

5-2018

Solving Atomic Wave Functions Using Artificial Neural Networks

Gaurav Gyawali
University of New Orleans

Follow this and additional works at: https://scholarworks.uno.edu/honors_theses



Part of the [Physics Commons](#)

Recommended Citation

Gyawali, Gaurav, "Solving Atomic Wave Functions Using Artificial Neural Networks" (2018). *Senior Honors Theses*. 104.

https://scholarworks.uno.edu/honors_theses/104

This Honors Thesis-Unrestricted is protected by copyright and/or related rights. It has been brought to you by ScholarWorks@UNO with permission from the rights-holder(s). You are free to use this Honors Thesis-Unrestricted in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you need to obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/or on the work itself.

This Honors Thesis-Unrestricted has been accepted for inclusion in Senior Honors Theses by an authorized administrator of ScholarWorks@UNO. For more information, please contact scholarworks@uno.edu.

**Solving Atomic Wave Functions Using Artificial
Neural Networks**

An Honors Thesis

Presented to

the Department of Physics

of the University of New Orleans

In Partial Fulfillment

of the Requirements for the Degree of Bachelor of Science,

with University High Honors and Honors in Physics

by

Gaurav Gyawali

May 2018

Acknowledgement

First and foremost, I am indebted to Dr. Steven Rick for his incessant support not only in my research endeavors but also in various areas of my academic pursuit. Working with his research group has been a key part of my undergraduate learning experience, and his unwavering composure and curiosity for physical sciences have been a constant source of my inspiration.

It was Dr. Juliette Ioup with whom I took my baby steps in the realm of quantum mechanics. A couple years after her PHYS 4401 course, I am now able to attempt writing a thesis in this confounding subject. Without her guidance and compassion, I'd never have mustered enough courage to embark on this arduous path. Dr. Kevin Stokes, Dr. Ralph Saxton, Dr. Joel Webb, and the entire Physics and Mathematics Departments have been extremely generous with their time and energy to impart to me a solid foundation in these subjects. I am also immensely grateful for the financial support provided by the UNO College of Science through COSURP that made this research possible.

My mentors, Dr. Monika VanGordon, Dr. Arjun Sharma, and Mr. Bryan Raubenolt- thank you for sharing your wisdom with me on a daily basis. Thanks to the University Honors Program for always been very supportive on my endeavors. I'd like to dedicate every single bit of this accomplishment to my family, who have worked very hard to offer me this opportunity. Special thanks to my friends and my roommates for always being there for me during the highs and the lows.

Contents

Acknowledgement	i
Abstract	iv
List of Figures	v
List of Tables	vi
1 Introduction	1
1.1 Background	1
1.2 Overview	3
2 Stationary States of Hydrogen	4
2.1 Separation of spatial and time variables	4
2.2 TISE in Spherical Coordinates	5
2.3 Angular Solution	6
2.4 Radial Solution	9
3 Hartree-Fock SCF Method	10
3.1 Hartree SCF Method	10
3.2 Slater Determinant	11
3.3 Hartree-Fock SCF Method	12
4 Variational Wave Function	13
4.1 Artificial Neural Network	13

4.2	Radial Basis Function Neural Network	14
4.3	Variational Representation of an Atomic Wave Function Using ANN	16
5	Optimization Scheme	20
5.1	Variational Principle	20
5.2	Variational Monte Carlo (VMC) Integration	20
5.3	Metropolis Algorithm	22
5.4	Stochastic Reconfiguration (SR)	23
5.4.1	SR Operators for Simple Harmonic Oscillator (SHO)	23
5.4.2	SR Operators for Helium Wave Function	24
6	Results	26
6.1	Simple Harmonic Oscillator (SHO)	26
6.2	Helium Atom	27
6.3	Lithium Atom	28
	Conclusion	30
	Bibliography	31

Abstract

Carleo and Troyer [3] have recently pointed out the possibility of solving quantum many-body problems by using Artificial Neural Networks (ANN). Their work is based on minimizing a variational wave function to obtain the ground states for various spin-dependent systems. This work is primarily focused on developing efficient method using ANN to solve the ground state wave function for atomic systems. We have developed a theoretical groundwork to represent the wave function of a many-electron atom by using artificial neural network while still preserving its antisymmetric property. By using the Metropolis algorithm, Variational Monte Carlo (VMC), and Stochastic Reconfiguration (SR) methods for minimization, we were able to obtain a highly accurate ground state wave function for the He atom. To verify our optimization algorithm, we reproduced the results for the ground state of a three dimensional Simple Harmonic Oscillator (SHO) given by Teng [18].

List of Figures

4.1	A feedforward Artificial Neural Network(ANN)	13
4.2	Notation for various neural layers	14
6.1	Convergence for neural networks with various N_h	27

List of Tables

2.1	Some of the low order spherical harmonics	9
6.1	Ground state energy for 3d SHO for various number of hidden nodes (Nh)	26

1. Introduction

1.1 Background

The world of tiny molecules and atoms can be described by Quantum Mechanics. Owing to the inherent uncertainty in the quantum world that persists beyond the precision of the measurement techniques, their description is given by a wave function Ψ rather than a fixed set of numerical values. The wave function contains all the information about a quantum system, which can either be a single particle or a system of many such particles. This wave function can be projected into various bases (i.e. position, momentum, spin) to obtain the required probabilistic description. Such a wave function can be obtained by solving the famous **Schrödinger equation**. It is a second order differential equation whose complexity increases exponentially with an increase in the number of particles in the system. Not surprisingly, most of the systems of interest to the scientific community are many-body, which are impossible to solve analytically. Thus, scientists rely on a variety of numerical techniques such as the variational methods. Most of these methods are based on an efficient representation of a many-body system by an appropriate function of many variables and its optimization to achieve the ground state. Optimization of too many variational parameters is still beyond the capability of the state-of-the-art computational resources, thereby requiring improved methods to solve the **quantum many-body problems**.

An **Artificial Neural Network (ANN)** is a widely used technique in the ma-

chine learning community. A vast oversimplification of a biological neural network, ANN has been very successful at learning the results to a given set of input variables. ANNs have been widely used by the computer science community in a variety of problems such as image and speech recognition, data prediction, and gaming. A popular example in today's context is AlphaGo [1], a computer program developed by Google, that was able to defeat the world's best player Ke Jie on a board game called Go. The varied application of ANN suggests its versatility in representing a function of many variables with a wide range of constraints, and its application to solve the quantum many-body problems has recently come to the attention of quantum scientists. In a paper by Lagaris et al. [2], a variational solution in the form of an ANN was assumed, and solved for simple systems such as Morse potential and an anharmonic potential. More recently, the efficiency resulting from the use of ANN in solving the quantum many-body problems has been amply demonstrated in a paper by Carleo and Troyer[3]. They introduced a representation of the spin-only dependent wave function in terms of an ANN specified by a set of variational parameters and used a stochastic framework for reinforcement learning of the parameters, allowing for the best possible representation of both ground state and time-dependent physical states of quantum Hamiltonians such as Ising and Heisenberg.

