

Storage Capacity of ultrametric committee machines

J. P. Neirotti

Non-linearity and Complexity Research Group, Aston University, Birmingham B4
7ET, UK

E-mail: `j.p.neirotti@aston.ac.uk`

Abstract. The problem of computing the storage capacity of a feed-forward network, with L hidden layers, N inputs, and K units in the first hidden layer, is analyzed using techniques from statistical mechanics. We found that the storage capacity strongly depends on the network architecture $\hat{\alpha}_c \sim (\log K)^{1-1/2^L}$ and that the number of units K limits the number of possible hidden layers L through the relationship $2^L - 1 < 2 \log K$.

PACS numbers: 89.70., 84.35.+i

Submitted to: *J. Phys. A: Math. Gen.*

1. Introduction

Understanding the efficiency and robustness by which the human brain processes information has been the motor force behind the development of neural networks models. Recent findings in Neuroscience have provided insight into the principles governing information representation in the mammalian brain [1, 2], motivating recent advances in the deep-learning subfield (for a review, see [3, 4] and references therein). Two characteristics are common to the mainstream deep-learning approach, namely hierarchical network architecture and feature extraction through relatively simple classifiers. It is only natural to infer that the level of sophistication reached by the deep-learning applications strongly motivates the study of similar systems, using the statistical mechanics technology developed during the past decades.

From the theoretical physics perspective, neural networks are the archetype of disordered systems. Similarities between networks of formal neurons and spin systems have suggested the application of statistical mechanics techniques for their study [5]. Particularly after [6], where it was demonstrated that the statistical-mechanics approach can be helpful for studying the properties of perceptrons [7, 8], most of the effort was concentrated on solving the generalization and the storage capacity problems for networks with more complex architectures [9–16]. Nonetheless, the feed-forward network with only one hidden layer of binary units [15, 16] has been the most complex architecture considered for the storage capacity problem.

Recently, more complex architectures have drawn attention, due to the possibility to obtain analytically some computational properties [17, 18]. In this article we are interested in computing the storage capacity of such architectures, known as *ultrametric committee machines* (UCMs).

1.1. UCMs

Let us consider a feed-forward network implementing a Boolean function $\sigma_{\mathbb{W}} : \{\pm 1\}^N \rightarrow \{\pm 1\}$, with L hidden layers and with $K \ll N$ hidden-to-input units in the first hidden layer. All units (output, hidden and input) in the committee are binary. Hidden-to-input links are implemented by synaptic vectors $\mathbf{w} \in \mathbb{R}^N$ (figure 1). The structure from the bottom up is composed by one output unit connected to K_L units in the L -th hidden layer, each of them connected to K_{L-1} units in the $(L-1)$ -th level. The total number of units in the $(L-1)$ -th level is then $K_L K_{L-1}$. Each node has an activation variable that is a function of the activation variables of the sub-tree with root at the node. Connections from units at the ℓ -th hidden layer to units at the $\ell+1$ -th layer are all set to one. To single out the variables of the ℓ -th layer we will use the notation $\mathbf{k}_\ell \equiv [k_L, k_{L-1}, \dots, k_\ell] = [[k_L, k_{L-1}, \dots, k_{\ell+1}] k_\ell] = [\mathbf{k}_{\ell+1} k_\ell]$, which runs over all hidden units of the ℓ -th layer. Thus

$$\sigma_{\mathbb{W}}(\mathbf{S}) \equiv \text{sgn} \left(\frac{1}{\sqrt{K_L}} \sum_{k_L=1}^{K_L} \sigma_{k_L}(\mathbf{S}) \right) \quad (1)$$

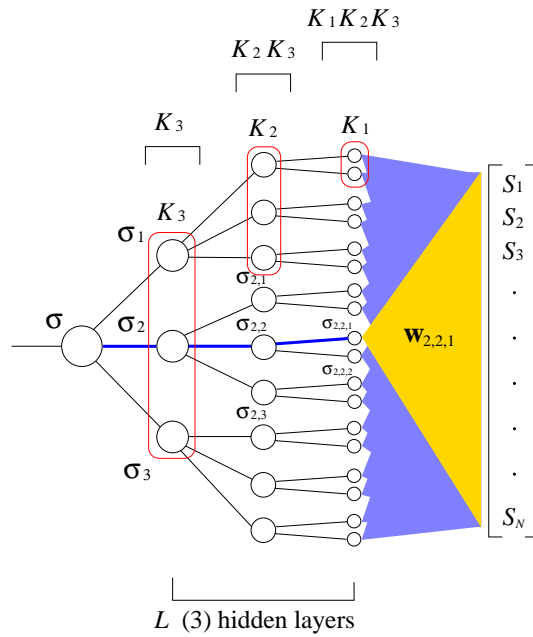


Figure 1. Schematic representation of an UCM with $L = 3$ hidden layers. The highlighted unit $\sigma_{2,2,1}$ is linked to the output unit through the path $\mathbf{k}_1 = [2, 2, 1]$.

$$\sigma_{\mathbf{k}_\ell}(\mathbf{S}) \equiv \text{sgn} \left(\frac{1}{\sqrt{K_{\ell-1}}} \sum_{k_{\ell-1}=1}^{K_{\ell-1}} \sigma_{[\mathbf{k}_\ell k_{\ell-1}]}(\mathbf{S}) \right) \quad (2)$$

$$\sigma_{\mathbf{k}_1}(\mathbf{S}) \equiv \text{sgn} \left(\frac{\mathbf{w}_{\mathbf{k}_1}^\top \mathbf{S}}{\sqrt{N}} \right), \quad (3)$$

where $\mathbb{W} = \{\mathbf{w}_{\mathbf{k}_1=[k_L, \dots, k_1]} \in \mathbb{R}^N, \mathbf{w}_{\mathbf{k}_1}^\top \mathbf{w}_{\mathbf{k}_1} = N, 1 \leq k_j \leq K_j\}$ is the set of synaptic vectors of the units in the first hidden layer, $\text{sgn}(x) = x/|x|$ if $x \neq 0$ and 0 otherwise, \mathbf{w}^\top is the transpose of the vector \mathbf{w} and $\mathbf{S} \in \{\pm 1\}^N$ is a pattern to be stored. As a last note over the architecture, we will impose the condition $K_j \gg \prod_{\ell=j+1}^L K_\ell$, which is equivalent to the initial imposition $N \gg K$.

The committee has been constructed drawing vectors from a suitable distribution over \mathbb{R}^N such that

$$\begin{aligned} [\Omega]_{\mathbf{k}_1, \mathbf{k}'_1} &\equiv \frac{\mathbf{w}_{\mathbf{k}_1}^\top \mathbf{w}_{\mathbf{k}'_1}}{N} \\ &= \delta_{\mathbf{k}_1 \mathbf{k}'_1} \left(1 - \tilde{\zeta}_1\right) + \delta_{\mathbf{k}_2 \mathbf{k}'_2} \left(\tilde{\zeta}_1 - \tilde{\zeta}_2\right) + \dots + \delta_{\mathbf{k}_L \mathbf{k}'_L} \left(\tilde{\zeta}_{L-1} - \tilde{\zeta}_L\right) + \tilde{\zeta}_L, \end{aligned} \quad (4)$$

where $\delta_{\mathbf{k}_\ell \mathbf{k}'_\ell} \equiv \prod_{m=\ell}^L \delta_{k_m k'_m}$ and $\delta_{ij} = 1$ if and only if $i = j$ and 0 otherwise. The structure of the matrix Ω is block-diagonal and resembles the matrices used to represent inter-replica interactions [19]. We impose the following scaling relationship to the elements of (4):

$$\tilde{\zeta}_\ell = \frac{\zeta_\ell}{\prod_{j=1}^\ell K_j}, \quad (5)$$

where $\zeta_j \sim O(1)$ and $\prod_{j=1}^L K_j = K$. In this way we can express the overlap matrix as:

$$[\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}'_1} = \delta_{\mathbf{k}_1 \mathbf{k}'_1} \Omega_0 + \delta_{\mathbf{k}_2 \mathbf{k}'_2} \frac{\Omega_1}{K_1} + \dots + \delta_{\mathbf{k}_L \mathbf{k}'_L} \frac{\Omega_{L-1}}{K_1 \dots K_{L-1}} + \frac{\Omega_L}{K}, \quad (6)$$

where

$$\Omega_\ell \equiv \zeta_\ell - \frac{\zeta_{\ell+1}}{K_{\ell+1}}. \quad (7)$$

Observe that the matrix $\mathbf{\Omega}$ has the following properties

- (i) $\mathbf{\Omega}$ is symmetric, i.e. $[\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}'_1} = [\mathbf{\Omega}]_{\mathbf{k}'_1, \mathbf{k}_1}$ for all paths \mathbf{k}_1 and \mathbf{k}'_1
- (ii) $\mathbf{\Omega}$ has non-negative entries, i.e. $[\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}'_1} > 0$ for all paths \mathbf{k}_1 and \mathbf{k}'_1
- (iii) $[\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}'_1} \geq \min \{ [\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}''_1}, [\mathbf{\Omega}]_{\mathbf{k}'_1, \mathbf{k}''_1} \}$ for all paths \mathbf{k}_1 , \mathbf{k}'_1 and \mathbf{k}''_1
- (iv) $1 = [\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}_1} \geq \max \{ [\mathbf{\Omega}]_{\mathbf{k}_1, \mathbf{k}'_1} \forall \mathbf{k}'_1 \neq \mathbf{k}_1 \}$,

therefore the matrix $\mathbf{\Omega}$ is *ultrametric* [20]. Given that the overlap matrix $\mathbf{\Omega}$ is ultrametric we dubbed these networks *ultrametric committee machines*.

