K}$ and $\mathbf{W} \in \mathbb{R}^{d\times K}$ have the same dimensionality and prior information $\mathrm{P}_{\mathrm{w}^\star} = \mathrm{P}_{\mathrm{w}}$, this setting is called the **Bayes-optimal setting**. Otherwise, it is called the **mismatched setting**.

In these notes, we focus on the **Bayes-optimal setting**, so that the committee machine estimation problem for the student committee machine (within the same hypothesis class and with weights $\mathbf{W} \in \mathbb{R}^{d\times K}$) reduces to learn the teacher rule generated by the ground truth weights $\mathbf{W}^\star$ from the supervised dataset $\{\mathbf{X}, \mathbf{y}\}$.

Remark 1.2. *Committee machines are a simple vectorized generalization of **GLM**, whose estimation is performed simultaneously with $K \geq 1$.*

1.2 Notations

- $\mathbf{X} \sim \mathrm{P}_{\mathrm{x}}(\mathbf{X}) = \prod_{i=1,\mu=1}^{d,n} \mathrm{P}_{\mathrm{x}}(x_{\mu i})$
- $\mathbf{W} \in \mathbb{R}^{d\times K}$ is the matrix of weights of the first layer with prior distribution: $\mathrm{P}_{\mathrm{w}}(\mathbf{W}) = \prod_{i=1}^{d} \mathrm{P}_{\mathrm{w}}(\mathbf{w}_i)$
- $\mathbf{y} \in \mathbb{R}^n$ is a set of n scalar observations generated according to (1).

- Indices $\mu \in [\![n]\!]$ and $i \in [\![d]\!]$ correspond respectively to data samples and input dimensions.
- $\alpha \equiv \frac{n}{d}$ with $d, n \to \infty$ and in the limit where $\alpha = \Theta(1)$
- $\boldsymbol{\xi}$ will denote later a Gaussian vectorial variable such that $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$

2 Bayesian inference and factor graph

Both **Minimum Mean-Squared Error (MMSE)** and **Maximum A Posteriori (MAP)** estimations boil down to the analysis of the posterior distribution $\mathrm{P}(\mathbf{W}|\mathbf{y}, \mathbf{X})$ expressed by the Bayes rule

$$\mathrm{P}(\mathbf{W}|\mathbf{y}, \mathbf{X}) = \frac{\mathbb{P}(\mathbf{y}|\mathbf{W}, \mathbf{X})\, \mathbb{P}(\mathbf{W})}{\mathrm{P}(\mathbf{y}, \mathbf{X})} = \frac{\mathrm{P}_{\text{out}}(\mathbf{y}|\mathbf{W}, \mathbf{X})\, \mathrm{P}_{\text{w}}(\mathbf{W})}{\mathscr{Z}_d(\{\mathbf{y}, \mathbf{X}\})}\,. \tag{2}$$

The joint distribution is also called the *partition function* $\mathrm{P}(\mathbf{y}, \mathbf{X}) \equiv \mathscr{Z}_d(\{\mathbf{y}, \mathbf{X}\})$. To connect with the statistical physics formalism, we introduce the corresponding Hamiltonian, for separable distributions $\mathrm{P}_{\text{out}}, \mathrm{P}_{\text{w}}$ along one dimension, by

$$\begin{aligned} \mathscr{H}_d(\mathbf{W}, \{\mathbf{y}, \mathbf{X}\}) &= -\log \mathrm{P}_{\text{out}}(\mathbf{y}|\mathbf{W}, \mathbf{X}) - \log \mathrm{P}_{\text{w}}(\mathbf{W})\,, \\ &= -\sum_{\mu=1}^{n} \log \mathrm{P}_{\text{out}}\left(y_\mu|\mathbf{W}, \mathbf{x}_\mu\right) - \sum_{i=1}^{d} \mathrm{P}_{\text{w}}(\mathbf{w}_i)\,. \end{aligned}$$

The spin variables represent the weights of the model $\mathbf{W} \in \mathbb{R}^{d\times K}$ and they interact through the random dataset $\{\mathbf{y}, \mathbf{X}\}$ that plays the role of the quenched exchange interactions. However here, the interactions are *fully connected*, meaning that each variable $\mathbf{w}_i \in \mathbb{R}^K$ is connected to every other spin $\{\mathbf{w}_j\}_{j\in\partial j\setminus i}$ as represented in the factor graph in Fig. 2.

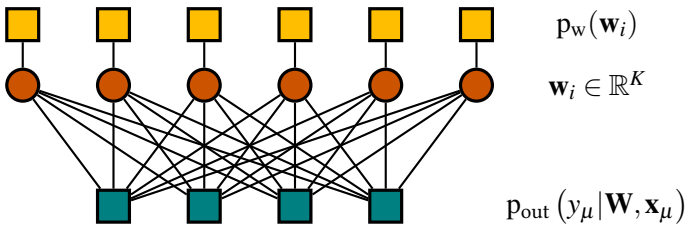

Figure 2 – Factor graph corresponding to the committee machines hypothesis class. The vectorial variables to infer $\mathbf{w}_i$ are fully connected through the quenched disorder $\mathbf{y} \sim \mathrm{P}_{\text{out}^\star}(.)$ and each variable follow a one-body interaction with a separable prior distribution $\mathrm{p_w}(\mathbf{w}_i)$.

The partition function at inverse temperature β is therefore defined by

$$\begin{aligned}\mathscr{Z}_d\left(\{\mathbf{y},\mathbf{X}\};\beta\right) \equiv \mathrm{P}\left(\mathbf{y},\mathbf{X}\right) &= \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}\; e^{-\beta\mathscr{H}_d(\mathbf{W},\{\mathbf{y},\mathbf{X}\})} \\ &= \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{w}\; e^{\beta\left(\log \mathrm{p}_{\mathrm{out}}(\mathbf{y}|\mathbf{W},\mathbf{X})+\log P_{\mathrm{w}}(\mathbf{W})\right)} \\ &= \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{w}\; \mathrm{p}_{\mathrm{out}}\left(\mathbf{y}|\mathbf{W},\mathbf{X}\right)\mathrm{p}_{\mathrm{w}}\left(\mathbf{W}\right)\,,\end{aligned} \tag{3}$$

and can be exactly mapped to Bayesian estimation for $\beta = 1$. In the context of ERM, MAP estimation can be analyzed by taking the limit $\beta \to \infty$.

In the considered modern high-dimensional regime with $d \to \infty$, $n \to \infty$, $\alpha = n/d = \Theta(1)$ and $K = \Theta(1)$, we are interested in computing the *free entropy* Φ *averaged* over the input data $\mathbf{X}$ and teacher weights $\mathbf{W}^\star$, or equivalently over the output labels $\mathbf{y}$ generated from it, defined as

$$\Phi(\alpha) \equiv \lim_{d\to\infty} \frac{1}{d}\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\log \mathscr{Z}_d\left(\mathbf{y},\mathbf{X}\right)\right]\,. \tag{4}$$

Remark 2.1.

- *The distribution* $\mathrm{P}\left(\mathbf{w}|\mathbf{y};\mathbf{X}\right)$ *is often intractable in the limit* $d \to \infty$
- *Moreover it is still hard to sample efficiently in the high-dimensional regime*

To circumvent this issue:

- We can focus only on marginals $\mathrm{P}(\mathbf{w}_i;\mathbf{y},\mathbf{X})$ which can been estimated with **Belief Propagation (BP)** equations and **Approximate Message Passing (AMP) algorithms**, as illustrated in Sec. 4.
- We may also try to compute directly the free entropy: we present the replica method in Sec. 3 that allows to compute the above average over the random dataset $\{\mathbf{y},\mathbf{X}\}$, that plays the role of the quenched disorder in usual spin glasses. We show the computation for the more involved committee machine model class and generalization of the GLM class, only for iid data.
- We try to give an idea how these two methods relate and we show they are complementary and consistent.

Remark 2.2. *The cumbersome computation for non iid data can be performed as well and lead to more complex expressions and has been performed in particular in [15] in the case of the GLM.*

3 Replica computation

3.1 Heuristic derivation

In this section, we present the heuristic derivation of the replica formula in the context of of the *committee machine*. This computation is necessary to properly guess the formula that can then be proven using the adaptive interpolation method. The reader interested in the replica approach to neural networks and the committee machine is invited to look as well to some of the classical papers [16, 17, 18, 19, 20, 21].

We present here the replica computation of the averaged free entropy $\Phi(\alpha)$ in eq. (4) for arbitrary *student* prior and channel distributions $\mathrm{P_w}, \mathrm{P_{w^\star}}$ and $\mathrm{P_{out}}, \mathrm{P_{out^\star}}$, so that the computation remains valid for both the Bayes-optimal and mismatched settings.

3.1.1 Replica trick

The average in eq. (4) is intractable in general, and the computation relies on the so called *replica trick*, that consists in applying the identity

$$\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\lim_{d\to\infty}\frac{1}{d}\log\mathscr{Z}_d(\mathbf{y},\mathbf{X})\right] = \lim_{r\to 0}\left[\lim_{d\to\infty}\frac{1}{d}\frac{\partial\log\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathscr{Z}_d(\mathbf{y},\mathbf{X})^r\right]}{\partial r}\right]. \tag{5}$$

The replica trick has been used in a series of previous works to compute the free energy density of GLM for separable distributions [22] and has been rigorous proven in this case by [23].

Eq. (5) is interesting in the sense that it reduces the intractable average to the computation of the moments of the averaged partition function, which are easier quantities to compute. Note that for $r \in \mathbb{N}$, $\mathscr{Z}_d(\mathbf{y},\mathbf{X})^r = \prod_{a=1}^r \mathscr{Z}_d(\mathbf{y},\mathbf{X})$ represents the partition function of r identical non-interacting copies of the initial system, called *replicas*. Taking the quenched average over the disorder will then correlate the replicas, before taking the number of replicas $r \to 0$. Therefore, we assume there exists an analytical continuation so that $r \in \mathbb{R}$ and the limit is well defined. Finally, notice that we exchanged the order of the limits $r \to 0$ and $d \to \infty$. These technicalities are crucial points but are not rigorously justified and we will ignore them in the rest of the computation.

First, in order to decouple the contributions of the channel $\mathrm{P_{out}}$ and the prior $\mathrm{P_w}$, we introduce the variable $\mathbf{Z} = \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}$ and a Dirac-delta integral:

$$\mathscr{Z}_d(\mathbf{y},\mathbf{X}) = \int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{z}\, \mathrm{p_{out}}(\mathbf{y}|\mathbf{Z}) \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{w}\, \mathrm{p_w}(\mathbf{W})\, \delta\left(\mathbf{Z} - \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}\right). \tag{6}$$

Thus the replicated partition function for an integer $r \in \mathbb{N}$ in eq. (5) can be written as

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathscr{Z}_d(\mathbf{y},\mathbf{X})^r\right] \\
&= \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\prod_{a=1}^{r}\int_{\mathbb{R}^n} \mathrm{d}\mathbf{Z}^a\, \mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a)\right. \qquad (7)\\
&\qquad \left.\times \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^a\, \mathrm{p}_{\mathrm{w}^a}(\mathbf{W}^a)\,\delta\left(\mathbf{Z}^a - \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}^a\right)\right] \\
&= \mathbb{E}_{\mathbf{X}}\int_{\mathbb{R}^n}\mathrm{d}\mathbf{y}\int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{Z}^\star\, \mathrm{p}_{\mathrm{out}^\star}(\mathbf{y}|\mathbf{Z}^\star) \qquad (8)\\
&\qquad \times \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^\star\, \mathrm{p}_{\mathrm{w}^\star}(\mathbf{W}^\star)\,\delta\left(\mathbf{Z}^\star - \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}^\star\right) \\
&\times \left[\prod_{a=1}^{r}\int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{Z}^a\, \mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a)\int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^a\, \mathrm{p}_{\mathrm{w}^a}(\mathbf{W}^a)\,\delta\left(\mathbf{Z}^a - \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}^a\right)\right] \\
&= \int_{\mathbb{R}^n}\mathrm{d}\mathbf{y}\prod_{a=0}^{r}\int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{Z}^a\, \mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a)\int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^a\, \mathrm{p}_{\mathrm{w}^a}(\mathbf{W}^a) \qquad (9)\\
&\qquad \times \underbrace{\mathbb{E}_{\mathbf{X}}\prod_{a=0}^{r}\delta\left(\mathbf{Z}^a - \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{W}^a\right)}_{(I)}.
\end{aligned}
$$

Note that the average over $\mathbf{y}$ is equivalent to the one over the ground truth vector $\mathbf{W}^\star$ in the case of a *teacher-student*, which can be conveniently grouped with the other terms by just extending the replica indices and considering it as a new replica $\mathbf{W}^0$ with index $a = 0$, leading to a total of $r+1$ replicas.

3.1.2 Average over the i.i.d input data X

Remains to compute the average over $\mathbf{X}$ in the term (I). We suppose that inputs are drawn from an iid distribution, for example a Gaussian $\mathrm{P}_{\mathrm{x}}(\mathbf{x}) = \mathscr{N}_{\mathbf{x}}(\mathbf{0}, \mathbf{I}_d)$. More precisely, for $(i,j) \in [\![d]\!]^2$, $(\mu,\nu) \in [\![n]\!]^2$, $\mathbb{E}_{\mathbf{X}}\left[x_{\mu i}x_{\nu j}\right] = \delta_{\mu\nu}\delta_{ij}$.
By definition, the average in (I) defines the probability density $\mathrm{p}_{z^a}(\mathbf{Z}^a)$ and as $\forall k \in [\![K]\!], \forall \mu \in [\![n]\!]$, $z^a_{\mu k} = \frac{1}{\sqrt{d}}\sum_{i=1}^{d} x_{\mu i}w^a_{ik}$ is the sum of iid random variables, the CLT insures that in the thermodynamic limit $d \to \infty$, $z^a_{\mu k}$ follows a Gaussian multivariate distribution, with first moments given by:

$$
\begin{aligned}
\mathbb{E}_{\mathbf{X}}[z^a_{\mu k}] &= \frac{1}{\sqrt{d}}\sum_{i=1}^{d}\mathbb{E}_{\mathbf{X}}\left[x_{\mu i}\right]w^a_{ik} = 0 \qquad (10)\\
\mathbb{E}_{\mathbf{X}}[z^a_{\mu k}z^b_{\nu k'}] &= \frac{1}{d}\sum_{ij}\mathbb{E}_{\mathbf{X}}\left[x_{\mu i}x_{\mu j}\right]w^a_{ik}w^b_{jk'} = \frac{1}{d}\sum_{ij}\delta_{ij}w^a_{ik}w^b_{jk'}\delta_{\mu\nu} \equiv \delta_{\mu\nu}\,Q^{ak}_{bk'}.
\end{aligned}
$$

Notice that averaging over the quenched disorder introduced correlations between replicas, which were initially independent, described by the symmetric

overlap matrix $\{Q^{ak}_{bk'}\}_{kk'}$ of size $(r+1)K \times (r+1)K$. This matrix order parameter measures the correlations between the replicated matrices $\{\mathbf{W}^a\}_{a=0}^r$ and is formally defined by

$$\mathbf{Q}(\{\mathbf{W}^a\}_{a=0}^r) \equiv \left(\frac{1}{d} \sum_{i=1}^{d} w^a_{ik} w^b_{ik'} \right)_{k,k'=1..K}^{a,b=0..r},$$

such that $\forall (a,b) \in [\![0:r]\!]^2$, $\mathbf{Q}^{ab} \in \mathbb{R}^{K \times K}$. Therefore, again by the CLT, in the limit $d \to \infty$, the hidden variable $\mathbf{Z}^a \in \mathbb{R}^{n \times K}$ converges in distribution to the multivariate distribution

$$\mathrm{P}_{z^a}(\mathbf{Z}^a|\mathbf{Q}) = \exp\left[-\frac{1}{2} \sum_{\mu=1}^{n} \sum_{a,b=0}^{r} \sum_{k,k'=1}^{K} z^a_{\mu k} z^b_{\mu k'} (\mathbf{Q}^{-1})^{ab}_{kk'} \right] / \left(\det(2\pi\mathbf{Q}) \right)^{\frac{n}{2}}. \quad (11)$$

Inserting this back in the replicated partition function finally writes

$$\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathscr{Z}_d(\mathbf{y},\mathbf{X})^r \right] = \int_{\mathbb{R}^n} \mathrm{d}\mathbf{y} \prod_{a=0}^{r} \int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{Z}^a \, \mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a) \, \mathrm{p}_{z^a}(\mathbf{Z}^a|\mathbf{Q}) \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^a \, \mathrm{p}_{w^a}(\mathbf{W}^a) \quad (12)$$

3.1.3 Fourier representation

Next we introduce the change of variable for the new order parameter $\mathbf{Q}^{ab}$ with a Dirac-δ distribution and its Fourier representation. For a variable $x \in \mathbb{R}$, the distribution $\delta(x)$ can be written as an integral over a purely imaginary parameter $\hat{x}$:

$$\delta(x) = \frac{1}{2i\pi} \int_{i\mathbb{R}} \mathrm{d}\hat{x} e^{-\hat{x}x}.$$

Applying the above identity to the change of variable, we obtain

$$\begin{aligned} 1 &= \int_{\mathbb{R}^{(K\times r+1)^2}} \mathrm{d}\mathbf{Q} \prod_{0\le a\le b\le r, 1\le k,k'\le K} \delta\left(dQ^{ab}_{kk'} - \sum_{i=1}^{d} w^a_{ik} w^b_{ik'} \right) \\ &\propto \int\int_{\mathbb{R}^{(K\times r+1)^2}} \mathrm{d}\mathbf{Q}\mathrm{d}\hat{\mathbf{Q}} \, \exp\left(-d \sum_{a=0}^{r} \sum_{k,k'}^{K} Q^{aa}_{kk'} \hat{Q}^{aa}_{kk'} - \frac{d}{2} \sum_{a\neq b}^{r} \sum_{k,k'}^{K} Q^{ab}_{kk'} \hat{Q}^{ab}_{kk'} \right) \quad (13) \\ &\quad \times \exp\left(\sum_{a=0}^{r} \sum_{k,k'}^{K} \hat{Q}^{aa}_{kk'} w^a_k w^a_{k'} + \frac{1}{2} \sum_{a\neq b}^{r} \sum_{k,k'}^{K} \hat{Q}^{ab}_{kk'} w^a_k w^b_{k'} \right), \end{aligned}$$

that involves a new ad-hoc purely imaginary matrix parameter $\hat{\mathbf{Q}} \in i\mathbb{R}^{(K\times(r+1))^2}$. Finally, multiplying the replicated partition function by 1, using the Cauchy theorem and rotating the integration, it becomes an integral over the symmetric

matrices $\mathbf{Q} \in \mathbb{R}^{(K\times r+1)^2}$ and $\hat{\mathbf{Q}} \in \mathbb{R}^{(K\times r+1)^2}$