In this work, we have attempted to expand this technique to solve for the electronic wave functions of multi-electron atoms. **Density Functional Theory (DFT)** and **Hartree-Fock**[5, 6], also known as Self Consistent Field (SCF), methods are being widely used to obtain the wave functions for atoms and molecules, but these methods are limited by computational resources for large systems. We have used a variational expression for the wave function consisting of four parts:

1. Neural network (Radial)
2. Exponential decay (Radial)

3. Spherical harmonics (Angular)

4. Spin part

The variational parameters are optimized by using the **Stochastic Reconfiguration** method [7] based on the variance minimization of the local energy. We calculated the expectation values for local energy as well as the operators associated with the variational parameters using the **Variational Monte Carlo (VMC)**[8]. The spatial samples for implementing the VMC algorithm were obtained stochastically through a simple **Metropolis** algorithm[9]. The details on the variational representation of the atomic wave function as well as the stochastic optimization technique will be discussed in more detail in chapter 5.

1.2 Overview

A brief introduction to the quantum many-body problems and our motivation for using ANN to solve these problems has been given in the preceding section. This section presents an overview of the structure of this thesis.

Chapter 2 will review the Schrödinger equation in three dimensions and its solutions to a Coulombic potential to obtain the stationary states of a hydrogen atom. Chapter 3 will briefly cover the anti-symmetrization of the wave function for a fermionic system by using **Slater determinants** and the Hartree-Fock procedure to solve for the electronic wave function for atoms. Chapter 4 will discuss ANN and its application to our variational representation of the atomic wave function. Chapter 5 will present the optimization scheme and the description of the various techniques used. In the last chapter, some of the results obtained from our preliminary calculations will be discussed.

2. Stationary States of Hydrogen

2.1 Separation of spatial and time variables

Schrödinger equation can be expressed in its most general form as follows:

$$\hat{H}\Psi(\mathbf{r}, t) = \hat{E}\Psi(\mathbf{r}, t) \quad (2.1)$$

where \hat{H} is the Hamiltonian operator and $\hat{E} = i\hbar\frac{\partial}{\partial t}$ is the energy operator. The wave function Ψ is a function of both space (\mathbf{r}), and time (t). Expressing the Hamiltonian as the sum of potential energy (V) and the kinetic energy ($T = -\frac{\hbar^2}{2m}\nabla^2$), we get the following Time Dependent Schrödinger Equation (TDSE):

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{r}, t) + V\Psi(\mathbf{r}, t) = i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{r}, t) \quad (2.2)$$

Using **separation of variables**, this equation can be separated into the spatial and time parts. The solution to the time part is simple and is given by:

$$C \exp\left(\frac{-iEt}{\hbar}\right) \quad (2.3)$$

where C is a proportionality constant and E is another constant used for the separation of variables. The spatial part depends on the potential, and the solution is unique for a specific potential function ($V(\mathbf{r})$). The following equation comprising only the spatial variables is known as the Time Independent Schrödinger

Equation(TISE):

$$-\frac{\hbar^2}{2m}\nabla^2\psi + V\psi = E\psi \quad (2.4)$$

where ψ is the time independent (spatial) wave function.

2.2 TISE in Spherical Coordinates

The potential energy between the electron and the proton in the H atom is Coulombic which is spherically symmetric. Thus, it makes more sense to proceed with change of coordinates from the usual rectilinear to spherical. The Laplacian operator (∇^2) in spherical coordinates is given as follows:

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \left(\frac{\partial^2}{\partial \phi^2} \right) \quad (2.5)$$

Using this operator, we can write the TISE in the following form:

$$-\frac{\hbar^2}{2m}\nabla^2(r, \theta, \phi)\psi(r, \theta, \phi) + V(r)\psi(r, \theta, \phi) = E\psi(r, \theta, \phi) \quad (2.6)$$

Again, we evoke the separation of variables to separate this equation into angular and radial parts:

$$\psi(r, \theta, \phi) = R(r)Y(\theta, \phi) \quad (2.7)$$

Substituting this relation into equation 2.6 gives:

$$-\frac{\hbar^2}{2m} \left[\frac{Y}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial R}{\partial r} \right) + \frac{R}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{R}{r^2 \sin^2 \theta} \left(\frac{\partial^2 Y}{\partial \phi^2} \right) \right] + VRY = ERY$$

Multiplying both sides of this equation by $\frac{-2mr^2}{RY\hbar^2}$, we get:

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{Y \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{Y \sin^2 \theta} \left(\frac{\partial^2 Y}{\partial \phi^2} \right) - \frac{2m}{\hbar^2} V = -\frac{2m}{\hbar^2} E$$

or, $\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{1}{Y \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{Y \sin^2 \theta} \left(\frac{\partial^2 Y}{\partial \phi^2} \right) - \frac{2m}{\hbar^2} (V(r) - E) = 0$

Rearranging the radial part to the Left Hand Side (LHS) and angular part to the Right Hand Side (RHS):

$$\left[\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) - \frac{2mr^2}{\hbar^2} (V(r) - E) \right] = \frac{-1}{Y} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \left(\frac{\partial^2 Y}{\partial \phi^2} \right) \right]$$

Here, the LHS depends on only the radial distance(r) whereas the RHS depends only on angles (θ and ϕ). Since changing either of these variables changes only one side of the equation, each side can be assumed to be a constant. Let this constant be $l(l + 1)$. We are being wise in our selection of this constant because it will simplify the further calculations. Thus, we get the following equations for angular and radial parts:

$$\frac{-1}{Y} \left[\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial Y}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \left(\frac{\partial^2 Y}{\partial \phi^2} \right) \right] = l(l + 1) \quad (2.8)$$

and

$$\left[\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) - \frac{2mr^2}{\hbar^2} (V(r) - E) \right] = l(l + 1) \quad (2.9)$$

2.3 Angular Solution

Again, we employ the separation of variables to separate the angular wave function(Y) into two parts:

$$Y(\theta, \phi) = \Theta(\theta)\Phi(\phi) \quad (2.10)$$

Substituting this expression into equation 2.8 gives:

$$\Phi \sin \theta \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \Theta \frac{d^2 \Phi}{d\phi^2} = -l(l + 1)\Theta\Phi \quad (2.11)$$

Dividing by $\Theta(\theta)\Phi(\phi)$, and rearranging we get:

$$\frac{1}{\Theta} \left[\sin \theta \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) \right] + l(l + 1) \sin^2 \theta = -\frac{1}{\Phi} \frac{d^2 \Phi}{d\phi^2} \quad (2.12)$$

We observe that the LHS depends only on θ and the RHS depends only on ϕ . Using the same argument as before, we can assume that each side is a constant given by m^2 . Again, we are being wise in choosing the constant. Now, we get the following two equations:

$$\frac{1}{\Phi} \frac{d^2\Phi}{d\phi^2} = -m^2 \quad (2.13)$$

and

$$\frac{1}{\Theta} \left[\sin \theta \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) \right] + l(l+1) \sin^2 \theta = m^2 \quad (2.14)$$

The solution to equation 2.13 is given by:

$$\Phi(\phi) = C e^{im\phi} \quad (2.15)$$

where C is a proportionality constant. We are not writing $e^{-im\phi}$ because we will let m have both positive and negative values. Since $e^{im\phi}$ is periodic with period 2π , $e^{im\phi} = e^{im(\phi+2\pi)} \Rightarrow e^{im\phi} = e^{im\phi} e^{i2m\pi}$. Thus, $m = \pm 1, \pm 2, \pm 3, \dots$.