1.2. The replica approach

Given a set of examples $\mathcal{S}_P = \{(\boldsymbol{\xi}_\mu)\}_{\mu=1}^P$ we want to compute the volume occupied by suitable synaptic vectors equally classifying the vectors in \mathcal{S}_P , according to the metric $d\mu(\mathbb{W})$:

$$V_{\mathcal{T}}(\mathcal{S}_P) \equiv \prod_{\mu=1}^P \Theta(\tau_\mu) \int d\mu(\mathbb{W}) \prod_{\mu=1}^P \prod_{\mathbf{k}_1} \Theta \left(\tau_{\mathbf{k}_1, \mu} \frac{\mathbf{w}_{\mathbf{k}_1}^\top \boldsymbol{\xi}_\mu}{\sqrt{N}} \right),$$

where the binary variables $\mathcal{T} \equiv \left\{ \tau_\mu, \{ \tau_{k_L, \mu} \}_{k_L=1}^{K_L}, \{ \tau_{\mathbf{k}_{L-1}, \mu} \}_{\mathbf{k}_{L-1}=1,1}^{K_L, K_{L-1}}, \dots, \{ \tau_{\mathbf{k}_1, \mu} \}_{\mathbf{k}_1=1, \dots, 1}^{K_L, \dots, K_1} \right\}_{\mu=1}^P$ conform the internal representation of $\boldsymbol{\xi}_\mu$ in the UCM with architecture determined by \mathbb{W} , and satisfy the following relationships

$$1 = \Theta \left(\frac{\tau_\mu}{\sqrt{K_L}} \sum_{k_L=1}^{K_L} \tau_{k_L, \mu} \right) = \Theta \left(\frac{\tau_{\mathbf{k}_{\ell+1}, \mu}}{\sqrt{K_\ell}} \sum_{k_\ell=1}^{K_\ell} \tau_{[\mathbf{k}_{\ell+1}, k_\ell], \mu} \right)$$

and Θ is the Heaviside function. Following [14], we compute the power β of the volume associated to the compatible internal representations:

$$V_{\mathcal{T}}^\beta(\mathcal{S}_P) \equiv \prod_{\mu=1}^P \Theta \left(\frac{1}{\sqrt{K_L}} \sum_{k_L=1}^{K_L} \tau_{k_L, \mu} \right) \left[\int d\mu(\mathbb{W}) \prod_{\mu=1}^P \prod_{\mathbf{k}_1} \Theta \left(\tau_{\mathbf{k}_1, \mu} \frac{\mathbf{w}_{\mathbf{k}_1}^\top \boldsymbol{\xi}_\mu}{\sqrt{N}} \right) \right]^\beta. \quad (8)$$

By replicating the products born from the powers we obtain the following expression:

$$V_{\mathcal{T}}^\beta(\mathcal{S}_P) = \prod_{\mu=1}^P \Theta \left(\frac{1}{\sqrt{K_L}} \sum_{k_L=1}^{K_L} \tau_{k_L, \mu} \right) \prod_{k_L=1}^{K_L} \sum_{\{ \tau_{\mathbf{k}_{L-1}, \mu} \}} \Theta \left(\frac{\tau_{k_L, \mu}}{\sqrt{K_{L-1}}} \sum_{k_{L-1}=1}^{K_{L-1}} \tau_{\mathbf{k}_{L-1}, \mu} \right) \dots \\ \prod_{k_2=1}^{K_2} \sum_{\{ \tau_{\mathbf{k}_1, \mu} \}} \Theta \left(\frac{\tau_{\mathbf{k}_2, \mu}}{\sqrt{K_1}} \sum_{k_1=1}^{K_1} \tau_{\mathbf{k}_1, \mu} \right) \prod_{\alpha=1}^{\beta} \prod_{k_1=1}^{K_1} \int d\mu(\mathbb{W}^\alpha) \Theta \left(\tau_{\mathbf{k}_1, \mu} \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha \top} \boldsymbol{\xi}_\mu}{\sqrt{N}} \right).$$

The statistical properties of the system can be obtained through the partition function:

$$Z_\beta(\mathcal{S}_P) \equiv \prod_{\mu=1}^P \sum_{\{\tau_{k_L,\mu}\}} V_{\mathcal{T}}^\beta(\mathcal{S}_P), \quad (9)$$

and, in particular, we are interested in optimizing the zero temperature specific entropy, which is a self averaging quantity, i.e.:

$$s(\beta) \equiv \frac{\langle \log Z_\beta(\mathcal{S}_P) \rangle_{\mathcal{S}_P}}{KN} = \frac{1}{KN} \lim_{n \rightarrow 0} \frac{\langle Z_\beta^n \rangle_{\mathcal{S}_P} - 1}{n}. \quad (10)$$

The replicated, quenched averaged partition function is then

$$\begin{aligned} \langle Z_\beta^n \rangle_{\mathcal{S}_P} &= \prod_{\mu=1}^P \prod_{a=1}^n \prod_{k_L=1}^{K_L} \sum_{\tau_{k_L,\mu}^a = \pm 1} \Theta \left(\frac{1}{\sqrt{K_L}} \sum_{k_L=1}^{K_L} \tau_{k_L,\mu}^a \right) \times \\ &\quad \prod_{k_{L-1}=1}^{K_{L-1}} \sum_{\tau_{k_{L-1},\mu}^a = \pm 1} \Theta \left(\frac{\tau_{k_L,\mu}^a}{\sqrt{K_{L-1}}} \sum_{k_L=1}^{K_L} \tau_{k_{L-1},\mu}^a \right) \cdots \\ &\quad \cdots \prod_{k_1=1}^{K_1} \sum_{\tau_{k_1,\mu}^a = \pm 1} \Theta \left(\frac{\tau_{k_2,\mu}^a}{\sqrt{K_1}} \sum_{k_1=1}^{K_1} \tau_{k_1,\mu}^a \right) \prod_{\alpha} \int d\mu(\mathbb{W}^{\alpha,a}) \left\langle \Theta \left(\tau_{\mathbf{k}_1,\mu}^a \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \boldsymbol{\xi}_\mu}{\sqrt{N}} \right) \right\rangle_{\mathcal{S}_P}. \end{aligned}$$

We can represent the Heaviside function by using the Fourier transform of the delta function

$$\Theta(x) = \int_0^\infty d\eta \delta(\eta - x) = \int \mathcal{D}(\eta, \hat{\eta}) \exp(i\hat{\eta}x),$$

where we have defined the notation $\mathcal{D}(x, \hat{x}) \equiv dx d\hat{x} \Theta(x) \exp(-ix\hat{x})/2\pi$. Thus

$$\begin{aligned} \langle Z_\beta^n \rangle_{\mathcal{S}_P} &= \prod_{\mu=1}^P \prod_{a=1}^n \int \mathcal{D}(\eta_\mu^a, \hat{\eta}_\mu^a) \prod_{k_L=1}^{K_L} \sum_{\tau_{k_L,\mu}^a = \pm 1} \exp \left(i \frac{\hat{\eta}_\mu^a \tau_{k_L,\mu}^a}{\sqrt{K_L}} \right) \int \mathcal{D}(\eta_{k_L,\mu}^a, \hat{\eta}_{k_L,\mu}^a) \cdots \\ &\quad \cdots \int \mathcal{D}(\eta_{k_2,\mu}^a, \hat{\eta}_{k_2,\mu}^a) \prod_{k_1=1}^{K_1} \sum_{\tau_{k_1,\mu}^a = \pm 1} \exp \left(i \frac{\tau_{k_2,\mu}^a \hat{\eta}_{k_2,\mu}^a \tau_{k_1,\mu}^a}{\sqrt{K_1}} \right) \\ &\quad \prod_{\alpha} \int \mathcal{D}(\eta_{\mathbf{k}_1,\mu}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1,\mu}^{\alpha,a}) \int d\mu(\mathbb{W}^{\alpha,a}) \left\langle \exp \left(i \tau_{\mathbf{k}_1,\mu}^a \hat{\eta}_{\mathbf{k}_1,\mu}^{\alpha,a} \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \boldsymbol{\xi}_\mu}{\sqrt{N}} \right) \right\rangle_{\mathcal{S}_P}. \end{aligned}$$

The average over patterns can be computed as follows

$$\begin{aligned} \left\langle \exp \left(i \sum_a \sum_{\alpha} \sum_{\mathbf{k}_1} \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a} \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \boldsymbol{\xi}}{\sqrt{N}} \right) \right\rangle_{\boldsymbol{\xi}} &= \prod_{j=1}^N \cos \left[\frac{1}{\sqrt{N}} \sum_a \sum_{\alpha} \sum_{\mathbf{k}_1} \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a} w_{\mathbf{k}_1,j}^{\alpha,a} \right] \\ &\simeq \exp \left(-\frac{1}{2} \sum_{a,b} \sum_{\alpha,\gamma} \sum_{\mathbf{k}_1,\mathbf{m}_1} \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a} \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b}}{N} \tau_{\mathbf{m}_1}^b \hat{\eta}_{\mathbf{m}_1}^{\gamma,b} \right) + O(N^{-2}). \end{aligned}$$

The synaptic overlaps are then arranged in a matrix with the following structure

$$[\mathbf{Q}]_{\mathbf{k}_1,\mathbf{m}_1}^{\alpha,a;\gamma,b} \equiv \frac{\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b}}{N}.$$