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\mathscr{Z}_d(\mathbf{y},\mathbf{X})^r\right] \\
&= \int\int_{\mathbb{R}^{(K\times r+1)^2}} \mathrm{d}\mathbf{Q}\mathrm{d}\hat{\mathbf{Q}}\, \exp\left(-d\sum_{a=0}^{r}\sum_{k,k'}^{K} Q^{aa}_{kk'}\hat{Q}^{aa}_{kk'} - \frac{d}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} Q^{ab}_{kk'}\hat{Q}^{ab}_{kk'}\right) \\
&\qquad\qquad \times \exp\left(\sum_{a=0}^{r}\sum_{k,k'} \hat{Q}^{aa}_{kk'} w^a_k w^a_{k'} + \frac{1}{2}\sum_{a\neq b}\sum_{k,k'} \hat{Q}^{ab}_{kk'} w^a_k w^b_{k'}\right) \\
&\int_{\mathbb{R}^n} \mathrm{d}\mathbf{y} \prod_{a=0}^{r} \int_{\mathbb{R}^{n\times K}} \mathrm{d}\mathbf{Z}^a\, \mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a)\, \mathrm{p}_{Z^a}(\mathbf{Z}^a|\mathbf{Q}) \int_{\mathbb{R}^{d\times K}} \mathrm{d}\mathbf{W}^a\, \mathrm{p}_{W^a}(\mathbf{W}^a) \\
&= \int\int_{\mathbb{R}^{(K\times r+1)^2}} \mathrm{d}\mathbf{Q}\mathrm{d}\hat{\mathbf{Q}}\, \exp\left(-d\sum_{a=0}^{r}\sum_{k,k'}^{K} Q^{aa}_{kk'}\hat{Q}^{aa}_{kk'} - \frac{d}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} Q^{ab}_{kk'}\hat{Q}^{ab}_{kk'}\right) \\
&\qquad\qquad \times \exp\left(\sum_{a=0}^{r}\sum_{k,k'}^{K} \hat{Q}^{aa}_{kk'} w^a_k w^a_{k'} + \frac{1}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} \hat{Q}^{ab}_{kk'} w^a_k w^b_{k'}\right) \\
&\left[\int_{\mathbb{R}} \mathrm{d}y \prod_{a=0}^{r} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^a\, \mathrm{p}_{\mathrm{out}^a}(y|\mathbf{z}^a)\, \mathrm{p}_{Z^a}(\mathbf{z}^a|\mathbf{Q})\right]^n \left[\prod_{a=0}^{r} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}^a\, \mathrm{p}_{W^a}(\mathbf{w}^a)\right]^d \\
&\simeq \int\int_{\mathbb{R}^{(K\times r+1)^2}} \mathrm{d}\mathbf{Q}\mathrm{d}\hat{\mathbf{Q}}\, e^{d\Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}})},
\end{aligned}
$$

where in the last step, we used a Laplace method [24] and omitted the sub-leading factors in the thermodynamic limit $d \to \infty$ to evaluate it as a function of the free entropy potential defined by

$$
\begin{aligned}
\Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}}) &\equiv -\sum_{a=0}^{r}\sum_{k,k'}^{K} Q^{aa}_{kk'}\hat{Q}^{aa}_{kk'} - \frac{1}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} Q^{ab}_{kk'}\hat{Q}^{ab}_{kk'} \\
&\quad + \log \Psi^{(r)}_{\mathrm{w}}(\hat{\mathbf{Q}}) + \alpha \log \Psi^{(r)}_{\mathrm{out}}(\mathbf{Q}), \\
\Psi^{(r)}_{\mathrm{w}}(\hat{\mathbf{Q}}) &\equiv \prod_{a=0}^{r} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}^a\, \mathrm{p}_{W^a}(\mathbf{w}^a) \\
&\qquad \times \exp\left(\sum_{a=0}^{r}\sum_{kk'}^{K} \hat{Q}^{aa}_{kk'} w^a_k w^a_{k'} + \frac{1}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} \hat{Q}^{ab}_{kk'} w^a_k w^b_{k'}\right), \\
\Psi^{(r)}_{\mathrm{out}}(\mathbf{Q}) &\equiv \prod_{a=0}^{r} \int_{\mathbb{R}} \mathrm{d}y \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^a\, \mathrm{p}_{\mathrm{out}^a}(y|\mathbf{z}^a)\, \mathrm{p}_{Z^a}(\mathbf{z}^a|\mathbf{Q}),
\end{aligned}
\tag{14}
$$

and where we decoupled the variable $\mathbf{Z}^a \in \mathbb{R}^{n\times K}$ and $\mathbf{W}^a \in \mathbb{R}^{d\times K}$ along the rows

$$
\begin{aligned}
\mathrm{p}_{\mathrm{out}^a}(\mathbf{y}|\mathbf{Z}^a) &= \prod_{\mu=1}^{n} \mathrm{p}_{\mathrm{out}^a}\left(y_\mu|\mathbf{z}_\mu^a\right), \text{ with } \mathbf{z}_\mu^a \in \mathbb{R}^K, \\
\mathrm{p}_{z^a}(\mathbf{Z}^a|\mathbf{Q}) &= \prod_{\mu=1}^{n} \mathrm{p}(\mathbf{z}_\mu^a|\mathbf{Q}), \\
\mathrm{p}_{\mathrm{w}^a}(\mathbf{W}^a) &= \prod_{i=1}^{d} \mathrm{p}_{\mathrm{w}^a}(\mathbf{w}_i^a), \text{ with } \mathbf{w}_i^a \in \mathbb{R}^K, \\
\mathrm{p}_{z^a}(\mathbf{z}^a|\mathbf{Q}) &= \exp\left[-\frac{1}{2}\sum_{a,b=0}^{r}\sum_{k,k'=1}^{K} z_k^a z_{k'}^b (\mathbf{Q}^{-1})_{kk'}^{ab}\right] / \left(\det(2\pi\mathbf{Q})\right)^{\frac{1}{2}}.
\end{aligned}
$$

Note that the averaged replicated partition function of this fully connected model can be expressed as a saddle point equation only because distributions $\mathrm{P}_{\mathrm{out}}, \mathrm{P}_{\mathrm{out}^\star}$ and $\mathrm{P}_{\mathrm{w}}, \mathrm{P}_{\mathrm{w}^\star}$ are separable so that a pre-factor scaling with the system size d dominates the exponential distribution. Finally, switching the two limits $r \to 0$ and $d \to \infty$, the quenched free entropy Φ simplifies as a saddle point equation

$$
\Phi(\alpha) = \mathbf{extr}_{\mathbf{Q},\hat{\mathbf{Q}}} \left\{ \lim_{r\to 0} \frac{\partial \Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}})}{\partial r} \right\}, \tag{15}
$$

over symmetric matrices $\mathbf{Q} \in \mathbb{R}^{(K\times r+1)^2}$ and $\hat{\mathbf{Q}} \in \mathbb{R}^{(K\times r+1)^2}$.

To summarize, we managed to get rid of the original high-dimensional integrals and replace them by an optimization in the space of matrices, which, in this form, is still intractable. We not only have to search in the space of $(r+1)\times(r+1)$ matrices to find the extremiser of $\Phi^{(r)}$, but we also need to compute the limit $r \to 0^+$. In the following we will assume a simple Ansatz for these matrices in order to first obtain an analytic expression in r before taking the derivative with respect to r.

3.2 Replica Symmetric free entropy

Our goal is to express the functional $\Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}})$ appearing in the free entropy as an analytical function of r, in order to perform the replica trick.

3.2.1 Replica symmetric ansatz

To do so, we will assume that the extremum of $\Phi^{(r)}$ is attained at a point in $\mathbf{Q},\hat{\mathbf{Q}}$ space such that a *replica symmetry* property is verified. More concretely, we

assume:

$$\begin{aligned}
\exists \mathbf{Q} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall a \in [\![0:r]\!] \quad \forall (k,k') \in [\![K]\!]^2 \quad Q^{aa}_{kk'} = Q_{kk'}\,, \\
\exists \mathbf{Q}^\star \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall (k,k') \in [\![K]\!]^2 \quad Q^{00}_{kk'} = Q^\star_{kk'}\,, \\
\exists \mathbf{q} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall (a<b) \in [\![0:r]\!]^2 \quad \forall (k,k') \in [\![K]\!]^2 \quad Q^{ab}_{kk'} = q_{kk'}\,, \\
\exists \mathbf{m} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall a \in [\![0:r]\!] \quad \forall (k,k') \in [\![K]\!]^2 \quad Q^{0a}_{kk'} = m_{kk'}\,,
\end{aligned} \tag{16}$$

and similarly for the ad-hoc parameter

$$\begin{aligned}
\exists \hat{\mathbf{Q}} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall a \in [\![0:r]\!] \quad \forall (k,k') \in [\![K]\!]^2 \quad \hat{Q}^{aa}_{kk'} = -\frac{1}{2}\hat{Q}_{kk'}\,, \\
\exists \hat{\mathbf{Q}}^\star \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall (k,k') \in [\![K]\!]^2 \quad \hat{Q}^{00}_{kk'} = \hat{Q}^\star_{kk'}\,, \\
\exists \hat{\mathbf{q}} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall (a<b) \in [\![0:r]\!]^2 \quad \forall (k,k') \in [\![K]\!]^2 \quad \hat{Q}^{ab}_{kk'} = \hat{q}_{kk'}\,, \\
\exists \hat{\mathbf{m}} \in \mathbb{R}^{K\times K} \text{ s.t} \quad \forall a \in [\![0:r]\!] \quad \forall (k,k') \in [\![K]\!]^2 \quad \hat{Q}^{0a}_{kk'} = \hat{m}_{kk'}\,.
\end{aligned} \tag{17}$$

The factor $-\frac{1}{2}$ is not necessary bu useful to recover commonly used formulations. This Ansatz can be represented by symmetric RS matrices $\mathbf{Q}^{(\mathrm{rs})} \in \mathbb{R}^{(K\times r+1)^2}$ and $\hat{\mathbf{Q}}^{(\mathrm{rs})} \in \mathbb{R}^{(K\times r+1)^2}$

$$\mathbf{Q}^{(\mathrm{rs})} = \begin{pmatrix} \mathbf{Q}^\star & \mathbf{m} & \cdots & \mathbf{m} \\ \mathbf{m}^\intercal & \mathbf{Q} & \mathbf{q} & \cdots \\ \vdots & \mathbf{q} & \ddots & \mathbf{q} \\ \mathbf{m}^\intercal & \cdots & \mathbf{q} & \mathbf{Q} \end{pmatrix} \text{ and } \hat{\mathbf{Q}}^{(\mathrm{rs})} = \begin{pmatrix} \hat{\mathbf{Q}}^\star & \hat{\mathbf{m}} & \cdots & \hat{\mathbf{m}} \\ \hat{\mathbf{m}}^\intercal & -\frac{1}{2}\hat{\mathbf{Q}} & \hat{\mathbf{q}} & \cdots \\ \vdots & \hat{\mathbf{q}} & \ddots & \hat{\mathbf{q}} \\ \hat{\mathbf{m}}^\intercal & \cdots & \hat{\mathbf{q}} & -\frac{1}{2}\hat{\mathbf{Q}} \end{pmatrix}, \tag{18}$$

where the *overlap* parameters may be reinterpreted as the scalar product between the replicas

$$\forall (a,b) \in [\![r]\!]^2,\ \mathbf{q} = \frac{1}{d}\mathbf{W}^{a\intercal}\mathbf{W}^b\,,$$

the self-overlap of each replica

$$\forall a \in [\![r]\!],\ \mathbf{Q} = \frac{1}{d}\mathbf{W}^{a\intercal}\mathbf{W}^a\,,$$

the scalar product with the ground truth

$$\forall a \in [\![r]\!],\ \mathbf{m} = \frac{1}{d}\mathbf{W}^{\star\intercal}\mathbf{W}^a\,,$$

and the second moment of the ground truth distribution

$$\mathbf{Q}^\star = \frac{1}{d}\mathbf{W}^{\star\intercal}\mathbf{W}^\star\,.$$

The above Ansatz simplifies in the scalar GLM case with $K=1$ to $q = \frac{1}{d}\mathbf{w}^a\cdot\mathbf{w}^b$ for $a\neq b$, a norm $Q = \frac{1}{d}\|\mathbf{w}^a\|_2^2$, an overlap with the ground truth $m = \frac{1}{d}\mathbf{w}^a\cdot\mathbf{w}^\star$ and a second moment $Q^\star = \frac{1}{d}\|\mathbf{w}^\star\|_2^2$.

3.2.2 Decomposition of the free entropy functional

Let's compute separately the terms involved in the functional $\Phi^{(r)}(\mathbf{Q}, \hat{\mathbf{Q}})$ in (14) by applying this Ansatz: the first is a trace term, the second term $\Psi_{\mathrm{w}}^{(r)}$ depends on the prior distributions $\mathrm{P_w}$, $\mathrm{P_{w^\star}}$ and finally the third term $\Psi_{\mathrm{out}}^{(r)}$ depends on the channel distributions $\mathrm{P_{out^\star}}$, $\mathrm{P_{out}}$.

Trace term The trace term in (14) can be easily computed at the RS fixed point and takes the following form

$$\begin{aligned} &-\sum_{a=0}^{r}\sum_{k,k'}^{K} Q_{kk'}^{aa}\hat{Q}_{kk'}^{aa} - \frac{1}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} Q_{kk'}^{ab}\hat{Q}_{kk'}^{ab}\Big|_{\mathrm{rs}} \\ &= -\mathrm{Tr}\left(\mathbf{Q}^\star\hat{\mathbf{Q}}^\star\right) + \frac{1}{2}r\,\mathrm{Tr}\left(\mathbf{Q}\hat{\mathbf{Q}}\right) - r\,\mathrm{Tr}\left(\mathbf{m}\hat{\mathbf{m}}\right) - \frac{r(r-1)}{2}\mathrm{Tr}\left(\mathbf{q}\hat{\mathbf{q}}\right), \end{aligned} \tag{19}$$

and taking the derivative and the limit $r \to 0$ we obtain

$$\begin{aligned} &\lim_{r\to 0}\partial_r\left(-\sum_{a=0}^{r}\sum_{k,k'}^{K} Q_{kk'}^{aa}\hat{Q}_{kk'}^{aa} - \frac{1}{2}\sum_{a\neq b}^{r}\sum_{k,k'}^{K} Q_{kk'}^{ab}\hat{Q}_{kk'}^{ab}\right)\Big|_{\mathrm{rs}} \\ &\qquad = \frac{1}{2}\mathrm{Tr}\left(\mathbf{Q}\hat{\mathbf{Q}}\right) - \mathrm{Tr}\left(\mathbf{m}\hat{\mathbf{m}}\right) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{q}\hat{\mathbf{q}}\right) \end{aligned} \tag{20}$$

Prior integral $\Psi_{\mathrm{w}}^{(r)}$ Evaluated at the RS fixed point the quadratic form reads

$$\begin{aligned} &\sum_{a=0}^{r}\sum_{kk'}\hat{Q}_{kk'}^{aa}w_k^a w_{k'}^a + \frac{1}{2}\sum_{a\neq b}\sum_{k,k'}\hat{Q}_{kk'}^{ab}w_k^a w_{k'}^b \\ &= \mathbf{w}^{\star\intercal}\hat{\mathbf{Q}}^\star\mathbf{w}^\star + \sum_{a=1}^{r}\mathbf{w}^{\star\intercal}\hat{\mathbf{m}}\mathbf{w}^a - \frac{1}{2}\sum_{a=1}^{r}\mathbf{w}^{a\intercal}\hat{\mathbf{Q}}\mathbf{w}^a + \frac{1}{2}\sum_{1\leq a\neq b\leq r}\mathbf{w}^{a\intercal}\hat{\mathbf{q}}\mathbf{w}^b \\ &= \mathbf{w}^{\star\intercal}\hat{\mathbf{Q}}^\star\mathbf{w}^\star + \sum_{a=1}^{r}\mathbf{w}^{\star\intercal}\hat{\mathbf{m}}\mathbf{w}^a - \frac{1}{2}\sum_{a=1}^{r}\mathbf{w}^{a\intercal}\left(\hat{\mathbf{Q}}+\hat{\mathbf{q}}\right)\mathbf{w}^a + \frac{1}{2}\left(\sum_{a=1}^{r}\mathbf{w}^a\right)^{\intercal}\hat{\mathbf{q}}\left(\sum_{a=1}^{r}\mathbf{w}^a\right). \end{aligned}$$

Using a Hubbard-Stratonovich transformation:

Proposition 3.1 (Hubbard-Stratonovich transformation)**.** *For $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and a symmetric positive definite matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$, for all $\mathbf{x} \in \mathbb{R}^d$*

$$\mathbb{E}_{\boldsymbol{\xi}}\exp\left(\boldsymbol{\xi}^\intercal\mathbf{A}^{1/2}\mathbf{x}\right) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\mathrm{d}\boldsymbol{\xi}\, e^{-\frac{1}{2}\mathbf{x}^\intercal\mathbf{x}+\boldsymbol{\xi}^\intercal\mathbf{A}^{1/2}\mathbf{x}} = e^{\frac{1}{2}\mathbf{x}^\intercal\mathbf{A}\mathbf{x}}, \tag{21}$$

the prior integral can be further simplified

$$\begin{aligned} &\Psi_{\mathrm{w}}^{(r)}(\hat{\mathbf{Q}})\Big|_{\mathrm{rs}} = \int_{\mathbb{R}^{(r+1)\times K}}\mathrm{d}\mathbf{W}\;\mathrm{p_w}(\mathbf{W})\, e^{\sum_{a=0}^{r}\sum_{k,k'}^{K}\hat{Q}_{kk'}^{aa}w_k^a w_{k'}^a + \frac{1}{2}\sum_{a\neq b}\sum_{k,k'}^{K}\hat{Q}_{kk'}^{ab}w_k^a w_{k'}^b} \\ &= \mathbb{E}_{\boldsymbol{\xi},\mathbf{w}^\star\sim\mathrm{P_{w^\star}}}\left[e^{\mathbf{w}^{\star\intercal}\hat{\mathbf{Q}}^\star\mathbf{w}^\star}\mathbb{E}_{\mathbf{w}\sim\mathrm{P_w}}\left[e^{\left(\mathbf{w}^\intercal\hat{\mathbf{m}}\mathbf{w}^\star - \frac{1}{2}\mathbf{w}^\intercal(\hat{\mathbf{Q}}+\hat{\mathbf{q}})\mathbf{w} + \mathbf{w}^\intercal\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}\right)}\right]^r\right]. \end{aligned} \tag{22}$$

Detailed computation of $\left(\mathbf{Q}^{(\mathrm{rs})}\right)^{-1}$ Let us focus on the matrix $\mathbf{Q}^{(\mathrm{rs})}$ involved in the expression of $\Psi_{\mathrm{out}}^{(r)}$ in (14). Given the replica symmetric ansatz, we must compute the determinant and inverse matrix of the matrix $\mathbf{\Sigma} \equiv \mathbf{Q}^{(\mathrm{rs})}$.

