Computational capabilities of multilayer committee machines

J P Neirotti[1] **and L Franco**[2]

[1] NCRG, Aston University, Birmingham, UK
[2] Depto. de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Spain

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

Abstract

We obtained an analytical expression for the computational complexity of many layered committee machines with a finite number of hidden layers ($L < \infty$) using the generalization complexity measure introduced by Franco *et al* (2006) *IEEE Trans. Neural Netw.* **17** 578. Although our result is valid in the large-size limit and for an overlap synaptic matrix that is ultrametric, it provides a useful tool for inferring the appropriate architecture a network must have to reproduce an arbitrary realizable Boolean function.

PACS numbers: 89.70.Eg, 84.35.+i

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In this paper we study the computational capabilities of a binary committee machine with an ultrametric synaptic overlap matrix. In many first-approach studies involving neural networks, the single-layered perceptron has been the preferred laboratory [2, 3]. Perceptrons are the most fundamental networks and, as is expected, their computation capabilities are limited. The next level in terms of architectural complexity is represented by networks with at least one hidden layer and a fixed hidden-to-output relation. A network with an arbitrary number, K, of units in its only hidden layer suffices to reproduce nontrivial scalar functions of N-dimensional variables. Exact representation of Boolean functions requires at most $K \sim O(2^N)$ units; continuous functions can be approximated with arbitrary accuracy if the number, K, of units is not constrained [4]. There is an extensive number of studies done on this type of network [5–8], mostly in the area of learning theory [9–18] and applications [19–22]. In the present paper we focus on quantifying the *complexity* of a binary committee machine, with no more hidden-to-input units K than inputs N and with a finite number of hidden layers L. This difficulty measure, or complexity, can be used to assess which network architecture is more convenient to reproduce a given realizable Boolean function.

Attempts to quantify the complexity of an object have given rise to many formal definitions of complexity [23–27]. Recently Franco [1] proposed to quantify a function's complexity by the size of the minimal set of inputs needed to train a feed-forward network, with a predetermined architecture, until reaching the zero prediction error. It was found that this minimal set is mainly formed by pairs of inputs with different classification and so adjacent to the classification boundary [28]. Further investigation showed that the average fraction of the neighboring pairs with different classification (or average discrepancy) is correlated with the generalization ability of the network implementing the function in such a way that the higher the fraction, the larger the effort required to reach complete generalization. They naturally concluded that the average discrepancy is a measure of how hard it is to reproduce the function's behavior [29].

The average discrepancy has been used successfully for inferring the architecture of very simple machines. In [2] we used Franco's complexity measure for inferring the dilution coefficient in perceptrons: results that have been recently confirmed by other means [3]. To use Franco's complexity for inferring the suitable committee for reproducing a Boolean function, we need to characterize both the Boolean functions to be mimicked and the committees to be used. In section 2 we present the expression for the average discrepancy for the $(L+2)$-layered committees. In section 3 we study some particular cases and finite-size dependences of the expression found in section 2. In section 4 we outline the Boolean functions that can be potentially described by the architectures explored in section 2. In the final section we discuss our conclusions and further possible investigations.

2. Average discrepancy for ultrametric committees

Let us consider a committee machine with L hidden layers and $K \leqslant N$ hidden-to-input units. All units of all layers are binary. These 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 Lth hidden layer, each connected to K_{L-1} units in the $(L-1)$ level. The total number of units in the $(L-1)$ 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 a root at the node. Connections from units at the ℓth hidden layer to units at the $(\ell+1)$ layer are all set to one. To single out the variables of the ℓth layer we will use the notation $\mathbf{k}_\ell \equiv [k_L, k_{L-1}, \ldots, k_\ell] = \mathbf{k}_{\ell+1} k_\ell$, which runs over all hidden units of the ℓth layer. Thus

$$\sigma(\mathbf{S}) \equiv \text{sgn}\left(\sum_{k_L=1}^{K_L} \sigma_{k_L}(\mathbf{S})\right) \tag{1a}$$

$$\sigma_{\mathbf{k}_\ell}(\mathbf{S}) \equiv \text{sgn}\left(\sum_{k_{\ell-1}=1}^{K_{\ell-1}} \sigma_{\mathbf{k}_\ell k_{\ell-1}}(\mathbf{S})\right) \tag{1b}$$

$$\sigma_{\mathbf{k}_1}(\mathbf{S}) \equiv \text{sgn}\left(\frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T} \mathbf{S}}{\sqrt{N}}\right), \tag{1c}$$

where $\mathbf{w}_{\mathbf{k}_1} \in \mathbb{S}^N$ is the synaptic vector of the $\mathbf{k}_1$th unit, $\mathbf{w}^\mathsf{T}$ is the transpose of the vector $\mathbf{w}$ and $\mathbf{S} \in \{\pm 1\}^N$. $\mathbb{S}^N$ is the surface of the N-hypersphere of radius $\sqrt{N}$ and $\{\pm 1\}^N$ is the N-hypercube centered at 0. The committee has been constructed by drawing vectors from a suitable distribution over $\mathbb{S}^N$ such that

$$[\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1'} \equiv \frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T} \mathbf{w}_{\mathbf{k}_1'}}{N} = \delta_{\mathbf{k}_1 \mathbf{k}_1'}(1 - \tilde{\zeta}_1) + \cdots + \delta_{\mathbf{k}_L \mathbf{k}_L'}(\tilde{\zeta}_{L-1} - \tilde{\zeta}_L) + \tilde{\zeta}_L, \tag{2}$$
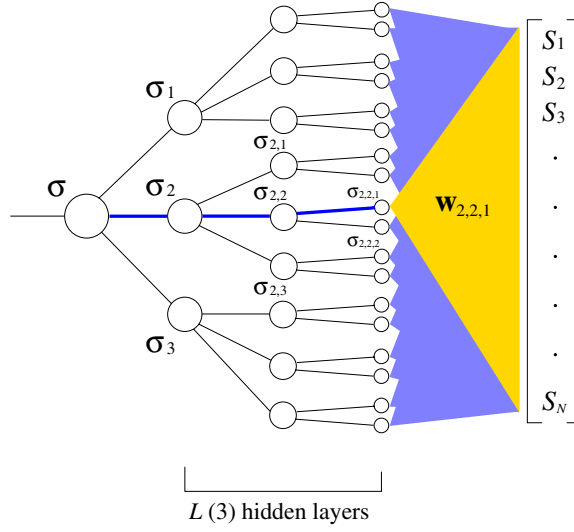
Figure 1. Typical feed-forward network architecture studied in this paper. This committee has $L = 3$ hidden layers with synaptic overlaps in the first hidden-to-input layer. All the synaptic weights linking hidden-to-hidden and hidden-to-output units are set to one. Observe that the highlighted synaptic vector $\mathbf{w}_{2,2,1}$ corresponds to the output of the input layer path $\mathbf{k}_1 = (2, 2, 1)$, i.e. the second unit of the third hidden layer, second unit second hidden layer, first unit first hidden layer.

where $\tilde{\zeta}_\ell$ is the overlap between synaptic vectors belonging to units that have a common root in the ℓth hidden layer (see figure 1) and $\delta_{\mathbf{k}_\ell \mathbf{k}'_\ell} \equiv \prod_{m=\ell}^{L} \delta_{k_m k'_m}$ and $\delta_{ij} = 1$ if and only if $i = j$ and zero otherwise. The structure of this matrix is block diagonal and resembles the matrices used to represent inter-replica interactions [30].

We suppose that the elements of the overlap matrix have a *natural* scaling relationship with the size of the system. The argument that gives support to this conjecture runs as follows: suppose we want to draw $K \leqslant N$ vectors from a uniform distribution over $\mathbb{S}^N$, with the constraint that for any pair of vectors $\mathbf{w}_i$ and $\mathbf{w}_j$ it is satisfied that $\mathbf{w}_i^\mathsf{T} \mathbf{w}_j = N \cos(\alpha)$ for all $i \neq j$ and fixed α. The first vector $\mathbf{w}_1$ drawn can be any vector on the surface of the sphere. The second vector $\mathbf{w}_2$ is one of the vectors located on the hyper-ring with center in $\mathbf{w}_1$ and radius $\sqrt{N}\sin(\alpha)$. Thus the probability $\mathcal{P}(\mathbf{w}_2|\mathbf{w}_1)$ has to be proportional to the volume of this hyper-ring, i.e. $\mathcal{P}(\mathbf{w}_2|\mathbf{w}_1) \propto \sin^{N-2}(\alpha)$. The third vector $\mathbf{w}_3$ has to sustain the same angle with the other two vectors; therefore, the probability has to have a second factor such that $\mathcal{P}(\mathbf{w}_3|\mathbf{w}_1, \mathbf{w}_2)\mathcal{P}(\mathbf{w}_2|\mathbf{w}_1) \propto \sin^{N-2}(\alpha)\sin^{N-3}(\alpha)$. After drawing K vectors we have that the probability of the set of vectors is $\mathcal{P}(\mathbf{w}_K|\mathbf{w}_1, \ldots, \mathbf{w}_{K-1})$ $\mathcal{P}(\mathbf{w}_{K-1}|\mathbf{w}_1, \ldots, \mathbf{w}_{K-2})\ldots\mathcal{P}(\mathbf{w}_2|\mathbf{w}_1) \propto \sin^{N-2}(\alpha)\sin^{N-3}(\alpha)\ldots\sin^{N-K}(\alpha)$. The expected overlap is then

$$\langle \text{overlap} \rangle \equiv \frac{\int_0^{\pi/2} d\alpha \cos(\alpha) \sin^z(\alpha)}{\int_0^{\pi/2} d\alpha \sin^z(\alpha)} = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\dfrac{2+z}{2}\right)}{\Gamma\left(\dfrac{3+z}{2}\right)}$$

$$\simeq \sqrt{\frac{2}{\pi z}} + O(z^{-1}),$$

with an asymptotic variance of

$$\text{var(overlap)} \simeq \frac{\pi - 2}{\pi z} + O(z^{-1}),$$

where $z = N - 2 + N - 3 + \cdots + N - K = \frac{1}{2}(K-1)[2N - K - 2]$. If $1 \geqslant \kappa \equiv K/N$ then

$$\langle \text{overlap} \rangle \sim \sqrt{\frac{4\kappa}{\pi(2-\kappa)}} \frac{1}{K} = \frac{\zeta_1}{K}, \tag{3}$$

where $\zeta_1 \sim O(1)$. Let us suppose now that from the $K = K_1 K_2$ vectors we make K_2 groups of K_1 vectors each, in such a way that the inter-group and extra-group overlaps are fixed to $\tilde{\zeta}_1$ and $\tilde{\zeta}_2$ respectively. We have proven that the inter-group overlap must scale with the inverse of the group size, i.e. $\tilde{\zeta}_1 = \zeta_1/K_1$ and we know that there are a total of $K_1 K_2$ inter-group overlaps and $K_1 K_2 (K_1 K_2 - 1)/2$ extra-group vector overlaps. A crude estimate of the average overlap is then

$$\langle \text{overlap} \rangle \sim \binom{K_1 K_2 + 1}{2}^{-1} \left[\binom{K_1 K_2}{1} \frac{\zeta_1}{K_1} + \binom{K_1 K_2}{2} \tilde{\zeta}_2 \right]$$

$$\sim \tilde{\zeta}_2 + O\left(1/K_1^2 K_2\right)$$

and, according to (3), we must have $\tilde{\zeta}_2 \sim O(1/K_1 K_2)$.

If the above argument is iterated L times we recover the matrix defined in (2) with elements satisfying the scaling relationship:

$$\tilde{\zeta}_\ell = \frac{\zeta_\ell}{\prod_{j=1}^\ell K_j}, \tag{4}$$

where $\zeta_j \sim O(1)$ and $\prod_{j=1}^L K_j = K$.

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

(i) $\mathbf{Q}$ is symmetric, i.e. $[\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1'} = [\mathbf{Q}]_{\mathbf{k}_1', \mathbf{k}_1}$ for all paths $\mathbf{k}_1$ and $\mathbf{k}_1'$.

(ii) $\mathbf{Q}$ only has non-negative entries, i.e. $[\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1'} > 0$ for all paths $\mathbf{k}_1$ and $\mathbf{k}_1'$.

(iii) $[\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1'} \geqslant \min \left\{ [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1''}, [\mathbf{Q}]_{\mathbf{k}_1', \mathbf{k}_1''} \right\}$ for all paths $\mathbf{k}_1$, $\mathbf{k}_1'$ and $\mathbf{k}_1''$.

(iv) $1 = [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1} \geqslant \max \left\{ [\mathbf{Q}]_{\mathbf{k}_1, \mathbf{k}_1'} \ \forall \ \mathbf{k}_1' \neq \mathbf{k}_1 \right\}$.

Therefore, the matrix $\mathbf{Q}$ is *ultrametric*. Given that the overlap matrix $\mathbf{Q}$ is ultrametric we dubbed these networks *ultrametric committee machines*.

The average discrepancy [1, 2] is defined as

$$\mathfrak{d}_P^{(N)}(\mathbf{Q}) = \sum_{\mathbf{S}, \mathbf{S}' \in \{\pm 1\}^N} \mathcal{P}(\mathbf{S}, \mathbf{S}') \frac{1 - \sigma(\mathbf{S})\sigma(\mathbf{S}')}{2},$$

where $\mathbf{S}$ and $\mathbf{S}'$ differ in exactly P entries. We assume that the components of the input vector $\mathbf{S}$ are i.i.d. variables, therefore $\mathcal{P}(\mathbf{S}) = \prod_{\ell=1}^N \mathcal{P}(S_\ell)$ and $\mathcal{P}(S = 1) = \mathcal{P}(S = -1) = \frac{1}{2}$. The input vector $\mathbf{S}'$ is constructed by flipping P entries randomly chosen from $\mathbf{S}$. Thus $\mathcal{P}(\mathbf{S}'|\mathbf{S}) = \binom{N}{P}^{-1} \delta_{N - \mathbf{S} \cdot \mathbf{S}', 2P}$.

