A Neural Network Approach to Quantum Chemistry, Finding the Ground State of Hydrogen Based Systems

Jean Michel Sellier[*a], Louis Maestrati[a]

[a]MILA, Montréal, Québec, Canada
[*]jeanmichel.sellier@gmail.com

Abstract

Recently, an alternative approach to represent the state of a quantum system has been suggested, which is based on the use of neural networks. In more details, it consists of a general method which is, in practice, specialized to the case of a spinglass which state is represented by restricted Boltzmann machines. This promising method has been subsequently generalized to find the ground state of any given quantum system provided with its corresponding Hamiltonian and simple validation tests have been presented. In this paper, a more complex validation testbed is considered consisting of two archetypal chemical systems: the hydrogen atom and the H_2 molecule. Both correspond to numerically difficult problems due to the well-known singularity in the hydrogen Coulomb potential at the position of the nucleus. In spite of those difficulties, such generalization correctly finds the ground state of a three-dimensional hydrogen atom without the need for any prior (physics) knowledge. Moreover, when applied to the H_2 molecule, this new method is able to correctly tackle this many-body problem as well. The authors believe that these promising results clearly show the potential to pave the way towards very different approaches to study quantum chemical systems.

Keywords: Quantum mechanics, Quantum Chemistry, Machine learning, Neural networks, Simulation of quantum systems

1. Introduction

The hydrogen atom is, by far, the simplest possible atom found in nature: it consists of only one proton and one electron. In spite of its apparent simplicity, though, it has had a role of paramount importance in the development of Physics, providing practical instances of quantum mechanical effects and, therefore, offering a powerful testbed. Moreover, two hydrogen atoms can easily bind by moving their nuclei at a distance equal to 0.7416 Å and by arranging their electrons to achieve the minimum energy. This forms a dihydrogen molecule, denoted as H_2, which dissociation energy is equal to 4.476 eV and which, once again, represents an important testbed for novel methods to be validated. As very well stated in [1]: "*It can be said without fear of contradiction that the*

two-electron bond is the single most important stereoelectronic feature of chemistry".

Recently, one of the authors of this paper (JMS) has suggested a new approach to find the ground state of a quantum system, which is based on the use of neural networks to represent its wave function [2]. While this method exploits the typical generalization capabilities of neural networks, as it reduces the complexity of the search for the ground state, it also keeps exactly the same explanatory capacities typical of Physics and Chemistry theories. In other words, this method is not affected by the problem of interpretability at all. There also are strong indications showing that it might not be affected by the sign problem as well, since the integrals involved in the mathematical expression of the energy can be calculated analitically. Finally, this method is embarassingly parallelizable, which certainly can help with the simulation of relatively big systems.

Previously, as a preliminary proof of concept, this new method has been validated against relatively simple single- and many-body quantum systems for which exact solutions are known. Within this context, the purpose of this paper is to continue such validation with a special focus on chemical systems. In particular, we study the applicability of such novel approach to two archetypal systems issued from the field of quantum chemistry, i.e. the hydrogen atom and the diatomic (homogeneous) H_2 molecule. In particular, finding the ground state of the hydrogen atom represents a quite challenging validation test because of the well known singularity appearing in the Coulomb potential generated by its nucleus. The same can be said about the H_2 molecule which also involves typical quantum phenomena such as tunnelling and exchange-correlation effects, which are known to require reliable methods to be properly tackled.

The paper is organized as follows. In the following section, we sketch the suggested method for the sake of completeness. Then, the method is applied to the cases of the hydrogen atom and the H_2 molecule. Finally, some conclusions and comments on possible future works are provided.