Let us define the vector components $[\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]_{\mathbf{k}_1}^{\alpha,a} \equiv \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}$. We have then

$$\begin{aligned} \langle Z_{\beta}^n \rangle_{\mathcal{S}_P} &= \int d\mathbf{Q} \left\langle \prod_{a,b} \prod_{\alpha,\gamma} \prod_{\mathbf{k}_1, \mathbf{m}_1} \delta \left(N [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{m}_1}^{\alpha,a;\gamma,b} - \mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b} \right) \right\rangle_{\mathbb{W}^{\beta n}} \\ &\quad \left\{ \prod_{a=1}^n \int \mathcal{D}(\eta^a, \hat{\eta}^a) \prod_{k_L=1}^{K_L} \sum_{\tau_{k_L}^a = \pm 1} \exp \left(i \frac{\hat{\eta}^a \tau_{k_L}^a}{\sqrt{K_L}} \right) \int \mathcal{D}(\eta_{k_L}^a, \hat{\eta}_{k_L}^a) \dots \right. \\ &\quad \dots \left. \int \mathcal{D}(\eta_{k_2}^a, \hat{\eta}_{k_2}^a) \prod_{k_1=1}^{K_1} \sum_{\tau_{k_1}^a = \pm 1} \exp \left(i \frac{\tau_{k_2}^a \hat{\eta}_{k_2}^a \tau_{k_1}^a}{\sqrt{K_1}} \right) \prod_{\alpha} \int \mathcal{D}(\eta_{\mathbf{k}_1}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}) \right. \\ &\quad \left. \exp \left(-\frac{1}{2} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^\top \mathbf{Q} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1] \right) \right\}^P. \end{aligned}$$

The average over the replicated set of synaptic vectors ($\mathbb{W}^{\beta n}$) can be expressed as:

$$\begin{aligned} \left\langle \prod_{a,b} \prod_{\alpha,\gamma} \prod_{\mathbf{k}_1, \mathbf{m}_1} \delta \left(N [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{m}_1}^{\alpha,a;\gamma,b} - \mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b} \right) \right\rangle_{\mathbb{W}^{\beta n}} &= \int \frac{d\hat{\mathbf{Q}}}{(2\pi)^{K\beta n}} \exp \left(\frac{N}{2} \text{tr} \mathbf{Q} \hat{\mathbf{Q}} \right) \times \\ &\times \left\langle \exp \left(-\frac{1}{2} \sum_{a,b} \sum_{\alpha,\gamma} \sum_{\mathbf{k}_1, \mathbf{m}_1} \mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b} [\hat{\mathbf{Q}}]_{\mathbf{k}_1, \mathbf{m}_1}^{\alpha,a;\gamma,b} \right) \right\rangle_{\mathbb{W}^{\beta n}}. \end{aligned}$$

By construction the measure of the synaptic vectors imposes that $\mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\alpha,a} = N [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{m}_1}$, so we have that the average $\left\langle \prod_{a,b} \prod_{\alpha,\gamma} \prod_{\mathbf{k}_1, \mathbf{m}_1} \delta \left(N q_{\mathbf{k}_1, \mathbf{m}_1}^{\alpha,a;\gamma,b} - \mathbf{w}_{\mathbf{k}_1}^{\alpha,a\top} \mathbf{w}_{\mathbf{m}_1}^{\gamma,b} \right) \right\rangle_{\mathbb{W}^{\beta n}}$ can be expressed as:

$$\int d\hat{\mathbf{Q}} \exp \left[-\frac{N}{2} \left(n\beta K - \text{tr} \mathbf{Q} \hat{\mathbf{Q}} + \log |\hat{\mathbf{Q}}| \right) \right].$$

By applying the Laplace method we observe that an extreme of the free energy will be obtained if $\hat{\mathbf{Q}} = \mathbf{Q}^{-1}$. If we consider $P = \alpha N$ then we have that the quenched average is

$$\langle Z_{\beta}^n \rangle_{\mathcal{S}_P} = \int d\mathbf{Q} \exp \left[N \left(\frac{1}{2} \log |\mathbf{Q}| + \alpha G_E(\mathbf{Q}) \right) \right]$$

where

$$\begin{aligned} G_E(\mathbf{Q}) &\equiv \log \left[\prod_{a=1}^n \int \mathcal{D}(\eta^a, \hat{\eta}^a) \prod_{k_L=1}^{K_L} \sum_{\tau_{k_L}^a = \pm 1} \exp \left(i \frac{\hat{\eta}^a \tau_{k_L}^a}{\sqrt{K_L}} \right) \int \mathcal{D}(\eta_{k_L}^a, \hat{\eta}_{k_L}^a) \dots \right. \\ &\quad \dots \left. \int \mathcal{D}(\eta_{k_2}^a, \hat{\eta}_{k_2}^a) \prod_{k_1=1}^{K_1} \sum_{\tau_{k_1}^{\alpha_2,a} = \pm 1} \exp \left(i \frac{\tau_{k_2}^a \hat{\eta}_{k_2}^a \tau_{k_1}^a}{\sqrt{K_1}} \right) \prod_{\alpha} \int \mathcal{D}(\eta_{\mathbf{k}_1}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}) \right. \\ &\quad \left. \exp \left(-\frac{1}{2} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^\top \mathbf{Q} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1] \right) \right]. \end{aligned}$$

Finally we have that:

$$\langle Z_{\beta}^n \rangle_{\mathcal{S}_P} \sim \text{extr}_{\mathbf{Q}} \left\{ \exp \left[N \left(\frac{1}{2} \log |\mathbf{Q}| + \alpha G_E(\mathbf{Q}) \right) \right] \right\}.$$

Let $\mathbb{O}^{K \times K}$ the commutative ring of ultrametric matrices with the block structure outline in (4). If we consider the Replica Symmetric (RS) Ansatz we can assume that

$$[\mathbf{Q}]^{\alpha, a; \gamma, b} = \delta^{a, b} \{ \delta^{\alpha, \gamma} [\mathbf{\Omega} - \mathbf{Y}] + \mathbf{Y} - \mathbf{R} \} + \mathbf{R}$$

where $\mathbf{\Omega}, \mathbf{Y}, \mathbf{R} \in \mathbb{O}^{K \times K}$. Observe that, although both have been originated from similar processes, the matrix \mathbf{Y} stores information about synaptic overlaps corresponding to vectors belonging to UCMs replicated due to the exponent β and \mathbf{R} stores similar information due to the replication process to compute the logarithm. Moreover, we have that:

$$\begin{aligned} [\mathbf{R}]_{\mathbf{k}_1, \mathbf{m}_1} &= \delta_{\mathbf{k}_1 \mathbf{m}_1} (\tilde{r}_0 - \tilde{r}_1) + \delta_{\mathbf{k}_2 \mathbf{m}_2} (\tilde{r}_1 - \tilde{r}_2) + \dots + \delta_{\mathbf{k}_L \mathbf{m}_L} (\tilde{r}_{L-1} - \tilde{r}_L) + \tilde{r}_L \\ &= \delta_{\mathbf{k}_1 \mathbf{m}_1} R_0 + \delta_{\mathbf{k}_2 \mathbf{m}_2} \frac{R_1}{K_1} + \dots + \delta_{\mathbf{k}_L \mathbf{m}_L} \frac{R_{L-1}}{K_1 \dots K_{L-1}} + \frac{R_L}{K} \\ [\mathbf{Y}]_{\mathbf{k}_1, \mathbf{m}_1} &= \delta_{\mathbf{k}_1 \mathbf{m}_1} (\tilde{y}_0 - \tilde{y}_1) + \delta_{\mathbf{k}_2 \mathbf{m}_2} (\tilde{y}_1 - \tilde{y}_2) + \dots + \delta_{\mathbf{k}_L \mathbf{m}_L} (\tilde{y}_{L-1} - \tilde{y}_L) + \tilde{y}_L \\ &= \delta_{\mathbf{k}_1 \mathbf{m}_1} Y_0 + \delta_{\mathbf{k}_2 \mathbf{m}_2} \frac{Y_1}{K_1} + \dots + \delta_{\mathbf{k}_L \mathbf{m}_L} \frac{Y_{L-1}}{K_1 \dots K_{L-1}} + \frac{Y_L}{K} \end{aligned}$$

where

$$\tilde{r}_\ell = \frac{r_\ell}{\prod_{j=1}^{\ell} K_j}, \quad R_\ell \equiv r_\ell - \frac{r_{\ell+1}}{K_{\ell+1}}, \quad \tilde{y}_\ell = \frac{y_\ell}{\prod_{j=1}^{\ell} K_j}, \quad Y_\ell \equiv y_\ell - \frac{y_{\ell+1}}{K_{\ell+1}},$$

and $r_\ell, y_\ell \sim O(1)$. Thus the logarithm of the determinant $\log |\mathbf{Q}|^{1/nK}$ can be expressed as:

$$\sum_{\ell=0}^L \frac{K_{\ell+1} - 1}{\prod_{j=1}^{\ell+1} K_j} \left[\beta \log \left(\sum_{j=0}^{\ell} \Delta_j \right) + \log \left(1 + \beta \frac{\sum_{j=0}^{\ell} D_j}{\sum_{j=0}^{\ell} \Delta_j} \right) + \beta \frac{\sum_{j=0}^{\ell} R_j}{\sum_{j=0}^{\ell} (\Delta_j + \beta D_j)} \right],$$

where $\Delta_j \equiv \Omega_j - Y_j$ and $D_j \equiv Y_j - R_j$.