Following the developments of appendix A and considering the large system-size limit $P, K \leqslant N \to \infty$ with $p \equiv P/N$ fixed, and by defining

$$\mathcal{G}_1(p) \equiv 1 - \frac{2}{\pi} \arccos(1 - 2p)$$

$$\gamma_\ell \equiv \left(\frac{2}{\pi}\right)^\ell \zeta_\ell \left[1 + \sum_{j=1}^\ell \left(\frac{2}{\pi}\right)^j \zeta_j \right]^{-1} \tag{5}$$

$$\mathcal{G}_{\ell+1}(p) \equiv 1 - \frac{2}{\pi} \arccos \left((1 - \gamma_\ell)\mathcal{G}_\ell(p) + \gamma_\ell(1 - 2p)\right)$$
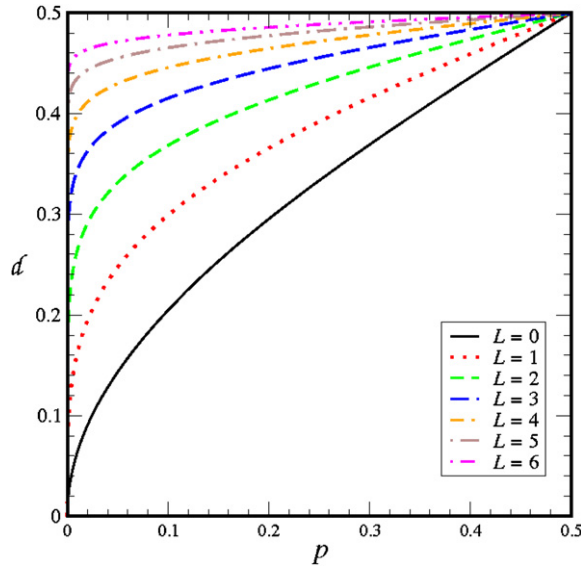
Figure 2. Discrepancy as a function of p for tree committee machines with $L = 0, \ldots, 6$ hidden layers, as obtained by applying (6).

we obtain the following expression:

$$\mathfrak{d}(p, \mathbf{Q}) = \frac{1}{\pi} \arccos\left((1 - \gamma_L)\mathcal{G}_L(p) + \gamma_L(1 - 2p)\right). \tag{6}$$

3. Particular cases and finite-size dependences

If we have that $\zeta_\ell = 0$ for all $\ell = 1, \ldots, L$, the matrix $\mathbf{Q}$ is the identity and the committee has a perfect tree structure. Let us introduce the notation

$$[\![f]\!]^{k+1}(x) = [\![f]\!]^k \circ f(x),$$

where $f \circ g(x) = f(g(x))$. Therefore, if we define $\psi(p) \equiv \arccos(1 - 2p)/\pi$, then

$$\mathfrak{d}(p, \mathbf{I}) = [\![\psi]\!]^{L+1}(p)$$

in agreement with [31] and references therein. Examples of curves obtained by applying (6) for tree committee machines are given in figure 2.

If the overlap of the ℓ_0th layer is too large, i.e. $\lim_{K_1, \ldots, K_{\ell_0} \to \infty} \left(\prod_{\ell=1}^{\ell_0} K_\ell\right)\tilde{\zeta}_{\ell_0}$ diverges, we have that

$$\gamma_\ell = \begin{cases} \left(\frac{2}{\pi}\right)^\ell \zeta_\ell \left[1 + \sum_{j=1}^\ell \left(\frac{2}{\pi}\right)^j \zeta_j\right]^{-1} & \forall \ell < \ell_0 \\ 1 & \ell = \ell_0 \\ 0 & \text{otherwise} \end{cases}$$

which immediately implies that

$$\lim_{\zeta_{\ell_0} \uparrow \infty} \mathfrak{d}(p, \mathbf{Q}) = [\![\psi]\!]^{L - \ell_0 + 1}(p).$$

Therefore, a too large an overlap reduces the computational capabilities of the network by effectively deleting the first ℓ_0 hidden layers.

Equation (6) is valid in the limit $P, K \leqslant N \to \infty$ keeping $P/N = p$ fixed. It is important to note that none of the ratios K_i/N are relevant to this formulation. Therefore, whatever the path taken to increase the number of hidden units, the asymptotic value of the discrepancy is always the same.

In order to check this statement we run a number of numerical experiments in networks with $L = 1, 2$. Firstly we developed an algorithm for generating the set of synaptic vectors satisfying the relationship (2) which has the following general form.

(i) Generate $K = \prod_{l=1}^{L} K_l$ orthonormal vectors characterized by a suitable set of indexes $|j_1, \ldots j_L) \in \mathbb{S}$ (the Gram–Schmidt algorithm may be used), with $1 \leqslant j_l \leqslant K_l$ and where $\mathbb{S}$ is a vector space over $\mathbb{R}$ with an inner product such that

$$(j'_1, \ldots, j'_L | j_1, \ldots j_L) = \prod_{l=1}^{L} \delta_{j'_l, j_l},$$

where $\delta_{j,k}$ is the Kronecker's delta (equal to 1 if $j = k$ and zero otherwise).

(ii) Define the block average vectors as

$$|\mu_l, j_{l+1}, \ldots j_L) \equiv \frac{1}{\sqrt{K_1}} \sum_{j_1=1}^{K_1} \cdots \frac{1}{\sqrt{K_l}} \sum_{j_l=1}^{K_l} |j_1, \ldots j_L).$$

(iii) Define the synaptic vectors as

$$\mathbf{w_{j_1}} = \mathbf{w}_{j_1, \ldots, j_L} \equiv \sqrt{N} \left[a_0 |j_1, \ldots j_L) + \sum_{l=1}^{L} a_l |\mu_l, j_{l+1}, \ldots j_L) \right],$$

where the real coefficients a_l, $0 \leqslant l \leqslant L$ are determined from the equations satisfying the definition of the matrix $\mathbf{Q}$ (2).

In this application we have to solve the following equations for the coefficients:

$$\tilde{\zeta}_1 = \frac{2a_0 a_1}{\sqrt{K_1}} + a_1^2$$

$$1 = \tilde{\zeta}_1 + a_0^2$$

for $L = 1$ and

$$\tilde{\zeta}_2 = \frac{2a_0 a_2}{\sqrt{K_1 K_2}} + \frac{2a_1 a_2}{\sqrt{K_2}} + a_2^2$$

$$\tilde{\zeta}_1 = \tilde{\zeta}_2 + \frac{2a_0 a_1}{\sqrt{K_1}} + a_1^2$$

$$1 = \tilde{\zeta}_1 + a_0^2$$

for $L = 2$.

By constructing networks like the one depicted in figure 1 with output σ given by (1a) and synaptic vectors chosen according to the algorithm described, we estimate the average discrepancy by computing

$$\mathfrak{d}_P^{(N)}[\sigma] \simeq \frac{1}{M} \sum_{m=1}^{M} \frac{1 - \sigma(\mathbf{S}_m)\sigma(\mathbf{S}'_m)}{2}$$

over $M = 10^4$ pairs of inputs $(\mathbf{S}_m, \mathbf{S}'_m)$ differing in $P \approx 0.03N$ bits. For $L = 1$ we worked with networks having $K = N = 279, 465, 651$ and 837 and overlap $\zeta_1 = 1$. The results (empty circles) are ploted in figure 3(a). The error bars have been obtained by repeating the
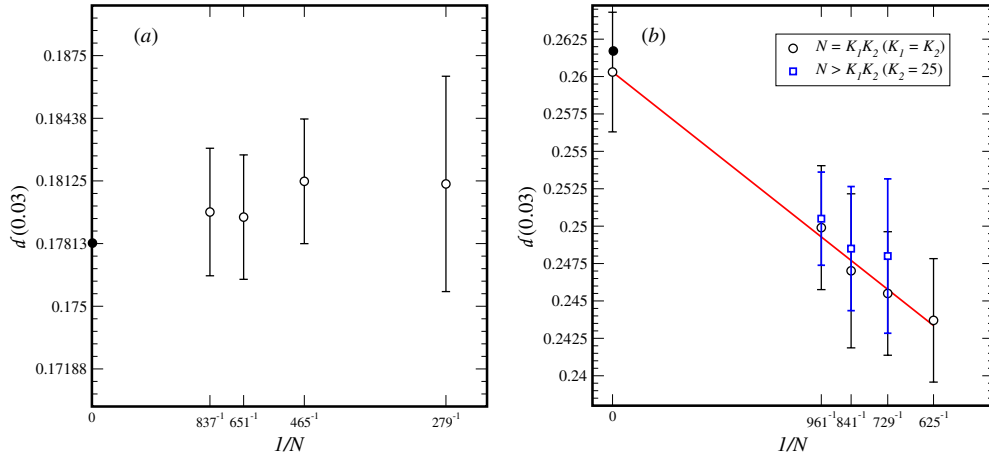
Figure 3. Finite-size dependences of the average discrepancy at $p \approx 0.03$, measured over $M = 10^4$ pairs. For $L = 1$ we worked with networks sizes of $K = N = 279$, 465, 651 and 837 and overlap $\zeta_1 = 1$. The results (empty circles) are plotted in panel (a). The error bars have been obtained by repeating the experiment ten times and computing the variance of the averages. All the results obtained are indistinguishable with the value obtained by applying equation (6) with $\zeta_1 = 1$ and $p = 0.03$, represented as a full circle. For $L = 2$ we perform two experiments, panel (b). We firstly considered networks with $\zeta_1 = \zeta_2 = 1$ and $K_1 = K_2 = \sqrt{N}$ for $N = 625$, 729, 841 and 961. The results are represented by empty circles and the error bars were computed as in the previous experiment. The second experiment was conducted with $\zeta_1 = \zeta_2 = 1$ and sizes of $25 = K_2 < K_1 = 27$, 29 and 31 and $N = 729$, 841 and 961 respectively (empty squares). Results with equal N have been found to be indistinguishable. Moreover, a linear regression produces an extrapolated value of the discrepancy of 0.260 ± 0.004 that is indistinguishable to 0.2617, the one obtained by applying (6) with $\zeta_1 = \zeta_2 = 1$ and $p = 0.03$ (full circle).

experiment ten times and computing the variance of the averages. All the results obtained are indistinguishable with the value obtained by applying equation (6) with $\zeta_1 = 1$ and $p = 0.03$, represented as a full circle.

For $L = 2$ we perform two experiments, with sampling size of 10^4. The first with $\zeta_1 = \zeta_2 = 1$ and $K_1 = K_2 = \sqrt{N}$ for $N = 625$, 729, 841 and 961. The results are represented by empty circles in figure 3(b). The error bars were computed as in the previous experiment. The second experiment was conducted with $\zeta_1 = \zeta_2 = 1$ and $25 = K_2 < K_1 = 27$, 29 and 31 and $N = 729$, 841 and 961 respectively (empty squares). Results with equal N have been found indistinguishable. Moreover, a linear regression produces an extrapolated value of the discrepancy of 0.260 ± 0.004 that is indistinguishable to 0.2617, the one obtained by applying (6) with $\zeta_1 = \zeta_2 = 1$ and $p = 0.03$ (full circle).

4. Boolean functions with continuous discrepancies

In this section we will study the Boolean functions that can be implemented by the the networks presented in figure 1. Expression (6) provides a family of curves that densely cover the set of possible discrepancy values. It is clear from (1b) that any function approached by these networks must be odd, i.e. $f(\mathbf{S}) = -f(-\mathbf{S})$. It is also clear that (6) is continuous in p; thus, we expect the discrepancy functional to be also continuous, i.e.

$$\forall\, 0 \leqslant P < \infty \text{ and } 0 < \varepsilon \in \mathbb{R}, \exists N_0 \in \mathbb{N} \Big/ \left| \mathfrak{d}_P^{(N_0)}[f] - \mathfrak{d}_{P+1}^{(N_0)}[f] \right| < \varepsilon. \tag{7}$$

In order to analyze what type of functions satisfy (7) we will use the decomposition of a function into an orthogonal basis set. The main development of this section is based on [32].

Let f and $g : \{\pm 1\}^N \to \mathbb{R}$. Consider the inner product defined by

$$\langle f | g \rangle \equiv \frac{1}{2^N} \sum_{\{\mathbf{S}\}} f(\mathbf{S}) g(\mathbf{S})$$

and also consider the set of parity functions $\chi_{\mathbb{I}}(\mathbf{S}) \equiv \prod_{k \in \mathbb{I}} S_k$, where $\mathbb{I} \subseteq [N] = \{1, 2, \ldots, N\}$. Observe that

$$\langle \chi_{\mathbb{I}} | \chi_{\mathbb{J}} \rangle = \frac{1}{2^N} \sum_{\{\mathbf{S}\}} \prod_{k \in \mathbb{I} \triangle \mathbb{J}} S_k = \delta_{\mathbb{I}, \mathbb{J}},$$

where $\mathbb{I} \triangle \mathbb{J} \equiv \mathbb{I} \cup \mathbb{J}/\mathbb{I} \cap \mathbb{J} = \{k \in [N] | k \in \mathbb{I} \mathbf{\,xor\,} k \in \mathbb{J}\}$ and $\delta_{\mathbb{I}, \mathbb{J}} = 1$ if both sets are equal (i.e. $\mathbb{I} \subset \mathbb{J}$ and $\mathbb{J} \subset \mathbb{I}$) and zero otherwise. Given a Boolean function $f : \{\pm 1\}^N \to \{\pm 1\}$ we define the Fourier amplitudes as $\hat{f}_{\mathbb{I}} \equiv \langle f | \chi_{\mathbb{I}} \rangle$. It is straightforward to prove that $1 = \sum_{\mathbb{I}} \hat{f}_{\mathbb{I}}^2$. The Fourier spectrum of f is defined as the set

$$\mathscr{S}[f] \equiv \{\hat{f}_{\mathbb{I}} | \mathbb{I} \subseteq [N]\} \tag{8}$$

By following the developments of appendix D and considering the definition

$$\mathscr{A}_r \equiv \sum_{\mathbb{I}_r} \hat{f}_{\mathbb{I}_r}^2, \tag{9}$$

where the sum is over all the sets of indexes $\mathbb{I}_r$ with exactly r elements, we have that the Fourier components of the function must satisfy the following condition:

$$\forall j \in \mathbb{N} \quad \lim_{N \to \infty} \frac{1}{N^j} \sum_{r=1}^N \mathscr{A}_r r^j = 0. \tag{10}$$

Thus the functions that can be approached by committees like the one depicted in figure 1 must have a Fourier spectrum satisfying the conditions elicited by (10). Observe that the functions not satisfying these conditions are those with Fourier components corresponding to parity functions constructed using a macroscopic (i.e. $O(N)$) number of bits. Trivially, the *full* parity machine $\chi_{[N]}$ does not satisfy (10) and its discrepancy curve is, in the large N limit, discontinuous everywhere.

Consider now the function $\mathsf{TRIBE}_n(\mathbf{S})$ which is defined by applying the AND function to the output of n OR functions, which in turn are applied to n equal and disjoint segments of the input vector $\mathbf{S}$ (see figure 4). Although the functions $\mathsf{TRIBE}_n(\mathbf{S})$ are not odd, they satisfy (10) (its Fourier spectrum is reported in [31]); they can easily be constructed from logical gates and have also been used in studies of noise sensitivity [31, 33]. We are currently working on developing methods for approximating TRIBE_n functions with ultrametric committees.

5. Conclusions

The most important result of this paper is expression (6). This is the expression of the average discrepancy of a binary committee machine with $L + 2$ layers as a function of the overlap parameters $\zeta_1, \ldots, \zeta_L$. It is clear that, for a fixed value of $p = P/N$, by varying these parameters the committee's average discrepancy may take any possible value between the discrepancy of a simple perceptron, i.e. $\frac{1}{\pi} \arccos(1 - 2p)$, and $\frac{1}{2}$. Moreover, for a fixed value of p, the larger the L and the smaller the overlap parameters, the larger the discrepancy. This suggests that the smaller the overlap parameters for large networks, the larger the ability of the committee to compute Boolean functions with a continuous discrepancy curve.
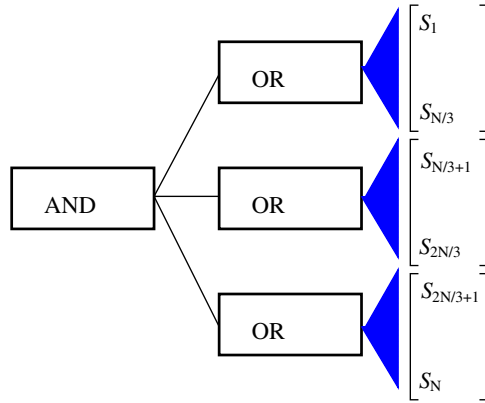
Figure 4. Graphic representation of the function $\mathsf{TRIBE}_3(\mathbf{S})$.