2. Neural Network Representation of Quantum States

The main problem of this work consists in finding the ground state of any kind of quantum system provided with an Hamiltonian. For the sake of simplicity, in the specific case of one electron in a three-dimensional space, this corresponds to obtaining the minimum eigenenergy of the following eigenproblem (along with its corresponding eigenfunction):

$$\hat{H}\psi\left(\mathbf{x}\right) = E\psi\left(\mathbf{x}\right), \tag{1}$$

where $\mathbf{x} = (x, y, z)$ are the spatial coordinates and:

$$\hat{H} = \frac{\hat{p}^2}{2m} + qV(\mathbf{x}), \tag{2}$$

(with $\hat{p} = -i\hbar\frac{\partial}{\partial x}$ being the usual momentum operator). The solutions of this equation can be written as a set of couples (E_n, ψ_n), ordered by eigenenergies, with the ground state completely described by the couple (E_0, ψ_0).

In order to reduce the complexity of this typical numerical problem, it is common practice to explicit a wave function in terms of a series defined over a orthonormal basis $\{\phi_l(\mathbf{x}), l \in [1, \ldots, N]\}$, usually chosen according to the problem at hand. In details, this series reads:

$$\psi(\mathbf{x}) = \sum_{l=1}^{N} a_l \phi_l(\mathbf{x}), \tag{3}$$

for some arbitrary integer N fixed before hand, and where the coefficients $\{a_l\}_{l \in [1,\ldots,N]}$ represent the new unknowns of the problem. This amounts to express the wave function simply by means of constants, i.e. $\psi = \psi(\{a_l\}_{l \in [1,\ldots,N]})$. On one hand, this highly simplifies the computational problem but, on the other hand, it introduces important limitations since we are now in front of a new problem which requires to choose:

- an orthonormal basis,

- an appropriate integer N which truncates this basis.

This actually introduces an extra layer of complexity for the user since one is forced to select hyperparameters in an infinite set of possible alternatives.

Now, the (well-known) universal approximation theorem [3], [4], [5] guarantees that a single-hidden-layer feedforward neural network (containing only a finite number of non-linear units) can approximate any continuous function defined over a real-valued n-dimensional compact set (under mild assumptions on the activation function). Therefore, one could exploit such specific networks to represent a wave function, as recently suggested in [6] (specifically applied to the case of restricted Boltzmann machines to predict the dynamics of a spinglass). More particularly, this novel approach leads to a new kind of discretization of the quantum state ψ which is now a function of the network parameters, known as weights and biases, and which completely defines it, i.e.

$$\psi = \psi(\mathbf{W}),$$

where $\mathbf{W} = \{w_{i,j}, b_i\}$ is the set of weights and biases defining the network. Although this theorem does not specify the actual number of parameters required to accurately approximate a function, the Reader should note that the selection process is now defined in terms of the number of hidden units and their activation function, which is relatively easier than the above approach (3). Moreover, the simplicity of the network structure, allows the representation of one-, two-, or three- dimensional ground states with relatively small computational resources. For the above reasons, in this work we limit ourselves to feedforward neural networks consisting of an input layer, one hidden layer with

non-linear units and one output layer with linear units. The input layer receives the coordinates $\mathbf{x} = (x, y, z)$ and the output layer returns two real values corresponding respectively to the real and the complex part of the function $\psi = \psi(\mathbf{x})$ (see Fig. 1 for a visual representation of the network).

It is relatively easy to show, by algebraically manipulating equ. (1), that the total energy of a system can be expressed as:

$$E = \frac{\int_\Omega d\mathbf{x}\, \psi^*(\mathbf{x})\hat{H}\psi(\mathbf{x})}{\int_\Omega d\mathbf{x}\, \psi^*(\mathbf{x})\psi(\mathbf{x})}, \tag{4}$$

where Ω represents the spatial (finite) domain of the system. Since, in this rather novel situation, the wave-function is represented by a neural network, it is now a function of both position and weights, i.e. $\psi = \psi(\mathbf{x}, \mathbf{W})$, and thus the energy solely depends on the weights as well, i.e. $E = E(\mathbf{W})$. At this point, we are able to define a target or fitness function which embeds the total energy (4) and forces both closed boundary and normalization conditions:

$$\begin{aligned} f &= f(\mathbf{W}) \\ &= E(\mathbf{W}) + \lambda \int_{\partial\Omega} d\mathbf{x}\, \psi^*(\mathbf{x}, \mathbf{W})\psi(\mathbf{x}, \mathbf{W}) + |\int_\Omega d\mathbf{x}\, \psi^*(\mathbf{x}, \mathbf{W})\psi(\mathbf{x}, \mathbf{W}) - 1| \end{aligned} \tag{5}$$

where $\lambda > 0$ is a real constant and $\partial\Omega$ represents the boundary of the domain Ω. This naturally ensures that the wave function is decreased at the boundaries of the spatial domain, and that its norm tends to unity.

On the remaining question of minimizing the quantity (5), one has to choose a suitable method to find out a well suited set of parameters $\mathbf{W}$. A classical machine learning approach would be to differentiate our energy according to the set of weights $\mathbf{W}$ and then to apply a stochastic gradient descent (backpropagation) [7]. However, one can easily notice that this method would be affected, in this particular situation, by two issues:

- Due to the denominator of (4), the expression is non-fully-differentiable and could lead to an unbounded derivative. Using gradient descent could therefore lead to unbounded quantities.

- Since the expression (4) is non-convex, there is no guarantee to find a global minimum of the function by means of the gradient descent method.

Indeed, finding the accurate $\mathbf{W}$, representing the ground state, requires a more exploratory and derivative-free method. On the other hand, Monte Carlo sampling techniques [8] have proven their efficiency in the Quantum Monte Carlo theory [9] and could therefore be an alternative approach. However, the recent Covariance Matrix Adaptation Evolution Strategy (CMA-ES) has shown, although empirically, its ability to be more efficient in terms of convergence towards the global optimum [10]. Thus, CMA-ES samplings of sets $\mathbf{W_i}$ are performed and evaluated by our network, until a global minimum of the energy E_0 is eventually reached.

We can now proceed with the validation of this suggested approach to the case of typical systems coming from the field of quantum chemistry, i.e. the hydrogen atom and the H_2 molecule.

3. Numerical Experiments

In this section, we focus on a further and more realistic validation of our suggested method in the context of quantum chemistry. More specifically, we look for the ground state of the hydrogen atom and the H_2 molecule. These systems are well understood in the communities of quantum physics and chemistry and can, therefore, be considered as classical testbeds. In both cases, we consider those systems in a three-dimensional space and restrict ourselves to the Born-Oppenheimer approximation [11], although the Reader should note that this does not represent a limitation for the method.

In practice, any system containing the H atom is affected by the typical difficulties of numerically treating the corresponding Coulomb potential, which contains a singularity, and reads:

$$V(r) = -\frac{e^2}{4\pi\epsilon_0 r} \qquad (6)$$

where r is the distance between the electron and the nucleus, ϵ_0 is the vacuum permittivity and e is the elementary charge. In the case of the H_2 molecule, the system is obviously affected by the same numerical issues but, carries an extra layer of complexity, coming from the interplay between the electron-ion attraction and the electron-electron repulsion.

Hydrogen atom. The hydrogen atom represents one of the few quantum chemical system provided with an analytical solution (in the non-relativistic case). In spherical coordinates, the exact solution reads:

$$\psi_{n,\ell,m}(r,\vartheta,\varphi) = \sqrt{\left(\frac{2}{na_0^*}\right)^3 \frac{(n-\ell-1)!}{2n(n+l)!}} e^{-\rho/2} \rho^\ell L_{n-\ell-1}^{2\ell+1}(\rho) Y_l^m(\vartheta,\varphi), \qquad (7)$$

where $\rho = \frac{2r}{na_0^*}$, (r,ϑ,φ) are the spherical coordinates, (n,ℓ,m) are the quantum numbers, a_0^* is the reduced Bohr radius, $L_{n-\ell-1}^{2\ell+1}(\rho)$ is a generalized Laguerre polynomial of degree $(n-\ell-1)$ and order $(2\ell+1)$ and $Y_\ell^m(\vartheta,\varphi)$ is a spherical harmonic function of degree ℓ and order m. The corresponding eigenenergy reads:

$$E_n = -\frac{E_0}{(n+1)^2}, \qquad (8)$$

where $E_0 \simeq 13.606$ eV. The ground state wave function of this atom is readily obtained from (7) with the quantum numbers $(n,\ell,m) = (1,0,0)$, i.e.

