Chapter 5

Deep learning variational Monte Carlo for solving the electronic Schrödinger equation

Leon Gerard[a], **Philipp Grohs**[a,b,*], **and Michael Scherbela**[a]

[a]*Faculty of Mathematics, University of Vienna, Vienna, Austria,* [b]*Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Linz, Austria*

[*]*Corresponding author: e-mail address: philipp.grohs@univie.ac.at*

Contents

Abstract

The electronic Schrödinger equation is one of the most fundamental models in physics, due to its capability of accurately predicting all properties of molecules. Having efficient numerical methods for its solution would revolutionize drug- or material discovery – among many other fields. However, while the equation itself is easily stated, its efficient and accurate numerical solution poses formidable challenges. This has sparked decades-long research on the development of a plethora of different highly specialized numerical schemes. In recent years, methods based on Deep Learning have been introduced and shown to outperform the previous state-of-the-art in terms of accuracy, allowing for increasingly large systems to be computable to within chemical accuracy. In this paper we

Handbook of Numerical Analysis, Volume 25, ISSN 1570-8659, https://doi.org/10.1016/bs.hna.2024.05.010
Copyright © 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

survey these exciting developments from the perspective of a (numerical) mathematician. To this end, we first provide an introduction into the mathematical theory of the electronic Schrödinger equation and then outline numerical methods to partially overcome some of its considerable challenges. Finally we survey recent work on Deep Learning-based Variational Monte Carlo methods and showcase some numerical results.

Keywords
Electronic Schrödinger equation, Deep learning, Variational Monte Carlo, Fermionic neural networks

MSC Codes
68T07, 81-08, 81Q10, 65-02, 46N50, 35-02, 35Q40, 65C05

1 Introduction

> *The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.*
>
> P. Dirac (1929)

This famous quote taken from Dirac (1929) refers to the Schrödinger equation of electronic systems, which is also the subject of the present survey article. The Schrödinger equation represents a partial differential equation capable of accurately describing all nonrelativistic properties of atoms and molecules. This means that one can – theoretically – simulate all properties of molecules from first principles without having to resort to expensive and time-consuming experiments. This has in principle the potential to lower cost and enable the search for new materials on a much greater scale than ever before. In view of this, the importance of developing efficient numerical methods for the electronic Schrödinger equation is hard to overstate and solving it can be considered the "holy grail" of computational chemistry.

Solving the equation – i.e., finding a ground-state wavefunction for a given molecule – is computationally challenging. Analytical solutions are only known for atoms with a single electron (i.e., a single Hydrogen atom) and thus any system of practical interest must be solved numerically. It has furthermore been shown that for model-Hamiltonians such as the Hubbard model, finding the ground-state wavefunction is a QMA-hard problem (Troyer and Wiese, 2005), making it at least as hard as any NP-complete problem.

This has motivated a diverse range of computational methods, including renowned techniques like Density Functional Theory (DFT), which was awarded the Nobel Prize in 1998. On the one hand, there exist methods like DFT

or Hartree Fock that can handle systems comprising hundreds of atoms with relatively crude but efficient approximations. On the other hand, methods such as CCSD(T), often named the gold standard when it comes to precise approximations, tend to exhibit scaling behavior typically in the order of $O(n_{el}^7)$, where n_{el} represents the number of electrons considered. In the case of the configuration interaction singles, doubles, triples, quadruples (CISDTQ) method the scaling can even go up to $O(n_{el}^{10})$ (Scherbela et al., 2022), limiting the method to rather small system sizes. For an introduction to computational methods see Szabo and Ostlund (1996).

In more recent years, methods from machine learning have made a significant impact in improving the tradeoff between accuracy and computational cost. These approaches can be roughly categorized into supervised and unsupervised. In the supervised regime one typically starts from a data set of high accuracy calculations for several different compounds. This data set is then used as training data for a machine learning regressor. Once trained, such a regressor is capable of speeding up calculations by many orders of magnitude while retaining the accuracy of the training data set. We refer to Schütt et al. (2020) for an overview focused on material discovery.

The supervised learning approach leaves open the key problem of generating highly accurate solutions to the Schrödinger equation that can be subsequently used as training data. To address this issue, and due to their excellent approximation properties (Elbrächter et al., 2021), deep neural networks have been explored as ansatz functions for electronic wave functions. These neural networks are then trained in an unsupervised fashion to approximate wave functions of electronic ground states. This approach – pioneered in Carleo and Troyer (2017); Pfau et al. (2020); Hermann et al. (2020) and coined *Deep Learning Variational Monte Carlo (DL-VMC)* – has been shown to significantly outperform classical methods in terms of accuracy. We refer to Hermann et al. (2022) for a summary of recent results in this direction. With its apparent ability to better describe wavefunctions of molecules when compared to classical representations and the great progress that could already be achieved within the span of only a few years, DL-VMC holds the potential to revolutionize the field of computational chemistry.

Broadly speaking, DL-VMC can be regarded as a specific instance of so-called PINNs (physics-informed-neural-networks (De Ryck and Mishra, 2024)) applied to the Schrödinger equation. While the analysis and development of PINN methods for solving PDEs constitutes a highly active field of research, these methods often lag behind more standard algorithms (such as finite elements or low rank methods) in terms of accuracy (Chuang and Barba, 2022) or sample complexity (Bayer et al., 2023; Berner et al., 2023). DL-VMC is a rare instance of a PINN-like method that actually surpasses conventional algorithms in terms of accuracy, making it all the more interesting.

The present paper aims to provide a survey of these exciting developments with a specific focus on a readership coming from the field of mathematical/nu-

merical analysis. While empirical results of DL-VMC are convincing, the mathematical theory behind these successes has yet to be developed. We therefore hope that our work can serve as an invitation to members of the mathematical community to contribute to this vibrant field.

1.1 The molecular and electronic Schrödinger equations

We start with an informal description of the molecular and electronic Schrödinger equations. A molecule can be described by n_{nuc} nuclei with $\mathbf{R} = \left(R_1, \dots, R_{n_{\text{nuc}}}\right) \in \mathbb{R}^{n_{\text{nuc}} \times 3}$ denoting the nuclear coordinates and $\mathbf{Z} = \left(Z_1, \dots, Z_{n_{\text{nuc}}}\right) \in \mathbb{N}^{n_{\text{nuc}}}$ denoting their respective charges. These nuclei are surrounded by a cloud of n_{el} electrons with $\mathbf{r} = \left(r_1, \dots, r_{n_{\text{el}}}\right) \in \mathbb{R}^{n_{\text{el}} \times 3}$ denoting the electron coordinates. In atomic units (meaning that electron mass, elementary charge, Planck's constant $\hbar$ and permittivity are equal to 1) we assume that the mass of the I-th nucleus is equal to M_I. The electrons and nuclei interact with each other through Coulomb attraction and repulsion forces, which leads to the molecular Hamiltonian

$$\mathcal{H}^{\text{mol}} = -\frac{1}{2}\sum_{I=1}^{n_{\text{nuc}}} \frac{1}{M_I}\nabla_{R_I}^2 - \frac{1}{2}\sum_{i=1}^{n_{\text{el}}} \nabla_{r_i}^2 + \sum_{i=1}^{n_{\text{el}}-1}\sum_{j=i+1}^{n_{\text{el}}} \frac{1}{|r_i - r_j|} + \sum_{I=1}^{n_{\text{nuc}}-1}\sum_{J=I}^{n_{\text{nuc}}} \frac{Z_I Z_J}{|R_I - R_J|} - \sum_{i=1}^{n_{\text{el}}}\sum_{I=1}^{n_{\text{nuc}}} \frac{Z_I}{|r_i - R_I|}. \tag{1}$$