We use tensor products notations $\otimes$ for matrices living in $\mathbb{R}^{r\times r} \otimes \mathbb{R}^{K\times K}$. Let us recall some basic properties of the tensor product of two matrices $\mathbf{A} \in \mathbb{R}^{r\times r}$ and $\mathbf{B} \in \mathbb{R}^{K\times K}$:

$$(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1} \tag{23}$$

$$\det(\mathbf{A}\otimes\mathbf{B}) = \det(\mathbf{A})^K \det(\mathbf{B})^r \tag{24}$$

$$\forall \mathbf{C},\mathbf{D} \quad (\mathbf{C}\otimes\mathbf{D})(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{C}\mathbf{A})\otimes(\mathbf{D}\mathbf{B}) \tag{25}$$

$$\mathrm{Sp}(\mathbf{A}\otimes\mathbf{B}) = \{\lambda_i\mu_j \quad / \lambda_i \in \mathrm{Sp}(\mathbf{A}), \mu_j \in \mathrm{Sp}(\mathbf{B})\} \tag{26}$$

Next we denote the *reduced* matrix

$$\mathbf{\Sigma}_r \equiv \begin{pmatrix} \mathbf{Q} & \mathbf{q} & \cdots & \mathbf{q} \\ \mathbf{q} & \mathbf{Q} & \cdots & \mathbf{q} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q} & \mathbf{q} & \cdots & \mathbf{Q} \end{pmatrix} = \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q}) + \mathbf{J}_r \otimes \mathbf{q}, \tag{27}$$

where $\mathbf{J}_r$ is the rank-1 matrix with all elements equal to 1. Finally, some basic linear algebra formulas allow us to reduce the problem of computing $\det(\mathbf{\Sigma})$ and $\mathbf{\Sigma}^{-1}$ to those of $\mathbf{\Sigma}_r$.

Calculation of S We define the $K \times K$ matrix :

$$\mathbf{S} \equiv \mathbf{Q}^\star - (\mathbf{1}_r^\intercal \otimes \mathbf{m})\mathbf{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal) \tag{28}$$

Let us compute $\mathbf{S}$:

$$\begin{aligned}
\mathbf{S} &\equiv \mathbf{Q}^\star - (\mathbf{1}_r^\intercal \otimes \mathbf{m})\mathbf{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal) \\
&= \mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal + r^2\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal \\
&= \mathbf{Q}^\star - r\mathbf{m}\left(\mathbf{I}_K - n(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{q}\right)(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal \\
&= \mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\left(\mathbf{Q}+(r-1)\mathbf{q}-r\mathbf{q}\right)(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal \\
&= \mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal
\end{aligned}$$

Calculation of $\mathbf{\Sigma}_r^{-1}$ $\mathbf{\Sigma}_r$ can be easily analyzed, thanks to a slight generalization of the Sherman-Morrison formula.

$$\mathbf{\Sigma}_r = \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q}) + \mathbf{J}_r \otimes \mathbf{q} = \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q}) + \mathbf{1}_r\mathbf{1}_r^\intercal \otimes \mathbf{q} \tag{29}$$

So that the determinant reads

$$\begin{aligned}
\det(\mathbf{\Sigma}_r) &= \det(\mathbf{Q}-\mathbf{q})^r \det\left(\mathbf{I}_K + (\mathbf{1}^\intercal \otimes \mathbf{q})(\mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q}))^{-1}(\mathbf{1}\otimes\mathbf{I}_K)\right) \\
&= \det(\mathbf{Q}-\mathbf{q})^r \det\left(\mathbf{I}_K + r\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1}\right) \\
\det(\mathbf{\Sigma}_r) &= \det(\mathbf{Q}-\mathbf{q})^{r-1} \det(\mathbf{Q}+(r-1)\mathbf{q}), \qquad (30)
\end{aligned}$$

and the inverse

$$\begin{aligned}\boldsymbol{\Sigma}_r^{-1} &= \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q})^{-1} &(31)\\ &- \left(\mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q})^{-1}\right)(\mathbf{1}_r \otimes \mathbf{I}_K)\left(\mathbf{I}_K + r\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1}\right)^{-1}(\mathbf{1}_r^\intercal \otimes \mathbf{q})\left(\mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q})^{-1}\right)\\ &= \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q})^{-1} - \mathbf{J}_r \otimes \left((\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1}\right) &(32)\end{aligned}$$

Determinant of $\mathbf{Q}^{(\mathrm{rs})}$ Therefore we obtain:

$$\begin{aligned}\det(\boldsymbol{\Sigma}) &= \det\left(\mathbf{Q}^{(\mathrm{rs})}\right) = \det(\boldsymbol{\Sigma}_r)\det(\mathbf{S})\\ &= \det(\mathbf{Q}-\mathbf{q})^{r-1}\det(\mathbf{Q}+(r-1)\mathbf{q})\det\left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)\end{aligned} \tag{33}$$

Inverse of $\mathbf{Q}^{(\mathrm{rs})}$ The inverse of the block matrix reads:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{S}^{-1} & -\mathbf{S}^{-1}(\mathbf{1}_r^\intercal \otimes \mathbf{m})\boldsymbol{\Sigma}_r^{-1} \\ -\boldsymbol{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal)\mathbf{S}^{-1} & \boldsymbol{\Sigma}_r^{-1} + \boldsymbol{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal)\mathbf{S}^{-1}(\mathbf{1}_r^\intercal \otimes \mathbf{m})\boldsymbol{\Sigma}_r^{-1} \end{pmatrix}$$

where each block can be computed separately. The top diagonal simply reads

$$\mathbf{S}^{-1} = \left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}$$

and the off-diagonal blocks read:

$$\begin{aligned}&-\boldsymbol{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal)\mathbf{S}^{-1} = &(34)\\ &\quad -\mathbf{1} \otimes \left(\mathbf{Q}+(r-1)\mathbf{q}\right)^{-1}\mathbf{m}^\intercal\left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}\Big)\\ &-\mathbf{S}^{-1}(\mathbf{1}_r^\intercal \otimes \mathbf{m})\boldsymbol{\Sigma}_r^{-1} = &(35)\\ &\quad -\mathbf{1}_r^\intercal \otimes \Big(\left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\Big)]\end{aligned}$$

The bottom diagonal block is a bit more involved :

$$\begin{aligned}&\boldsymbol{\Sigma}_r^{-1} + \boldsymbol{\Sigma}_r^{-1}(\mathbf{1}_r \otimes \mathbf{m}^\intercal)\mathbf{S}^{-1}(\mathbf{1}_r^\intercal \otimes \mathbf{m})\boldsymbol{\Sigma}_r^{-1}\\ &\quad = \mathbf{I}_r \otimes (\mathbf{Q}-\mathbf{q})^{-1} - \mathbf{J}_r \otimes \mathbf{A} &(36)\end{aligned}$$

with

$$\begin{aligned}\mathbf{A} \equiv\ &(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1} &(37)\\ &-(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\end{aligned}$$

Finally we observe that $\boldsymbol{\Sigma}^{-1}$ is also a block matrix of the form

$$\boldsymbol{\Sigma}^{-1} = \left(\mathbf{Q}^{(\mathrm{rs})}\right)^{-1} = \begin{bmatrix} \tilde{\mathbf{Q}}^\star & \tilde{\mathbf{m}} & \cdots & \tilde{\mathbf{m}} \\ \tilde{\mathbf{m}}^\intercal & \tilde{\mathbf{Q}} & \tilde{\mathbf{q}} & \cdots \\ \vdots & \tilde{\mathbf{q}} & \ddots & \tilde{\mathbf{q}} \\ \tilde{\mathbf{m}}^\intercal & \cdots & \tilde{\mathbf{q}} & \tilde{\mathbf{Q}} \end{bmatrix} \tag{38}$$

with

$$
\begin{aligned}
\tilde{\mathbf{Q}}^\star &= \left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1} \\
\tilde{\mathbf{m}} &= -\left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1} \\
\tilde{\mathbf{Q}} &= (\mathbf{Q}-\mathbf{q})^{-1} - (\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{q}(\mathbf{Q}-\mathbf{q})^{-1} \\
&\quad + (\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal \\
&\qquad \times \left(\mathbf{Q}^\star - r\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1}\mathbf{m}^\intercal\right)^{-1}\mathbf{m}(\mathbf{Q}+(r-1)\mathbf{q})^{-1} \\
\tilde{\mathbf{q}} &= \tilde{\mathbf{Q}} - (\mathbf{Q}-\mathbf{q})^{-1}
\end{aligned}
$$

Channel integral $\Psi_{\text{out}}^{(r)}$ The quadratic form in $\mathrm{p}_{z^a}(\mathbf{z}^a|\mathbf{Q}^{(\mathrm{rs})}$ reads

$$
\begin{aligned}
&-\frac{1}{2}\sum_{a,b}\sum_{k,k'} z_k^a z_{k'}^b (\mathbf{Q}^{-1})_{kk'}^{ab} \\
&= -\frac{1}{2}\mathbf{z}^{\star\intercal}\tilde{\mathbf{Q}}^\star\mathbf{z} - \sum_{a=1}^r \mathbf{z}^{\star\intercal}\tilde{\mathbf{m}}\mathbf{z}^a \\
&\qquad -\frac{1}{2}\sum_{a=1}^r \mathbf{z}^{a\intercal}\left(\tilde{\mathbf{Q}}-\tilde{\mathbf{q}}\right)\mathbf{z}^a - \frac{1}{2}\left(\sum_a^r \mathbf{z}^a\right)^\intercal \tilde{\mathbf{q}}\left(\sum_a^r \mathbf{z}^a\right),
\end{aligned}
$$

and using another Gaussian transformation, see Prop. 3.1, we finally obtain

$$
\begin{aligned}
&\Psi_{\text{out}}^{(r)}(\mathbf{Q})\Big|_{\text{rs}} = \int \mathrm{d}y \int_{\mathbb{R}^{(r+1)\times K}} \mathrm{d}\mathbf{Z}\ \mathrm{p}_{\text{out}}(y|\mathbf{Z})\,\mathrm{p}(\mathbf{Z}|\mathbf{Q}) \\
&= \int \mathrm{d}y \int_{\mathbb{R}^{(r+1)\times K}} \mathrm{d}\mathbf{Z}\mathrm{p}_{\text{out}}(y|\mathbf{Z})\, e^{-\frac{1}{2}\sum_{a,b=0}^r \sum_{k,k'=1}^K z_k^a z_{k'}^b (\mathbf{Q}^{-1})_{kk'}^{ab}} / \left(\det(2\pi\mathbf{Q})^{(\mathrm{rs})}\right)^{\frac{1}{2}} \\
&= \int \mathrm{d}y\ \mathbb{E}_{\boldsymbol{\xi}} \mathrm{e}^{-\frac{1}{2}\log\left(\det\left(2\pi\mathbf{Q}^{(\mathrm{rs})}\right)\right)} \times \int \mathrm{d}\mathbf{z}^\star\ \mathrm{p}_{\text{out}^\star}(y|\mathbf{z}^\star)\, e^{-\frac{1}{2}\mathbf{z}^{\star\intercal}\tilde{\mathbf{Q}}^\star\mathbf{z}^\star} \\
&\times \left[\int \mathrm{d}\mathbf{z}\ \mathrm{p}_{\text{out}}(y|\mathbf{z}) \exp\left(-\mathbf{z}^{\star\intercal}\tilde{\mathbf{m}}\mathbf{z} - \frac{1}{2}\mathbf{z}^\intercal\left(\tilde{\mathbf{Q}}-\tilde{\mathbf{q}}\right)\mathbf{z} + \mathbf{z}^\intercal(-\tilde{\mathbf{q}})^{1/2}\boldsymbol{\xi}\right)\right]^r,
\end{aligned}
\tag{39}
$$

with $\det\left(\mathbf{Q}^{(\mathrm{rs})}\right)$ given by (33).

3.2.3 Consistency conditions $r \to 0$: $\Theta(1)$ terms

It remains to take the limit $r \to 0^+$ of the expressions for $\Psi_{\text{w}}^{(r)}$ and $\Psi_{\text{out}}^{(r)}$ that are now analytical in r. First, our assumptions must be consistent and thus we need to check the consistency conditions in the limit $r \to 0$. Indeed, if $\Phi^{(r)}$ is finite we could obtain divergence taking the limit $\lim_{r\to 0}\frac{1}{r}\Phi^{(r)} = \infty$. Therefore to avoid such divergence, we must at least impose that $\lim_{r\to 0}\Phi^{(r)} = 0$:

$$
\lim_{r\to 0}\Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}}) = -\mathrm{Tr}\left(\mathbf{Q}^\star\hat{\mathbf{Q}}^\star\right) + \log\Psi_{\text{w}}^0(\hat{\mathbf{Q}}^\star) + \alpha\log\Psi_{\text{out}}^0(\mathbf{Q}^\star)
$$

with

$$\Psi_{\mathrm{w}}^{0}(\hat{\mathbf{Q}}^{\star}) \equiv \mathbb{E}_{\mathbf{w}^{\star}} \exp\left(\mathbf{w}^{\star\intercal}\hat{\mathbf{Q}}^{\star}\mathbf{w}^{\star}\right),$$
$$\Psi_{\mathrm{out}}^{0}(\mathbf{Q}^{\star}) \equiv \int_{\mathbb{R}} \mathrm{d}y \int \mathrm{d}\mathbf{z}^{\star}\, \mathrm{P}_{\mathrm{out}^{\star}}(y|\mathbf{z}^{\star})\, \mathscr{N}_{\mathbf{z}^{\star}}(\mathbf{0}, \mathbf{Q}^{\star}) = 1\,.$$

Taking the saddle point equations over $\mathbf{Q}^{\star}$ and $\hat{\mathbf{Q}}^{\star}$, imposing the consistency condition $\lim_{r\to 0}\Phi^{(r)}(\mathbf{Q},\hat{\mathbf{Q}}) = 0$, we finally obtain

$$\hat{\mathbf{Q}}^{\star} = \mathbf{0} \quad \text{and} \quad \mathbf{Q}^{\star} = \mathbb{E}_{\mathbf{w}^{\star}}\left[\mathbf{w}^{\star\intercal}\mathbf{w}^{\star}\right]. \tag{40}$$

3.2.4 Replica trick $r \to 0$ limit: $\Theta(r)$ terms

Imposing the conditions (40) avoids divergence in the replica trick, and we can therefore proceed with the $\Theta(r)$ terms.

Prior integral $\Psi_{\mathrm{w}}^{(r)}$ The limit $r \to 0$ and the derivative of the logarithm of the prior integral (22) can be trivially computed

$$\begin{aligned}
&\lim_{r\to 0} \partial_r \log \Psi_{\mathrm{w}}^{(r)}(\hat{\mathbf{Q}})\Big|_{\mathrm{rs}} \\
&= \mathbb{E}_{\boldsymbol{\xi},\mathbf{w}^{\star}} \\
&\times \log\left[\mathbb{E}_{\mathbf{w}} \exp\left(\left[\mathbf{w}^{\star\intercal}\hat{\mathbf{m}}\mathbf{w} - \frac{1}{2}\mathbf{w}^{\intercal}(\hat{\mathbf{Q}}+\hat{\mathbf{q}})\mathbf{w} + \boldsymbol{\xi}^{\intercal}\hat{\mathbf{q}}^{1/2}\mathbf{w}\right]\right)\right],
\end{aligned}$$

with $\boldsymbol{\xi} \sim \mathscr{N}(\mathbf{0}, \mathbf{I}_K)$ and $\mathbf{w}^{\star} \sim \mathrm{P}_{\mathrm{w}^{\star}}$. To conclude, we can symmetrize and decouple the *teacher* and *student* expectations $\mathbb{E}_{\mathbf{w}^{\star}}, \mathbb{E}_{\mathbf{w}}$. By performing the change of variable $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} + \hat{\mathbf{q}}^{-1/2}\hat{\mathbf{m}}\mathbf{w}^{\star}$. We finally obtain

$$\begin{aligned}
&\lim_{r\to 0} \partial_r \log \Psi_{\mathrm{w}}^{(r)}(\hat{\mathbf{Q}})\Big|_{\mathrm{rs}} \\
&= \mathbb{E}_{\boldsymbol{\xi},\mathbf{w}^{\star}} \exp\left(-\frac{1}{2}\mathbf{w}^{\star\intercal}\hat{\mathbf{m}}^{\intercal}\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\mathbf{w}^{\star} + \boldsymbol{\xi}^{\intercal}\hat{\mathbf{q}}^{-1/2}\hat{\mathbf{m}}\mathbf{w}^{\star}\right) \\
&\qquad \times \log\left[\mathbb{E}_{\mathbf{w}} \exp\left(\left[-\frac{1}{2}\mathbf{w}^{\intercal}(\hat{\mathbf{Q}}+\hat{\mathbf{q}})\mathbf{w} + \boldsymbol{\xi}^{\intercal}\hat{\mathbf{q}}^{1/2}\mathbf{w}\right]\right)\right] \\
&\equiv \mathbb{E}_{\boldsymbol{\xi},\mathbf{w}^{\star}} \mathscr{Z}_{\mathrm{w}^{\star}}\left(\hat{\mathbf{m}}\hat{\mathbf{q}}^{-1/2}\boldsymbol{\xi}, \hat{\mathbf{m}}^{\intercal}\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\right) \log \mathscr{Z}_{\mathrm{w}}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{Q}}+\hat{\mathbf{q}}\right),
\end{aligned} \tag{41}$$

with the corresponding denoising distribution Q_{w} and functions $\mathscr{Z}_{\mathrm{w}^{\star}}, \mathscr{Z}_{\mathrm{w}}$ defined by

$$\begin{aligned}
\mathrm{Q}_{\mathrm{w}}(\mathbf{w};\boldsymbol{\gamma},\boldsymbol{\Lambda}) &\equiv \frac{\mathrm{P}_{\mathrm{w}}(\mathbf{w})}{\mathscr{Z}_{\mathrm{w}}(\boldsymbol{\gamma},\boldsymbol{\Lambda})} e^{-\frac{1}{2}\mathbf{w}^{\intercal}\boldsymbol{\Lambda}\mathbf{w} + \boldsymbol{\gamma}^{\intercal}\mathbf{w}}, \\
\mathscr{Z}_{\mathrm{w}}(\boldsymbol{\gamma},\boldsymbol{\Lambda}) &\equiv \mathbb{E}_{\mathbf{w}\sim\mathrm{P}_{\mathrm{w}}}\left[e^{-\frac{1}{2}\mathbf{w}^{\intercal}\boldsymbol{\Lambda}\mathbf{w} + \boldsymbol{\gamma}^{\intercal}\mathbf{w}}\right] = \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}\; \mathrm{p}_{\mathrm{w}}(\mathbf{w}) e^{-\frac{1}{2}\mathbf{w}^{\intercal}\boldsymbol{\Lambda}\mathbf{w} + \boldsymbol{\gamma}^{\intercal}\mathbf{w}},
\end{aligned} \tag{42}$$

respectively with distribution $\mathrm{P}_{\mathrm{w}^{\star}}$ and P_{w}.

