

PAPER

Pattern capacity of a single quantum perceptron

To cite this article: Fabio Benatti *et al* 2022 *J. Phys. A: Math. Theor.* **55** 155301

View the [article online](#) for updates and enhancements.

You may also like

- [Unitary quantum perceptron as efficient universal approximator](#)
E. Torrontegui and J. J. Garcia-Ripoll
- [Capacity of the covariance perceptron](#)
David Dahmen, Matthieu Gilson and Moritz Helias
- [Quantitative analysis of different volatile organic compounds using an improved electronic nose](#)
Daqi Gao, Jiuming Ji, Jiayu Gong *et al.*



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Pattern capacity of a single quantum perceptron

Fabio Benatti^{1,2}, Giovanni Gramegna^{1,2,*} and Stefano Mancini^{3,4}

¹ Dipartimento di Fisica, Università 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 School of Science and Technology, University of Camerino, I-62032 Camerino, Italy

⁴ Istituto Nazionale di Fisica Nucleare, Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy

E-mail: giovanni.gramegna@units.it

Received 21 January 2022, revised 15 February 2022

Accepted for publication 25 February 2022

Published 29 March 2022



Abstract

Recent developments in quantum machine learning have seen the introduction of several models to generalize the classical perceptron to the quantum regime. The capabilities of these quantum models need to be determined precisely in order to establish if a quantum advantage is achievable. Here we use a statistical physics approach to compute the pattern capacity of a particular model of quantum perceptron realized by means of a continuous variable quantum system.

Keywords: quantum perceptron, perceptron capacity, quantum neural networks

(Some figures may appear in colour only in the online journal)

1. Introduction

Artificial neural networks have proven to be an extremely efficient computational model in specific tasks such as pattern recognition or image classification and have revolutionized the field of data analysis on classical computers [1–4]. Artificial neural networks are commonly implemented by classical algorithms on ordinary computers [5]; however, there has been recently a surge of interest towards the architecture of physical neural networks, i.e. implemented on a dedicated hardware [6, 7]. At the same time, the advent of quantum computation has shown

*Author to whom any correspondence should be addressed.

that purely quantum mechanical features such as coherence and entanglement allow to address hard computational tasks with an exponential improvement of the performances compared to classical computation [8]. The great success achieved in these two fields is currently motivating a rapid increase in the studies concerning quantum machine learning, exploring the interaction between machine learning and quantum computation, with the aim to understand whether the two fields can benefit from each other.

The simplest model of an artificial neuron traces back to the classical Rosenblatt's perceptron [9], which can be seen as the simplest learning algorithm for binary classification. Several possibilities can be considered to implement a perceptron by means of a quantum architecture [10–16]. In this context it is important to investigate the capability of a particular model of quantum perceptron to achieve quantum advantages with respect to its classical counterparts.

The main limitation affecting a single classical perceptron is due to the fact that the classification task is performed through a separation of patterns belonging to different classes through a hyperplane in the vector space containing the N features defining the pattern. In particular, it was soon pointed out that a simple perceptron is unable to compute the XOR function [17], since this corresponds to a classification problem where different classes cannot be separated with a line in the plane. However, it was found that when a large number of features is considered, i.e. for patterns in a vector space with a large dimension N , given p random labeled patterns it is extremely unlikely that a perceptron cannot classify them if $p < 2N$ for large N [18, 19]. On the contrary, the probability that p random labeled patterns can be classified by a simple perceptron becomes vanishingly small for $p > 2N$ in the large N regime. It became clear that the important parameter to characterize the performances of a perceptron is then the ratio $\alpha = p/N$, and led to identify the critical value of this ratio as the pattern capacity of a classical perceptron which thus is $\alpha_c = 2$.

In the seminal work [20], Gardner took a new approach to the pattern capacity of neural networks, adopting tools of statistical physics and in particular methods from the theory of disordered systems. The possibility to find a hyperplane separating randomly labeled patterns belongs in fact to the class of random constraint satisfaction problems [16, 21, 22], which can be investigated using the statistical theory of spin-glasses. In this approach, the parameter α gives rise to a phase transition in the high-dimensional case, and the pattern capacity is determined by the critical value α_c separating the SAT-phase, for $\alpha < \alpha_c$, where it is possible to satisfy all the constraints, i.e. classify all the patterns, from the UNSAT-phase, $\alpha > \alpha_c$, where the minimum number of unsatisfied constraints is larger than zero.

Here, we will follow Gardner's statistical approach to derive the pattern capacity of a particular model of quantum perceptron, based on a continuous variable multi-mode quantum system, which was introduced in [14]. We show that this model offers no quantum advantage over its classical counterpart, since its capacity is always smaller than that of its classical limit.

The article is structured as follows. In section 2 we introduce the classical perceptron and the definition of its pattern capacity. In section 3 we describe the model of quantum perceptron under investigation, and present the resulting pattern capacity. In section 5 we explain in detail the techniques employed, based on the same statistical approach used by Gardner to determine the pattern capacity of a classical perceptron. Finally, in section 4 we discuss the results herewith achieved comparing them with the pattern capacity also obtained by means of the statistical approach, but for a different model of quantum perceptron.

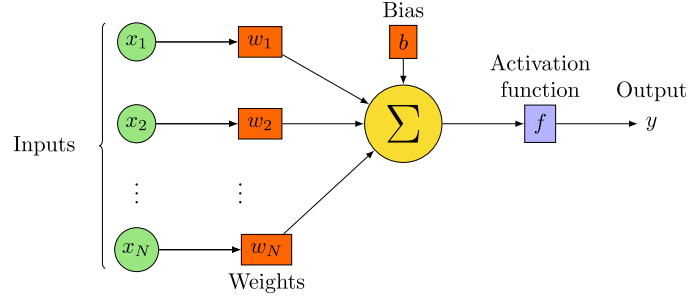


Figure 1. Schematic representation of the mathematical model of a perceptron.

2. The classical perceptron and its capacity

A classical perceptron is a mathematical model that mimics the functioning of a physical neuron. By referring to figure 1, given an input pattern, namely an N -dimensional vector $\mathbf{x} = (x_1, \dots, x_n)$ with x_i real or discrete (eventually binary), a perceptron computes an affine transformation

$$\mathbf{x} \mapsto z := \mathbf{x} \cdot \mathbf{w} + b, \quad (1)$$

with real parameters $\mathbf{w} = (w_1, \dots, w_n)$ and b , called *weights* and *bias*, respectively. Subsequently, the perceptron evaluates on the output z an *activation function* $f: \mathbb{R} \rightarrow \mathbb{R}$, eventually yielding the final result $y := f(z)$.

There exist different possible activation functions, some being more computationally efficient like $\text{sgn}(z)$, and others more biologically inspired as the hyperbolic tangent $f(z) = (e^z - e^{-z}) / (e^z + e^{-z})$, or the sigmoid function $f(z) = 1 / (1 + e^{-z})$.