$$\psi_0(r) = \frac{1}{\sqrt{\pi a_0^{*3}}} e^{-\frac{r}{a_0^*}}, \qquad (9)$$

with an energy equal to $E_0 \simeq -13.606$ eV.

In this context, our validation test consists in comparing the exact solution (9) with the one obtained by means of our suggested approach in a (finite) domain represented by a rectangular box of dimensions $(L_x, L_y, L_z) = (1, 1, 1)$ nm (the H nucleus is positioned in the center). In Fig. 2, we show the probability density obtained with the neural network approach compared to the theoretical solution, as a cut in the x-direction ($y = \frac{L_y}{2}$ and $z = \frac{L_z}{2}$). A good agreement is observed. In Fig. 3, we obtain the same accuracy for the cuts in the y- (left-hand side) and z-directions (right-hand side). Finally, Fig. 4 shows three different planar cuts (top left: $x = 0.5$ nm, top right: $y = 0.5$ nm, bottom left: $z = 0.5$ nm). In particular, this figure clearly demonstrates that the spherical symmetry is fully respected and, thus, it is clear that *the network is actually able to reconstruct the wave function from scratch without any prior knowledge.* The corresponding ground state energy is computed using equation (4) and is equal to -13.4101 eV, which is, again, in good agreement with the theory.

The diatomic H_2 molecule. Due to its complexity, the H_2 molecule does not have an exact solution which could be utilized for comparison purposes. However, numerical approximations can be obtained. For instance, the density functional theory (DFT), which consists in solving a system of coupled single-electron equations can provide such approximated solution. In practice, this reduces to solve the Kohn-Sham system of equations, based on the assumption that the electron-electron interaction can be reliably approximated by an exchange-correlation potential [12]. Many possible choices are possible for such term and, in this work, we used the Perdew-Wang functional (PW91) [13]. Thus, in order to compute an approximated solution, we utilized the basis set TZ2P (triple zeta with 2 polarization functions) to represent the wave function and solved the two (coupled) equations by means of a deterministic method. Then, we solved the same set of equations by representing the ground state of the two electrons by means of two separated neural networks with $\mathbf{W_1}$ and $\mathbf{W_2}$ their respective sets of weights. Finally, CMA sampling of the two sets were performed as usual to minimize the corresponding energies E_1 and E_2 from which one recovers the total energy. The results of such numerical experiments are reported in Fig. 5 in function of the (varying) distance between the two nuclei. Once again, one can observe that the two methods provide results which are in very good agreement.