Prior integral $\Psi^{(r)}_{\text{out}}$ The limit $r \to 0$ and the derivative of the logarithm of the channel integral (39) is more tricky. First, the limit of the determinant (33) simplifies easily and yields

$$\det\left(\mathbf{Q}^{(\mathrm{rs})}\right) \underset{r\to 0}{\longrightarrow} \det\left(\mathbf{Q}^{\star}\right)$$

and the matrix elements of $\left(\mathbf{Q}^{(\mathrm{rs})}\right)^{-1}$ in this limit become

$$\begin{aligned}
\tilde{\mathbf{Q}}^{\star} &\underset{r\to 0}{\longrightarrow} \left(\mathbf{Q}^{\star}\right)^{-1}, \\
\tilde{\mathbf{m}} &\underset{r\to 0}{\longrightarrow} -\left(\mathbf{Q}^{\star}\right)^{-1}\mathbf{m}(\mathbf{Q}-\mathbf{q})^{-1}, \\
\tilde{\mathbf{q}} &\underset{r\to 0}{\longrightarrow} -(\mathbf{Q}-\mathbf{q})^{-1}\left(\mathbf{q}-\mathbf{m}^{\intercal}\left(\mathbf{Q}^{\star}\right)^{-1}\mathbf{m}\right)(\mathbf{Q}-\mathbf{q})^{-1}, \\
\tilde{\mathbf{Q}} &\underset{r\to 0}{\longrightarrow} \tilde{\mathbf{q}} + (\mathbf{Q}-\mathbf{q})^{-1}.
\end{aligned}$$

To take properly the $r \to 0$ limit, let us first perform the change of variable $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} - (-\tilde{\mathbf{q}})^{1/2}\tilde{\mathbf{m}}^{\intercal}\mathbf{z}^{\star}$ in (39) so that

$$\begin{aligned}
\Psi^{(r)}_{\text{out}}(\mathbf{Q})\Big|_{\text{rs}} &= \int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi},r) \times J_1(y,\boldsymbol{\xi},r)^r \\
J_0(y,\boldsymbol{\xi},r) &= \mathrm{e}^{-\frac{1}{2}\log\left(\det\left(2\pi\mathbf{Q}^{(\mathrm{rs})}\right)\right)} \\
&\quad \times \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^{\star}\, \mathrm{p}_{\mathrm{out}^{\star}}\left(y|\mathbf{z}^{\star}\right) e^{-\frac{1}{2}\mathbf{z}^{\star\intercal}\left(\tilde{\mathbf{Q}}^{\star}-\tilde{\mathbf{m}}^{\intercal}\tilde{\mathbf{q}}^{-1}\tilde{\mathbf{m}}\right)\mathbf{z}^{\star}+\boldsymbol{\xi}^{\intercal}(-\tilde{\mathbf{q}})^{1/2}\tilde{\mathbf{m}}^{\intercal}\mathbf{z}^{\star}} \\
J_1(y,\boldsymbol{\xi},r) &= \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}\, \mathrm{p}_{\mathrm{out}}\left(y|\mathbf{z}\right) e^{-\frac{1}{2}\mathbf{z}^{\intercal}\left(\tilde{\mathbf{Q}}-\tilde{\mathbf{q}}\right)\mathbf{z}+\mathbf{z}^{\intercal}(-\tilde{\mathbf{q}})^{1/2}\boldsymbol{\xi}}
\end{aligned} \tag{43}$$

so that the integral may be written as follows:

$$\begin{aligned}
&\log\left(\Psi^{(r)}_{\text{out}}\right) = \log\int \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi},r) J_1(y,\boldsymbol{\xi},r)^r \\
&= \log\int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}}\left[J_0(y,\boldsymbol{\xi},0) + r\partial_r J_0(y,\boldsymbol{\xi},0) + rJ_0(y,\boldsymbol{\xi},0)\log J_1(y,\boldsymbol{\xi},0) + \Theta(r^2)\right] \\
&= \log\left(1 + r\int_{\mathbb{R}} \mathrm{d}y \times \mathbb{E}_{\boldsymbol{\xi}}\left[\partial_r J_0(y,\boldsymbol{\xi},0) + J_0(y,\boldsymbol{\xi},0)\log J_1(y,\boldsymbol{\xi},0) + \Theta(r^2)\right]\right) \\
&= r\int_{\mathbb{R}} \mathrm{d}y \times \mathbb{E}_{\boldsymbol{\xi}}\left[\partial_r J_0(y,\boldsymbol{\xi},0) + J_0(y,\boldsymbol{\xi},0)\log J_1(y,\boldsymbol{\xi},0)\right] + \Theta(r^2)
\end{aligned}$$

where we used the consistency condition $\int_{\mathbb{R}} \mathrm{d}y \times \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi},0) = 1$. The first term can be evaluated similarly

$$\begin{aligned}
&\int_{\mathbb{R}} \mathrm{d}y \times \mathbb{E}_{\boldsymbol{\xi}} \partial_r J_0(y,\boldsymbol{\xi},0) = \partial_r\left[\int_{\mathbb{R}} \mathrm{d}y \times \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi},r)\right]_{r=0} \\
&= \partial_r\left[\mathrm{e}^{-\frac{1}{2}\log\left(\det\left(2\pi\mathbf{Q}^{(\mathrm{rs})}\right)\right)} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^{\star} e^{-\frac{1}{2}\mathbf{z}^{\star\intercal}\left(\tilde{\mathbf{Q}}^{\star}\right)\mathbf{z}^{\star}}\right]_{r=0} \\
&= \partial_r\left[\mathrm{e}^{-\frac{1}{2}\log\left(\det\left(\tilde{\mathbf{Q}}^{\star}\right)\det\left(\mathbf{Q}^{(\mathrm{rs})}\right)\right)}\right]_{r=0} = \ldots = 0
\end{aligned}$$

Finally we obtain:

$$\lim_{r\to 0}\partial_r \log\left(\Psi_{\text{out}}^{(r)}\right) = \int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi}) \log J_1(y,\boldsymbol{\xi}) = \lim_{r\to 0}\int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} J_0(y,\boldsymbol{\xi}) J_1(y,\boldsymbol{\xi})^r$$

with

$$\begin{aligned}
J_0(y,\boldsymbol{\xi}) &\equiv \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^\star \mathrm{p}_{\text{out}^\star}\left(y|\mathbf{z}^\star\right) e^{-\frac{1}{2}\mathbf{z}^{\star\intercal}\mathbf{N}\mathbf{z}^\star + \boldsymbol{\xi}^\intercal \mathbf{M}\mathbf{z}^\star} \\
J_1(y,\boldsymbol{\xi}) &\equiv \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}\, \mathrm{p}_{\text{out}}\left(y|\mathbf{z}\right) e^{-\frac{1}{2}\mathbf{z}^\intercal(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{z} + \mathbf{z}^\intercal \mathbf{P}\boldsymbol{\xi}} \\
&= \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\, \mathrm{p}_{\text{out}}\left(y|\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z}\right) e^{\boldsymbol{\xi}^\intercal \mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{z}}
\end{aligned}$$

where we changed the variable $\mathbf{z} \leftarrow (\mathbf{Q}-\mathbf{q})^{-1/2}\mathbf{z}$ and used the fact that with the second formulation the determinant term at the power r goes away in the limit $r \to 0$ and with

$$\begin{aligned}
\mathbf{P} &\equiv (-\tilde{\mathbf{q}})^{1/2} = (\mathbf{Q}-\mathbf{q})^{-1/2}\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)^{1/2}(\mathbf{Q}-\mathbf{q})^{-1/2} \\
\mathbf{M} &\equiv (-\tilde{\mathbf{q}})^{-1/2}\tilde{\mathbf{m}}^\intercal = -(-\tilde{\mathbf{q}})^{-1/2}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1} \\
&= -\mathbf{P}^{-1}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1} \\
\mathbf{N} &\equiv \tilde{\mathbf{Q}}^\star - \tilde{\mathbf{m}}^\intercal\tilde{\mathbf{q}}^{-1}\tilde{\mathbf{m}} \\
&= (\mathbf{Q}^\star)^{-1} + (\mathbf{Q}^\star)^{-1}\mathbf{m}(\mathbf{Q}-\mathbf{q})^{-1}(\mathbf{Q}-\mathbf{q})\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)^{-1}(\mathbf{Q}-\mathbf{q})(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1} \\
&= (\mathbf{Q}^\star)^{-1}\left[\mathbf{I}_K + \mathbf{m}\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\right] \\
&= (\mathbf{Q}^\star)^{-1}\left[\mathbf{I}_K + \left[(\mathbf{Q}^\star)^{-1}\mathbf{m}^{-\intercal}\mathbf{q}\mathbf{m}^{-1} - \mathbf{I}_K\right]^{-1}\right] \\
&= \left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{-1}
\end{aligned}$$

Then we rescale $\mathbf{z}^\star \leftarrow \mathbf{N}^{1/2}\mathbf{z}^\star$ such that

$$\begin{aligned}
J_0(y,\boldsymbol{\xi}) &= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}^\star \mathrm{p}_{\text{out}^\star}\left(y|\mathbf{z}^\star\right) e^{-\frac{1}{2}\mathbf{z}^{\star\intercal}\mathbf{N}\mathbf{z}^\star + \boldsymbol{\xi}^\intercal \mathbf{M}\mathbf{z}^\star} \\
&= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)} \\
&\quad \times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\text{out}^\star}\left(y|\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star\right) e^{\boldsymbol{\xi}^\intercal \mathbf{M}\mathbf{N}^{-1/2}\mathbf{z}^\star}\right]
\end{aligned}$$

Moreover notice that:

$$\begin{aligned}
\mathbf{M}\mathbf{N}^{-1/2} &= \mathbf{P}^{-1}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{1/2} \\
&= \mathbf{P}\mathbf{P}^{-2}(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{1/2} \\
&= \mathbf{P}(\mathbf{Q}-\mathbf{q})\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)^{-1}(\mathbf{Q}-\mathbf{q})(\mathbf{Q}-\mathbf{q})^{-1}\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{1/2} \\
&= \mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}
\end{aligned}$$

with

$$\mathbf{R} = (\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{q}^{-1}\mathbf{m}^\intercal\left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{-1/2}$$

and notice that $\mathbf{R} = \mathbf{I}_K$ in the Bayes-optimal setting. Therefore:

$$\begin{aligned}
J_0(y,\boldsymbol{\xi}) &= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)} \\
&\quad\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\mathrm{out}^\star}\left(y|\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star\right)e^{\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}\mathbf{z}^\star}\right]
\end{aligned}$$

As a summary, we have:

$$\lim_{r\to 0}\partial_r\log\left(\Psi_{\mathrm{out}}^{(r)}\right) = \lim_{r\to 0}\int_{\mathbb{R}}\mathrm{d}y\mathbb{E}_{\boldsymbol{\xi}}J_0(y,\boldsymbol{\xi})J_1(y,\boldsymbol{\xi})^r$$

with

$$\begin{aligned}
J_0(y,\boldsymbol{\xi}) &= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)} \\
&\quad\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\mathrm{out}^\star}\left(y|\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star\right)e^{\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}\mathbf{z}^\star}\right] \\
J_1(y,\boldsymbol{\xi}) &= \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\,\mathrm{p}_{\mathrm{out}}\left(y|\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z}\right)e^{\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{z}}
\end{aligned}$$

We can shift $\mathbf{z}\leftarrow\mathbf{z}-(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{P}\boldsymbol{\xi}$ to obtain

$$\begin{aligned}
J_1(y,\boldsymbol{\xi}) &= e^{\frac{1}{2}\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{P}\boldsymbol{\xi}} \\
&\quad\times \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\,\mathrm{p}_{\mathrm{out}}\left(y|\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z}+(\mathbf{Q}-\mathbf{q})\mathbf{P}\boldsymbol{\xi}\right)
\end{aligned}$$

and $\mathbf{z}^\star\leftarrow\mathbf{z}^\star-\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{-1/2}\mathbf{m}\mathbf{q}^{-1}\left(\mathbf{Q}-\mathbf{q}\right)\mathbf{P}\boldsymbol{\xi} = \mathbf{z}^\star-\mathbf{R}^\intercal\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{P}\boldsymbol{\xi}$:

$$\begin{aligned}
J_0(y,\boldsymbol{\xi}) &= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)} \\
&\quad\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\mathrm{out}^\star}\left(y|\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star\right)e^{\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}\mathbf{z}^\star}\right] \\
&= \mathrm{e}^{-\frac{1}{2}\log(\det(2\pi\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)}e^{\frac{1}{2}\boldsymbol{\xi}^\intercal\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}\mathbf{R}^\intercal(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{P}\boldsymbol{\xi}} \\
&\quad\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\mathrm{out}^\star}\left(y|\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star+\mathbf{m}\mathbf{q}^{-1}\left(\mathbf{Q}-\mathbf{q}\right)\mathbf{P}\boldsymbol{\xi}\right)\right]
\end{aligned}$$

We obtain (taking the limit $r\to 0$ to the term in front of J_1):

$$\begin{aligned}
&\lim_{r\to 0}\partial_r\log\left(\Psi_{\mathrm{out}}^{(r)}\right) \\
&= \int_{\mathbb{R}}\mathrm{d}y\int_{\mathbb{R}^K}\mathrm{d}\boldsymbol{\xi}\,\mathrm{e}^{-\frac{1}{2}\log(\det(\mathbf{Q}^\star))}\mathrm{e}^{\frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right)}e^{-\frac{1}{2}\boldsymbol{\xi}^\intercal\left[\mathbf{I}_K-\mathbf{P}(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{R}\mathbf{R}^\intercal(\mathbf{Q}-\mathbf{q})^{1/2}\mathbf{P}\right]\boldsymbol{\xi}} \\
&\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\mathrm{out}^\star}\left(y|\left(\mathbf{Q}^\star-\mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star+\mathbf{m}\mathbf{q}^{-1}\left(\mathbf{Q}-\mathbf{q}\right)\mathbf{P}\boldsymbol{\xi}\right)\right] \\
&\times \log\left[\mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\,\mathrm{p}_{\mathrm{out}}\left(y|\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z}+(\mathbf{Q}-\mathbf{q})\mathbf{P}\boldsymbol{\xi}\right)\right]
\end{aligned}$$

Computing

$$
\begin{aligned}
&\left[\mathbf{I}_K - \mathbf{P}\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{R}\mathbf{R}^\intercal\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{P}\right] \\
&= \mathbf{P}\left[\mathbf{P}^{-2} - \left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{R}\mathbf{R}^\intercal\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\right]\mathbf{P} \\
&= \mathbf{P}(\mathbf{Q}-\mathbf{q})\left[\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)^{-1} - \mathbf{q}^{-1}\mathbf{m}^\intercal\left[\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right]^{-1}\mathbf{m}\mathbf{q}^{-1}\right](\mathbf{Q}-\mathbf{q})\mathbf{P} \\
&= \mathbf{P}(\mathbf{Q}-\mathbf{q})\mathbf{q}^{-1}(\mathbf{Q}-\mathbf{q})\mathbf{P}
\end{aligned}
$$

By finally changing the Gaussian measure $\boldsymbol{\xi} \leftarrow \mathbf{q}^{-1/2}\left(\mathbf{Q}-\mathbf{q}\right)\mathbf{P}\boldsymbol{\xi}$ we obtain

$$
\begin{aligned}
&\lim_{r\to 0}\partial_r \log\left(\Psi_{\text{out}}^{(r)}\right) \\
&= \exp\left(C\right) \times \int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} \\
&\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\left[\mathrm{p}_{\text{out}^\star}\left(y|\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star + \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}\right)\right] \\
&\times \log\left[\mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)}\,\mathrm{p}_{\text{out}}\left(y|\left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z} + \mathbf{q}^{1/2}\boldsymbol{\xi}\right)\right]
\end{aligned}
$$

with

$$
\begin{aligned}
C &= -\frac{1}{2}\log(\det\left(\mathbf{Q}^\star\right)) + \frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right) \\
&\quad - \frac{1}{2}\log\det\left(\mathbf{P}^2\right) - \log\det\left(\mathbf{Q}-\mathbf{q}\right) + \frac{1}{2}\log\det\left(\mathbf{q}\right) \\
&= -\frac{1}{2}\log(\det\left(\mathbf{Q}^\star\right)) + \frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right) \\
&\quad - \frac{1}{2}\log\det\left(\left(\mathbf{Q}-\mathbf{q}\right)\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)\left(\mathbf{Q}-\mathbf{q}\right)\right) - \log\det\left(\mathbf{Q}-\mathbf{q}\right) + \frac{1}{2}\log\det\left(\mathbf{q}\right) \\
&= -\frac{1}{2}\log(\det\left(\mathbf{Q}^\star\right)) + \frac{1}{2}\log\left(\det\left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)\right) \\
&\quad - \frac{1}{2}\log\det\left(\left(\mathbf{q}-\mathbf{m}^\intercal(\mathbf{Q}^\star)^{-1}\mathbf{m}\right)\right) + \frac{1}{2}\log\det\left(\mathbf{q}\right) \\
&= -\frac{1}{2}\log(\det\left(\mathbf{Q}^\star\right)) - \frac{1}{2}\log\det\left(\mathbf{m}\mathbf{m}^{-1}\right) - \frac{1}{2}\log\det\left(\mathbf{q}\right) + \frac{1}{2}\log(\det\left(\mathbf{Q}^\star\right)) + \frac{1}{2}\log(\det\left(\mathbf{q}\right)) \\
&= 0
\end{aligned}
$$

We obtain:

$$
\begin{aligned}
&\lim_{r\to 0} \partial_r \log \Psi_{\text{out}}^{(r)}(\mathbf{Q})\Big|_{\text{rs}} \\
&= \int_{\mathbb{R}} \mathrm{d}y \mathbb{E}_{\boldsymbol{\xi}} \\
&\times \mathbb{E}_{\mathbf{z}^\star\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)} \left[\mathrm{p}_{\text{out}^\star}\left(y | \left(\mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right)^{1/2}\mathbf{z}^\star + \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}\right)\right] \\
&\times \log\left[\mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)} \, \mathrm{p}_{\text{out}}\left(y | \left(\mathbf{Q}-\mathbf{q}\right)^{1/2}\mathbf{z} + \mathbf{q}^{1/2}\boldsymbol{\xi}\right)\right] \\
&= \int_{\mathbb{R}} \mathrm{d}y \, \mathbb{E}_{\boldsymbol{\xi}} \mathscr{Z}_{\text{out}^\star}\left(\mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}, \mathbf{Q}^\star - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right) \log \mathscr{Z}_{\text{out}}\left(\mathbf{q}^{1/2}\boldsymbol{\xi}, \mathbf{Q}-\mathbf{q}\right),
\end{aligned}
\tag{44}
$$

where the denoising distribution Q_{out} and functions $\mathscr{Z}_{\text{out}^\star}$, $\mathscr{Z}_{\text{out}}$ are defined by

$$
\begin{aligned}
\mathrm{Q}_{\text{out}}(\mathbf{z}; y, \boldsymbol{\omega}, \mathbf{V}) &\equiv \frac{\mathrm{P}_{\text{out}}(y|\mathbf{z})}{\mathscr{Z}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V})} \frac{e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\omega})^\intercal \mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\omega})}}{\sqrt{\det(2\pi\mathbf{V})}}, \\
\mathscr{Z}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V}) &\equiv \mathbb{E}_{\mathbf{z}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)} \left[\mathrm{P}_{\text{out}}\left(y | \mathbf{V}^{1/2}\mathbf{z} + \boldsymbol{\omega}\right)\right] \\
&= \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z} \, \mathrm{p}_{\text{out}}(y|\mathbf{z}) \frac{e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\omega})^\intercal \mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\omega})}}{\sqrt{\det(2\pi\mathbf{V})}}.
\end{aligned}
\tag{45}
$$

for the distributions $\mathrm{P}_{\text{out}^\star}$, P_{out}.