In the following, we consider the activation function $\text{sgn}(z)$ and, for the sake of simplicity, $b = 0$ (note however that one can always include a bias by adding a component to patterns and weights, and consider for the additional component $x_0 = 1$ and $w_0 = b$). In such a context, a binary classification problem amounts to assigning an input pattern to one of two possible classes, indexed by a dichotomic variable $\xi \in \{-1, 1\}$. In the following, we focus on input patterns with binary entries,

$$\Pi := \{\mathbf{x}^\mu \in \{-1, 1\}^N : 1 \leq \mu \leq p\}, \quad (2)$$

and denote with

$$\Xi := \{\xi = \{\xi^\mu\}_{\mu=1}^p : \xi^\mu = \pm 1\} \quad (3)$$

the set containing the $2^p = \text{card}(\Xi)$ possible classifications of the given p patterns. Such a classical perceptron classifies the patterns by means of a suitably chosen weight vector $\mathbf{w} \in \mathbb{R}^N$ which performs a linear separation of the input space choosing the activation function $f(z)$ to be $f(z) = \text{sgn}(z)$. We then define:

$$y^\mu := \text{sgn}(\mathbf{w} \cdot \mathbf{x}^\mu), \quad 1 \leq \mu \leq p. \quad (4)$$

If a particular target classification $\xi \in \Xi$ is fixed, we say that the perceptron correctly classifies the input patterns if $y^\mu = \xi^\mu$ for each $\mu = 1, \dots, p$. For later convenience it is useful to express

this requirement in a different way introducing the pattern stabilities

$$\Delta^\mu = \xi^\mu \frac{\mathbf{w} \cdot \mathbf{x}^\mu}{\|\mathbf{w}\|}, \tag{5}$$

where $\|\mathbf{w}\| = \sqrt{\sum_j w_j^2}$ denotes the weight Euclidean norm. The classification of the input patterns is correct with respect to a chosen target classification ξ if

$$\Delta^\mu \geq 0 \quad \text{for each } \mu = 1, \dots, p. \tag{6}$$

From a geometric point of view, equation (6) expresses the fact that the hyperplane in \mathbb{R}^N perpendicular to \mathbf{w} separates the patterns \mathbf{x}^μ with $\xi^\mu = +1$ from patterns with $\xi^\mu = -1$.

Example 1. The four binary patterns

$$\Pi = \{\mathbf{x}^1 = (-1, -1), \mathbf{x}^2 = (-1, 1), \mathbf{x}^3 = (1, -1), \mathbf{x}^4 = (1, 1)\}$$

are associated with $2^4 = 16$ dichotomies. Sending $x \mapsto (1+x)/2$, the chosen patterns constitute a two-bit random variable with values (i, j) , $i, j = 0, 1$, and the choice $\xi = (-1, 1, 1, -1)$ implements the XOR binary gate

$$\text{XOR}(0, 0) = 0, \quad \text{XOR}(0, 1) = 1, \quad \text{XOR}(1, 0) = 1, \quad \text{XOR}(1, 1) = 0.$$

The so-called XOR problem [17] in machine-learning exactly corresponds to the impossibility of computing the function XOR by means of a perceptron, since it is not possible to find a line in \mathbb{R}^2 which separates \mathbf{x}^1 and \mathbf{x}^4 from \mathbf{x}^2 and \mathbf{x}^3 .

The XOR problem indicates that it is not always possible to associate to a set of patterns Π the correct target classification ξ by using a perceptron. The existence of a particular choice of weights which realizes the desired classification depends on the particular set Π and on the target classification ξ considered. A natural way to proceed is to consider a statistical approach, where one fixes the total number of patterns p and constructs the pattern set Π by selecting randomly p points from $\{-1, 1\}^N$, and choose independently the corresponding target classifications ξ^μ .

In the approach introduced by Gardner in [20], the capacity of a perceptron is related to the volume of the subset of \mathbb{R}^N consisting of those weights $\mathbf{w} \in \mathbb{R}^N$ of definite length $\|\mathbf{w}\|$ that correctly classify the given input patterns Π . In this context the condition (6) is strengthened in order to include the possibility of the additional requirement of *finite stability* $\kappa > 0$, which amounts to requiring that the quantity defined in (5)

$$\Delta^\mu > \kappa \quad \text{for each } \mu = 1, \dots, p. \tag{7}$$

Remark 2. From a practical perspective, the introduction of a finite stability requirement allows one to obtain a perceptron that may diminish classification errors in the presence of wrong input patterns. For example, because of the presence of some noise, it can happen that instead of the correct pattern \mathbf{x}^μ , an altered version of it is presented to the perceptron, say $\tilde{\mathbf{x}}^\mu$, where some sign is changed, i.e. $\tilde{x}_j^\mu = -x_j^\mu$ for some $j = 1, \dots, N$. In such a case, the request $\kappa = 0$ might lead to an error if the corresponding stability $\Delta^\mu > 0$. On the other hand, the more restrictive request $\kappa > 0$ would not if $\Delta^\mu < \kappa$.

Due to the definition of the quantities Δ^μ , the validity of the conditions (6) is independent on the normalization of the weight vector \mathbf{w} , which implies that we can suitably fix it and compute the relative volume of weights that satisfy the conditions (7) among the weights with a fixed norm. We choose it to be $\|\mathbf{w}\|^2 = N$ (which corresponds to taking vectors \mathbf{w} whose components are $O(1)$), so that the relative volume of interest can be written as:

$$V := V(\{\xi^\mu \mathbf{x}^\mu\}_{\mu=1}^p) = \frac{1}{C_N} \int_{\mathbb{R}^N} d\mathbf{w} \delta(\|\mathbf{w}\|^2 - N) \prod_{\mu=1}^p \theta(\Delta^\mu - \kappa), \quad (8)$$

where θ denotes the Heaviside step function, $\theta(z) = 1$ for $z \geq 0$ and zero otherwise,

$$C_N = \int_{\mathbb{R}^N} d\mathbf{w} \delta(\|\mathbf{w}\|^2 - N) = \frac{\pi^{N/2} N^{N/2-1}}{\Gamma(N/2)} \sim \sqrt{\frac{(2\pi e)^N}{4\pi N}} \quad (9)$$

is the reference volume of weights contained in the hypersphere of radius \sqrt{N} , $\Gamma(z)$ is the Gamma function and the large N behaviour follows from the Stirling approximation. The expression (8) resembles a statistical-mechanical partition function, where the conventional exponential weight is replaced by an all-or-nothing weight given by the step functions. In analogy with statistical mechanics, the relevant quantity is actually $\ln V$ rather than V itself, of which we will compute the average $\langle \ln V \rangle$ with respect to the identically and independently distributed stochastic variables ξ^μ, x_j^μ , with $\mu = 1, 2, \dots, p$, and $j = 1, 2, \dots, N$.

Using the replica method [20], Gardner showed the existence of a critical value of α , given by

$$\alpha_c(\kappa) = \left[\int_{-\kappa}^{\infty} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} (t + \kappa)^2 \right]^{-1}, \quad (10)$$

such that for $\alpha < \alpha_c(\kappa)$ the following limit holds:

$$\begin{aligned} \lim_{\substack{N, p \rightarrow \infty \\ p/N = \alpha}} \frac{\langle \ln V \rangle}{N} &= \min_{0 \leq q \leq 1} \left\{ \alpha \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \ln \left[1 - \Phi \left(\frac{t\sqrt{q} + \kappa}{\sqrt{1-q}} \right) \right] + \frac{q}{2(1-q)} + \frac{\ln(1-q)}{2} \right\} \\ &\equiv \mathcal{F}(\alpha, \kappa), \end{aligned} \quad (11)$$

where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt, \quad (12)$$

while for $\alpha > \alpha_c(\kappa)$:

$$\lim_{\substack{N, p \rightarrow \infty \\ p/N = \alpha}} \frac{\langle \ln V \rangle}{N} = -\infty. \quad (13)$$

Gardner's statistical approach was later reformulated in a mathematically rigorous setting [23, 24], where it was pointed out that the random variable $N^{-1} \ln V$ is self-averaging. Namely, deviations from the average $N^{-1} \langle \ln V \rangle$ become vanishingly small in the limit $N \rightarrow \infty$. In other

words, the average $N^{-1}\langle \ln V \rangle$ is a good representative of what happens for almost all realizations of the random variable $N^{-1} \ln V$, each realization corresponding to a particular choice of the random patterns and classifications $\{\mathbf{x}^\mu, \xi^\mu\}_{\mu=1}^p$.

According to such an observation, equations (11) and (13) can be interpreted as follows: for $\alpha < \alpha_c(\kappa)$ the relative volume of weights that are able to correctly classify a random choice of the patterns is approximately $V \sim \exp(-\mathcal{F}(\alpha, \kappa)N)$, while above the critical value, that is when $\alpha > \alpha_c(\kappa)$, the relative volume is more than exponentially vanishing with N , i.e. $V = o(\exp(-cN))$.