Most results in committee machines [14, 17, 18] are obtained with a finite number of hidden units. In our case we consider the regime with both large number of input and hidden units. Although the average discrepancy appears to be a quantity that develops finite-size dependences, expression (6) represents a good approach even for the smallest system size studied in this paper.

Also observe that the relative errors in our numerical calculations are of the order of 2%. We intend to continue the study of these finite-size effects in order to obtain scaling laws that may be useful for computing the properties of finite systems. These proposed studies are limited by the number of hidden layers considered, given that the more hidden layers the larger the number of parameters $(K_1, \ldots, K_L)$ we need to increment to infinity.

Acknowledgments

Appendix A. Calculation of (6)

By using the definition (1b) and the properties of the Heaviside function we can see that

$$
\Theta(\pm \sigma(\mathbf{S})) = \Theta\left(\pm \sum_{k_L=1}^{K_L} \sigma_{k_L}(\mathbf{S})\right)
$$

$$
= \sum_{\tau_1=\pm 1} \cdots \sum_{\tau_{K_L}=\pm 1} \Theta\left(\pm \sum_{k_L=1}^{K_L} \tau_{k_L}\right) \prod_{k_L=1}^{K_L} \Theta\left(\tau_{k_L}\sigma_{k_L}(\mathbf{S})\right)
$$

$$
= \sum_{\{\tau_L\}} \Theta\left(\pm \sum_{k_L=1}^{K_L} \tau_{k_L}\right) \prod_{k_L=1}^{K_L} \Theta\left(\tau_{k_L}\sigma_{k_L}(\mathbf{S})\right),
$$

where $\boldsymbol{\tau}_L = (\tau_1, \ldots, \tau_{K_L})$ are binary variables used to represent the internal state of the hidden units in the Lth layer and $\sum_{\{\tau_L\}} \equiv \sum_{\tau_1=\pm 1} \cdots \sum_{\tau_{K_L}=\pm 1}$. Applying this identity again

9

we obtain

$$\Theta(\pm\sigma(\mathbf{S})) = \sum_{\{\tau_L\}} \Theta\left(\pm\sum_{k_L=1}^{K_L}\tau_{k_L}\right) \prod_{k_L=1}^{K_L}\sum_{\{\tau_{L-1}^{k_L}\}} \Theta\left(\tau_{k_L}\sum_{k_{L-1}=1}^{K_{L-1}}\tau_{k_L,k_{L-1}}\right)\prod_{k_{L-1}=1}^{K_{L-1}}\Theta\left(\tau_{k_L,k_{L-1}}\sigma_{k_L,k_{L-1}}(\mathbf{S})\right),$$

where $\tau_{L-1}^{k_L} \equiv \left(\tau_{k_L,1}, \tau_{k_L,2}, \ldots, \tau_{k_L,K_{L-1}}\right)$ is the set of binary variables used to represent the state of the hidden units at the $(L-1)$ layer, with root at the k_L unit in the Lth layer. If we use the vectorial notation for the indexes, we can define the vector $\tau_{\ell-1}^{\mathbf{k}_\ell} \equiv \left(\tau_{\mathbf{k}_\ell,1}, \tau_{\mathbf{k}_\ell,2}, \ldots, \tau_{\mathbf{k}_\ell,K_{\ell-1}}\right)$ as the vector where the component $\tau_{\mathbf{k}_\ell,m}$ represents the state of the mth unit at the $(\ell-1)$ layer with root at the $\mathbf{k}_\ell$th unit in the ℓth layer. In this way we obtain that

$$\Theta(\pm\sigma(\mathbf{S})) = \sum_{\{\tau_L\}} \Theta\left(\pm\sum_{k_L=1}^{K_L}\tau_{k_L}\right) \prod_{k_L=1}^{K_L}\prod_{\ell=1}^{L-1}\sum_{\{\tau_{L-\ell}^{\mathbf{k}_{L-\ell+1}}\}} \Theta\left(\tau_{\mathbf{k}_{L-\ell+1}}\sum_{k_{L-\ell}=1}^{K_{L-\ell}}\tau_{\mathbf{k}_{L-\ell+1},k_{L-\ell}}\right)$$

$$\times \prod_{k_{L-\ell}=1}^{K_{L-\ell}}\prod_{k_1=1}^{K_1}\Theta\left(\tau_{\mathbf{k}_1}\frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T}\mathbf{S}}{\sqrt{N}}\right). \tag{A.1}$$

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

$$\Theta(x) = \int_0^\infty \mathrm{d}\eta\,\delta(\eta-x) = \int_0^\infty \mathrm{d}\eta \int_{-\infty}^\infty \frac{\mathrm{d}\hat\eta}{2\pi} \exp\left[-\mathrm{i}\hat\eta(\eta-x)\right] = \int \mathcal{D}(\eta,\hat\eta)\exp(\mathrm{i}\hat\eta x),$$

where we have defined the notation $\int\mathcal{D}(x,\hat x) \equiv \int_0^\infty \mathrm{d}x \int_{-\infty}^\infty \mathrm{d}\hat x\,\exp(-\mathrm{i}x\hat x)/2\pi$. Applying these expressions we have that, by redistributing the products,

$$\Theta(\pm\sigma(\mathbf{S})) = \int \mathcal{D}(\eta_{L+1},\hat\eta_{L+1}) \sum_{\{\tau_L\}} \exp\left(\pm\mathrm{i}\hat\eta_{L+1}\sum_{k_L=1}^{K_L}\tau_{k_L}\right)$$

$$\times \int\left[\prod_{k_L=1}^{K_L}\mathcal{D}(\eta_{k_L},\hat\eta_{k_L})\right]\left(\prod_{k_L=1}^{K_L}\sum_{\{\tau_{L-1}^{k_L}\}}\right)\exp\left(\mathrm{i}\sum_{k_L=1}^{K_L}\hat\eta_{k_L}\tau_{k_L}\sum_{k_{L-1}=1}^{K_{L-1}}\tau_{k_L,k_{L-1}}\right)$$

$$\times \int\left[\prod_{\mathbf{k}_{L-1}}\mathcal{D}(\eta_{\mathbf{k}_{L-1}},\hat\eta_{\mathbf{k}_{L-1}})\right]\left(\prod_{\mathbf{k}_{L-1}}\sum_{\{\tau_{L-2}^{\mathbf{k}_{L-1}}\}}\right)$$

$$\times \exp\left(\mathrm{i}\sum_{\mathbf{k}_{L-1}}\hat\eta_{\mathbf{k}_{L-1}}\tau_{\mathbf{k}_{L-1}}\sum_{k_{L-2}=1}^{K_{L-2}}\tau_{\mathbf{k}_{L-1},k_{L-2}}\right)\cdots$$

$$\times \int\left[\prod_{\mathbf{k}_2}\mathcal{D}(\eta_{\mathbf{k}_2},\hat\eta_{\mathbf{k}_2})\right]\left(\prod_{\mathbf{k}_2}\sum_{\{\tau_1^{\mathbf{k}_2}\}}\right)\exp\left(\mathrm{i}\sum_{\mathbf{k}_2}\hat\eta_{\mathbf{k}_2}\tau_{\mathbf{k}_2}\sum_{k_1=1}^{K_1}\tau_{\mathbf{k}_2,k_1}\right)$$

$$\times \int\left[\prod_{\mathbf{k}_1}\mathcal{D}(\eta_{\mathbf{k}_1},\hat\eta_{\mathbf{k}_1})\right]\exp\left(\mathrm{i}\sum_{\mathbf{k}_1}\hat\eta_{\mathbf{k}_1}\tau_{\mathbf{k}_1}\frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T}\mathbf{S}}{\sqrt{N}}\right),$$

where $\prod_{\mathbf{k}_\ell} \equiv \prod_{k_L=1}^{K_L}\prod_{k_{L-1}=1}^{K_{L-1}}\cdots\prod_{k_\ell=1}^{K_\ell}$ and $\sum_{\mathbf{k}_\ell} \equiv \sum_{k_L=1}^{K_L}\sum_{k_{L-1}=1}^{K_{L-1}}\cdots\sum_{k_\ell=1}^{K_\ell}$. Observe that $\prod_{\mathbf{k}_{\ell+1}}\sum_{\{\tau_\ell^{\mathbf{k}_{\ell+1}}\}}$ is the trace over all the internal representations at the ℓth layer; therefore, if we define the vector τ_ℓ such that $[\tau_\ell]_{\mathbf{k}_\ell} = \tau_{\mathbf{k}_\ell}$ we can re-write the trace as $\sum_{\{\tau_\ell\}} \equiv \prod_{\mathbf{k}_{\ell+1}}\sum_{\{\tau_\ell^{\mathbf{k}_{\ell+1}}\}}$. Equivalently, we can redefine the differentials such that $\mathcal{D}(\eta_\ell,\hat\eta_\ell) \equiv \prod_{\mathbf{k}_\ell}\mathcal{D}(\eta_{\mathbf{k}_\ell},\hat\eta_{\mathbf{k}_\ell})$. Finally

we want to remark that $\sum_{\mathbf{k}_{\ell+1}} \hat{\eta}_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_{\ell+1}} \sum_{k_\ell=1}^{K_\ell} \tau_{\mathbf{k}_{\ell+1},k_\ell} = \sum_{\mathbf{k}_\ell} \hat{\eta}_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_\ell}$. With all things considered and defining the variable $\tau_{k_{L+1}} \equiv \tau_{L+1} = \pm 1$ according to the prefactor of the argument, we can write

$$
\Theta(\pm\sigma(\mathbf{S})) = \sum_{\{\tau_{L+1}\}} \delta_{\tau_{L+1},\pm 1} \int \mathcal{D}(\eta_{L+1}, \hat{\eta}_{L+1}) \prod_{\ell=L}^{1} \int \mathcal{D}(\boldsymbol{\eta}_\ell, \hat{\boldsymbol{\eta}}_\ell)
$$
$$
\times \sum_{\{\tau_\ell\}} \exp\left(i \sum_{\mathbf{k}_\ell} \hat{\eta}_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_\ell} \right) \exp\left(i \sum_{\mathbf{k}_1} \hat{\eta}_{\mathbf{k}_1} \tau_{\mathbf{k}_1} \frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T} \mathbf{S}}{\sqrt{N}} \right). \tag{A.2}
$$

As is presented in [2], we can define the sets of indexes $\mathbb{I}_P$ such that for each set $\mathbb{I}_P$ there is a vector $\mathbf{S}_{\mathbb{I}_P}$ such that $[\mathbf{S}_{\mathbb{I}_P}]_i = -S_i \ \forall \ i \in \mathbb{I}_P$ and S_i otherwise. We can write the discrepancy as

$$
\mathfrak{d}_P^{(N)}(\mathbf{Q}) = \frac{1}{2^N} \sum_{\{\mathbf{S}\}} \binom{N}{P}^{-1} \sum_{\mathbb{I}_P} [\Theta(-\sigma(\mathbf{S}))\Theta(\sigma(\mathbf{S}_{\mathbb{I}_P})) + \Theta(\sigma(\mathbf{S}))\Theta(-\sigma(\mathbf{S}_{\mathbb{I}_P}))]
$$
$$
= \frac{1}{2^N} \sum_{\{\mathbf{S}\}} \binom{N}{P}^{-1} \sum_{\mathbb{I}_P} 2\Theta(-\sigma(\mathbf{S}))\Theta(\sigma(\mathbf{S}_{\mathbb{I}_P}))
$$
$$
\times 2 \sum_{\{\tau_{L+1},\overline{\tau}_{L+1}\}} \delta_{\tau_{L+1},-1} \delta_{\overline{\tau}_{L+1},1} \int \mathcal{D}(\eta_{L+1},\hat{\eta}_{L+1}) \int \mathcal{D}(\eta'_{L+1},\hat{\eta}'_{L+1})
$$
$$
\times \prod_{\ell=L}^{1} \int \mathcal{D}(\boldsymbol{\eta}_\ell,\hat{\boldsymbol{\eta}}_\ell)\mathcal{D}(\boldsymbol{\eta}'_\ell,\hat{\boldsymbol{\eta}}'_\ell) \sum_{\{\tau_\ell,\overline{\tau}_\ell\}} \exp\left[i\sum_{\mathbf{k}_\ell} \left(\hat{\eta}_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_{\ell+1}} \tau_{\mathbf{k}_\ell} + \hat{\eta}'_{\mathbf{k}_{\ell+1}} \overline{\tau}_{\mathbf{k}_{\ell+1}} \overline{\tau}_{\mathbf{k}_\ell} \right) \right]
$$
$$
\times \frac{1}{2^N} \sum_{\mathbf{S}\in\mathbf{Z}^N} \binom{N}{P}^{-1} \sum_{\mathbb{I}_P} \exp\left[i\sum_{\mathbf{k}_1} \frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T}}{\sqrt{N}} \left(\hat{\eta}_{\mathbf{k}_1} \tau_{\mathbf{k}_1} \mathbf{S} + \hat{\eta}'_{\mathbf{k}_1} \overline{\tau}_{\mathbf{k}_1} \mathbf{S}_{\mathbb{I}_P} \right) \right].
$$

Observing that $\mathbf{w}^\mathsf{T}\mathbf{S}_{\mathbb{I}_P} = \mathbf{w}^\mathsf{T}\mathbf{S} - 2\sum_{j\in\mathbb{I}_P} w_j S_j$, we can write

$$
\kappa_{P,N}(\boldsymbol{v}_1,\mathbf{S}) \equiv \exp\left\{ i\sum_{\mathbf{k}_1} \left[(\hat{\eta}_{\mathbf{k}_1}\tau_{\mathbf{k}_1} + \hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1}) \sum_{j=1}^{N} \frac{w_{\mathbf{k}_1,j}S_j}{\sqrt{N}} - 2\hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1} \sum_{j\in\mathbb{I}_p} \frac{w_{\mathbf{k}_1,j}S_j}{\sqrt{N}} \right] \right\},
$$

where $\boldsymbol{v}_1$ represents the set of variables in the first hidden and input layers.

