The Hintons in your Neural Network:
a Quantum Field Theory View of Deep Learning

Roberto Bondesan [1]   **Max Welling** [1]

Abstract

In this work we develop a quantum field theory formalism for deep learning, where input signals are encoded in Gaussian states, a generalization of Gaussian processes which encode the agent's uncertainty about the input signal. We show how to represent linear and non-linear layers as unitary quantum gates, and interpret the fundamental excitations of the quantum model as particles, dubbed "Hintons". On top of opening a new perspective and techniques for studying neural networks, the quantum formulation is well suited for optical quantum computing, and provides quantum deformations of neural networks that can be run efficiently on those devices. Finally, we discuss a semi-classical limit of the quantum deformed models which is amenable to classical simulation.

1. Introduction and Contributions

Since its inception more than 10 years ago, deep learning has achieved some stunning successes, transforming entire fields such as speech recognition, automated translation, computer vision and protein folding. It's impact on industry and the economy is so large that dedicated chips are being developed just to process neural network workloads. A natural question arises, what will drive the next wave of innovation in this field?

One candidate for the next disruption in AI is quantum computing. Quantum computing is in its nascent phase with only a modest number of qubits or qumodes (i.e. states of light) available to perform computations. Moreover, it is not at all clear how to map neural networks onto a quantum computer in ways that provide clear benefits. Yet, the confluence of deep learning and quantum computing holds promise and deserves exploration.

In this paper we propose a direct mapping of a deep neural network onto a optical quantum computer through the language of quantum field theory. We consider data, e.g. an image, as a finite sampling of an underlying continuous signal. We model this using quantum optical Gaussian states and show that they represent a generalization of a Gaussian process. We show how to compute the posterior of this Quantum Gaussian Process given input data. Then we show how both linear layers and nonlinear layers can be represented as unitary transformations on these Gaussian states.

Because nonlinearities map us outside of the scope of Gaussian states we discuss a number of special cases which are tractable to simulate classically. The full architecture can only be simulated on an optical quantum computer.

We also provide details on how to map our model onto an optical quantum computer, including quantum Gaussian state preparation, and the implementation of linear and nonlinear layers in terms of elementary universal gates.

While we have implemented and tested our tractable approximations we emphasize that our contribution is theoretical; we do not expect that these tractable approximations perform better than their classical cousins. By describing deep learning in the language of quantum field theory we pave the way for the development of novel quantum neural network architectures in the future. We find the match between neural networks and quantum field theory very natural indeed. And perhaps amusingly, in the process of formulating deep learning in the language of quantum field theory we discovered a new particle: The "Hinton" is the elementary excitation of the quantum field from which optical quantum neural networks are made.

2. Related Work

Several works have recently investigated quantum neural networks as a variational quantum circuit using qubit architectures, e.g. (Farhi & Neven, 2018; Verdon et al., 2018; Beer et al., 2019; Cong et al., 2019; Huggins et al., 2019; Bondesan & Welling, 2020). Quantum optical neural networks similar to those described in this work have been discussed in (Killoran et al., 2019; Steinbrecher et al., 2019;

Lau et al., 2017; Das et al., 2018b). Optical devices have also been considered for implementations of classical neural networks on photonic hardware due to the efficiency of matrix multiplication with optical instruments in (Shen et al., 2017). While the implementation of linear layers considered in all these works is the same as ours, our work differs from previous literature in the following two main aspects: 1) we employ Gaussian states to reason about uncertainty due to discretization errors; 2) we show how to implement quantum analogs of popular nonlinearities such as softplus using unitary gates. Technical details of these differences will be discussed after we introduce our framework below. Using Gaussian states for interpolating data that is further processed by a neural network is inspired by the recent works (Li & Marlin, 2016; Finzi et al., 2020), which however are concerned only with classical networks.

Quantum algorithms for Gaussian processes have been studied in (Zhao et al., 2019; Das et al., 2018a). However, the focus there is speeding up GP regression and not devising a quantum neural network. Further, none of these works discusses the connection between quantum optical Gaussian states and Gaussian processes.

3. Background

3.1. Probabilistic Numeric Neural Networks

We start by reviewing the framework of probabilistic numeric neural networks for classifying an input signal with missing data (Finzi et al., 2020). This model uses a GP on a continuous space $\mathcal{X} = \mathbb{R}^D$ to interpolate the input signal and defines a neural network on this GP representation. To simplify the discussion we consider here a finite dimensional input space. This also allows us to present a simple numerical procedure to study the models, and does not change the conceptual findings of the present paper. Note that our approach is also related to (Li & Marlin, 2016) but we shall keep the term probability numeric to emphasize the connection with the field of probabilistic numerics (Cockayne et al., 2019).

The input to the model is $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ corresponding to the observation y_i of a signal (field) $\varphi : \mathcal{X} \to \mathbb{R}$ at location $x_i \in \mathcal{X}$. We assume an underlying finite input grid $\mathcal{X} = \{1, \dots, l\}^{\times d}$ and denote by φ the vector with components φ_x. We consider a prior GP with zero mean and kernel k, and compute the posterior $\mathcal{GP}(\mu', k')$ to interpolate the signal (Rasmussen et al., 2006):

$$\mu'_x = \mathbf{k}_x^\top A^{-1} \mathbf{y}, \quad k'_{x,x'} = k_{x,x'} - \mathbf{k}_x^\top A^{-1} \mathbf{k}_{x'}. \quad (1)$$

Here $\mathbf{k}_x = \{k_{x,x_i}\}_{i=1}^N$, $A_{i,j} = k_{x_i,x_j} + \delta_{i,j}\sigma_n^2$, where σ_n^2 is measurement noise. We then apply a sequence of linear layers and nonlinearities on the input random vector

$\varphi^{(1)} \sim \mathcal{GP}(\mu', k')$:

$$\varphi^{(\ell+1)} = \sigma(\mathcal{B}^{(\ell)}(\mathcal{A}^{(\ell)}\varphi^{(\ell)})). \quad (2)$$

In the intermediate layers, φ has shape $N_C \times |\mathcal{X}|$, with N_C being the number of channels, and components $\varphi_{a,x}$. σ a point-wise non-linearity, $\mathcal{B}$ adds a bias, and $\mathcal{A}$ is linear. In the case of translation equivariance, (Finzi et al., 2020) chose

$$\mathcal{A} = \sum_k W_k e^{\mathcal{D}_k}, \quad (3)$$

with $[W_k]^{ab}$ a matrix of parameters, where a, b index the channels, and

$$\mathcal{D}_k = \sum_{i_1,\dots,i_d \geq 0} \alpha_{i_1,\dots,i_d} \partial_1^{i_1} \cdots \partial_d^{i_d}, \quad (4)$$

where, denoting the unit vector in direction $\mu \in \{1, \dots, d\}$ by e_μ, $(\partial_\mu \varphi)_x = \varphi_{x+e_\mu} - \varphi_x$ and $\alpha_{i_1,\dots,i_d}$ is constant.

