Journal of Physics A: Mathematical and Theoretical

PAPER • OPEN ACCESS

On the capacity of a quantum perceptron for storing biased patterns

To cite this article: Fabio Benatti *et al* 2024 *J. Phys. A: Math. Theor.* **57** 025301

View the article online for updates and enhancements.

You may also like

- Crystallite growth limits in amorphous oxides
Seth David Linker, Christopher Ausbeck, Riccardo DeSalvo et al.
- Probing magnetoelectric effect in the spin-modulated magnet Fe2GeO4
Guanzhong Zhou, Yongsen Tang, Lin Lin et al.
- Experimental scalings of scrape-off layer particle flux width by outboard divertor Langmuir probes for deuterium and helium plasmas on EAST
Xiang Liu, Lingyi Meng, Jichan Xu et al.

This content was downloaded from IP address 193.204.9.102 on 20/12/2023 at 12:01

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **57** (2024) 025301 (25pp)
https://doi.org/10.1088/1751-8121/ad125b

On the capacity of a quantum perceptron for storing biased patterns

Fabio Benatti[1,2], **Giovanni Gramegna**[3,4,*], **Stefano Mancini**[5,6] **and Gibbs Nwemadji**[7,8]

[1] Dipartimento di Fisica, Universitá di Trieste, Strada Costiera 11, I-34151 Trieste, Italy
[2] Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Strada Costiera 11, I-34151 Trieste, Italy
[3] Dipartimento di Fisica, Università di Bari, I-70126 Bari, Italy
[4] INFN, Sezione di Bari, Bari 70126, Italy
[5] Scuola di Scienze e Tecnologie, Università di Camerino, I-62032 Camerino, Italy
[6] Istituto Nazionale di Fisica Nucleare, Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy
[7] Theoretical and Scientific Data Science Group, Scuola Internazionale Superiore di Studi Avanzati (SISSA), Via Bonomea 265, I-34136 Trieste, Italy
[8] Quantitative Life Sciences, The Abdus Salam International Centre for Theoretical Physics (ICTP), 34151 Trieste, Italy

E-mail: giovanni.gramegna@ba.infn.it

Received 10 August 2023; revised 15 November 2023
Accepted for publication 5 December 2023
Published 20 December 2023

Abstract

Although different architectures of quantum perceptrons have been recently put forward, the capabilities of such quantum devices versus their classical counterparts remain debated. Here, we consider random patterns and targets independently distributed with biased probabilities and investigate the storage capacity of a continuous quantum perceptron model that admits a classical limit, thus facilitating the comparison of performances. Such a more general context extends a previous study of the quantum storage capacity where using statistical mechanics techniques in the limit of a large number of inputs, it was proved that no quantum advantages are to be expected concerning the storage properties. This outcome is due to the fuzziness inevitably introduced by the intrinsic stochasticity of quantum devices. We strengthen such an indication by

* Author to whom any correspondence should be addressed.

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

© 2023 The Author(s). Published by IOP Publishing Ltd

showing that the possibility of indefinitely enhancing the storage capacity for highly correlated patterns, as it occurs in a classical setting, is instead prevented at the quantum level.

Keywords: quantum perceptron, storage capacity, quantum neural networks

1. Introduction

Machine learning aims at building methods that are able to make predictions or decisions based on sample data, without being explicitly programmed to do so. Quantum information theory studies the storage and transmission of information encoded in quantum states. Nowadays these two disciplines are becoming intertwined giving rise to the field of quantum machine learning.

The flow of ideas runs both ways: on the one hand applications of machine learning techniques are envisaged to analyze quantum systems [1, 2], on the other hand, the implementation of machine learning concepts on quantum hardware is also actively investigated [3–5]. Along this latter avenue quantum advantages are expected in terms of higher storage capabilities and an increased information processing power [6–9].

The task of precisely comparing the power of quantum and classical neural networks as probabilistic models for information processing and storage is thus becoming pressing. In particular, the issue of determining precisely the storage capacity of the most elementary constituent of a neural network, namely the perceptron [10], has been addressed in the classical scenario without referring to any specific learning rule using several approaches, ranging from combinatorics [11, 12] to statistical mechanics methods [13–16]. The latter has been used recently to generalize the calculation to some models of quantum perceptrons [17–19]. However, the results depend on the specific model used (see e.g. [18, 19], based respectively on the models [5, 20]).

Here, by referring to the continuous variable quantum perceptron model introduced in [20], we study the storage capacity of random classical binary patterns. The components of the patterns and their assigned output classification are taken to be independent and identically distributed (*i.i.d.*) according to a probability with a bias $-1 \leqslant m_{\rm in} \leqslant 1$ for the patterns and $-1 \leqslant m_{\rm out} \leqslant 1$ for their classification. Such a model admits the classical perceptron as a classical limit, thus enabling a direct comparison of the storage performances in the two cases. For classical perceptrons, simultaneously large biases for patterns and output classification allow to greatly enhance the storage capacity, which diverges when $m_{\rm in} = m_{\rm out} = m \to 1$ [13, 21–23]. We show that this possibility is prevented at the quantum level. Moreover, we also find that, when the biases $m_{\rm in}$ and $m_{\rm out}$ are varied separately, the quantum storage capacity depends on both of them, unlike in the classical case, where the storage capacity is a function only of the output bias. However, also in the quantum setting, when the output correlations are maximal, that is when $m_{\rm out} \to 1$, the asymptotic behavior is no more dependent on $m_{\rm in}$, exactly as in the classical case. The dependence of the quantum storage capacity on $m_{\rm in}$ in such a case is through the velocity with which the limit behavior is reached. Overall, the performances of the continuous quantum model remain below the classical ones. These results thus corroborate those found in [19] with unbiased patterns. They confirm that, at the level of a simple, that is one-layer, quantum perceptron, the uncertainties brought about by pattern encoding via Gaussian states and homodyne measurements cannot be counteracted by linear super-positions of pattern states.

2. A continuous quantum perceptron model

The continuous variable model of a quantum perceptron proposed in [20] is characterized by N bosonic input modes and one bosonic output mode. We summarize here the main features of this model, for more details on the circuit implementation of such a perceptron the interested reader can refer to [20]. The components x_j of an input pattern $\boldsymbol{x} \in \mathbb{R}^N$ are encoded by states of the form

$$|\psi(x_j)\rangle = \frac{1}{\left(2\pi\sigma_j^2\right)^{1/4}} \int_{-\infty}^{+\infty} \mathrm{d}q_j \exp\left(-\frac{(q_j - x_j)^2}{4\sigma_j^2}\right) |q_j\rangle, \tag{1}$$

which are Gaussian weighted normalized super-positions of pseudo-eigenstates $|q_j\rangle$ of position-like operators q_j, centered around the pattern components x_j with widths σ_j. As a result, a pattern $\boldsymbol{x}$ is encoded into

$$|\Psi(\boldsymbol{x})\rangle = \bigotimes_{j=1}^{N} |\psi(x_j)\rangle. \tag{2}$$

Such a state is then given as input to a quantum circuit which first operates with a series of independent squeezing operators

$$S_j(r_j) = \mathrm{e}^{\mathrm{i}\, r_j (q_j p_j + p_j q_j)}\ , \quad r_j \in \mathbb{R}\ , \quad \mathrm{e}^{-2r_j} = w_j\ , \tag{3}$$

where p_j is a momentum-like operator conjugated to q_j ($[q_j, p_j] = \mathrm{i}$) and r_j is the squeezing parameter implementing the weight w_j.[9] Notice that

$$S_j(r_j)\,|q_j\rangle = \sqrt{w_j}\,|w_j q_j\rangle\ . \tag{4}$$

Then, the circuit consists of entangling controlled addition gates CX on pairs of consecutive modes:

$$\mathrm{CX}_{j,j+1} := \exp\left(-\mathrm{i}\, q_j \otimes p_{j+1}\right)\ , \qquad \mathrm{CX}_{j,j+1}\,|q_j, q_{j+1}\rangle = |q_j, q_j + q_{j+1}\rangle\ . \tag{5}$$

Their combined action on the attenuated multi-mode position eigenstates gives

$$\begin{aligned} |w_1 q_1, w_2 q_2, \ldots, w_N q_N\rangle \to |w_1 q_1, w_1 q_1 + w_2 q_2, \ldots, w_N q_N\rangle \to \ldots \\ \ldots \to |w_1 q_1, w_1 q_1 + w_2 q_2, \ldots, \sum_{j=1}^{N} w_j q_j\rangle\ . \end{aligned} \tag{6}$$

The sequence of operations described by equations (3)–(6) implements a parameterized unitary $U(\boldsymbol{w})$ on the initial state (2). Finally, we focus on the last mode, which is described by the density matrix

$$\rho_N(\boldsymbol{w}, \boldsymbol{x}) = \mathrm{Tr}_{1,\ldots,N-1}\left(U(\boldsymbol{w})\,|\Psi(\boldsymbol{x})\rangle\langle\Psi(\boldsymbol{x})|\,U^\dagger(\boldsymbol{w})\right), \tag{7}$$

and perform homodyne detection of the position-like quadrature, which yields a value s with probability density [20]

$$P_{\boldsymbol{w},\boldsymbol{x},\sigma}(s) = \langle s|\rho_N(\boldsymbol{w}, \boldsymbol{x})\,|s\rangle = \frac{1}{\sqrt{2\pi}\|\boldsymbol{w}\|\sigma} \exp\left(-\frac{(s - \boldsymbol{w}\cdot\boldsymbol{x})^2}{2\|\boldsymbol{w}\|^2\sigma^2}\right), \tag{8}$$

[9] In order to implement negative weights, a phase shift gate $\mathrm{e}^{\mathrm{i}\frac{\pi}{2}(q_j^2 + p_j^2)}$ is operated after the squeezing.

where, for sake of simplicity, we have set $\sigma_j^2 = \sigma^2$ for all j and thus encoded the input patterns by Gaussian states of the same width.

Remark 1. With a slight modification of the above protocol, it is possible to obtain a description of such a continuous variable quantum perceptron as controlled unitary acting on the tensor product $\mathscr{H} = \mathscr{H}_{\rm in} \otimes \mathscr{H}_{\rm out}$ where $\mathscr{H}_{\rm in}$ is the Hilbert space of the N bosonic modes encoding the input, while $\mathscr{H}_{\rm out}$ is the Hilbert space of the additional ancilla mode storing the output. Then, the present model can be connected with other models investigated in the literature, in particular [24], where it was pointed out that a perceptron acting as a controlled unitary has as particular cases also the models considered in [25, 26]. Actually, the action of the continuous variable quantum perceptron here investigated can be described with the unitary

$$U(\boldsymbol{w}) := \prod_{j=1}^{N} \mathrm{CX}_{j,\rm out} S_j(r_j), \tag{9}$$

with $S_j(r)$ as in equation (3), while $\mathrm{CX}_{j,\rm out} = \exp(-\mathrm{i}\, q_j \otimes p_{\rm out})$ is the controlled addition gate involving the jth bosonic mode of the input and the output mode.

During the training phase of a single perceptron, a set of p pairs $\{(\boldsymbol{x}^\mu, \xi^\mu)\}_{\mu=1,\dots,p}$ is provided, where $\boldsymbol{x}^\mu \in \{-1,1\}^N$ is the μth input pattern and $\xi^\mu \in \{-1,1\}$ is its corresponding classification, assumed to be known for all the elements in the training set. The storage of the p patterns in the training set by a single perceptron requires that a suitable choice of weights $\boldsymbol{w}$ exists for which the association $\boldsymbol{x}^\mu \mapsto \xi^\mu$ can be realized for each $\mu = 1,\dots,p$. Classically, the classification of a pattern $\boldsymbol{x}^\mu$ as ± 1 by a weight vector $\boldsymbol{w}$ is obtained by checking the sign of $\boldsymbol{w}\cdot\boldsymbol{x}^\mu$. Then, a correct classification relative to a prescribed target $\xi^\mu = \pm 1$ is obtained when $\xi^\mu\, \boldsymbol{w}\cdot\boldsymbol{x}^\mu \geqslant \kappa\|\boldsymbol{w}\|$ where κ is a stabilizing threshold. It renders the classification more robust against noise affecting the weights that, when $\kappa = 0$, might make $\xi^\mu\, \boldsymbol{w}\cdot\boldsymbol{x}^\mu$ jump from positive to negative values and vice versa. In the case of the quantum perceptron model outlined above, a pattern $\boldsymbol{x}^\mu$ is classified as $\xi^\mu = +1$ (resp. $\xi^\mu = -1$) if the measurement outcome is above the threshold $\kappa\|\boldsymbol{w}\|$ (resp. below $-\kappa\|\boldsymbol{w}\|$), while the pattern is not classified when the measurement outcome is in between $(-\kappa\|\boldsymbol{w}\|, \kappa\|\boldsymbol{w}\|)$. Therefore, even when classically $\mathrm{sign}(\boldsymbol{w}\cdot\boldsymbol{x}^\mu - \kappa\|\boldsymbol{w}\|) = +1$, quantumly, the pattern is classified as -1 if $s < -\kappa\|\boldsymbol{w}\|$ and such errors occur with probability density $P_{\boldsymbol{w},\boldsymbol{x}^\mu,\sigma}(s)$.