3. Continuous variable quantum perceptron: pattern capacity

We will now use the classical statistical approach outlined above to compute the pattern capacity of the particular quantum perceptron introduced in [14] and schematically represented by the quantum circuit in figure 2. In this model, the components x_j of an input pattern \mathbf{x} are encoded by states of the form

$$|\psi_j\rangle = \frac{1}{(2\pi\sigma_j^2)^{1/4}} \int_{-\infty}^{+\infty} dq_j \exp\left(-\frac{(q_j - x_j)^2}{4\sigma_j^2}\right) |q_j\rangle, \quad (14)$$

which are Gaussian weighted normalized superpositions of pseudo-eigenstates $|q_j\rangle$ of position-like operators q , centered around the pattern components x_j with width σ_j . A pattern \mathbf{x} is then encoded into the tensor product state

$$|\Psi\rangle = \bigotimes_{j=1}^N |\psi_j\rangle = \prod_{j=1}^N \frac{1}{(2\pi\sigma_j^2)^{1/4}} \int_{-\infty}^{+\infty} dq_j \exp\left(-\frac{(q_j - x_j)^2}{4\sigma_j^2}\right) \bigotimes_{j=1}^N |q_j\rangle. \quad (15)$$

Such state is then provided as input to the quantum circuit in figure 2, which firstly operates with a series of independent squeezing operators

$$S_j(r_j) = e^{i r_j (q_j p_j + p_j q_j)}, \quad e^{-2r_j} = w_j, \quad (16)$$

where p_j is a momentum-like operator conjugated to q_j , $[q_j, p_j] = i$. In order to implement negative weights, a phase shift gate $e^{i\frac{\pi}{2}(q^2 + p^2)}$ is operated after the squeezing, thus sending position eigenstates $|q_j\rangle$ into $|w_j q_j\rangle$.

Then, the circuit acts with an entangling controlled addition gate CX on pairs of consecutive states:

$$\text{CX} := \exp(-i q_j \otimes p_{j+1}), \quad \text{CX} |q_j, q_{j+1}\rangle = |q_j, q_j + q_{j+1}\rangle. \quad (17)$$

The combined action on the attenuated multi-mode position eigen-state of the $n - 1$ CX gates of the circuit in figure 2 is then given by

$$\begin{aligned} |w_1 q_1, w_2 q_2, \dots, w_n q_n\rangle &\rightarrow |w_1 q_1, w_1 q_1 + w_2 q_2, \dots, w_n q_n\rangle \rightarrow \dots \\ &\dots \rightarrow \left| w_1 q_1, w_1 q_1 + w_2 q_2, \dots, \sum_{j=1}^N w_j q_j \right\rangle. \end{aligned} \quad (18)$$

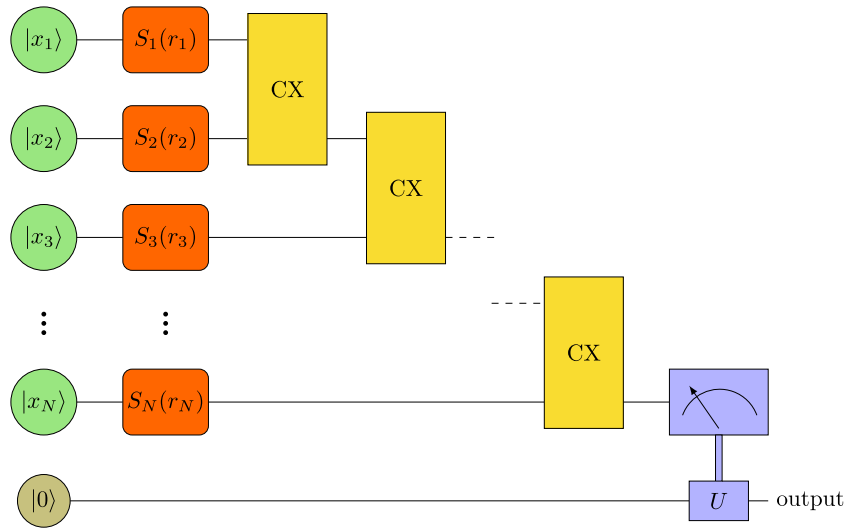


Figure 2. Scheme for a continuous valued quantum perceptron.

Then, the amplitude associated to the last mode position eigenstate is given by the Gaussian distribution centered around $\mathbf{w} \cdot \mathbf{x}^\mu$:

$$\psi_{\mathbf{w},\mathbf{x}^\mu}(s) = \frac{1}{(2\pi \sum_j w_j^2 \sigma_j^2)^{1/4}} \exp\left(-\frac{(s - \mathbf{w} \cdot \mathbf{x}^\mu)^2}{4 \sum_j w_j^2 \sigma_j^2}\right). \tag{19}$$

The variances σ_j^2 can be suitably chosen. Here, we will consider the case $\sigma_j^2 = \sigma^2$ for all j . Homodyne detection operated on the last mode yields the outcome probability

$$P_{\mathbf{w},\mathbf{x}^\mu,\sigma}(s) = |\psi_{\mathbf{w},\mathbf{x}^\mu}(s)|^2 = \frac{1}{\sqrt{2\pi} \|\mathbf{w}\| \sigma} \exp\left(-\frac{(s - \mathbf{w} \cdot \mathbf{x}^\mu)^2}{2 \|\mathbf{w}\|^2 \sigma^2}\right). \tag{20}$$

Consequently a pattern \mathbf{x}^μ is classified as $\sigma = +1$ if the result is above some threshold $\kappa \|\mathbf{w}\|$, and $\sigma = -1$ if it is below $-\kappa \|\mathbf{w}\|$, while the pattern is not classified when the measurement result is between $(-\kappa \|\mathbf{w}\|, \kappa \|\mathbf{w}\|)$. Then, whether a pattern is correctly classified or not becomes a binary stochastic variable for which the probability of correct classification of the pattern μ is

$$R^\mu(\kappa) = \int_{-\infty}^{+\infty} ds P_{\mathbf{w},\mathbf{x}^\mu,\sigma}(s) \theta\left(\xi^\mu \frac{s}{\|\mathbf{w}\|} - \kappa\right). \tag{21}$$

Depending on the actual outcome of the measurement process, the ancilla mode appended to the initialized first N ones will then be changed accordingly. For this classification problem the ancilla can be taken as a qutrit, which is initialized in some reference state, while two orthogonal states are used to store respectively the classifications $\sigma = +1$ and $\sigma = -1$.

Remark 3. Notice that the classical perceptron action is exactly encoded by the position-like pseudo-autokets $|q_1, q_2, \dots, q_N\rangle$. However, such an ideal situation can only be approximated in practice. One useful way to achieve it is via Gaussian smoothening, the narrowing of the

weights thus corresponding to the classical deterministic limit of the proposed continuous quantum perceptron.

3.1. Results

Notice that a pattern which would be correctly classified by a classical perceptron (i.e. a pattern for which $\Delta^\mu > 0$) has a non-vanishing probability of being misclassified by such a model of quantum perceptron, namely $R^\mu < 1$ even if $\Delta^\mu > \kappa$. Viceversa, even if the classical perceptron misclassifies ξ^μ yielding $\Delta^\mu < \kappa$, the quantum one would classify it correctly with some probability. This kind of error comes from the random character of the outcomes s of the homodyne measurement which yields the value s either in $(-\infty, \kappa\|\mathbf{w}\|] \cup [\kappa\|\mathbf{w}\|, \infty)$ or in $(-\kappa\|\mathbf{w}\|, \kappa\|\mathbf{w}\|)$. Because of the unavoidable presence of the stochastic error due to the quantum setting and the corresponding rendering of the perceptron non-linearity by a homodyne quantum measurement, the computation by Gardner requires the introduction of an additional parameter ϵ establishing a threshold for the acceptable probability of error in the classification of each pattern. Then, the relative volume we are interested in is given by

$$V := V(\{\mathbf{x}^\mu, \xi^\mu\}_{\mu=1}^p) = \frac{1}{C_N} \int_{\mathbb{R}^N} d\mathbf{w} \delta(\|\mathbf{w}\|^2 - N) \prod_{\mu=1}^p \theta(R^\mu(\kappa) - 1 + \epsilon). \quad (22)$$

Remark 4. In equation (22) we are computing the volume in the space of weights which guarantees that the probability of making an error *on each* pattern is smaller than ϵ , that will then be chosen smaller than $1/2$. This does not imply that $\epsilon p \leq p/2$ errors will occur on average for the p patterns, but rather that $\epsilon p \leq p/2$ is an upper bound to the number of the patterns which will be misclassified by the perceptron.