After L transformations of type (2), the output feature undergoes a global average pooling, $\varphi^{(L+1)} = \mathcal{P}\varphi^{(L)}$, where

$$(\mathcal{P}\varphi)_{a,x} = \begin{cases} \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \varphi_{a,x} & \text{first element} \\ \varphi_{a,x} & \text{else} \end{cases}. \quad (5)$$

This implements an invertible version of the usual operation. Finally, the first element of $\varphi^{(L+1)}$ is passed through a linear layer that produces an output whose mean is interpreted as logits for classification. Denoting $\mathcal{L} = \mathcal{B} \circ \mathcal{A}$, the chain of operations of the PNCNN is

$$\Phi = \mathcal{L}^{(L)} \circ \mathcal{P} \circ \sigma \circ \mathcal{L}^{(L-1)} \circ \cdots \sigma \circ \mathcal{L}^{(0)}. \quad (6)$$

3.2. Quantum Mechanics

In this section we review some basics of quantum mechanics, in particular quantum fields. We refer the reader to e.g. (Sakurai & Napolitano, 2017) for a standard introduction. In the quantum formalism we associate to every classical configuration of a random field $\varphi = \{\varphi_x\}_{x \in \mathcal{X}} \in \mathbb{R}^{|\mathcal{X}|}$ (for simplicity we ignore the channels in this section) a vector $|\varphi\rangle$. The span of all these vectors is a vector space $\mathcal{H}$, with elements given by superpositions:

$$|\Psi\rangle = \int D(\varphi)\, \psi(\varphi) |\varphi\rangle, \quad D(\varphi) = \prod_{x \in \mathcal{X}} d\varphi_x. \quad (7)$$

Note that $\psi(\varphi)$ are the coefficients $\in \mathbb{C}$ that are used to combine vectors $|\varphi\rangle$. The bra-ket notation might feel a little strange to readers unfamiliar with quantum mechanics, and is used to represent abstract elements in $\mathcal{H}$.

$\mathcal{H}$ is a Hilbert space equipped with a scalar product $\langle \varphi | \varphi' \rangle = \delta(\varphi - \varphi')$, so that $\langle \varphi | \Psi \rangle = \psi(\varphi)$. We call

a linear operator on $\mathcal{H}$ a quantum field (indicated with a hat). The algebra of quantum fields is generated by the pairs $\{\widehat{\varphi}_x, \widehat{\pi}_x\}_{x \in \mathcal{X}}$:

$$\langle \boldsymbol{\varphi} | \, \widehat{\varphi}_x \, | \Psi \rangle = \varphi_x \psi(\boldsymbol{\varphi}) \,, \tag{8}$$

$$\langle \boldsymbol{\varphi} | \, \widehat{\pi}_x \, | \Psi \rangle = -i \frac{\partial}{\partial \varphi_x} \psi(\boldsymbol{\varphi}) \,. \tag{9}$$

They are self-adjoint and satisfy the canonical commutation relations

$$[\widehat{\varphi}_x, \widehat{\pi}_{x'}] := \widehat{\varphi}_x \widehat{\pi}_{x'} - \widehat{\pi}_{x'} \widehat{\varphi}_x = i \delta_{x,x'} \,. \tag{10}$$

We will often need to compute analytic functions of operators O, e.g. $f(O) = e^O$. These expressions are defined by the Taylor expansion of f.

A quantum state is a normalized superposition: $\langle \Psi | \Psi \rangle = 1$. We can define expectation values of an operator O by its matrix elements in a state $\langle \Psi | O | \Psi \rangle$. This reduces to classical expectation values when O is diagonal: $\langle \Psi | O | \Psi \rangle = \int D(\boldsymbol{\varphi}) D(\boldsymbol{\varphi}') \langle \Psi | \boldsymbol{\varphi} \rangle \langle \boldsymbol{\varphi} | O | \boldsymbol{\varphi}' \rangle \langle \boldsymbol{\varphi}' | \Psi \rangle = \mathbb{E}_{\boldsymbol{\varphi} \sim |\psi(\boldsymbol{\varphi})|^2}[O(\boldsymbol{\varphi})]$.

Quantum dynamics needs to preserve the norm of quantum states and acts by unitary operators $\widehat{U}^t = e^{-it\widehat{H}}$, where $\widehat{H}$ is a self-adjoint Hamiltonian. Instead of evolving states we can equivalently evolve the observables $\widehat{A}$ we are going to measure: $\widehat{A}(t) = (\widehat{U}^t)^\dagger \widehat{A} \widehat{U}^t$, which can be computed using the Baker-Campbell-Hausdorff formula:

$$e^{+it\widehat{H}} \widehat{A} e^{-it\widehat{H}} = \widehat{A} + it[\widehat{H}, \widehat{A}] - \frac{t^2}{2}[\widehat{H}, [\widehat{H}, \widehat{A}]] + \dots \tag{11}$$

$\widehat{A}(t)$ satisfies the Heisenberg equation of motion:

$$\frac{d\widehat{A}(t)}{dt} = i[\widehat{H}, \widehat{A}(t)] \,. \tag{12}$$

Measurements reduce a quantum superposition to a classical configuration [1]. This projection $|\Psi\rangle \mapsto |\bar{\varphi}\rangle$ occurs with probability $|\psi(\bar{\varphi})|^2$. We shall make use below of the following more general fact about partial measurements (see e.g. (Watrous, 2018)).

Proposition 3.1. *Let $|\Psi\rangle$ be a prior state in $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$. If a measurement on $\mathcal{H}_1$ gives outcome y_1, the posterior state is $|\zeta_{y_1}\rangle = (P_{y_1} \otimes \mathbf{1}_2) |\Psi\rangle / \sqrt{\langle \Psi | P_{y_1} \otimes \mathbf{1}_2 | \Psi \rangle}$, with $P_y = |y\rangle \langle y|$ the projector on state $|y\rangle$.*

We note that Bayes rule follows from proposition 3.1: a subsequent measurement of an observable on $\mathcal{H}_2$ with outcome

Figure 1. The covariance ellipse $\frac{1}{2}(\boldsymbol{z} - \boldsymbol{m})^T C(\boldsymbol{z} - \boldsymbol{m}) \leq 1$, also known as "quantum blob" (de Gosson, 2006), in a 2d phase space $\boldsymbol{z} = (\varphi, \pi)$. The area is proportional to $\text{Det}(C) > 1$ (uncertainty principle) and is preserved by (20), (21).

y_2 on the state $|\zeta_{y_1}\rangle$ will give outcome probability:

$$p(y_2|y_1) = \langle \zeta_{y_1} | \, \mathbf{1}_1 \otimes P_{y_2} \, | \zeta_{y_1} \rangle = \frac{\langle \Psi | \, P_{y_1} \otimes P_{y_2} \, | \Psi \rangle}{\langle \Psi | \, P_{y_1} \otimes \mathbf{1}_2 \, | \Psi \rangle} \,, \tag{13}$$

which coincides with $p(y_1, y_2)/p(y_1)$.