3.3 Summary

3.3.1 Summary - Mismatched case

In the mismatched case, where the teacher and the student have not the same prior distributions, we finally obtain the replica symmetric free entropy Φ_{rs} for the committee machine hypothesis class:

$$
\begin{aligned}
\Phi_{\text{rs}}(\alpha) &\equiv \mathbb{E}_{\mathbf{y},\mathbf{X}}\left[\lim_{d\to\infty} \frac{1}{d} \log\left(\mathscr{Z}_d(\mathbf{y}, \mathbf{X})\right)\right] \\
&= \mathbf{extr}_{\mathbf{Q},\hat{\mathbf{Q}},\mathbf{q},\hat{\mathbf{q}},\mathbf{m},\hat{\mathbf{m}}} \Big\{ -\mathrm{Tr}\left(\mathbf{m}\hat{\mathbf{m}}\right) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{Q}\hat{\mathbf{Q}}\right) + \frac{1}{2}\mathrm{Tr}\left(\mathbf{q}\hat{\mathbf{q}}\right) \\
&\qquad + \Psi_{\text{w}}\left(\hat{\mathbf{Q}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}\right) + \alpha \Psi_{\text{out}}\left(\mathbf{Q}, \mathbf{m}, \mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right)\Big\},
\end{aligned}
\tag{46}
$$

where $\boldsymbol{\rho}_{\text{w}^\star} \equiv \lim_{d\to\infty} \mathbf{Q}^\star = \lim_{d\to\infty} \mathbb{E}_{\mathbf{w}^\star} \frac{1}{d} \mathbf{W}^{\star\intercal}\mathbf{W}^\star$ and the channel and prior integrals are defined by

$$
\begin{aligned}
\Psi_{\text{w}}\left(\hat{\mathbf{Q}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}\right) &\equiv \mathbb{E}_{\boldsymbol{\xi}}\Big[\mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{m}}\hat{\mathbf{q}}^{-1/2}\boldsymbol{\xi}, \hat{\mathbf{m}}\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\right) \\
&\qquad \times \log \mathscr{Z}_{\text{w}}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{Q}} + \hat{\mathbf{q}}\right)\Big],
\end{aligned}
\tag{47}
$$

$$
\begin{aligned}
\Psi_{\text{out}}\left(\mathbf{Q}, \mathbf{m}, \mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right) &\equiv \mathbb{E}_{y,\boldsymbol{\xi}}\Big[\mathscr{Z}_{\text{out}^\star}\left(y, \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{m}\mathbf{q}^{-1}\mathbf{m}^\intercal\right) \\
&\qquad \times \log \mathscr{Z}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \mathbf{Q}-\mathbf{q}\right)\Big],
\end{aligned}
\tag{48}
$$

where again $\mathscr{Z}_{\text{out}^\star}, \mathscr{Z}_{\text{w}^\star}$ and $\mathscr{Z}_{\text{out}}, \mathscr{Z}_{\text{w}}$ are defined in (42)- (45) and depend respectively on channel and prior distributions of the *teacher* and *student*.

3.3.2 Bayes optimal MMSE estimation

For MMSE estimation in the Bayes-optimal setting, the student has access to the ground truth distributions of the teacher $\mathrm{P}_{\text{out}}(\mathbf{y}|\mathbf{Z}) = \mathrm{P}_{\text{out}^\star}(\mathbf{y}|\mathbf{Z})$ and $\mathrm{P}_{\text{w}}(\mathbf{W}) = \mathrm{P}_{\text{w}^\star}(\mathbf{W})$, and therefore $\mathscr{Z}_{\text{out}} = \mathscr{Z}_{\text{out}^\star}$, $\mathscr{Z}_{\text{w}} = \mathscr{Z}_{\text{w}^\star}$. In this idealistic setting, the Nishimori conditions imply that

$$\mathbf{Q} = \mathbf{Q}_{\text{w}^\star}\,, \quad \hat{\mathbf{Q}} = \mathbf{0}\,, \quad \mathbf{m} = \mathbf{q} \equiv \mathbf{q}_{\text{b}}\,, \quad \hat{\mathbf{m}} = \hat{\mathbf{q}} \equiv \hat{\mathbf{q}}_{\text{b}}\,. \tag{49}$$

Therefore the free entropy in eq. (47) simplifies as an optimization problem over the *overlaps* $\mathbf{q}_{\text{b}}, \hat{\mathbf{q}}_{\text{b}} \in \mathbb{R}^{K\times K}$

$$\Phi^{\text{b}}_{\text{rs}}(\alpha) = \mathbf{extr}_{\mathbf{q}_{\text{b}},\hat{\mathbf{q}}_{\text{b}}} \left\{ -\frac{1}{2}\mathrm{Tr}\left(\mathbf{q}_{\text{b}}\hat{\mathbf{q}}_{\text{b}}\right) + \Psi^{\text{b}}_{\text{w}}\left(\hat{\mathbf{q}}_{\text{b}}\right) + \alpha\Psi^{\text{b}}_{\text{out}}\left(\mathbf{q}_{\text{b}}; \boldsymbol{\rho}_{\text{w}^\star}\right) \right\}, \tag{50}$$

with free entropy terms $\Psi^{\text{b}}_{\text{w}}$ and $\Psi^{\text{b}}_{\text{out}}$ given by

$$\begin{aligned} \Psi^{\text{b}}_{\text{w}}\left(\hat{\mathbf{q}}\right) &= \mathbb{E}_{\xi}\left[\mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{q}}\right) \log \mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{q}}\right)\right], \\ \Psi^{\text{b}}_{\text{out}}\left(\mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right) &= \mathbb{E}_{y,\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}\right) \right. \\ &\qquad \left. \times \log \mathscr{Z}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}\right)\right]. \end{aligned} \tag{51}$$

Application to the GLM For the GLM hypothesis class, the same equations are valid if we take $K = 1$ for both the teacher and the student. As a result, we recover the replica symmetric free entropy in the Bayes-optimal setting rigorously proven in [23].

3.4 Fixed point equations

The overlaps parameters, such as $\mathbf{m}, \mathbf{q}$, play a crucial role since they measure the performances of the statistical estimation. Their behaviours are respectively characterized by the extremization of the free entropy (46) in the mismatched setting and (50) in the Bayes-optimal case. In this section, we give the expressions of the corresponding fixed point equations, whose derivations can be found in [25] Appendix. IV.4-5 for $K = 1$ which can be extended to $K \geq 1$.

3.4.1 Mismatched setting

Extremizing the free entropy eq. (46), we easily obtain the set of six fixed point equations

$$\begin{aligned} \hat{\mathbf{Q}} &= -2\alpha\partial_{\mathbf{Q}}\Psi_{\text{out}}\left(\mathbf{Q}, \mathbf{m}, \mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right), & \mathbf{Q} &= -2\partial_{\hat{\mathbf{Q}}}\Psi_{\text{w}}\left(\hat{\mathbf{Q}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}\right) \\ \hat{\mathbf{q}} &= -2\alpha\partial_{\mathbf{q}}\Psi_{\text{out}}\left(\mathbf{Q}, \mathbf{m}, \mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right), & \mathbf{q} &= -2\partial_{\hat{\mathbf{q}}}\Psi_{\text{w}}\left(\hat{\mathbf{Q}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}\right), \\ \hat{\mathbf{m}} &= \alpha\partial_{\mathbf{m}}\Psi_{\text{out}}\left(\mathbf{Q}, \mathbf{m}, \mathbf{q}; \boldsymbol{\rho}_{\text{w}^\star}\right), & \mathbf{m} &= \partial_{\hat{\mathbf{m}}}\Psi_{\text{w}}\left(\hat{\mathbf{Q}}, \hat{\mathbf{m}}, \hat{\mathbf{q}}\right). \end{aligned} \tag{52}$$

Interestingly, these equations can be reformulated as functions of $\mathscr{Z}_{\text{out}^\star}$, $\mathscr{Z}_{\text{w}^\star}$ and the denoising functions $f_{\text{out}^\star}, f_{\text{w}^\star}, f_{\text{out}}, f_{\text{w}}$ defined by

$$
\begin{aligned}
\mathbf{f}_{\text{w}}(\boldsymbol{\gamma}, \boldsymbol{\Lambda}) &\equiv \partial_{\boldsymbol{\gamma}} \log\left(\mathscr{Z}_{\text{w}}(\boldsymbol{\gamma}, \boldsymbol{\Lambda})\right) = \mathbb{E}_{\text{Q}_{\text{w}}}\left[\mathbf{w}\right], && (53)\\
\partial_{\boldsymbol{\gamma}} \mathbf{f}_{\text{w}}(\boldsymbol{\gamma}, \boldsymbol{\Lambda}) &\equiv \mathbb{E}_{\text{Q}_{\text{w}}}\left[\mathbf{w}\mathbf{w}^\intercal\right] - \mathbf{f}_{\text{w}}(\boldsymbol{\gamma}, \boldsymbol{\Lambda})^{\otimes 2}, && (54)\\
\mathbf{f}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V}) &\equiv \partial_{\boldsymbol{\omega}} \log\left(\mathscr{Z}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V})\right) = \mathbf{V}^{-1}\mathbb{E}_{\text{Q}_{\text{out}}}\left[\mathbf{z} - \boldsymbol{\omega}\right], && (55)\\
\partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V}) &\equiv \frac{\partial \mathbf{f}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V})}{\partial \boldsymbol{\omega}} && (56)\\
&= \mathbf{V}^{-1}\mathbb{E}_{\text{Q}_{\text{out}}}\left[(\mathbf{z} - \boldsymbol{\omega})^{\otimes 2}\right]\mathbf{V}^{-1} - \mathbf{V}^{-1} - \mathbf{f}_{\text{out}}(y, \boldsymbol{\omega}, \mathbf{V})^{\otimes 2}.
\end{aligned}
$$

Defining the natural variables $\boldsymbol{\Sigma} = \mathbf{Q} - \mathbf{q}$ and $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{Q}} + \hat{\mathbf{q}}$ they can reformulated as

$$
\begin{aligned}
\hat{\mathbf{m}} &= \alpha \mathbb{E}_{y,\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{out}^\star} \times \mathbf{f}_{\text{out}^\star}\left(y, \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{m}^\intercal\mathbf{q}^{-1}\mathbf{m}\right)\right.\\
&\qquad\qquad \left.\times \mathbf{f}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \boldsymbol{\Sigma}\right)^\intercal\right],\\
\hat{\mathbf{q}} &= \alpha \mathbb{E}_{y,\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{out}^\star}\left(y, \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{m}^\intercal\mathbf{q}^{-1}\mathbf{m}\right)\mathbf{f}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \boldsymbol{\Sigma}\right)^{\otimes 2}\right],\\
\hat{\boldsymbol{\Sigma}} &= -\alpha \mathbb{E}_{y,\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{out}^\star}\left(y, \mathbf{m}\mathbf{q}^{-1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{m}^\intercal\mathbf{q}^{-1}\mathbf{m}\right)\right.\\
&\qquad\qquad \left.\times \partial_{\boldsymbol{\omega}}\mathbf{f}_{\text{out}}\left(y, \mathbf{q}^{1/2}\boldsymbol{\xi}, \boldsymbol{\Sigma}\right)\right],\\
\mathbf{m} &= \mathbb{E}_{\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{w}^\star} \times f_{\text{w}^\star}\left(\hat{\mathbf{m}}\hat{\mathbf{q}}^{-1/2}\boldsymbol{\xi}, \hat{\mathbf{m}}^\intercal\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\right)\mathbf{f}_{\text{w}}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\boldsymbol{\Sigma}}\right)\right],\\
\mathbf{q} &= \mathbb{E}_{\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{m}}\hat{\mathbf{q}}^{-1/2}\boldsymbol{\xi}, \hat{\mathbf{m}}^\intercal\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\right)\mathbf{f}_{\text{w}}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\boldsymbol{\Sigma}}\right)^2\right],\\
\boldsymbol{\Sigma} &= \mathbb{E}_{\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{m}}\hat{\mathbf{q}}^{-1/2}\boldsymbol{\xi}, \hat{\mathbf{m}}^\intercal\hat{\mathbf{q}}^{-1}\hat{\mathbf{m}}\right)\partial_{\boldsymbol{\gamma}}\mathbf{f}_{\text{w}}\left(\hat{\mathbf{q}}^{1/2}\boldsymbol{\xi}, \hat{\boldsymbol{\Sigma}}\right)\right],
\end{aligned}
\tag{57}
$$

where we use the abusive notation $\mathbb{E}_y = \int_{\mathbb{R}} \mathrm{d}y$.

3.4.2 Bayes-optimal estimation

Extremizing the Bayes-optimal free entropy eq. (50), we easily obtain the set of fixed point equations over the scalar parameters $\mathbf{q}_{\text{b}}, \hat{\mathbf{q}}_{\text{b}}$. It can be deduced from eq. (57) using the Nishimori conditions $\mathbf{f}_{\text{w}} = \mathbf{f}_{\text{w}^\star}$, $\mathbf{f}_{\text{out}} = \mathbf{f}_{\text{out}^\star}$, $\mathbf{m} = \mathbf{q}, \boldsymbol{\Sigma} = \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}, \hat{\mathbf{m}} = \hat{\mathbf{q}}$ and $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{q}}$ that lead to

$$
\begin{aligned}
\hat{\mathbf{q}}_{\text{b}} &= \alpha \mathbb{E}_{y,\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{out}^\star}\left(y, \mathbf{q}_{\text{b}}^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}_{\text{b}}\right)\mathbf{f}_{\text{out}^\star}\left(y, \mathbf{q}_{\text{b}}^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}_{\text{b}}\right)^{\otimes 2}\right],\\
\mathbf{q}_{\text{b}} &= \mathbb{E}_{\boldsymbol{\xi}}\left[\mathscr{Z}_{\text{w}^\star}\left(\hat{\mathbf{q}}_{\text{b}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{q}}_{\text{b}}\right)\mathbf{f}_{\text{w}^\star}\left(\hat{\mathbf{q}}_{\text{b}}^{1/2}\boldsymbol{\xi}, \hat{\mathbf{q}}_{\text{b}}\right)^{\otimes 2}\right].
\end{aligned}
\tag{58}
$$

4 Belief Propagation and Approximate Message Passing

The Approximate Message Passing (AMP) algorithm can be seen as Taylor expansion of the loopy Belief Propagation (BP) approach [1, 26, 27], similar to the so-called TAP equation in spin glass theory [28]. While the behavior of AMP can be rigorously studied [29, 30, 31], it is useful and instructive to see how the derivation can be performed in the framework of BP and the cavity method, as was pioneered in [17] for the single layer problem. The derivation uses the GAMP notations of [32] and follows closely the one of [33]. The computation is presented for the committee machine hypothesis class, which is the vectorized version of the GLM, with $K \geq 1$ vectorial parameters $\mathbf{W} = \{\mathbf{w}_k\}_{k=1}^K \in \mathbb{R}^{d\times K}$.

4.1 Factor graph and joint probability distribution

As a central illustration, we present the instructive derivation of the rBP equations starting with the BP equations in the context of committee machines. We recall the Joint Probability Distribution (JPD)

$$\mathrm{P}_d\left(\mathbf{W}|\mathbf{y},\mathbf{X}\right) = \frac{\mathrm{P}_{\mathrm{out}}(\mathbf{y}|\mathbf{Z})\mathrm{P}_{\mathrm{w}}(\mathbf{W})}{\mathscr{Z}_d(\mathbf{y},\mathbf{X})} = \frac{\prod_{\mu=1}^n \mathrm{P}_{\mathrm{out}}(y_\mu|\mathbf{z}_\mu)\prod_{i=1}^d \mathrm{P}_{\mathrm{w}}(\mathbf{w}_i)}{\mathscr{Z}_d(\mathbf{y},\mathbf{X})}\,, \tag{59}$$

where we defined $\mathbf{Z} = \frac{1}{\sqrt{d}}\mathbf{X}\mathbf{Z} \in \mathbb{R}^{n\times K}$ and we assume that the channel and prior distributions factorize over factors $\mathrm{P}_{\mathrm{out}}(y_\mu|\mathbf{z}_\mu)$ and variables $\mathrm{P}_{\mathrm{w}}(\mathbf{w}_i)$.

The posterior distribution may be represented by the following bipartite factor graph in Fig. 3. In the following, we attach a set of *messages* $\{m_{i\to\mu}, \tilde{m}_{\mu\to i}\}_{i=1..n}^{\mu=1..m}$ to

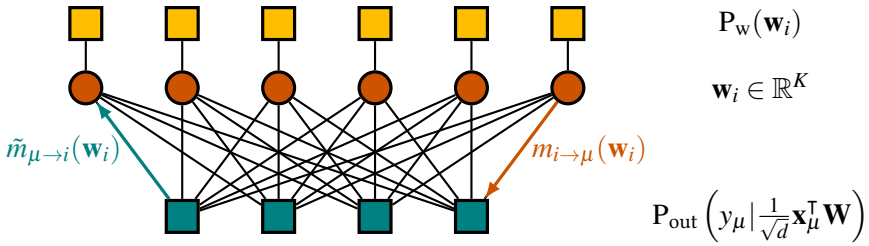

Figure 3 – Factor graph representation of the joint distribution for committee machines.

the edges of this bipartite factor graph. These messages correspond to the marginal probabilities of $\mathbf{w}_i \in \mathbb{R}^K$ if we remove the edges $(i \to \mu)$ or $(\mu \to i)$. We define the auxiliary variable $\mathbf{z}_\mu = \frac{1}{\sqrt{d}}\mathbf{x}_\mu^\intercal\mathbf{W} \in \mathbb{R}^K$ which is $\Theta(1)$ thanks to the pre-factor rescaling $1/\sqrt{d}$. This scaling is crucial as it allows the BP equations to hold true even though the factor graph is not tree-like and is instead fully connected with short loops.

4.2 Belief Propagation equations

The BP equations (also called the sum-product equations) for $\mathbf{w}_i = (w_{ik})_{k=1..K} \in \mathbb{R}^K$ on the factor graph Fig. 3 can be formulated as:

$$
\begin{aligned}
m^{t+1}_{i\to\mu}(\mathbf{w}_i) &= \frac{1}{\mathscr{Z}_{i\to\mu}} \mathrm{P_w}(\mathbf{w}_i) \prod_{\nu\neq\mu}^{n} \tilde{m}^t_{\nu\to i}(\mathbf{w}_i) \\
\tilde{m}^t_{\mu\to i}(\mathbf{w}_i) &= \frac{1}{\mathscr{Z}_{\mu\to i}} \int_{\mathbb{R}^K} \prod_{j\neq i}^{d} \mathrm{d}\mathbf{w}_j \; \mathrm{P_{out}}\left(y_\mu | \frac{1}{\sqrt{d}} \sum_{j=1}^{d} x_{\mu j}\mathbf{w}_j\right) m^t_{j\to\mu}(\mathbf{w}_j)\,,
\end{aligned}
\tag{60}
$$

The BP equations assume that incoming messages are independent. Hence these equations are exact on a tree (no loop), but they remain exact if "correlations decrease fast enough / long loops". We assume in the following that the hypothesis is true in our model.