Following the developments presented in Appendix A we have that

$$G_E(\mathbf{Q}) \simeq n \left\{ K \sqrt{\frac{\Delta_0}{\beta Y_0}} + \int \mathcal{D}z \log \mathcal{H} \left(-\sqrt{\frac{\mathcal{G}_L}{1 - \mathcal{G}_L}} z \right) \right\} + O(n^2),$$

where

$$\begin{aligned} \Sigma_0 &\equiv 1 \\ \Sigma_\ell &\equiv 1 + \sum_{j=1}^{\ell} \left(\frac{2}{\pi} \right)^j \left[\frac{\Omega_j}{Y_0} + \frac{1 - \beta \Delta_j}{\beta Y_0} \right] \end{aligned} \quad (11)$$

$$\begin{aligned} \mathcal{G}_0 &\equiv \frac{R_0}{Y_0} \\ \mathcal{G}_\ell &\equiv \frac{\Sigma_{\ell-1}}{\Sigma_\ell} \left[\frac{2}{\pi} \arcsin(\mathcal{G}_{\ell-1}) + \left(\frac{2}{\pi} \right)^\ell \frac{R_\ell}{\Sigma_{\ell-1} Y_0} \right]. \end{aligned} \quad (12)$$

By considering $\hat{\alpha} \equiv \alpha/K$, the entropy of the system is:

$$s(\beta) = \frac{1}{2} \sum_{\ell=0}^L \frac{K_{\ell+1} - 1}{\prod_{j=1}^{\ell+1} K_j} \left[\beta \log \left(\sum_{j=0}^{\ell} \Delta_j \right) + \log \left(1 + \beta \frac{\sum_{j=0}^{\ell} D_j}{\sum_{j=0}^{\ell} \Delta_j} \right) + \beta \frac{\sum_{j=0}^{\ell} R_j}{\sum_{j=0}^{\ell} (\Delta_j + \beta D_j)} \right]$$

$$+ \hat{\alpha} \left[\sqrt{\frac{\Delta_0 K^2}{\beta Y_0}} + \int \mathcal{D}z \log \mathcal{H} \left(-\sqrt{\frac{\mathcal{G}_L}{1 - \mathcal{G}_L}} z \right) \right].$$

2. Asymptotic behavior

In the limit of $\beta \rightarrow 0$ we expect $Y_\ell, R_\ell \rightarrow \Omega_\ell$, which implies that $\Delta_\ell, D_\ell \rightarrow 0$ and $\mathcal{G}_\ell \uparrow 1$, so we define the new parameters

$$\omega_\ell \equiv \frac{\Omega_\ell}{Y_0}, \quad \bar{\omega}_\ell \equiv \sum_{j=0}^{\ell} \omega_j, \quad \delta_\ell \equiv \frac{D_\ell}{Y_0}, \quad \bar{\delta}_\ell \equiv \sum_{j=0}^{\ell} \delta_j \quad (13)$$

and

$$m_0^2 \equiv \lim_{\beta \rightarrow 0} \frac{K^2 \Delta_0}{\beta Y_0}, \quad m_{\ell>0} \equiv \lim_{\beta \rightarrow 0} \frac{\Delta_\ell}{\beta Y_0}, \quad \bar{m}_{\ell>0} \equiv \frac{m_0^2}{K^2} + \sum_{j=1}^{\ell} m_j \simeq \sum_{j=1}^{\ell} m_j \quad (14)$$

thus we define the entropy at $\beta = 0$ as:

$$s_0(\hat{\alpha}; \{\delta_j\}, \{m_j\}) \simeq \frac{K_1 - 1}{K_1} \left[\log K - \log m_0 + \frac{1}{2} \log \delta_0 + \frac{\omega_0}{2\delta_0} \right] + \\ + \frac{1}{2} \sum_{\ell=1}^L \frac{K_{\ell+1} - 1}{\prod_{j=1}^{\ell+1} K_j} \left[\log \left(1 + \frac{\bar{\delta}_\ell}{\bar{m}_\ell} \right) + \frac{\bar{\omega}_\ell - \bar{\delta}_\ell}{\bar{m}_\ell + \bar{\delta}_\ell} \right] + \hat{\alpha} \left[m_0 - \frac{1}{4} \frac{\mathcal{G}_L}{1 - \mathcal{G}_L} \right].$$

The optimization process implies solving the saddle point equations $\partial_\lambda s_0 = 0$, where λ is any of the parameters $\{m_j, \delta_j\}$. The equation for m_0 is

$$0 = -\frac{1}{m_0} + \hat{\alpha} + O(K_1^{-1})$$

which implies that $\hat{\alpha} m_0 \simeq 1$. For all $j = 1, \dots, L$ we have that

$$0 = \frac{1}{2} \sum_{\ell=1}^L \frac{K_{\ell+1} - 1}{\prod_{j=1}^{\ell+1} K_j} \left[\frac{1}{\bar{m}_\ell + \bar{\delta}_\ell} - \frac{1}{\bar{m}_\ell} - \frac{\bar{\omega}_\ell - \bar{\delta}_\ell}{(\bar{m}_\ell + \bar{\delta}_\ell)^2} \right] \frac{\partial \bar{m}_\ell}{\partial m_j} - \frac{\hat{\alpha}}{4} \frac{1}{(1 - \mathcal{G}_L)^2} \frac{\partial \mathcal{G}_L}{\partial m_j} \\ 0 = -\frac{1}{2} \sum_{\ell=1}^L \frac{K_{\ell+1} - 1}{\prod_{j=1}^{\ell+1} K_j} \frac{\bar{\omega}_\ell - \bar{\delta}_\ell}{(\bar{m}_\ell + \bar{\delta}_\ell)^2} \frac{\partial \bar{\delta}_\ell}{\partial \delta_j} - \frac{\hat{\alpha}}{4} \frac{1}{(1 - \mathcal{G}_L)^2} \frac{\partial \mathcal{G}_L}{\partial \delta_j}.$$

The solutions to these equations are such that $\max\{m_j, \delta_j\} \lesssim O\left(\sqrt{\delta_0/K_1 \hat{\alpha}}\right)$. This implies that the dominant variable in \mathcal{G}_L is δ_0 , thus

$$\mathcal{G}_L \simeq 1 - \mathcal{A}_L \delta_0^{1/2^L},$$

where, by redefining Σ_ℓ , we have that:

$$\Sigma_{\ell>0} \equiv 1 + \sum_{j=1}^{\ell} \left(\frac{2}{\pi} \right)^j \zeta_j \\ \mathcal{A}_L \equiv \left(\frac{8}{\pi^2} \right)^{1-1/2^L} \frac{\prod_{j=1}^{L-1} \Sigma_j^{1/2^{L-j}}}{\Sigma_L}. \quad (15)$$

Thus, considering that $m_0 \hat{\alpha} \simeq 1$ and disregarding additive constants, the effective entropy becomes:

$$s_0(\hat{\alpha}; \delta_0) \simeq \log K + \log \hat{\alpha} + \frac{1}{2} \log \delta_0 + \frac{1}{2\delta_0} - \frac{\hat{\alpha}}{4\mathcal{A}_L} \delta_0^{-1/2^L} + O(K_1^{-1}).$$

The equation $\partial_{\delta_0} s_0 = 0$ is

$$0 = 1 - \frac{1}{\delta_0} + \frac{\hat{\alpha}}{2^{L+1}\mathcal{A}_L} \delta_0^{-1/2^L}$$

thus

$$\frac{1}{\delta_0} \simeq \frac{\hat{\alpha}}{2^{L+1}\mathcal{A}_L} \delta_0^{-1/2^L} \quad (16)$$

and

$$s_0(\hat{\alpha}; \delta_0) \simeq \log K + \log \hat{\alpha} + \frac{1}{2} \log \delta_0 - \frac{\hat{\alpha}}{4\mathcal{A}_L} \frac{2^L - 1}{2^L} \delta_0^{-1/2^L} + O(K_1^{-1}),$$

so, disregarding terms of $O(\log \hat{\alpha})$, $O(\log \delta_0)$, we have that the entropy $s_0(\hat{\alpha}_c; \delta_0)$ gets negligibly small at:

$$\hat{\alpha}_c \simeq \frac{2^{L+2}}{2^L - 1} \mathcal{A}_L \delta_0^{1/2^L} \log K. \quad (17)$$

Close to the criticality we expect that δ_0 , which is a measure of the largest difference between the matrices \mathbf{Y} and \mathbf{R} , to be small. Equations (16) and (17) lead to

$$\frac{1}{\delta_0} \simeq \frac{2 \log K}{2^L - 1} \quad (18)$$

$$\hat{\alpha}_c(L, K) \simeq \frac{2^{L+2-1/2^L}}{(2^L - 1)^{1-1/2^L}} \mathcal{A}_L (\log K)^{1-1/2^L}. \quad (19)$$

From (18) we extract the following relationship for L and K :

$$L < \log_2(2 \log K + 1), \quad (20)$$

which limits the number of hidden layers we may put in the UCM.

These are the results found by Monasson and Urbanczik for $L = 1$ [13–15]. The upper bound for this expression is given by:

$$\hat{\alpha}_c(L, K) \lesssim \frac{32}{\pi^2} \log K \simeq 2.25 \log_2 K.$$

Although this capacity is larger than $\log_2 K$, there is no violation to the Mithchinson-Durbin bound [21], which is only applied to networks with $L = 1$.

3. Conclusions

We computed the storage capacity per unit, for an ultrametric committee machine with $K \gg 1$ units in the first of its L hidden layers. Our results are compatible with previous works (with $L = 1$ [13–15]) and represents a step forward in the level of complexity of tractable architectures. Our results, represented by the equations (19) and (20), are only valid if all the quantities represented by (12) can reach the value 1 (i.e. $\mathcal{G}_\ell \uparrow 1$). This can

only be achieved for finite values of the overlaps [$\Sigma_j < \infty$ given by (11)]. A divergence in Σ_ℓ occurs if the overlap ζ_ℓ is too large (particularly if the scaling law given by (5) is not respected). As it was observed in [17, 18], too large an overlap effectively reduces the complexity of the machine, i.e. the UCM can be effectively replaced by another UCM with less hidden layers, rendering the subsequent calculations meaningless.