The dependence on $\mathbf{S}$ can be eliminated by averaging over all possible inputs,

$$
\frac{1}{2^N} \sum_{\{\mathbf{S}\}} \kappa_{P,N}(\boldsymbol{v}_1,\mathbf{S}) = \prod_{j\in\mathbb{I}_P} \cos\left(\sum_{\mathbf{k}_1} (\hat{\eta}_{\mathbf{k}_1}\tau_{\mathbf{k}_1} - \hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1}) \frac{w_{\mathbf{k}_1,j}}{\sqrt{N}} \right)
$$
$$
\times \prod_{j\notin\mathbb{I}_P} \cos\left(\sum_{\mathbf{k}_1} (\hat{\eta}_{\mathbf{k}_1}\tau_{\mathbf{k}_1} + \hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1}) \frac{w_{\mathbf{k}_1,j}}{\sqrt{N}} \right)
$$
$$
\simeq \exp\left[-\frac{1}{2} \sum_{\mathbf{k}_1,\mathbf{m}_1} \left((\hat{\eta}_{\mathbf{k}_1}\tau_{\mathbf{k}_1} + \hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1}) \frac{\mathbf{w}_{\mathbf{k}_1}^\mathsf{T}\mathbf{w}_{\mathbf{m}_1}}{N} (\hat{\eta}_{\mathbf{m}_1}\tau_{\mathbf{m}_1} + \hat{\eta}'_{\mathbf{m}_1}\overline{\tau}_{\mathbf{m}_1}) \right.\right.
$$
$$
\left.\left. -4\hat{\eta}'_{\mathbf{k}_1}\overline{\tau}_{\mathbf{k}_1} \sum_{j\in\mathbb{I}_P} \frac{w_{\mathbf{k}_1,j}w_{\mathbf{m}_1,j}}{N} \hat{\eta}'_{\mathbf{m}_1}\overline{\tau}_{\mathbf{m}_1} \right) \right],
$$

for large enough N. By using definition (2) and the result of appendix B we have that in the thermodynamic limit $P \leqslant N \to \infty$ and $P/N = p$

$$
\kappa_p(\boldsymbol{v}_1) \equiv \lim_{P \leqslant N \to \infty} \binom{N}{P}^{-1} \sum_{\mathbb{I}_P} \frac{1}{2^N} \sum_{\{\mathbf{S}\}} \kappa_{P,N}(\boldsymbol{v}_1, \mathbf{S})
$$

$$
\simeq \exp \left\{ -\frac{1}{2}([\hat{\boldsymbol{\eta}}_1 \boldsymbol{\tau}_1] - (1-2p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1])^{\mathsf{T}} \mathbf{Q}([\hat{\boldsymbol{\eta}}_1 \boldsymbol{\tau}_1] - (1-2p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]) \right.
$$

$$
\left. - 2p(1-p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]^{\mathsf{T}} \mathbf{Q}[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1] \right\}, \tag{A.3}
$$

where $[\hat{\boldsymbol{\eta}}_1 \boldsymbol{\tau}_1]_{\mathbf{k}_1} = \hat{\eta}_{\mathbf{k}_1} \tau_{\mathbf{k}_1}$, $[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]_{\mathbf{k}_1} = \hat{\eta}_{\mathbf{k}_1}' \overline{\tau}_{\mathbf{k}_1}$.

By the use of the Hubbard–Stratonovitch identity, we can simplify this expression. Observe that

$$
\exp\left[-\frac{\phi_\ell^2}{2} \sum_{\mathbf{k}_\ell} \left(\sum_{k_{\ell-1}} \cdots \sum_{k_1} v_{\mathbf{k}_1} \right)^2 \right] = \int \mathcal{D}\boldsymbol{x}_\ell \exp\left(\mathrm{i}\phi_\ell \sum_{\mathbf{k}_1} x_{\mathbf{k}_\ell} v_{\mathbf{k}_1} \right),
$$

where $\phi_\ell \equiv \sqrt{\tilde{\xi}_{\ell-1} - \tilde{\xi}_\ell}$, $\phi_1 \equiv \sqrt{1 - \tilde{\xi}_1}$, $\phi_{L+1} \equiv \sqrt{\tilde{\xi}_L}$, $\mathcal{D}x \equiv \mathrm{d}x \exp(-x^2/2)/\sqrt{2\pi}$ and $\mathcal{D}\boldsymbol{x}_\ell \equiv \prod_{\mathbf{k}_\ell} \mathcal{D}x_{\mathbf{k}_\ell}$. Therefore

$$
\exp\left(-\frac{1}{2}\boldsymbol{v}_1^{\mathsf{T}} \mathbf{Q} \boldsymbol{v}_1 \right) = \exp\left(-\frac{\phi_1^2}{2}\boldsymbol{v}_1^{\mathsf{T}} \boldsymbol{v}_1 \right) \int \left(\prod_{\ell=2}^{L+1} \mathcal{D}\boldsymbol{x}_\ell \right) \exp\left(\mathrm{i} \sum_{\mathbf{k}_1} \overline{X}_{[\mathbf{k}_2]} v_{\mathbf{k}_1} \right),
$$

where $\overline{X}_{[\mathbf{k}_2]} \equiv \sum_{\ell=2}^{L+1} \phi_\ell x_{\mathbf{k}_\ell}$ and $x_{\mathbf{k}_{L+1}} \equiv x_{L+1}$. Given that we have two quadratic terms in the exponential of (A.3) we will need two sets of integrals. Thus

$$
\kappa_p(\boldsymbol{v}_1) = \exp\left\{ -\frac{\phi_1^2}{2}[([\hat{\boldsymbol{\eta}}_1 \boldsymbol{\tau}_1] + (1-2p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1])^{\mathsf{T}}([\hat{\boldsymbol{\eta}}_1 \boldsymbol{\tau}_1] + (1-2p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]) \right.
$$

$$
\left. + 4p(1-p)[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]^{\mathsf{T}}[\hat{\boldsymbol{\eta}}_1' \overline{\boldsymbol{\tau}}_1]] \right\}
$$

$$
\times \int \left[\prod_{\ell=2}^{L+1} \mathcal{D}\boldsymbol{x}_\ell \mathcal{D}\boldsymbol{x}_\ell' \right] \exp\left[\mathrm{i} \sum_{\mathbf{k}_1} \overline{X}_{[\mathbf{k}_2]} \left(\hat{\eta}_{\mathbf{k}_1} \tau_{\mathbf{k}_1} + (1-2p)\hat{\eta}_{\mathbf{k}_1}' \overline{\tau}_{\mathbf{k}_1} \right) \right]
$$

$$
\times \exp\left[\mathrm{i}\sqrt{4p(1-p)} \sum_{\mathbf{k}_1} \overline{X}_{[\mathbf{k}_2]}' \hat{\eta}_{\mathbf{k}_1}' \overline{\tau}_{\mathbf{k}_1} \right].
$$

We can integrate over $\mathcal{D}(\boldsymbol{\eta}_1, \hat{\boldsymbol{\eta}}_1)$ and $\mathcal{D}(\boldsymbol{\eta}_1', \hat{\boldsymbol{\eta}}_1')$

$$
\int \mathcal{D}(\boldsymbol{\eta}_1, \hat{\boldsymbol{\eta}}_1) \mathcal{D}(\boldsymbol{\eta}_1', \hat{\boldsymbol{\eta}}_1') \kappa_p(\boldsymbol{v}_1) = \prod_{\ell=2}^{L+1} \int \mathcal{D}\boldsymbol{x}_\ell \mathcal{D}\boldsymbol{x}_\ell' \prod_{\mathbf{k}_1} \int_0^\infty \frac{\mathrm{d}\eta}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\eta - \tau_{\mathbf{k}_1} X_{[\mathbf{k}_2]} \right)^2 \right]
$$

$$
\times \mathcal{H}\left(\overline{\tau}_{\mathbf{k}_1} \left(\beta(p)\tau_{\mathbf{k}_1}\eta + X_{[\mathbf{k}_2]}' \right) \right),
$$

where $X_{[\mathbf{k}_2]} \equiv \overline{X}_{[\mathbf{k}_2]}/\phi_1$, $\beta(p) \equiv (1-2p)/\sqrt{4p(1-p)}$ and $\mathcal{H}(x) \equiv \int_x^\infty \mathcal{D}y$ is Gardner's error function.

Finally

$$
\kappa_p(\boldsymbol{v}_2) \equiv \sum_{\{\tau_1, \overline{\tau}_1\}} \exp[\mathrm{i}([\hat{\boldsymbol{\eta}}_2 \boldsymbol{\tau}_2]^{\mathsf{T}} \boldsymbol{\tau}_1 + [\hat{\boldsymbol{\eta}}_2' \overline{\boldsymbol{\tau}}_2]^{\mathsf{T}} \overline{\boldsymbol{\tau}}_1)] \int \mathcal{D}(\boldsymbol{\eta}_1, \hat{\boldsymbol{\eta}}_1) \mathcal{D}(\boldsymbol{\eta}_1', \hat{\boldsymbol{\eta}}_1') \kappa_p(\boldsymbol{v}_1)
$$

$$= \prod_{\ell=2}^{L+1} \int \mathcal{D}\boldsymbol{x}_\ell \mathcal{D}\boldsymbol{x}'_\ell \prod_{\mathbf{k}_2} \left\{ \sum_{\tau, \overline{\tau} = \pm 1} \exp \left(i \hat{\eta}_{\mathbf{k}_2} \tau_{\mathbf{k}_2} \tau + i \hat{\eta}'_{\mathbf{k}_2} \overline{\tau}_{\mathbf{k}_2} \overline{\tau} \right) \right.$$

$$\left. \times \int_0^\infty \frac{d\eta}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\eta - \tau X_{[\mathbf{k}_2]} \right)^2 \right] \mathcal{H} \left(\overline{\tau} \left(\beta(p) \tau \eta + X'_{[\mathbf{k}_2]} \right) \right) \right\}^{K_1}$$

and by using the identities

$$1 = \int_0^\infty \frac{d\eta}{\sqrt{2\pi}} \left\{ \exp \left[-\frac{1}{2} \left(\eta - X_{[\mathbf{k}_2]} \right)^2 \right] + \exp \left[-\frac{1}{2} \left(\eta + X_{[\mathbf{k}_2]} \right)^2 \right] \right\}$$

$$\mathrm{erf} \left(\frac{X_{[\mathbf{k}_2]}}{\sqrt{2}} \right) = \int_0^\infty \frac{d\eta}{\sqrt{2\pi}} \left\{ \exp \left[-\frac{1}{2} \left(\eta - X_{[\mathbf{k}_2]} \right)^2 \right] - \exp \left[-\frac{1}{2} \left(\eta + X_{[\mathbf{k}_2]} \right)^2 \right] \right\}$$

and the definitions

$$\mathcal{G}_p^+ \left(X_{[\mathbf{k}_2]}, X'_{[\mathbf{k}_2]} \right) = \int \mathcal{D}\eta \, \mathrm{sgn}(\eta + X_{[\mathbf{k}_2]}) \, \mathrm{erf} \left(\frac{\beta(p) \left(\eta + X_{[\mathbf{k}_2]} \right) + X'_{[\mathbf{k}_2]}}{\sqrt{2}} \right) \tag{A.4}$$

$$\mathcal{G}_p^- \left(X_{[\mathbf{k}_2]}, X'_{[\mathbf{k}_2]} \right) = \int \mathcal{D}\eta \, \mathrm{erf} \left(\frac{\beta(p) \left(\eta + X_{[\mathbf{k}_2]} \right) + X'_{[\mathbf{k}_2]}}{\sqrt{2}} \right), \tag{A.5}$$

we can define the function $\Lambda_p(\boldsymbol{v}; X, X')$, where $\boldsymbol{v} = \{\hat{\eta}, \tau, \hat{\eta}', \overline{\tau}\}$

$$\Lambda_p(\boldsymbol{v}; X, X') \equiv R_p(\boldsymbol{v}; X, X') + i \, I_p(\boldsymbol{v}; X, X')$$

$$R_p(\boldsymbol{v}; X, X') \equiv \cos(\hat{\eta}) \cos(\hat{\eta}') - \tau \sin(\hat{\eta}) \overline{\tau} \sin(\hat{\eta}') \mathcal{G}_p^+(X, X') \tag{A.6}$$

$$I_p(\boldsymbol{v}; X, X') \equiv \tau \sin(\hat{\eta}) \cos(\hat{\eta}') \mathrm{erf} \left(\frac{X}{\sqrt{2}} \right) + \cos(\hat{\eta}) \overline{\tau} \sin(\hat{\eta}') \mathcal{G}_p^-(X, X').$$

With this definition we have that

$$\kappa_p(\boldsymbol{v}_2) = \prod_{\ell=2}^{L+1} \int \mathcal{D}\boldsymbol{x}_\ell \mathcal{D}\boldsymbol{x}'_\ell \prod_{\mathbf{k}_2} \left[\Lambda_p \left(\boldsymbol{v}_{\mathbf{k}_2}; X_{[\mathbf{k}_2]}, X'_{[\mathbf{k}_2]} \right) \right]^{K_1}, \tag{A.7}$$

where $\boldsymbol{v}_{\mathbf{k}_2}$ represents the set of variables $\hat{\eta}_{\mathbf{k}_2}$, $\hat{\eta}'_{\mathbf{k}_2}$, $\tau_{\mathbf{k}_2}$, $\overline{\tau}_{\mathbf{k}_2}$.

　　By using the kernel of expression (A.7) we can reconstruct the average discrepancy component

$$\mathfrak{d}(p, \mathbf{Q}) = 2 \int \mathcal{D}(\eta_{L+1}, \hat{\eta}_{L+1}) \mathcal{D}(\eta'_{L+1}, \hat{\eta}'_{L+1}) \int \mathcal{D}x_{L+1} \mathcal{D}x'_{L+1}$$

$$\times \sum_{\{\tau_L, \overline{\tau}_L\}} \exp \left[i \sum_{\mathbf{k}_L} \left(\hat{\eta}_{L+1} \tau_{\mathbf{k}_\ell} - \hat{\eta}'_{L+1} \overline{\tau}_{\mathbf{k}_\ell} \right) \right] \int \mathcal{D}(\boldsymbol{\eta}_L, \hat{\boldsymbol{\eta}}_L) \mathcal{D}(\boldsymbol{\eta}'_L, \hat{\boldsymbol{\eta}}'_L) \int \mathcal{D}\boldsymbol{x}_L \mathcal{D}\boldsymbol{x}'_L$$

$$\times \prod_{\ell=L-1}^2 \sum_{\{\tau_\ell, \overline{\tau}_\ell\}} \exp \left[i ([\hat{\boldsymbol{\eta}}_{\ell+1} \boldsymbol{\tau}_{\ell+1}]^\mathsf{T} \boldsymbol{\tau}_\ell + [\hat{\boldsymbol{\eta}}'_{\ell+1} \overline{\boldsymbol{\tau}}_{\ell+1}]^\mathsf{T} \overline{\boldsymbol{\tau}}_\ell) \right] \int \mathcal{D}(\boldsymbol{\eta}_\ell, \hat{\boldsymbol{\eta}}_\ell) \mathcal{D}(\boldsymbol{\eta}'_\ell, \hat{\boldsymbol{\eta}}'_\ell)$$

$$\times \int \mathcal{D}\boldsymbol{x}_\ell \mathcal{D}\boldsymbol{x}'_\ell \prod_{\mathbf{k}_2} \left[\Lambda_p \left(\boldsymbol{v}_{\mathbf{k}_2}; X_{[\mathbf{k}_2]}, X'_{[\mathbf{k}_2]} \right) \right]^{K_1}.$$

Observe that the function $\Lambda_p \left(\boldsymbol{v}_{\mathbf{k}_2}; X_{[\mathbf{k}_2]}, X'_{[\mathbf{k}_2]} \right)$ depends only on the variables described by the path $\mathbf{k}_2$ and does not depend on any interaction linking variables with different indices.