The solution to equation 2.14 is non-trivial, and is given by **associated Legendre's polynomials** as follows:

$$\Theta(\theta) = A P_l^m(\cos \theta) \quad (2.16)$$

where A is a constant of proportionality and $P_l^m(\cos \theta)$ is an associated Legendre's polynomial given by:

$$P_l^m(\cos \theta) = (-1)^m (\sin \theta)^m \frac{d^m}{d(\cos \theta)^m} P_l(\cos \theta) \quad (2.17)$$

where

$$P_l(x) = \frac{1}{2^l l!} \left(\frac{d}{dx} \right)^l (x^2 - 1)^l \quad (2.18)$$

is called **Legendre's polynomial** given by this Rodrigue's formula. For these formulas to make physical sense, $l > 0$, and $P_l^m = 0$ whenever $|m| > l$. Thus, $m = -l, -l + 1, \dots, 0, \dots, l - 1, l$.

In spherical coordinates, the volume element is given by $d^3\mathbf{r} = r^2 \sin \theta dr d\theta d\phi$. We can impose the normalization as follows:

$$\int |\psi|^2 r^2 \sin \theta dr d\theta d\phi = \int |R|^2 r^2 dr \int |Y|^2 \sin \theta d\theta d\phi = 1 \quad (2.19)$$

It is often convenient to separately normalize both the radial part (R) and the angular part (Y) separately so that the overall wave function is normalized also. The normalization condition for angular part is:

$$\int_0^{2\pi} \int_0^\pi |Y|^2 \sin \theta d\theta d\phi = 1 \quad (2.20)$$

The normalized angular wave functions are called **spherical harmonics**, and are given by:

$$Y_l^m(\theta, \phi) = \epsilon \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} e^{im\phi} P_l^m(\cos \theta) \quad (2.21)$$

where,

$$\epsilon = (-1)^m \text{ if } m \geq 0$$

$$\epsilon = 1 \text{ if } m \leq 0$$

Since the angular equation 2.8 does not depend on the potential, spherical harmonics give the general solution to the angular part for spherically symmetric potentials. These are important to our work because our variational wave function consists of spherical harmonics in addition to the neural network and exponential decay radial terms and spin terms.

$Y_0^0 = \left(\frac{1}{4\pi}\right)^{1/2}$	$Y_2^{\pm 2} = \left(\frac{15}{32\pi}\right)^{1/2} \sin^2 \theta e^{\pm 2i\phi}$
$Y_1^0 = \left(\frac{3}{4\pi}\right)^{1/2} \cos \theta$	$Y_3^0 = \left(\frac{7}{16\pi}\right)^{1/2} (5 \cos^3 \theta - 3 \cos \theta)$
$Y_1^{\pm 1} = \mp \left(\frac{3}{8\pi}\right)^{1/2} \sin \theta e^{\pm i\phi}$	$Y_3^{\pm 1} = \mp \left(\frac{21}{64\pi}\right)^{1/2} \sin \theta (5 \cos^2 \theta - 1) e^{\pm i\phi}$
$Y_2^0 = \left(\frac{5}{16\pi}\right)^{1/2} (3 \cos^2 \theta - 1)$	$Y_3^{\pm 2} = \left(\frac{105}{32\pi}\right)^{1/2} \sin^2 \theta \cos \theta e^{\pm 2i\phi}$
$Y_2^{\pm 1} = \mp \left(\frac{5}{8\pi}\right)^{1/2} \sin \theta \cos \theta e^{\pm i\phi}$	$Y_3^{\pm 3} = \left(\frac{35}{64\pi}\right)^{1/2} \sin^3 \theta e^{\pm 3i\phi}$

Table 2.1: Some of the low order spherical harmonics

2.4 Radial Solution

For an H atom, we substitute the Coulombic potential for V in equation 2.9. The solution to this equation is again non-trivial, so only the results will be mentioned here:

$$R_{nl} = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}} e^{-r/na} \left(\frac{2r}{na}\right)^l \left[L_{n-l-1}^{2l+1} \left(\frac{2r}{na}\right) \right] \quad (2.22)$$

where

$$L_p^{q-p}(x) = (-1)^p \left(\frac{d}{dx}\right)^p L_q(x) \quad (2.23)$$

is an **associated Laguerre Polynomial**, and

$$L_q(x) = e^x \left(\frac{d}{dx}\right)^q (e^{-x} x^q) \quad (2.24)$$

is the q^{th} **Laguerre Polynomial**. n stands for the order of the stationary state, and $l = 0, \dots, n-1$ comes from the part we omitted. We can obtain the stationary wavefunctions for the H atom by multiplying equation 2.21 and 2.22, i.e. $\psi_{nlm} = R_{nl} \cdot Y_m^l$

3. Hartree-Fock SCF Method

The Self Consistent Field (SCF) method is based on the assumption that the motion of each electron in the effective field of the (n-1) others is governed by a one-particle Schrödinger equation[10]. From the studies of the motions of electrons in atoms done in early 1920's, it was found possible to reproduce the energy levels of a valence electron if the Bohr orbits of the inner electrons were smeared out into a continuous, spherically symmetrical charge distribution[10]. Using this premise that a bound electron behaves as a charge cloud, with charge density for the k-th electron given by $\rho_k(\mathbf{r}) = -e|f_k(\mathbf{r})|^2$, where $f_k(\mathbf{r})$ is the wave function for k-th electron, Hartree worked out a self-consistent field theory for atoms which is described in the following section[11].