Consequently, the inherent randomness due to the quantum encoding of the patterns is such that the correct classification of pattern μ becomes a binary stochastic variable with probability distribution given by

$$R_{\kappa,\sigma}(\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu) = \int_{-\infty}^{+\infty} \mathrm{d}s\, P_{\boldsymbol{w},\boldsymbol{x}^\mu,\sigma}(s)\, \Theta\left(\xi^\mu \frac{s}{\|\boldsymbol{w}\|} - \kappa\right), \tag{10}$$

where $\Theta(\cdot)$ denotes the Heaviside function. Finally, an ancilla mode is appended to the initialized N ones and its state changed according to the actual outcome of a suitable homodyne measurement. The result of the measurement can then be used, to implement the non-linear activation function, that is, in the case of binary classification we are interested in, a sign function.

One of the advantages of the continuous quantum model just presented is that it allows to recover the functioning of the classical perceptron when $\sigma \to 0$, i.e. by encoding a pattern $\boldsymbol{x}^\mu$ into the position-like pseudo-autokets $|x_1^\mu, x_2^\mu, \dots, x_N^\mu\rangle$. Indeed, in this limit the Gaussian probability density in equation (10) becomes a Dirac delta centered around $\boldsymbol{w}\cdot\boldsymbol{x}^\mu$. Another advantage of using continuous variable architectures is that they are easily implementable experimentally, for example using photonic quantum computers. A recent proposal with a

focus on the experimental implementation is given for example in [27], while extensive reviews on this topic can also be found in [1, 28].

3. Statistical mechanics derivation of the storage capacity

3.1. Gardner's approach

According to Gardner's statistical approach [13], the optimal storage capacity of a simple perceptron can be obtained from the fraction of weights which correctly reproduces the desired input–output relations normalized to the total volume of allowed vectors $\boldsymbol{w}$. Indeed, the storage capacity is defined as the *critical* value α_c of the ratio

$$\alpha \equiv \frac{p}{N}, \tag{11}$$

of the number of patterns p to the dimension of the input space N such that the storage condition

$$\xi^{\mu} \frac{\boldsymbol{w}\cdot\boldsymbol{x}^{\mu}}{\|\boldsymbol{w}\|} \geqslant \kappa \tag{12}$$

cannot be satisfied anymore.

In fact, by increasing the number of patterns, the volume of vectors $\boldsymbol{w}$ realizing the condition (12) typically shrinks, and the relative volume of such weights vanishes. Then, it is exactly the limit of vanishing relative volume that leads to the storage capacity of the perceptron.

We shall consider weights for which $\|\boldsymbol{w}\|^2 = N$ so that their components are typically of order 1. Then, the fraction of weights $\boldsymbol{w}$ of length $\sqrt{N}$ in $\mathbb{R}^N$ that classify p binary patterns $\boldsymbol{x}^{\mu} \in \{+1,-1\}^N$, up to an error ϵ, is given by:

$$V_N^Q\left(\boldsymbol{x}^{\mu},\xi^{\mu},\kappa,\sigma,\epsilon\right) := \frac{1}{Z_N}\int_{\mathbb{R}^N} \mathrm{d}^N\boldsymbol{w}\;\delta\left(||\boldsymbol{w}||^2 - N\right)\prod_{\mu=1}^{p}\Theta\left(R_{\kappa,\sigma}\left(\boldsymbol{w},\boldsymbol{x}^{\mu},\xi^{\mu}\right) - 1 + \epsilon\right). \tag{13}$$

where the total volume of the space of the weights is

$$Z_N = \int_{\mathbb{R}^N} \mathrm{d}^N\boldsymbol{w}\;\delta\left(||\boldsymbol{w}||^2 - N\right) \overset{N\gg 1}{\simeq} \sqrt{\frac{(2\pi e)^N}{4\pi N}}\,, \tag{14}$$

namely the volume of the sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$. The relation to the classification of the pattern $\boldsymbol{x}^{\mu}$ is due to the fact that equation (8) represents the probability distribution of the measurement outcomes of the quantum perceptron encoding the patterns $\boldsymbol{x}^{\mu}$ into Gaussian states of variance σ.

It should be noted also that even if the volume defined by (13) contains the quantity $R_{\kappa,\sigma}$ defined in equation (10) as some sort of 'average' over the possible results obtained from the quantum measurement, this should not be interpreted operationally as performing many measurements and then inferring the classification based on many shots. Rather, we are only requiring that the probability of success in the single-shot classification is at least $1-\epsilon$ for each pattern we want to store.

Remark 2. The probability $R_{\kappa,\sigma}(\boldsymbol{w},\boldsymbol{x}^{\mu},\xi^{\mu})$ depends on the pattern $\boldsymbol{x}^{\mu}$, on the target classification ξ^{μ}, on the weights $\boldsymbol{w}$, on the threshold parameter κ and on the Gaussian width σ. Therefore, the fraction of volume V_N^Q depends on patterns, targets, threshold, width and also on the allowed statistical error ϵ: when the width σ vanishes, from the distributional limit

$$\lim_{\sigma\to 0} P_{\boldsymbol{w},\boldsymbol{x}^{\mu},\sigma}\left(s\right) = \delta\left(s - \boldsymbol{w}\cdot\boldsymbol{x}^{\mu}\right)\,, \tag{15}$$

one recovers the expression of the fraction of weights of the classical perceptron

$$V_N^C(\boldsymbol{x}^\mu,\xi^\mu,\kappa)=\frac{1}{Z_N}\int_{\mathbb{R}^N}\mathrm{d}^N\boldsymbol{w}\;\delta\left(\|\boldsymbol{w}\|^2-N\right)\prod_{\mu=1}^{p}\Theta\left(\xi^\mu\frac{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}{\|\boldsymbol{w}\|}-\kappa\right). \tag{16}$$

Notice that this expression does not depend on the statistical error ϵ that needs to be introduced in the quantum setting. Indeed, in this latter case, the measured parameter s is statistically distributed around the classical scalar product $\boldsymbol{w}\cdot\boldsymbol{x}^\mu/\|\boldsymbol{w}\|$. Therefore, $R_{\kappa,\sigma}(\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu)$ cannot be equal to 1 unless the Gaussian distribution becomes a Dirac delta peaked around it. Note that the statistical error ϵ is an upper bound to the perceptron allowed errors. The value $\epsilon=1/2$ for the bound to the errors is a particular one: in such a case, as far as the storage capacity is concerned, the quantum perceptron is expected to behave classically, in spite of the quantum pattern encoding. To see this, one can simply note that the condition $R_{\kappa,\sigma}(\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu)>1/2$ implemented by (13) is equivalent to

$$\int_{\kappa\|\boldsymbol{w}\|}^{\infty}\mathrm{d}sP_{\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu}(s)<\frac{1}{2}\text{ if }\xi^\mu=1,\quad\int_{-\infty}^{-\kappa\|\boldsymbol{w}\|}\mathrm{d}sP_{\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu}(s)<\frac{1}{2}\text{ if }\xi^\mu=-1. \tag{17}$$

Then, using the fact that $\boldsymbol{w}\cdot\boldsymbol{x}^\mu$ is the median point of (8), i.e.

$$\int_{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}^{\infty}\mathrm{d}sP_{\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu}(s)=\int_{-\infty}^{\boldsymbol{w}\cdot\boldsymbol{x}^\mu}\mathrm{d}sP_{\boldsymbol{w},\boldsymbol{x}^\mu,\xi^\mu}(s)=\frac{1}{2}, \tag{18}$$

condition (17) is satisfied if and only if $\xi^\mu\boldsymbol{x}^\mu\cdot\boldsymbol{w}>\kappa\|\boldsymbol{w}\|$, which is the condition implemented in (16).

In analogy with the partition function of statistical mechanics, we take $\log V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)$ as the relevant quantity, since it has the important property of being *self-averaging*, i.e. its average $\langle\log V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)\rangle$ is a good representative of its typical behavior for random choices of input patterns and targets [14–16]. In particular, this average will be computed considering the components of the input patterns as well as targets, to be binary stochastic variables distributed according to

$$\Pr\left(x_j^\mu=\pm1\right)=\frac{1\pm m_{\text{in}}}{2}\,,\qquad\Pr\left(\xi^\mu=\pm1\right)=\frac{1\pm m_{\text{out}}}{2}\,. \tag{19}$$

The parameters $-1\leqslant m_{\text{in}},m_{\text{out}}\leqslant1$ measure the bias between the binary values of patterns and targets, respectively, and thus of their correlations. The smaller the bias is, the greater the independence of their two possible values. In many of the previous works where a biased distribution was considered, it was assumed that input patterns and corresponding classifications in the training set have the same bias m. To the best of our knowledge, the first distinction between input bias and classification bias was considered in [23], but an extensive investigation of the dependence of the storage capacity on m_{in} and m_{out} when $\kappa>0$ is missing in the literature. Since this regime will be important for the quantum perceptron model we consider, we are going to investigate it.

Following the classical approach by Gardner we will derive a critical value α_c^Q such that for $\alpha<\alpha_c^Q$ we obtain a finite value (potentially vanishing) for the limit

$$\lim_{\substack{N,p\to\infty\\ p/N=\alpha}}\frac{\langle\log V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon))\rangle}{N}\,, \tag{20}$$

while for $\alpha > \alpha_c^Q$:

$$\lim_{\substack{N,p\to\infty \\ p/N=\alpha}} \frac{\langle \log V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)\rangle}{N} = -\infty \ . \tag{21}$$

3.2. Replica method and saddle point equations

The quenched average appearing in the limit (20) can be computed by means of the replica-trick [13, 29, 30]:

$$\langle \log V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)\rangle = \lim_{n\to 0} \frac{\langle [V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)]^n\rangle - 1}{n} = \lim_{n\to 0} \frac{\log\langle [V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)]^n\rangle}{n}. \tag{22}$$

The relevant quantity $\langle [V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)]^n\rangle$ involves n replicas indexed by the subscript γ:

$$\langle [V_N^Q(\boldsymbol{x}^\mu, \xi^\mu, \kappa, \sigma, \epsilon)]^n\rangle = \frac{1}{Z_N^n}\int_{\mathbb{R}^{nN}} \mathrm{d}^{nN}\boldsymbol{W}\, \delta_N^{(n)}\left(\|\boldsymbol{W}\|^2\right) \left\langle \prod_{\gamma=1}^{n}\prod_{\mu=1}^{p} \Theta\left(R_\gamma^\mu - 1 + \epsilon\right)\right\rangle_{\boldsymbol{x},\xi}, \tag{23}$$

where, for the sake of compactness, we introduced the symbols $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n) \in \mathbb{R}^{nN}$ and

$$\mathrm{d}^{nN}\boldsymbol{W} \equiv \prod_{\gamma=1}^{n} \mathrm{d}^N \boldsymbol{w}_\gamma \ , \quad \delta_N^{(n)}\left(\|\boldsymbol{w}\|^2\right) \equiv \prod_{\gamma=1}^{n} \delta\left(\|\boldsymbol{w}_\gamma\|^2 - N\right) \ . \tag{24}$$

Moreover, in equation (23), it is made explicit that the mean value is computed with respect to the patterns $\boldsymbol{x}^\mu$ and targets ξ^μ, $\mu = 1, \ldots, p$.

The lengthy calculations of the mean value in equation (23) by means of the replica-symmetric ansatz and of the saddle point approximation are reported in appendix A. The main difference with respect to the classical calculation performed by Gardner consists in the dependence on the random variables $\{\boldsymbol{x}^\mu, \xi^\mu\}$ over which the average is taken in (23).

The replica method introduces several order parameters, the most important one being the average overlap of two randomly chosen weights $\boldsymbol{w}_\gamma$ and $\boldsymbol{w}_\delta$ in different replicas:

$$q_{\gamma\delta} = \frac{1}{N}\sum_{j=1}^{N} w_j^\gamma w_j^\delta \ . \tag{25}$$

In the replica symmetric (RS) ansatz it is assumed that, for the solution of the saddle point equations, the average overlap is the same for each pair of replicas, i.e. $q_{\gamma\delta} = q$ for all $\gamma \neq \delta$. The RS ansatz is expected to be valid when the space of solutions for the problem is connected [16, 29]. This is not true, for example, for the case of the 'negative perceptron', corresponding to $\kappa < 0$, or for the binary perceptron (i.e. with binary weights, $w_j \in \{-1, 1\}$) where the solution space can become disconnected [31] and replica symmetry breaking is required to get the correct solution. However, we obtain (see below) that the quantum perceptron considered behaves essentially in the same way as a classical perceptron with a shifted stability $\kappa \to \tilde{\kappa}$, at least from the point of view of the storage capacity. Since in the meaningful range of the parameters ($0 \leqslant \epsilon \leqslant 1/2$) we get $\tilde{\kappa} \geqslant \kappa \geqslant 0$, the RS solution is sufficient for our purposes.

Notice that by increasing the ratio p/N, the number of weights satisfying (12) diminishes, hence their average overlap increases. The critical value of α_c both in the classical and the quantum scenario is then obtained in the limit of maximal overlap $q \to 1$.