As before, the statistically relevant quantity is $\langle \ln V \rangle$, which can be computed using the replica method together with the replica symmetric ansatz, as detailed in the next section. We obtain that the pattern capacity of the quantum model presented is given by the critical value

$$\alpha_c^q(\kappa, \epsilon, \sigma) = \alpha_c(\tilde{\kappa}), \quad (23)$$

where

$$\tilde{\kappa} = \kappa + \sigma \Phi^{-1}(1 - \epsilon). \quad (24)$$

In particular, the following limits hold (they can be obtained from (44), (46) and (52) in section 5):

$$\lim_{\substack{N, p \rightarrow \infty \\ p/N = \alpha}} \frac{\langle \ln V \rangle}{N} = \mathcal{F}(\alpha, \tilde{\kappa}) \quad (25)$$

for $\alpha < \alpha_c(\tilde{\kappa})$, where \mathcal{F} is the same as in equation (11), while for $\alpha > \alpha_c(\tilde{\kappa})$

$$\lim_{\substack{N, p \rightarrow \infty \\ p/N = \alpha}} \frac{\langle \ln V \rangle}{N} = -\infty. \quad (26)$$

Since $\alpha_c(\kappa)$ is decreasing in κ , equation (24) implies that (recall that $0 < \epsilon < 1/2$) the quantum pattern capacity cannot be larger than its classical counterpart (see also figure 3). The proof of the above result is detailed in a dedicated section after the discussion.

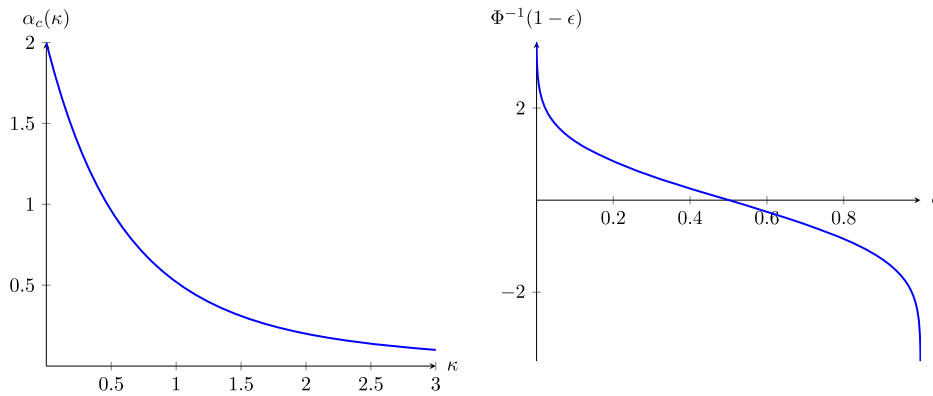


Figure 3. $\alpha_c(\kappa)$ is monotonically decreasing in κ , and $\Phi^{-1}(1 - \epsilon)$ is positive for $0 < \epsilon < 1/2$. Then, for $0 < \epsilon < 1/2$, $\alpha_c^g(\kappa, \epsilon, \sigma) < \alpha_c(\kappa)$.

4. Discussion

In the specific setting considered, a continuous variable quantum perceptron cannot thus provide any quantum advantage as far as its pattern capacity is concerned. The main reason underlying such conclusion is that the quantum perceptron closely resembles a stochastic classical perceptron due to the uncertainty in the outcomes of the (homodyne) measurements necessary to extract the relevant information, namely the stabilities Δ^μ in (5). Notice that, according to remark 3, the functioning of the classical perceptron is recovered from the quantum model proposed by letting the Gaussian widths $\sigma \rightarrow 0$, so that (20) gives $\lim_{\sigma \rightarrow 0} P_{w, x^\mu, \sigma} = \delta(s - w \cdot x)$, and $\tilde{\kappa} \rightarrow \kappa$, as it should be.

The statistical, replica-trick techniques adopted in our investigation are however quite flexible, so that they could be possibly employed to explore different and less classically inspired models of quantum perceptrons. Indeed, upon completion of this work, we discovered that one such attempt has been recently made in [25]. The model considered there is a discrete variable model where the step functions are computed not on $\Delta^\mu - \kappa$, rather on $|\Delta^\mu|^2 - \kappa$. In this way, Gardner’s statistical approach results insensitive to the sign of the stability. Instead, in our case we infer the value of $\text{sgn}(\Delta^\mu)$ by measuring a Gaussianly distributed parameter (see (20) and the corresponding discussion), which mimics the functioning of a stochastic classical perceptron. Using the same statistical techniques employed here, the authors of [25] find that the pattern capacity of that model is twice the classical capacity. Also, differently from the case studied in this work, the use of the measured probabilities as arguments of activations functions makes somewhat difficult to provide a classical limit for this discrete-variable quantum perceptron with which to compare the advantages it brings about.

In this respect, in order to better understand whether quantum advantages can be expected and how they can be achieved, an avenue certainly to be explored is the more information-rooted approach [26] that considers the information stored in a perceptron rather than the number of patterns. Furthermore, the fact that in the continuous model of binary classification by means of a single quantum perceptron here discussed there is no quantum advantage, does not exclude that quantum advantages could come from networks of continuous quantum perceptrons. Assemblies consisting of more than one perceptron should indeed be used for non-binary classification schemes; in such cases constructive quantum interferences could in line of principle lead to higher pattern capacities than in the classical case. However, for that

purpose Gardner’s statistical approach should be extended to networks first in the classical and then in the quantum scenario, which is outside of the scope of the present work and will be matter of further investigation.

5. Methods

In the following we present the derivation of equation (23). Using the replica trick, one has to compute

$$\langle \ln V \rangle = \lim_{n \rightarrow 0} \frac{\langle V^n \rangle - 1}{n}, \tag{27}$$

by evaluating $\langle V^n \rangle$ for n integer, and then taking the limit $n \rightarrow 0$. The quantity we need is then:

$$\langle V^n \rangle = \frac{1}{C_N^n} \left\langle \prod_{\gamma=1}^n \int_{\mathbb{R}^N} d\mathbf{w}^\gamma \delta(\|\mathbf{w}^\gamma\|^2 - N) \prod_{\mu=1}^p \theta(R_\gamma^\mu(\kappa) - 1 + \epsilon) \right\rangle. \tag{28}$$

Using the integral representation

$$\theta(x - \gamma) = \int_{\gamma}^{\infty} dz \delta(z - x) = \int_{\gamma}^{\infty} dz \int_{-\infty}^{\infty} \frac{dy}{2\pi} e^{iy(z-x)}, \tag{29}$$

one writes

$$\theta(R_\gamma^\mu(\kappa) - 1 + \epsilon) = \int_{1-\epsilon}^{\infty} \frac{dz_\gamma^\mu}{2\pi} \int_{-\infty}^{\infty} dy_\gamma^\mu e^{iy_\gamma^\mu z_\gamma^\mu} e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)}. \tag{30}$$