To understand the information encompassed by (19) consider the following scenarios. Firstly, in a biological context, the mammalian brain may have up to 10^{10} neurons, arrayed in hierarchical structures, having no more than six layers. These numbers are in agreement with the bound (20), which indicates that these structures may have emerged as an efficient means to increase the brain computation capabilities. Secondly, suppose that we construct an UCM with only one hidden layer and no overlaps between synaptic vectors (a true tree committee machine) and a mole ($K = 6.02 \cdot 10^{23}$) of units. For such a machine the critical capacity is $\hat{\alpha}_1 = \frac{16}{\pi} \sqrt{\log K} \simeq 37.70$. If we take 10% of those units and construct a second hidden layer, leaving $0.9K$ in the first hidden layer, the capacity of the new UCM is $\hat{\alpha}_2 = 2^{15/4} \left(\frac{8}{3\pi^2} \log(0.9K) \right)^{3/4} \simeq 101.39$, almost three times the capacity of the one-hidden-layer UCM. If the units in the second layer are re-arranged to construct a third and fourth layers, we obtain UCMs with the following capacities: $\hat{\alpha}_3 \simeq 147.52$ and $\hat{\alpha}_4 \simeq 169.23$. Observe that we cannot continue with this process without breaking the bound imposed by (20) and making δ_0 too large. A note of caution here. The analysis presented is valid if quantities that represent differences between matrix elements corresponding to original and replicated systems are sufficiently small. In the present scenario, where the replica symmetric approach has been applied, there is only one of such quantities that remains relevant, namely δ_0 . Equation (18) links δ_0 with K and L (the network's architecture). The necessity to keep δ_0 small produces the upper bound (20) for the total number of hidden layers given K . We cannot determine if this limitation is real, i.e. that the capacity of the system cannot be increased any further by adding hidden layers to the architecture, or a byproduct of the replica symmetric approach. Anyway, it is clear that re-arranging the network architecture, without adding extra resources, results in more than a 400% gain in the network capacity.

It is important to note that, given the architectural constraints imposed by (4), we cannot increase the capacity any further by dilution [22]. The most diluted UCM is represented by one with all its synaptic overlaps ζ_l set to zero (a true committee machine), which is precisely the case presented. The asymptotic behavior obtained for the capacity is $\hat{\alpha} \sim \log(K)$. Such behavior is expected in machines where the architecture is arranged in such a way that the closer to the output the layer the smaller the number of processing units in it. Given that UCMs are constructed following this pattern, the upper bound $2.25 \log_2(K)$ may represent the true upper bound for these kind of architectures. To construct machines with capacities beyond $\log(K)$ would require to consider hidden layers with larger number of processing units, i.e. $K > N$ which goes beyond the scope of the present article.

The possibility to predict these quantities will help the development of real systems

where the balance of resources and computational gain are very practical issues.

Acknowledgments

The author would like to acknowledge the discussions with Dr Roberto C. Alamino that help the improvement of the early versions of this work.

Appendix A. Calculation of the energetic part

Let us define the vector components $[\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]_{\mathbf{k}_1}^{\alpha,a} \equiv \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}$. The exponential in G_E becomes then

$$\begin{aligned} \exp\left(-\frac{1}{2}[\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^\top \mathbf{Q} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]\right) &= \exp\left[-\frac{1}{2} \sum_a \sum_\alpha [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\alpha,a \top} [\boldsymbol{\Omega} - \mathbf{Y}] [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\alpha,a} - \right. \\ &\quad \left. - \frac{1}{2} \sum_a \sum_\alpha \sum_\gamma [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\alpha,a \top} [\mathbf{Y} - \mathbf{R}] [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\gamma,a} - \right. \\ &\quad \left. - \frac{1}{2} \sum_a \sum_\alpha \sum_b \sum_\gamma [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\alpha,a \top} \mathbf{R} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^{\gamma,b}\right]. \end{aligned}$$

Analyzing term by term and by the use of the Hubbard-Stratonovitch (HS) identity, we have that RHS can be expressed as:

$$\int \mathcal{D}\text{HS} \exp\left[-\frac{1}{2} \Delta_0 \sum_{a,\alpha} \sum_{\mathbf{k}_1} (\hat{\eta}_{\mathbf{k}_1}^{\alpha,a})^2 + i \sum_{a,\alpha} \sum_{\mathbf{k}_1} \left(\Gamma_{\mathbf{k}_1}^a + \frac{\mathcal{Y}_{\mathbf{k}_2}^{\alpha,a}}{\sqrt{K_1}}\right) \tau_{\mathbf{k}_1}^a \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}\right],$$

where $\mathcal{D}x \equiv dx \exp(-x^2/2)/\sqrt{2\pi}$ is the Gaussian metric, $\mathcal{D}\text{HS}$ represents the Gaussian metric in all the Hubbard-Stratonovitch variables and

$$\begin{aligned} \Gamma_{\mathbf{k}_1}^a &\equiv \sqrt{D_0} x_{\mathbf{k}_1}^a + \sqrt{R_0} x_{\mathbf{k}_1} \\ \mathcal{Y}_{\mathbf{k}_2}^{\alpha,a} &\equiv \sum_{m=1}^L \left[\sqrt{\frac{\Delta_m}{\prod_{j=2}^m K_j}} x_{\mathbf{k}_{m+1}}^{\alpha,a} + \sqrt{\frac{D_m}{\prod_{j=2}^m K_j}} x_{\mathbf{k}_{m+1}}^a + \sqrt{\frac{R_m}{\prod_{j=2}^m K_j}} x_{\mathbf{k}_{m+1}} \right]. \end{aligned}$$

The integrals over $\mathcal{D}(\eta_{\mathbf{k}_1}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1}^{\alpha,a})$ produce the following expression

$$\int \mathcal{D}(\eta_{\mathbf{k}_1}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}) \exp\left(-\frac{1}{2}[\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^\top \mathbf{Q} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]\right) = \mathcal{H}\left(-\frac{\tau_{\mathbf{k}_1}^a}{\sqrt{\Delta_0}} \left(\Gamma_{\mathbf{k}_1}^a + \mathcal{Y}_{\mathbf{k}_2}^{\alpha,a}/\sqrt{K_1}\right)\right),$$

where $\mathcal{H}(x) \equiv \int \mathcal{D}u \Theta(u - x)$. Thus, by applying the formula (B.3), we have:

$$\begin{aligned} \mathcal{A}_{\mathbf{k}_1}^a &\equiv \int \mathcal{D}x_{\mathbf{k}_1}^a \prod_\alpha \int \mathcal{D}(\eta_{\mathbf{k}_1}^{\alpha,a}, \hat{\eta}_{\mathbf{k}_1}^{\alpha,a}) \exp\left(-\frac{1}{2}[\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]^\top \mathbf{Q} [\boldsymbol{\tau}_1 \hat{\boldsymbol{\eta}}_1]\right) \\ &\simeq \mathcal{H}\left(-\tau_{\mathbf{k}_1}^a \left(\sqrt{\frac{R_0}{D_0}} x_{\mathbf{k}_1} + \frac{\sum_\alpha \mathcal{Y}_{\mathbf{k}_2}^{\alpha,a}}{\sqrt{K_1 D_0 \beta}}\right)\right) + \\ &\quad + \frac{1}{2} \sqrt{\frac{\Delta_0}{\beta D_0}} \exp\left\{-\frac{1}{2} \left(\sqrt{\frac{R_0}{D_0}} x_{\mathbf{k}_1} + \frac{\sum_\alpha \mathcal{Y}_{\mathbf{k}_2}^{\alpha,a}}{\sqrt{K_1 D_0 \beta}}\right)^2\right\} \quad (\text{A.1}) \end{aligned}$$

Observe that the dependency on $x_{\mathbf{k}_\ell}^{\alpha,a}$ is concentrated on the term $\sum_\alpha \Upsilon_{\mathbf{k}_2}^{\alpha,a}$, thus we can make a change of variables such that:

$$\Upsilon_{\mathbf{k}_2}^a \equiv \frac{1}{\beta} \sum_\alpha \Upsilon_{\mathbf{k}_2}^{\alpha,a} = \sum_{m=1}^L \left[\sqrt{\frac{\Delta_m/\beta + D_m}{\prod_{j=2}^m K_j}} x_{\mathbf{k}_{m+1}}^a + \sqrt{\frac{R_m}{\prod_{j=2}^m K_j}} x_{\mathbf{k}_{m+1}} \right].$$

We can disregard terms of $O(\Delta_0/(\beta D_0))$ and $O(\sqrt{\Delta_0}/(\beta D_0 K_1))$, such that:

$$\mathcal{A}_{\mathbf{k}_1}^a \simeq \mathcal{H} \left(-\sqrt{\frac{R_0}{D_0}} \tau_{\mathbf{k}_1}^a x_{\mathbf{k}_1} \right) + \left[\frac{1}{2} \sqrt{\frac{\Delta_0}{\beta D_0}} + \frac{\tau_{\mathbf{k}_1}^a \Upsilon_{\mathbf{k}_2}^a}{\sqrt{2\pi K_1 D_0}} \right] \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right).$$