Here the term

$$-\frac{1}{2}\sum_{I=1}^{n_{\text{nuc}}} \frac{1}{M_I}\nabla_{R_I}^2 - \frac{1}{2}\sum_{i=1}^{n_{\text{el}}} \nabla_{r_i}^2$$

represents the kinetic energy and the term

$$\sum_{i=1}^{n_{\text{el}}-1}\sum_{j=i+1}^{n_{\text{el}}} \frac{1}{|r_i - r_j|} + \sum_{I=1}^{n_{\text{nuc}}-1}\sum_{J=I}^{n_{\text{nuc}}} \frac{Z_I Z_J}{|R_I - R_J|} - \sum_{i=1}^{n_{\text{el}}}\sum_{I=1}^{n_{\text{nuc}}} \frac{Z_I}{|r_i - R_I|},$$

which should be understood as a multiplication operator, represents the potential energy induced by Coulomb forces. The state of a molecule can be quantum mechanically described by its molecular wave function $\Psi = \Psi(\mathbf{R}, \mathbf{r})$. The time evolution of such a state is governed by the *time-dependent Schrödinger Equation*

$$i\frac{\partial}{\partial t}\Psi(t) = \mathcal{H}^{\text{mol}}\Psi(t). \tag{2}$$

While (2) in principle allows for a full description of a given molecule, it is often too complicated to be solved, even by numerical methods.

A common simplification is given by the *Born-Oppenheimer approximation* which essentially assumes that the positions of the nuclei stay fixed so that the

term corresponding to the kinetic energy of the nuclei can be omitted in the Hamiltonian. Heuristically, this approximation is justified since nuclei are much heavier than electrons and therefore the motion of the nuclei $\mathbf{R}$ can be assumed to occur on a much slower timescale than the motion of the electrons. Hence (to within a certain accuracy) the motion of the nuclei can be neglected. For a fixed geometrical conformation described by nuclear coordinates and charges $(\mathbf{R}, \mathbf{Z})$ this leads to the new Hamiltonian

$$\begin{aligned}\mathcal{H}^{\mathrm{BO}}_{(\mathbf{R},\mathbf{Z})} = &-\frac{1}{2}\sum_{i=1}^{n_{\mathrm{el}}} \nabla^2_{r_i} + \sum_{i=1}^{n_{\mathrm{el}}-1}\sum_{j=i+1}^{n_{\mathrm{el}}} \frac{1}{|r_i - r_j|} \\ &+ \sum_{I=1}^{n_{\mathrm{nuc}}-1}\sum_{J=I}^{n_{\mathrm{nuc}}} \frac{Z_I Z_J}{|R_I - R_J|} - \sum_{i=1}^{n_{\mathrm{el}}}\sum_{I=1}^{n_{\mathrm{nuc}}} \frac{Z_I}{|r_i - R_I|},\end{aligned} \tag{3}$$

which is now applied to *electronic wavefunctions* $\Psi^{\mathrm{e}} : \mathbb{R}^{n_{\mathrm{el}}\times 3} \ni (r_1, \ldots, r_{n_{\mathrm{el}}}) \to \mathbb{C}$, depending on the electron coordinates only, whereas the geometric conformation $(\mathbf{R}, \mathbf{Z})$ is treated as a parameter. We have therefore reduced the dimensionality by $3 \cdot n_{\mathrm{nuc}}$. Studying the simplified quantum system described by the Hamiltonian (3) is then referred to as the Born-Oppenheimer approximation.

Of specific interest is the *time-independent electronic Schrödinger Equation*, which amounts to solving the eigenvalue problem

$$\mathcal{H}^{\mathrm{BO}}_{(\mathbf{R},\mathbf{Z})}\Psi^{\mathrm{e}} = \lambda_{(\mathbf{R},\mathbf{Z})}\Psi^{\mathrm{e}}, \quad \lambda_{(\mathbf{R},\mathbf{Z})} \in \mathbb{R}. \tag{4}$$

The Eigenvalues $\lambda_{(\mathbf{R},\mathbf{Z})}$ correspond to the energies that the electronic system can assume without breaking apart (Hunziker and Sigal, 2000). The corresponding Eigenvectors describe precisely those electronic states Ψ^{e} whose energy $E = \lambda_{(\mathbf{R},\mathbf{Z})} = \left\langle \mathcal{H}^{\mathrm{BO}}\Psi^{\mathrm{e}}, \Psi^{\mathrm{e}} \right\rangle$ can be measured without uncertainty. Their time evolution under the time-dependent electronic Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi^{\mathrm{e}}(t) = \mathcal{H}^{\mathrm{BO}}\Psi^{\mathrm{e}}(t)$$

can be easily calculated to yield $\Psi^{\mathrm{e}}(t) = e^{-iEt}\Psi^{\mathrm{e}}(0)$. Therefore, knowledge of the eigenstates allows for a simple solution of the time-dependent electronic Schrödinger equation whenever the initial state is given as a superposition of eigenstates.

The eigenvectors $\Psi^{\mathrm{e},0}_{(\mathbf{R},\mathbf{Z})}$ corresponding to the smallest possible eigenvalue $\lambda^0_{(\mathbf{R},\mathbf{Z})}$ in (4) are called *electronic ground states* and generally represent the most stable and likely electronic states. The eigenvectors corresponding to higher eigenvalues are called *excited states*. For a fixed number of nuclei n_{nuc} and fixed nuclear charges $\mathbf{Z}$, the mapping $\mathcal{E}_{\mathbf{Z}} : \mathbb{R}^{n_{\mathrm{nuc}}\times 3} \ni \mathbf{R} \mapsto \lambda^0_{(\mathbf{R},\mathbf{Z})}$ is called the *potential energy surface (PES)*.

The PES contains a wealth of information about the chemical properties of a given molecule. For example, it can be used as an approximation of

the potential energy of the nuclear coordinates $\mathbf{R}$ leading to the force field $F_{\mathbf{Z}}(\mathbf{R}) := -\nabla_{\mathbf{R}}\mathcal{E}_{\mathbf{Z}}(\mathbf{R})$ acting on the nuclei. Fig. 1 shows a calculation of a PES of hydrogen chains using a deep learning ansatz (Scherbela et al., 2022). Using these forces, the dynamics of the nuclei can be approximately simulated classically by solving Newton's equations

$$M\ddot{\mathbf{R}}(t) = F_{\mathbf{Z}}(\mathbf{R}(t)), \quad M = \mathrm{diag}(M_1, \ldots, M_{n_{\mathrm{nuc}}}) \in \mathbb{R}^{n_{\mathrm{nuc}} \times n_{\mathrm{nuc}}}.$$

This in turn allows for the determination of the structure of molecules or the ab initio simulation of chemical reactions.