Now resorting to the fact that each pattern is independent from the other ones, we can factor the average into:

$$\begin{aligned} \left\langle \prod_{\mu=1}^p \prod_{\gamma=1}^n \theta(R_\gamma^\mu(\kappa) - 1 + \epsilon) \right\rangle &= \prod_{\mu=1}^p \left\langle \prod_{\gamma=1}^n \theta(R_\gamma^\mu(\kappa) - 1 + \epsilon) \right\rangle \\ &= \prod_{\mu=1}^p \int_{[1-\epsilon, \infty)^n} \left(\prod_{\gamma=1}^n \frac{dz_\gamma^\mu}{2\pi} \right) \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n dy_\gamma^\mu \right) \\ &\quad \times e^{i \sum_{\gamma=1}^n y_\gamma^\mu z_\gamma^\mu} \left\langle \prod_{\gamma=1}^n e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)} \right\rangle. \end{aligned} \tag{31}$$

Then, we have to go through the computation of:

$$\left\langle \prod_{\gamma=1}^n e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)} \right\rangle_{\Pi, \xi}, \tag{32}$$

where we wrote explicitly the variables over which we are averaging. The details of the calculation are given in appendix A, the final result being:

$$\left\langle \prod_{\mu=1}^p \prod_{\gamma=1}^n \theta (R_{\gamma}^{\mu}(\kappa) - 1 + \epsilon) \right\rangle = \left[\int_{[1-\epsilon, \infty)^n} \left(\prod_{\gamma=1}^n \frac{dz_{\gamma}}{2\pi} \right) \int_{\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n} \left(\prod_{\gamma=1}^n \frac{d\lambda_{\gamma} dy_{\gamma} d\omega_{\gamma}}{2\pi} \right) e^{K(\{\lambda_{\gamma}\}, \{y_{\gamma}\}, \{\omega_{\gamma}\}, \{q_{\gamma\delta}\})} \right]^p, \quad (33)$$

where

$$K(\{\lambda_{\gamma}\}, \{y_{\gamma}\}, \{\omega_{\gamma}\}, \{q_{\gamma\delta}\}) := i \sum_{\gamma=1}^n y_{\gamma} [z_{\gamma} - \Phi(\lambda_{\gamma})] - i \sum_{\gamma=1}^n \left(\frac{\kappa}{\sigma} + \lambda_{\gamma} \right) \omega_{\gamma}^{\mu} - \frac{1}{2\sigma^2} \sum_{\gamma, \delta=1}^n q_{\gamma\delta} \omega_{\gamma} \omega_{\delta} \quad (34)$$

and

$$q_{\gamma\delta} = \frac{1}{N} \sum_{j=1}^N w_j^{\gamma} w_j^{\delta}, \quad \text{for } \gamma, \delta = 1, \dots, n, \quad (35)$$

(satisfying $q_{\gamma\gamma} = 1$ due to the constraint $\|\mathbf{w}^{\gamma}\| = \sqrt{N}$) are order parameters introduced by means of delta functions,

$$\delta \left(q_{\gamma\delta} - \frac{1}{N} \sum_{j=1}^N w_j^{\gamma} w_j^{\delta} \right) = N \int_{-\infty}^{+\infty} \frac{dF_{\gamma\delta}}{2\pi} \exp \left(-iNq_{\gamma\delta}F_{\gamma\delta} + iF_{\gamma\delta} \sum_{j=1}^N w_j^{\gamma} w_j^{\delta} \right), \quad (36)$$

which involves the introduction of further parameters $F_{\gamma\delta}$, for $\gamma \neq \delta$. Similarly, from the integral representation

$$\delta \left(\sum_{j=1}^N (w_j^{\gamma})^2 - N \right) = \int_{-\infty}^{+\infty} \frac{dE_{\gamma}}{4\pi} \exp \left[i \frac{E_{\gamma}}{2} \left(N - \sum_{j=1}^N (w_j^{\gamma})^2 \right) \right], \quad (37)$$

we get further order parameters E_{γ} , for $\gamma = 1, \dots, n$. Then, $\langle V^n \rangle$ can be globally rewritten as:

$$\langle V^n \rangle = \frac{1}{C_N^n} \int_{\mathbb{R}^n \times \mathbb{R}^{n(n-1)/2} \times \mathbb{R}^{n(n-1)/2}} \left(\prod_{\gamma=1}^n dE_{\gamma} \right) \left(\prod_{\substack{\gamma, \delta=1 \\ \gamma < \delta}}^n dq_{\gamma\delta} dF_{\gamma\delta} \right) e^{NG(\{q_{\gamma\delta}\}, \{F_{\gamma\delta}\}, \{E_{\gamma}\})}, \quad (38)$$

where

$$G(\{q_{\gamma\delta}\}, \{F_{\gamma\delta}\}, \{E_{\gamma}\}) = \alpha G_1(\{q_{\gamma\delta}\}) + G_2(\{F_{\gamma\delta}\}, \{E_{\gamma}\}) + G_3(\{q_{\gamma\delta}\}, \{F_{\gamma\delta}\}, \{E_{\gamma}\}), \quad (39)$$

with

$$G_1(\{q_{\gamma\delta}\}) = \ln \left[\int_{[1-\epsilon, \infty)^n} \left(\prod_{\gamma=1}^n \frac{dz_\gamma}{2\pi} \right) \int_{\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n} \left(\prod_{\gamma=1}^n \frac{d\lambda_\gamma dy_\gamma d\omega_\gamma}{2\pi} \right) e^{K(\{\lambda_\gamma\}, \{y_\gamma\}, \{\omega_\gamma\}, \{q_{\gamma\delta}\})} \right] \quad (40)$$

$$G_2(\{F_{\gamma\delta}\}, \{E_\gamma\}) = \ln \left[\int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n dw^\gamma \right) \exp \left(-\frac{i}{2} \sum_{\gamma=1}^n E_\gamma (w^\gamma)^2 + i \sum_{\substack{\gamma, \delta=1 \\ \gamma < \delta}}^n F_{\gamma\delta} w^\gamma w^\delta \right) \right] \quad (41)$$

$$G_3(\{q_{\gamma\delta}\}, \{F_{\gamma\delta}\}, \{E_\gamma\}) = -i \sum_{\substack{\gamma, \delta=1 \\ \gamma < \delta}}^n F_{\gamma\delta} q_{\gamma\delta} + \frac{i}{2} \sum_{\gamma=1}^n E_\gamma. \quad (42)$$

When N is large, the behaviour of $\langle V^n \rangle$ can be obtained using the saddle-point approximation:

$$\langle V^n \rangle \simeq \frac{1}{C_N^n} e^{NG(z_S)} \sqrt{\frac{2\pi}{N|\det G''(z_S)|}}, \quad (43)$$

where $z_S = (\{q_{\gamma\delta}^S\}, \{F_{\gamma\delta}^S\}, \{E_\gamma^S\})$ is the stationary point of G , and $G''(z_S)$ denotes the Hessian matrix of G at the stationary point.

Using (43), and the fact that $\langle V^n \rangle \rightarrow 1$ as $n \rightarrow 0$, we can write:

$$\begin{aligned} \frac{\langle \ln V \rangle}{N} &= \lim_{n \rightarrow 0} \frac{\langle V^n \rangle - 1}{nN} \\ &= \lim_{n \rightarrow 0} \frac{\ln \langle V^n \rangle}{nN} \\ &= \lim_{n \rightarrow 0} \frac{1}{n} \left[G(z_S) - \frac{\ln C_N^n}{N} + \frac{1}{2N} \ln \left(\frac{2\pi}{N|\det G''(z_S)|} \right) \right] \\ &= \lim_{n \rightarrow 0} \frac{G(z_S)}{n} - \frac{\ln(2\pi e)}{2} + O\left(\frac{\ln N}{N}\right). \end{aligned} \quad (44)$$