The next step involves the trace over the variables $\tau_{\mathbf{k}_1}^a$. Thus

$$\begin{aligned} & \sum_{\tau_{\mathbf{k}_1}^a = \pm 1} \exp \left(i \frac{\tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \tau_{\mathbf{k}_1}^a}{\sqrt{K_1}} \right) \mathcal{A}_{\mathbf{k}_1}^a \simeq \cos \left(\frac{\hat{\eta}_{\mathbf{k}_2}^a}{\sqrt{K_1}} \right) \left[1 + \sqrt{\frac{\Delta_0}{\beta D_0}} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right) \right] + \\ & + i \tau_{\mathbf{k}_2}^a \sin \left(\frac{\hat{\eta}_{\mathbf{k}_2}^a}{\sqrt{K_1}} \right) \left[\operatorname{erf} \left(\sqrt{\frac{R_0}{2D_0}} x_{\mathbf{k}_1} \right) + \sqrt{\frac{2}{\pi K_1 D_0}} \Upsilon_{\mathbf{k}_2}^a \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right) \right] \\ & \simeq 1 + \sqrt{\frac{\Delta_0}{\beta D_0}} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right) + i \frac{\tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a}{\sqrt{K_1}} \operatorname{erf} \left(\sqrt{\frac{R_0}{2D_0}} x_{\mathbf{k}_1} \right) - \\ & - \frac{(\hat{\eta}_{\mathbf{k}_2}^a)^2}{2K_1} + i \sqrt{\frac{2}{\pi D_0}} \frac{\tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \Upsilon_{\mathbf{k}_2}^a}{K_1} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right), \end{aligned}$$

thus

$$\begin{aligned} & \prod_a \sum_{\tau_{\mathbf{k}_1}^a = \pm 1} \exp \left(i \frac{\tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \tau_{\mathbf{k}_1}^a}{\sqrt{K_1}} \right) \mathcal{A}_{\mathbf{k}_1}^a \simeq 1 + n \sqrt{\frac{\Delta_0}{\beta D_0}} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right) + \\ & + \frac{i}{\sqrt{K_1}} \left(\frac{1}{\sqrt{D_0}} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right) \sum_a \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \right) \operatorname{erf} \left(\sqrt{\frac{R_0}{2D_0}} x_{\mathbf{k}_1} \right) - \\ & - \frac{1}{K_1} \left(\sum_{a < b} \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \tau_{\mathbf{k}_2}^b \hat{\eta}_{\mathbf{k}_2}^b \right) \operatorname{erf}^2 \left(\sqrt{\frac{R_0}{2D_0}} x_{\mathbf{k}_1} \right) - \frac{1}{2K_1} \sum_a (\hat{\eta}_{\mathbf{k}_2}^a)^2 + \\ & + \frac{i}{K_1} \sqrt{\frac{2}{\pi}} \left[\sum_a \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \Upsilon_{\mathbf{k}_2}^a \right] \frac{1}{\sqrt{D_0}} \exp \left(-\frac{R_0 x_{\mathbf{k}_1}^2}{2D_0} \right), \end{aligned}$$

and

$$\begin{aligned} \mathcal{A}_{\mathbf{k}_2} & \equiv \int \mathcal{D}x_{\mathbf{k}_1} \prod_a \sum_{\tau_{\mathbf{k}_1}^a = \pm 1} \exp \left(i \frac{\tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \tau_{\mathbf{k}_1}^a}{\sqrt{K_1}} \right) \mathcal{A}_{\mathbf{k}_1}^a \\ & \simeq 1 + \frac{1}{K_1} \left[n K_1 \sqrt{\frac{\Delta_0}{\beta Y_0}} - \frac{1}{2} \sum_a (\hat{\eta}_{\mathbf{k}_2}^a)^2 - \frac{2}{\pi} \arcsin \left(\frac{R_0}{Y_0} \right) \left(\sum_{a < b} \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \tau_{\mathbf{k}_2}^b \hat{\eta}_{\mathbf{k}_2}^b \right) \right] + \\ & \quad + i \sqrt{\frac{2}{\pi Y_0}} \sum_a \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \Upsilon_{\mathbf{k}_2}^a \\ & \simeq \int \mathcal{D}z_{\mathbf{k}_2} \exp \left[n K_1 \sqrt{\frac{\Delta_0}{\beta Y_0}} - \frac{1}{2} \left[1 - \frac{2}{\pi} \arcsin \left(\frac{R_0}{Y_0} \right) \right] \sum_a (\hat{\eta}_{\mathbf{k}_2}^a)^2 + \right. \end{aligned}$$

$$+i\sqrt{\frac{2}{\pi}} \sum_a \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \left(\sqrt{\arcsin\left(\frac{R_0}{Y_0}\right)} z_{\mathbf{k}_2} + \sqrt{\frac{1}{Y_0}} \mathcal{R}_{\mathbf{k}_2}^a \right).$$

Thus

$$\begin{aligned} \mathcal{A}_{\mathbf{k}_2} \equiv & \exp \left[nK_1 \sqrt{\frac{\Delta_0}{\beta Y_0}} - \frac{1}{2} \left[1 - \frac{2}{\pi} \arcsin\left(\frac{R_0}{Y_0}\right) \right] \sum_a (\hat{\eta}_{\mathbf{k}_2}^a)^2 + \right. \\ & \left. + i\sqrt{\frac{2}{\pi}} \sum_a \tau_{\mathbf{k}_2}^a \hat{\eta}_{\mathbf{k}_2}^a \left(\sqrt{\arcsin\left(\frac{R_0}{Y_0}\right)} z_{\mathbf{k}_2} + \frac{\mathcal{R}_{\mathbf{k}_2}^a}{\sqrt{Y_0}} \right) \right], \end{aligned}$$

which marks the end of the integration over the variables associated to the first layer.

The first integration associated to the second layer is over $\mathcal{D}(\eta_{\mathbf{k}_2}^a, \hat{\eta}_{\mathbf{k}_2}^a)$. Thus

$$\int \prod_a \mathcal{D}(\eta_{\mathbf{k}_2}^a, \hat{\eta}_{\mathbf{k}_2}^a) \mathcal{A}_{\mathbf{k}_2} \simeq \exp \left(nK_1 \sqrt{\frac{\Delta_0}{\beta Y_0}} \right) \prod_a \mathcal{H} \left(-\frac{\tau_{\mathbf{k}_2}^a (\Gamma_{\mathbf{k}_2}^a + \mathcal{R}_{\mathbf{k}_3}^a / \sqrt{K_2})}{\sqrt{1 - (2/\pi) \arcsin(R_0/Y_0)}} \right), \quad (\text{A.2})$$

where the dominant part of the argument of the error function can be expressed as (joining over $z_{\mathbf{k}_2}$ and $x_{\mathbf{k}_2}$)

$$\begin{aligned} \Gamma_{\mathbf{k}_2}^a & \equiv \sqrt{\frac{2}{\pi}} \left(\frac{\Delta_1}{\beta Y_0} + \frac{D_1}{Y_0} \right) x_{\mathbf{k}_2}^a + \sqrt{\frac{2}{\pi}} \left(\arcsin\left(\frac{R_0}{Y_0}\right) + \frac{R_1}{Y_0} \right) x_{\mathbf{k}_2} \\ \mathcal{R}_{\mathbf{k}_3}^a & \equiv \sum_{m=2}^L \left(\prod_{j=3}^m K_j^{-1/2} \right) \left[\sqrt{\frac{2}{\pi}} \left(\frac{\Delta_m}{\beta Y_0} + \frac{D_m}{Y_0} \right) x_{\mathbf{k}_{m+1}}^a + \sqrt{\frac{2}{\pi}} \frac{R_m}{Y_0} x_{\mathbf{k}_{m+1}} \right]. \end{aligned}$$

Observe that

$$\int \mathcal{D}x \mathcal{H}(ax + y) = \mathcal{H}\left(\frac{y}{\sqrt{1+a^2}}\right)$$

so

$$\begin{aligned} & \int \mathcal{D}x_{\mathbf{k}_2}^a \mathcal{H} \left(-\frac{\tau_{\mathbf{k}_2}^a (\Gamma_{\mathbf{k}_2}^a + \mathcal{R}_{\mathbf{k}_3}^a / \sqrt{K_2})}{\sqrt{1 - (2/\pi) \arcsin(R_0/Y_0)}} \right) = \\ & = \mathcal{H} \left(-\tau_{\mathbf{k}_2}^a \left(\sqrt{\frac{\mathcal{G}_1}{1 - \mathcal{G}_1}} x_{\mathbf{k}_2} + \frac{\mathcal{R}_{\mathbf{k}_3}^a}{\sqrt{K_2 \Sigma_1 (1 - \mathcal{G}_1)}} \right) \right), \end{aligned}$$

where

$$\begin{aligned} \Sigma_1 & \equiv 1 + \frac{2}{\pi} \left[\frac{\Omega_1}{Y_0} + \frac{1 - \beta \Delta_1}{\beta Y_0} \right] \\ \mathcal{G}_1 & \equiv \frac{1}{\Sigma_1} \left[\frac{2}{\pi} \arcsin\left(\frac{R_0}{Y_0}\right) + \frac{2}{\pi} \frac{R_1}{Y_0} \right], \end{aligned}$$

thus:

$$\begin{aligned} \mathcal{B}_{\mathbf{k}_2}^a & \equiv \sum_{\tau_{\mathbf{k}_2}^a = \pm 1} \exp \left(i \frac{\tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a \tau_{\mathbf{k}_2}^a}{\sqrt{K_2}} \right) \int \mathcal{D}x_{\mathbf{k}_2}^a \mathcal{H} \left(-\frac{\tau_{\mathbf{k}_2}^a (\Gamma_{\mathbf{k}_2}^a + \mathcal{R}_{\mathbf{k}_3}^a / \sqrt{K_2})}{\sqrt{1 - (2/\pi) \arcsin(R_0/Y_0)}} \right) \\ & \simeq 1 + i \frac{\tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a}{\sqrt{K_2}} \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1 - \mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\mathcal{R}_{\mathbf{k}_3}^a}{\sqrt{2K_2 \Sigma_1 (1 - \mathcal{G}_1)}} \right) - \frac{(\hat{\eta}_{\mathbf{k}_3}^a)^2}{2K_2} + O(K_2^{-3/2}). \end{aligned}$$