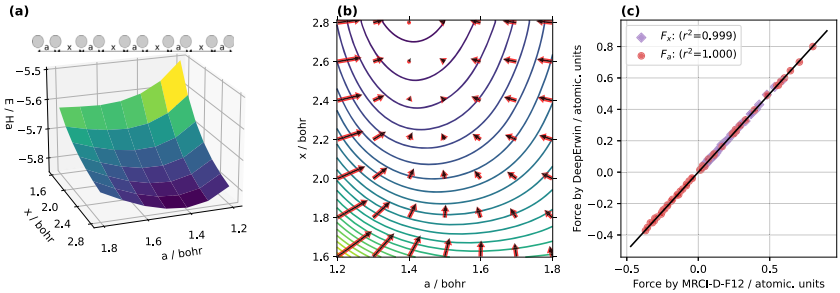

FIGURE 1 (a) Potential energy surface (PES) for the H_{10} chain. The variable a is representing the distance between two H-atoms and x is the distance between two H_2 molecules. The lowest ground state of the PES describes the dimerization. (b) Cubic interpolation of the PES with force vectors for the H_{10} chain. Red arrows depict force vectors computed by DeepErwin via the Hellmann-Feynman theorem, whereas black arrows represent numerical gradients that are based on finite differences of MRCI-F12 reference calculations. (c) Forces computed by DeepErwin plotted against the respective forces obtained from finite differences of the MRCI-F12 reference calculation. Figure is taken from Scherbela et al. (2022).

Furthermore, one can approximate the molecular ground state $\Psi^0(\mathbf{R}, \mathbf{r})$ by the separated product $\Psi^0(\mathbf{R}, \mathbf{r}) \sim \Psi^{\mathrm{n},0}_{\mathbf{Z}}(\mathbf{R}) \cdot \Psi^{\mathrm{e},0}_{(\mathbf{R},\mathbf{Z})}(\mathbf{r})$, where $\Psi^{\mathrm{n},0}_{\mathbf{Z}}(\mathbf{R})$ solves the time-independent nuclear Schrödinger Equation

$$\Psi^{\mathrm{n}} = \lambda \left(-\sum_{I}^{n_{\mathrm{nuc}}} \frac{1}{2M_I} \nabla^2_{\mathbf{R}_I} + \mathcal{E}_{\mathbf{Z}} \right) \Psi^{\mathrm{n}}, \quad \lambda \in \mathbb{R}$$

with minimal eigenvalue. The degree to which these approximations are valid depends on the structure of the eigenvalues of the electronic Hamiltonian, see for example Jecko (2014) and the references therein. At any rate, it should be clear that the PES is an extremely important quantity since it allows for the computation of properties of molecules without having to conduct actual experiments.

A grand goal of computational chemistry is thus as follows.

Find efficient and accurate algorithms for evaluating the PES $(\mathbf{R}, \mathbf{Z}) \mapsto \mathcal{E}_{\mathbf{Z}}(\mathbf{R})$

Efficient evaluation of the PES allows, for example, to perform a computational structure search or to compute chemical reaction rates. However, for these types of tasks, highly accurate estimates of the ground-state energy become essential. For example, a study by Barone et al. (2013) demonstrated that the geometrical conformations of the smallest amino acid, Glycine, are partitioned by transition barriers of only ∼20 millihartree (mHa). To put the scale of the transition barrier into perspective, the ground-state energy is typically several Hartree (see Fig. 2 for a visual representation).

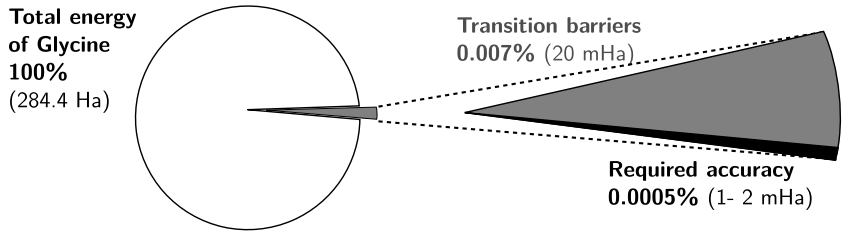

FIGURE 2 Conceptual visualization of the relevant energy-scales on the example of Glycine. Total energies are often hundreds of Hartrees, whereas required accuracy to predict experiments is often only milli-Hartrees. Figure adapted from Gerard et al. (2022).

The requirement of having to achieve very high accuracy in combination with the high dimensionality of the problem (4) (the dimension being $3\times$ the number of electrons) renders the efficient evaluation of the PES a formidable challenge. An additional complication is given by the *Pauli Exclusion Principle*, which posits that the electronic wave function must satisfy certain antisymmetry conditions, see Section 2.3 below.

As already mentioned, a plethora of methods have been developed over the last decades to solve the Schrödinger Equation approximately. Some methods such as the Hartree-Fock (HF) method are computationally cheap and scale well with system size, but are only accurate for a limited class of systems. Other methods such as Configuration Interaction (CI) and Coupled Cluster (CC) often yield highly-accurate results that are in good agreement with experiments, but scale poorly with system size and are thus limited to small systems (cf. Fig. 3). Density Functional Theory (DFT) has emerged as a breakthrough, and has been awarded the Nobel Prize in chemistry in 1998, since it yields surprisingly high accuracy while scaling well with system size. However, DFT requires the use of an essentially uncontrolled approximation, which can fail for many systems of interest and is thus not universally applicable.

Broadly speaking, the aim of DL-VMC is to represent the wave function as a neural network Ψ_θ and aim to approximate the electronic ground state by minimizing the Raleigh-Ritz quotient corresponding to the Eigenvalue problem (4), which amounts to minimizing the loss

$$\mathcal{L}(\theta) := \frac{\langle \mathcal{H}^{\mathrm{BO}} \Psi_\theta, \Psi_\theta \rangle}{\langle \Psi_\theta, \Psi_\theta \rangle}.$$

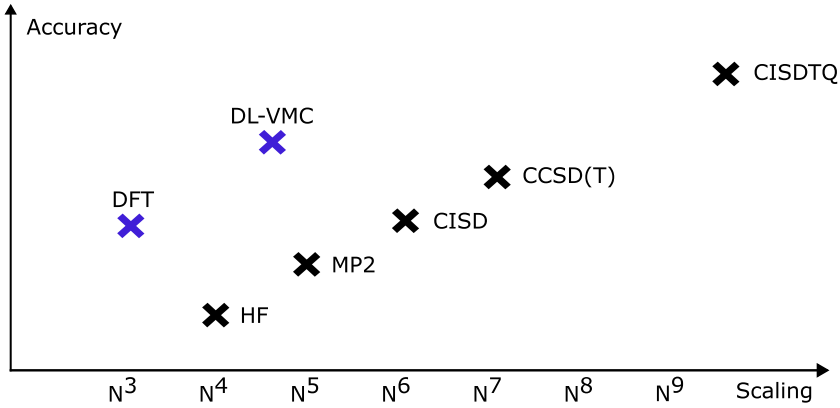

FIGURE 3 A conceptual visualization of the computational scaling with respect to the estimated accuracy for different Quantum Chemistry methods. The abbreviations stand for DFT: Density Functional Theory; DL-VMC: Deep-Learning-based Variational Monte Carl; HF: Hartree-Fock; MP2: Møller-Plesset 2nd order; CISD: Configuration Interaction (doubles); CCSD(T): Coupled-Cluster; CISDTQ: Configuration Interaction (quadruples). The figure is adapted from Hermann et al. (2022).