4. Conclusions and Future Works

In this paper, we presented further validations of a newly suggested method based on the representation of the ground state of quantum systems by means of neural networks, and the embarassingly parallelizable and derivative-free CMA-ES. More specifically, we applied this new approach to archetypal systems coming from the field of quantum chemistry, i.e. the hydrogen atom and the H_2 molecule which are both affected by the singularity in the Coulomb potential of the H atom. We have shown that this new method is able to tackle such
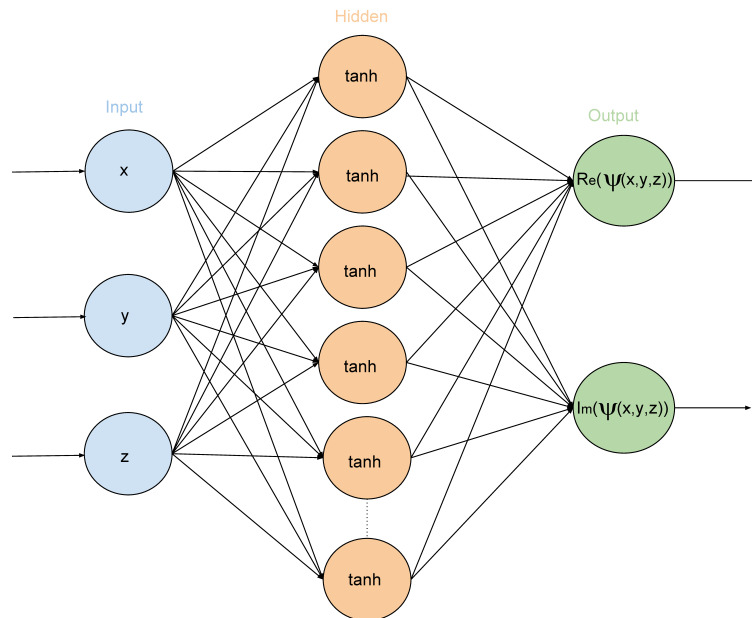
Figure 1: Architecture of the feedforward neural network used in this paper, 'tanh' units are examples of activation functions that can be used in the hidden layer.

numerically peculiar situations and provides accurate ground states and energies. However, in spite of these preliminary encouraging results, many aspects remain to be investigated. For instance, it would be very important to explore other networks architecture, especially deep ones which might bring further advantages. Different sampling algorithms could be explored as well. This will be the subject of our next future efforts.

References

[1] F.A. Cotton, D.G. Nocera, The Whole Story of the Two-Electron Bond, with the δ Bond as a Paradigm, Acc. Chem. Res. 33, p. 483, (2000).

[2] J.M. Sellier, Could the Hilbert Space Be a Smaller Place? A Neural Network Perspective, submitted to the Journal of Computational Physics, preprint available on arXiv:1902.08577, (2019).
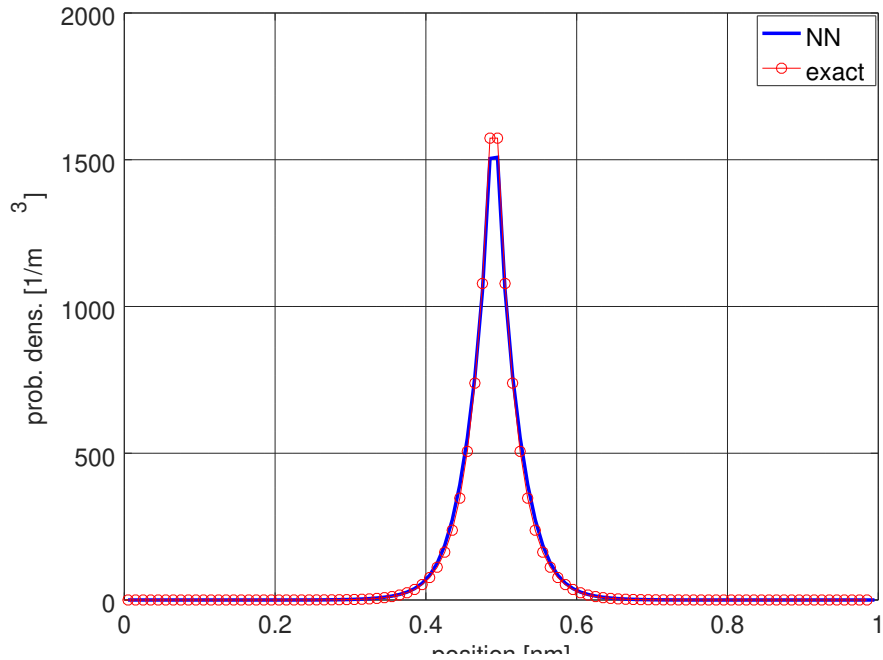
Figure 2: Probability density obtained with the neural network, compared to the theoretical values as a cut in the x-direction ($y = \frac{L_y}{2}$ and $z = \frac{L_z}{2}$). The nucleus of the atom H is centered on the position $x = y = z = 0.5nm$, a perfect symmetry of the curve around the nucleus position is obtained. This confirms our theoretical expectations and demonstrates an ability of the network to understand by itself intrinsic properties of the quantum system.
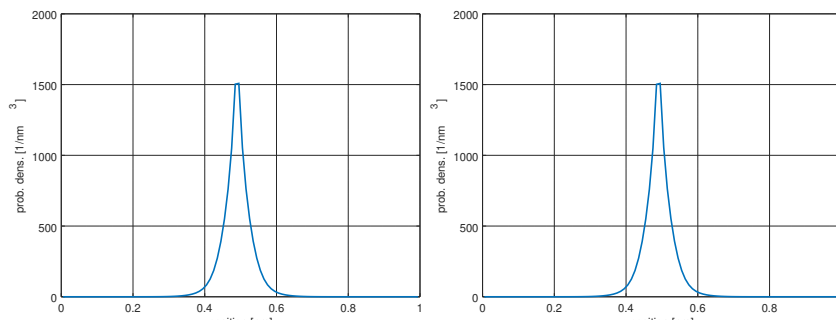