3.1 Hartree SCF Method

For an atomic system consisting of more than one electron, because of the interelectronic repulsion terms, the Schrödinger equation is not separable, but we can obtain a zeroth-order wave function by neglecting these repulsions. The Schrödinger equation would then separate into n one-electron H-like wavefunctions. The zeroth-order wave function would then be a product of n hydrogenlike (one-electron) orbitals:

$$\psi = f_1(\mathbf{r}_1)f_2(\mathbf{r}_2) \cdots f_n(\mathbf{r}_n) \quad (3.1)$$

where the hydrogenlike k -th orbitals are in the form:

$$f_k = R_{n_k l_k}(r_k) \cdot Y_{l_k}^{m_k}(\theta_k, \phi_k) \quad (3.2)$$

Electron 1 is assumed to move in the combined field of the nucleus taken as a point charge Ze and $(n - 1)$ other electrons taken as a continuous negative charge distribution, in which the density of the k -th electron is given by: $\rho_k(\mathbf{r}) = -e|f_k(\mathbf{r})|^2$. Now, we can write Hartree's equation for the k -th electron as follows:

$$\hat{H}_k^{eff} f(\mathbf{r}_k) = \epsilon_k f(\mathbf{r}_k) \quad (3.3)$$

where

$$\hat{H}_k^{eff} = -\frac{\hbar^2}{2m} \nabla^2 + V_k(f_1, f_2 \cdots, f_n) \quad (3.4)$$

and

$$V_k(f_1, f_2 \cdots, f_n) = -\frac{Ze^2}{r} + \sum_{l \neq k} e^2 \int d\tau' \frac{|f_l(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} \quad (3.5)$$

Here, the prime ($'$) coordinates represent the coordinates of the l -th electron with $l \neq k$. To use the Hartree's method, we make the **central-field approximation** that the potential is dependent on the radial distance only, and solve the Hartree's equation for the first electron to obtain an improved orbital by using $(n - 1)$ trial orbitals. Then we use 1 improved orbital and $(n - 2)$ trial orbitals, and solve the Hartree's equation for the second electron to obtain the second improved orbital. We continue the process until we have improved orbitals for all n electrons. Then we go back to electron 1 and repeat the process until convergence is achieved.

3.2 Slater Determinant

In the preceding discussion, we paid no attention to the Pauli exclusion principle which states that no two electrons can occupy the same quantum space. For a wave function to represent the correct physical picture, it has to be antisymmetric.

An antisymmetric wave function can be constructed for an n electron system by writing the Slater determinant of n orbitals as follows:

$$\psi(\mathbf{r}_1, \mathbf{r}_2 \cdots, \mathbf{r}_n) = \begin{vmatrix} f_1(\mathbf{r}_1) & f_2(\mathbf{r}_1) & \cdots & f_n(\mathbf{r}_1) \\ f_1(\mathbf{r}_2) & f_2(\mathbf{r}_2) & \cdots & f_n(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(\mathbf{r}_n) & f_2(\mathbf{r}_n) & \cdots & f_n(\mathbf{r}_n) \end{vmatrix} \quad (3.6)$$

Often, we see the following condensed notation used for the Slater determinant:

$$\psi(\mathbf{r}_1, \mathbf{r}_2 \cdots, \mathbf{r}_n) = |f_1(\mathbf{r}_1) \ f_2(\mathbf{r}_1) \ \cdots \ f_n(\mathbf{r}_1)| \quad (3.7)$$

3.3 Hartree-Fock SCF Method

In 1930, Fock and Slater [6, 13] independently worked out a SCF method consisting of determinantal, antisymmetric wave functions. The procedure is the same as before except that we are using a slightly different form of Hamiltonian in equation 2.3 consisting of particle exchange operators (\hat{P}_{ij}). In 1951, Roothaan introduced a representation of the Hartree-Fock orbitals as linear combinations of a complete set of known functions called **basis functions**, i.e.

$$f_k(\mathbf{r}) = \sum_{\alpha=1}^{n'} c_{i\alpha} \chi_{\alpha}(\mathbf{r}) \quad (3.8)$$

Here n' is the number of linearly independent functions $\chi_1, \chi_2, \cdots, \chi_{n'}$ in the basis set. If n is the number of occupied orbitals, it is necessary that $n' \geq n$. For $n' = n$ it is called the minimal basis set. Commonly used basis sets are **Slater Type Orbitals(STOs)** and **Gaussian** orbitals. Using this method, we can compute the integrals for the expectation values for the kinetic and potential energies in advance, and the problem reduces to n eigenvalue problems for $n' \times n'$ matrices.

4. Variational Wave Function

4.1 Artificial Neural Network

A standard Artificial Neural Network (ANN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. The input neurons are activated through sensors perceiving the environment while other neurons are activated through the weighted connections from already active neurons. Such interconnection of neurons leads to output neurons, the final result of all the neuronal processing on the input sensors[15]. An ANN may be either recurrent or feedforward, of which only the latter is discussed in this thesis. A simple example of a feedforward neural network is given in the figure below:

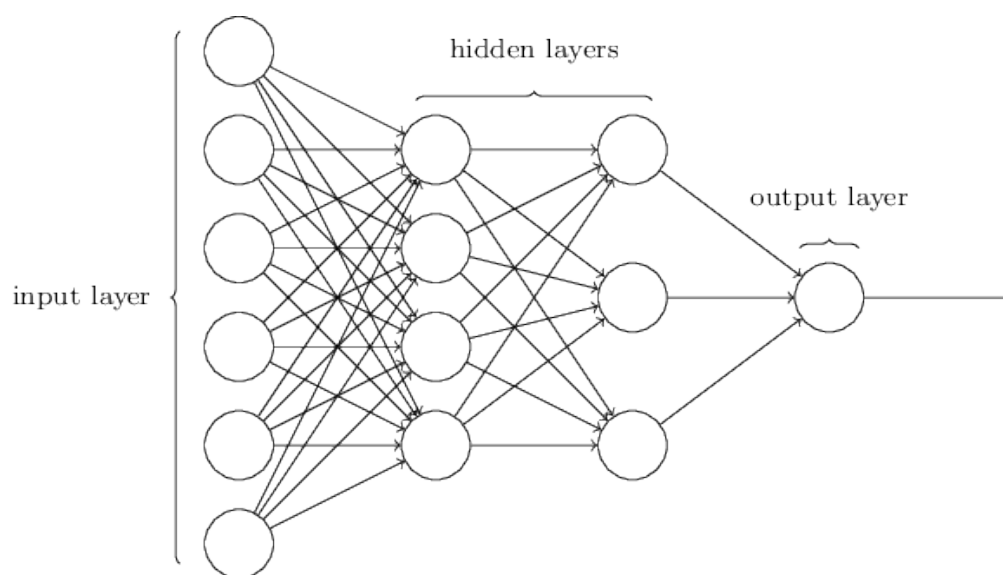


Figure 4.1: A feedforward Artificial Neural Network(ANN)

Suppose $x_i^{(1)}$ is the input, $w_{ji}^{(1)}$ is the weight and $b_j^{(1)}$ us the bias for the input layer.