Eventually, one arrives at the following equation that must be satisfied by the critical ratio α_c^Q of number of patterns to weight dimension, which according to (11) defines the quantum storage capacity:

$$\alpha_c^Q \left[\frac{1+m_{\text{out}}}{2} \int_{a_-(M)}^{+\infty} \mathrm{D}x \, (x - a_-(M))^2 + \frac{1-m_{\text{out}}}{2} \int_{a_+(M)}^{+\infty} \mathrm{D}x \, (x - a_+(M))^2 \right] = 1. \tag{26}$$

In the above expression,

$$a_\pm(M) \equiv -\frac{\tilde{\kappa} \pm m_{\text{in}} M}{\sqrt{1-m_{\text{in}}^2}}, \quad \mathrm{D}x \equiv \frac{\mathrm{d}x}{\sqrt{2\pi}} e^{-x^2/2}, \tag{27}$$

where

$$\tilde{\kappa} := \kappa + \sigma \Phi^{-1}(1-\epsilon), \tag{28}$$

and Φ^{-1} is the inverse function of

$$\Phi(x) \equiv \int_{-\infty}^{x} \mathrm{D}u = \frac{1 + \mathrm{erf}\left(x/\sqrt{2}\right)}{2} \quad \text{with} \quad \mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}u \, \mathrm{e}^{-u^2}, \tag{29}$$

while the quantity M satisfies

$$(1+m_{\text{out}}) \int_{a_-(M)}^{+\infty} \mathrm{D}x \, (x - a_-(M)) = (1-m_{\text{out}}) \int_{a_+(M)}^{+\infty} \mathrm{D}x \, (x - a_+(M)) \, . \tag{30}$$

Thus, in order to compute α_c^Q from (26), one has to first solve (30) in terms of M.

Remark 3. Notice that when $m_{\text{in}} = 0$, that is when the patterns are unbiased, we have $a_+(M) = a_-(M)$ so that equation (30) can be satisfied only for $m_{\text{out}} = 0$, and the storage capacity is fixed by (26) only, which coincides with the expression found in [19]. This is due to the fact that a perceptron cannot match unbiased patterns with biased classifications. In fact, considering the mapping realized by a perceptron with weights $\boldsymbol{w}$, one finds that for a random input $\boldsymbol{x} \in \mathbb{R}^N$, with independent components distributed according to $\Pr(x_j = \pm 1) = 1/2$ for each $j = 1, \ldots, N$, the distribution of the output σ is given by:

$$\Pr(\sigma = +1) = \Pr(\boldsymbol{w} \cdot \boldsymbol{x} \geqslant 0) = \Pr(\boldsymbol{w} \cdot \boldsymbol{x} \leqslant 0) = \Pr(\sigma = -1), \tag{31}$$

which implies $\Pr(\sigma = \pm 1) = 1/2$ (note that $\Pr(\boldsymbol{w} \cdot \boldsymbol{x} = 0) = 0$ for all $\boldsymbol{w} \in \mathbb{R}^N$ except for those belonging to a set with zero Lebesgue measure on the sphere with radius $\sqrt{N}$). However, this holds true only if $m_{\text{in}} = 0$ *exactly*. If we consider $m_{\text{in}} > 0$ and take the limit $m_{\text{in}} \to 0$ one can see that equation (26) can be solved without asking $m_{\text{out}} \to 0$, since in such a case equation (30) can also be satisfied by choosing $M \sim \frac{C}{m_{\text{in}}}$. This is clearly shown in the middle panel of figure 1, where the storage capacity does not vanish in the limit $m_{\text{in}} \to 0$ even for $m_{\text{out}} \neq 0$.

In practice, the combined effects of quantum pattern encodings and measurements is to replace the classical stabilizing threshold κ in $\tilde{\kappa}$ defined in (28). Then, the classical storage capacity obtains not only by eliminating the errors due to quantum pattern encoding, that is by letting $\sigma \to 0$, but also, confirming the argument in remark 2, when $\sigma \neq 0$, so that the pattern encoding is not sharp and carries quantum fuzziness; however, $\epsilon = 1/2$ so that $\Phi^{-1}(1-\epsilon) = 0$ and $\tilde{\kappa} = \kappa$.

Note that when $\tilde{\kappa} = 0$, equation (27) yields $a_\pm(M) \propto M$, with a prefactor dependent on m_{in}. Since only the value of $a_\pm(M)$ enters into equations (26) and (30), one can recognize that a change in m_{in} will rescale the value of M which solves (30), but it will not affect the critical storage capacity.

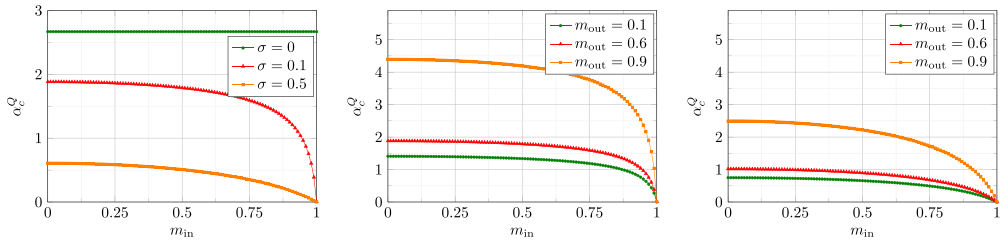

Figure 1. Storage capacity α_c^Q vs $m_{\rm in}$, for $\kappa = 0$, and $\epsilon = 0.01$. Only values $0 < m_{\rm in} < 1$ are shown here, but the results are symmetric with respect to $m_{\rm in} = 0$. (Left) The bias on the target classification is fixed to $m_{\rm out} = 0.6$, the shown curves corresponding to different values of σ. In the classical case ($\sigma = 0$) α_c^Q does not depend on $m_{\rm in}$. In the quantum case ($\sigma > 0$), increasing the value of $|m_{\rm in}|$ decreases α_c^Q. Furthermore, increasing σ always decreases the storage capacity. (Center) The value $\sigma = 0.1$ is fixed, the shown curves corresponding to different values of $m_{\rm out}$. Increasing the value of $|m_{\rm out}|$ always increases the storage capacity (Right) As before, with $\sigma = 0.3$, showing the lowering of the perceptron performance with increasing quantum fuzziness in the pattern encoding.

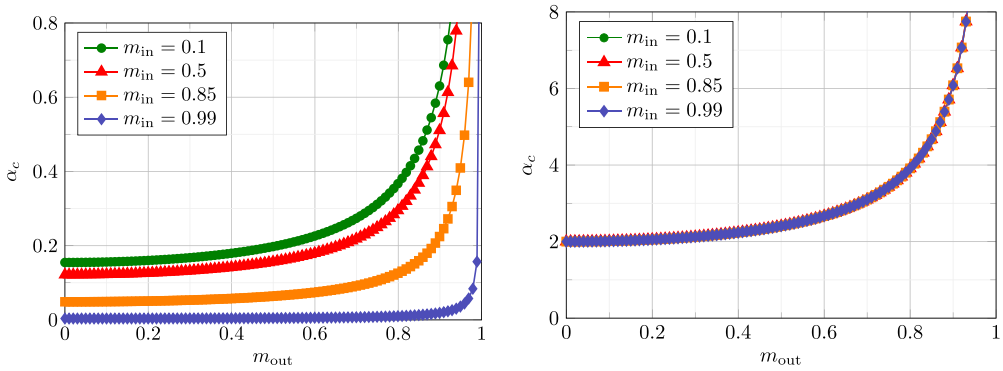

Figure 2. (Left) Storage capacity α_c^Q vs $m_{\rm out}$ for different values of $m_{\rm in}$, $\kappa = 0$, $\sigma = 1$ and $\epsilon = 0.01$. Increasing $|m_{\rm out}|$ always increases α_c^Q, while increments of $|m_{\rm in}|$ have the opposite effect. There is a divergence for $|m_{\rm out}| \to 1$ for each value of $m_{\rm in}$, although higher values of $|m_{\rm out}|$ are required to observe the divergence if $|m_{\rm in}|$ is increased. (Right) Storage capacity in the classical limit, obtained for $\sigma = 0$ (all the values of the other parameters are unchanged). All the curves corresponding to different values of $m_{\rm in}$ collapse into each other since in this case there is no dependence on $m_{\rm in}$ (recall that here $\kappa = 0$.

4. Results

The numerical results obtained by solving equations (26) and (30) for several values of $m_{\rm in}$, $m_{\rm out}$ and $\tilde{\kappa}$ are shown in figures 1 and 2. Since the storage capacity depends solely on $\widetilde{\kappa} = \kappa + \sigma\Phi^{-1}(1-\epsilon)$, we kept fixed the values $\epsilon = 0.01, \kappa = 0$ and considered different values of σ, which also allows us to recover the classical limit for $\sigma = 0$. A striking feature which distinguishes the quantum perceptron from the classical one is the dependence of the storage capacity on the bias $m_{\rm in}$, which is not present in the classical case with zero stability $\kappa = 0$ (see the left panel of figure 1). More precisely, as soon as $\sigma > 0$, increasing the value of $|m_{\rm in}|$ while keeping fixed the value of $m_{\rm out}$ always decreases the storage capacity α_c^Q. On the other hand, a common feature with the classical case is the divergence of the storage capacity when

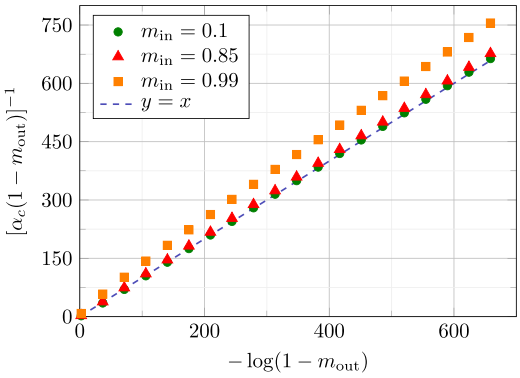

Figure 3. Limit α_c for $m_{\rm out} \to 1$, showing that the convergence to the asymptotic behavior (32), represented here by the line $y = x + q$, holds for several values of $m_{\rm in}$, although the onset of the asymptotic behavior appears later for $m_{\rm in}$ close to one.

$m_{\rm out} \to 1$, for each fixed value of $m_{\rm in}$ (see figure 2). The asymptotic behavior in this limit (see appendix B for the analytic derivation) is given by (see also figure 3):

$$\alpha_c^Q \simeq -\frac{1}{(1 - m_{\rm out}) \log (1 - m_{\rm out})} , \tag{32}$$

confirming the result obtained for the classical scenario.

Remark 4. An interesting feature emerging by differentiating the bias of the patterns $m_{\rm in}$ from the bias of target classification $m_{\rm out}$ is that even if the quantum storage capacity depends on $m_{\rm in}$ for each fixed $0 \leqslant m_{\rm out} \leqslant 1$, the asymptotic behavior when $m_{\rm out} \to 1^-$ does not depend on the pattern biases $m_{\rm in}$.

Even if the asymptotic behavior (32) does not show a dependence on the patterns bias $m_{\rm in}$, one can see from figures 2 and 3 that as the input bias $m_{\rm in}$ increases, higher values of $m_{\rm out}$ are required to observe the asymptotic behavior for the quantum perceptron. This is in contrast with the behavior of the classical perceptron, where there is no dependence at all on $m_{\rm in}$ (see again figure 2). In other words, values of $m_{\rm in}$ closer to 1 slow down the attainment of the asymptotic behavior in the quantum case, which motivates the investigation of the joint limit $m_{\rm in} = m_{\rm out} = m \to 1$. The results obtained (see figure 4) show another striking difference between the classical and the quantum perceptron, that is, while classically the storage capacity diverges when $m \to 1$, in the quantum case this divergence is suppressed. In particular, from the analytic asymptotic expressions (see appendix C) we obtain that the asymptotic behavior in the quantum scenario reads

$$\alpha_c^Q \sim \frac{1}{\tilde{\kappa}^2}, \tag{33}$$

which is finite for all values of $\sigma > 0$, $0 \leqslant \epsilon < 1/2$, while we recover the classical divergence (for $\kappa = 0$) in the classical limit $\sigma = 0$.

Remark 5. Figure 1 shows that for $\kappa = 0$ the separation between curves corresponding to different values of $m_{\rm out}$ is reduced in the quantum regime. This is in contrast to what happens for curves corresponding to different values of $m_{\rm in}$, as shown in figure 2.

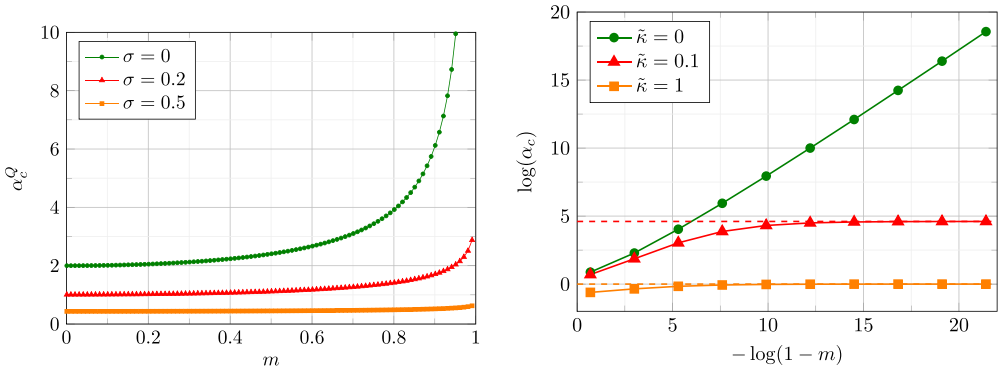

Figure 4. (Left) α_c^Q plotted against $m = m_{\text{in}} = m_{\text{out}}$. Increasing the value of the bias $|m|$ always increases the storage capacity. (Right) Limit α_c^Q for $m_{\text{out}} = m_{\text{in}} = m \to 1$, for several values of $\tilde{\kappa}$. The case $\tilde{\kappa} = 0$ corresponds to the classical case with $\kappa = 0$, where the storage capacity diverges in the limit. In the quantum scenario ($\tilde{\kappa} > 0$) the divergence is suppressed. The dashed lines correspond the asymptotic value $\alpha_c \sim 1/\tilde{\kappa}^2$.