3.3. Gaussian States

Define the $2|\mathcal{X}|$ dimensional vector of operators:

$$\widehat{\boldsymbol{R}} = (\widehat{\varphi}_1, \dots, \widehat{\varphi}_{|\mathcal{X}|}, \widehat{\pi}_1, \dots, \widehat{\pi}_{|\mathcal{X}|}) \,. \tag{14}$$

After introducing the symplectic form J, (10) reads as:

$$[\widehat{R}_i, \widehat{R}_j] = iJ_{ij} \,, \quad J = \begin{pmatrix} 0 & \mathbf{1}_{|\mathcal{X}|} \\ -\mathbf{1}_{|\mathcal{X}|} & 0 \end{pmatrix} \,. \tag{15}$$

Gaussian states are specified uniquely by their mean and covariance defined as:

$$\boldsymbol{m} = \langle \Psi | \, \widehat{\boldsymbol{R}} \, | \Psi \rangle \tag{16}$$

$$\tfrac{1}{2} C_{ij} = \langle \Psi | \, \tfrac{1}{2}(\widehat{R}_i \widehat{R}_j + \widehat{R}_j \widehat{R}_i) \, | \Psi \rangle - m_i m_j \,. \tag{17}$$

Here and below we shall denote by $1, 2$ the first and second $|\mathcal{X}|$ components related to $\widehat{\varphi}, \widehat{\pi}$ sectors:

$$\boldsymbol{m} = (\boldsymbol{m}^1, \boldsymbol{m}^2) \,, \quad C = \begin{pmatrix} C^{11} & C^{12} \\ C^{21} & C^{22} \end{pmatrix} \,. \tag{18}$$

Proposition 3.2. *The covariance matrix C satisfies:*

$$C = C^T \,, \quad C > 0 \,, \quad C + iJ \geq 0 \,. \tag{19}$$

See Appendix A for a proof. The condition $C + iJ \geq 0$ encodes the uncertainty principle and distinguishes quantum Gaussian states from classical Gaussian distributions on phase space (Bartlett et al., 2012). Figure 1 gives a visualization of this fact.

We denote a Gaussian state by $|\boldsymbol{m}, C\rangle$ [2]. The wave-function $\langle \varphi | \boldsymbol{m}, C \rangle$ of a Gaussian state is a Gaussian function of $\boldsymbol{\varphi}$, albeit with complex quadratic form (Hudson, 1974).

---

[1] One can consider measurements of non-diagonal self-adjoint fields O, which project onto an eigenstate of O, but in this work we shall restrict to measurements of diagonal operators.
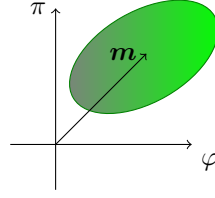
[2] More general mixed Gaussian states can be defined, see e.g. (Adesso et al., 2014; Weedbrook et al., 2012), but we do not need such generality here.

We now show that the following unitary transformations, whose Hamiltonians are at most quadratic in $\widehat{R}_i$, implement the most general transformations among Gaussian states:

$$\widehat{D}(\boldsymbol{\xi}) = \mathrm{e}^{i\widehat{\boldsymbol{R}}^T J \boldsymbol{\xi}}, \quad \boldsymbol{\xi} \in \mathbb{R}^{2|\mathcal{X}|} \tag{20}$$

$$\widehat{\omega}(S) = \mathrm{e}^{\frac{i}{2}\widehat{\boldsymbol{R}}^T J X \widehat{\boldsymbol{R}}}, \quad S = \mathrm{e}^X \in \mathrm{Sp}_{2|\mathcal{X}|}(\mathbb{R}). \tag{21}$$

The Lie group $\mathrm{Sp}_{2|\mathcal{X}|}(\mathbb{R})$ is the symplectic group of matrices satisfying $SJS^T = J$, which include rotations and scaling. An element X in its Lie algebra, $S = \mathrm{e}^X$, satisfies $JX + X^T J = JX - (JX)^T = 0$, so

$$\left(\widehat{\boldsymbol{R}}^T J X \widehat{\boldsymbol{R}}\right)^\dagger = \widehat{\boldsymbol{R}}^T J X \widehat{\boldsymbol{R}}, \tag{22}$$

ensuring unitarity of $\widehat{\omega}(S)$. $\widehat{D}$ and $\widehat{\omega}$ implement symmetry transformations (resp. translations and linear symplectic transforms [3]) on $\mathcal{H}$. The following proposition shows that $\widehat{\boldsymbol{R}}$ intertwines their action on $\mathcal{H}$ and the fundamental action on $\mathbb{R}^{2|\mathcal{X}|}$:

Proposition 3.3. *The unitaries of* (20) *and* (21) *represent symplectic affine transformations of the canonical operators*

$$\widehat{D}(\boldsymbol{\xi})^\dagger \widehat{\boldsymbol{R}} \widehat{D}(\boldsymbol{\xi}) = \widehat{\boldsymbol{R}} + \boldsymbol{\xi} \tag{23}$$

$$\widehat{\omega}(S)^\dagger \widehat{\boldsymbol{R}} \widehat{\omega}(S) = S \widehat{\boldsymbol{R}}. \tag{24}$$

This result follows from formula (11) and the commutation relations (15). See Appendix A for details. This related proposition gives the effect on the mean and covariance of Gaussian states and is also proved in the Appendix A.

Proposition 3.4. *Under the unitaries of* (20) *and* (21)*, the Gaussian states transform as*

$$\widehat{D}(\boldsymbol{\xi})\,|\boldsymbol{m}, C\rangle = |\boldsymbol{m} + \boldsymbol{\xi}, C\rangle \tag{25}$$

$$\widehat{\omega}(S)\,|\boldsymbol{m}, C\rangle = \left|S\boldsymbol{m}, SCS^T\right\rangle. \tag{26}$$

4. Quantum Extensions of Probabilistic Numeric NNs

We introduce in this section a series of quantum operations that generalize the classical layers of a probabilistic numeric NN.

4.1. State Preparation

We start by showing how to perform Bayesian inference with Gaussian states. From 3.1 we have the following result.

Proposition 4.1 (Quantum GP Inference). *Let* $|0, C\rangle$ *be a Gaussian prior state such that* $C_{x,x'}^{11} = k_{x,x'}$. *Given data* $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ *we consider the posterior:*

$$|\zeta_\mathcal{D}\rangle = \frac{P_\mathcal{D}\,|0, C\rangle}{||P_\mathcal{D}\,|0, C\rangle||}, \quad P_\mathcal{D} = \prod_{i=1}^N |y_i\rangle_{x_i\,x_i}\langle y_i|, \tag{27}$$

where $|y\rangle_x$ *is an eigenstate of* $\widehat{\varphi}_x$. *We have:*

$$|\zeta_\mathcal{D}\rangle = |\boldsymbol{m}', C'\rangle, \quad (m')_x^1 = \mu_x', \quad (C')_{x,x'}^{11} = k_{x,x'}' \tag{28}$$

with μ', k' *as in* (1).

Proof. It follows directly from Prop. 3.1 with $|\Psi\rangle = |0, C\rangle$, $\mathcal{H}_1$ the space corresponding to locations x_i and the formulas for the Gaussian conditionals (Rasmussen et al., 2006). $\square$

The quantum GP inference step is thus equivalent to the measurement postulate in quantum mechanics and allows one to encode a classical signal in a quantum state in such a way that quantum entanglement represents an agent's uncertainty about discretization errors.