13

In order to solve the problem analytically in the large number of hidden units regime, let us define the operators $\mathcal{U}_\ell$ for all $\ell = 2, 3, \ldots, L+1$

$$\mathcal{U}_\ell \equiv \sum_{\{\tau_\ell, \overline{\tau}_\ell\}} \exp(i\hat{\eta}_{\ell+1}\tau_{\ell+1}\tau_\ell + i\hat{\eta}'_{\ell+1}\overline{\tau}_{\ell+1}\overline{\tau}_\ell) \int \mathcal{D}x_\ell \mathcal{D}x'_\ell \lim_{K_{\ell-1}\uparrow\infty} \int \mathcal{D}(\eta_\ell, \hat{\eta}_\ell) \mathcal{D}(\eta'_\ell, \hat{\eta}'_\ell)$$

$$\mathcal{U}_{L+1} \equiv \sum_{\{\tau_{L+1}, \overline{\tau}_{L+1}\}} \delta_{\tau_{L+1},1}\delta_{\overline{\tau}_{L+1},-1} \int \mathcal{D}x_{L+1}\mathcal{D}x'_{L+1} \lim_{K_L\uparrow\infty} \int \mathcal{D}(\eta_{L+1}, \hat{\eta}_{L+1}) \mathcal{D}(\eta'_{L+1}, \hat{\eta}'_{L+1}).$$

With these operators we can write

$$\mathfrak{d}_p(\mathbf{Q}) = 2\mathcal{U}_{L+1}[\Lambda_p(\boldsymbol{v}_{L+1}; X_{L+1}, X'_{L+1})]^{K_L},$$

where

$$\Lambda_p(\boldsymbol{v}_{\ell+1}; X_{\ell+1}, X'_{\ell+1}) \equiv \mathcal{U}_\ell[\Lambda_p(\boldsymbol{v}_\ell; X_\ell, X'_\ell)]^{K_{\ell-1}}, \tag{A.8}$$

where $\boldsymbol{v}_\ell$ represents the set of variables $\hat{\eta}_\ell$, $\hat{\eta}'_\ell$, τ_ℓ, $\overline{\tau}_\ell$ and $X_\ell \equiv \sum_{j=\ell}^{L+1} \frac{\phi_j}{\phi_1} x_j$ (and equivalently for X'_ℓ).

The set of equations (A.8) with the iteration of order 1 given by (A.6) can be analytically solved if the number of hidden units at each layer is large (i.e. in the limit of $K_\ell \to \infty$ for all $\ell = 1, 2, \ldots, L$). First of all let us define the variables

$$\tilde{X}_\ell \equiv \left(\zeta_{\ell-1} - \frac{\zeta_\ell}{K_\ell}\right)^{\frac{1}{2}} x_\ell + K_\ell^{-\frac{1}{2}} \tilde{X}_{\ell+1}$$

with $\tilde{X}_{L+1} \equiv \zeta_L^{\frac{1}{2}} x_{L+1}$ such that

$$X_\ell = \sum_{j=\ell}^{L+1} \frac{\phi_j}{\phi_1} x_j$$

$$= \left(1 - \frac{\zeta_1}{K_1}\right)^{-\frac{1}{2}} K_{\ell-1}^{-\frac{1}{2}} \tilde{X}_\ell$$

$$\simeq K_{\ell-1}^{-1/2} \left[\left(\zeta_{\ell-1} - \frac{\zeta_\ell}{K_\ell}\right)^{\frac{1}{2}} x_\ell + K_\ell^{-\frac{1}{2}} \tilde{X}_{\ell+1}\right] + O\left((K_1 K_{\ell-1})^{-\frac{1}{2}}\right),$$

and

$$\tilde{Y}_\ell \equiv \frac{\beta \tilde{X}_\ell + \tilde{X}'_\ell}{\sqrt{1 + \beta^2}}.$$

Using the fact that the number of hidden units is large we can expand (A.4), (A.5) and the error function; therefore,

$$\mathcal{G}_p^+(X_2, X'_2) \simeq \mathcal{G}_1(p) + O\left(K_1^{-1}\right) \tag{A.9}$$

$$\mathcal{G}_p^-(X_2, X'_2) \simeq \mathrm{erf}\left(\alpha_1 \tilde{Y}_2 K_1^{-\frac{1}{2}}\right) + O\left(K_1^{-\frac{3}{2}}\right) \tag{A.10}$$

$$\mathrm{erf}\left(\frac{X_2}{\sqrt{2}}\right) = \mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right), \tag{A.11}$$

where $\mathcal{G}_1(p) \equiv 1 - \frac{2}{\pi}\arccos(1 - 2p)$ and $\alpha_1 = 2^{-\frac{1}{2}}$. Before proceeding with the calculation, let us express the real and imaginary parts of the function Λ_p using these approximations, and disregarding terms of $O\left(K_1^{-1}\right)$

$$R_p(\boldsymbol{v}_2; X_2, X'_2) \simeq \cos(\hat{\eta}_2) \cos(\hat{\eta}'_2) - \tau_2\overline{\tau}_2 \sin(\hat{\eta}_2) \sin(\hat{\eta}'_2)\mathcal{G}_1(p)$$

$$I_p(\boldsymbol{v}_2; X_2, X'_2) \simeq \tau_2 \sin(\hat{\eta}_2) \cos(\hat{\eta}'_2)\mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right) + \overline{\tau}_2 \cos(\hat{\eta}_2) \sin(\hat{\eta}'_2)\mathrm{erf}\left(\alpha_1 \tilde{Y}_2 K_1^{-\frac{1}{2}}\right);$$

14

therefore,

$$\Lambda_p^{K_1}(\boldsymbol{v}_2; X_2, X_2') = [R_p(\boldsymbol{v}_2; X_2, X_2') + \mathrm{i}I_p(\boldsymbol{v}_2; X_2, X_2')]^{K_1}$$

$$= \sum_{k=0}^{K_1}\sum_{r=0}^{K_1-k}\sum_{j=0}^{k}\mathrm{i}^k\binom{K_1}{k}\binom{K_1-k}{r}\binom{k}{j}\tau_2^{r+k-j}\overline{\tau}_2^{r+j}(-\mathcal{G}_p^+(X_2, X_2'))^r(\mathcal{G}_p^-(X_2, X_2'))^j$$

$$\times\,\mathrm{erf}\big(\alpha_1\tilde{X}_2 K_1^{-\frac{1}{2}}\big)^{2m+1}\cos(\hat{\eta}_2)^{K_1-k-r+j}\sin(\hat{\eta}_2)^{k+r-j}\cos(\hat{\eta}_2')^{K_1-r-j}\sin(\hat{\eta}_2')^{r+j};$$

thus,

$$\int\mathcal{D}(\eta_2, \hat{\eta}_2)\,\mathcal{D}(\eta_2', \hat{\eta}_2')\Lambda_p^{K_1}(\boldsymbol{v}_2; X_2, X_2') = \sum_{k=0}^{K_1}\sum_{r=0}^{K_1-k}\sum_{j=0}^{k}\mathrm{i}^k\binom{K_1}{k}\binom{K_1-k}{r}\binom{k}{j}\tau_2^{r+k-j}\overline{\tau}_2^{r+j}$$

$$\times\,(-\mathcal{G}_p^+(X_2, X_2'))^r\,(\mathcal{G}_p^-(X_2, X_2'))^j\,\mathrm{erf}\big(\alpha_1\tilde{X}_2 K_1^{-\frac{1}{2}}\big)^{k-j}$$

$$\times\int\mathcal{D}(\eta_2, \hat{\eta}_2)\cos(\hat{\eta}_2)^{K_1-k-r+j}\sin(\hat{\eta}_2)^{k+r-j}$$

$$\times\int\mathcal{D}(\eta_2', \hat{\eta}_2')\cos(\hat{\eta}_2')^{K_1-r-j}\sin(\hat{\eta}_2')^{r+j}, \tag{A.12}$$

and in agreement with (C.4), (C.5) and (C.6) we have that there are four situations where the integrals in (A.12) are not zero.

(i) $k + r - j = r + j = 0$, which implies that $k = r = j = 0$. The correspondent term to the series evaluates to $\frac{1}{4}$.

(ii) $k + r - j = 0$ and $r + j \neq 0$ is odd, which implies that $r = 0$, given that $j \leqslant k$. Thus $k = j$ is odd.

(iii) $k + r - j \neq 0$ is odd, and $r + j = 0$ which implies that $r = j = 0$, given that $0 \leqslant r, j$. Thus k is odd.

(iv) $k + r - j \neq 0$ is odd, and $r + j \neq 0$ is odd, which implies that k is even and r and j have different parities.

The first term, with all the indexes equal to 0 is straightforwardly $A_0 = \frac{1}{4}$. The second term has the indexes $j = k = 2m + 1$ and $r = 0$. Thus, using that $K_1 = 2M + 1$

$$A_1^M(\overline{\tau}_2, \tilde{Y}_2) = \sum_{m=0}^{M}\binom{2M+1}{2m+1}(\mathrm{i}\overline{\tau}_2)^{2m+1}(\mathcal{G}_p^-(X_2, X_2'))^{2m+1}\frac{1}{2}a_m^M$$

$$= \frac{\overline{\tau}_2}{2^{2M+2}}\sum_{m=0}^{M}(-1)^m\binom{2M+1}{2m+1}\binom{M}{m}^{-1}\binom{2m}{m}\binom{2(M-m)}{M-m}(\mathcal{G}_p^-(X_2, X_2'))^{2m+1},$$

where a_m^M is given by (C.6)

$$a_m^M = -\frac{\mathrm{i}}{2^{2M+1}}\binom{M}{m}^{-1}\binom{2m}{m}\binom{2(M-m)}{M-m}.$$

By using (A.10) we can take the limit of $K_1 \to \infty$

$$A_1(\overline{\tau}_2, \tilde{Y}_2) = \frac{\overline{\tau}_2}{4}\lim_{M\uparrow\infty}\sum_{m=0}^{M}(-1)^m\frac{\binom{2M+1}{2m+1}\binom{2m}{m}\binom{2(M-m)}{M-m}}{2^{2M}\binom{M}{m}(2M+1)^{m+\frac{1}{2}}}\left(\frac{2}{\sqrt{\pi}}\alpha_1\tilde{Y}_2\right)^{2m+1}$$

$$= \frac{\overline{\tau}_2}{4}\lim_{M\uparrow\infty}\frac{2}{\sqrt{\pi}}\sum_{m=0}^{M}\left[\frac{(-1)^m}{(2m+1)m!}\left(\sqrt{\frac{2}{\pi}}\alpha_1\tilde{Y}_2\right)^{2m+1} + O(M^{-1})\right]$$

$$= \frac{\overline{\tau}_2}{4}\mathrm{erf}\left(\sqrt{\frac{2}{\pi}}\alpha_1\tilde{Y}_2\right).$$

Therefore, the integral over the Hubbard–Stratonovitch variables x_2 and x_2' is using the recursive nature of $\tilde{X}_2$ and $\tilde{X}_2'$. Therefore

$$\int \mathcal{D}x_2 \mathcal{D}x_2' A_1(\overline{\tau}_2, \tilde{Y}_2) = \frac{\overline{\tau}_2}{4} \mathscr{A}_1(\tilde{Y}_3),$$

such that

$$\mathscr{A}_1(\tilde{Y}_3) \equiv \int \mathcal{D}x_2 \mathcal{D}x_2' \mathrm{erf}\left(\sqrt{\frac{2}{\pi}}\alpha_1 \frac{\beta \tilde{X}_2 + \tilde{X}_2'}{\sqrt{1+\beta^2}}\right)$$

$$\simeq \int \mathcal{D}x_2 \mathcal{D}x_2' \left\{ \mathrm{erf}\left(\sqrt{\frac{2\zeta_1}{\pi}}\alpha_1 \frac{\beta x_2 + x_2'}{\sqrt{1+\beta^2}}\right) + \frac{2^{\frac{3}{2}}\alpha_1}{\pi}\tilde{Y}_3 K_2^{-\frac{1}{2}} \exp\left[-\frac{2\zeta_1\alpha_1^2}{\pi}\frac{(\beta x_2 + x_2')^2}{1+\beta^2}\right]\right\}$$

$$\quad + O\left(K_2^{-\frac{3}{2}}\right)$$

$$\simeq \frac{2}{\sqrt{\pi}}\sqrt{\frac{2\alpha_1^2}{\pi + 4\zeta_1\alpha_1^2}}\tilde{Y}_3 K_2^{-\frac{1}{2}} + O\left(K_2^{-\frac{3}{2}}\right)$$

$$\simeq \mathrm{erf}\left(\alpha_2 \tilde{Y}_3 K_2^{-\frac{1}{2}}\right) + O\left(K_2^{-\frac{3}{2}}\right),$$

where

$$\alpha_2 \equiv \sqrt{\frac{2\alpha_1^2}{\pi + 4\zeta_1\alpha_1^2}}.$$

The third term is similar to the second. It is formed by the following contributions:

$$A_2^M(\tau_2, \tilde{X}_2) = \sum_{m=0}^{M}\binom{2M+1}{2m+1}(\mathrm{i}\tau_2)^{2m+1}\mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right)^{2m+1}\frac{1}{2}a_m^M,$$

and by expanding the error function, taking the limit of large M and integrating over x_2 and x_2', we arrive at

$$\int \mathcal{D}x_2 \mathcal{D}x_2' A_2(\tau_2, \tilde{X}_2) = \frac{\tau_2}{4}\mathscr{A}_2(\tilde{X}_3),$$

$$\mathscr{A}_2(\tilde{X}_3) \simeq \mathrm{erf}\left(\alpha_2 \tilde{X}_3 K_2^{-\frac{1}{2}}\right) + O\left(K_2^{-\frac{3}{2}}\right).$$

The last contribution involves terms with even values of k and non-zero, uneven values of r and j. Therefore

$$A_3^M(\tau_2, \overline{\tau}_2, \tilde{X}_2, \tilde{Y}_2) = \tau_2\overline{\tau}_2 \sum_{m=0}^{M}\sum_{n=0}^{M-m}\sum_{s=0}^{m-1}(-1)^m\binom{2M+1}{2m}\binom{2(M-m)+1}{2n}\binom{2m}{2s+1}$$

$$\quad \times a_{n+s}^M a_{m+n-s-1}^M (\mathcal{G}_p^+(X_2, X_2'))^{2n}\mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right)^{2m-2s-1}(\mathcal{G}_p^-(X_2, X_2'))^{2s+1}$$

$$\quad - \tau_2\overline{\tau}_2 \sum_{m=0}^{M}\sum_{n=0}^{M-m}\sum_{s=0}^{m}(-1)^m\binom{2M+1}{2m}\binom{2(M-m)+1}{2n+1}\binom{2m}{2s}$$

$$\quad \times a_{n+s}^M a_{m+n-s}^M (\mathcal{G}_p^+(X_2, X_2'))^{2n+1}\mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right)^{2m-2s}(\mathcal{G}_p^-(X_2, X_2'))^{2s}.$$