4.3 Relaxed Belief Propagation equations

The idea of the relaxed BP equations is to simply expand in the limit $d \to \infty$ the set of $\Theta(d^2)$ messages $\tilde{m}$ of the BP equations in (60) before plugging them in m. Truncating the expansion and keeping only terms of order $\Theta(1/d)$, messages become *Gaussian*. Hence messages are therefore parametrized only by the mean $\hat{\mathbf{w}}^t_{i\to\mu}$ and the covariance matrix $\hat{\mathbf{C}}^t_{i\to\mu}$ of the marginal distribution at time t:

$$
\begin{aligned}
\hat{\mathbf{w}}^t_{i\to\mu} &\equiv \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}_i \; m^t_{i\to\mu}(\mathbf{w}_i)\; \mathbf{w}_i \\
\hat{\mathbf{C}}^t_{i\to\mu} &\equiv \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}_i \; m^t_{i\to\mu}(\mathbf{w}_i)\; \mathbf{w}_i\mathbf{w}_i^\intercal - \hat{\mathbf{w}}^t_{i\to\mu}(\hat{\mathbf{w}}^t_{i\to\mu})^\intercal
\end{aligned}
\tag{61}
$$

To decouple the argument of $\mathrm{P_{out}}$, we first by introducing its Fourier transform $\hat{\mathrm{P}}_{\mathrm{out}}$ according to

$$
\begin{aligned}
\mathrm{P_{out}}\left(y_\mu | \frac{1}{\sqrt{d}} \sum_{j=1}^{d} x_{\mu j}\mathbf{w}_j\right) &= \frac{1}{(2\pi)^{K/2}} \\
&\times \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\xi} \exp\left(i\boldsymbol{\xi}^\intercal \left(\frac{1}{\sqrt{d}} \sum_{j=1}^{d} x_{\mu j}\mathbf{w}_j\right) \hat{\mathrm{P}}_{\mathrm{out}}(y_\mu, \boldsymbol{\xi})\right).
\end{aligned}
$$

Injecting this representation in the BP equations, (60) becomes:

$$
\begin{aligned}
\tilde{m}^t_{\mu\to i}(\mathbf{w}_i) = \frac{1}{(2\pi)^{K/2}\mathscr{Z}_{\mu\to i}} \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\xi}\; \hat{\mathrm{P}}_{\mathrm{out}}(y_\mu, \boldsymbol{\xi})\; \exp\left(i\boldsymbol{\xi}^\intercal \frac{1}{\sqrt{d}} x_{\mu i}\mathbf{w}_i\right) \\
\times \prod_{j\neq i}^{d} \underbrace{\int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}_j \; m^t_{j\to\mu}(\mathbf{w}_j)\; \exp\left(i\boldsymbol{\xi}^\intercal \frac{1}{\sqrt{d}} x_{\mu j}\mathbf{w}_j\right)}_{\equiv I_j}
\end{aligned}
\tag{62}
$$

In the limit $d \to \infty$ the term I_j can be easily expanded and expressed using $\hat{\mathbf{w}}$ and $\hat{\mathbf{C}}$ in (61):

$$
\begin{aligned}
I_j &= \int_{\mathbb{R}^K} \mathrm{d}\mathbf{w}_j \, m^t_{j\to\mu}(\mathbf{w}_j) \, \exp\left(i\frac{x_{\mu j}}{\sqrt{d}}\boldsymbol{\xi}^\intercal \mathbf{w}_j)\right) \\
&\simeq \exp\left(i\frac{x_{\mu j}}{\sqrt{d}}\boldsymbol{\xi}^\intercal \hat{\mathbf{w}}^t_{j\to\mu} - \frac{1}{2}\frac{x^2_{\mu j}}{d}\boldsymbol{\xi}^\intercal \hat{\mathbf{C}}^t_{j\to\mu}\boldsymbol{\xi}\right).
\end{aligned}
$$

Finally using the inverse Fourier transform:

$$
\begin{aligned}
\tilde{m}^t_{\mu\to i}(\mathbf{w}_i) &= \frac{1}{(2\pi)^{K/2}\mathscr{Z}_{\mu\to i}} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z} \mathrm{P}_{\text{out}}(y_\mu|\mathbf{z}) \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\xi}\, e^{-i\boldsymbol{\xi}^\intercal \mathbf{z}} e^{ix_{\mu i}\boldsymbol{\xi}^\intercal \mathbf{w}_i} \\
&\qquad \times \prod_{j\neq i}^{d} \exp\left(i\frac{x_{\mu j}}{\sqrt{d}}\boldsymbol{\xi}^\intercal \hat{\mathbf{w}}^t_{j\to\mu} - \frac{1}{2}\frac{x^2_{\mu j}}{d}\boldsymbol{\xi}^\intercal \hat{\mathbf{C}}^t_{j\to\mu}\boldsymbol{\xi}\right) \\
&= \frac{1}{(2\pi)^K \mathscr{Z}_{\mu\to i}} \int_{\mathbb{R}} \mathrm{d}\mathbf{z}\, \mathrm{P}_{\text{out}}(y_\mu|\mathbf{z}) \\
&\qquad \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\xi}\; e^{-i\boldsymbol{\xi}^\intercal \mathbf{z}} e^{ix_{\mu i}\boldsymbol{\xi}^\intercal \mathbf{w}_i} e^{i\sum_{j\neq i}^{d}\frac{x_{\mu j}}{\sqrt{d}}\boldsymbol{\xi}^\intercal \hat{\mathbf{w}}^t_{j\to\mu}} e^{-\frac{1}{2}\sum_{j\neq i}^{d}\frac{x^2_{\mu j}}{d}\boldsymbol{\xi}^\intercal \hat{\mathbf{C}}^t_{j\to\mu}\boldsymbol{\xi}} \\
&= \frac{1}{(2\pi)^K \mathscr{Z}_{\mu\to i}} \int_{\mathbb{R}^K} \mathrm{d}\mathbf{z}\, \mathrm{P}_{\text{out}}(y_\mu|\mathbf{z}) \\
&\qquad \times \sqrt{\frac{(2\pi)^K}{\det\left(V^t_{\mu\to i}\right)}} \underbrace{e^{-\frac{1}{2}\left(\mathbf{z}-\frac{x_{\mu i}}{\sqrt{d}}\mathbf{w}_i-\boldsymbol{\omega}^t_{\mu\to i}\right)^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}\left(\mathbf{z}-\frac{x_{\mu i}}{\sqrt{d}}\mathbf{w}_i-\boldsymbol{\omega}^t_{\mu\to i}\right)}}_{\equiv H_{\mu\to i}},
\end{aligned}
$$

where we defined the mean and variance, depending on the node i:

$$
\boldsymbol{\omega}^t_{\mu\to i} \equiv \frac{1}{\sqrt{d}}\sum_{j\neq i}^{d} x_{\mu j}\hat{\mathbf{w}}^t_{j\to\mu}\,, \qquad\qquad \mathbf{V}^t_{\mu\to i} \equiv \frac{1}{d}\sum_{j\neq i}^{d} x^2_{\mu j}\hat{\mathbf{C}}^t_{j\to\mu}\,.
$$

Again, in the limit $d \to \infty$, the term $H_{\mu\to i}$ can be expanded as

$$
\begin{aligned}
H_{\mu\to i} &\simeq e^{-\frac{1}{2}\left(\mathbf{z}-\boldsymbol{\omega}^t_{\mu\to i}\right)^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}\left(\mathbf{z}-\boldsymbol{\omega}^t_{\mu\to i}\right)} \\
&\qquad \times \left(1 + \frac{x_{\mu i}}{\sqrt{d}}\mathbf{w}_i^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}(\mathbf{z}-\boldsymbol{\omega}^t_{\mu\to i}) - \frac{1}{2}\frac{x^2_{\mu i}}{d}\mathbf{w}_i^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}\mathbf{w}_i \right. \\
&\qquad \left. + \frac{1}{2}\frac{x^2_{\mu i}}{d}\mathbf{w}_i^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}(\mathbf{z}-\boldsymbol{\omega}^t_{\mu\to i})(\mathbf{z}-\boldsymbol{\omega}^t_{\mu\to i})^\intercal (\mathbf{V}^t_{\mu\to i})^{-1}\mathbf{w}_i\right).
\end{aligned}
$$

Putting all pieces together, the message $\tilde{m}_{\mu\to i}$ can be expressed using definitions of $\mathbf{f}_{\text{out}}$ and $\partial_{\boldsymbol{\omega}}\mathbf{f}_{\text{out}}$ in Appendix. ??. We finally obtain

$$\begin{aligned}
\tilde{m}^t_{\mu\to i}(\mathbf{w}_i) \sim \frac{1}{\mathscr{Z}_{\mu\to i}} \Bigg\{ & 1 + \frac{x_{\mu i}}{\sqrt{d}} \mathbf{w}_i^\intercal \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) \\
& + \frac{1}{2}\frac{x_{\mu i}^2}{d} \mathbf{w}_i^\intercal \mathbf{f}_{\text{out}} \mathbf{f}_{\text{out}}^\intercal(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) \mathbf{w}_i \\
& + \frac{1}{2}\frac{x_{\mu i}^2}{d} \mathbf{w}_i^\intercal \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) \mathbf{w}_i \Bigg\} \\
= \frac{1}{\mathscr{Z}_{\mu\to i}} & \left\{ 1 + \mathbf{w}_i^\intercal \mathbf{b}^t_{\mu\to i} + \frac{1}{2} \mathbf{w}_i^\intercal \mathbf{b}^t_{\mu\to i} (\mathbf{b}^t_{\mu\to i})^\intercal (\mathbf{w}_i) - \frac{1}{2} \mathbf{w}_i^\intercal \mathbf{A}^t_{\mu\to i} w_i \right\} \\
= & \sqrt{\frac{\det\left(\mathbf{A}^t_{\mu\to i}\right)}{(2\pi)^K}} e^{-\frac{1}{2}\left(\mathbf{w}_i^\intercal - (\mathbf{A}^t_{\mu\to i})^{-1}\mathbf{b}^t_{\mu\to i}\right)^\intercal \mathbf{A}^t_{\mu\to i}\left(\mathbf{w}_i^\intercal - (\mathbf{A}^t_{\mu\to i})^{-1}\mathbf{b}^t_{\mu\to i}\right)}
\end{aligned}$$

with the following definitions of $\mathbf{A}_{\mu\to i}$ and $\mathbf{b}_{\mu\to i}$

$$\begin{aligned}
\mathbf{b}^t_{\mu\to i} &\equiv \frac{x_{\mu i}}{\sqrt{d}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) \,, \\
\mathbf{A}^t_{\mu\to i} &\equiv -\frac{x_{\mu i}^2}{d} \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) \,.
\end{aligned}$$

The set of BP equations can finally be closed over the Gaussian messages $\{m_{i\to\mu}\}_{i=1..d}^{\mu=1..n}$ according to

$$\begin{aligned}
m^{t+1}_{i\to\mu}(\mathbf{w}_i) = \frac{1}{\mathscr{Z}_{i\to\mu}} \mathrm{P_w}(\mathbf{w}_i) \prod_{\nu\neq\mu}^{n} & \sqrt{\frac{\det\left(\mathbf{A}^t_{\nu\to i}\right)}{(2\pi)^K}} \\
& \times e^{-\frac{1}{2}\left(\mathbf{w}_i - (\mathbf{A}^t_{\nu\to i})^{-1}\mathbf{b}^t_{\nu\to i}\right)^\intercal \mathbf{A}^t_{\nu\to i}\left(w_i - (\mathbf{A}^t_{\nu\to i})^{-1}\mathbf{b}^t_{\nu\to i}\right)} .
\end{aligned}$$

In the end, computing the mean and variance of the product of Gaussians, the messages are updated using $\mathbf{f}_{\text{w}}$ and $\partial_{\boldsymbol{\gamma}}\mathbf{f}_{\text{w}}$, defined in Appendix. ??, according to

$$\hat{\mathbf{w}}^{t+1}_{i\to\mu} = \mathbf{f}_{\text{w}}(\boldsymbol{\gamma}^t_{\mu\to i}, \boldsymbol{\Lambda}^t_{\mu\to i}) \,, \qquad \hat{\mathbf{C}}^{t+1}_{i\to\mu} = \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\text{w}}(\boldsymbol{\gamma}^t_{\mu\to i}, \boldsymbol{\Lambda}^t_{\mu\to i}) \,,$$

with

$$\boldsymbol{\gamma}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n} \mathbf{b}^t_{\nu\to i} \,, \qquad \boldsymbol{\Lambda}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n} \mathbf{A}^t_{\nu\to i} \,.$$

Summary of the rBP equations In the end, the rBP equations are simply the following set of equations:

$$
\begin{aligned}
&\hat{\mathbf{w}}^{t+1}_{i\to\mu} = \mathbf{f}_{\mathrm{w}}(\boldsymbol{\gamma}^t_{\mu\to i}, \boldsymbol{\Lambda}^t_{\mu\to i})\,, && \hat{\mathbf{C}}^{t+1}_{i\to\mu} = \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\mathrm{w}}(\boldsymbol{\gamma}^t_{\mu\to i}, \boldsymbol{\Lambda}^t_{\mu\to i})\\
&\boldsymbol{\gamma}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n} \mathbf{b}^t_{\nu\to i}\,, && \boldsymbol{\Lambda}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n} \mathbf{A}^t_{\nu\to i}\\
&\mathbf{b}^t_{\mu\to i} = \frac{x_{\mu i}}{\sqrt{d}}\mathbf{f}_{\mathrm{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i})\,, && \\
&\mathbf{A}^t_{\mu\to i} = -\frac{x^2_{\mu i}}{d}\partial_{\boldsymbol{\omega}}\mathbf{f}_{\mathrm{out}}(y_\mu, \boldsymbol{\omega}^t_{\mu\to i}, \mathbf{V}^t_{\mu\to i}) && \\
&\boldsymbol{\omega}^t_{\mu\to i} = \sum_{j\neq i}^{d} \frac{x_{\mu j}}{\sqrt{d}}\hat{\mathbf{w}}^t_{j\to\mu}\,, && \mathbf{V}^t_{\mu\to i} = \sum_{j\neq i}^{d} \frac{x^2_{\mu j}}{d}\hat{\mathbf{C}}^t_{j\to\mu}\,.
\end{aligned}
\tag{63}
$$

4.4 AMP algorithm

The rBP equations eq. (63) contains $\Theta(d^2)$ messages. However all the messages depend weakly on the target node. The missing message is negligible in the limit $d \to \infty$, that allows us to expand the rBP around the *full* messages:

$$
\begin{aligned}
&\boldsymbol{\omega}^t_{\mu} \equiv \sum_{j=1}^{d} \frac{x_{\mu j}}{\sqrt{d}}\hat{\mathbf{w}}^t_{j\to\mu}\,, && \mathbf{V}^t_{\mu} \equiv \sum_{j=1}^{d} \frac{x^2_{\mu j}}{d}\hat{\mathbf{C}}^t_{j\to\mu}\\
&\boldsymbol{\gamma}^t_{i} \equiv \sum_{\mu=1}^{n} \mathbf{b}^t_{\mu\to i}\,, && \boldsymbol{\Lambda}^t_{i} \equiv \sum_{\mu=1}^{n} \mathbf{A}^t_{\mu\to i}\,.
\end{aligned}
\tag{64}
$$

By completing the sum, we naturally remove the target node dependence and reduce the set of messages to $\Theta(d)$. Let us now perform the expansion of the rBP messages.

Partial covariance $\mathbf{f}_{\mathrm{w}}$: $\boldsymbol{\Lambda}^t_{\mu\to i}$

$$
\begin{aligned}
\boldsymbol{\Lambda}^t_{\mu\to i} &= \sum_{\nu\neq\mu}^{n} \mathbf{A}^t_{\nu\to i} = \sum_{\nu=1}^{n} \mathbf{A}^t_{\nu\to i} - \mathbf{A}^t_{\mu\to i}\\
&= \boldsymbol{\Lambda}^t_i - \mathbf{A}^t_{\mu\to i} = \boldsymbol{\Lambda}^t_i + \Theta\left(\frac{1}{d}\right)\,.
\end{aligned}
$$

Partial mean $\mathbf{f}_{\mathrm{w}}$: $\boldsymbol{\gamma}^t_{\mu\to i}$

$$
\boldsymbol{\gamma}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n} \mathbf{b}^t_{\nu\to i} = \sum_{\nu=1}^{n} \mathbf{b}^t_{\nu\to i} - \mathbf{b}^t_{\mu\to i} = \boldsymbol{\gamma}^t_i - \mathbf{b}^t_{\mu\to i} + \Theta\left(\frac{1}{d}\right)\,.
$$

Mean $\hat{\mathbf{w}}_{i\to\mu}^{t+1}$ update

$$\begin{aligned}\hat{\mathbf{w}}_{i\to\mu}^{t+1} &= \mathbf{f}_{\mathrm{w}}(\boldsymbol{\gamma}_{\mu\to i}^{t}, \boldsymbol{\Lambda}_{\mu\to i}^{t}) = \mathbf{f}_{\mathrm{w}}\left(\boldsymbol{\gamma}_{i}^{t} - \mathbf{b}_{\mu\to i}^{t}, \boldsymbol{\Lambda}_{i}^{t}\right) + \Theta\left(\frac{1}{d}\right) \\ &= \mathbf{f}_{\mathrm{w}}\left(\boldsymbol{\gamma}_{i}^{t}, \boldsymbol{\Lambda}_{i}^{t}\right) - \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\mathrm{w}}\left(\boldsymbol{\gamma}_{i}^{t}, \boldsymbol{\Lambda}_{i}^{t}\right)\mathbf{b}_{\mu\to i}^{t} + \Theta\left(\frac{1}{d}\right) \\ &= \hat{\mathbf{w}}_{i}^{t+1} - \hat{\mathbf{C}}_{i}^{t+1}\mathbf{b}_{\mu\to i}^{t} + \Theta\left(\frac{1}{d}\right) \\ &= \hat{\mathbf{w}}_{i}^{t+1} - \frac{x_{\mu i}}{\sqrt{d}}\hat{\mathbf{C}}_{i}^{t+1}\mathbf{f}_{\mathrm{out}}(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t}) + \Theta\left(\frac{1}{d}\right).\end{aligned}$$

where we defined the prior updates

$$\hat{\mathbf{w}}_{i}^{t+1} \equiv \mathbf{f}_{\mathrm{w}}\left(\boldsymbol{\gamma}_{i}^{t}, \boldsymbol{\Lambda}_{i}^{t}\right), \qquad \hat{\mathbf{C}}_{i}^{t+1} \equiv \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\mathrm{w}}\left(\boldsymbol{\gamma}_{i}^{t}, \boldsymbol{\Lambda}_{i}^{t}\right),$$

and used the fact that $\mathbf{b}_{\mu\to i}^{t} \simeq \frac{x_{\mu i}}{\sqrt{d}}\hat{\mathbf{C}}_{i}^{t+1}\mathbf{f}_{\mathrm{out}}(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t})$ by expanding the equation over $\mathbf{b}_{\mu\to i}^{t}$ in (63).