4.2. Quantum Linear Layers

Next, we show how to perform the quantum equivalent of a linear layer that acts on the quantum fields $\widehat{\boldsymbol{R}}$ in the same way as a classical linear layer acts on a classical field $\boldsymbol{R}$. From Prop. 3.3 we have the following natural definition.

Definition 4.1. *A quantum linear layer is the unitary:*

$$\widehat{U}_{lin}(\boldsymbol{\xi}, S) = \widehat{D}(\boldsymbol{\xi})\widehat{\omega}(S), \tag{29}$$

where $\widehat{D}(\boldsymbol{\xi})$ *and* $\widehat{\omega}(S)$ *generalize the bias and multiplication by the weight matrix respectively.*

4.3. Quantum Non-Linearity

The definition of a unitary operator whose time evolution correspond to the action of a non-linearity is more involved, and in fact has remained elusive in the previous quantum neural network literature. Similarly to classical non-linearities, a quantum non-linearity acts pointwise on the quantum fields. This restricts the associated Hamiltonian to be $\sum_{x,a} \widehat{H}_{x,a}$, where $\widehat{H}_{x,a}$ acts non-trivially only on the quantum fields at x, a. As a design principle, we consider the following class of time evolutions which map $\widehat{\varphi}_{x,a}$ to a function $\sigma(\widehat{\varphi}_{x,a})$. (Recall that a function of an operator is defined by its Taylor series.) This will provide a simple way to embed classical neural networks in our framework.

Proposition 4.2. *Under the time evolution generated by*

$$\widehat{U}_\sigma = \exp\left(-i\sum_{x,a} \widehat{H}_{x,a}\right), \tag{30}$$

$$\widehat{H}_{x,a} = \tfrac{1}{2}(\widehat{\pi}_{x,a} f(\widehat{\varphi}_{x,a}) + f(\widehat{\varphi}_{x,a})\widehat{\pi}_{x,a}), \tag{31}$$

the fields evolve according to the equations of motion:

$$\dot{\widehat{\varphi}}_{x,a}(t) = f(\widehat{\varphi}_{x,a}(t)), \tag{32}$$

$$\dot{\widehat{\pi}}_{x,a}(t) = -\tfrac{1}{2}(\widehat{\pi}_{x,a}(t)f'(\widehat{\varphi}_{x,a}(t)) + h.c.), \tag{33}$$

where h.c. means the Hermitian conjugate of the expression preceding it.

Proof. The equations of motion follow from (12) upon using the commutator $[\widehat{\pi}_{x,a}, f(\widehat{\varphi}_{x,a})] = -if(\widehat{\varphi}_{x,a})$. $\square$

The following proposition relates σ to f (the proof checks the time derivatives and is in Appendix B).

Proposition 4.3. *The ODEs* (32), (33) *have solutions*

$$\widehat{\varphi}_{x,a}(t) = F^{-1}\left(F(\widehat{\varphi}_{x,a}(0)) + t\right), \tag{34}$$

$$\widehat{\pi}_{x,a}(t) = \frac{1}{2}\left(\widehat{\pi}_{x,a}(0)\frac{f(\widehat{\varphi}_{x,a}(0))}{f(\widehat{\varphi}_{x,a}(t))} + h.c.\right) \tag{35}$$

where F is such that $\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \frac{1}{f(x)}$.

Prop. 4.3 gives a general, albeit implicit, solution to the problem of constructing a quantum non-linearity. Next, we give an explicit solution for the case of softplus, a popular smooth version of ReLU[4].

Lemma 4.4 (Quantum Softplus). *The softplus non-linearity with temperature β,*

$$\sigma_\beta(x) = \frac{1}{\beta}\log\left(1 + e^{\beta x}\right), \tag{36}$$

corresponds to time evolution from time 0 to time 1 under

$$\widehat{H}_{x,a} = \frac{1}{2\beta}\left(\widehat{\pi}_{a,x}e^{-\beta\widehat{\varphi}_{a,x}} + e^{-\beta\widehat{\varphi}_{a,x}}\widehat{\pi}_{a,x}\right). \tag{37}$$

Proof. This Hamiltonian is of the general form (30) with $f(x) = \beta^{-1}e^{-\beta x}$. The antiderivative of its inverse, $\beta e^{+\beta x}$, is $F(x) = e^{\beta x}$. So $F^{-1}(y) = \beta^{-1}\log(y)$ and the result then follows immediately from Prop. 4.3. $\square$

We now move on to discussing architectures and symmetries of quantum neural networks. We shall resume the study of non-linearities in section 7 where we will look at a semi-classical limit of the quantum models.

4.4. Architecture

We define a quantum neural network by putting these pieces together:

$$\widehat{U}_{\mathrm{NN}} = \widehat{U}_{\mathrm{lin}}(\boldsymbol{\xi}^{(L)}, S^{(L)})\widehat{U}_{\mathcal{P}}\prod_{\ell=0}^{L-1}\widehat{U}_\sigma\widehat{U}_{\mathrm{lin}}(\boldsymbol{\xi}^{(\ell)}, S^{(\ell)}), \tag{38}$$

where $\widehat{U}_\sigma$ and $\widehat{U}_{\mathrm{lin}}$ are as in Prop. 4.2 and Def. 4.1, while

$$\widehat{U}_{\mathcal{P}} = \widehat{\omega}(S = \mathcal{P} \oplus (\mathcal{P}^{-1})^T), \tag{39}$$

is a global average pooling operator, where $\mathcal{P}$ is as in (5). To make a prediction we proceed similarly to (Finzi et al.,

2020). We discard the spatial locations that have not been aggregated over by averaging in $\widehat{U}_{\mathcal{P}}$, act with a final linear classifier and finally measure the means $\widehat{\varphi}_a$ for the $a = 1, \ldots, C$ channels, C being the number of classes:

$$l_a = \langle\zeta_{\mathcal{D}}|\widehat{U}_{\mathrm{NN}}^\dagger\widehat{\varphi}_a\widehat{U}_{\mathrm{NN}}|\zeta_{\mathcal{D}}\rangle. \tag{40}$$

These are interpreted as the logits for the classification task at hand, analogously to the classical case of section 3.1. (Regression can be treated similarly by changing the last layer as standard practice for neural networks.) For simplicity we introduced the state $|\zeta_{\mathcal{D}}\rangle$ in Prop. 4.1 without referring to the channel dimension. To make sense of the action of $\widehat{U}_{\mathrm{NN}}$ on it, we add extra registers for the channel dimension which are initialized to the vacuum state, which is a Gaussian state with zero mean and unit covariance (Adesso et al., 2014). While we considered here only a global average pooling at the end following (Finzi et al., 2020), it is possible to pool features in intermediate layers as well by simply discarding registers associated to the modes to be discarded. Computing the logits of (40) is in general intractable classically and requires a quantum computer. Training can be done via gradient descent – see (Beer et al., 2019; Verdon et al., 2018) for examples of quantum neural network training and (Shen et al., 2017) for classical optical neural network training.