5.1. Replica-symmetric ansatz

At this point the replica-symmetric ansatz is made, which consists in the assumption that the stationary point of G can be found among the points $z = (\{q_{\gamma\delta}\}, \{F_{\gamma\delta}\}, \{E_\gamma\})$ such that:

$$q_{\gamma\delta} = q \quad F_{\gamma\delta} = F \quad E_\gamma = E, \quad (45)$$

for all $\gamma, \delta = 1, \dots, n$, with $\gamma \neq \delta$. In the following, we denote with

$$G^{\text{RS}}(q, F, E) = \alpha G_1^{\text{RS}}(q) + G_2^{\text{RS}}(F, E) + G_3^{\text{RS}}(q, F, E)$$

the restriction of G onto the replica-symmetric subspace. The details of the derivation of $G_1^{\text{RS}}(q)$ are given in appendix B, which yields:

$$G_1^{\text{RS}} = \ln \left[\int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(1 - \Phi \left(\frac{\tilde{\kappa} + t\sqrt{q}}{\sqrt{1-q}} \right) \right)^n \right]. \tag{46}$$

In the limit $n \rightarrow 0$ we obtain the asymptotic expansion

$$G_1^{\text{RS}} = n \int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \ln \left(1 - \Phi \left(\frac{\tilde{\kappa} + t\sqrt{q}}{\sqrt{1-q}} \right) \right) + O(n^2), \tag{47}$$

which follows from the fact that for $n \rightarrow 0$:

$$\begin{aligned} \ln \left[\int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} f(t)^n \right] &= \ln \left[1 + \int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} (f(t)^n - 1) \right] \\ &= \int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} (f(t)^n - 1) + O(n^2) \\ &= n \int_{-\infty}^{\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \ln(f(t)) + O(n^2). \end{aligned} \tag{48}$$

The derivation of G_2^{RS} follows simply by a Gaussian integration:

$$G_2^{\text{RS}}(F, E) = \frac{1}{2} \ln \left(\frac{(2\pi i)^n}{\det \Lambda} \right), \tag{49}$$

where $\Lambda_{\gamma\delta} = -(E + F)\delta_{\gamma\delta} + F$. The determinant is easily computed by noticing that the matrix Λ has one non-degenerate eigenvalue $\lambda_1 = -E + nF$ and a $(n - 1)$ -degenerate eigenvalue $\lambda_2 = -(E + F)$, so that we can write

$$\begin{aligned} G_2^{\text{RS}} &= \frac{1}{2} \ln \left[\frac{(2\pi i)^n}{(-1)^n (E + F)^{n-1} (E - (n - 1)F)} \right] \\ &= \frac{n}{2} \left(\ln(2\pi) - \ln(iE + iF) + \frac{F}{E + F} \right) + O(n^2). \end{aligned} \tag{50}$$

Finally, the last term is simply given by:

$$G_3^{\text{RS}}(q, F, E) = i \frac{n}{2} (E + qF) + O(n^2). \tag{51}$$

Therefore, in the limit $n \rightarrow 0$ we have:

$$\begin{aligned} \frac{1}{n}G^{\text{RS}}(q, F, E) &= \alpha \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \ln \left[1 - \Phi \left(\frac{t\sqrt{q} + \tilde{\kappa}}{\sqrt{1-q}} \right) \right] \\ &+ \frac{1}{2} \left(\ln(2\pi) - \ln(iE + iF) + \frac{F}{E+F} + iE + iqF \right). \end{aligned} \quad (52)$$

The saddle point equations $\frac{\partial G^{\text{RS}}}{\partial E} = 0$ and $\frac{\partial G^{\text{RS}}}{\partial F} = 0$ can readily be solved, giving

$$F_S(q) = -\frac{iq}{(1-q)^2}, \quad E_S(q) = -i\frac{1-2q}{(1-q)^2}, \quad (53)$$

which, substituted into G^{RS} , yield:

$$\begin{aligned} \frac{1}{n}G^{\text{RS}}(q, F_S(q), E_S(q)) &= \alpha \int_{-\infty}^{+\infty} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \ln \left[1 - \Phi \left(\frac{t\sqrt{q} + \tilde{\kappa}}{\sqrt{1-q}} \right) \right] \\ &+ \frac{\ln(2\pi)}{2} + \frac{\ln(1-q)}{2} + \frac{q}{2(1-q)} + \frac{1}{2}. \end{aligned} \quad (54)$$

Setting $\partial G^{\text{RS}}/\partial q = 0$ in (54) we get the saddle point equation

$$\alpha \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \mathcal{A} \left(\frac{\tilde{\kappa} + t\sqrt{q}}{\sqrt{1-q}} \right) \frac{t + \tilde{\kappa}\sqrt{q}}{2\sqrt{q}(1-q)^{3/2}} = \frac{q}{2(1-q)^2}, \quad (55)$$

where

$$\mathcal{A}(u) := \frac{1}{\sqrt{2\pi}} \frac{e^{-u^2/2}}{1 - \Phi(u)}. \quad (56)$$

The most important order parameter is q , whose value at the replica symmetric-saddle point represents the most probable average overlap (35) between a pair of solutions to (7). For $\alpha \rightarrow 0$, equation (55) gives $q \rightarrow 0$: in this case almost all w 's solve (7), and the typical overlap between random pairs of w in the space of interactions is vanishing. As α grows, it becomes harder and harder to find solutions, hence the typical overlap between them increases. The optimal perceptron corresponds to the limit $q \rightarrow 1$, when there is only a single solution solving the problem with the given stability κ , and the corresponding value of α is the critical storage capacity $\alpha_c^q(\kappa, \epsilon, \sigma)$. The critical value of $\alpha_c^q(\kappa, \epsilon, \sigma)$ is obtained thus by equation (55) taking the limit $q \rightarrow 1$:

$$\frac{1}{\alpha_c^q(\kappa, \epsilon, \sigma)} = \lim_{q \rightarrow 1} \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \mathcal{A} \left(\frac{\tilde{\kappa} + t\sqrt{q}}{\sqrt{1-q}} \right) \frac{\sqrt{1-q}}{q^{3/2}} (t + \tilde{\kappa}\sqrt{q}). \quad (57)$$

Using the asymptotic expansion

$$\mathcal{A}(u) = u + O\left(\frac{1}{u}\right) \quad \text{for } u \rightarrow \infty \quad (58)$$

and $\mathcal{A}(u) \rightarrow 0$ as $u \rightarrow -\infty$, one obtains that

$$\lim_{q \rightarrow 1} \int \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} \mathcal{A} \left(\frac{\tilde{\kappa} + t\sqrt{q}}{\sqrt{1-q}} \right) \frac{\sqrt{1-q}}{q^{3/2}} (t + \tilde{\kappa}\sqrt{q}) = (t + \tilde{\kappa})^2 \theta(t + \tilde{\kappa}). \quad (59)$$

Then, by the dominated convergence theorem, equation (57) gives the final result:

$$\alpha_c^q(\kappa, \epsilon, \sigma) = \left[\int_{-\tilde{\kappa}}^{\infty} \frac{dt}{\sqrt{2\pi}} e^{-t^2/2} (t + \tilde{\kappa})^2 \right]^{-1} = \alpha_c(\tilde{\kappa}). \tag{60}$$

Acknowledgments

FB and SM warmly thank Samad Khabbazi-Oskouei for useful discussions in the early stage of this work. GG acknowledges full support from the Italian Region FVG-Fondo Sociale Europeo, Programma specifico 89/19, FP 1910491601—CUP J74I19000760002.

Data availability statement

No new data were created or analysed in this study.