In recent years it has been demonstrated empirically (see for example Fig. 8) that a judiciously chosen DL-VMC ansatz is capable of achieving considerably more accurate ground state energies than previous methods, bringing us closer to the aforementioned "grand goal".

1.2 Outline

In Section 2 we provide a mathematical introduction into the spectral theory of the electronic Schrödinger equation. In particular, we will see that the electronic Hamiltonian is self-adjoint and induces a bounded and (up to a translation) coercive bilinear form on H^1, characterizes its spectrum, and elaborates on the Pauli exclusion principle.

Section 3 provides a concise introduction into variational Monte Carlo (VMC) and introduces ansatz spaces of functions that satisfy the Pauli exclusion principle. We furthermore present results and techniques related to efficient Markov chain Monte Carlo (MCMC) quadrature methods and preconditioning methods for optimization.

In Section 4 we describe various neural network architectures used in DL-VMC. This includes constructions that are transferable among different geometric conformations.

Finally, in Section 5 we review numerical results.

1.3 Notation

We denote with $L^2(\mathbb{R}^d)$, $H^k(\mathbb{R}^d)$ the usual Lebesgue space, resp. Sobolev spaces, see Evans (2022). The symbols $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ typically denote the norm

and inner product in L^2, unless stated otherwise. The inner product of two vectors $v, w \in \mathbb{R}^d$ will be denoted by $v \cdot w$ and the norm of $v \in \mathbb{R}^d$ by $|v|$.

2 Mathematical preliminaries

So far the main problem (4) does not stand on a rigorous mathematical footing. Specifically, it is a priori not clear in which sense the eigenvalue equation (4) is well-posed, what would be the correct domain of definition of $\mathcal{H}^{\mathrm{BO}}$, whether this operator is self adjoint and whether an isolated minimal eigenvalue exists. Addressing these issues requires studying the spectral theory of the electronic Hamiltonian $\mathcal{H}^{\mathrm{BO}}_{(\mathbf{R},\mathbf{Z})}$. This will be done in Section 2.2 after a brief introduction into some mathematical foundations of quantum mechanics in Section 2.1.

Furthermore, electrons carry an additional variable, namely the *spin*, which can assume the values $\pm\frac{1}{2}$. Therefore, the electronic wave function actually depends on *spin coordinates*

$$\left((\mathbf{r}_1, \sigma_1), \ldots, (\mathbf{r}_{n_{\mathrm{el}}}, \sigma_{n_{\mathrm{el}}})\right), \quad \sigma = (\sigma_1, \ldots, \sigma_{n_{\mathrm{el}}}) \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}^{n_{\mathrm{el}}}.$$

The *Pauli Exclusion Principle* states that every admissible wave function must be antisymmetric with respect to permutations of different spin coordinates. We elaborate on the consequences in Section 2.3.

For notational convenience we will often drop sub- and superscripts such as in $\mathcal{H}^{\mathrm{BO}}$. However, it is important to note that all objects that follow will depend on the geometric conformation $(\mathbf{R}, \mathbf{Z})$.

2.1 Basic mathematical setting

In classical mechanics, the state of a single particle in $\mathbb{R}$ is completely determined as a point (x, p) in phase-space $\mathbb{R}^2$. Its position is uniquely determined as x, its momentum as p and, if it has mass m and is subjected to a force field $F(x) = -\nabla V(x)$, its energy as

$$E = \frac{p^2}{2m} + V(x). \tag{5}$$

Its motion is governed by Newton's law $m\ddot{x}(t) = F(x(t))$, which conserves its energy.

Quantum states are modeled as elements of a complex Hilbert space H. Two elements $\Psi_1, \Psi_2 \in H$ represent the same state if there is $c \in \mathbb{C} \setminus \{0\}$ with $\Psi_1 = c\Psi_2$ and we can therefore assume that states are represented by unit vectors in H.

The position of a single particle

As an example we may consider $H = L^2(\mathbb{R})$, where each $\Psi \in L^2(\mathbb{R})$ with $\|\Psi\| = 1$ describes the wave function of a single particle. Similar to classical

mechanics, we would like to infer properties of quantum states. For example, for a wave function $\Psi \in L^2(\mathbb{R})$ of a single particle we may be interested in determining its position, its momentum or its energy, etc. However, as it is well-known, properties of quantum states are typically not fully determined. What quantum mechanics provides instead is a probability distribution that models the likelihood of a particular observable assuming its value in a specific set. Returning to our example of $\Psi \in L^2(\mathbb{R})$ modeling a single particle in $\mathbb{R}$ we may define the probability that the position of Ψ lies in a measurable set $E \subset \mathbb{R}$ as

$$\mathbb{P}\big[\text{The position of } \Psi \text{ is observed in } E\big] = \int_E |\Psi(x)|^2 dx,$$

implying that the expected position is given as

$$\mathbb{E}\left[\text{Position of } \Psi\right] = \int_{\mathbb{R}} x|\Psi(x)|^2 dx = \langle \mathcal{X}\Psi, \Psi\rangle$$

with $\mathcal{X}\Psi := (x \mapsto x \cdot \Psi(x))$ denoting the *position operator*, which is defined on the dense subset $\mathcal{D}(\mathcal{X}) := \left\{\Psi \in L^2(\mathbb{R}) : \ x \cdot \Psi(x) \in L^2(\mathbb{R})\right\} \subset L^2(\mathbb{R})$. For $E \subset \mathbb{R}$ measurable define the projection operator $\mu^{\mathcal{X}}(E) : \Psi(x) \mapsto \chi_E(x)\Psi(x)$ with χ_E denoting the indicator function of E. The mapping $E \mapsto \mu^{\mathcal{X}}(E)$ is called *projection-valued measure*. This projection-valued measure allows to formally decompose $\mathcal{X}$ as

$$\mathcal{X} = \int_{\mathbb{R}} \lambda d\mu^{\mathcal{X}}(\lambda) := \lim_{h\to 0} \sum_{i\in\mathbb{Z}} hi\mu^{\mathcal{X}}([ih, (i+1)h]). \tag{6}$$

It is not hard to verify that this Riemann sum converges in the sense that

$$\lim_{h\to 0} \left\langle \sum_{i\in\mathbb{Z}} hi\mu^{\mathcal{X}}([ih, (i+1)h])\Psi, \Psi \right\rangle = \langle \mathcal{X}\Psi, \Psi\rangle \quad \forall \Psi \in \mathcal{D}(\mathcal{X}).$$

Given our projection-valued measure we also have

$$\mathbb{P}[\mathcal{X} \in E] = \langle \mu^{\mathcal{X}}(E)\Psi, \Psi\rangle. \tag{7}$$

The decomposition (6) also defines a functional calculus that allows for the study of functions of $\mathcal{X}$: indeed for a measurable function $f : \mathbb{R} \to \mathbb{R}$ we may define the operator

$$f(\mathcal{X}) := \int_{\mathbb{R}} f(\lambda) d\mu^{\mathcal{X}}(\lambda). \tag{8}$$

In our example it can be seen that $f(\mathcal{X})\Psi(x) = f(x) \cdot \Psi(x)$, as expected. Finally we have that

$$\mathbb{E}\left[f\left(\text{Position of } \Psi\right)\right] = \langle f(\mathcal{X})\Psi, \Psi\rangle =: \langle f(\mathcal{X})\rangle_\Psi \tag{9}$$

which provides a simple expression for moments of the distribution of the position of Ψ. For instance, the variance of the position of Ψ is given by $\mathbb{V}[\text{Position of } \Psi] = \left\langle (\mathcal{X} - \langle \mathcal{X} \rangle_\Psi)^2 \right\rangle_\Psi$.