In this notation, the index of the layer is denoted by the superscript. The input layer 1 is now fed to the hidden layer 2, in which each nodal value is given by the following relation:

$$y_j^{(1)} = \sum_i w_{ji}^{(1)} x_i^{(1)} + b_j^{(1)} \quad (4.1)$$

Before feeding these values to the next layer, they are transformed by an activation function. If $a(x)$ is the activation function, the transformed value of each node in layer 2 is given by:

$$x_j^2 = a(y_j^{(1)}) \quad (4.2)$$

Activation function in use these days include the sigmoid function, rectified linear unit, gaussian, ramp function and so on. In this work, we have used the sigmoidal activation function:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4.3)$$

In a similar fashion, a weighted sum is carried out to calculate the transformed input for layer 2 to obtain $y_j^{(2)}$, which is transformed into $x_j^{(3)}$, the input to layer 3. The following schematic explains the notation clearly:

$$x^{(1)} \rightarrow y^{(1)} \xrightarrow{a} x^{(2)} \rightarrow y^{(2)} \xrightarrow{a} \dots \xrightarrow{a} y^{(n)}$$

Figure 4.2: Notation for various neural layers

4.2 Radial Basis Function Neural Network

Another type of neural network used in this work is the radial basis function neural network. We'll use the radial basis function neural networks to approximate the ground state wave function for a three dimensional simple harmonic oscillator. For a three-layer neural network with one input layer, one output layer, and one

hidden layer, the output is given by the following equation:

$$y(\mathbf{x}) = \sum_i^M a_i \rho_i(\|\mathbf{x} - \mathbf{c}_i\|) \quad (4.4)$$

where \mathbf{a}_i and \mathbf{c}_i are the parameters of the neural network, \mathbf{x}_i is the input vector with the same dimension as \mathbf{c}_i , M is the number of neurons in the hidden layer and $\rho(\|\ \ \|)$ is a radial basis function. Two of the commonly used radial basis functions are :

$$\rho_i(\|\mathbf{x} - \mathbf{c}_i\|) = e^{-b_i \|\mathbf{x} - \mathbf{c}_i\|^2} \quad (4.5)$$

and

$$\rho_i(\|\mathbf{x} - \mathbf{c}_i\|) = e^{-b_i \|\mathbf{x} - \mathbf{c}_i\|} \quad (4.6)$$

In a recent paper, Teng[18] has used the radial function network to carry out the Variational Monte Carlo (VMC) calculations for the ground state of some simple Hamiltonians such as a Simple Harmonic Oscillator (SHO). In this work, too, we tried to reproduce his results to get familiarized with the optimization techniques. The results obtained from these calculations are presented in the results section. The reason for using the neural networks as the wave function approximators is their representation capability. Hornik [16] showed that a standard multilayer feedforward networks with as few as a single hidden layer and an arbitrary bounded and nonconstant activation function are universal approximators for arbitrary finite input environment measures, provided that sufficiently many hidden units are available. His results of representing an arbitrary function by neural networks can be thought of being inspired by Kolmogorov's representation theorem [17] that states for any integer $n \geq 2$ there are continuous real functions $\psi^{pq}(x)$ on the closed unit interval $E^1 = [0, 1]$ such that each continuous real function $f(x_1, \dots, x_n)$ on

the n -dimensional unit cube E^n is representable as:

$$f(x_1, \dots, x_n) = \sum_{q=1}^{q=2n+1} X_q \left[\sum_{p=1}^n \psi^{pq}(x_p) \right] \quad (4.7)$$

where $X_q(y)$ are continuous real functions.

4.3 Variational Representation of an Atomic Wave Function Using ANN

To derive the variational representation of an atomic wave function using ANN, we will rewrite equations 3.1 and 3.2. The zeroth order wave function, that does not involve any correlation terms, for a many-electron atom can be written as follows:

$$\psi = f_1(\mathbf{r}_1) f_2(\mathbf{r}_2) \cdots f_n(\mathbf{r}_n) \quad (4.8)$$

where the hydrogenlike k -th orbitals are in the form:

$$f_k = R_{n_k l_k}(r_k) Y_{l_k}^{m_k}(\theta_k, \phi_k) \quad (4.9)$$

Here n_k , l_k , and m_k are the principal, angular momentum, and magnetic quantum numbers, respectively. In our approach, rather than approximating the radial functions for each of the orbitals differently, we approximate them in a combined fashion by using a single feedforward artificial neural network with one hidden layer and m sigmoid hidden units, i.e.

$$N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) = \sum_{i=1}^m a_i \sigma \left(\sum_{j=1}^n b_{ij} r_j + c_i \right) = \sum_{i=1}^m a_i \sigma_i(r_1 \cdots r_n; \mathbf{b}_i; c_i) \quad (4.10)$$

In this equation, $\boldsymbol{\lambda} = (a_1, \dots, a_m; \mathbf{b}_1, \dots, \mathbf{b}_m; c_1, \dots, c_m)$ is a collective representation for all the network parameters a , b , and c . σ is the sigmoid function widely

used by the machine learning community, given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (4.11)$$

An advantage of using this function can be seen while differentiating it. The evaluation of the function gets us very close to the evaluation of its derivative, resulting in a huge computational advantage:

$$\sigma'(x) = \sigma(x) \cdot [1 - \sigma(x)] \quad (4.12)$$

If we look up the radial wave functions for an electron in H-like atoms, we will see that these consist of an exponential decay term given by $e^{-\frac{Zr}{nka_0}}$, where a_0 is the Bohr's radius. Using this observation, we will multiply our neural network radial wave function by the following term:

$$e^{-d_1 r_1} e^{-d_2 r_2} \dots \quad (4.13)$$

To make the our formulation tidier, we will rewrite the wave function as follows:

$$\Psi = N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) \underbrace{e^{-d_1 r_1} Y_{l_1}^{m_1}(\theta_1, \phi_1)}_{f_1(r_1, \theta_1, \phi_1) = f_1(\mathbf{r}_1)} \dots \underbrace{e^{-d_n r_n} Y_{l_n}^{m_n}(\theta_n, \phi_n)}_{f_n(r_n, \theta_n, \phi_n) = f_n(\mathbf{r}_n)} \quad (4.14)$$

$$= N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) f_1(\mathbf{r}_1) \dots f_n(\mathbf{r}_n) \quad (4.15)$$

Here, the f_i 's are not to be confused with the actual orbitals. These are one-electron functions consisting of both radial and angular part, but are missing the polynomial radial part of H-like orbitals.

Our formulation of the wave function for a many-electron atom in the form of equation 4.8 so far does not take into account the antisymmetric requirement required by the Pauli exclusion principle. This formulation is similar to the Hatree theory whose successor is the Hartree-Fock theory includes the Pauli exclusion principle while still letting each electron have its own orbital. In this theory, the

assumed form of the wave function is a Slater determinant (see section 3.2) . However, our formulation does not consist of distinct orbitals rather the product of the radial parts from each of the orbitals is represented by the neural network. It is important to note that, in addition to the neural network representation, we also have an exponential decay factor (equation 4.13) for each of the orbitals. We could have included the effect of these factors into the neural network itself, however, doing so would eliminate the possibility of exploiting the Slater determinant to achieve the antisymmetric wave function, which is so far the only method known to us to enforce the Pauli exclusion principle in a many-electron wave function.