The product over replica indexes leads to:

$$\begin{aligned} \prod_a \mathcal{B}_{\mathbf{k}_2}^a &\simeq 1 - \frac{1}{2K_2} \sum_a (\hat{\eta}_{\mathbf{k}_3}^a)^2 + \\ &+ \frac{i}{\sqrt{K_2}} \sum_a \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right) - \\ &- \frac{1}{K_2} \sum_{a<b} \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right) \times \\ &\times \tau_{\mathbf{k}_3}^b \hat{\eta}_{\mathbf{k}_3}^b \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^b}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right) + O(K_2^{-3/2}). \end{aligned}$$

By integrating over $x_{\mathbf{k}_2}$, and disregarding terms of $O(K_2^{-3/2})$, we obtain:

$$\begin{aligned} \int \mathcal{D}x_{\mathbf{k}_2} \prod_a \mathcal{B}_{\mathbf{k}_2}^a &\simeq 1 - \frac{1}{2K_2} \sum_a (\hat{\eta}_{\mathbf{k}_3}^a)^2 + \\ &+ \frac{i}{\sqrt{K_2}} \sum_a \mathcal{I}(\Upsilon_{\mathbf{k}_3}^a) \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a - \frac{1}{K_2} \sum_{a<b} \mathcal{I}(\Upsilon_{\mathbf{k}_3}^a, \Upsilon_{\mathbf{k}_3}^b) \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a \tau_{\mathbf{k}_3}^b \hat{\eta}_{\mathbf{k}_3}^b, \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}(\Upsilon_{\mathbf{k}_3}^a) &\equiv \int \mathcal{D}x_{\mathbf{k}_2} \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right) \\ \mathcal{I}(\Upsilon_{\mathbf{k}_3}^a, \Upsilon_{\mathbf{k}_3}^b) &\equiv \int \mathcal{D}x_{\mathbf{k}_2} \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right) \times \\ &\times \operatorname{erf} \left(\sqrt{\frac{\mathcal{G}_1}{1-\mathcal{G}_1}} \frac{x_{\mathbf{k}_2}}{\sqrt{2}} + \frac{\Upsilon_{\mathbf{k}_3}^b}{\sqrt{2K_2\Sigma_1(1-\mathcal{G}_1)}} \right). \end{aligned}$$

The first integral, though can be exactly obtained, should be computed up to leading order in K_2 :

$$\mathcal{I}(\Upsilon_{\mathbf{k}_3}^a) = \operatorname{erf} \left(\frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1}} \right) \simeq \sqrt{\frac{2}{\pi K_2\Sigma_1}} \Upsilon_{\mathbf{k}_3}^a + O(K_2^{-3/2}).$$

The second integral produces the following expression:

$$\begin{aligned} \mathcal{I}(\Upsilon_{\mathbf{k}_3}^a, \Upsilon_{\mathbf{k}_3}^b) &= -1 + \operatorname{erf} \left(\frac{\Upsilon_{\mathbf{k}_3}^a}{\sqrt{2K_2\Sigma_1}} \right) + \operatorname{erf} \left(\frac{\Upsilon_{\mathbf{k}_3}^b}{\sqrt{2K_2\Sigma_1}} \right) + \\ &+ 4 \int_{\frac{\Upsilon_{\mathbf{k}_3}^a - \Upsilon_{\mathbf{k}_3}^b / \mathcal{G}_1}{\sqrt{2K_2\Sigma_1}}}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(z + \frac{\Upsilon_{\mathbf{k}_3}^b}{\mathcal{G}_1 \sqrt{2K_2\Sigma_1}} \right)^2 \right] \mathcal{H} \left(-\sqrt{\frac{\mathcal{G}_1^2}{1-\mathcal{G}_1^2}} z \right) \\ &\simeq \frac{2}{\pi} \arcsin(\mathcal{G}_1) + O(K_1^{-1/2}). \end{aligned}$$

Thus

$$\begin{aligned} \prod_{\mathbf{k}_2} \int \mathcal{D}x_{\mathbf{k}_2} \prod_a \mathcal{B}_{\mathbf{k}_2}^a &\simeq \int \mathcal{D}z_{\mathbf{k}_3} \exp \left[-\frac{1}{2} \left[1 - \frac{2}{\pi} \arcsin(\mathcal{G}_1) \right] \sum_a (\hat{\eta}_{\mathbf{k}_3}^a)^2 + \right. \\ &\left. + i \sqrt{\frac{2}{\pi}} \arcsin(\mathcal{G}_1) \sum_a \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a z_{\mathbf{k}_3} + i \sqrt{\frac{2}{\pi\Sigma_1}} \sum_a \tau_{\mathbf{k}_3}^a \hat{\eta}_{\mathbf{k}_3}^a \Upsilon_{\mathbf{k}_3}^a \right] + O(K_2^{-1/2}). \end{aligned}$$

By redefining $\Upsilon_{\mathbf{k}_3}^a$ as

$$\begin{aligned}\Upsilon_{\mathbf{k}_3}^a &\equiv \Gamma_{\mathbf{k}_3}^a + \Upsilon_{\mathbf{k}_4}^a / \sqrt{K_3} \\ \Gamma_{\mathbf{k}_3}^a &\equiv \sqrt{\left(\frac{2}{\pi}\right)^2 \frac{1}{\Sigma_1} \left(\frac{\Omega_2}{Y_0} + \frac{1-\beta}{\beta} \frac{\Delta_2}{Y_0} - \frac{R_2}{Y_0}\right) x_{\mathbf{k}_3}^a + \frac{2}{\pi} \arcsin(\mathcal{G}_1) + \left(\frac{2}{\pi}\right)^2 \frac{1}{\Sigma_1} \frac{R_2}{Y_0} x_{\mathbf{k}_3}^a} \\ \Upsilon_{\mathbf{k}_4}^a &\equiv \sum_{m=3}^L \left(\prod_{j=4}^m K_j^{-1/2}\right) \frac{2}{\pi \sqrt{\Sigma_1}} \left[\sqrt{\frac{\Omega_m}{Y_0} + \frac{1-\beta}{\beta} \frac{\Delta_m}{Y_0} - \frac{R_m}{Y_0}} x_{\mathbf{k}_{m+1}}^a + \sqrt{\frac{R_m}{Y_0}} x_{\mathbf{k}_{m+1}}^a \right],\end{aligned}$$

we obtain that

$$\int \prod_a \mathcal{D}(\eta_{\mathbf{k}_3}^a, \hat{\eta}_{\mathbf{k}_3}^a) \prod_{\mathbf{k}_2} \mathcal{B}_{\mathbf{k}_2}^a \simeq \prod_a \mathcal{H} \left(-\frac{\tau_{\mathbf{k}_3}^a (\Gamma_{\mathbf{k}_3}^a + \Upsilon_{\mathbf{k}_4}^a / \sqrt{K_3})}{\sqrt{1 - (2/\pi) \arcsin(\mathcal{G}_1)}} \right),$$

and the process can be repeated, following the procedure starting at (A.2), until:

$$\exp(G_E) \simeq \exp \left(nK \sqrt{\frac{\Delta_0}{\beta Y_0}} \right) \int \mathcal{D}z \mathcal{H}^n \left(-\sqrt{\frac{\mathcal{G}_L}{1 - \mathcal{G}_L}} z \right),$$

where

$$\begin{aligned}\Sigma_0 &\equiv 1 \\ \Sigma_\ell &\equiv 1 + \sum_{j=1}^{\ell} \left(\frac{2}{\pi}\right)^j \left[\frac{\Omega_j}{Y_0} + \frac{1-\beta}{\beta} \frac{\Delta_j}{Y_0} \right] \\ \mathcal{G}_0 &\equiv \frac{R_0}{Y_0} \\ \mathcal{G}_\ell &\equiv \frac{\Sigma_{\ell-1}}{\Sigma_\ell} \left[\frac{2}{\pi} \arcsin(\mathcal{G}_{\ell-1}) + \left(\frac{2}{\pi}\right)^\ell \frac{R_\ell}{\Sigma_{\ell-1} Y_0} \right],\end{aligned}$$

thus, for small n

$$G_E \simeq n \left\{ K \sqrt{\frac{\Delta_0}{\beta Y_0}} + \int \mathcal{D}z \log \mathcal{H} \left(-\sqrt{\frac{\mathcal{G}_L}{1 - \mathcal{G}_L}} z \right) \right\} + O(n^2).$$

Appendix B. Expansion for the Gardner's error function

The integral we need to compute is

$$I^a \equiv \int \mathcal{D}x_{\mathbf{k}_1}^a \prod_{\alpha=1}^{\beta} \mathcal{H} \left(-\frac{\tau_{\mathbf{k}_1}^a}{\sqrt{\Delta_0}} \left(\Gamma_{\mathbf{k}_1}^a + \Upsilon_{\mathbf{k}_2}^{\alpha,a} / \sqrt{K_1} \right) \right),$$

where

$$\Gamma_{\mathbf{k}_1}^a \equiv \sqrt{D_0} x_{\mathbf{k}_1}^a + \Gamma_{\mathbf{k}_2}^a$$

and $\Upsilon_{\mathbf{k}_2}^{\alpha,a}$ and $\Gamma_{\mathbf{k}_2}^a$ are variables of $O(1)$, independent on $x_{\mathbf{k}_1}^a$. By defining the parameters

$$\delta \equiv \sqrt{\frac{\Delta_0}{D_0}}, \quad G \equiv \tau_{\mathbf{k}_1}^a \frac{\Gamma_{\mathbf{k}_2}^a}{\sqrt{D_0}}, \quad \epsilon^\alpha \equiv \tau_{\mathbf{k}_1}^a \frac{\Upsilon_{\mathbf{k}_2}^{\alpha,a}}{\sqrt{K_1 D_0}}$$

we can re-write the integral as

$$\begin{aligned} I &\equiv \int \mathcal{D}\mu \prod_{\alpha=1}^{\beta} \mathcal{H}\left(-\frac{\mu + G + \epsilon^\alpha}{\delta}\right). \\ &= \frac{1}{\sqrt{2\pi}} \int d\mu \exp\left[-\frac{1}{2}(\mu - G)^2\right] \prod_{\alpha=1}^{\beta} \mathcal{H}\left(-\frac{\mu + \epsilon^\alpha}{\delta}\right) \end{aligned} \quad (\text{B.1})$$