Figure 3: As shown in the figure 2, probability density of the hydrogen electron obtained with the neural network, for the cuts in the y- and z-direction (left- and right-hand side respectively). We obtain the same curves as in Fig. 2 and this demonstrates an ability of the network to understand by itself the **spherical** symmetry of the problem.
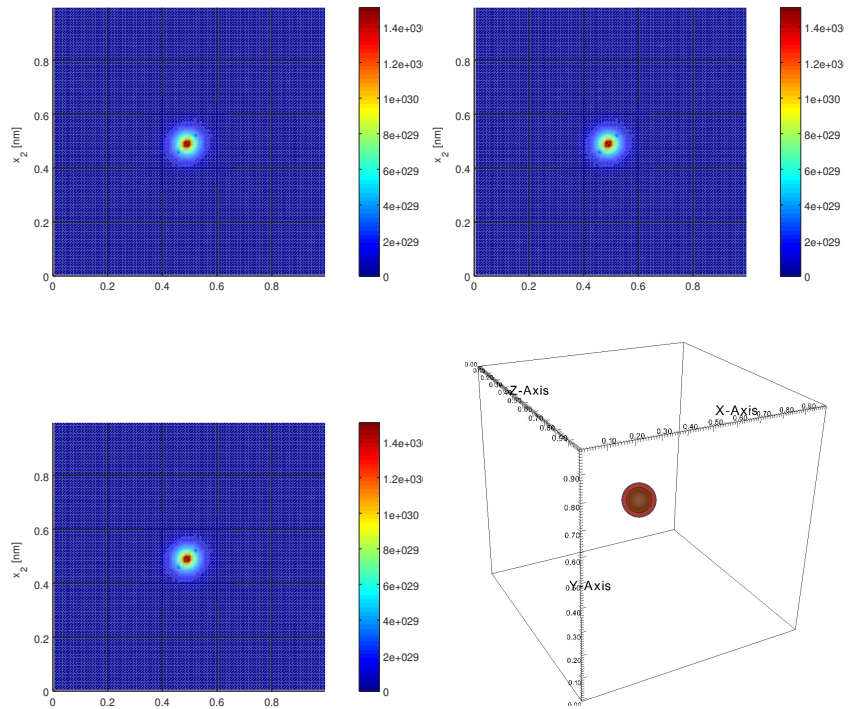
8