Each of the orbitals can accommodate two electrons with opposite spins. Spin can be regarded as an internal degree of freedom having two possible states, denoted by the spin functions α or β . Unless magnetic interactions are included in the Hamiltonian, space and spin variables for each electron are separable. Notice that n orbitals can account for $2n$ electrons. Now, we are in a position to express the many-electron wave function in the form of Slater determinant:

$$\Psi = N(r_1, r_2, \dots r_n; \boldsymbol{\lambda}) \begin{vmatrix} f_1(\mathbf{r}_1)\alpha & f_1(\mathbf{r}_1)\beta & f_2(\mathbf{r}_2)\alpha & f_2(\mathbf{r}_2)\beta & \dots \\ f_1(\mathbf{r}_2)\alpha & f_1(\mathbf{r}_2)\beta & f_2(\mathbf{r}_2)\alpha & f_2(\mathbf{r}_2)\beta & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ f_1(\mathbf{r}_n)\alpha & f_1(\mathbf{r}_n)\beta & f_2(\mathbf{r}_n)\alpha & f_2(\mathbf{r}_n)\beta & \dots \end{vmatrix} \quad (4.16)$$

This notation can be abbreviated as follows:

$$\Psi = N(r_1, r_2, \dots r_n; \boldsymbol{\lambda}) \begin{vmatrix} f_1(\mathbf{r}_1)\alpha & f_1(\mathbf{r}_1)\beta & f_2(\mathbf{r}_1)\alpha & f_2(\mathbf{r}_1)\beta & \dots \end{vmatrix} \quad (4.17)$$

$$= N(r_1, r_2, \dots r_n; \boldsymbol{\lambda}) \begin{vmatrix} f_1(\mathbf{r}_1) & \overline{f_1(\mathbf{r}_1)} & f_2(\mathbf{r}_1) & \overline{f_2(\mathbf{r}_1)} & \dots \end{vmatrix} \quad (4.18)$$

Here, the unbarred function is multiplied by the spin variable α while the barred function is multiplied by the spin variable β . Note that the normalization factor $\frac{1}{\sqrt{N!}}$ has been omitted from the Slater determinant as it is taken into account in the

coefficient of the neural network function. In this variational notation, we expect to have fewer parameters for the variational representation because the radial part of each of the orbitals is not optimized separately rather we are optimizing their product while still preserving the antisymmetric property of the wave function. Following is a simple illustration of this property. Consider the He wave function:

$$\begin{aligned}\Psi &= N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) \begin{vmatrix} f_1(\mathbf{r}_1)\alpha & f_1(\mathbf{r}_1)\beta \\ f_1(\mathbf{r}_2)\alpha & f_1(\mathbf{r}_2)\beta \end{vmatrix} \\ &= N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) \begin{vmatrix} e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1)\alpha & e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1)\beta \\ e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2)\alpha & e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2)\beta \end{vmatrix}\end{aligned}$$

If $\mathbf{r}_1 = \mathbf{r}_2$ i.e. $(r_1, \theta_1, \phi_1) = (r_2, \theta_2, \phi_2)$, rows 1 and 2 are equal resulting in a zero determinant.

5. Optimization Scheme

5.1 Variational Principle

The variational principle in quantum mechanics allows us to find the ground state of a quantum system that satisfies the Schrödinger equation:

$$\hat{H}\psi = E\psi \tag{5.1}$$

It states that, for any choice of trial function Ψ_T the energy calculated is always greater than or equal to the ground state energy i.e.

$$E_0 \leq \frac{\langle \psi_T | \hat{H} | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} \tag{5.2}$$

Since our goal is to find the ground state, the problem is now reduced to a optimization problem in which we are supposed to minimize the energy by finding optimal values of the variational parameters for the wave function as described in the previous section.

5.2 Variational Monte Carlo (VMC) Integration

To calculate the expectation value of an operator \hat{O} , we need to integrate over all spatial and spin variables. The number of spatial variables (x, y, z) increases by a factor of 3 with the addition of each new particle to the system. Consequently, grid-based integration method can be computationally very expensive in higher

dimensional calculations. Although grid-based integration is a very accurate, depending on how fine a mesh we choose, most of the time, we are sampling over configurations in which the probability of finding the particles is extremely low. Thus, it makes more sense to use a method that spends less time sampling over low probability configurations and more time sampling over high probability configurations. Variational Monte Carlo Integration computes the expectation value of an operator by taking the average value of the the operator over the configurations sampled based on the probability distribution function given by $|\psi(\mathbf{R}; \boldsymbol{\lambda})|^2$, which is literally the probability of finding the particles in the given configuration. Suppose \mathbf{R} represents the set of all the spatial variables $(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$ and $\boldsymbol{\lambda}$ represents all the variational parameters. The calculation of the expectation value proceeds as follows:

$$\begin{aligned}
\langle \hat{O} \rangle &= \frac{\langle \psi(\mathbf{R}; \boldsymbol{\lambda}) | \hat{O} | \psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}{\langle \psi(\mathbf{R}; \boldsymbol{\lambda}) | \psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle} \\
&= \frac{\int |\psi(\mathbf{R}; \boldsymbol{\lambda})|^2 \frac{\hat{O} |\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}{|\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle} d\mathbf{R}}{\underbrace{\int |\psi(\mathbf{R}; \boldsymbol{\lambda})|^2 d\mathbf{R}}_1} \\
&= \int \underbrace{|\psi(\mathbf{R}; \boldsymbol{\lambda})|^2}_{\substack{\text{Probability} \\ \text{Distribution} \\ \text{Function}}} \frac{\hat{O} |\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}{|\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle} d\mathbf{R} \tag{5.3}
\end{aligned}$$

There is no spin part in this expression because we are assuming that the operator \hat{O} does not act on spin. We can see that that the quantity being sampled is $\frac{\hat{O} |\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}{|\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}$ rather than \hat{O} .

The following is an example that illustrates the calculation of the expectation value for the hamiltonian (\hat{H}):

$$\begin{aligned}
E &= \langle \psi(\mathbf{R}; \boldsymbol{\lambda}) | \hat{H} | \psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle \\
&= \int |\psi(\mathbf{R}; \boldsymbol{\lambda})|^2 \frac{\hat{H} |\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle}{|\psi(\mathbf{R}; \boldsymbol{\lambda}) \rangle} d\mathbf{R} \tag{5.4}
\end{aligned}$$

Suppose

$$E_L(\mathbf{R}) = \frac{\hat{H} |\psi(\mathbf{R}; \boldsymbol{\lambda})\rangle}{|\psi(\mathbf{R}; \boldsymbol{\lambda})\rangle} \quad (5.5)$$

is the local energy. Then we can write the energy as:

$$E = \frac{1}{N} \sum_{i=1}^N E_L(\mathbf{R}_i) \quad (5.6)$$

Here, E_L is calculated for N samples sampled according to the probability distribution function $|\psi(\mathbf{R}; \boldsymbol{\lambda})|^2$. The sampling is carried out by using the Metropolis algorithm which is described in the following section. In theory, the accurate value of energy is obtained as $N \rightarrow \infty$

5.3 Metropolis Algorithm

The Metropolis algorithm is the most popular form of **Markov Chain Monte Carlo(MCMC)** process. We start with some random configuration \mathbf{R}_{old} , pick a trial configuration $\mathbf{R}_{\text{trial}}$, and now attempt to make the move $\mathbf{R}_{\text{old}} \rightarrow \mathbf{R}_{\text{trial}}$. The probability of this move being accepted is given by the following factor:

$$w = \min \left(1, \left| \frac{\psi(\mathbf{R}_{\text{trial}}, \boldsymbol{\lambda})}{\psi(\mathbf{R}_{\text{old}}, \boldsymbol{\lambda})} \right|^2 \right) \quad (5.7)$$

The acceptance ratio is given by the ratio of number of accepted moves to the total number of attempted moves. It is common to adjust the range of the move vector such that the acceptance ration is around 0.5. However, this method is still inefficient because we are still calculating the value of the wave function at a lot of configurations which are not very probable. To overcome this issue, an improved version of this method called the **Metropolis-Hastings** algorithm [9] is used. This MCMC process uses Green's function to improve the acceptance ratio. It is not described in detail here because we used the Metropolis algorithm for all our calculations.