5. Discussion and conclusion

Summarizing, we studied the storage capacity of the continuous variable model of quantum perceptron presented in [20] in the presence of a bias in the distribution of the patterns and their corresponding classifications. Besides the advantage of allowing an almost entirely analytical study, such a model admits the classical perceptron as a classical limit, thus allowing for a direct comparison of the storage performances in the two cases. We found that the additional randomness introduced in the quantum model gives rise to an effective increment in the stability parameter used in Gardner's statistical approach, $\kappa \to \tilde{\kappa}$, which gives rise to several peculiar features that are not observed in the classical case with zero stability.

For instance, the possibility of indefinitely enhancing the storage capacity by increasing the bias of the patterns and their classifications is prevented at the quantum level. Moreover, in contrast to the classical case, when the bias of the patterns and the bias in their corresponding classifications are varied separately, a dependence of the storage capacity on the input patterns' bias appears even when the stability parameter reduces to zero. Overall, however, the performance of the quantum perceptron model remains below that of the classical one. This is likely due to the fact that the considered quantum model introduces two sources of randomness: one due to the encoding of patterns by means of non-zero width Gaussian states and another one due to the final measurement operation implementing the classical non-linear activation function.

It should be noted that all the features characterizing the storage capacity of the quantum perceptron are obtained exclusively as a consequence of an effective stability $\tilde{\kappa} > 0$. In other words, one would be able to find these same features from the investigation of the classical perceptron in the regime of $\kappa > 0$, such as the different dependence on m_{in}, m_{out} and the different asymptotics in the limit $m = m_{\text{in}} = m_{\text{out}} \to 1$. It is worth stressing that the modification in the effective threshold $\tilde{\kappa} - \kappa = \sigma\Phi^{-1}(1-\epsilon)$ contains both the contribution of randomness coming from the width σ of the Gaussian encoding of the patterns, and the statistical error due to the quantum measurement ϵ: as a consequence, the worsening of the quantum storage capacity with respect to the classical one cannot be ascribed to only one of them. In this regard, it should be noted also that even if the volume defined by (13) contains the quantity

$R_{\kappa,\sigma}$ defined in equation (10) as some sort of 'average' over the possible results obtained from the quantum measurement, this should not be interpreted operationally as performing many measurements and then inferring the classification based on many shots. Rather, we are only requiring that the probability of success in the single-shot classification is at least $1-\epsilon$ for each pattern we want to store. Different approaches might be considered, such as for example considering the regime $\alpha > \alpha_c$ and focusing on the average success probability involving the classification of all the patterns to be stored. Another different venue would be to consider multi-layer quantum perceptrons, where one could hope for quantum advantages of the sort coming from linear superpositions and entanglement.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

F B, G G and S M acknowledge financial support from PNRR MUR Project PE0000023-NQSTI. S M also acknowledges financial support from the European Union's Horizon 2020 research and innovation programme, under Grant Agreement QUARTET No. 86264.

Appendix A. Derivation of the main result

In order to extract the large N behavior of $\langle[V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\rangle$ in equation (34), we recast it as

$$\left\langle\left[V_N^Q\left(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon\right)\right]^n\right\rangle = \frac{1}{Z_N^n}\int_{\mathbb{R}^{nN}} \mathrm{d}^{nN}\boldsymbol{W}\,\delta_N^{(n)}\left(\|\boldsymbol{W}\|^2\right)A\left(\boldsymbol{W}\right), \tag{34}$$

with

$$A\left(\boldsymbol{W}\right) := \left\langle \prod_{\mu=1}^{p}\prod_{\gamma=1}^{n}\Theta\left(R_{\kappa,\sigma}\left(\boldsymbol{w}_\gamma,\boldsymbol{x}^\mu,\xi^\mu\right)-1+\epsilon\right)\right\rangle_{\boldsymbol{x},\xi}. \tag{35}$$

Using the distributional expression

$$\Theta\left(x-u\right) = \int_u^{+\infty}\mathrm{d}\lambda\,\delta\left(\lambda-x\right) = \int_u^{+\infty}\mathrm{d}\lambda\int_{-\infty}^{+\infty}\frac{\mathrm{d}y}{2\pi}\;\mathrm{e}^{\mathrm{i}y(\lambda-x)}, \tag{36}$$

we express the Heaviside function in (35) as

$$\Theta\left(R_{\kappa,\sigma}\left(\boldsymbol{w}_\gamma,\boldsymbol{x}^\mu,\xi^\mu\right)-1+\epsilon\right) = \int_{1-\epsilon}^{+\infty}\mathrm{d}\lambda_\gamma^\mu\int_{-\infty}^{+\infty}\frac{\mathrm{d}y_\gamma^\mu}{2\pi}\;\mathrm{e}^{\mathrm{i}y_\gamma^\mu\left(\lambda_\gamma^\mu-R_{\kappa,\sigma}\left(\boldsymbol{w}_\gamma,\boldsymbol{x}^\mu,\xi^\mu\right)\right)}, \tag{37}$$

so that

$$A(\boldsymbol{W}) = \frac{1}{(2\pi)^{nP}}\int_{[1-\epsilon,+\infty)^{nP}}\mathrm{d}^{nP}\boldsymbol{\Lambda}\int_{\mathbb{R}^{nP}}\mathrm{d}^{nP}\boldsymbol{Y}\;\mathrm{e}^{\,\mathrm{i}\sum_{\mu=1}^{p}\sum_{\gamma=1}^{n}y_\gamma^\mu\lambda_\gamma^\mu}B(\boldsymbol{Y},\boldsymbol{W}), \tag{38}$$

where we introduced the symbols

$$\mathrm{d}^{nP}\mathbf{\Lambda} \equiv \prod_{\mu=1}^{p}\prod_{\gamma=1}^{n} \mathrm{d}\lambda_{\gamma}^{\mu}\ ,\ \ \mathrm{d}^{nP}\boldsymbol{Y} \equiv \prod_{\mu=1}^{p}\prod_{\gamma=1}^{n} \mathrm{d}y_{\gamma}^{\mu}\ , \tag{39}$$

and the quantity

$$B\left(\boldsymbol{Y},\boldsymbol{W}\right) \equiv \left\langle \prod_{\gamma=1}^{n}\prod_{\mu=1}^{p}\ \mathrm{e}^{-\,\mathrm{i}y_{\gamma}^{\mu}R_{\kappa,\sigma}\left(\boldsymbol{w}_{\gamma},\boldsymbol{x}^{\mu},\xi^{\mu}\right)} \right\rangle_{\boldsymbol{x},\xi}\ . \tag{40}$$

To proceed, it is convenient to use (see (29))

$$\Phi\left(x\right) \equiv \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{d}u\ e^{-u^{2}/2}$$

and rewrite (10) as

$$R_{\kappa,\sigma}\left(\boldsymbol{w}_{\gamma},\boldsymbol{x}^{\mu},\xi^{\mu}\right) = \int_{-\infty}^{+\infty} \mathrm{d}s\ P_{\boldsymbol{w}_{\gamma},\boldsymbol{x}^{\mu},\sigma}\left(s\right)\ \Theta\left(\xi^{\mu}\frac{s}{||\boldsymbol{w}||}-\kappa\right) = \Phi\left(-\frac{\kappa}{\sigma}+\xi^{\mu}\frac{\boldsymbol{x}^{\mu}\cdot\boldsymbol{w}_{\gamma}}{||\boldsymbol{w}_{\gamma}||\sigma}\right). \tag{41}$$

Then, using the exponential representation of the Dirac delta and the statistical independence of patterns and targets with different indices, one gets

$$\begin{aligned} B\left(\boldsymbol{Y},\boldsymbol{W}\right) &= \left\langle \prod_{\mu=1}^{p}\prod_{\gamma=1}^{n} \mathrm{e}^{-\,\mathrm{i}y_{\gamma}^{\mu}\Phi\left(-\frac{\kappa}{\sigma}+\xi^{\mu}\frac{\boldsymbol{x}^{\mu}\cdot\boldsymbol{w}_{\gamma}}{||\boldsymbol{w}_{\gamma}||\sigma}\right)} \right\rangle_{\boldsymbol{x}^{\mu},\xi^{\mu}} \\ &= \left\langle \prod_{\mu=1}^{p}\prod_{\gamma=1}^{n}\int_{\mathbb{R}} \mathrm{d}\eta_{\gamma}^{\mu}\ \delta\left(\eta_{\gamma}^{\mu}+\frac{\kappa}{\sigma}-\xi^{\mu}\frac{\boldsymbol{x}^{\mu}\cdot\boldsymbol{w}_{\gamma}}{||\boldsymbol{w}_{\gamma}||\sigma}\right)\ \mathrm{e}^{-\,\mathrm{i}y_{\gamma}^{\mu}\Phi\left(\eta_{\gamma}^{\mu}\right)} \right\rangle_{\boldsymbol{x},\xi} \\ &= \frac{1}{\left(2\pi\right)^{nP}}\int_{\mathbb{R}^{2nP}} \mathrm{d}^{nP}\mathbf{\Omega}\ \mathrm{d}^{nP}\boldsymbol{H}\ C\left(\mathbf{\Omega},\boldsymbol{W}\right) \\ &\quad\times\exp\left[-i\sum_{\mu=1}^{p}\sum_{\gamma=1}^{n}\left(\eta_{\gamma}^{\mu}+\frac{\kappa}{\sigma}\right)\omega_{\gamma}^{\mu}-i\sum_{\mu=1}^{p}\sum_{\gamma=1}^{n}y_{\gamma}^{\mu}\Phi\left(\eta_{\gamma}^{\mu}\right)\right]\ , \end{aligned} \tag{42}$$

with

$$C\left(\mathbf{\Omega},\boldsymbol{W}\right) := \prod_{\mu=1}^{p}\left\langle \exp\left(\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\omega_{\gamma}^{\mu}\frac{\boldsymbol{x}^{\mu}\cdot\boldsymbol{w}_{\gamma}}{\sigma\sqrt{N}}\right)\right\rangle_{\boldsymbol{x},\xi}\ , \tag{43}$$

where $\mathbf{\Omega}=\{\omega_{\gamma}^{\mu}\}_{\mu,\gamma}$, $\boldsymbol{H}=\{\eta_{\gamma}^{\mu}\}_{\mu,\gamma}$ and

$$\mathrm{d}^{nP}\mathbf{\Omega}\ \mathrm{d}^{nP}\boldsymbol{H} = \prod_{\mu=1}^{P}\prod_{\gamma=1}^{n}\mathrm{d}\omega_{\gamma}^{\mu}\ \mathrm{d}\eta_{\gamma}^{\mu}\ . \tag{44}$$

Since the patterns have components $x_{j}^{\mu}=\pm1$ which are statistically independent and identically distributed, using (19), one computes the mean over the patterns in $C(\mathbf{\Omega},\boldsymbol{W})$ as follows:

$$
\begin{aligned}
C(\boldsymbol{\Omega},\boldsymbol{W}) &= \prod_{\mu=1}^{p}\prod_{j=1}^{N}\left\langle \exp\left(\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\omega_{\gamma}^{\mu}\frac{w_{\gamma,j}x_{j}^{\mu}}{\sigma\sqrt{N}}\right)\right\rangle_{x_{j}^{\mu},\xi^{\mu}} \\
&= \prod_{\mu=1}^{p}\prod_{j=1}^{N}\frac{1}{2}\left\langle \exp\left(\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)+\exp\left(-\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)\right. \\
&\quad \left.+m_{\mathrm{in}}\left(\exp\left(\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)-\exp\left(-\mathrm{i}\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)\right)\right\rangle_{\xi^{\mu}} \\
&= \prod_{\mu=1}^{p}\prod_{j=1}^{N}\left\langle \cos\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)+\mathrm{i}m_{\mathrm{in}}\sin\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)\right\rangle_{\xi^{\mu}} \\
&= \prod_{\mu=1}^{p}\prod_{j=1}^{N}\left\langle \exp\left(\log\left[\cos\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)+\mathrm{i}m_{\mathrm{in}}\sin\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)\right]\right)\right\rangle_{\xi^{\mu}}.
\end{aligned}
$$