Our definition of a quantum neural network is similar to that of (Killoran et al., 2019). In particular, the construction of quantum linear layers is the same. However, w.r.t. that work we introduce the following two main novelties: 1) we use Gaussian states for data interpolation (Prop. 4.1); 2) we use a unitary gate implementing the non-linearity (see Sec. 4.3). (Killoran et al., 2019) discusses two sets of non-linearities: the first uses a quantum channel to implement a non-linearity of the type $\widehat{\varphi}_{x,a} \mapsto \sigma(\widehat{\varphi}_{x,a})$, but, w.r.t. our setting, that implementation requires to double the number of quantum registers and then discard half of them.[5] The second strategy is to employ Hamiltonians such as $\widehat{H}_{\mathrm{cubic}} = \widehat{\varphi}_{a,x}^3$ or $\widehat{H}_{\mathrm{Kerr}} = (\widehat{\varphi}_{a,x}^2 + \widehat{\pi}_{a,x}^2)^2$ as non-linearities, since either of them, together with the unitaries of (20) and (21), form a simple set of universal gates for quantum computation with continuous variables (Lloyd & Braunstein, 1999), i. e. any other quantum gate can be expressed in terms of those.[6] While requiring minimal resources, these choices are likely to perform worse than our choice. Indeed in the semi-classical limit discussed in section 7, these correspond to low degree polynomial non-linearities, which are not efficient for classical neural network approximation (Pinkus, 1999).

---

[4]The ability to implement ReLU and linear layers gives us a recipe to implement max pooling as well. Indeed for two inputs, $\max(x_1, x_2) = \mathrm{ReLU}(x_1 - x_2, 0) + x_2$, and the many inputs case can be reduced to comparisons of pairs of inputs.

[5]In our setting σ is constrained to be an ODE flow. However, this is not restrictive from the point of view of expressivity of the neural network due to the presence of linear layers.

[6]The quantum linear layers together with $\widehat{U}_\sigma$ also provide a universal set if f is a polynomial of degree at least 3.

4.5. Symmetries

Generally, symmetries in classical neural networks are realized as linear maps $g \in G$ that acts on the activations φ as $\rho(g)\varphi$, where ρ is a representation matrix. On top of translations, prominent examples of G in ML are rotations (Cohen et al., 2018) and permutations (Maron et al., 2018). Having replaced the linear action on activations with $\widehat{\omega}$, we define unitary representations of G on quantum states by $\widehat{\omega}(S_g := \rho(g) \oplus \rho^*(g))$, where $\rho^*(g) = \rho(g^{-1})^T$ is the dual representation, ensuring symplecticity of S_g. We have

$$\widehat{\omega}(S_g)^\dagger \begin{pmatrix} \widehat{\varphi} \\ \widehat{\pi} \end{pmatrix} \widehat{\omega}(S_g) = \begin{pmatrix} \rho(g)\widehat{\varphi} \\ \rho^*(g)\widehat{\pi} \end{pmatrix}. \tag{41}$$

For example, in case of translations along the $\mu \in \{1, \ldots, d\}$ axis, $(\tau_\mu \varphi)_{a,x} = \varphi_{a,x+e_\mu}$, we have $S_{\tau_\mu} = \tau_\mu \oplus \tau_\mu$, which translates both $\widehat{\varphi}, \widehat{\pi}$ variables equally.

Equivariance of a quantum linear layer $\widehat{\omega}(S)$ now amounts to the commutation relations:

$$\widehat{\omega}(S)\widehat{\omega}(S_g) = \widehat{\omega}(S_g)\widehat{\omega}(S) \Rightarrow SS_g = S_g S \tag{42}$$

where the second formula follows from the group homomorphism property: $\widehat{\omega}(S)\widehat{\omega}(S') = \widehat{\omega}(SS')$. The characterization of symmetries presented here completely solves the problem of designing equivariant quantum linear layers by reducing the problem to designing equivariant classical linear layers with symplectic weight matrices S which commute with $\rho(g) \oplus \rho^*(g)$. As in the classical case, since the non-linearities act pointwise, they will be invariant under operations that permute the coordinates $(\widehat{\varphi}_{x,a}, \widehat{\pi}_{x,a}) \mapsto (\widehat{\varphi}_{x',a'}, \widehat{\pi}_{x',a'})$, such as spatial symmetries, ensuring equivariance of the whole architecture.

As an illustration, the condition $SS_{\tau_\mu} = S_{\tau_\mu}S$ restricts each $M \times M$ block of

$$S = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{43}$$

to be a convolution. Similarly, for rotations $G = \mathrm{SO}(3)$, irreps are self-dual, so $S_g = \rho(g) \oplus \rho(g)$ and each block of S is a group convolution (Cohen et al., 2018).

5. Hintons

We discuss now a particle interpretation of the formalism introduced for neural networks. We introduce the operators:

$$\widehat{b} = \frac{1}{\sqrt{2}}(\widehat{\varphi} + i\widehat{\pi}), \quad \widehat{b}^\dagger = \frac{1}{\sqrt{2}}(\widehat{\varphi} - i\widehat{\pi}). \tag{44}$$

We can then use a particle basis (a.k.a. Fock space) for $\mathcal{H}$, where we identify a zero mean and unit covariance Gaussian state with no particles (vacuum) $|\Omega\rangle = |0, \mathbf{1}\rangle$ such that $\widehat{b}|\Omega\rangle = 0$ and create an orthogonal basis by acting with

different monomials $(\widehat{b}_{a_1,x_1}^\dagger)^{n_i} \cdots (\widehat{b}_{a_m,x_m}^\dagger)^{n_m}$ on $|\Omega\rangle$, n_i being the number of particles at channel a_i and location x_i. In the quantum optical setting, the particles are called photons. In the neural network context we dub these fundamental excitations "Hintons" after G. Hinton, a founding father of the field of deep learning.

From the quantum field theory point of view, locality depends on whether the logarithm of S in the quantum linear layer of definition 4.1 has non-zero matrix elements between sites that are far apart. Convolutional layers lead to a local theory, while fully connected layers introduce non-local interactions.

In the next sections, we will analyze tractable limits of the quantum neural network defined so far to get more insights into the architectural design.

6. The Case of Classical Probabilistic NNs

Here we show how a classical probabilistic NN can be embedded in the quantum model of Sec. 4.4. First we prove the following quantum representation of the push forward of a GP under a generic classical (invertible) map (proof in the Appendix C).

Lemma 6.1. *Let $|m, C\rangle$ be a Gaussian state and $\widehat{U}$ a unitary such that:*

$$\widehat{U}^\dagger \widehat{\varphi} \widehat{U} = F(\widehat{\varphi}). \tag{45}$$

Then:

$$|\langle \varphi | \widehat{U} | m, C \rangle|^2 = \left(F \# \mathcal{GP}(m^1, C^{11}) \right)(\varphi) \tag{46}$$

where $f \# p$ denotes the push forward of p under f.

We already know that $\widehat{U}_\sigma$ has this property, so we only need to constrain the quantum linear layers so that they do not mix $\widehat{\varphi}$ with $\widehat{\pi}$.