Self-adjoint operators

The properties of having a decomposition (6) which defines a probability distribution (7) and a functional calculus (8) are enjoyed by every self-adjoint operator.

Definition 2.1. Let $\mathcal{A}$ be a linear operator on a Hilbert space H, defined on a dense subset $\mathcal{D}(\mathcal{A}) \subset H$. Let

$$\mathcal{D}(\mathcal{A}^*) := \{\Phi \in H : \exists C \in [0, \infty) : \forall \Psi \in \mathcal{D}(\mathcal{A}) : |\langle \Phi, \mathcal{A}\Psi \rangle| \leqslant C \cdot \|\Psi\|\}. \tag{10}$$

In other words, $\mathcal{D}(\mathcal{A}^*)$ consists of those $\Phi \in H$ such that the linear functional $L_\Phi : \Psi \mapsto \langle \Phi, \mathcal{A}\Psi \rangle$ is bounded on the dense set $\mathcal{D}$. The boundedness implies that L_Φ can be extended to a bounded linear functional on all of H and by the Riesz representation theorem there exists a unique $\chi \in H$ with $L_\Phi(\cdot) = \langle \chi, \cdot \rangle$. We can then define the *adjoint operator* by $\mathcal{A}^*\Psi = \chi$. $\mathcal{A}$ is called *symmetric* if $\mathcal{A} = \mathcal{A}^*$ on $\mathcal{D}(\mathcal{A})$ and *self-adjoint* if additionally $\mathcal{D}(\mathcal{A}) = \mathcal{D}(\mathcal{A}^*)$. Any self-adjoint operator is called a *quantum observable*.

Remark 2.2. Note that for every symmetric operator it holds that $\mathcal{D}(\mathcal{A}) \subset \mathcal{D}(\mathcal{A}^*)$. This is because for $\Phi, \Psi \in \mathcal{D}(\mathcal{A})$ we have that

$$\langle \Phi, \mathcal{A}\Psi \rangle = \langle \mathcal{A}\Phi, \Psi \rangle \leqslant \|\mathcal{A}\Phi\| \cdot \|\Psi\|.$$

Self-adjointness is a somewhat subtle definition since it involves a correct specification of the domain of definition of $\mathcal{A}$. It is however the right concept in the context of quantum mechanics, as evidenced by the spectral theorem that we only state in an informal way. For more details please consult the excellent monograph by Hall (2013).

Definition 2.3. The *resolvent set* $\rho(\mathcal{A})$ of an operator $\mathcal{A} : \mathcal{D}(\mathcal{A}) \to \mathcal{H}$ consists of all $\lambda \in \mathbb{C}$ such that $\mathcal{A} - \lambda : \mathcal{D}(\mathcal{A}) \to H$ is boundedly invertible. The spectrum $\sigma(\mathcal{A})$ of $\mathcal{A}$ is defined as

$$\sigma(\mathcal{A}) = \mathbb{C} \setminus \rho(\mathcal{A}).$$

$\lambda \in \mathbb{C}$ is called an Eigenvalue of $\mathcal{A}$ if there is $\Psi \in \mathcal{D}(\mathcal{A}) \setminus \{0\}$ with

$$\mathcal{A}\Psi = \lambda\Psi. \tag{11}$$

The subspace $E_\lambda \subset H$ containing all $\Psi \in H$ such that (11) holds is called the Eigenspace of $\mathcal{A}$ corresponding to the Eigenvalue λ. All $\Psi \in E_\lambda \setminus \{0\}$ are called Eigenvector of $\mathcal{A}$ corresponding to the Eigenvalue λ. The dimension $m_\lambda := \dim(E_\lambda)$ of E_λ is called the multiplicity of the Eigenvalue λ and λ is

of finite multiplicity if $m_\lambda \in \mathbb{N}$. An Eigenvalue λ is called isolated if there is an open neighborhood B of λ in $\mathbb{C}$ with $\sigma(\mathcal{A}) \cap B = \{\lambda\}$. If $\mathcal{A}$ is self-adjoint, the discrete spectrum $\sigma_{\text{disc}}(\mathcal{A})$ consists of all isolated Eigenvalues with finite multiplicity and the essential spectrum equals $\sigma_{\text{ess}}(\mathcal{A}) := \sigma(\mathcal{A}) \setminus \sigma_{\text{disc}}(\mathcal{A})$.

If $\mathcal{A}$ is self-adjoint, then the spectrum has the following intuitive characterization of consisting of those values λ for which there are states that are "almost" Eigenvectors.

Lemma 2.4. *For a self-adjoint operator $\mathcal{A}$ it holds that*

$$\lambda \in \sigma(\mathcal{A}) \Leftrightarrow \forall \varepsilon \in (0, \infty)\ \exists \Psi \in \mathcal{D}(\mathcal{A}) :\ \|(\mathcal{A} - \lambda)\,\Psi\| \leqslant \varepsilon \|\Psi\|. \tag{12}$$

Proof. If the right hand side of (12) holds true, then $\mathcal{A} - \lambda$ cannot be boundedly invertible.

On the other hand, suppose that λ does not satisfy the right hand side of (12). Then there exists $\delta \in (0, \infty)$ with

$$\|(\mathcal{A} - \lambda)\Psi\| \geqslant \delta \|\Psi\| \quad \forall \Psi \in \mathcal{D}(\mathcal{A}). \tag{13}$$

This clearly implies that $\mathcal{A} - \lambda$ is injective. It also has closed range because if we have a Cauchy sequence $\Phi_n = (\mathcal{A} - \lambda)\Psi_n$ in the range of $\mathcal{A} - \lambda$ with $\Phi = \lim_{n\to\infty} \Phi_n$, then by (13) the limit $\Psi = \lim_{n\to\infty} \Psi_n$ exists.

We now show that $\Psi \in \mathcal{D}(\mathcal{A} - \lambda)$ and $(\mathcal{A} - \lambda)\Psi = \Phi$. Let $\mathcal{B} := \mathcal{A} - \lambda$. Then $\mathcal{B}$ is self-adjoint. Let $\Xi \in \mathcal{D}(\mathcal{A})$ be arbitrary. Then

$$\langle \Psi, \mathcal{B}\Xi \rangle = \lim_{n\to\infty} \langle \Psi_n, \mathcal{B}\Xi \rangle = \lim_{n\to\infty} \langle \mathcal{B}\Psi_n, \Xi \rangle = \langle \Phi, \Xi \rangle$$

Therefore it holds that $\Psi \in \mathcal{D}(\mathcal{B}^*) = \mathcal{D}(\mathcal{B})$ and $\mathcal{B}\Psi = \mathcal{B}^*\Psi = \Phi$.