In order to continue the development we introduce the following identity:

$$\int_{-\infty}^{\infty}\mathcal{D}y\, y^{2\ell} = \begin{cases} 1 & \ell = 0 \\ (2\ell-1)!! & \text{otherwise.} \end{cases} \tag{A.13}$$

Therefore, for large values of M and using expansions (A.10) and (A.11) we have that the leading terms in M are

$$\binom{2M+1}{2m}\binom{2(M-m)+1}{2n}a_{n+s}^M a_{m+n-s-1}^M \mathrm{erf}\left(\alpha_1 \tilde{X}_2 K_1^{-\frac{1}{2}}\right)^{2m-2s-1}(\mathcal{G}_p^-(X_2, X_2'))^{2s+1}$$

$$\simeq -\frac{1}{2\pi}\frac{1}{(2m)!(2n)!}\left(\frac{2}{\sqrt{\pi}}\alpha_1 \tilde{X}_2\right)^{2m-2s-1}\left(\frac{2}{\sqrt{\pi}}\alpha_1 \tilde{Y}_2\right)^{2s+1}\int \mathcal{D}y\mathcal{D}z\, y^{2(m+n-s-1)}z^{2(n+s)}$$

and

$$\binom{2M+1}{2m}\binom{2(M-m)+1}{2n+1}a_{n+s}^M a_{m+n-s}^M \mathrm{erf}\left(\alpha_2 \tilde{X}_2 K_1^{-\frac{1}{2}}\right)^{2m-2s}(\mathcal{G}_p^-(X_2, X_2'))^{2s}$$

$$\simeq -\frac{1}{2\pi}\frac{1}{(2m)!(2n+1)!}\left(\frac{2}{\sqrt{\pi}}\alpha_1 \tilde{X}_2\right)^{2m-2s}\left(\frac{2}{\sqrt{\pi}}\alpha_1 \tilde{Y}_2\right)^{2s}\int \mathcal{D}y\mathcal{D}z\, y^{2(m+n-s)}z^{2(n+s)}$$

so we can reconstruct the expression for A_3^M such that

$$A_3^M(\tau_2, \overline{\tau}_2, \tilde{X}_2, \tilde{Y}_2) \simeq -\frac{\tau_2 \overline{\tau}_2}{2\pi}\int \frac{\mathcal{D}y}{y}\frac{\mathcal{D}z}{z}\sum_{m=0}^{M}\frac{(-1)^m}{(2m)!}\sum_{n=0}^{M-m}\frac{(yz\mathcal{G}_1(p))^{2n}}{(2n)!}$$

$$\times \sum_{s=0}^{m-1}\binom{2m}{2s+1}\left(\frac{2}{\sqrt{\pi}}y\alpha_1 \tilde{X}_2\right)^{2m-2s-1}\left(\frac{2}{\sqrt{\pi}}z\alpha_1 \tilde{Y}_2\right)^{2s+1}$$

$$+\frac{\tau_2 \overline{\tau}_2}{2\pi}\int \frac{\mathcal{D}y}{y}\frac{\mathcal{D}z}{z}\sum_{m=0}^{M}\frac{(-1)^m}{(2m)!}\sum_{n=0}^{M-m}\frac{(yz\mathcal{G}_1(p))^{2n+1}}{(2n+1)!}$$

$$\times \sum_{s=0}^{m}\binom{2m}{2s}\left(\frac{2}{\sqrt{\pi}}y\alpha_1 \tilde{X}_2\right)^{2m-2s}\left(\frac{2}{\sqrt{\pi}}z\alpha_1 \tilde{Y}_2\right)^{2s} + O\left(K_1^{-1}\right).$$

If $M(K_1)$ is large enough, the sums in n can be approximated by the hyperbolic cosine and sine, respectively. Thus, in the limit,

$$A_3(\tau_2, \overline{\tau}_2, \tilde{X}_2, \tilde{Y}_2) = \frac{\tau_2 \overline{\tau}_2}{2\pi}\int \frac{\mathcal{D}y}{y}\frac{\mathcal{D}z}{z}\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m)!}$$

$$\times \left[-\cosh(yz\mathcal{G}_1(p))\sum_{s=0}^{m-1}\binom{2m}{2s+1}\left(\frac{2}{\sqrt{\pi}}y\alpha_1 \tilde{X}_2\right)^{2m-2s-1}\left(\frac{2}{\sqrt{\pi}}z\alpha_1 \tilde{Y}_2\right)^{2s+1}\right.$$

$$\left.+\sinh(yz\mathcal{G}_1(p))\sum_{s=0}^{m}\binom{2m}{2s}\left(\frac{2}{\sqrt{\pi}}y\alpha_1 \tilde{X}_2\right)^{2m-2s}\left(\frac{2}{\sqrt{\pi}}z\alpha_1 \tilde{Y}_2\right)^{2s}\right]$$

which can be rearranged such that

$$A_3(\tau_2, \overline{\tau}_2, \tilde{X}_2, \tilde{Y}_2) = \frac{\tau_2 \overline{\tau}_2}{4\pi}\int \frac{\mathcal{D}y}{y}\frac{\mathcal{D}z}{z}\left[\exp(yz\,\mathcal{G}_1(p))\cos\left(\frac{2\alpha_1}{\sqrt{\pi}}(y\tilde{X}_2 - z\tilde{Y}_2)\right)\right.$$

$$\left.-\exp(-yz\,\mathcal{G}_1(p))\cos\left(\frac{2\alpha_1}{\sqrt{\pi}}(y\tilde{X}_2 + z\tilde{Y}_2)\right)\right]. \tag{A.14}$$

The next step involves the integration of the Hubbard–Stratonovitch variables. Observe that only even functions in such variables will survive the process. A_3 can be expressed as a

linear combination of an odd and even functions in x_2 and x'_2. The even part is, disregarding corrections of $O(K_2^{-1})$,

$$
A_3^{\text{even}}(\tau_2, \bar{\tau}_2, \tilde{X}_2, \tilde{Y}_2) \simeq \frac{\tau_2 \bar{\tau}_2}{4\pi} \int \frac{\mathcal{D}y}{y} \frac{\mathcal{D}z}{z}
$$

$$
\times \left\{ \exp(yz\mathcal{G}_1(p)) \cos\left[\sqrt{\frac{4\zeta_1\alpha_1^2}{\pi}} \left(y - \frac{\beta z}{\sqrt{1+\beta^2}} \right) x_2 \right] \cos\left(\sqrt{\frac{4\zeta_1\alpha_1^2}{\pi}} \frac{z}{\sqrt{1+\beta^2}} x'_2 \right) \right.
$$

$$
\left. - \exp(-yz\mathcal{G}_1(p)) \cos\left[\sqrt{\frac{4\zeta_1\alpha_1^2}{\pi}} \left(y + \frac{\beta z}{\sqrt{1+\beta^2}} \right) x_2 \right] \cos\left(\sqrt{\frac{4\zeta_1\alpha_1^2}{\pi}} \frac{z}{\sqrt{1+\beta^2}} x'_2 \right) \right\},
$$

which implies that

$$
\frac{\tau_2 \bar{\tau}_2}{4} \mathscr{A}_3(\tilde{X}_3, \tilde{Y}_3) \equiv \int \mathcal{D}x_2 \mathcal{D}x'_2 A_3(\tau_2, \bar{\tau}_2, \tilde{X}_2, \tilde{Y}_2)
$$

$$
\mathscr{A}_3(\tilde{X}_3, \tilde{Y}_3) \simeq \frac{1}{\pi} \int \frac{\mathcal{D}y}{y} \frac{\mathcal{D}z}{z} \left\{ \exp\left[yz\mathcal{G}_1(p) - \frac{2\zeta_1\alpha_1^2}{\pi} \left(y - \frac{\beta z}{\sqrt{1+\beta^2}} \right)^2 - \frac{2\zeta_1\alpha_1^2}{\pi} \frac{z^2}{1+\beta^2} \right] \right.
$$

$$
\left. - \exp\left[-yz\mathcal{G}_1(p) - \frac{2\zeta_1\alpha_1^2}{\pi} \left(y + \frac{\beta z}{\sqrt{1+\beta^2}} \right)^2 - \frac{2\zeta_1\alpha_1^2}{\pi} \frac{z^2}{1+\beta^2} \right] \right\} + O(K_2^{-1})
$$

and by defining the variables $u \equiv \sqrt{1 + \frac{4}{\pi}\zeta_1\alpha_1^2}\, y$ and $t \equiv \sqrt{1 + \frac{4}{\pi}\zeta_1\alpha_1^2}\, z$, and using that $1 - 2p = \beta/\sqrt{1+\beta^2}$, we have that

$$
\mathscr{A}_3(\tilde{X}_3, \tilde{Y}_3) \simeq \frac{2}{\pi} \tau_2 \bar{\tau}_2 \int \frac{\mathcal{D}u}{u} \frac{\mathcal{D}t}{t} \sinh\left(\frac{\mathcal{G}_1(p) + \frac{4}{\pi}\zeta_1\alpha_1^2(1-2p)}{1 + \frac{4}{\pi}\zeta_1\alpha_1^2} ut \right) + O(K_2^{-1}).
$$

This integral can be solved by expanding the hyperbolic sine in power series and using (A.13) backwards. Thus

$$
\mathcal{J}(b) \equiv \frac{2}{\pi} \int \frac{\mathcal{D}u}{u} \frac{\mathcal{D}t}{t} \sinh(but)
$$

$$
= \frac{2}{\pi} \sum_{j=0}^{\infty} \frac{b^{2j+1}}{(2j+1)!} \left(\int \mathcal{D}u u^{2j} \right)^2
$$

$$
= \frac{2}{\pi} \left(b + \sum_{j=1}^{\infty} \frac{[(2j-1)!!]^2}{(2j+1)!} b^{2j+1} \right)
$$

$$
= 1 - \frac{2}{\pi} \arccos(b).
$$

In the end we obtain

$$
\mathscr{A}_3(\tilde{X}_3, \tilde{Y}_3) \simeq 1 - \frac{2}{\pi} \arccos\left(\frac{\mathcal{G}_1(p) + \frac{4}{\pi}\zeta_1\alpha_1^2(1-2p)}{1 + \frac{4}{\pi}\zeta_1\alpha_1^2} \right) + O(K_2^{-1})
$$

$$
\simeq \mathcal{G}_2(p) + O(K_2^{-1})
$$

where

$$
\gamma_1 \equiv \frac{4\zeta_1\alpha_1^2}{\pi + 4\zeta_1\alpha_1^2}
$$

$$
\mathcal{G}_2(p) \equiv 1 - \frac{2}{\pi} \arccos\left((1-\gamma_1)\mathcal{G}_1(p) + \gamma_1(1-2p) \right).
$$

The last operation we have to perform is the trace over the activation variables, i.e.

$$\Lambda_p(\boldsymbol{v}_3; X_3, X_3') = \mathcal{U}_2[\Lambda_p(\boldsymbol{v}_2; X_2, X_2')]^{K_1}$$

$$= \frac{1}{4} \sum_{\{\tau_2, \overline{\tau}_2\}} \exp(i\hat{\eta}_3 \tau_3 \tau_2 + i\hat{\eta}_3' \overline{\tau}_3 \overline{\tau}_2)\left(1 + \overline{\tau}_2 \mathrm{erf}\left(\alpha_2 \tilde{Y}_3 K_2^{-\frac{1}{2}}\right) + \tau_2 \mathrm{erf}\left(\alpha_2 \tilde{X}_3 K_2^{-\frac{1}{2}}\right) + \tau_2 \overline{\tau}_2 \mathcal{G}_2(p)\right)$$

$$= R_p(\boldsymbol{v}_3; X_3, X_3') + iI_p(\boldsymbol{v}_3; X_3, X_3'),$$

where the real and imaginary part of the function Λ_p are, disregarding terms of $O\left(K_2^{-1}\right)$ and $O\left(K_2^{-\frac{3}{2}}\right)$ respectively,

$$R_p(\boldsymbol{v}_3; X_3, X_3') \simeq \cos(\hat{\eta}_3)\cos(\hat{\eta}_3') - \tau_3 \overline{\tau}_3 \sin(\hat{\eta}_3)\sin(\hat{\eta}_3')\mathcal{G}_2(p)$$

$$I_p(\boldsymbol{v}_3; X_3, X_3') \simeq \tau_3 \sin(\hat{\eta}_3)\cos(\hat{\eta}_3')\mathrm{erf}\left(\alpha_2 \tilde{X}_3 K_2^{-\frac{1}{2}}\right) + \overline{\tau}_3 \cos(\hat{\eta}_3)\sin(\hat{\eta}_3')\mathrm{erf}\left(\alpha_2 \tilde{Y}_3 K_2^{-\frac{1}{2}}\right).$$

The iteration that produces the functions of the variables at the $(\ell + 1)$ layer will have the following parameters:

$$\alpha_\ell^2 \equiv \frac{2\alpha_{\ell-1}^2}{\pi + 4\zeta_{\ell-1}\alpha_{\ell-1}^2} = \frac{1}{2}\left(\frac{2}{\pi}\right)^{\ell-1}\left[1 + \sum_{j=1}^{\ell-1}\left(\frac{2}{\pi}\right)^j \zeta_j\right]^{-1}$$

$$\gamma_\ell \equiv \frac{4\zeta_\ell \alpha_\ell^2}{\pi + 4\zeta_\ell \alpha_\ell^2} = \left(\frac{2}{\pi}\right)^\ell \zeta_\ell \left[1 + \sum_{j=1}^{\ell}\left(\frac{2}{\pi}\right)^j \zeta_j\right]^{-1} \qquad\qquad \text{(A.15)}$$

$$\mathcal{G}_{\ell+1}(p) \equiv 1 - \frac{2}{\pi}\arccos\left((1 - \gamma_\ell)\mathcal{G}_\ell(p) + \gamma_\ell(1 - 2p)\right)$$

with the initial conditions

$$\alpha_1^2 = \frac{1}{2}$$

$$\mathcal{G}_1(p) = 1 - \frac{2}{\pi}\arccos(1 - 2p).$$

Observe that this process can be repeated until the very last iteration, where $\tilde{X}_{L+2} = \tilde{Y}_{L+2} = 0$. Therefore

$$\mathfrak{d}(p, \mathbf{Q}) = \frac{1}{\pi}\arccos\left((1 - \gamma_L)\mathcal{G}_L(p) + \gamma_L(1 - 2p)\right).$$

Appendix B. Calculation of the partial overlap matrix probability

Let us suppose that we have selected K vectors from $\mathbb{S}^N$ such that there is a constraint on the angle between any pair of them, and as a result we can construct the matrix $\mathbf{Q}$. Let us find now the distribution probability of the matrix elements correspondent to the partial sums of the type $\frac{1}{N}\sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1, j} w_{\mathbf{k}_1', j}$. To this end we define the following quantities:

$$r_{\mathbf{k}_1}^2 \equiv \frac{1}{N}\sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1, j}^2$$

$$r_{\mathbf{k}_1} r_{\mathbf{k}_1'}\cos(\varphi_{\mathbf{k}_1\mathbf{k}_1'}) \equiv \frac{1}{N}\sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1, j} w_{\mathbf{k}_1', j}$$

so the task is to find the distribution of the variables $r_{\mathbf{k}_1}^2$ (diagonal elements) and $r_{\mathbf{k}_1} r_{\mathbf{k}_1'}\cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)$ (off-diagonal elements).