5.4 Stochastic Reconfiguration (SR)

Stochastic Reconfiguration is commonly used to optimize the quantum wave function [7, 18, 3]. To use this method, we need to define the following operator:

$$O_{\lambda_i} = \frac{\partial_i \psi(\mathbf{R}, \boldsymbol{\lambda})}{\psi(\mathbf{R}, \boldsymbol{\lambda})} \quad (5.8)$$

The covariance matrix and forces are defined as:

$$\begin{aligned} S_{ij} &= \langle O_i^* O_j \rangle - \langle O_i^* \rangle \langle O_j \rangle \\ F_i &= \langle E_{local} O_i^* \rangle - \langle E_{local} \rangle \langle O_i^* \rangle \end{aligned}$$

The parameters are updated as follows:

$$\boldsymbol{\lambda}'_j = \boldsymbol{\lambda}_j + \alpha \mathbf{S}^{-1} \mathbf{F} \quad (5.9)$$

Since the covariance matrix (\mathbf{S}) can be non-invertible, we take its Moore-Penrose pseudo-inverse for regularization purposes.

5.4.1 SR Operators for Simple Harmonic Oscillator (SHO)

As mentioned earlier, we replicated the results for a SHO produced by Teng [18] in which we used a Gaussian basis function to estimate the ground state. The wave function can be approximated for this neural network as:

$$\psi(\mathbf{R}; \mathbf{a}, \mathbf{b}, \mathbf{c}) = \sum_i^M a_i \rho_i(\|\mathbf{R} - \mathbf{c}_i\|) \quad (5.10)$$

where

$$\rho_i(\|\mathbf{R} - \mathbf{c}_i\|) = e^{-|b_i| \|\mathbf{R} - \mathbf{c}_i\|^2} \quad (5.11)$$

is the Gaussian basis and we've used $\mathbf{R} = (x, y, z)$ for a 3-dimensional SHO.

Now we can define the SR operators as follows:

$$\hat{O}_{a_i}(\mathbf{R}) = \frac{\rho_i}{\psi} \quad (5.12)$$

$$\hat{O}_{b_i}(\mathbf{R}) = -\frac{a_i b_i |\mathbf{R} - \mathbf{c}_i|^2 \rho_i}{|b_i| \psi} \quad (5.13)$$

$$\hat{O}_{c_{ij}}(\mathbf{R}) = \frac{2a_i |b_i| (r_j - c_{ij}) \rho_i}{\psi} \quad (5.14)$$

5.4.2 SR Operators for Helium Wave Function

Here, we define the SR Operators for a Helium wave function only because it is fairly simple; for bigger atoms, the number of terms in the Slater determinant increases and calculation of the operators involving derivatives gets more complicated. As mentioned in the previous chapter, the wave function for a He atom is given by the following determinant:

$$\begin{aligned} \psi &= N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) \begin{vmatrix} f_1(\mathbf{r}_1)\alpha & f_1(\mathbf{r}_1)\beta \\ f_1(\mathbf{r}_2)\alpha & f_1(\mathbf{r}_2)\beta \end{vmatrix} \\ &= N(r_1, r_2, \dots, r_n; \boldsymbol{\lambda}) \begin{vmatrix} e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1)\alpha & e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1)\beta \\ e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2)\alpha & e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2)\beta \end{vmatrix} \\ &= \sum_{i=1}^m a_i \sigma \left(\sum_{j=1}^n b_{ij} r_j + c_i \right) e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1) e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2) \cdot \underbrace{(\alpha\beta - \beta\alpha)}_{=1} \quad (5.15) \end{aligned}$$

The spin functions integrate to 1 when summed over all the spin configurations so we are eliminating the spin part from our calculation for the Helium wave function. However, for larger atoms, it is wrong to do so because not all the spin terms integrate to 1. In order to avoid such complexity arising from the spin terms, VMC calculations exploit the technique of separating the Slater determinant into two determinants with spin up and spin down terms [8]. For He, the SR operators

for each of the parameters are as follows:

$$\begin{aligned}\hat{O}_{a_i}(\mathbf{R}) &= \frac{\sigma_i}{N} \\ \hat{O}_{b_{ij}}(\mathbf{R}) &= \frac{\sigma'_i r_j a_i}{N} \\ \hat{O}_{c_{ij}}(\mathbf{R}) &= \frac{\sigma'_j a_j}{N}\end{aligned}$$

where,

$$\sigma' = \sigma(\sigma - 1) \tag{5.16}$$

is the derivative of the sigmoid function.

6. Results

6.1 Simple Harmonic Oscillator (SHO)

The Hamiltonian for the 3d simple harmonic oscillator is given by:

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \frac{1}{2}k(x^2 + y^2 + z^2) \quad (6.1)$$

In our simplified units, $\hbar = m = k = 1$. Consequently,

$$\hat{H} = -\frac{1}{2}\nabla^2 + \frac{1}{2}(x^2 + y^2 + z^2) \quad (6.2)$$

In these units, the ground state is given by:

$$\psi_0(x, y, z) = e^{-\frac{1}{2}(x^2+y^2+z^2)} \quad (6.3)$$

and the n-th energy level is given by $E_n = (n + \frac{3}{2})$. So, the ground state energy is equal to 1.5 units. We carried out the SR optimization for a Gaussian basis neural network of various hidden nodes (Nh) which are listed in Table 6.1:

Nh	Energy
1	1.50
2	1.50
5	1.50
10	1.50

Table 6.1: Ground state energy for 3d SHO for various number of hidden nodes (Nh)

The optimization using SR for a 3d SHO was very fast on the computer. Even for the neural networks with 10 hidden nodes, the ground state was reached in less than 10 steps. The plot below shows the convergence for neural networks with various N_h .

The average value of all the calculations with varying number of hidden nodes

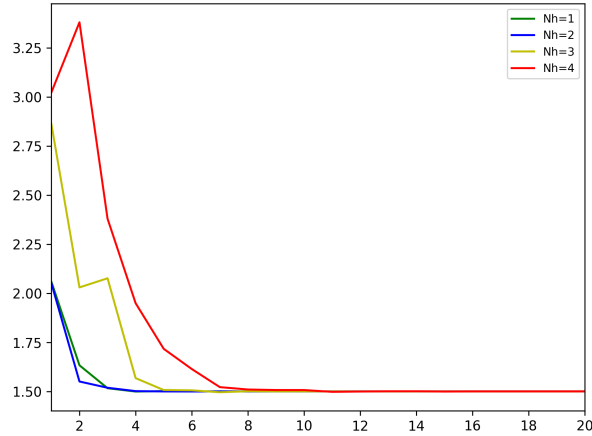


Figure 6.1: Convergence for neural networks with various N_h

was 1.50, so the neural network wave function was able to accurately represent the ground state of the SHO. It is not quite surprising because the ground state of a SHO is a Gaussian given by equation 6.3 so only one hidden node should be able to represent the ground state. The main purpose of this calculation was to get acquainted with application of the various methods discussed in the previous section by reproducing the results of Teng[18].