Theorem 6.2. *Consider the quantum network* (38) *with:*

$$S^{(\ell)} = \begin{pmatrix} \mathcal{A}^{(\ell)} & 0 \\ 0 & ((\mathcal{A}^{(\ell)})^{-1})^T \end{pmatrix}, \quad \xi^{(\ell)} = (b^{(\ell)}, 0). \tag{47}$$

We have the quantum–classical duality:

$$|\langle \varphi | \widehat{U}_{NN} | \zeta_\mathcal{D} \rangle|^2 = (\Phi \# \mathcal{GP}(\mu', k'))(\varphi), \tag{48}$$

where the rhs is the push forward of the GP posterior (1) *under the map Φ of* (6). *That is, the logits of Eq.* (40) *computed by the quantum neural network coincide exactly with those computed by the probabilistic numeric NN of section 3.1 with weights and biases $\mathcal{A}^{(\ell)}, \xi^{(\ell)}$.*

Proof. Recalling the linear transformation law of Prop. 3.3 and the non-linear action of $\widehat{U}_\sigma$, the proof then follows from lemma 6.1 with $F = \Phi$. $\square$

7. The Semiclassical Limit

Let us denote by $\boldsymbol{R} = (\boldsymbol{\varphi}, \boldsymbol{\pi})$ the classical fields corresponding to the quantum operators introduced above. Recall that notationally we distinguish between classical and quantum fields by the absence or presence of a hat.

Note that despite producing entangled states, the quantum linear layers acting on Gaussian states can be simulated efficiently on a classical computer as the action amounts to the matrix multiplications of Prop. 3.3. See also (Bartlett et al., 2002). In fact, at the linear level the only difference between a quantum evolution and a probabilistic classical evolution of a Gaussian Liouville measure in phase space is the covariance condition $C + iJ \geq 0$ (Prop. 3.2) coming from the non-commutativity of position and momenta in quantum mechanics. We remark also that this condition is preserved by classical evolution thanks to the symplectic nature of classical mechanics (De Gosson, 2009). In Sec. 6 we showed how restricting the linear layers to block diagonal matrices led to the classical model, which corresponds to the push forward of an initial Gaussian Liouville distribution under a neural network, but this time only involving the φ field. In this section, as an intermediate step towards studying the full quantum model, we change the non-linearity in such a way that the modified model corresponds to the push forward of a initial Gaussian measure under a neural network, involving both the φ and the π fields. We can interpret the resulting model as a semi-classical limit of the quantum model since it uses elements of quantum mechanics (uncertainty relation for the covariance) as well as classical mechanics (for the non-linearity). Conceptually, we have the inclusion of models as special cases depicted in figure 2.
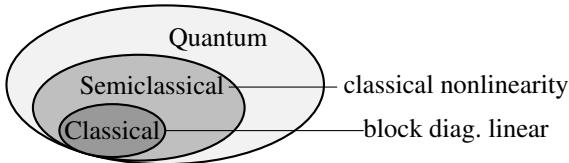
Figure 2. Hierarchy of the neural networks considered.

We define the semiclassical model by simply replacing $\widehat{U}_\sigma$ with a classical Hamiltonian evolution under which the phase space measure evolves into a new classical phase space measure, which corresponds to the classical limit of the equation of motion (Sakurai & Napolitano, 2017).

Proposition 7.1. *Under the classical time evolution generated by the Hamiltonian*

$$H = \sum_{x,a} \pi_{x,a} f(\varphi_{x,a}) \,, \tag{49}$$

the fields transform as:

$$\varphi_{x,a}(t) = F^{-1}\left(F(\varphi_{x,a}(0)) + t\right) \,, \tag{50}$$

$$\pi_{x,a}(t) = \pi_{x,a}(0) \frac{f(\varphi_{x,a}(0))}{f(\varphi_{x,a}(t))} \tag{51}$$

where $F'(x) = 1/f(x)$.

Proof. The proof follows by simply checking the classical equations of motion:

$$\dot{\varphi}_{x,a}(t) = f(\varphi_{x,a}(t)) \,, \quad \dot{\pi}_{x,a}(t) = -\pi_{x,a}(t) f'(\varphi_{x,a}(t))$$

$\square$

We note that classical and quantum equations of motions and solutions (Prop. 4.3) look identical. This is a consequence of the correspondence between quantum and classical mechanics under the identification $[\widehat{A}, \widehat{H}] \longleftrightarrow i\hbar\{A, H\}$, where $\{\cdot, \cdot\}$ is the Poisson bracket.

In particular the following is the classical counterpart of lemma 4.4.

Proposition 7.2 (Symplectic Softplus). *Replacing operators with classical variables in lemma 4.4 we have*

$$U_\sigma : \begin{pmatrix} \varphi \\ \pi \end{pmatrix} \mapsto \begin{pmatrix} \frac{1}{\beta}\log\left(1 + e^{\beta\varphi}\right) \\ \pi(1 + e^{-\beta\varphi}) \end{pmatrix} \,. \tag{52}$$

We then define a neural network that pushes forward the input Gaussian Liouville distribution on phase space to an output distribution p_{out} by alternating linear layers with non-linear classical layers U_σ. Analogously to the original construction of (40), its mean is then used as the logits for classification:

$$l_c = \mathbb{E}_{\varphi_c \sim p_{\text{out}}}(\varphi_c) \,. \tag{53}$$

The semiclassical neural network is a type of learnable Hamiltonian flow (Bondesan & Lamacraft, 2019; Rezende et al., 2019; Toth et al., 2020). Adding momenta can be interpreted as an augmentation strategy for neural ODEs (Dupont et al., 2019; Massaroli et al., 2021). However we do not expect that the semiclassical neural network outperforms the classical ones of (Li & Marlin, 2016; Finzi et al., 2020), since already in the classical case we avoid the expressivity restrictions of ODEs thanks to the presence of extra channels.

We implemented and tested the semiclassical neural network with symplectic softplus non-linearity. The experiments are performed on a simple classification task for irregularly sampled time series. As already remarked, we do not claim that the semi-classical network performs better than a classical probabilistic numeric neural network with similar capacity.

We therefore emphasize that the purpose of our experiments is simply to check that after adding momenta and using a different non-linearity than standard, the model performs similarly to a classical baseline. In the implementation we saw that taking $\beta = 0.1$ for the non-linearity avoids numerical instabilities for negative φ due to the presence of $\mathrm{e}^{-\beta\varphi}$ in (52). We refer to Appendix D for details of architecture, task and learning algorithm tested. We leave the study of the fully quantum case as an outstanding open problem which will most likely be tackled only when error corrected quantum optical computers will become available.

8. Quantum Optical Implementation

We here describe practical considerations for implementing our model on an optical quantum computer. Optical devices are attractive since one can implement matrix multiplication in a very efficient way (Shen et al., 2017) and quantum optical implementations of quantum neural networks have been discussed in (Killoran et al., 2019; Steinbrecher et al., 2019; Lau et al., 2017; Das et al., 2018b). A graphical depiction of our proposal is in Fig. 3. While the implementation of linear layers is common to these works, the quantum non-linearities and the data embedding in Gaussian states are original to our work.