It is then

$$
\mathcal{P}\big(r_{\mathbf{k}_1}^2\big) = \int d\mathbf{w}_{\mathbf{k}_1}\, \mathcal{P}\big(r_{\mathbf{k}_1}^2\big|\mathbf{w}_{\mathbf{k}_1}\big)\mathcal{P}\big(\mathbf{w}_{\mathbf{k}_1}\big)
$$

$$
\propto \int d\mathbf{w}_{\mathbf{k}_1}\delta\left(Nr_{\mathbf{k}_1}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j}^2\right)\delta\left(\sum_{j=1}^N w_{\mathbf{k}_1,j}^2 - N\right)
$$

$$
\propto \int d\mathbf{w}_{\mathbf{k}_1}\delta\left(Nr_{\mathbf{k}_1}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j}^2\right)\delta\left(Nr_{\mathbf{k}_1}^2 + \sum_{j\notin\mathbb{I}_P} w_{\mathbf{k}_1,j}^2 - N\right)
$$

$$
\propto \int \prod_{j\in\mathbb{I}_P} dw_{\mathbf{k}_1,j}\delta\left(\sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j}^2 - Nr_{\mathbf{k}_1}^2\right)\int \prod_{j\notin\mathbb{I}_P} dw_{\mathbf{k}_1,j}\delta\left(\sum_{j\notin\mathbb{I}_P} w_{\mathbf{k}_1,j}^2 - N\big(1 - r_{\mathbf{k}_1}^2\big)\right)
$$

$$
\propto r_{\mathbf{k}_1}^{P-1}\big(1 - r_{\mathbf{k}_1}^2\big)^{(N-P-1)/2}
$$

$$
= \left[NB\left(\frac{P+1}{2}, \frac{N-P+1}{2}\right)\right]^{-1} r_{\mathbf{k}_1}^{P-1}\big(1 - r_{\mathbf{k}_1}^2\big)^{(N-P-1)/2},
$$

where $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ is the beta function. For the off-diagonal elements we proceed in the same manner:

$$
\mathcal{P}\big(r_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\big(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\big)\big) = \int d\mathbf{w}_{\mathbf{k}_1} d\mathbf{w}_{\mathbf{k}_1'}\, \mathcal{P}\big(r_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\big(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\big)\big|\mathbf{w}_{\mathbf{k}_1}, \mathbf{w}_{\mathbf{k}_1'}\big)\mathcal{P}\big(\mathbf{w}_{\mathbf{k}_1}, \mathbf{w}_{\mathbf{k}_1'}\big)
$$

$$
= \int d\mathbf{w}_{\mathbf{k}_1} d\mathbf{w}_{\mathbf{k}_1'}\, \mathcal{P}\big(r_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\big(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\big)\big|\mathbf{w}_{\mathbf{k}_1}, \mathbf{w}_{\mathbf{k}_1'}\big)\mathcal{P}\big(\mathbf{w}_{\mathbf{k}_1'}\big|\mathbf{w}_{\mathbf{k}_1}\big)\mathcal{P}\big(\mathbf{w}_{\mathbf{k}_1}\big)
$$

$$
\propto \int d\mathbf{w}_{\mathbf{k}_1} d\mathbf{w}_{\mathbf{k}_1'}\delta\left(Nr_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\big(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\big) - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j} w_{\mathbf{k}_1',j}\right)\delta\left(Nr_{\mathbf{k}_1}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j}^2\right)
$$

$$
\times \delta\left(Nr_{\mathbf{k}_1'}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1',j}^2\right)\delta\left(\sum_{j=1}^N w_{\mathbf{k}_1,j} w_{\mathbf{k}_1',j} - N\tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'}\right)
$$

$$
\times \delta\left(\sum_{j=1}^N w_{\mathbf{k}_1,j}^2 - N\right)\delta\left(\sum_{j=1}^N w_{\mathbf{k}_1',j}^2 - N\right),
$$

where $\tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'}$ is the overlap between the vectors $\mathbf{w}_{\mathbf{k}_1}$ and $\mathbf{w}_{\mathbf{k}_1'}$. Observe that as in the case of the diagonal elements, we can separate the variables into two groups, according to whether their indices are in $\mathbb{I}_P$ or not. Thus

$$
\mathcal{P}\big(r_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\big(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\big)\big) = \int \prod_{j\in\mathbb{I}_P} dw_{\mathbf{k}_1,j} dw_{\mathbf{k}_1',j}\,\delta\left(Nr_{\mathbf{k}_1}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j}^2\right)\delta\left(Nr_{\mathbf{k}_1'}^2 - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1',j}^2\right)
$$

$$
\times \delta\left(Nr_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos(\varphi_{\mathbf{k}_1\mathbf{k}_1'}) - \sum_{j\in\mathbb{I}_P} w_{\mathbf{k}_1,j} w_{\mathbf{k}_1',j}\right)\int \prod_{j\notin\mathbb{I}_P} dw_{\mathbf{k}_1,j} dw_{\mathbf{k}_1',j}
$$

$$
\times \delta\left(\sum_{j\notin\mathbb{I}_P} w_{\mathbf{k}_1,j}^2 - N\big(1 - r_{\mathbf{k}_1}^2\big)\right)\delta\left(\sum_{j\notin\mathbb{I}_P} w_{\mathbf{k}_1',j}^2 - N\big(1 - r_{\mathbf{k}_1'}^2\big)\right)
$$

$$\times \delta \left(\sum_{j \notin \mathbb{I}_P} w_{\mathbf{k}_1, j} w_{\mathbf{k}'_1, j} - N \big(\tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} - r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \big) \right).$$

Let us define the overlap $\cos \big(\tilde{\varphi}_{\mathbf{k}_1 \mathbf{k}'_1} \big)$ by

$$\cos \big(\tilde{\varphi}_{\mathbf{k}_1 \mathbf{k}'_1} \big) \equiv \frac{\tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} - r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)}{\sqrt{\big(1 - r_{\mathbf{k}_1}^2 \big) \big(1 - r_{\mathbf{k}'_1}^2 \big)}}. \tag{B.1}$$

The integrals to be solved are of the form

$$\mathcal{I} \equiv \int_{\mathbb{R}^d} \mathrm{d}\boldsymbol{x}_1 \mathrm{d}\boldsymbol{x}_2 \delta(\boldsymbol{x}_1 \cdot \boldsymbol{x}_2 - N r_1 r_2 \cos(\varphi)) \, \delta \big(|\boldsymbol{x}_1|^2 - N r_1^2 \big) \delta \big(|\boldsymbol{x}_2|^2 - N r_2^2 \big)$$

which is proportional to the length of the hyper-ring around $\boldsymbol{x}_2$ with radius $|\boldsymbol{x}_1| \sin(\varphi)$ times the length of the hyper-ring around $\boldsymbol{x}_1$ with radius $|\boldsymbol{x}_2| \sin(\varphi)$. Therefore

$$\mathcal{I} \propto (r_1 r_2)^{d-1} \sin^{d-2}(\varphi)$$

thus

$$\mathcal{P} \big(r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \big) \propto \big(r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \big)^{P-1} \sin^{P-2} \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)$$
$$\times \big[\big(1 - r_{\mathbf{k}_1}^2 \big) \big(1 - r_{\mathbf{k}'_1}^2 \big) \big]^{(N-P-1)/2} \sin^{N-P-2} \big(\tilde{\varphi}_{\mathbf{k}_1 \mathbf{k}'_1} \big),$$

and, by (B.1) we have that

$$\big(1 - r_{\mathbf{k}_1}^2 \big) \big(1 - r_{\mathbf{k}'_1}^2 \big) \sin^2 \big(\tilde{\varphi}_{\mathbf{k}_1 \mathbf{k}'_1} \big) = \big(1 - \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1}^2 \big) - r_{\mathbf{k}_1}^2 - r_{\mathbf{k}'_1}^2$$
$$+ 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) + r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big).$$

Therefore

$$\mathcal{P} \big(r_{\mathbf{k}_1}^2 \big) \propto r_{\mathbf{k}_1}^{P-1} \big(1 - r_{\mathbf{k}_1}^2 \big)^{(N-P-1)/2}$$

$$\mathcal{P} \big(r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \big) \propto \big(r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \big)^{P-1} \sin^{P-2} \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \sqrt{\big(1 - r_{\mathbf{k}_1}^2 \big) \big(1 - r_{\mathbf{k}'_1}^2 \big)}$$
$$\times \big[1 - \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1}^2 - r_{\mathbf{k}_1}^2 - r_{\mathbf{k}'_1}^2 + 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) + r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1}^2 \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \big]^{(N-P-2)/2}.$$

In the thermodynamic limit, where $P \leqslant N \to \infty$ and $P/N = p$, the saddle point equations of the Laplace method are

$$0 = \frac{p}{r_{\mathbf{k}_1}} - \frac{1-p}{1 - r_{\mathbf{k}_1}^2} r_{\mathbf{k}_1} \quad \forall k_1 \tag{B.2}$$

for the diagonal elements, and

$$0 = \frac{p}{r_{\mathbf{k}_1}} - \frac{1-p}{2} \frac{2 r_{\mathbf{k}_1} - 2 r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) - 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1}^2 \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)}{1 - \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1}^2 - r_{\mathbf{k}_1}^2 - r_{\mathbf{k}'_1}^2 + 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) + r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1}^2 \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)} \tag{B.3}$$

$$0 = \frac{p}{r_{\mathbf{k}'_1}} - \frac{1-p}{2} \frac{2 r_{\mathbf{k}'_1} - 2 r_{\mathbf{k}_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) - 2 r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1} \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)}{1 - \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1}^2 - r_{\mathbf{k}_1}^2 - r_{\mathbf{k}'_1}^2 + 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) + r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1}^2 \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)} \tag{B.4}$$

$$0 = p \cot \mathrm{g} \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)$$
$$- \frac{1-p}{2} \frac{2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \sin \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) - 2 r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1}^2 \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) \sin \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)}{1 - \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1}^2 - r_{\mathbf{k}_1}^2 - r_{\mathbf{k}'_1}^2 + 2 r_{\mathbf{k}_1} r_{\mathbf{k}'_1} \tilde{\zeta}_{\mathbf{k}_1 \mathbf{k}'_1} \cos \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big) + r_{\mathbf{k}_1}^2 r_{\mathbf{k}'_1}^2 \sin^2 \big(\varphi_{\mathbf{k}_1 \mathbf{k}'_1} \big)}, \tag{B.5}$$

$\forall \mathbf{k}_1 \neq \mathbf{k}_1'$. From (B.2) we have that $r_{\mathbf{k}_1}^2 = p \ \forall \mathbf{k}_1$. Equations (B.3) and (B.4) are identical; therefore, we can suppose that in the limit $r_{\mathbf{k}_1} = r_{\mathbf{k}_1'} = r$ and the system gets reduced to only two equations

$$0 = p - (1-p)r^2 \frac{1 - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'} \cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) - r^2 \sin^2\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)}{1 - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'}^2 - 2r^2\left(1 - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'} \cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)\right) + r^4 \sin^2\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)}$$

$$0 = p \cot g\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) - (1-p)r^2 \frac{\tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'} \sin\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) - r^2 \cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) \sin\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)}{1 - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'}^2 - 2r^2\left(1 - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'} \cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)\right) + r^4 \sin^2\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)}$$

which accepts the solution $r^2 = p$ and $\cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) = \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'} \ \forall \mathbf{k}_1 \mathbf{k}_1'$. Therefore, in the thermodynamic limit, we have that

$$\mathcal{P}\left(r_{\mathbf{k}_1}^2\right) \propto \delta\left(r_{\mathbf{k}_1}^2 - p\right)$$

$$\mathcal{P}\left(r_{\mathbf{k}_1} r_{\mathbf{k}_1'} \cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right)\right) \propto \delta\left(r_{\mathbf{k}_1}^2 - p\right)\delta\left(r_{\mathbf{k}_1'}^2 - p\right)\delta\left(\cos\left(\varphi_{\mathbf{k}_1\mathbf{k}_1'}\right) - \tilde{\zeta}_{\mathbf{k}_1\mathbf{k}_1'}\right)$$

which implies that the expected value of the matrix of partial overlaps given these distributions of elements is

$$\lim_{P \leqslant N \to \infty} \binom{N}{P}^{-1} \sum_{\mathbb{I}_P} \frac{1}{N} \sum_{j \in \mathbb{I}_P} w_{\mathbf{k}_1, j} w_{\mathbf{k}_1', j} = p \, [Q]_{\mathbf{k}_1, \mathbf{k}_1'} . \tag{B.6}$$

Appendix C. Integrals

In here we present the calculation of several integrals that appear in the main development.