When $N\gg 1$, the leading order expansion in $1/N$ of each factor in the product over the μ index reads

$$
\begin{aligned}
&\prod_{j=1}^{N}\left\langle \exp\left(\log\left[1+\mathrm{i}m_{\mathrm{in}}\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)-\frac{1}{2}\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)^{2}\right]\right)\right\rangle_{\xi^{\mu}} \\
&\quad = \prod_{j=1}^{N}\left\langle \exp\left(\log\left(1+\mathrm{i}m_{\mathrm{in}}\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)-\frac{1}{2\sigma^{2}N}\sum_{\beta=1}^{n}w_{\beta,j}\omega_{\beta}^{\mu}\sum_{\gamma=1}^{n}w_{\gamma,j}\omega_{\gamma}^{\mu}\right)\right)\right\rangle_{\xi^{\mu}} \\
&\quad = \prod_{j=1}^{N}\left\langle \exp\left(\mathrm{i}m_{\mathrm{in}}\left(\xi^{\mu}\sum_{\gamma=1}^{n}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}\right)-\frac{1-m_{\mathrm{in}}^{2}}{2\sigma^{2}N}\sum_{\beta=1}^{n}w_{\beta,j}\omega_{\beta}^{\mu}\sum_{\gamma=1}^{n}w_{\gamma,j}\omega_{\gamma}^{\mu}\right)\right\rangle_{\xi^{\mu}} \\
&\quad = \left\langle \exp\left(\mathrm{i}m_{\mathrm{in}}\xi^{\mu}\sum_{\gamma=1}^{n}\sum_{j=1}^{N}\frac{w_{\gamma,j}\omega_{\gamma}^{\mu}}{\sigma\sqrt{N}}-\frac{1-m_{\mathrm{in}}^{2}}{2\sigma^{2}N}\sum_{\gamma,\beta=1}^{n}\omega_{\gamma}^{\mu}\omega_{\beta}^{\mu}\sum_{j=1}^{N}w_{\gamma,j}w_{\beta,j}\right)\right\rangle_{\xi^{\mu}},
\end{aligned}
\tag{45}
$$

the remaining terms vanishing as $O\left(1/\sqrt{N}\right)$. Using $\|\boldsymbol{w}_{\gamma}\|^{2}=N$ and setting

$$
M_{\gamma}\equiv\frac{1}{\sqrt{N}}\sum_{j=1}^{N}w_{\gamma,j}\,,\quad q_{\gamma\beta}\equiv\frac{1}{N}\sum_{j=1}^{N}w_{\gamma,j}w_{\beta,j}
\tag{46}
$$

for $\gamma,\beta=1,\ldots,n$, $\gamma>\beta$, we finally focus upon

$$
C(\boldsymbol{\Omega},\boldsymbol{W})\simeq\prod_{\mu=1}^{p}\left\langle \exp\left(\frac{\mathrm{i}m_{\mathrm{in}}\xi^{\mu}}{\sigma}\sum_{\gamma=1}^{n}M_{\gamma}\omega_{\gamma}^{\mu}-\frac{1-m_{\mathrm{in}}^{2}}{2\sigma^{2}}\left[\sum_{\gamma=1}^{n}\left(\omega_{\gamma}^{\mu}\right)^{2}+2\sum_{\gamma>\beta=1}^{n}\omega_{\beta}^{\mu}\omega_{\gamma}^{\mu}q_{\gamma\beta}\right]\right)\right\rangle_{\xi^{\mu}},
\tag{47}
$$

neglecting corrections of order $O\left(1/\sqrt{N}\right)$. Therefore, to leading order in $1/N$, equation (42) becomes

$$B(\boldsymbol{Y},\boldsymbol{W}) \simeq \frac{1}{(2\pi)^{nP}} \int_{\mathbb{R}^{2np}} \mathrm{d}^{nP}\boldsymbol{\Omega}\, \mathrm{d}^{nP}\boldsymbol{H} \exp\left[-i\sum_{\mu=1}^{p}\sum_{\gamma=1}^{n}\left(\eta_\gamma^\mu + \frac{\kappa}{\sigma}\right)\omega_\gamma^\mu - i\sum_{\mu=1}^{p}\sum_{\gamma=1}^{n} y_\gamma^\mu \Phi(\eta_\gamma^\mu)\right] \times \prod_{\mu=1}^{p}\left\langle \exp\left(\frac{\mathrm{i}m_{\mathrm{in}}\xi^\mu}{\sigma}\sum_{\gamma=1}^{n} M_\gamma \omega_\gamma^\mu - \frac{1-m_{\mathrm{in}}^2}{2\sigma^2}\left[\sum_{\gamma=1}^{n}(\omega_\gamma^\mu)^2 + 2\sum_{\gamma>\beta=1}^{n}\omega_\beta^\mu\omega_\gamma^\mu q_{\gamma\beta}\right]\right)\right\rangle_{\xi^\mu}\Bigg). \tag{48}$$

Since the integrals for different μ's are the same and the targets ξ^μ are statistically independent and equally distributed, equation (38) reduces to

$$A(\boldsymbol{W}) = \left(\frac{1}{(2\pi)^{2n}}\int_{[1-\epsilon,+\infty)^n} \mathrm{d}^n\boldsymbol{\lambda}, \int_{\mathbb{R}^{3n}} \mathrm{d}^n\boldsymbol{y}\, \mathrm{d}^n\boldsymbol{\eta}\, \mathrm{d}^n\boldsymbol{\omega} \left\langle \mathrm{e}^{K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{Q},\boldsymbol{M})}\right\rangle_\xi\right)^p, \tag{49}$$

where $\boldsymbol{\eta} = \{\eta_\gamma\}$, $\boldsymbol{y} = \{y_\gamma\}$, $\boldsymbol{\lambda} = \{\lambda_\gamma\}$, $\boldsymbol{\omega} = \{\omega_\gamma\}$, $\boldsymbol{Q} = \{q_{\alpha\beta}\}$, $\boldsymbol{M} = \{M_\gamma\}$, for $\alpha,\beta,\gamma = 1,\ldots,n$,

$$\mathrm{d}^n\boldsymbol{\lambda} \equiv \prod_{\gamma=1}^{n}\mathrm{d}\lambda_\gamma\ ,\ \mathrm{d}^n\boldsymbol{y} \equiv \prod_{\gamma=1}^{n}\mathrm{d}y_\gamma\ ,\ \mathrm{d}^n\boldsymbol{\eta} \equiv \prod_{\gamma=1}^{n}\mathrm{d}\eta_\gamma\ ,\ \mathrm{d}^n\boldsymbol{\omega} \equiv \prod_{\gamma=1}^{n}\mathrm{d}\omega_\gamma$$

and

$$K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{Q},\boldsymbol{M}) \equiv \mathrm{i}\sum_{\gamma=1}^{n}\left[y_\gamma\lambda_\gamma - y_\gamma\Phi(\eta_\gamma) - \left(\frac{\kappa - \xi\, m_{\mathrm{in}}M_\gamma}{\sigma} + \eta_\gamma\right)\omega_\gamma\right] \tag{50}$$

$$- \frac{1-m_{\mathrm{in}}^2}{2\sigma^2}\left(\sum_{\gamma=1}^{n}(\omega_\gamma)^2 + 2\sum_{\gamma\neq\beta=1}^{n}\omega_\beta\omega_\gamma q_{\gamma\beta}\right). \tag{51}$$

Writing the various Dirac deltas appearing in (34) as

$$\delta\left(||\boldsymbol{w}_\gamma||^2 - N\right) = \int_{-\infty}^{+\infty}\frac{\mathrm{d}E_\gamma}{4\pi}\ \mathrm{e}^{-\mathrm{i}\frac{E_\gamma}{2}N + \mathrm{i}\frac{E_\gamma}{2}\sum_{j=1}^{N} w_{\gamma,j}^2}, \tag{52}$$

as well as those Dirac deltas whose integration over $q_{\gamma\beta}$, respectively M_γ, implements the constraints (46) within (49), as

$$\delta\left(q_{\gamma\beta} - \frac{1}{N}\sum_{j=1}^{N} w_{\gamma,j}w_{\beta,j}\right) = N\int_{-\infty}^{+\infty}\frac{\mathrm{d}F_{\gamma\beta}}{2\pi}\ \exp\left(-\mathrm{i}Nq_{\gamma\beta}F_{\alpha\beta} + \mathrm{i}F_{\gamma\beta}\sum_{j=1}^{N} w_{\gamma,j}w_{\beta,j}\right), \tag{53}$$

$$\delta\left(M_\gamma - \frac{1}{\sqrt{N}}\sum_{j=1}^{N} w_{\gamma,j}\right) = \sqrt{N}\int_{-\infty}^{+\infty}\frac{\mathrm{d}L_\gamma}{2\pi}\ \exp\left(-\mathrm{i}\sqrt{N}L_\gamma M_\gamma + \mathrm{i}L_\gamma\sum_{j=1}^{N} w_{\gamma,j}\right), \tag{54}$$

and inserting them into (34), one finally arrives at the following explicit integral expression

$$\begin{aligned}\left\langle [V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\right\rangle = \frac{(N)^{n^2/2}}{Z_N^n}\int_{\mathbb{R}^{nN+n^2+2n}}\prod_{\gamma=1}^n\prod_{j=1}^N\prod_{\gamma<\beta=1}^n \mathrm{d}w_{\gamma,j}\,\frac{\mathrm{d}E_\gamma}{4\pi}\,\mathrm{d}M_\gamma\,\mathrm{d}q_{\gamma\beta}\,\frac{\mathrm{d}F_{\gamma\beta}}{2\pi}\,\frac{\mathrm{d}L_\gamma}{2\pi}\\ \times\exp\left(-i\frac{E_\gamma}{2}N+\mathrm{i}\frac{E_\gamma}{2}\sum_{j=1}^N w_{\gamma,j}^2\right)\exp\left(-\mathrm{i}Nq_{\gamma\beta}F_{\gamma\beta}+\mathrm{i}F_{\gamma\beta}\sum_{j=1}^N w_{\gamma,j}w_{\beta,j}\right)\\ \times\exp\left(-\mathrm{i}\sqrt{N}M_\gamma L_\gamma+\mathrm{i}L_\gamma\sum_{j=1}^N w_{\gamma,j}\right)\\ \times\left[\int_{1-\epsilon}^{+\infty}\frac{1}{(2\pi)^{2n}}\int_{[1-\epsilon,+\infty)^n}\mathrm{d}^n\boldsymbol{\lambda}\int_{\mathbb{R}^{3n}}\mathrm{d}^n\boldsymbol{y}\,\mathrm{d}^n\boldsymbol{\eta}\,\mathrm{d}^n\boldsymbol{\omega}\left\langle \mathrm{e}^{K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{Q},\boldsymbol{M})}\right\rangle_\xi\right]^p .\end{aligned}$$

Regrouping together the integrals over $w_{\gamma,j}$ with different j and same γ, one writes

$$\begin{aligned}\left\langle [V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\right\rangle = \frac{N^{n^2/2}}{Z_N^n}\int_{\mathbb{R}^{n^2+2n}}\prod_{\gamma=1}^n\prod_{\gamma<\beta=1}^n \frac{\mathrm{d}E_\gamma}{4\pi}\,\mathrm{d}M_\gamma\,\mathrm{d}q_{\gamma\beta}\,\frac{\mathrm{d}F_{\gamma\beta}}{2\pi}\,\frac{\mathrm{d}L_\gamma}{2\pi}\\ \times\left[\exp\left(-\frac{\mathrm{i}}{2}\sum_{\gamma=1}^n E_\gamma-\mathrm{i}\sum_{\gamma<\beta=1}^n F_{\gamma\beta}q_{\gamma,\beta}-\frac{\mathrm{i}}{\sqrt{N}}\sum_{\gamma=1}^n L_\gamma M_\gamma\right)\right]^N\\ \times\left[\int_{\mathbb{R}^n}\prod_{\gamma=1}^n \mathrm{d}w_\gamma\exp\left(\mathrm{i}\sum_{\gamma<\beta=1}^n F_{\gamma\beta}w_\gamma w_\beta+\frac{\mathrm{i}}{2}\sum_{\gamma=1}^n E_\gamma\,w_\gamma^2+\mathrm{i}\sum_{\gamma=1}^n L_\gamma\,w_\gamma\right)\right]^N\\ \times\left[\frac{1}{(2\pi)^{2n}}\int_{[1-\epsilon,+\infty)^n}\mathrm{d}^n\boldsymbol{\lambda}\int_{\mathbb{R}^{3n}}\mathrm{d}^n\boldsymbol{y}\,\mathrm{d}^n\boldsymbol{\eta}\,\mathrm{d}^n\boldsymbol{\omega}\left\langle \mathrm{e}^{K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{Q},\boldsymbol{M})}\right\rangle_\xi\right]^p .\end{aligned}$$