Covariance $\hat{\mathbf{C}}_{i\to\mu}^{t+1}$ update

$$\begin{aligned}\hat{\mathbf{C}}_{i\to\mu}^{t+1} &= \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\mathrm{w}}(\boldsymbol{\gamma}_{\mu\to i}^{t}, \boldsymbol{\Lambda}_{\mu\to i}^{t}) \\ &\simeq \partial_{\boldsymbol{\gamma}}\mathbf{f}_{\mathrm{w}}(\boldsymbol{\gamma}_{i}^{t}, \boldsymbol{\Lambda}_{i}^{t}) + \Theta\left(\frac{1}{\sqrt{d}}\right) = \hat{\mathbf{C}}_{i}^{t+1} + \Theta\left(\frac{1}{\sqrt{d}}\right).\end{aligned}$$

Channel update function $\mathbf{f}_{\mathrm{out}}(y_{\mu}, \boldsymbol{\omega}_{\mu\to i}^{t}, \mathbf{V}_{\mu\to i}^{t})$

$$\begin{aligned}&\mathbf{f}_{\mathrm{out}}(y_{\mu}, \boldsymbol{\omega}_{\mu\to i}^{t}, \mathbf{V}_{\mu\to i}^{t}) = \mathbf{f}_{\mathrm{out}}\left(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t} - \frac{x_{\mu i}}{\sqrt{d}}\hat{\mathbf{w}}_{i\to\mu}^{t}, \mathbf{V}_{\mu}^{t} - \frac{x_{\mu i}^{2}}{d}\hat{\mathbf{C}}_{i\to l}^{t}\right) \\ &= \mathbf{f}_{\mathrm{out}}\left(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t}\right) - \frac{x_{\mu i}}{\sqrt{d}}\partial_{\boldsymbol{\omega}}\mathbf{f}_{\mathrm{out}}\left(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t}\right)\underbrace{\hat{\mathbf{w}}_{i\to\mu}^{t}}_{=\hat{\mathbf{w}}_{i}^{t}+\Theta\left(\frac{1}{\sqrt{d}}\right)} + \Theta\left(\frac{1}{d}\right) \\ &= \mathbf{f}_{\mathrm{out}}\left(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t}\right) - \frac{x_{\mu i}}{\sqrt{d}}\partial_{\boldsymbol{\omega}}\mathbf{f}_{\mathrm{out}}\left(y_{\mu}, \boldsymbol{\omega}_{\mu}^{t}, \mathbf{V}_{\mu}^{t}\right)\hat{\mathbf{w}}_{i}^{t} + \Theta\left(\frac{1}{d}\right).\end{aligned}$$

Covariance $\mathbf{f}_{\mathrm{out}}$: $\mathbf{V}_{\mu}^{t}$

$$\mathbf{V}_{\mu}^{t} \equiv \sum_{j=1}^{d}\frac{x_{\mu j}^{2}}{d}\hat{\mathbf{C}}_{j\to\mu}^{t} = \sum_{j=1}^{d}\frac{x_{\mu j}^{2}}{d}\hat{\mathbf{C}}_{j\to\mu}^{t} + \Theta\left(\frac{1}{d^{3/2}}\right).$$

Mean $\mathbf{f}_{\text{out}}$: $\boldsymbol{\omega}_\mu^t$

$$\begin{aligned}\boldsymbol{\omega}_\mu^t &= \sum_{i=1}^d \frac{x_{\mu i}}{\sqrt{d}} \hat{\mathbf{w}}_{i\to\mu}^t \\ &= \sum_{i=1}^d \frac{x_{\mu i}}{\sqrt{d}} \left(\hat{\mathbf{w}}_i^t - x_{\mu i} \hat{\mathbf{C}}_i^t \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_\mu^{t-1}, \mathbf{V}_\mu^{t-1}) + \Theta\left(\frac{1}{d}\right) \right) \\ &= \sum_{i=1}^d \frac{x_{\mu i}}{\sqrt{d}} \hat{\mathbf{w}}_i^t - \sum_{i=1}^d \frac{x_{\mu i}^2}{d} \hat{\mathbf{C}}_i^t \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_\mu^{t-1}, \mathbf{V}_\mu^{t-1}) + \Theta\left(\frac{1}{d^{3/2}}\right).\end{aligned}$$

Covariance $\mathbf{f}_{\text{w}}$: $\boldsymbol{\Lambda}_i^t$

$$\begin{aligned}\boldsymbol{\Lambda}_i^t &\equiv \sum_{\mu=1}^n \mathbf{A}_{\mu\to i}^t = \sum_{\nu=1}^n -\frac{x_{\mu i}^2}{d} \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_{\mu\to i}^t, \mathbf{V}_{\mu\to i}^t) \\ &= \sum_{\mu=1}^n -\frac{x_{\mu i}^2}{d} \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_\mu^t, \mathbf{V}_\mu^t) + \Theta\left(\frac{1}{d^{3/2}}\right).\end{aligned}$$

Mean $\mathbf{f}_{\text{w}}$: $\boldsymbol{\gamma}_i^t$

$$\begin{aligned}\boldsymbol{\gamma}_i^t &= \sum_{\mu=1}^n \mathbf{b}_{\mu\to i}^t = \sum_{\mu=1}^n \frac{x_{\mu i}}{\sqrt{d}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_{\mu\to i}^t, \mathbf{V}_{\mu\to i}^t) \\ &= \sum_{\mu=1}^n \frac{x_{\mu i}}{\sqrt{d}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_\mu^t, \mathbf{V}_\mu^t) \\ &\quad - \frac{x_{\mu i}^2}{d} \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(y_\mu, \boldsymbol{\omega}_\mu^t, \mathbf{V}_\mu^t) \hat{\mathbf{w}}_i^t + \Theta\left(\frac{1}{d^{3/2}}\right).\end{aligned}$$

Summary - AMP algorithm

We finally obtain the AMP algorithm as a reduced set of $\Theta(d)$ messages in Algo. 1.

4.5 State evolution equations of AMP

In this section we derive the behavior of the AMP algorithm in Algo. 1 in the thermodynamic limit $d \to \infty$. This average asymptotic behavior can be tracked with some overlap parameters at time t, $\mathbf{m}^t$, $\mathbf{q}^t$, $\boldsymbol{\Sigma}^t$, that respectively measure the correlation of the AMP estimator with the ground truth, the norms of student and teacher weights, the estimator variance and the second moment of the teacher network $\boldsymbol{\rho}_{\text{w}^\star}$, defined by

$$\begin{aligned}&\mathbf{m}^t \equiv \mathbb{E} \lim_{d\to\infty} \frac{1}{d} \hat{\mathbf{W}}^{t\intercal} \mathbf{W}^\star, \quad \mathbf{q}^t \equiv \mathbb{E} \lim_{d\to\infty} \frac{1}{d} \hat{\mathbf{W}}^{t\intercal} \hat{\mathbf{W}}^t, \\ &\boldsymbol{\Sigma}^t \equiv \mathbb{E} \lim_{d\to\infty} \frac{1}{d} \sum_{i=1}^d \hat{\mathbf{C}}_i^t, \qquad \boldsymbol{\rho}_{\text{w}^\star} \equiv \mathbb{E} \lim_{d\to\infty} \frac{1}{d} \mathbf{W}^{\star\intercal} \mathbf{W}^\star,\end{aligned} \tag{65}$$

Input: vector $\mathbf{y} \in \mathbb{R}^n$ and matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$:
Initialize: $\hat{\mathbf{w}}_i$, $\mathbf{f}_{\text{out},\mu} \in \mathbb{R}^K$ and $\hat{\mathbf{V}}_i$, $\partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out},\mu} \in \mathbb{R}^{K \times K}$ for $1 \leq i \leq d$ and $1 \leq \mu \leq n$ at $t = 0$.
repeat
Channel: Update the mean $\omega_\mu \in \mathbb{R}^K$ and variance $V_\mu \in \mathbb{R}^{K \times K}$:

$$\mathbf{V}_\mu^t = \sum_{i=1}^{d} \frac{x_{\mu j}^2}{d} \hat{\mathbf{C}}_i^t$$

$$\boldsymbol{\omega}_\mu^t = \sum_{i=1}^{d} \frac{x_{\mu i}}{\sqrt{d}} \hat{\mathbf{w}}_i^t - \mathbf{V}_\mu^t \mathbf{f}_{\text{out},\mu}^{t-1} ,$$

Update $\mathbf{f}_{\text{out},\mu}$ and $\partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out},\mu}$:

$$\mathbf{f}_{\text{out},\mu}^t = \mathbf{f}_{\text{out}} \left(y_\mu, \boldsymbol{\omega}_\mu^t, \mathbf{V}_\mu^t \right), \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out},\mu}^t = \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}} \left(y_\mu, \boldsymbol{\omega}_\mu^t, \mathbf{V}_\mu^t \right)$$

Prior: Update the mean $\boldsymbol{\gamma}_i \in \mathbb{R}^K$ and variance $\boldsymbol{\Lambda}_i \in \mathbb{R}^{K \times K}$:

$$\boldsymbol{\Lambda}_i^t = \sum_{\mu=1}^{n} -\frac{x_{\mu i}^2}{d} \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out},\mu}$$

$$\boldsymbol{\gamma}_i^t = \sum_{\mu=1}^{n} \frac{x_{\mu i}}{\sqrt{d}} \mathbf{f}_{\text{out},\mu} + \boldsymbol{\Lambda}_i^t \hat{\mathbf{w}}_i^t ,$$

Update the estimated marginals $\hat{\mathbf{w}}_i \in \mathbb{R}$ and $\hat{\mathbf{C}}_i \in \mathbb{R}^+$:

$$\hat{\mathbf{w}}_i^{t+1} = \mathbf{f}_{\text{w}} \left(\boldsymbol{\gamma}_i^t, \boldsymbol{\Lambda}_i^t \right), \hat{\mathbf{C}}_i^{t+1} = \partial_{\boldsymbol{\gamma}} \mathbf{f}_{\text{w}} \left(\boldsymbol{\gamma}_i^t, \boldsymbol{\Lambda}_i^t \right)$$

$t \leftarrow t + 1$
until Convergence on $\hat{\mathbf{w}}_i$, $\hat{\mathbf{C}}_i$.
Output: $\{\hat{\mathbf{w}}_i\}_{i=1}^d$ and $\{\hat{\mathbf{C}}_i\}_{i=1}^d$.

Algorithm 1: Approximate Message Passing algorithm for committee machines.

where the expectation is over ground truth signals $\mathbf{W}^\star$ and input data $\mathbf{X}$. The aim is to derive the asymptotic behavior of these overlap parameters, called SE. The idea is simply to compute the overlap distributions starting with the set of rBP equations in (63).

4.5.1 Messages distribution

In order to get the asymptotic behavior of the overlap parameters, we first need to compute the distribution of $\mathbf{W}^{t+1}$ and, as a result, of the mean $\boldsymbol{\gamma}^t_{\mu\to i}$ and covariance $\boldsymbol{\Lambda}^t_{\mu\to i}$. Recalling that under the BP assumption incoming messages are independent, the messages $\boldsymbol{\omega}^t_{\mu\to i}$ and $\mathbf{z}_\mu$ are the sum of independent variables and follow Gaussian distributions. However, these two variables are correlated and we need to compute correctly the covariance matrix.

To compute it, we will make use of different ingredients. First, we recall that in the T-S scenario, the output has been generated by a teacher such that $\forall \mu \in [\![n]\!],\ y_\mu = \varphi_{\text{out}^\star}\left(\frac{1}{\sqrt{d}}\mathbf{x}_\mu^\intercal \mathbf{W}^\star\right)$. By convenience, we define $\mathbf{z}_\mu \equiv \frac{1}{\sqrt{d}}\mathbf{x}_\mu^\intercal \mathbf{W}^\star = \frac{1}{\sqrt{d}}\sum_{i=1}^d x_{\mu i}\mathbf{w}_i^\star$ and $z_{\mu\to i} \equiv \frac{1}{\sqrt{d}}\sum_{j\neq i}^d x_{\mu j}\mathbf{w}_j^\star$. Second, in the case the input data are i.i.d Gaussian, we have $\mathbb{E}_{\mathbf{X}}[x_{\mu i}] = 0$ and $\mathbb{E}_{\mathbf{X}}[x_{\mu i}^2] = 1$.

Partial mean $\mathbf{f}_{\text{out}}$: $\boldsymbol{\omega}^t_{\mu\to i}$ Let's compute the first two moments, using expansions of the rBP equations (63):

$$\mathbb{E}\left[\boldsymbol{\omega}^t_{\mu\to i}\right] = \frac{1}{\sqrt{d}}\sum_{j\neq i}^d \mathbb{E}_{\mathbf{X}}\left[x_{\mu j}\right]\mathbb{E}\left[\hat{\mathbf{w}}^t_{j\to\mu}\right] = \mathbf{0}\,,$$

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\omega}^t_{\mu\to i}\left(\boldsymbol{\omega}^t_{\mu\to i}\right)^\intercal\right] &= \frac{1}{d}\sum_{j\neq i}^d \mathbb{E}_{\mathbf{X}}\left[x_{\mu j}^2\right]\mathbb{E}\left[\hat{\mathbf{w}}^t_{j\to\mu}\left(\hat{\mathbf{w}}^t_{j\to\mu}\right)^\intercal\right]\\ &= \frac{1}{d}\sum_{i=1}^d \mathbb{E}_{\mathbf{X}}\left[x_{\mu j}^2\right]\mathbb{E}\left[\hat{\mathbf{w}}^t_i(\hat{\mathbf{w}}_i)^\intercal\right] + \Theta\left(d^{-3/2}\right) \underset{d\to\infty}{\longrightarrow} \mathbf{q}^t\,.\end{aligned}$$

Hidden variable $\mathbf{z}_\mu$ Let us compute the first moments of the hidden variable $\mathbf{z}_\mu$:

$$\begin{aligned}\mathbb{E}\left[\mathbf{z}_\mu\right] &= \frac{1}{\sqrt{d}}\sum_{i=1}^d \mathbb{E}_{\mathbf{X}}\left[x_{\mu i}\right]\mathbb{E}_{\mathbf{W}^\star}\left[\mathbf{w}_i^\star\right] = \mathbf{0}\,,\\ \mathbb{E}\left[\mathbf{z}_\mu\mathbf{z}_\mu^\intercal\right] &= \frac{1}{d}\sum_{i=1}^d \mathbb{E}_{\mathbf{X}}\left[x_{\mu i}^2\right]\mathbb{E}_{\mathbf{W}^\star}\left[\mathbf{w}_i^\star(\mathbf{w}_i^\star)^\intercal\right] \underset{d\to\infty}{\longrightarrow} \boldsymbol{\rho}_{\mathrm{w}^\star}\,.\end{aligned}$$

Correlation between $\mathbf{z}_\mu$ and $\boldsymbol{\omega}^t_{\mu\to i}$ The cross correlation is given by

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\omega}^t_{\mu\to i}\mathbf{z}_\mu^\intercal\right] = \frac{1}{d}\sum_{j\neq i,k=1}^{d}\mathbb{E}_{\mathbf{X}}\left[x_{\mu j}x_{\mu k}\right]\mathbb{E}_{\mathbf{W}^\star}\left[\hat{\mathbf{w}}^t_{j\to\mu}(\mathbf{w}^\star_k)^\intercal\right]\\
&= \frac{1}{d}\sum_{j\neq i}^{d}\mathbb{E}_{\mathbf{W}^\star}\left[\hat{\mathbf{w}}^t_{j\to\mu}(\mathbf{w}^\star_j)^\intercal\right] = \frac{1}{d}\sum_{i}^{d}\mathbb{E}_{\mathbf{W}^\star}\left[\hat{\mathbf{w}}^t_{i}(\mathbf{w}^\star_i)^\intercal\right] + \Theta\left(d^{-3/2}\right)\\
&\quad\underset{d\to\infty}{\longrightarrow}\mathbf{m}^t\,.
\end{aligned}$$

Hence asymptotically the random vector $(\mathbf{z}_\mu, \boldsymbol{\omega}^t_{\mu\to i})$ follow a multivariate Gaussian distribution with covariance matrix $\mathbf{Q}^t = \begin{bmatrix}\boldsymbol{\rho}_{\mathrm{w}^\star} & \mathbf{m}^t\\ \mathbf{m}^t & \mathbf{q}^t\end{bmatrix} \in \mathbb{R}^{(2K)\times(2K)}$.

Partial variance $\mathbf{f}_{\mathbf{out}}$: $\mathbf{V}_{\mu\to i}$ $\mathbf{V}_{\mu\to i}$ concentrates around its mean:

$$\mathbb{E}\left[\mathbf{V}^t_{\mu\to i}\right] = \frac{1}{d}\sum_{j\neq i}^{d}\mathbb{E}_{\mathbf{X}}\left[x^2_{\mu j}\right]\hat{\mathbf{C}}^t_{j\to\mu} = \frac{1}{d}\sum_{i}^{d}\hat{\mathbf{C}}^t_i + \Theta(d^{-3/2}) \underset{d\to\infty}{\longrightarrow}\boldsymbol{\Sigma}^t\,.$$

Ad-hoc overlaps Let us define some other ad-hoc order parameters, that will appear in the following:

$$\begin{aligned}
\hat{\mathbf{q}}^t &\equiv \alpha\mathbb{E}_{\boldsymbol{\omega},\mathbf{z}}\left[\mathbf{f}_{\mathrm{out}}(\varphi_{\mathrm{out}^\star}(\mathbf{z}),\boldsymbol{\omega},\boldsymbol{\Sigma}^t)^{\otimes 2}\right]\,,\\
\hat{\mathbf{m}}^t &\equiv \alpha\mathbb{E}_{\boldsymbol{\omega},\mathbf{z}}\left[\partial_{\mathbf{z}}\mathbf{f}_{\mathrm{out}}(\varphi_{\mathrm{out}^\star}(\mathbf{z}),\boldsymbol{\omega},\boldsymbol{\Sigma}^t)\right]\,,\\
\hat{\boldsymbol{\chi}}^t &\equiv \alpha\mathbb{E}_{\boldsymbol{\omega},\mathbf{z}}\left[-\partial_{\boldsymbol{\omega}}\mathbf{f}_{\mathrm{out}}(\varphi_{\mathrm{out}^\star}(\mathbf{z}),\boldsymbol{\omega},\boldsymbol{\Sigma}^t)\right]\,.
\end{aligned} \tag{66}$$

Partial mean $\mathbf{f}_{\mathrm{w}}$: $\boldsymbol{\gamma}^t_{\mu\to i}$ Using the expression $y_\nu = \varphi_{\mathrm{out}^\star}\left(\mathbf{z}_{\nu\to i} + \frac{1}{\sqrt{d}}x_{\nu i}\mathbf{w}^\star_i\right)$ and expanding $\boldsymbol{\gamma}^t_{\mu\to i}$, we obtain

$$\begin{aligned}
&\boldsymbol{\gamma}^t_{\mu\to i} = \sum_{\nu\neq\mu}^{n}\mathbf{b}^t_{\nu\to i} = \sum_{\nu\neq\mu}^{n}\frac{x_{\nu i}}{\sqrt{d}}\mathbf{f}_{\mathrm{out}}\left(y_\nu,\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}\right)\\
&= \frac{1}{\sqrt{d}}\sum_{\nu\neq\mu}^{n}x_{\nu i}\mathbf{f}_{\mathrm{out}}\left(\varphi_{\mathrm{out}^\star}\left(\mathbf{z}_{\nu\to i}\right),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}\right)\\
&+\frac{1}{d}\sum_{\nu\neq\mu}^{n}x^2_{\nu i}\partial_{\mathbf{z}}\mathbf{f}_{\mathrm{out}}\left(\varphi_{\mathrm{out}^\star}\left(\mathbf{z}_{\nu\to i}\right),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}\right)\mathbf{w}^\star_i\,.
\end{aligned}$$

Thus, taking the average

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\gamma}^t_{\mu\to i}\right] &= \mathbf{0} + \frac{1}{d}\sum_{\nu\neq\mu}^{n}\mathbb{E}_{\mathbf{z},\boldsymbol{\omega}}\left[\partial_{\mathbf{z}}\mathbf{f}_{\text{out}}\left(\varphi_{\text{out}^\star}\left(\mathbf{z}_{\nu\to i}\right),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}\right)\right]\mathbf{w}^\star_i \\ &\underset{d\to\infty}{\longrightarrow} \hat{\mathbf{m}}^t\mathbf{w}^\star_i\,, \\ \mathbb{E}\left[\left(\boldsymbol{\gamma}^t_{\mu\to i}\right)^{\otimes 2}\right] &= \frac{1}{d}\sum_{\nu\neq\mu}^{n}\mathbb{E}_{\mathbf{z},\boldsymbol{\omega}}\left[\mathbf{f}_{\text{out}}\left(\varphi_{\text{out}^\star}\left(\mathbf{z}_{\nu\to i}\right),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}\right)^{\otimes 2}\right] \\ &\underset{d\to\infty}{\longrightarrow} \hat{\mathbf{q}}^t\,.\end{aligned}$$

Hence $\boldsymbol{\gamma}^t_{\mu\to i}\sim\hat{\mathbf{m}}^t\mathbf{w}^\star_i+(\hat{\mathbf{q}}^t)^{1/2}\boldsymbol{\xi}$ with $\boldsymbol{\xi}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_K)$.