Appendix A. Derivation of (33)

Since x_j^μ and ξ^μ are independent random variables, the average over the ξ^μ 's can be carried out immediately, yielding:

$$\begin{aligned} \left\langle \prod_{\gamma=1}^n e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)} \right\rangle_{\Pi, \xi} &= \frac{1}{2} \left[\left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \int_{\kappa\|\mathbf{w}^\gamma\|}^{\infty} ds P_{\mathbf{w}^\gamma, \mathbf{x}^\mu}(s) \right) \right\rangle_{\Pi} \right. \\ &\quad \left. + \left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \int_{-\infty}^{-\kappa\|\mathbf{w}^\gamma\|} ds P_{\mathbf{w}^\gamma, \mathbf{x}^\mu}(s) \right) \right\rangle_{\Pi} \right]. \tag{A.1} \end{aligned}$$

Now noting that

$$\int_{\kappa\|\mathbf{w}\|}^{\infty} ds P_{\mathbf{w}, \mathbf{x}}(s) = 1 - \Phi \left(\frac{\kappa}{\sigma} - \frac{\mathbf{w} \cdot \mathbf{x}}{\|\mathbf{w}\|\sigma} \right), \tag{A.2}$$

$$\int_{-\infty}^{-\kappa\|\mathbf{w}\|} ds P_{\mathbf{w}, \mathbf{x}}(s) = \Phi \left(-\frac{\kappa}{\sigma} - \frac{\mathbf{w} \cdot \mathbf{x}}{\|\mathbf{w}\|\sigma} \right), \tag{A.3}$$

using the fact that $1 - \Phi(x) = \Phi(-x)$, and recalling that we are integrating over the sphere with $\|\mathbf{w}\| = \sqrt{N}$, we have

$$\begin{aligned} \left\langle \prod_{\gamma=1}^n e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)} \right\rangle_{\Pi, \xi} &= \frac{1}{2} \left[\left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \Phi \left(\frac{\mathbf{w}^\gamma \cdot \mathbf{x}^\mu}{\sigma\sqrt{N}} - \frac{\kappa}{\sigma} \right) \right) \right\rangle_{\Pi} \right. \\ &\quad \left. + \left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \Phi \left(-\frac{\kappa}{\sigma} - \frac{\mathbf{w}^\gamma \cdot \mathbf{x}^\mu}{\sigma\sqrt{N}} \right) \right) \right\rangle_{\Pi} \right]. \tag{A.4} \end{aligned}$$

In order to compute the averages of the two terms inside the square brackets, we write the functions $f_y(\lambda) = e^{-iy\Phi(\lambda)}$ in terms of their Fourier transforms:

$$f_y(\lambda) = \frac{1}{\sqrt{2\pi}} \int d\omega e^{i\omega\lambda} \hat{f}_y(\omega), \quad \hat{f}_y(\omega) = \frac{1}{\sqrt{2\pi}} \int dt e^{-i\omega t} f_y(\lambda). \quad (\text{A.5})$$

We can then write:

$$\begin{aligned} & \left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \Phi \left(\frac{\mathbf{w}^\gamma \cdot \mathbf{x}^\mu}{\sigma\sqrt{N}} - \frac{\kappa}{\sigma} \right) \right) \right\rangle_{\Pi} \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_\gamma^\mu \right) \left\langle \exp \left(i \sum_{\gamma=1}^n \omega_\gamma^\mu \frac{\mathbf{w}^\gamma \cdot \mathbf{x}^\mu}{\sigma\sqrt{N}} \right) \right\rangle_{\Pi} e^{-i\frac{\kappa}{\sigma} \sum_{\gamma=1}^n \omega_\gamma^\mu} \prod_{\gamma=1}^n \hat{f}_{y_\gamma^\mu}(\omega_\gamma^\mu) \\ &= \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_\gamma^\mu \right) \exp \left[\sum_{j=1}^N \ln \cos \left(\frac{1}{\sigma\sqrt{N}} \sum_{\gamma=1}^n w_j^\gamma \omega_\gamma^\mu \right) \right] e^{-i\frac{\kappa}{\sigma} \sum_{\gamma=1}^n \omega_\gamma^\mu} \prod_{\gamma=1}^n \hat{f}_{y_\gamma^\mu}(\omega_\gamma^\mu) \\ &\simeq \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_\gamma^\mu \right) \exp \left(-\frac{1}{2\sigma^2} \sum_{\gamma,\delta=1}^n q_{\gamma\delta} \omega_\gamma^\mu \omega_\delta^\mu \right) e^{-i\frac{\kappa}{\sigma} \sum_{\gamma=1}^n \omega_\gamma^\mu} \prod_{\gamma=1}^n \hat{f}_{y_\gamma^\mu}(\omega_\gamma^\mu), \end{aligned} \quad (\text{A.6})$$

where $q_{\gamma\delta}$ are the order parameters introduced in (35).

Using equation (A.5) we get:

$$\begin{aligned} & \left\langle \prod_{\gamma=1}^n \exp \left(-iy_\gamma^\mu \Phi \left(\frac{\mathbf{w}^\gamma \cdot \mathbf{x}^\mu}{\sigma\sqrt{N}} - \frac{\kappa}{\sigma} \right) \right) \right\rangle_{\Pi} \simeq \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\lambda_\gamma^\mu \right) e^{-i\sum_{\gamma=1}^n y_\gamma^\mu \Phi(\lambda_\gamma^\mu)} \\ & \quad \times \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_\gamma^\mu \right) e^{-i\sum_{\gamma=1}^n (\kappa/\sigma + \lambda_\gamma^\mu) \omega_\gamma^\mu} \\ & \quad \times \exp \left(-\frac{1}{2\sigma^2} \sum_{\gamma,\delta=1}^n q_{\gamma\delta} \omega_\gamma^\mu \omega_\delta^\mu \right). \end{aligned} \quad (\text{A.7})$$

The computation of the second term in square brackets of equation (A.4) is completely analogous and leads to the same result. We can thus write

$$\begin{aligned} & \left\langle \prod_{\gamma=1}^n e^{-iy_\gamma^\mu R_\gamma^\mu(\kappa)} \right\rangle_{\Pi,\xi} \simeq \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n \frac{d\lambda_\gamma^\mu}{2\pi} \right) \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_\gamma^\mu \right) \\ & \quad \times \exp \left(-i \sum_{\gamma=1}^n y_\gamma^\mu \Phi(\lambda_\gamma^\mu) - i \sum_{\gamma=1}^n \left(\frac{\kappa}{\sigma} + \lambda_\gamma^\mu \right) \omega_\gamma^\mu - \frac{1}{2\sigma^2} \sum_{\gamma,\delta=1}^n q_{\gamma\delta} \omega_\gamma^\mu \omega_\delta^\mu \right) \end{aligned} \quad (\text{A.8})$$

which can be inserted into (31) to finally obtain:

$$\begin{aligned} \left\langle \prod_{\mu=1}^p \prod_{\gamma=1}^n \theta(R_{\gamma}^{\mu}(\kappa) - 1 + \epsilon) \right\rangle &\simeq \prod_{\mu=1}^p \int_{[1-\epsilon, \infty)^n} \left(\prod_{\gamma=1}^n \frac{dz_{\gamma}^{\mu}}{2\pi} \right) \\ &\times \int_{\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n} \left(\prod_{\gamma=1}^n \frac{d\lambda_{\gamma}^{\mu} dy_{\gamma}^{\mu} d\omega_{\gamma}^{\mu}}{2\pi} \right) e^{K(\{\lambda_{\gamma}^{\mu}\}, \{y_{\gamma}^{\mu}\}, \{\omega_{\gamma}^{\mu}\}, \{q_{\gamma\delta}\})}, \end{aligned} \tag{A.9}$$

where K is given in (34). Then, using the fact that the factors in the product corresponding to each $\mu = 1, \dots, p$ are all equal, we can drop the μ index from the integration variables and finally obtain (33).