8.1. State preparation

The state preparation step encodes the input signal $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$ into quantum registers. We compute the GP posterior as in (1) for a set of points $x \in \mathcal{X}$ and then we create an input Gaussian state by acting on the vacuum state $|\Omega\rangle$ defined in Sec. 5 with the linear layer $\widehat{D}(\boldsymbol{\xi} = (\boldsymbol{\mu}', 0))\widehat{\omega}(S = A \oplus A^{-1})$, where A is a square root of k'. The vacuum state can be created using lasers (Nielsen & Chuang, 2000) and we defer to the next section the implementation of the linear layer. This procedure incurs a complexity similar to the classical GP inference, that is $O(N^3)$. In Appendix E.1 we give further comments on the state preparation step. We see developing a more efficient quantum implementation of quantum GP inference as in interesting future direction.

8.2. Linear layer

The implementation of a unitary $\widehat{\omega}(S)$ on a quantum optical computer is a well studied problem and we limit ourselves here to simply say that it can be done by a sequence of beamsplitters and phase shifters. Appendix E.2 provides more details of the procedure and its complexity.

8.3. Non linearity

Quantum computers can perform arbitrary computations if given a set of universal gates. For quantum optical com-

puters, one can take the quadratic Hamiltonians and the cubic Hamiltonian $\widehat{\varphi}^3$ (Lloyd & Braunstein, 1999). We can implement a non-linearity with Hamiltonian (30) by approximating it with the truncation of the Taylor series of the function f to order k:

$$\widehat{H}^{(k)} = \sum_{x,a} \sum_{\ell=0}^{k} f_\ell \widehat{H}_\ell \,, \quad \widehat{H}_\ell = \frac{1}{2}\left(\widehat{\pi}_{x,a}\widehat{\varphi}_{x,a}^\ell + \widehat{\varphi}_{x,a}^\ell \widehat{\pi}_{x,a}\right)\,.$$

Note that the non-linear gate Hamiltonian is unbounded from below but that is not a problem here, and the construction gives a well-defined quantum operation. In fact, evidence that it is possible to implement and compute with a Hamiltonian with unbounded spectrum from below is provided by the squeezing Hamiltonian $\widehat{\phi}\widehat{\pi} + \widehat{\pi}\widehat{\phi} = i(\widehat{b}^2 - (\widehat{b}^\dagger)^2)$, which is a special case of the linear layer. (Physically, it can be implemented using optical parametric amplifiers, see App. E.2.)

We then use the standard procedures for quantum simulation (Nielsen & Chuang, 2000) to implement $\mathrm{e}^{i\widehat{H}^{(k)}}$. This is detailed in Appendix E.3, where we derive an explicit decomposition in terms of gates which use only the universal Hamiltonians $\widehat{\pi}_{x,a}$, $\widehat{\pi}_{x,a}^2$ and $\widehat{\varphi}_{x,a}^3$. This fully specifies a protocol to implement our model on a quantum optical device.

9. Conclusion and Outlook

In this paper we introduced a new formalism for deep learning in terms of quantum fields. We made the following contributions: 1) we showed how to use Gaussian states for Bayesian inference, 2) we devised unitary operators that implement standard non-linearities, 3) we presented tractable limits of the quantum network, 4) we discussed how to implement our models on a quantum computer. We also implemented the semi-classical architecture to check that it performs on par with classical models based on GPs (which it did).

Exciting directions for the near future are: studying approximate solutions that get closer to the full quantum model; finding efficient ways to do quantum GP inference on quantum hardware; developing further quantum non-linearities and the quantum formalism for classical models.
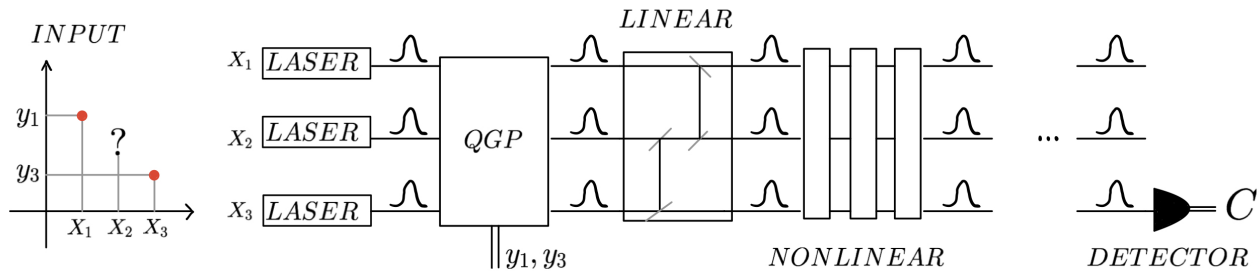
Figure 3. High level depiction of the implementation of our model in quantum optical hardware. The input on the left are observations y_1, y_3 of a signal at locations x_1, x_3 – information at the intermediate value x_2 is missing. We then prepare laser beams for all locations x_1, x_2, x_3 and use quantum GP (QGP) inference to create a posterior state. We then apply a series of linear and non-linear layers, till we measure an observable with a detector to get a class C for classifying the input signal.

References

Adesso, G., Ragy, S., and Lee, A. R. Continuous variable quantum information: Gaussian states and beyond. *Open Systems & Information Dynamics*, 21(01n02):1440001, 2014.

Bartlett, S. D., Sanders, B. C., Braunstein, S. L., and Nemoto, K. Efficient classical simulation of continuous variable quantum information processes. *Physical Review Letters*, 88(9), Feb 2002. ISSN 1079-7114. doi: 10.1103/physrevlett.88.097904. URL http://dx.doi.org/10.1103/PhysRevLett.88.097904.

Bartlett, S. D., Rudolph, T., and Spekkens, R. W. Reconstruction of gaussian quantum mechanics from liouville mechanics with an epistemic restriction. *Physical Review A*, 86(1), Jul 2012. ISSN 1094-1622. doi: 10.1103/physreva.86.012103. URL http://dx.doi.org/10.1103/PhysRevA.86.012103.

Beer, K., Bondarenko, D., Farrelly, T., Osborne, T. J., Salzmann, R., and Wolf, R. Efficient Learning for Deep Quantum Neural Networks. *arXiv e-prints*, art. arXiv:1902.10445, February 2019.

Bondesan, R. and Lamacraft, A. Learning symmetries of classical integrable systems, 2019.

Bondesan, R. and Welling, M. Quantum deformed neural networks. 2020.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. The ucr time series classification archive, July 2015. www.cs.ucr.edu/~eamonn/time_series_data/.

Cockayne, J., Oates, C. J., Sullivan, T. J., and Girolami, M. Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789, 2019.

Cohen, T. S., Geiger, M., Koehler, J., and Welling, M. Spherical cnns, 2018.

Cong, I., Choi, S., and Lukin, M. D. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, August 2019. doi: 10.1038/s41567-019-0648-8.

Das, S., Siopsis, G., and Weedbrook, C. Continuous-variable quantum gaussian process regression and quantum singular value decomposition of nonsparse low-rank matrices. *Physical Review A*, 97(2), Feb 2018a. ISSN 2469-9934. doi: 10.1103/physreva.97.022315. URL http://dx.doi.org/10.1103/PhysRevA.97.022315.