Partial covariance $\mathbf{f}_{\text{w}}$: $\boldsymbol{\Lambda}^t_{\mu\to i}$

$$\begin{aligned}\boldsymbol{\Lambda}^t_{\mu\to i} &= \sum_{\nu\neq\mu}^{n}\mathbf{A}^t_{\nu\to i} = -\frac{1}{d}\sum_{\nu\neq\mu}^{n}x^2_{\mu i}\partial_{\boldsymbol{\omega}}\mathbf{f}_{\text{out}}(y_\nu,\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}) \\ &= -\frac{1}{d}\sum_{\nu\neq\mu}^{n}x^2_{\mu i}\partial_{\boldsymbol{\omega}}\mathbf{f}_{\text{out}}(\varphi_{\text{out}^\star}(\mathbf{z}_{\nu\to i}),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i}) + \Theta\left(d^{-3/2}\right)\end{aligned}$$

and taking the average

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\Lambda}^t_{\mu\to i}\right] &= -\frac{1}{d}\sum_{\nu\neq\mu}^{n}\mathbb{E}_{\mathbf{z},\boldsymbol{\omega}}\left[\partial_{\boldsymbol{\omega}}\mathbf{f}_{\text{out}}(\varphi_{\text{out}^\star}(\mathbf{z}_{\nu\to i}),\boldsymbol{\omega}^t_{\nu\to i},\mathbf{V}^t_{\nu\to i})\right] \\ &\underset{d\to\infty}{\longrightarrow}\hat{\boldsymbol{\chi}}^t\,,\end{aligned}$$

so that in the thermodynamic limit $\boldsymbol{\Lambda}^t_{\mu\to i}\sim\hat{\boldsymbol{\chi}}^t$.

4.5.2 Summary of the SE - mismatched setting

Using the definition of the overlaps in (65) at time $t+1$ and the message distributions, we finally obtain the set of SE equations of the AMP algorithm in Algo. 1 in the mismatched setting:

$$\begin{aligned}\mathbf{m}^{t+1} &\equiv \mathbb{E}\lim_{d\to\infty}\frac{1}{d}\hat{\mathbf{W}}^{t+1\intercal}\hat{\mathbf{W}}^\star = \mathbb{E}_{\mathbf{w}^\star,\boldsymbol{\xi}}\left[\mathbf{f}_{\text{w}}\left(\hat{\mathbf{m}}^t\mathbf{w}^\star+(\hat{\mathbf{q}}^t)^{1/2}\boldsymbol{\xi},\hat{\boldsymbol{\chi}}^t\right)\mathbf{w}^{\star\intercal}\right], \\ \mathbf{q}^{t+1} &\equiv \mathbb{E}\lim_{d\to\infty}\frac{1}{d}\hat{\mathbf{W}}^{t+1\intercal}\hat{\mathbf{W}}^{t+1} = \mathbb{E}_{\mathbf{w}^\star,\boldsymbol{\xi}}\left[\mathbf{f}_{\text{w}}\left(\hat{\mathbf{m}}^t\mathbf{w}^\star+(\hat{\mathbf{q}}^t)^{1/2}\boldsymbol{\xi},\hat{\boldsymbol{\chi}}^t\right)^{\otimes 2}\right], \\ \boldsymbol{\Sigma}^{t+1} &\equiv \mathbb{E}\lim_{d\to\infty}\frac{1}{d}\sum_{i=1}^{d}\hat{\mathbf{C}}^{t+1}_i = \mathbb{E}_{\mathbf{w}^\star,\boldsymbol{\xi}}\left[\partial_{\boldsymbol{\gamma}}\mathbf{f}_{\text{w}}\left(\hat{\mathbf{m}}^t\mathbf{w}^\star+(\hat{\mathbf{q}}^t)^{1/2}\boldsymbol{\xi},\hat{\boldsymbol{\chi}}^t\right)\right],\end{aligned} \tag{67}$$

and

$$\begin{aligned}
\hat{\mathbf{q}}^t &= \alpha \int_{\mathbb{R}^K} \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{z}\, \mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}_{2K}, \mathbf{Q}^t\right) \mathbf{f}_{\text{out}}(\varphi_{\text{out}^\star}(\mathbf{z}), \boldsymbol{\omega}, \boldsymbol{\Sigma}^t)^{\otimes 2} \\
\hat{\mathbf{m}}^t &= \alpha \int_{\mathbb{R}^K} \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{z}\, \mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}_{2K}, \mathbf{Q}^t\right)) \partial_{\mathbf{z}} \mathbf{f}_{\text{out}}(\varphi_{\text{out}^\star}(\mathbf{z}), \boldsymbol{\omega}, \boldsymbol{\Sigma}^t)\,, \\
\hat{\boldsymbol{\chi}}^t &= -\alpha \int_{\mathbb{R}^K} \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{z}\, \mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}_{2K}, \mathbf{Q}^t\right) \partial_{\boldsymbol{\omega}} \mathbf{f}_{\text{out}}(\varphi_{\text{out}^\star}(\mathbf{z}), \boldsymbol{\omega}, \boldsymbol{\Sigma}^t)\,.
\end{aligned} \tag{68}$$

with $\mathbf{Q}^t = \begin{bmatrix} \boldsymbol{\rho}_{\text{w}^\star} & \mathbf{m}^t \\ \mathbf{m}^t & \mathbf{q}^t \end{bmatrix} \in \mathbb{R}^{(2K)\times(2K)}$.

4.5.3 Summary of the SE - Bayes-optimal setting

In the Bayes-optimal setting, the student $\mathrm{P}_{\text{w}} = \mathrm{P}_{\text{w}^\star}$ and $\mathrm{P}_{\text{out}} = \mathrm{P}_{\text{out}^\star}$, so that we have $\mathbf{f}_{\text{w}} = \mathbf{f}_{\text{w}^\star}$ and $\mathbf{f}_{\text{out}} = \mathbf{f}_{\text{out}^\star}$. Moreover, the Nishimori conditions, recalled in Appendix. ??, imply that

$$\mathbf{m}^t = \mathbf{q}^t \equiv \mathbf{q}_{\text{b}}^t\,, \qquad \hat{\mathbf{q}}^t = \hat{\mathbf{m}}^t = \hat{\boldsymbol{\chi}}^t \equiv \hat{\mathbf{q}}_{\text{b}}^t\,, \qquad \boldsymbol{\Sigma}^t = \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}^t\,.$$

Therefore the set of SE equations simplify and reduce to

$$\begin{aligned}
\mathbf{q}_{\text{b}}^{t+1} &= \mathbb{E}_{\mathbf{w}^\star, \boldsymbol{\xi}} \left[\mathbf{f}_{\text{w}^\star}\left(\hat{\mathbf{q}}_{\text{b}}^t \mathbf{w}^\star + (\hat{\mathbf{q}}_{\text{b}}^t)^{1/2} \boldsymbol{\xi}, \hat{\mathbf{q}}_{\text{b}}^t \right)^{\otimes 2} \right] \\
\hat{\mathbf{q}}_{\text{b}}^t &= \alpha \int_{\mathbb{R}^K} \int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{z}\, \mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}_{2K}, \mathbf{Q}_{\text{b}}^t\right) \mathbf{f}_{\text{out}^\star}(\varphi_{\text{out}^\star}(\mathbf{z}), \boldsymbol{\omega}, \boldsymbol{\rho}_{\text{w}^\star} - \mathbf{q}_{\text{b}}^t)^{\otimes 2}
\end{aligned} \tag{69}$$

with the simplified covariance matrix $\mathbf{Q}_{\text{b}}^t = \begin{bmatrix} \boldsymbol{\rho}_{\text{w}^\star} & \mathbf{q}_{\text{b}}^t \\ \mathbf{q}_{\text{b}}^t & \mathbf{q}_{\text{b}}^t \end{bmatrix}$.

5 Consistency between AMP and the replica computation

Very surprisingly, the SE of the AMP algorithm can be obtained in a convoluted and more rapid way. It turns out that in the Bayes-optimal setting, AMP performs a gradient ascent on the RS free entropy in (50). Meaning that at convergence, and under good initialization, the AMP overlaps are given by the saddle point equations of the RS free entropy $\Phi^{(\text{rs})}$. To see this, we shall start performing the change of variable $\boldsymbol{\xi} \leftarrow \boldsymbol{\xi} + \left(\hat{\mathbf{q}}_{\text{b}}^t\right)^{1/2} \mathbf{w}^\star$ in (69) so that we directly obtain the first equation of (58) with the corresponding time indices

$$\mathbf{q}_{\text{b}}^{t+1} = \mathbb{E}_{\mathbf{w}^\star, \boldsymbol{\xi}} \left[\mathscr{Z}_{\text{w}^\star}\left((\hat{\mathbf{q}}_{\text{b}}^t)^{1/2} \boldsymbol{\xi}, \hat{\mathbf{q}}_{\text{b}}^t \right) \mathbf{f}_{\text{w}^\star}\left((\hat{\mathbf{q}}_{\text{b}}^t)^{1/2} \boldsymbol{\xi}, \hat{\mathbf{q}}_{\text{b}}^t \right)^{\otimes 2} \right]. \tag{70}$$

Moreover in this setting, we notice that variables $\boldsymbol{\omega}^t_{\mu\to i}$ and $\mathbf{z}_\mu - \boldsymbol{\omega}^t_{\mu\to i}$ become independent since

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\omega}^t_{\mu\to i}\left(\mathbf{z}_\mu - \boldsymbol{\omega}^t_{\mu\to i}\right)^\intercal\right] &\underset{d\to\infty}{\longrightarrow} \mathbf{m}^t - \mathbf{q}^t = \mathbf{q}^t_{\mathrm{b}} - \mathbf{q}^t_{\mathrm{b}} = \mathbf{0}\,,\\
\mathbb{E}\left[\boldsymbol{\omega}^t_{\mu\to i}(\boldsymbol{\omega}^t_{\mu\to i})^\intercal\right] &\underset{d\to\infty}{\longrightarrow} \mathbf{q}^t_{\mathrm{b}}\,,\\
\mathbb{E}\left[\left(\mathbf{z}_\mu - \boldsymbol{\omega}^t_{\mu\to i}\right)\left(\mathbf{z}_\mu - \boldsymbol{\omega}^t_{\mu\to i}\right)^\intercal\right] &\underset{d\to\infty}{\longrightarrow} \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\,,
\end{aligned}$$

so that the multivariate Gaussian distribution factorize to

$$\mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}, \mathbf{Q}^t_{\mathrm{b}}\right) = \mathscr{N}_{\boldsymbol{\omega}}\left(\mathbf{0}_K, \mathbf{q}^t_{\mathrm{b}}\right)\mathscr{N}_{\mathbf{z}}\left(\boldsymbol{\omega}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\right).$$

Using $\mathrm{P}_{\mathrm{out}^\star}(y|\mathbf{z}) = \delta\left(y - \varphi_{\mathrm{out}^\star}(\mathbf{z})\right)$ the second equation of (69) becomes

$$\begin{aligned}
\hat{\mathbf{q}}^t &= \alpha\int_{\mathbb{R}^K}\int_{\mathbb{R}^K} \mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\mathbf{z}\;\mathscr{N}_{(\mathbf{z},\boldsymbol{\omega})}\left(\mathbf{0}_{2K}, \mathbf{Q}^t_{\mathrm{b}}\right)\mathbf{f}_{\mathrm{out}^\star}(\varphi_{\mathrm{out}^\star}(\mathbf{z}), \boldsymbol{\omega}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}})^{\otimes 2}\\
&= \alpha\int_{\mathbb{R}}\mathrm{d}y\int_{\mathbb{R}^K}\mathrm{d}\boldsymbol{\omega}\;\mathscr{N}_{\boldsymbol{\omega}}\left(\mathbf{0}_K, \mathbf{q}^t_{\mathrm{b}}\right)\\
&\qquad\times\int_{\mathbb{R}^K}\mathrm{d}\mathbf{z}\;\mathrm{p}_{\mathrm{out}^\star}(y|\mathbf{z})\mathscr{N}_{\mathbf{z}}\left(\boldsymbol{\omega}; \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\right)\mathbf{f}_{\mathrm{out}^\star}(y, \boldsymbol{\omega}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}})^{\otimes 2}\\
&= \alpha\int_{\mathbb{R}}\mathrm{d}y\int_{\mathbb{R}^K}\mathrm{d}\boldsymbol{\xi}\;\mathscr{N}_{\boldsymbol{\xi}}\left(\mathbf{0}; \mathbf{I}_K\right)\int_{\mathbb{R}^K}\mathrm{d}\mathbf{z}\;\mathrm{p}_{\mathrm{out}^\star}(y|\mathbf{z})\\
&\qquad\times\mathscr{N}_{\mathbf{z}}\left((\mathbf{q}^t_{\mathrm{b}})^{1/2}\boldsymbol{\xi}; \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\right)\mathbf{f}_{\mathrm{out}^\star}(y, (\mathbf{q}^t_{\mathrm{b}})^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}})^{\otimes 2}\\
&\qquad\qquad\text{(Change of variable } \boldsymbol{\xi} \leftarrow (\mathbf{q}^t_{\mathrm{b}})^{-1/2}\boldsymbol{\omega}^t)\\
&= \alpha\int_{\mathbb{R}}\mathrm{d}y\;\mathbb{E}_{\boldsymbol{\xi}}\,\mathscr{Z}_{\mathrm{out}^\star}\left(y, (\mathbf{q}^t_{\mathrm{b}})^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\right)\\
&\qquad\qquad\times\mathbf{f}_{\mathrm{out}^\star}\left(y, (\mathbf{q}^t_{\mathrm{b}})^{1/2}\boldsymbol{\xi}, \boldsymbol{\rho}_{\mathrm{w}^\star} - \mathbf{q}^t_{\mathrm{b}}\right),
\end{aligned}$$

which is exactly the second fixed point equation of the RS free entropy (58).

References

[1] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[2] Tommaso Castellani and Andrea Cavagna. Spin-glass theory for pedestrians. *Journal of Statistical Mechanics: Theory and Experiment*, (5):215–266, 2005.

[3] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Phase Transitions, Optimal Errors and Optimality of Message-Passing in Generalized Linear Models. pages 1–59, 2017.

[4] JS Yedidia and WT Freeman. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, pages 239 – 269, 2001.

[5] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: Algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(8):1–42, 2012.

[6] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

[7] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7), 2017.

[8] S Patarnello and P Carnevali. Learning networks of neurons with boolean logic. *Europhysics Letters (EPL)*, 4(4):503–508, aug 1987.

[9] Elizabeth Gardner and Bernard Derrida. Three unfinished works on the optimal storage capacity of networks. *Journal of Physics A: Mathematical and General*, 22(12):1983, 1989.

[10] Naftali Tishby, Esther Levin, and Sara A. Solla. Consistent inference of probabilities in layered networks: Predictions and generalization. In Anon, editor, *IJCNN Int Jt Conf Neural Network*, pages 403–409. Publ by IEEE, December 1989. IJCNN International Joint Conference on Neural Networks ; Conference date: 18-06-1989 Through 22-06-1989.

[11] H. Sompolinsky, N. Tishby, and H. S. Seung. Learning from examples in large neural networks. *Physical Review Letters*, 65(13):1683–1686, 1990.

[12] Sebastian Seung, Haim Sompolinsky, and Naftali Tishby. Statistical mechanics of learning from examples. *Physical Review A*, 45(8):6056, 1992.

[13] Timothy LH Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65(2):499, 1993.

[14] G. Györgyi. Techniques of replica symmetry breaking and the storage problem of the McCulloch-Pitts neuron. *Physics Report*, 342(4-5):263–392, 2001.

[15] Yoshiyuki Kabashima. Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels. In *Journal of Physics: Conference Series*, volume 95, page 012001. IOP Publishing, 2008.

[16] Elizabeth Gardner and Bernard Derrida. Optimal storage properties of neural network models. *Journal of Physics A: Mathematical and general*, 21(1):271, 1988.

[17] Marc Mézard. The space of interactions in neural networks: Gardner's computation with the cavity method. *Journal of Physics A: Mathematical and General*, 22(12):2181, 1989.

[18] Henry Schwarze and John Hertz. Generalization in a large committee machine. *EPL (Europhysics Letters)*, 20(4):375, 1992.

[19] Henry Schwarze and John Hertz. Generalization in fully connected committee machines. *EPL (Europhysics Letters)*, 21(7):785, 1993.

[20] Henry Schwarze. Learning a rule in a multilayer neural network. *Journal of Physics A: Mathematical and General*, 26(21):5781, 1993.

[21] Rémi Monasson and Riccardo Zecchina. Learning and generalization theories of large committee-machines. *Modern Physics Letters B*, 9(30):1887–1897, 1995.

[22] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012.

[23] 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.

[24] R. Wong. *Asymptotic Approximations of Integrals Computer Science and Scientific Computing*. 1989.

[25] Benjamin Aubin, Florent Krzakala, Yue M Lu, and Lenka Zdeborová. Generalization error in high-dimensional perceptrons: Approaching bayes error with convex optimization. In *Advances in Neural Information Processing Systems 33*. 2020.

[26] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[27] Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

[28] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

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

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

[31] Mohsen Bayati, Marc Lelarge, Andrea Montanari, et al. Universality in polytope phase transitions and message passing algorithms. *The Annals of Applied Probability*, 25(2):753–822, 2015.

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

[33] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.