Appendix B. Derivation of $G_1^{\text{RS}}(q)$

The restriction of (34) to the replica symmetric subspace gives

$$\begin{aligned} K^{\text{RS}}(\{\lambda_{\gamma}\}, \{y_{\gamma}\}, \{\omega_{\gamma}\}, q) &= i \sum_{\gamma=1}^n y_{\gamma} [z_{\gamma} - \Phi(\lambda_{\gamma})] - i \sum_{\gamma=1}^n \left(\frac{\kappa}{\sigma} + \lambda_{\gamma} \right) \omega_{\gamma} \\ &\quad - \frac{1-q}{2\sigma^2} \sum_{\gamma=1}^n \omega_{\gamma}^2 - \frac{q}{2\sigma^2} \left(\sum_{\gamma=1}^n \omega_{\gamma} \right)^2 \end{aligned} \tag{B.1}$$

and the integration over the ω_{γ} variables in (40) can now be carried out, yielding:

$$\begin{aligned} &\int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\omega_{\gamma} \right) \exp \left(-i \sum_{\gamma=1}^n \left(\frac{\kappa}{\sigma} + \lambda_{\gamma} \right) \omega_{\gamma}^{\mu} - \frac{q}{2\sigma^2} \sum_{\substack{\gamma, \delta=1 \\ \gamma \neq \delta}}^n \omega_{\gamma} \omega_{\delta} - \frac{1}{2\sigma^2} \sum_{\gamma=1}^n (\omega_{\gamma})^2 \right) \\ &= \sigma^n \int_{\mathbb{R}^n} \left(\prod_{\gamma=1}^n d\tilde{\omega}_{\gamma} \right) \exp \left(-i \sum_{\gamma=1}^n (\kappa + \sigma \lambda_{\gamma}) \tilde{\omega}_{\gamma}^{\mu} - \frac{q}{2} \left(\sum_{\gamma=1}^n \tilde{\omega}_{\gamma} \right)^2 - \frac{1-q}{2} \sum_{\gamma=1}^n (\tilde{\omega}_{\gamma})^2 \right) \\ &= \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \prod_{\gamma=1}^n \left[\sqrt{\frac{2\pi\sigma^2}{1-q}} \exp \left(-\frac{(\kappa + \lambda_{\gamma}\sigma + \sqrt{qt})^2}{2(1-q)} \right) \right], \end{aligned} \tag{B.2}$$

where the first equality follows from a change of variables and the second one follows from a Gaussian integration after a linearization of the quadratic term using the Gaussian integral trick

$$e^{-\frac{q}{2} \left(\sum_{\gamma=1}^n \omega_{\gamma} \right)^2} = \int_{-\infty}^{\infty} \frac{dt}{\sqrt{2\pi}} e^{-\frac{t^2}{2} + i\sqrt{q} \sum_{\gamma=1}^n \omega_{\gamma}}. \tag{B.3}$$

Using (B.2) into (40) and a rescaling of the λ variables, we can write the integral inside the logarithm in (40) as:




$$\begin{aligned}
 & \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \prod_{\gamma=1}^n \int_{-\infty}^{+\infty} \frac{d\lambda_\gamma}{\sqrt{2\pi(1-q)}} \int_{1-\epsilon}^{\infty} \frac{dz_\gamma}{2\pi} \int dy_\gamma e^{-iy_\gamma[z_\gamma - \Phi(\lambda_\gamma/\sigma)]} \\
 & \quad \times \exp\left(-\frac{(\kappa + \lambda_\gamma + \sqrt{qt})^2}{2(1-q)}\right) \\
 & = \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \int_{1-\epsilon}^{\infty} \frac{dz}{2\pi} \int dy e^{-iy[z - \Phi(\lambda/\sigma)]} \right. \\
 & \quad \left. \times \exp\left(-\frac{(\kappa + \lambda + \sqrt{qt})^2}{2(1-q)}\right) \right)^n \\
 & = \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(\int_{-\infty}^{+\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \theta\left[\Phi\left(\frac{\lambda}{\sigma}\right) - 1 + \epsilon\right] \right. \\
 & \quad \left. \times \exp\left(-\frac{(\kappa + \lambda + \sqrt{qt})^2}{2(1-q)}\right) \right)^n, \tag{B.4}
 \end{aligned}$$

where the first equality follows from the fact that factors corresponding to different γ 's are all equal to each other, and the second equality follows from the integral representation of the θ function. After a further change of the integration variables, (B.4) can be rewritten as

$$\begin{aligned}
 & \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(\int_{\kappa + \sigma\Phi^{-1}(1-\epsilon)}^{\infty} \frac{d\lambda}{\sqrt{2\pi(1-q)}} \exp\left(-\frac{(\lambda + \sqrt{qt})^2}{2(1-q)}\right) \right)^n \\
 & = \int_{-\infty}^{+\infty} dt \frac{e^{-t^2/2}}{\sqrt{2\pi}} \left(1 - \Phi\left(\tilde{\kappa} + \frac{\sqrt{qt}}{\sqrt{1-q}}\right) \right)^n, \tag{B.5}
 \end{aligned}$$

which is the argument of the logarithm in (46).

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>

References

- [1] Hertz J, Anders K and Palmer R G 2018 *Introduction to the Theory of Neural Computation* (Boca Raton, FL: CRC Press)
- [2] Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Netw.* **61** 85–117
- [3] Abiodun O I *et al* 2019 Comprehensive review of artificial neural network applications to pattern recognition *IEEE Access* **7** 158820–46
- [4] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
- [5] Rojas R 1996 *Neural Networks: A Systematic Introduction* (Berlin: Springer)
- [6] Merolla P A *et al* 2014 A million spiking-neuron integrated circuit with a scalable communication network and interface *Science* **345** 668–73
- [7] Schuman C D, Potok T E, Patton R M, Birdwell J D, Dean M E, Rose G S and Plank J S 2017 A survey of neuromorphic computing and neural networks in hardware (arXiv:1705.06963)

- [8] Nielsen M A and Chuang I L 2010 *Quantum Computation and Quantum Information* 10th edn (Cambridge: Cambridge University Press)
- [9] Rosenblatt F 1957 The perceptron: a perceiving and recognizing automaton *Technical Report Inc. Report No. 85-460-1* Cornell Aeronautical Laboratory
- [10] Lewenstein M 1994 Quantum perceptrons *J. Mod. Opt.* **41** 2491–501
- [11] Kak S 1995 On quantum neural computing *Inf. Sci.* **83** 143–60
- [12] Zak M and Williams C P 1998 Quantum neural nets *Int. J. Theor. Phys.* **37** 651–84
- [13] Wan K H, Dahlsten O, Kristjánsson H, Gardner R and Kim M S 2017 Quantum generalisation of feedforward neural networks *npj Quantum Inf.* **3** 36
- [14] Benatti F, Mancini S and Mangini S 2019 Continuous variable quantum perceptron *Int. J. Quantum Inf.* **17** 1941009
- [15] Tacchino F, Macchiavello C, Gerace D and Bajoni D 2019 An artificial neuron implemented on an actual quantum processor *npj Quantum Inf.* **5** 26
- [16] Artiaco C, Balducci F, Parisi G and Scardicchio A 2021 Quantum jamming: critical properties of a quantum mechanical perceptron *Phys. Rev. A* **103** L040203
- [17] Minsky M and Papert S A 2017 *Perceptrons: An Introduction to Computational Geometry* (Cambridge, MA: MIT Press)
- [18] Cover T M 1965 Geometrical and Statistical properties of systems of linear inequalities with applications in pattern recognition *IEEE Trans. Electron. Comput.* **EC-14** 326–34
- [19] Venkatesh S S and Psaltis D 1992 On reliable computation with formal neurons *IEEE Trans. Pattern Anal. Machine Intell.* **14** 87–91
- [20] Gardner E 1988 The space of interactions in neural network models *J. Phys. A: Math. Gen.* **21** 257–70
- [21] Franz S and Parisi G 2016 The simplest model of jamming *J. Phys. A: Math. Theor.* **49** 145001
- [22] Franz S, Parisi G, Sevelev M, Urbani P and Zamponi F 2017 Universality of the SAT-UNSAT (jamming) threshold in non-convex continuous constraint satisfaction problems *SciPost Phys.* **2** 19
- [23] Talagrand M 1999 Self averaging and the space of interactions in neural networks *Random Struct. Alg.* **14** 199–213
- [24] Shcherbina M and Tirozzi B 2003 Rigorous solution of the Gardner problem *Commun. Math. Phys.* **234** 383–422
- [25] Gratsea A, Kasper V and Lewenstein M 2021 Storage properties of a quantum perceptron (arXiv:2111.08414)
- [26] Brunel N, Nadal J-P and Toulouse G 1992 *J. Phys. A: Math. Gen.* **25** 5017