Das, S., Siopsis, G., and Weedbrook, C. Continuous-variable quantum gaussian process regression and quantum singular value decomposition of nonsparse low-rank matrices. *Phys. Rev. A*, 97:022315, Feb 2018b. doi: 10.1103/PhysRevA.97.022315. URL https://link.aps.org/doi/10.1103/PhysRevA.97.022315.

de Gosson, M. *Symplectic Geometry and Quantum Mechanics*. Operator Theory: Advances and Applications. Birkhäuser Basel, 2006. ISBN 9783764375751. URL https://books.google.nl/books?id=q9SHRvay75IC.

De Gosson, M. A. The symplectic camel and the uncertainty principle: The tip of an iceberg? *Foundations of Physics*, 39(2):194–214, 2009.

Dupont, E., Doucet, A., and Teh, Y. W. Augmented neural odes. 2019.

Farhi, E. and Neven, H. Classification with Quantum Neural Networks on Near Term Processors. *arXiv e-prints*, art. arXiv:1802.06002, February 2018.

Finzi, M., Bondesan, R., and Welling, M. Probabilistic numeric convolutional neural networks. 2020.

Gardner, J. R., Pleiss, G., Bindel, D., Weinberger, K. Q., and Wilson, A. G. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 2018.

Harrow, A. W., Hassidim, A., and Lloyd, S. Quantum algorithm for linear systems of equations. *Physical Review Letters*, 103(15), Oct 2009. ISSN 1079-7114. doi: 10.1103/physrevlett.103.150502. URL http://dx.doi.org/10.1103/PhysRevLett.103.150502.

Hudson, R. When is the wigner quasi-probability density non-negative? *Reports on Mathematical Physics*, 6(2):249 – 252, 1974. ISSN 0034-4877. doi: https://doi.org/10.1016/0034-4877(74)90007-X. URL http://www.sciencedirect.com/science/article/pii/003448777490007X.

Huggins, W., Patil, P., Mitchell, B., Whaley, K. B., and Miles Stoudenmire, E. Towards quantum machine learning with tensor networks. *Quantum Science and Technology*, 4(2):024001, Apr 2019. doi: 10.1088/2058-9565/aaea94.

Kerenidis, I. and Prakash, A. Quantum recommendation systems, 2016.

Killoran, N., Bromley, T. R., Arrazola, J. M., Schuld, M., Quesada, N., and Lloyd, S. Continuous-variable quantum neural networks. *Physical Review Research*, 1(3):033063, 2019.

Lau, H.-K., Pooser, R., Siopsis, G., and Weedbrook, C. Quantum machine learning over infinite dimensions. *Phys. Rev. Lett.*, 118:080501, Feb 2017. doi: 10.1103/PhysRevLett.118.080501. URL https://link.aps.org/doi/10.1103/PhysRevLett.118.080501.

Li, S. C.-X. and Marlin, B. A scalable end-to-end gaussian process adapter for irregularly sampled time series classification, 2016.

Li, S. C.-X. and Marlin, B. M. Classification of sparse and irregularly sampled time series with mixtures of expected gaussian kernels and random features. In *UAI*, pp. 484–493, 2015.

Lloyd, S. and Braunstein, S. L. Quantum computation over continuous variables. *Physical Review Letters*, 82(8):1784–1787, Feb 1999. ISSN 1079-7114. doi: 10.1103/physrevlett.82.1784. URL http://dx.doi.org/10.1103/PhysRevLett.82.1784.

Maron, H., Ben-Hamu, H., Shamir, N., and Lipman, Y. Invariant and equivariant graph networks. *arXiv preprint arXiv:1812.09902*, 2018.

Massaroli, S., Poli, M., Park, J., Yamashita, A., and Asama, H. Dissecting neural odes. 2021.

Nielsen, M. and Chuang, I. *Quantum Computation and Quantum Information*. Cambridge Series on Information and the Natural Sciences. Cambridge University Press, 2000. ISBN 9780521635035. URL https://books.google.co.uk/books?id=aai-P4V9GJ8C.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019.

Pinkus, A. Approximation theory of the mlp model in neural networks. *Acta Numerica*, 8:143–195, 1999. doi: 10.1017/S0962492900002919.

Rasmussen, C., Williams, C., Press, M., Bach, F., and (Firm), P. *Gaussian Processes for Machine Learning*. Adaptive computation and machine learning. MIT Press, 2006. ISBN 9780262182539. URL https://books.google.nl/books?id=Tr34DwAAQBAJ.

Reck, M., Zeilinger, A., Bernstein, H. J., and Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73:58–61, Jul 1994. doi: 10.1103/PhysRevLett.73.58. URL https://link.aps.org/doi/10.1103/PhysRevLett.73.58.

Rezende, D. J., Racanière, S., Higgins, I., and Toth, P. Equivariant hamiltonian flows, 2019.

Sakurai, J. and Napolitano, J. *Modern Quantum Mechanics*. Cambridge University Press, 2017. ISBN 9781108422413. URL https://books.google.nl/books?id=010yDwAAQBAJ.

Shen, Y., Harris, N. C., Skirlo, S., Prabhu, M., Baehr-Jones, T., Hochberg, M., Sun, X., Zhao, S., Larochelle, H., Englund, D., et al. Deep learning with coherent nanophotonic circuits. *Nature Photonics*, 11(7):441, 2017.

Simon, R., Sudarshan, E. C. G., and Mukunda, N. Gaussian-wigner distributions in quantum mechanics and optics. *Phys. Rev. A*, 36:3868–3880, Oct 1987. doi: 10.1103/PhysRevA.36.3868. URL https://link.aps.org/doi/10.1103/PhysRevA.36.3868.

Steinbrecher, G. R., Olson, J. P., Englund, D., and Carolan, J. Quantum optical neural networks. *npj Quantum Information*, 5(1):1–9, 2019.

Toth, P., Rezende, D. J., Jaegle, A., Racanière, S., Botev, A., and Higgins, I. Hamiltonian generative networks, 2020.

Verdon, G., Pye, J., and Broughton, M. A Universal Training Algorithm for Quantum Deep Learning. *arXiv e-prints*, art. arXiv:1806.09729, June 2018.

Wang, Z., Yan, W., and Oates, T. Time series classification from scratch with deep neural networks: A strong baseline, 2016.

Watrous, J. *The Theory of Quantum Information*. Cambridge University Press, 2018. ISBN 9781107180567. URL https://books.google.nl/books?id=GRNSDwAAQBAJ.

Weedbrook, C., Pirandola, S., García-Patrón, R., Cerf, N. J., Ralph, T. C., Shapiro, J. H., and Lloyd, S. Gaussian quantum information. *Reviews of Modern Physics*, 84(2):621–669, May 2012. ISSN 1539-0756. doi: 10.1103/revmodphys.84.621. URL http://dx.doi.org/10.1103/RevModPhys.84.621.

Zhao, Z., Fitzsimons, J. K., and Fitzsimons, J. F. Quantum-assisted gaussian process regression. *Physical Review A*, 99(5):052331, 2019.