C.1. Integrals related to (A.12)

In order to simplify the following developments, we will suppose that $\exists \, M \in \mathbb{N} \, | \, K = 2M + 1$. The integrals to be solved are

$$b_0^M \equiv \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2M+1}$$

$$b_m^M \equiv \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2(M-m)+1} \sin(\hat{\eta})^{2m}$$

$$a_m^M \equiv \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2(M-m)} \sin(\hat{\eta})^{2m+1}.$$

Before computing the integrals observe that for all $A > 0$ and $B \geqslant 0$

$$\int \mathcal{D}(\eta, \hat{\eta}) \sin(A\hat{\eta}) = -\frac{\mathrm{i}}{4\pi} \int_0^\infty \mathrm{d}\eta \int_{-\infty}^\infty \mathrm{d}\hat{\eta} \exp(-\mathrm{i}\hat{\eta}\eta)[\exp(\mathrm{i}\hat{\eta}A) - \exp(-\mathrm{i}\hat{\eta}A)]$$

$$= -\frac{\mathrm{i}}{4\pi} \int_0^\infty \mathrm{d}\eta \int_{-\infty}^\infty \mathrm{d}\hat{\eta}[\exp[-\mathrm{i}\hat{\eta}(\eta - A)] - \exp[-\mathrm{i}\hat{\eta}(\eta + A)]]$$

$$= -\frac{\mathrm{i}}{2}[\Theta(A) - \Theta(-A)]$$

$$= -\frac{\mathrm{i}}{2}, \tag{C.1}$$

similarly

$$\int \mathcal{D}(\eta, \hat{\eta}) \cos(A\hat{\eta}) \cos(B\hat{\eta}) = \frac{1}{4}[\Theta(A+B) + \Theta(-A-B) + \Theta(A-B) + \Theta(-A+B)]$$

$$= \frac{1}{2} \tag{C.2}$$

and

$$\int \mathcal{D}(\eta, \hat{\eta}) \cos(A\hat{\eta}) \sin(B\hat{\eta}) = -\frac{i}{4} [\Theta(A+B) - \Theta(A-B) + \Theta(-A+B) - \Theta(-A-B)]$$

$$= -\frac{i}{2} \Theta(B-A). \tag{C.3}$$

The first integral is

$$b_0^M = \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2M+1} = \frac{1}{2^{2M}} \sum_{k=0}^{M} \binom{2M+1}{k} \int \mathcal{D}(\eta, \hat{\eta}) \cos[(2M - 2k + 1)\hat{\eta}]$$

$$= \frac{1}{2^{2M+1}} \sum_{k=0}^{M} \binom{2M+1}{k} = \frac{1}{2} \quad \forall M. \tag{C.4}$$

The second integral is

$$b_m^M = \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2(M-m)+1} \sin(\hat{\eta})^{2m}$$

$$= \int \mathcal{D}(\eta, \hat{\eta}) \frac{1}{2^{2(M-m)}} \sum_{k=0}^{M-m} \binom{2(M-m)+1}{k} \cos[(2M - 2(k+m) + 1)\hat{\eta}]$$

$$\times \frac{1}{2^{2m}} \left\{ 2 \sum_{j=0}^{m-1} (-1)^{m-j} \binom{2m}{j} \cos[(2m - 2j)\hat{\eta}] + \binom{2m}{m} \right\}$$

$$= \frac{1}{2^{2M}} \sum_{k=0}^{M-m} \binom{2(M-m)+1}{k} \sum_{j=0}^{m-1} (-1)^{m-j} \binom{2m}{j} + \frac{1}{2^{2m+1}} \binom{2m}{m}$$

$$= \frac{1}{2^{2M}} \frac{2^{2(M-m)+1}}{2} \left[-\frac{1}{2} \binom{2m}{m} \right] + \frac{1}{2^{2m+1}} \binom{2m}{m}$$

$$= 0 \quad \forall M, m > 0. \tag{C.5}$$

And the last integral is then

$$a_m^M = \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2(M-m)} \sin(\hat{\eta})^{2m+1}$$

$$= \int \mathcal{D}(\eta, \hat{\eta}) \cos(\hat{\eta})^{2(M-m)} [1 - \cos^2(\hat{\eta})]^m \sin(\hat{\eta})$$

$$= \int \mathcal{D}(\eta, \hat{\eta}) \sum_{\ell=0}^{m} (-1)^\ell \binom{m}{\ell} \cos(\hat{\eta})^{2(M-m+\ell)} \sin(\hat{\eta})$$

$$= \sum_{\ell=0}^{m} (-1)^\ell \binom{m}{\ell} \frac{(-i)}{2^{2(M-m+\ell)}} \left\{ \frac{1}{2} \binom{2(M-m+\ell)}{M-m+\ell} + \right.$$

$$\left. + \sum_{k=0}^{M-m+\ell-1} \binom{2(M-m+\ell)}{k} \Theta[1 - 2(M-m+\ell-k)] \right\}$$

$$= -\frac{i}{2^{2(M-m)+1}} \sum_{\ell=0}^{m} \left(-\frac{1}{4} \right)^\ell \binom{m}{\ell} \binom{2(M-m+\ell)}{M-m+\ell}$$

$$= -\frac{i}{2^{2M+1}} \binom{M}{m}^{-1} \binom{2m}{m} \binom{2(M-m)}{M-m}. \tag{C.6}$$

Appendix D. Derivation of condition (10)

We start by writing the average discrepancy component as

$$\mathfrak{d}_P^{(N)}[f] = \frac{1}{2} - \frac{1}{2} \sum_{\{\mathbf{S}\},\{\mathbf{S}'\}} \mathcal{P}(\mathbf{S},\mathbf{S}') f(\mathbf{S}) f(\mathbf{S}'),$$

where $\mathcal{P}(\mathbf{S},\mathbf{S}') = 2^{-N} \binom{N}{P}^{-1} \delta_{\mathbf{S}^{\mathsf{T}}\mathbf{S}', N-2P}$. The second term can be then expressed as

$$\sum_{\{\mathbf{S}\},\{\mathbf{S}'\}} \mathcal{P}(\mathbf{S},\mathbf{S}') f(\mathbf{S}) f(\mathbf{S}') = 2^{-N} \binom{N}{P}^{-1} \sum_{\{\mathbf{S}\},\{\mathbf{S}'\}} \delta_{\mathbf{S}^{\mathsf{T}}\mathbf{S}', N-2P} \sum_{\mathbb{I},\mathbb{J}} \hat{f}_{\mathbb{I}} \hat{f}_{\mathbb{J}} \prod_{k\in\mathbb{I}} \prod_{l\in\mathbb{J}} S_k S_l'.$$

By expressing the Kronecker delta as

$$\delta_{r,m} = \frac{1}{2\pi i} \oint \frac{\mathrm{d}Z}{Z^{m+1}} Z^r$$

we have

$$\sum_{\{\mathbf{S}\},\{\mathbf{S}'\}} \mathcal{P}(\mathbf{S},\mathbf{S}') f(\mathbf{S}) f(\mathbf{S}') = 2^{-N} \binom{N}{P}^{-1} \sum_{\mathbb{I},\mathbb{J}} \hat{f}_{\mathbb{I}} \hat{f}_{\mathbb{J}} \frac{1}{2\pi i} \oint \frac{\mathrm{d}Z}{Z^{P+1}}$$

$$\times \prod_{k\in\mathbb{I}} \prod_{l\in\mathbb{J}} \prod_{j=1}^{N} \frac{1}{2} \sum_{S_j=\pm 1} \sum_{S_j'=\pm 1} Z^{\frac{1}{2}(1-S_j S_j')} S_k S_l'.$$

Now we have to take the average over the input variables. To do so we define the following sets: $\mathbb{A} \equiv \mathbb{I} \cap \mathbb{J}$, $\mathbb{B} \equiv \mathbb{I}/\mathbb{A}$, $\mathbb{C} \equiv \mathbb{J}/\mathbb{A}$ and $\mathbb{D} \equiv [N]/(\mathbb{I} \cup \mathbb{J})$. These four sets completely cover $[N]$ without intersecting each other. The spin average is then

$$g(Z) \equiv \prod_{j=1}^{N} \frac{1}{2} \sum_{S_j=\pm 1} \sum_{S_j'=\pm 1} Z^{\frac{1}{2}(1-S_j S_j')} \prod_{k\in\mathbb{I}} S_k \prod_{l\in\mathbb{J}} S_l'$$

$$= (1 - Z)^{|\mathbb{A}|} \delta_{\mathbb{B},\varnothing} \delta_{\mathbb{C},\varnothing} (1 + Z)^{|\mathbb{D}|}.$$

If $\mathbb{B} = \mathbb{C} = \varnothing$, then $\mathbb{I} = \mathbb{J}$ and then

$$g(Z) = \delta_{\mathbb{I},\mathbb{J}} (1 - Z)^{|\mathbb{I}|} (1 + Z)^{N-|\mathbb{J}|}.$$

Let us write the sum over the sets of indexes $\sum_{\mathbb{I}}$ as $\sum_{r=1}^{N} \sum_{\mathbb{I}_r}$, where $|\mathbb{I}_r| = r$. Then, if

$$\mathscr{A}_r \equiv \sum_{\mathbb{I}_r} \hat{f}_{\mathbb{I}_{y_r}}^2 , \tag{D.1}$$

we can write

$$\mathfrak{d}_P^{(N)}[f] = \frac{1}{2} - \frac{1}{2} \binom{N}{P}^{-1} \sum_{r=1}^{N} \mathscr{A}_r \frac{1}{2\pi i} \oint \frac{\mathrm{d}Z}{Z^{P+1}} (1-Z)^r (1+Z)^{N-r}$$

$$= \frac{1}{2} - \frac{1}{2} \binom{N}{P}^{-1} \sum_{r=1}^{N} \mathscr{A}_r \sum_{j=0}^{r} (-1)^j \binom{r}{j} \sum_{k=0}^{N-r} \binom{N-r}{k} \delta_{j+k,P}$$

$$= \frac{1}{2} - \frac{1}{2} \binom{N}{P}^{-1} \sum_{r=1}^{N} \mathscr{A}_r \sum_{j=0}^{\min\{r,P\}} (-1)^j \binom{r}{j} \binom{N-r}{P-j}.$$

In this case the difference of average discrepancy components is

$$\Delta_P^N[f] \equiv \mathfrak{d}_P^{(N)}[f] - \mathfrak{d}_{P+1}^{(N)}[f] = \frac{1}{2} \sum_{r=1}^{N} \mathscr{A}_r \sum_{j=0}^{r} (-1)^j \binom{r}{j} \left[\frac{\binom{N-r}{P-j}}{\binom{N}{P}} - \frac{\binom{N-r}{P+1-j}}{\binom{N}{P+1}} \right].$$

Observe that

$$\binom{r}{j} = \frac{\prod_{k=0}^{j-1}(r-k)}{j!} = \frac{\mathsf{P}_j(r)}{j!},$$

where $\mathsf{P}_j(r)$ is a polynomial of degree j in r. If $P \ll N$ then

$$\binom{N}{P} \simeq \frac{N^P}{P!} + O(N^{-1}).$$

Thus

$$\binom{N}{P}^{-1}\binom{r}{j}\binom{N-r}{P-j} \simeq \binom{P}{j}\frac{\mathsf{P}_j(r)\mathsf{P}_{P-j}(N-r)}{N^P} + O(N^{-1})$$

$$\simeq \binom{P}{j}\frac{a_P r^j (N-r)^{P-j}}{N^P} + O(N^{-1}),$$

where $a_P \in \mathbb{R}$ is a constant. We then have the following expression (disregarding terms of $O(1/N)$):

$$\Delta_P^N[f] \simeq \frac{1}{2}\sum_{r=1}^{N}\mathscr{A}_r\sum_{j=0}^{r}(-1)^j\left[\binom{P}{j}a_P\left(\frac{r}{N}\right)^j\left(1-\frac{r}{N}\right)^{P-j}\right.$$

$$\left.-\binom{P+1}{j}a_{P+1}\left(\frac{r}{N}\right)^j\left(1-\frac{r}{N}\right)^{P+1-j}\right]$$

and observe that both terms in the sum over j are polynomials of order P and $P+1$ in r/N. Thus

$$\Delta_P^N[f] \simeq \frac{1}{2}\sum_{r=1}^{N}\mathscr{A}_r\sum_{j=0}^{P+1}b_j^{(P)}\left(\frac{r}{N}\right)^j + O(N^{-1})$$

$$\simeq \frac{1}{2}\sum_{j=1}^{P+1}b_j^{(P)}\sum_{r=1}^{N}\mathscr{A}_r\left(\frac{r}{N}\right)^j + O(N^{-1}),$$

where $b_j^{(P)} \in \mathbb{R}$ are the polynomial coefficients and depend only on the value of P. From this expression it is clear that in order to satisfy condition (7) for all values of P we need to require that

$$\forall\, j \in \mathbb{N} \quad \lim_{N\to\infty}\frac{1}{N^j}\sum_{r=1}^{N}\mathscr{A}_r\, r^j = 0, \tag{D.2}$$

which is condition (10).

References

[1] Franco L and Anthony M 2006 *IEEE Trans. Neural Netw.* **17** 578
[2] Neirotti J P 2010 *J. Phys. A: Math. Theor.* **43** 015101
[3] Neirotti J P 2010 *J. Phys. A: Math. Theor.* **43** 125101
[4] Cybenko G 1988 *Math. Control Signals Syst.* **2** 303
[5] Engel A 1996 *J. Phys. A: Math. Gen.* **29** L323
[6] Urbanczik R 1997 *J. Phys. A: Math. Gen.* **30** L387
[7] Ahr M, Biehl M and Urbanczik R 1999 *J. Phys. A: Math. Gen.* **32** L531
[8] Rosen-Zvi M, Engel A and Kanter I 2001 *Phys. Rev. Lett.* **87** 078101
[9] Schwarze H and Hertz J 1993 *J. Phys. A: Math. Gen.* **26** 4919
[10] Schwarze H 1993 *J. Phys. A: Math. Gen.* **26** 5781

[11] Opper M 1994 *Phys. Rev. Lett.* **72** 2113
[12] Copelli M and Caticha N 1995 *J. Phys. A: Math. Gen.* **28** 1615
[13] Urbanczik R 1995 *J. Phys. A: Math. Gen.* **28** 7097
[14] Saad D and Solla S A 1995 *Phys. Rev. Lett.* **74** 4337
[15] Monasson R and Zecchina R 1995 *Phys. Rev. Lett.* **75** 2432
[16] Vicente R and Caticha N 1997 *J. Phys. A: Math. Gen.* **30** L599
[17] Saad D and Rattray M 1998 Optimal on-line learning in neural networks *On-Line Learning in Neural Networks*
        ed D Saad (Cambridge: Cambridge University Press)
[18] Copelli M and Caticha N 1998 On-line learning in neural networks *Universal Asymptotics in Committee
        Machines With Tree Architecture* ed D Saad (Cambridge: Cambridge University Press)
[19] Kinzel W and Kanter I 2002 *Interacting Neural Networks and Cryptography (Advances in Solid State Physics
        vol 42)* (Berlin: Springer)
[20] Kurant M and Thiran P 2007 *Phys. Rev.* E **76** 026102
[21] Cousseau F, Mimura K and Okada M 2008 Statistical Mechanics of Lossy Compression for Non-Monotonic
        Multilayer Perceptrons *ISIT2008: Proc. 2008 IEEE Int. Symp. on Information Theory (Toronto, Canada)*
        pp 509–13
[22] Cousseau F, Mimura K and Okada M 2010 *Phys. Rev.* E **81** 021104
[23] Church A 1936 *Am. J. Math.* **58** 345
[24] Turing A M 1937 *Proc. Lond. Math. Soc.* **42** 230
[25] Hartmanis J and Stearns R E 1965 *Trans. Am. Math. Soc.* **117** 285
[26] Kolmogorov A N 1965 *Probl. Inf. Transm.* **1** 1
[27] Li M and Vitányi P 1997 *An Introduction to Kolmogorov Complexity and its Applications* 2nd edn (Berlin:
        Springer)
[28] Franco L and Cannas S A 2000 *Neural Comput.* **12** 2405
[29] Franco L 2006 *Neurocomputing* **70** 351
[30] Mezard M, Parisi G and Virasoro M A 1987 *Spin Glasses Theory and Beyond* (Singapore: World Scientific)
[31] O'Donnel R W 2003 Computational applications of noise sensitivity *PhD Thesis* MIT
[32] Stefankovic D 2000 Fourier transform in computer science *Master Thesis* University of Chicago
[33] Ben-Or M and Linial N 1995 *Collective Coin Flipping (Randomness and Computation)* (New York: Academic)