6.2 Helium Atom

The Hamiltonian for an atomic system in general is given by:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^n \nabla_i^2 - \sum_{i=1}^n \frac{Z}{r_i} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (6.4)$$

In this reduced system of units, we get the energy in the units of Hartrees (1 Hartree = 27.21138602 eV). For the He atom, the Hamiltonian is given by:

$$\hat{H} = -\frac{1}{2}\nabla_1^2 + -\frac{1}{2}\nabla_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}} \quad (6.5)$$

The calculated value of ground state energy by using a neural network wave function given by equation 5.15 is 2.86 Hartrees. The experimentally determined value is 2.90 Hartrees and the value determined from SCF calculations is 2.86168 Hartrees [19]. Our calculation is very close to the SCF calculation; however, this alone is not a proof of concept for the neural network based atomic wave functions. The main reason is because the He ground state only involves 1s orbital. The radial part for the 2s orbital for a hydrogen atom is $2a^{-3/2}e^{-\frac{r}{a}}$ where a is the Bohr's radius. This expression is purely an exponential decay, and our variational expression of the wave function consists of an exponential decay part already. So, the neural network is only left to represent a constant function i.e. the normalization factor. So, the problem predominantly turns into an optimization problem for the exponential decay factor d rather than the actual neural network parameters. Since we are trying to get the neural network to represent a constant function, the optimization is completed in only 3-4 steps by using SR method.

6.3 Lithium Atom

The Hamiltonian for a Li atom is given by:

$$\hat{H} = -\frac{1}{2}\nabla_1^2 + -\frac{1}{2}\nabla_2^2 - \frac{3}{r_1} - \frac{3}{r_2} - \frac{3}{r_3} + \frac{1}{r_{12}} + \frac{1}{r_{23}} + \frac{1}{r_{13}} \quad (6.6)$$

As mentioned earlier, the wave function for Li atom obtained using Slater determinant involves spin terms which is fairly non intuitive to optimize by using the VMC algorithm that involves splitting the determinant into two determinants,

one each for spin up and spin down configurations. We have not yet been able to apply this method, so we tried to optimize our variational wave function for the Li atom by assuming it as a product of the orbitals:

$$\psi = \sum_{i=1}^m a_i \sigma \left(\sum_{j=1}^n b_{ij} r_j + c_i \right) e^{-d_1 r_1} Y_0^0(\theta_1, \phi_1) e^{-d_1 r_2} Y_0^0(\theta_2, \phi_2) e^{-d_2 r_3} Y_2^0(\theta_3, \phi_3) \quad (6.7)$$

Now the neural network is not representing just the normalization factor but rather a polynomial function. For an H atom, the radial part for the 2s orbital is given by:

$$R_{20} = \frac{1}{\sqrt{2}} a^{-3/2} \left(1 - \frac{1}{2} \frac{r}{a} \right) e^{-\frac{r}{2a}} \quad (6.8)$$

We already have the exponential decay part in our variational form so the only part that the neural network is left to approximate is the polynomial part. It should be noted that our variational form in equation 6.7 is not entirely accurate because we can have multiple electrons occupying the same position because there is no antisymmetric property here. We calculate the ground state energy to be -7.41 Hartrees, while the value calculated by using highly accurate SCF method on Hylleraas basis set is -7.4780603239041 [20]. Although our value is not quite the same as the SCF calculated value, it is more accurate than the non-antisymmetrized wave function optimized value which is -7.403 Hartrees as calculated by Davis[21].

Conclusion

In this work, we have outlined a theoretical model to calculate the ground state wave function for atomic systems by using an artificial neural network. Although this model has not been fully tested, we developed the computational groundwork for the wave function calculation and its optimization. In this process, we reproduced the results for solving the SHO ground state using radial basis function neural network first proposed by Teng [18]. In all the calculations, the Metropolis algorithm was used to sample the spatial configurations, VMC was used to calculate the expectation values of the operators, and SR was used to minimize the wave function. We tested this approach for the He atom and the results seem to be competent with highly accurate SCF calculations. However, for the Li atom, we have not fully worked out the VMC calculation involving the Slater determinant, so we tested the un-antisymmetrized wave function which gave ground state energy value better than the one calculated by a similar approach [21]. Through this work, we have demonstrated the case made by Carleo and Troyer [3] that novel methods involving artificial neural networks can be developed to solve quantum many body problems.

Bibliography

- [1] Paul Mozur, *The New York Times*. May 23, 2017
- [2] I.E. Lagaris, A. Lokas, D.I. Fotiadis, *Computer Physics Communications* **104**, 1-14 (1997).
- [3] Giuseppe Carle and Matthias Troyer, *Science* **355**, 602-606 (2017).
- [4] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [5] D. R. Hartree, *The Calculation of Atomic Structure* John Wiley and Sons, Inc. New York. (1957)
- [6] Vladimir Fock, *Z.Physik.* **61**, 126 (1930).
- [7] Sandro Sorella, Michele Casula, Dario Rocca, *Journal of Chemical Physics* **127**, 014105 (2007).
- [8] B. L. Hammond, Jr. W. A. Lester, and P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*. World Scientific, Singapore (1994).
- [9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, *Journal of Chemical Physics.* **21**, 1087-1092(1953).
- [10] S.M. Blinder, *American Journal of Physics.* **336**,6, 431-443(1965)
- [11] D. R. Hartree, *Proceedings of Cambridge Philosophical Society.* **24**, 89, 111, 246(1928)
- [12] Ira N. Levine, *Pearson Education, Inc.*, New Jersey (2009).

- [13] John C. Slater, *Physical Review*. **35**, 210 (1930).
- [14] C.C.J. Roothan, *Reviews of Modern Physics*. **23**, 69 (1951).
- [15] Jürgen Schmidhuber, *arXiv:1404.7828 v4* . (2014).
- [16] Kurt Hornik, *Neural Networks*. **4**, 251-257 (1991)
- [17] A.N. Kolmogorov, *Dokl. Akad. Nauk SSSR* textbf108:2, 179–182 (1961)
- [18] Peiyuan Teng, *arXiv:1710.03213*, 2017 (2017)
- [19] <http://www.umich.edu/~chem461/QMChap8.pdf>
- [20] Mariusz Puchalski, Krzysztof Pachucki, <http://arxiv.org/abs/physics/0601213v1>
(2006).
- [21] Stephen L.Davis, *Journal of Chemical Education* **84:4**, 711-720 (2007).