We are interested in the case where $0 < \delta \ll 1$ and $|\epsilon^\alpha| \ll 1$. Without loss of generality we can consider that $\epsilon^i > \epsilon^{i+1}$. Gardner's error function presents two distinctive behaviors depending on its argument. Given $\mathcal{H}(-x/\delta)$, for all $x \in (-C\delta, C\delta)$, for a suitable $1 \ll C$ and $C\delta \ll 1$, the function increases rapidly. This is the so-called active region. For all $x \notin (-C\delta, C\delta)$, the function admits the following expansion

$$\mathcal{H}(-x/\delta) = \Theta(x) - \frac{\delta e^{-x^2/(2\delta^2)}}{\sqrt{2\pi}x} + D\delta^3 \quad (\text{B.2})$$

for a suitable $D \in \mathbb{R}$. Observe that for all $\mu < \varepsilon - C\delta$, the product is dominated by an exponentially decreasing behavior, whereas for all $\mu > \varepsilon + C\delta$, the product is almost one. Let us define the quantities: $\varepsilon \equiv \epsilon^\beta$ and $E = \epsilon^1$. In order to find bounds for this integral, and using the analysis of the previous paragraph, we have that:

$$\begin{aligned} c\delta &< \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon-C\delta}^{-\varepsilon+C\delta} d\mu \exp\left[-\frac{1}{2}(\mu - G)^2\right] \prod_{\alpha=1}^{\beta} \mathcal{H}\left(-\frac{\mu + \epsilon^\alpha}{\delta}\right) < 2C\delta \\ 0 &< \frac{1}{\sqrt{2\pi}} \int_{-E-C\delta}^{-\varepsilon-C\delta} d\mu \exp\left[-\frac{1}{2}(\mu - G)^2\right] \prod_{\alpha=1}^{\beta} \mathcal{H}\left(-\frac{\mu + \epsilon^\alpha}{\delta}\right) < e^{-C^2/2}(E - \varepsilon), \end{aligned}$$

where $0 < c < C \in \mathbb{R}$. In the region that remains, i.e. $\mathbb{O}_1 \cup \mathbb{O}_2$ with $\mathbb{O}_1 \equiv (-\infty, -E - C\delta)$ and $\mathbb{O}_2 \equiv (-\varepsilon + C\delta, \infty)$, we have that

$$\begin{aligned} I_{\mathbb{O}} &\equiv I_{\mathbb{O}_1} + I_{\mathbb{O}_2} \\ I_{\mathbb{O}_1} &\equiv \frac{1}{\sqrt{2\pi}} \int_{\mathbb{O}_1} d\mu \exp\left[-\frac{1}{2}(\mu - G)^2\right] \prod_{\alpha=1}^{\beta} \left[\frac{\delta \exp(-(\mu + \epsilon^\alpha)^2/(2\delta^2))}{\sqrt{2\pi}|\mu + \epsilon^\alpha|} + D\delta^3 \right] \\ I_{\mathbb{O}_2} &\equiv \frac{1}{\sqrt{2\pi}} \int_{\mathbb{O}_2} d\mu \exp\left[-\frac{1}{2}(\mu - G)^2\right] \prod_{\alpha=1}^{\beta} \left[1 - \frac{\delta \exp(-(\mu + \epsilon^\alpha)^2/(2\delta^2))}{\sqrt{2\pi}(\mu + \epsilon^\alpha)} + D\delta^3 \right] \end{aligned}$$

where the first term can be bounded by

$$e^{-c\beta\delta} \mathcal{H}(-G - \varepsilon + C\delta) < T_1 < e^{c\beta\delta} \mathcal{H}(-G - E + C\delta).$$

The second term can be rearranged by considering the following:

$$\exp\left[-\frac{1}{2\delta^2} \sum_{\alpha} (\mu + \epsilon^\alpha)^2 + \beta d' \log \delta\right] < \prod_{\alpha=1}^{\beta} \left[\frac{\delta \exp(-(\mu + \epsilon^\alpha)^2/(2\delta^2))}{\sqrt{2\pi}|\mu + \epsilon^\alpha|} + D\delta^3 \right]$$

and

$$\prod_{\alpha=1}^{\beta} \left[\frac{\delta \exp(-(\mu + \epsilon^\alpha)^2/(2\delta^2))}{\sqrt{2\pi}|\mu + \epsilon^\alpha|} + D\delta^3 \right] < \exp\left[-\frac{1}{2\delta^2} \sum_{\alpha} (\mu + \epsilon^\alpha)^2 + \beta d \log \delta\right]$$

for suitable $0 < d, d' \in \mathbb{R}$. Thus, disregarding terms of $O(\beta\delta^2)$, the second integral can be bounded from below by:

$$\sqrt{\frac{\delta^2}{\beta + \delta^2}} \exp \left[-\frac{1}{2} \frac{\beta}{\beta + \delta^2} (G + \bar{\varepsilon})^2 - \frac{1}{2} \frac{\beta}{\delta^2} \bar{\sigma}_\varepsilon^2 + \beta d' \log \delta \right] \times \\ \times \mathcal{H} \left(\sqrt{\frac{\delta^2}{\beta + \delta^2}} G - \frac{\beta}{\delta} \frac{\bar{\varepsilon}}{\sqrt{\beta + \delta^2}} + E + C\delta \right)$$

and from above by:

$$\sqrt{\frac{\delta^2}{\beta + \delta^2}} \exp \left[-\frac{1}{2} \frac{\beta}{\beta + \delta^2} (G + \bar{\varepsilon})^2 - \frac{1}{2} \frac{\beta}{\delta^2} \bar{\sigma}_\varepsilon^2 + \beta d \log \delta \right] \times \\ \times \mathcal{H} \left(\sqrt{\frac{\delta^2}{\beta + \delta^2}} G - \frac{\beta}{\delta} \frac{\bar{\varepsilon}}{\sqrt{\beta + \delta^2}} + E + C\delta \right),$$

where

$$\bar{\varepsilon} \equiv \frac{1}{\beta} \sum_{\alpha} \epsilon^{\alpha} \\ \bar{\sigma}_\varepsilon^2 \equiv \frac{1}{\beta} \sum_{\alpha} (\epsilon^{\alpha})^2 - \bar{\varepsilon}^2.$$

Thus, by defining $B_\delta \equiv \beta/\delta^2$, and taking the limits $\beta, \delta, \epsilon^\alpha, B_\delta \bar{\sigma}_\varepsilon^2 \rightarrow 0$ and $1 \ll B_\delta$, the integral approaches the following expression

$$I = \mathcal{H}(-G - \bar{\varepsilon}) + \frac{1}{2} \sqrt{\frac{1}{B_\delta}} \exp \left\{ -\frac{(G + \bar{\varepsilon})^2}{2} \right\}, \quad (\text{B.3})$$

where we substituted ε by $\bar{\varepsilon}$ in the first term of the RHS.

References

- [1] Lee T and Munford D 2003 *J. Opt. Soc. Amer.* **20** 1434
- [2] Lee T Munford D Romero R and Lamme V 1998 *Vision. Res.* **38** 2429
- [3] Hinton G E 2007 *Trends Cog. Scs.* **11** 428
- [4] Bengio 2009 *Fund. and Trends in Machine Learning* **2** 1
- [5] Amit D J 1989 *Modeling brain function* (Cambridge Univ. Press, New York)
- [6] Gardner E 1987 *Europhys. Lett.* **4** 481; 1988 *J. Phys. A* **21** 257; Gardner E and Derrida B 1988 *ibid.* **21** 271
- [7] Seung H S Sompolinsky H and Tishby N 1992 *Phys. Rev. A* **45** 6056
- [8] Watkin T L H Rau A and Biehl M 1993 *Rev. Mod. Phys.* **65** 499
- [9] Barkai E Hansel D and Kanter I 1990 *Phys. Rev. Lett.* **65** 2312
- [10] Barkai E Hansel D and Sompolinsky H 1992 *Phys. Rev. A* **45** 4146
- [11] Engel A Köhler H M Tschepke F Vollmayr H and Zippelius A 1992 *Phys. Rev. A* **45** 7590
- [12] Oppen M 1994 *Phys. Rev. Lett.* **72** 2113
- [13] Monasson R and O'Kane D 1994 *Europhys. Lett.* **27** 85
- [14] Monasson R and Zecchina R 1995 *Phys. Rev. Lett.* **75** 2432
- [15] Urbanczik R 1997 *J. Phys. A* **30** L387
- [16] Kwon C and Oh J-H 1997 *J. Phys. A* **30** 6273
- [17] Neirrotti J P and Franco L 2010 *J. Phys. A* **43** 445103

- [18] Neirrotti J P 2012 *J. Stat. Phys.* **149** 887
- [19] Mezard M Parisi G and Virasoro M A 1987, *Spin Glasses and Beyond* (World Scientific)
- [20] Rammal R Toulouse G and Virasoro M A 1986 *Rev. Mod. Phys.* **58** 765
- [21] Mitchinson G J and Durbin R M 1989 *Biol. Cybern.* **60** 345
- [22] Lopez B and Kinzel W 1997, *J. Phys. A* **30** 7753