We still need to show that $\mathcal{B}$ is surjective. To this end suppose that $\Xi \in \text{range}(\mathcal{B})^\perp$. Then

$$\|\mathcal{B}\Xi\|^2 = \langle \mathcal{B}\Xi, \mathcal{B}\Xi \rangle = \langle \Xi, \mathcal{B}^*\mathcal{B}\Xi \rangle = 0.$$

By (13) this implies that $\Xi = 0$. Therefore, the range of $\mathcal{B}$ is dense and closed and hence $\mathcal{B}$ is surjective. By (13) we conclude by the Closed Graph Theorem that $\mathcal{B} = \mathcal{A} - \lambda$ is boundedly invertible. □

Remark 2.5. An examination of the proof of Lemma 2.4 shows the useful property that every self-adjoint operator is *closed*, that is, the set $\{(\Psi, \mathcal{A}\Psi) : \Psi \in \mathcal{D}(\mathcal{A})\}$ is closed in $H \times H$.

Furthermore, one can easily deduce from Lemma 2.4 that the spectrum is always closed.

By working a bit harder one can obtain a more detailed characterization of the essential spectrum of self-adjoint operators.

Theorem 2.6. *For a self-adjoint operator $\mathcal{A}$ it holds that*

$$\lambda \in \sigma_{ess}(\mathcal{A}) \Leftrightarrow \forall n \in \mathbb{N}\ \exists \Psi^n \in \mathcal{D}(\mathcal{A}) : \left\| (\mathcal{A} - \lambda)\, \Psi^n \right\| \leqslant \frac{1}{n} \cdot \|\Psi^n\| \wedge$$
$$\forall \Xi \in H : \lim_{n\to\infty} \langle \Psi^n, \Xi \rangle = 0. \tag{14}$$

Proof. $\Rightarrow$**:** Suppose that $\lambda \in \sigma_{ess}(\mathcal{A})$.

We distinguish three cases.

Case 1: λ is an Eigenvalue with infinite multiplicity. Then there exist countably many orthonormal corresponding Eigenvectors $(\Psi^n)_{n\in\mathbb{N}}$. Let $\Xi \in H$. Then $\sum_{n=1}^{\infty} \langle \Psi^n, \Xi \rangle^2 \leqslant \|\Xi\|^2$ which implies that $\lim_{n\to\infty} \langle \Psi^n, \Xi \rangle = 0$.

Case 2: λ is not an Eigenvalue. Since $\lambda \in \sigma(\mathcal{A})$ there exists by Lemma 2.4 a sequence $\Psi^n \in H$ with $\|\Psi^n\| = 1$ and $\|(\mathcal{A} - \lambda)\Psi^n\| \leqslant \frac{1}{n}$. Since the Ψ^n's are bounded, by the Banach-Alaoglu Theorem there exists a weak limit (possibly after passing to a subsequence), e.g. some $\Omega \in H$ with $\lim_{n\to\infty} \langle \Psi^n, \Xi \rangle = \langle \Omega, \Xi \rangle$ for all $\Xi \in H$. It holds for all $\Xi \in \mathcal{D}(\mathcal{A} - \lambda)$ that

$$\langle \Omega, (\mathcal{A} - \lambda)\Xi \rangle = \lim_{n\to\infty} \langle \Psi^n, (\mathcal{A} - \lambda)\Xi \rangle = \lim_{n\to\infty} \langle (\mathcal{A} - \lambda)\Psi^n, \Xi \rangle = 0.$$

Therefore it holds that $\Omega \in \mathcal{D}((\mathcal{A} - \lambda)^*) = \mathcal{D}(\mathcal{A} - \lambda)$ and $(\mathcal{A} - \lambda)^*\Omega = (\mathcal{A} - \lambda)\Omega = 0$. This implies that either Ω is an Eigenvalue or $\Omega = 0$. Since we already excluded the former, it must hold that $\Omega = 0$ which is what we wanted to show.

Case 3: λ is an Eigenvalue with finite multiplicity that is not isolated. W.l.o.g. assume $\lambda \neq 0$. Let $\mu^{\mathcal{A}}(\{\lambda\})$ be the orthogonal projection on the corresponding Eigenspace. Consider $\tilde{\mathcal{A}} := \mathcal{A} - \lambda\mu^{\mathcal{A}}(\{\lambda\})$. Then λ cannot be an Eigenvalue of $\tilde{\mathcal{A}}$. Since it holds that $\sigma(\mathcal{A}) \setminus \{\lambda, 0\} = \sigma(\tilde{\mathcal{A}}) \setminus \{\lambda, 0\}$ and due to the assumption that λ is an accumulation point of $\sigma(\mathcal{A})$, λ must be an accumulation point of $\sigma(\tilde{\mathcal{A}})$. Therefore (since the spectrum is always closed) it must hold that $\lambda \in \sigma(\tilde{\mathcal{A}})$. By Case 2 there must thus exist Ψ^n with $\|\Psi^n\| = 1$, $\|(\tilde{A} - \lambda)\Psi^n\| \leqslant \frac{1}{n}$ and $\forall \Xi \in H :\ \lim_{n\to\infty} \langle \Psi^n, \Xi \rangle = 0$. The last condition implies that for all $\Xi \in H :\ \lim_{n\to\infty} \langle \mu^{\mathcal{A}}(\{\lambda\})\Psi^n, \Xi \rangle = \lim_{n\to\infty} \langle \Psi^n, \mu^{\mathcal{A}}(\{\lambda\})\Xi \rangle = 0$ and, since the range of $\mu^{\mathcal{A}}$ is finite-dimensional, together with the fact that on finite dimensional spaces strong and weak convergence coincide, it follows that $\lim_{n\to\infty} \|\mu^{\mathcal{A}}(\{\lambda\})\Psi^n\| = 0$. This implies that $\lim_{n\to\infty} \|(\mathcal{A} - \lambda)\Psi^n\| = 0$ which proves that the right hand side of (14) holds true.

$\Leftarrow$**:** For the other direction we assume that $\lambda \in \sigma_{\text{disc}}(\mathcal{A})$ and suppose w.l.o.g that $\lambda \neq 0$. Define again $\tilde{\mathcal{A}} := \mathcal{A} - \lambda\mu^{\mathcal{A}}(\{\lambda\})$. Since λ is an isolated Eigenvalue of $\mathcal{A}$, it holds that $\lambda \notin \sigma(\tilde{\mathcal{A}})$. This can be argued as follows. If λ was in the spectrum of $\tilde{\mathcal{A}}$, then λ would necessarily have to be isolated. This implies that λ would be an Eigenvalue of $\tilde{\mathcal{A}}$ (the reader can verify this as an exercise. It will probably be convenient to use the Spectral Theorem 2.11). But this would imply that there exists an Eigenvector of $\mathcal{A}$ that lies in the orthogonal complement of the Eigenspace range$(\mu^{\mathcal{A}}(\{\lambda\}))$ of $\mathcal{A}$ w.r.t. λ, which is a contradiction. Therefore, the right hand side of (14) cannot hold for the operator $\tilde{\mathcal{A}}$ and λ.