In the large N limit, the contribution $\frac{1}{\sqrt{N}}\sum_{\gamma=1}^n L_\gamma M_\gamma$ can be neglected; using (13), at leading order in $N\gg 1$, one gets

$$\begin{aligned}\left\langle [V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\right\rangle = \frac{N^{n^2/2}}{Z_N^n}\int_{\mathbb{R}^{3n}}\left(\prod_{\gamma=1}^n\frac{\mathrm{d}E_\gamma}{4\pi}\,\mathrm{d}M_\gamma\,\frac{\mathrm{d}L_\gamma}{2\pi}\right)\int_{\mathbb{R}^{n^2-n}}\left(\prod_{\gamma>\beta=1}^n \mathrm{d}q_{\gamma\beta}\,\frac{\mathrm{d}F_{\gamma\beta}}{2\pi}\right)\\ \times\exp\left(NG(\mathbf{E},\mathbf{F},\mathbf{L},\mathbf{M},\mathbf{Q})\right) ,\end{aligned}\tag{55}$$

where $\mathbf{E}=\{E_\gamma\}_{\gamma=1}^n$, $\mathbf{F}=\{F_{\gamma\beta}\}_{\gamma>\beta=1}^n$, $\mathbf{L}=\{L_\gamma\}_{\gamma=1}^n$, and

$$G(\mathbf{E},\mathbf{F},\mathbf{L},\mathbf{M},\mathbf{Q})\equiv\frac{p}{N}G_1(\mathbf{M},\mathbf{Q})+G_2(\mathbf{E},\mathbf{F},\mathbf{L})+G_3(\mathbf{E},\mathbf{F},\mathbf{Q}) ,\tag{56}$$

with

$$G_1(\mathbf{M},\mathbf{Q})=\log\left[\int_{[1-\epsilon,+\infty)^n}\left(\prod_{\gamma=1}^n\frac{\mathrm{d}\lambda_\gamma}{2\pi}\right)\int\left(\prod_{\gamma=1}^n\frac{\mathrm{d}y_\gamma\,\mathrm{d}\eta_\gamma\,\mathrm{d}\omega_\gamma}{2\pi}\right)\left\langle \mathrm{e}^{K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\boldsymbol{y},\boldsymbol{\omega},\boldsymbol{Q},\boldsymbol{M})}\right\rangle_\xi\right] ,\tag{57}$$

$$G_2(\mathbf{E},\mathbf{F},\mathbf{L})=\log\left[\int_{\mathbb{R}^n}\prod_{\gamma=1}^n \mathrm{d}w_\gamma\,\mathrm{e}^{\mathrm{i}\sum_{\gamma<\beta}F_{\gamma\beta}w_\gamma w_\beta+\mathrm{i}\sum_{\gamma=1}^n\frac{E_\gamma}{2}w_\gamma^2+\mathrm{i}\sum_{\gamma=1}^n w_\gamma L_\gamma}\right] ,\tag{58}$$

$$G_3(\mathbf{E},\mathbf{F},\mathbf{Q})=-\frac{\mathrm{i}}{2}\sum_{\gamma=1}^n E_\gamma-\mathrm{i}\sum_{\gamma>\beta}F_{\gamma\beta}q_{\gamma,\beta} .\tag{59}$$

A.1. Saddle-point approximation

When N is large, the behavior of $\langle[V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\rangle$ can be obtained using the saddle-point approximation, as follows. Setting $\mathbf{z}\equiv(\mathbf{E},\mathbf{F},\mathbf{L},\mathbf{M},\mathbf{Q})$ and considering it as a vector in $\mathbb{C}^{n^2+2n}$, one expands $G(\mathbf{z})\equiv G(\mathbf{E},\mathbf{F},\mathbf{L},\mathbf{M},\mathbf{Q})$ around the stationary point $\mathbf{z}^0=(\mathbf{E}^0,\mathbf{F}^0,\mathbf{L}^0,\mathbf{M}^0,\mathbf{Q}^0)$ such that $\dfrac{\partial G(\mathbf{z}^0)}{\partial z_k}=0$:

$$G(\mathbf{z})\simeq G\left(\mathbf{z}^0\right)+\frac{1}{2}\sum_{j,k=1}^{n^2+2n}\frac{\partial^2 G\left(\mathbf{z}^0\right)}{\partial z_j\partial z_k}\left(z_j-z_j^0\right)\left(z_k-z_k^0\right)\ .$$

Let $G''(\mathbf{z}^0)\equiv\left[\frac{\partial^2 G(\mathbf{z}^0)}{\partial z_j\partial z_k}\right]$ be the Hessian $(n^2+2n)\times(n^2+2n)$ matrix at the stationary point. If such a matrix is negative semi-definite, by suitably deforming the integration paths into the complex domain, one can perform n^2+2n Gaussian integrations by rescaling with $\sqrt{N}$ the corresponding integration variables and approximate:

$$\langle[V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\rangle\simeq\frac{1}{Z_N^n}\frac{1}{N^n}\left(\sqrt{\frac{2\pi}{|\det G''(\mathbf{z}^0)|}}\right)^{n^2+2n}\mathrm{e}^{NG(\mathbf{z}^0)}\ . \tag{60}$$

From equations (22) and (20), one needs to control the behavior of the ratio $\dfrac{1}{nN}\log\langle[(V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\rangle$ for $n\to0^+$ and $N\to+\infty$. From equation (60) and using equation (14) one gets:

$$\frac{1}{nN}\log\langle[V_N^Q(\boldsymbol{x}^\mu,\xi^\mu,\kappa,\sigma,\epsilon)]^n\rangle=\frac{1}{n}G\left(\mathbf{z}^0\right)-\frac{1}{2}\log(2\pi\,\mathrm{e})+O\left(\frac{\log N}{N}\right)\ . \tag{61}$$

A.2. RS ansatz

Making use of the replica-symmetric ansatz which states that the stationary point $\mathbf{z}_0$ is replica-insensitive, one seeks it setting

$$q_{\gamma\beta}=q,\quad F_{\gamma\beta}=F,\quad E_\gamma=E\quad M_\gamma=M\quad\text{and}\quad L_\gamma=L, \tag{62}$$

so that (50) becomes

$$\begin{aligned}K_\xi(\boldsymbol{\eta},\boldsymbol{\lambda},\mathbf{y},\boldsymbol{\omega},q,\boldsymbol{M})=i\sum_{\gamma=1}^{n}\left[y_\gamma\lambda_\gamma-y_\gamma\Phi(\eta_\gamma)-\left(\frac{\kappa-\xi pM}{\sigma}+\eta_\gamma\right)\omega_\gamma\right]\\-\frac{(1-q)\left(1-m_{\mathrm{in}}^2\right)}{2\sigma^2}\sum_{\gamma=1}^{n}\omega_\gamma^2-\frac{q\left(1-m_{\mathrm{in}}^2\right)}{2\sigma^2}\left(\sum_{\gamma=1}^{n}\omega_\gamma\right)^2\ .\end{aligned} \tag{63}$$

Notice that the argument of the logarithm in (57) is the average of the following quantity

$$\Delta(n) \equiv \int_{[1-\epsilon,+\infty)^n} \left(\prod_{\gamma=1}^{n} \frac{\mathrm{d}\lambda_\gamma}{2\pi}\right) \int_{\mathbb{R}^{2n}} \left(\prod_{\gamma=1}^{n} \frac{\mathrm{d}\eta_\gamma\, \mathrm{d}y_\gamma}{2\pi}\right) \int_{\mathbb{R}^n} \prod_{\gamma=1}^{n} \mathrm{d}\omega_\gamma \, \exp\left(\mathrm{i}\frac{\xi m_{\mathrm{in}} M}{\sigma} \sum_{\gamma=1}^{n} \omega_\gamma\right)$$
$$\times \exp\left[\mathrm{i}\sum_{\gamma=1}^{n}\Big(y_\gamma\lambda_\gamma - y_\gamma\Phi(\eta_\gamma)\Big) - \mathrm{i}\sum_{\gamma=1}^{n}\Big(\frac{\kappa}{\sigma}+\eta_\gamma\Big)\omega_\gamma - \frac{(1-q)(1-m_{\mathrm{in}}^2)}{2\sigma^2}\sum_{\gamma=1}^{n}\omega_\gamma^2\right.$$
$$\left. - \frac{q(1-m_{\mathrm{in}}^2)}{2\sigma^2}\left(\sum_{\gamma=1}^{n}\omega_\gamma\right)^2\right]. \tag{64}$$

Such a quantity can then be manipulated as follows: using the Gaussian representation

$$\exp\left(-\frac{q\left(1-m_{\mathrm{in}}^2\right)}{2\sigma^2}\left(\sum_{\gamma=1}^{n}\omega_\gamma\right)^2\right) = \int_{\mathbb{R}} \mathrm{D}t \exp\left(-\frac{t^2}{2} - \mathrm{i}\,t\frac{\sqrt{q\left(1-m_{\mathrm{in}}^2\right)}}{\sigma}\sum_{\gamma=1}^{n}\omega_\gamma\right), \tag{65}$$

via straightforward Gaussian integration over the variables ω_γ, one writes

$$\Delta(n) = \int_{[1-\epsilon,+\infty)^2} \left(\prod_{\gamma=1}^{n} \frac{\mathrm{d}\lambda_\gamma}{2\pi}\right) \int_{\mathbb{R}^{2n}} \left(\prod_{\gamma} \frac{\mathrm{d}\eta_\gamma\, \mathrm{d}y_\gamma}{2\pi}\right) \prod_{\gamma=1}^{n} \mathrm{e}^{\mathrm{i}(y_\gamma\lambda_\gamma - y_\gamma\Phi(\eta_\gamma))}$$
$$\times \int_{\mathbb{R}} \mathrm{D}x \prod_{\gamma=1}^{n} \sqrt{\frac{2\pi\sigma^2}{(1-q)(1-m_{\mathrm{in}}^2)}} \exp\left(-\frac{\left(\kappa - \xi m_{\mathrm{in}} M + \sigma\eta_\gamma + x\sqrt{q(1-m_{\mathrm{in}}^2)}\right)^2}{2(1-q)(1-m_{\mathrm{in}}^2)}\right).$$

Notice that the argument of the logarithm in (57) amounts to $\langle\Delta(n)\rangle_\xi$. Then, one sees that $\Delta(n)$ consists in n independent integrals with respect to $\mathrm{d}\eta_\gamma$, $\mathrm{d}\lambda_\gamma$ and $\mathrm{d}y_\gamma$:

$$\Delta\,(n) = \int_{\mathbb{R}} \mathrm{D}x \left[\int_{\mathbb{R}} \frac{\sigma\ \mathrm{d}\eta}{\sqrt{2\pi\,(1-q)\left(1-m_{\mathrm{in}}^2\right)}} \int_{1-\epsilon}^{+\infty} \mathrm{d}\lambda \int_{\mathbb{R}} \frac{\mathrm{d}y}{2\pi}\, \mathrm{e}^{-\,\mathrm{i}y(\lambda-\Phi(\eta/\sigma))}\right.$$
$$\left.\times \exp\left(-\frac{\left(\kappa - \xi\, m_{\mathrm{in}} M + \sigma\eta + x\sqrt{q\left(1-m_{\mathrm{in}}^2\right)}\right)^2}{2\,(1-q)\left(1-m_{\mathrm{in}}^2\right)}\right)\right]^n$$
$$= \int_{\mathbb{R}} \mathrm{D}x \left[\int_{\mathbb{R}} \frac{\sigma\ \mathrm{d}\eta}{\sqrt{2\pi\,(1-q)\left(1-m_{\mathrm{in}}^2\right)}} \Theta\left[\Phi\left(\frac{\eta}{\sigma}\right) - 1 + \epsilon\right]\right.$$
$$\left.\times \exp\left(-\frac{\left(\kappa - \xi\, m_{\mathrm{in}} M + \sigma\eta + x\sqrt{q\left(1-m_{\mathrm{in}}^2\right)}\right)^2}{2\,(1-q)\left(1-m_{\mathrm{in}}^2\right)}\right)\right]^n$$

Due to the monotonicity of the error function, the function $\Phi(x)$ in (29) is invertible and one has

$$\Theta\left[\Phi\left(\frac{\eta}{\sigma}\right) - 1 + \epsilon\right] = \Theta\left[\eta - \sigma\Phi^{-1}\,(1-\epsilon)\right].$$

Then, by changing the integration variable η into $\lambda = \sigma\eta + \kappa$ and using again (29), one gets

$$\Delta(n) = \int_{\mathbb{R}} \mathrm{D}x \left[\int_{\kappa+\sigma\Phi^{-1}(1-\epsilon)}^{+\infty} \frac{\mathrm{d}\lambda}{\sqrt{2\pi\,(1-q)\,(1-m_{\rm in}^2)}} \exp\left(-\frac{\left(\lambda - \xi\, m_{\rm in} M + x\sqrt{q\,(1-m_{\rm in}^2)}\right)^2}{2\,(1-q)\,(1-m_{\rm in}^2)} \right) \right]^n$$
$$= \int_{\mathbb{R}} \mathrm{D}x \left[1 - \Phi\left(\frac{\tilde{\kappa} - \xi\, m_{\rm in} M + \sqrt{q\,(1-m_{\rm in}^2)}\,x}{\sqrt{(1-q)\,(1-m_{\rm in}^2)}} \right) \right]^n$$

where $\tilde{\kappa} = \kappa + \sigma\Phi^{-1}(1-\epsilon)$. Since the replica trick lets n vanish as a continuous quantity, we can use the first order approximation $z^n \simeq 1 + n\log z$, valid for $n \to 0$, and write:

$$\Delta(n) = 1 + n \int_{\mathbb{R}} \mathrm{D}x\, \log\left[1 - \Phi\left(\frac{\tilde{\kappa} - \xi\, m_{\rm in} M + \sqrt{q\,(1-m_{\rm in}^2)}\,x}{\sqrt{(1-q)\,(1-m_{\rm in}^2)}} \right) \right]. \tag{66}$$