Figure 4: Cut of the three-dimensional ground state for the atom H computed by means of a feedforward neural network. The cuts are performed on the planes x = 0.5 nm (top left), y = 0.5 nm (top right) and z = 0.5 nm (bottom left) respectively. The spherical symmetry is very well respected.
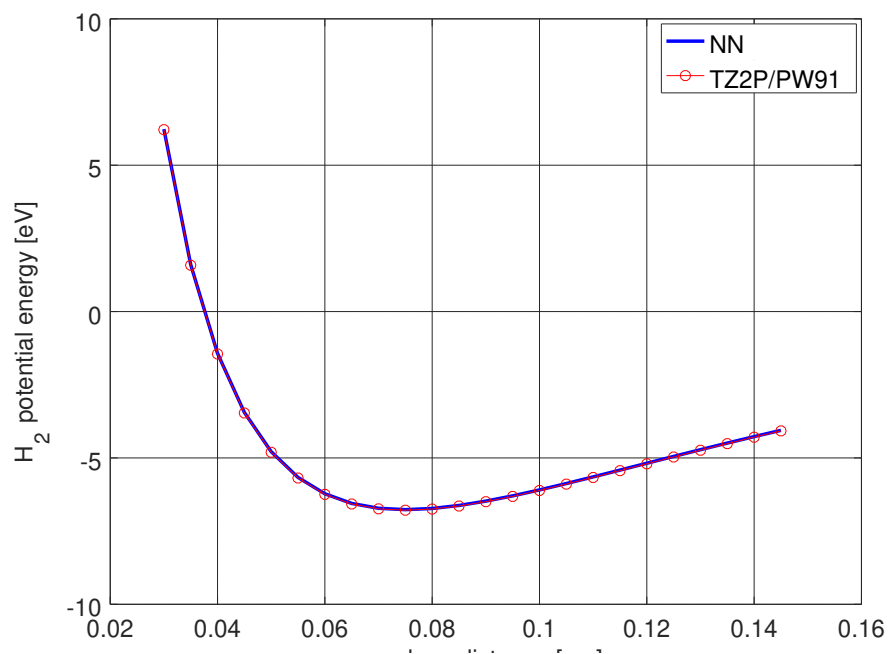
9

Figure 5: H_2 potential energy found with the DFT method and with our neural network method. The energy of the two-body system evolves with the distance between the two atoms. The two methods provide the same minimum, which corresponds to the stable state of the system.

[3] A.N. Kolmogorov, On the Representation of Continuous Functions of Several Variables by Superposition of Continuous Functions of one Variable and Addition, Doklady Akademii. Nauk USSR, 114, pp. 679-681, (1957).

[4] G. Cybenko, Approximations by Superpositions of Sigmoidal Functions, Mathematics of Control, Signals, and Systems, 2(4), pp. 303314, (1989).

[5] K. Hornik, Approximation Capabilities of Multilayer Feedforward Networks, Neural Networks, 4(2), pp. 251257, (1991).

[6] G. Carleo, M. Troyer, Solving the Quantum Many-body Problem with Artificial Neural Networks, Science 355, pp. 602606, (2017).

[7] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, The MIT Press, (2016).

[8] I.T. Dimov, Monte Carlo Methods for Applied Scientists, World Scientific, Singapore, (2007).

[9] B.L. Hammond, J.W.A. Lester, and P.J. Reynolds. Monte Carlo Methods in Ab Initio Quantum Chemistry. World Scientific, Singapore, (1994).

[10] N. Hansen, The CMA Evolution Strategy: a Comparing Review, Towards a New Evolutionary Computation. Studies in Fuzziness and Soft Computing, vol 192, Springer, (2006).

[11] M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln, Annalen der Physik, (1927).

[12] W. Kohn, L.J. Sham Self-Consistent Equations Including Exchange and Correlation Effects, Physical Review 140, no. 4A, (1965).

[13] J. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, Atoms, Molecules, Solids, and Surfaces: Applications of the Generalized Gradient Approximation for Exchange and Correlation, Phys. Rev. B 46, p. 6671, (1992).