Now assume that the right hand side of (14) holds true for the operator $\mathcal{A}$ and λ. Then, using the fact that the range of $\mu^{\mathcal{A}}(\{\lambda\})$ is finite dimensional, we can argue exactly as in Case 3 above to deduce that the right hand side of (14) must also hold for $\tilde{\mathcal{A}}$ and λ, which would give a contradiction. □

Corollary 2.7. *For a self-adjoint operator $\mathcal{A}$ it holds that*

$$\sigma(\mathcal{A}) \subset \mathbb{R}. \tag{15}$$

Proof. Let $\lambda = a + ib$ with $b \neq 0$ and $\Psi \in \mathcal{D}(\mathcal{A})$. Then it holds that

$$\begin{aligned}\langle (\mathcal{A}-\lambda)\,\Psi, (\mathcal{A}-\lambda)\,\Psi\rangle &= \langle (\mathcal{A}-a)\,\Psi, (\mathcal{A}-a)\,\Psi\rangle + ib\,\langle \Psi, (\mathcal{A}-\lambda)\,\Psi\rangle \\ &\quad + ib\,\langle \Psi, (\mathcal{A}-\lambda)\,\Psi\rangle + b^2\|\Psi\|^2 \\ &= \langle (\mathcal{A}-a)\,\Psi, (\mathcal{A}-a)\,\Psi\rangle + b^2\|\Psi\|^2 \geqslant b^2\|\Psi\|^2,\end{aligned}$$

where the last equality follows from the self-adjointness of $\mathcal{A}$. Due to Lemma 2.4 it follows that λ cannot be in the spectrum. □

Remark 2.8. It might be instructive to go back to the position operator $\mathcal{X}$ of a single particle. It is not hard to check that $\mathcal{X}$ is self-adjoint. Its spectrum is all of $\mathbb{R}$ but there are no eigenvalues. Indeed, for $\lambda \in \mathbb{R}$, any Ψ with $(\mathcal{X}-\lambda)\Psi = 0$ would have to be vanished on all of $\mathbb{R} \setminus \{\lambda\}$ which is impossible. On the other hand, it is easy to construct function Ψ with $\|\Psi\| = 1$ which are arbitrarily well localized near λ in the sense of Ψ being supported on $[\lambda - \varepsilon/2, \lambda + \varepsilon/2]$. These functions then satisfy the right hand side of (12) and therefore λ must be in the spectrum of $\mathcal{X}$. Furthermore it is easy to see that these functions converge weakly to 0 as $\varepsilon \to 0$. Therefore, $\sigma(\mathcal{X}) = \sigma_{\text{ess}}(\mathcal{X}) = \mathbb{R}$.

The spectral theorem

We will now state the spectral theorem for unbounded self-adjoint operators. Recall the definition of a projection-valued measure.

Definition 2.9. A mapping μ from the Borel sigma algebra on a set $A \subset \mathbb{R}$ to the set of orthogonal projections on H is called a *projection-valued measure* supported on A if

1. $\mu(\varnothing) = 0$ and $\mu(A) = I$,
2. For all disjoint and measurable subsets $(E_i)_{i\in\mathbb{N}}$ of A it holds that $\mu(\bigcup_{i\in\mathbb{N}} E_i) = \sum_{i\in\mathbb{N}} \mu(E_i)$, and
3. For E_1, E_2 measurable we have $\mu(E_1 \cap E_2) = \mu(E_1)\mu(E_2)$.

For $\Psi \in H$ we denote $\mu_\Psi : E \mapsto \langle \mu(E)\Psi, \Psi\rangle$ and note that μ_Ψ is a positive Borel measure on A.

We have the following important result

Theorem 2.10. *For μ a projection-valued measure supported on A and $f : A \to \mathbb{C}$ measurable, let*

$$\mathcal{D}_f := \left\{ \Psi \in H : \int_A |f(\lambda)|^2 d\mu_\Psi(\lambda) < \infty \right\}.$$

Then $\mathcal{D}_f$ is dense in H and there exists a unique (unbounded) operator, denoted $\int_A f(\lambda) d\mu(\lambda)$ with $\mathcal{D}\left(\int_A f(\lambda) d\mu(\lambda)\right) = \mathcal{D}_f$ and

$$\left\langle \Psi, \int_A f(\lambda) d\mu(\lambda) \Psi \right\rangle = \int_A f(\lambda) d\mu_\Psi(\lambda). \tag{16}$$

Furthermore it holds that

$$\int_A f(\lambda) \cdot g(\lambda) d\mu(\lambda) = \int_A f(\lambda) d\mu(\lambda) \int_A g(\lambda) d\mu(\lambda) \tag{17}$$

and

$$\int_A \overline{f(\lambda)} d\mu(\lambda) = \left(\int_A f(\lambda) d\mu(\lambda) \right)^*. \tag{18}$$

Proof. See Hall (2013, Proposition 10.1 and Proposition 10.2). □

We can now state the spectral theorem.

Theorem 2.11 (Spectral Theorem for unbounded self-adjoint Operators). *For every self-adjoint operator $\mathcal{A}$ the spectrum $\sigma(\mathcal{A}) \subset \mathbb{R}$ there exists a unique spectral measure $\mu^{\mathcal{A}}$ supported on $\sigma(\mathcal{A})$ such that*

$$\mathcal{A} = \int_{\sigma(\mathcal{A})} \lambda d\mu^{\mathcal{A}}(\lambda). \tag{19}$$

Proof. The proof is quite long and involved. See Hall (2013, Section 10) for an excellent exposition. □

The spectral theorem readily allows for the definition of a functional calculus on self-adjoint operators.

Definition 2.12 (Functional calculus). Let $\mathcal{A}$ be self-adjoint with spectral measure $\mu^{\mathcal{A}}$. For $f : \sigma(\mathcal{A}) \to \mathbb{C}$ measurable we define

$$f(\mathcal{A}) := \int_{\sigma(\mathcal{A})} f(\lambda) d\mu^{\mathcal{A}}(\lambda).$$

We have now shown that all previous desirable properties of the position operator $\mathcal{X}$ can be established for every self-adjoint operator. This motivates the following definition.

Definition 2.13. A self-adjoint operator $\mathcal{A}$ on a Hilbert space H is called a *quantum observable* on H.

A quantum observable thus observes (or measures) a certain property of a state Ψ. The outcome of this measurement is random and the spectral measure represents the probability measure on the measurement outcome lying in a set E if the system is in the state Ψ with $\|\Psi\| = 1$:

$$\mathbb{P}[\text{The measurement } \mathcal{A} \text{ of a state } \Psi \text{ lies in } E] = \mu_\Psi^{\mathcal{A}}(E).$$

Remark 2.14. Note that a measurement can only be realized deterministically if the measure $\mu_\Psi^{\mathcal{A}}$ is concentrated in a single point λ. This can only occur if λ is an Eigenvalue of $\mathcal{A}$ and Ψ a corresponding Eigenvector.