Finally, we notice that the RS ansatz makes all matrix and vector entries equal. Then, averaging over the target parameter ξ according to the distribution in (19) yields the following leading behavior for the function $G_1(M,q)$ when $n \to 0^+$:

$$G_1(M,q) \simeq \log\left(1 + n\, g(M,q)\right) \simeq n\, g(M,q)\ , \tag{67}$$

where

$$g(M,q) \equiv \frac{1+m_{\rm out}}{2} \int \mathrm{D}x \log\left[1 - \Phi\left(\frac{x\sqrt{q} - a_-(M)}{\sqrt{1-q}} \right) \right]$$
$$+ \frac{1-m_{\rm out}}{2} \int \mathrm{D}x \log\left[1 - \Phi\left(\frac{x\sqrt{q} - a_+(M)}{\sqrt{1-q}} \right) \right]\ , \tag{68}$$

with

$$a_\pm(M) \equiv -\frac{\tilde{\kappa} \pm m_{\rm in} M}{\sqrt{1-m_{\rm in}^2}}\ . \tag{69}$$

Because of the RS ansatz, $G_2(E,F,L)$ in (58) can be recast as

$$G_2(E,F,L) = \log\left[\int_{\mathbb{R}^n} \mathrm{d}^n \boldsymbol{w} \exp\left(\mathrm{i} \sum_{\gamma<\beta=1}^{n} F\, w_\gamma w_\beta + \mathrm{i}\,\frac{E}{2} \sum_{\gamma=1}^{n} w_\gamma^2 + \mathrm{i}\, L \sum_{\gamma=1}^{n} w_\gamma \right) \right]\ , \tag{70}$$

where now $\boldsymbol{w} = (w_1, \ldots, w_n)$. Then, rewriting

$$\sum_{\gamma<\beta=1}^{n} w_\gamma w_\beta = \frac{1}{2} \left(\left(\sum_{\gamma=1}^{n} w_\gamma \right)^2 - \sum_{\gamma=1}^{n} w_\gamma^2 \right)\ ,$$

one finds

$$\mathrm{i}F \sum_{\gamma<\beta=1}^{n} w_\gamma\, w_\beta + \mathrm{i}\,\frac{E}{2} \sum_{\gamma=1}^{n} w_\gamma^2 = \mathrm{i}\,\frac{F}{2} \left(\sum_{\gamma=1}^{n} w_\gamma \right)^2 - \mathrm{i}\,\frac{F-E}{2} \sum_{\gamma=1}^{n} w_\gamma^2\ ,$$

and, using (65),

$$\exp\left(\mathrm{i}\sum_{\gamma<\beta=1}^{n} F w_\gamma w_\beta + \mathrm{i}\,\frac{E}{2}\sum_{\gamma=1}^{n} w_\gamma^2\right) = \int_{\mathbb{R}} \frac{\mathrm{d}x}{\sqrt{2\pi}}\, \mathrm{e}^{-x^2/2 - x\sqrt{\mathrm{i}F}\sum_\gamma w_\gamma}\, \mathrm{e}^{-\,\mathrm{i}\frac{F-E}{2}\sum_{\gamma=1}^{n} w_\gamma^2}\,. \tag{71}$$

Then, after straightforward Gaussian manipulations and integration,

$$\begin{aligned} G_2(F,E,L) &= \log\left(\int_{\mathbb{R}} \mathrm{D}x \left(\int_{\mathbb{R}} \mathrm{d}w\, \exp\left(-x\sqrt{\mathrm{i}F}w - \mathrm{i}\frac{F-E}{2}w^2 + \mathrm{i}wL\right)\right)^n\right) \\ &= \frac{n}{2}\log\frac{2\pi}{\mathrm{i}(F-E)} + \log\left(\int_{\mathbb{R}} \mathrm{D}x\, \exp\left(-\mathrm{i}n\,\frac{\left(x\sqrt{\mathrm{i}F} - \mathrm{i}L\right)^2}{2(F-E)}\right)\right) \\ &= \frac{n}{2}\log\frac{2\pi}{\mathrm{i}(F-E)} - \frac{1}{2}\log\left(1 - n\,\frac{F}{F-E}\right) - \mathrm{i}\,\frac{FL^2n^2}{2(F-E)\left((n-1)F+E\right)} + \mathrm{i}\,\frac{nL^2}{2(F-E)}\,. \end{aligned}$$

At leading order in $n \to 0^+$ one gets

$$G_2(F,E,L) \simeq \frac{n}{2}\left(\log\frac{2\pi}{\mathrm{i}(F-E)} + \frac{F}{F-E} + \mathrm{i}\frac{L^2}{F-E}\right)\,. \tag{72}$$

Finally, the replica symmetry ansatz yields the following leading order behavior for (59),

$$G_3(E,F,q) \simeq -\,\mathrm{i}n\frac{E}{2} - \mathrm{i}Fq\,\frac{n(n-1)}{2} \simeq -\,\mathrm{i}n\,\frac{1}{2}\,(E - Fq)\,. \tag{73}$$

Using (67), (72) and (73) into (56), we get, at the leading order in $n \to 0^+$:

$$G(E,F,L,M,q) \simeq n\alpha\, g(M,q) + \frac{n}{2}\left(\log\frac{2\pi}{\mathrm{i}(F-E)} + \frac{F}{F-E} + \mathrm{i}\frac{L^2}{F-E} + \mathrm{i}\,(qF-E)\right)\,. \tag{74}$$

The stationary point $\mathbf{z}^0$ is then found by asking that $\frac{\partial G(\mathbf{z})}{\partial E} = \frac{\partial G(\mathbf{z})}{\partial F} = \frac{\partial G(\mathbf{z})}{\partial L} = 0$ yielding the stationary point components:

$$F^0 = -\frac{\mathrm{i}q}{(1-q)^2}\,, \quad E^0 = \mathrm{i}\frac{1-2q}{(1-q)^2}\,, \quad L^0 = 0\,. \tag{75}$$

Then, from (74) one obtains

$$\frac{G\left(E^0,F^0,L^0,M,q\right)}{n} = \alpha\, g(M,q) + \frac{1}{2}\left(\log\left(2\pi\,(1-q)\right) + \frac{1}{(1-q)}\right)\,. \tag{76}$$

The sought after stationary point of $G(\mathbf{z})$ is finally obtained by setting

$$\partial_q G\left(E^0,F^0,L^0,M,q\right) = \partial_M G\left(E^0,F^0,L^0,M,q\right) = 0\,. \tag{77}$$

Using (68), one explicitly computes, with $\beta = q, M$,

$$\begin{aligned} \partial_\beta\, g(M,q) = &-\frac{1+m_{\mathrm{out}}}{2\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{D}x\, \frac{\exp\left(-\frac{\left(x\sqrt{q}-a_-(M)\right)^2}{2(1-q)}\right)}{1-\Phi\left(\frac{x\sqrt{q}-a_-(M)}{\sqrt{1-q}}\right)}\,\partial_\beta\,\frac{x\sqrt{q}-a_-(M)}{\sqrt{1-q}} \\ &-\frac{1-m_{\mathrm{out}}}{2\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{D}x\, \frac{\exp\left(-\frac{\left(x\sqrt{q}-a_+(M)\right)^2}{2(1-q)}\right)}{1-\Phi\left(\frac{x\sqrt{q}-a_+(M)}{\sqrt{1-q}}\right)}\,\partial_\beta\,\frac{x\sqrt{q}-a_+(M)}{\sqrt{1-q}}\,. \end{aligned} \tag{78}$$

As observed at the end of section 3, the optimal storage capacity is obtained when q in (46), namely the average overlap of random weights, tends to 1. The condition $q \to 1$ implies that the arguments of the functions Φ in (68) tend to $\pm\infty$. Then, one can use the asymptotic behavior of the error function,

$$\mathrm{erf}\,(x) \simeq \pm 1 - \frac{\mathrm{e}^{-x^2}}{\sqrt{\pi}}\left(\frac{1}{x} - \frac{1}{2x^3}\right) \text{ when } \quad x \to \pm\infty\,, \tag{79}$$

together with (29) to get that

$$\frac{1}{1-\Phi\left(\frac{x\sqrt{q}-a_\pm(M)}{\sqrt{1-q}}\right)} \simeq \begin{cases} \exp\left(\frac{\left(x\sqrt{q}-a_\pm(M)\right)^2}{2(1-q)}\right)\sqrt{2\pi}\frac{x\sqrt{q}-a_\pm(M)}{\sqrt{1-q}} & \ldots \quad x\sqrt{q}-a_\pm(M) > 0 \\ 1 & \ldots \quad x\sqrt{q}-a_\pm(M) \leqslant 0 \end{cases}. \tag{80}$$

Therefore, for $q \to 1^-$, the vanishing Gaussian terms in (78) can only be compensated for $x\sqrt{q} \geqslant a_\pm(M)$, so that

$$\begin{aligned} \partial_\beta\, g\,(M,q) \simeq & -\frac{1+m_{\mathrm{out}}}{2}\int_{a_-(M)/\sqrt{q}}^{+\infty} \mathrm{D}x\,\frac{x\sqrt{q}-a_-\,(M)}{\sqrt{1-q}}\,\partial_\beta\frac{x\sqrt{q}-a_-\,(M)}{\sqrt{1-q}} \\ & -\frac{1-m_{\mathrm{out}}}{2}\int_{a_+(M)/\sqrt{q}}^{+\infty} \mathrm{D}x\,\frac{x\sqrt{q}-a_+\,(M)}{\sqrt{1-q}}\,\partial_\beta\frac{x\sqrt{q}-a_+\,(M)}{\sqrt{1-q}}\,. \end{aligned} \tag{81}$$

Furthermore, using (69),

$$\begin{aligned} \partial_M\, g(M,q) \simeq & \frac{m_{\mathrm{in}}}{\sqrt{(1-m_{\mathrm{in}}^2)(1-q)}}\left(-\frac{1+m_{\mathrm{out}}}{2}\int_{a_-(M)/\sqrt{q}}^{+\infty}\mathrm{D}x\,(x\sqrt{q}-a_-(M))\right. \\ & \left. +\frac{1-m_{\mathrm{out}}}{2}\int_{a_+(M)/\sqrt{q}}^{+\infty}\mathrm{D}x\,(x\sqrt{q}-a_+(M))\right), \end{aligned} \tag{82}$$

which, together with (77) yields (30). On the other hand, (76) and (77) yields

$$\alpha\,\partial_q\, g\,(M,q) = -\frac{q}{2\,(1-q)^2}\,. \tag{83}$$

Thus, from

$$\partial_q\frac{x\sqrt{q}-a_\pm\,(M)}{\sqrt{1-q}} = \frac{x}{2\sqrt{q\,(1-q)}} - \frac{x\sqrt{q}-a_\pm\,(M)}{2\,(1-q)^{3/2}}\,,$$

and (81) one retrieves (26) as the leading term in $(1-q)^{-1}$ in the limit $q \to 1^-$.

Appendix B. Large output bias limit

In order to extract the leading order behavior of the quantum critical storage capacity when the target bias $m_{\mathrm{out}} \to 1^-$, we distinguish two possibilities. Firstly, in this appendix, we keep the input bias m_{in} fixed and let $m_{\mathrm{out}} \to 1^-$; then, in the next one we treat the case when $m_{\mathrm{in}} = m_{\mathrm{out}} = m \to 1^-$. Consider the equation

$$(1+m_{\mathrm{out}})\int_{a_-(M)}^{+\infty}\mathrm{D}x\,(x-a_-\,(M)) = (1-m_{\mathrm{out}})\int_{a_+(M)}^{+\infty}\mathrm{D}x\,(x-a_+\,(M))\,,$$

with $a_\pm(M)$ as in (27). If $m_{\rm out} \to 1^-$ and $m_{\rm in}$ is kept fixed, M must diverge to $+\infty$. Otherwise, the left-hand side cannot vanish, being the integral of a positive function. Then, in the following, we shall consider $m_{\rm out}$ close to 1 so that $a_+(M) < 0$, $a_-(M) > 0$ and

$$\mp a_\pm(M) = \frac{m_{\rm in} M}{\sqrt{1-m_{\rm in}^2}} \pm \frac{\tilde\kappa}{\sqrt{1-m_{\rm in}^2}} \gg 1 \,,$$

yielding

$$M \gg \frac{\sqrt{1-m_{\rm in}^2}}{m_{\rm in}} + \frac{\tilde\kappa}{m_{\rm in}}. \tag{84}$$

In the classical case $\tilde\kappa = \kappa$; moreover, the limit behavior of the critical classical storage capacity for $m_{\rm out} \to 1^+$,

$$\alpha_c^C \simeq -\frac{1}{(1-m_{\rm out})\log(1-m_{\rm out})} \,,$$

is obtained for $\kappa = 0$, namely for $M \gg \dfrac{\sqrt{1-m_{\rm in}^2}}{m_{\rm in}}$.

We proceed by recasting the two storage capacity defining equations (26) and (30) in terms of Gaussian and error functions:

$$\alpha_c^Q \left[\frac{1+m_{\rm out}}{4} \left(\left(1+a_-^2(M)\right) \left(1-{\rm erf}\left(\frac{a_-(M)}{\sqrt{2}}\right)\right) - a_-(M)\sqrt{\frac{2}{\pi}} {\rm e}^{-a_-^2(M)/2} \right) \right.$$
$$\left. + \frac{1-m_{\rm out}}{4} \left(\left(1+a_+^2(M)\right) \left(1-{\rm erf}\left(\frac{a_+(M)}{\sqrt{2}}\right)\right) - a_+(M)\sqrt{\frac{2}{\pi}} {\rm e}^{-a_+^2(M)/2} \right) \right] = 1 \,, \tag{85}$$

respectively

$$\frac{1-m_{\rm out}}{1+m_{\rm out}} = \frac{\sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_-^2(M)/2} - a_-(M)\left(1-{\rm erf}\left(\frac{a_-(M)}{\sqrt{2}}\right)\right)}{\sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_+^2(M)/2} - a_+(M)\left(1-{\rm erf}\left(\frac{a_+(M)}{\sqrt{2}}\right)\right)} \,. \tag{86}$$

Equation (86) can be satisfied in the limit $m_{\rm out} \to 1^-$ only if the right-hand side vanishes, which can happen only for values of M such that $a_\pm(M) \to \mp\infty$. For small but finite values of $1-m_{\rm out}$, the solution of (86) is obtained for $\mp a_\pm(M) \gg 1$, which implies

$$M \gg \frac{\sqrt{1-m_{\rm in}^2}}{m_{\rm in}} + \frac{\tilde\kappa}{m_{\rm in}}. \tag{87}$$

Using the asymptotic behavior in (79) with $\mp a_\pm(M) \gg 1$, one gets

$$1-{\rm erf}\left(\frac{a_-(M)}{\sqrt{2}}\right) \simeq \sqrt{\frac{2}{\pi}}{\rm e}^{-a_-^2(M)/2}\left(\frac{1}{a_-(M)} - \frac{1}{a_-^3(M)}\right) \,, \tag{88}$$

$$1-{\rm erf}\left(\frac{a_+(M)}{\sqrt{2}}\right) \simeq 2 + \sqrt{\frac{2}{\pi}}{\rm e}^{-a_+^2(M)/2}\left(\frac{1}{a_+(M)} - \frac{1}{a_+^3(M)}\right) \simeq 2 \,. \tag{89}$$

Hence, for $m_{\rm out} \to 1^-$, at leading order, (86) and (85) read

$$1-m_{\rm out} \simeq -\frac{{\rm e}^{-a_-^2(M)/2}}{\sqrt{\frac{\pi}{2}}a_-^2(M)\,a_+(M)} \,, \quad 1 \simeq \alpha_c^Q \frac{1-m_{\rm out}}{2} a_+^2(M) \,. \tag{90}$$

Since $m_{\rm out} \to 1^-$ implies $a_-(M) \simeq -a_+(M)$, the first asymptotic behavior in (90) yields

$$\log(1-m_{\rm out}) \simeq -\frac{a_-^2(M)}{2} - \log\left(\sqrt{\frac{\pi}{2}} a_-^2(M)\left(-a_+(M)\right)\right) \simeq -\frac{a_-^2(M)}{2} \simeq -\frac{a_+^2(M)}{2}\,. \tag{91}$$

from which the limit of strongly correlated targets, $m_{\rm out} \to 1^-$, is as in (32), for all pattern biases $0 \leqslant m_{\rm in} \leqslant 1$. However, it must be emphasized that, because of (84) the limit is reached with possibly quite different slopes depending on both the quantum parameter $\sigma\Phi^{-1}(1-\epsilon)$ and on the degree of independence of the input patterns $m_{\rm in}$.

Appendix C. Simultaneously large input and output bias

Setting $m_{\rm in} = m_{\rm out} = m$, the quantity M has to be chosen such that

$$\frac{1-m}{1+m} = \frac{\sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_-^2(M)/2} - a_-(M)\left(1-{\rm erf}\left(\frac{a_-(M)}{\sqrt{2}}\right)\right)}{\sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_+^2(M)/2} - a_+(M)\left(1-{\rm erf}\left(\frac{a_+(M)}{\sqrt{2}}\right)\right)} = \frac{I_-(M)}{I_+(M)}\,, \tag{92}$$

where we set

$$I_\pm(M) = \sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_\pm^2(M)/2} - a_\pm(M)\left(1-{\rm erf}\left(\frac{a_\pm(M)}{\sqrt{2}}\right)\right) \tag{93}$$

In the limit $m \to 1^-$, the quantities $a_\pm(M)$ in (27) behave as

$$a_\pm(M) \simeq -\frac{\tilde\kappa \pm M}{\sqrt{2(1-m)}} \tag{94}$$

and both diverge unless $M = \tilde\kappa$. Notice however, that inserting $M = \tilde\kappa$ into (92) and taking the limit, the equality (92) cannot be satisfied. Indeed, $a_-(\tilde\kappa) = 0$ and

$$a_+(\tilde\kappa) \simeq -\frac{2\tilde\kappa}{\sqrt{2(1-m)}} \Longrightarrow \frac{I_-(M)}{I_+(M)} \simeq \sqrt{1-m}\,,$$

whereas the right-hand side vanishes as $1-m$. Therefore, we need first to consider the asymptotic behavior of the right-hand side of (92) when $m \to 1^-$ and then properly choose M which will thus depend on m.

We distinguish the following cases

$$M > \tilde\kappa \Longrightarrow \begin{cases} a_-(M) > 0 \Rightarrow I_-(M) \simeq \sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_-^2(M)/2}\frac{1}{a_-^2(M)} \\ a_+(M) < 0 \Rightarrow I_+(M) \simeq -2a_+(M) \end{cases}, \tag{95}$$

$$-\tilde\kappa < M < \tilde\kappa \Longrightarrow \begin{cases} a_-(M) < 0 \Rightarrow I_-(M) \simeq -2a_-(M) \\ a_+(M) < 0 \Rightarrow I_+(M) \simeq -2a_+(M) \end{cases}, \tag{96}$$

$$M < -\tilde\kappa \Longrightarrow \begin{cases} a_-(M) < 0 \Rightarrow I_-(M) \simeq -2a_-(M) \\ a_+(M) > 0 \Rightarrow I_+(M) \simeq \sqrt{\frac{2}{\pi}}\,{\rm e}^{-a_+^2(M)/2}\frac{1}{a_+^2(M)} \end{cases}, \tag{97}$$

where we made explicit the asymptotic behaviors (88) and (89) when $m \to 1^-$. Then, one finds

$$M > \tilde\kappa \Longrightarrow \frac{I_-(M)}{I_+(M)} \simeq \frac{2}{\sqrt{\pi}}\frac{(1-m)^{3/2}}{(\tilde\kappa - M)^2(\tilde\kappa + M)}\exp\left(-\frac{(\tilde\kappa - M)^2}{4(1-m)}\right), \tag{98}$$

$$-\tilde\kappa < M < \tilde\kappa \Longrightarrow \frac{I_-(M)}{I_+(M)} \simeq \frac{\tilde\kappa - M}{\tilde\kappa + M}\,, \tag{99}$$

$$M < -\tilde{\kappa} \Longrightarrow \frac{I_-(M)}{I_+(M)} \simeq \frac{\sqrt{\pi}}{2} \frac{(\tilde{\kappa}-M)^2(\tilde{\kappa}+M)}{(1-m)^{3/2}} \exp\left(+\frac{(\tilde{\kappa}-M)^2}{4(1-m)}\right) . \tag{100}$$

Since (92) asks for $\frac{I_-(M)}{I_+(M)} \simeq \frac{1-m}{2}$, together with $a_-(M) \to +\infty$ and $a_+(M) \to -\infty$, the only behavior compatible with these request is the one in (98). Indeed, the third one is clearly to be excluded, while the second one asks for

$$M \simeq \tilde{\kappa} m \Longrightarrow \begin{cases} a_-(M) \simeq -\tilde{\kappa}\sqrt{\frac{1-m}{2}} \\ a_+(M) \simeq -\tilde{\kappa}\sqrt{\frac{2}{1-m}} \end{cases} , \tag{101}$$

which requires $a_-(M) \to 0$ instead of $a_-(M) \to +\infty$. Then, the only possible remaining behavior yields

$$\frac{4}{\sqrt{\pi}} \frac{(1-m)^{1/2}}{(\tilde{\kappa}-M)^2(\tilde{\kappa}+M)} \exp\left(-\frac{(\tilde{\kappa}-M)^2}{4(1-m)}\right) \simeq 1 \tag{102}$$

or, in terms of $a_\pm(M)$,

$$\sqrt{\frac{2}{\pi}} \frac{1}{a_-(M)} \exp\left(-\frac{a_-^2(M)}{2}\right) \simeq a_-(M)\, a_+(M)\,(1-m) . \tag{103}$$

The functional dependence of M on m when $m \to 1^-$, implicitly determined by (102), cannot be given in terms of simple functions and can be obtained only numerically; however, (102) implies that

$$\lim_{m\to 1^-} M(m) = \tilde{\kappa} . \tag{104}$$

Finally, using (103) and (104), (88) and (89), together with (94), from (85), one gets

$$\begin{aligned} \frac{1}{\alpha_c^Q} &\simeq \left[-\frac{1}{\sqrt{2\pi}} \frac{1}{a_-^3(M)} \exp\left(-\frac{a_-^2(M)}{2}\right) + \frac{1-m}{2} a_+^2(M)\right] = \frac{1-m}{2} \frac{a_+(M)}{a_-(M)} \left(a_-(M)\, a_+(M) - 1\right) \\ &\simeq \frac{1-m}{2} \frac{\tilde{\kappa}+M}{\tilde{\kappa}-M} \frac{\tilde{\kappa}^2-M^2}{2(1-m)} = \frac{(\tilde{\kappa}+M)^2}{4} \simeq \tilde{\kappa}^2 . \end{aligned} \tag{105}$$

ORCID iDs

Fabio Benatti https://orcid.org/0000-0002-0712-2057
Giovanni Gramegna https://orcid.org/0000-0001-7532-1704
Stefano Mancini https://orcid.org/0000-0002-3797-3987
Gibbs Nwemadji https://orcid.org/0000-0001-7994-4810

References

[1] Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L 2019 Machine learning and the physical sciences *Rev. Mod. Phys.* **91** 045002
[2] Käming N *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 035037
[3] Schuld M, Sinayskiy I and Petruccione F 2014 *Quantum Inf. Process.* **13** 2567
[4] Dunjko V and Briegel H J 2018 *Rep. Prog. Phys.* **81** 074001
[5] Tacchino F, Macchiavello C, Gerace D and Bajoni D 2019 *npj Quantum Inf.* **5** 26
[6] Wittek P 2014 *Quantum Machine Learning* (Elsevier Science & Technology)

[7] Lewenstein M, Gratsea A, Riera-Campeny A, Aloy A, Kasper V and Sanpera A 2021 *Quantum Sci. Technol.* **6** 045002
[8] Huang H-Y, Broughton M, Mohseni M, Babbush R, Boixo S, Neven H and McClean J R 2021 *Nat. Commun.* **12** 2631
[9] Ban Y, Chen Xi, Torrontegui E, Solano E and Casanova J 2021 *Sci. Rep.* **11** 5783
[10] Rosenblatt F 1957 The perceptron: a perceiving and recognizing automaton *Technical Report Inc. Report* No. 85-460-1 (Cornell Aeronautical Laboratory)
[11] Cover T M 1965 *IEEE Trans. Electron. Comput.* **EC-14** 326–34
[12] Venkatesh S S and Psaltis D 1992 *IEEE Trans. Pattern Anal. Mach. Intell.* **14** 87–91
[13] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[14] Shcherbina M and Tirozzi B 2003 *Commun. Math. Phys.* **234** 383
[15] Talagrand M 2011 *Mean Field Models for Spin Glasses. Volume II* (*Ergebnisse der Mathematik und Ihrer Grenzgebiete 3* vol 55) (Springer)
[16] Engel A and Van den Broeck C 2001 *Statistical Mechanics of Learning* (Cambridge University Press)
[17] Artiaco C, Balducci F, Parisi G and Scardicchio A 2021 *Phys. Rev.* A **103** L040203
[18] Gratsea A, Kasper V and Lewenstein M 2021 Storage properties of a quantum perceptron (arXiv:2111.08414)
[19] Benatti F, Gramegna G and Mancini S 2022 *J. Phys. A: Math. Theor.* **55** 155301
[20] Benatti F, Mancini S and Mangini S 2019 *Int. J. Quantum Inf.* **17** 1941009
[21] Gardner E 1987 *Europhys. Lett.* **4** 481
[22] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[23] West A H L and Saad D 1997 *J. Phys. A: Math. Gen.* **30** 3471
[24] Beer K, Bondarenko D, Farrelly T, Osborne T J, Salzmann R, Scheiermann D and Wolf R 2020 *Nat. Commun.* **11** 808
[25] Torrontegui E and García-Ripoll J J 2019 *Europhys. Lett.* **125** 30004
[26] Cao Y, Guerreschi G G and Aspuru-Guzik A 2017 Quantum neuron: an elementary building block for machine learning on quantum computers (arXiv:1711.11240)
[27] Bangar S, Sunny L, Yeter-Aydeniz K and Siopsis G 2023 Experimentally realizable continuous-variable quantum neural networks (arXiv:2306.02525)
[28] Killoran N, Bromley T R, Arrazola J M, Schuld M, Quesada N and Lloyd S 2019 Continuous-variable quantum neural networks *Phys. Rev. Res.* **1** 033063
[29] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* vol 9 (World Scientific Publishing Company)
[30] Castellani T and Cavagna A 2005 *J. Stat. Mech.* P05012
[31] Malatesta E M 2023 High-dimensional manifold of solutions in neural networks: insights from statistical physics (arXiv:2309.09240)