We close this paragraph with the following result that will be used later to turn the Eigenvalue problem into an optimization problem.

Lemma 2.15. *For a self-adjoint operator $\mathcal{A}$ let $\mathcal{W}(\mathcal{A}) := \{\Psi \in H : \langle \mathcal{A}\Psi, \Psi\rangle < \infty\}$. Then it holds that*

$$\inf \sigma(\mathcal{A}) = \inf\{\langle \mathcal{A}\Psi, \Psi\rangle : \Psi \in \mathcal{W}(\mathcal{A}),\ \|\Psi\| = 1\}. \tag{20}$$

Proof. Let $\lambda \in \sigma(\mathcal{A})$. Then, by Lemma 2.4 for every $\varepsilon > 0$ there is $\Psi \in \mathcal{D}((A))$ with $\|\Psi\| = 1$ and $\|(\mathcal{A} - \lambda)\Psi\| \leqslant \varepsilon$. Therefore, by Cauchy-Schwarz

$$|\langle \mathcal{A}\Psi, \Psi\rangle - \lambda| = |\langle (\mathcal{A} - \lambda)\,\Psi, \Psi\rangle| \leqslant \varepsilon$$

This implies that

$$\begin{aligned}\inf\{\langle \mathcal{A}\Psi, \Psi\rangle : \Psi \in \mathcal{W}(\mathcal{A}),\ \|\Psi\| = 1\}& \\ \leqslant \inf\{\langle \mathcal{A}\Psi, \Psi\rangle : \Psi \in \mathcal{D}(\mathcal{A}), \|\Psi\| = 1\} &\leqslant \inf\sigma(\mathcal{A}).\end{aligned}$$

For the other direction we note that by the spectral theorem we have for $\Psi \in H$ with $\|\Psi\| = 1$ that

$$\langle \mathcal{A}\Psi, \Psi\rangle = \int_{\sigma(\mathcal{A})} \lambda d\mu_\Psi^{\mathcal{A}}(\lambda) \geqslant \inf\sigma(\mathcal{A}) \cdot \int_{\sigma(\mathcal{A})} 1 d\mu_\Psi^{\mathcal{A}}(\lambda) = \inf\sigma(\mathcal{A})$$

which implies that

$$\inf\{\langle \mathcal{A}\Psi, \Psi\rangle : \Psi \in \mathcal{W}(\mathcal{A}),\ \|\Psi\| = 1\} \geqslant \inf\sigma(\mathcal{A}). \qquad \square$$

Other observables and the Schrödinger equation

Returning to our example of a particle moving in $\mathbb{R}$ we would now like to find a quantum mechanical analog of other observables, such as momentum or energy. The momentum operator is defined as $\mathcal{P} := -i\frac{d}{dx}$. Intuitively this makes sense

because $\mathcal{P}$ is simply the position operator in the Fourier domain. This also shows that $\mathcal{P}$ is self-adjoint. Having position and momentum operators we can now define the energy operator, or the Hamiltonian in analogy with (5) via

$$\mathcal{H} = \frac{\mathcal{P}^2}{2m} + V(\mathcal{X}) = -\frac{1}{2m}\frac{d^2}{dx^2} + V(\mathcal{X}).$$

Remark 2.16. The process of constructing self-adjoint quantum observables from classical observables $f(x, p)$ is called *quantization*. Due to the fact that the position and momentum operators do not commute, this is a highly nontrivial problem, see Hall (2013).

One can show that the Hamiltonian is self-adjoint (this is not completely trivial due to the conditions on the domains of self-adjoint operators!) and therefore, the energy of a state $\Phi \in \mathcal{D}(\mathcal{H})$ with $\|\Psi\| = 1$ is distributed according to $\mu_\Psi^{\mathcal{H}}$.

The evolution of quantum systems is governed by the Schrödinger equation

$$\dot{\Psi}(t) = \frac{1}{i}\mathcal{H}\Psi(t). \tag{21}$$

Using our functional calculus one can show that $\Phi(t) = e^{-it\mathcal{H}}\Psi(0)$ and that $U(t) = e^{-it\mathcal{H}}$ is a unitary one-parameter group which is defined on all of $L^2(\mathbb{R})$ (this is one part of the famous Stone-von Neumann Theorem (Stone, 1932; von Neumann, 1932)). Using (21) we can now model a quantum particle moving in $\mathbb{R}$.

Molecules consist of several particles, each with coordinates in $\mathbb{R}^3$. It is straightforward how the position and momentum operators can be extended to particles in $\mathbb{R}^3$. This naturally leads to the definition of the molecular Hamiltonian $\mathcal{H}^{\text{mol}}$ and $\mathcal{H}^{\text{BO}}$ from the introduction which is our main interest of study.

2.2 The Hamiltonian of the electronic Schrödinger equation

At this point it is not clear that the electronic Hamiltonian $\mathcal{H}^{\text{BO}}_{(\mathbf{R},\mathbf{Z})}$ is self-adjoint. Furthermore, we would like to know more about the precise nature of the spectrum of $\mathcal{H}^{\text{BO}}_{(\mathbf{R},\mathbf{Z})}$. Addressing these issues are the subject of this section. For notational convenience we will omit the subscript describing the geometric conformation and simply write $\mathcal{H} = \mathcal{H}^{\text{BO}}_{(\mathbf{R},\mathbf{Z})}$.

Our main tool in asserting the self-adjointness of $\mathcal{H}$ will be the Kato-Rellich theorem.

Theorem 2.17 (Kato-Rellich Theorem). *Let $\mathcal{A}$, $\mathcal{B}$ be self-adjoint with $\mathcal{D}(\mathcal{A}) \subset \mathcal{D}(\mathcal{B})$. Suppose that there are $a \in (0, 1)$ and $b \in (0, \infty)$ with*

$$\|\mathcal{B}\Psi\| \leqslant a\|\mathcal{A}\Psi\| + b\|\Psi\| \quad \forall \Psi \in \mathcal{D}(\mathcal{A}).$$

Then $\mathcal{A} + \mathcal{B}$ is self-adjoint on $\mathcal{D}(\mathcal{A})$.

Proof. See Kato (2013, Section V, Theorem 4.13) or Hall (2013, Theorem 9.37). □

We rewrite the operator $\mathcal{H}$ of (3) as

$$\mathcal{H} = -\frac{1}{2}\Delta + V(\mathbf{r})$$

and aim to apply the Kato-Rellich Theorem with $\mathcal{A} = -\frac{1}{2}\Delta$ and $\mathcal{B} = V(\mathbf{r})$. To this end the following result will prove useful.

Proposition 2.18 (Hardy Inequality). *For any smooth and compactly supported function u on $\mathbb{R}^3$ it holds that*

$$\int_{\mathbb{R}^3} \frac{1}{|r|^2} u(r)^2 dr \leqslant 4 \cdot \int_{\mathbb{R}^3} |\nabla u(r)|^2 dr.$$

Proof. See Yserentant (2010, Lemma 4.1). □

Theorem 2.19. *The operator $\mathcal{H}$ is self-adjoint on $L^2(\mathbb{R}^{n_{el}\times 3})$ with $\mathcal{D}(\mathcal{H}) = H^2(\mathbb{R}^{n_{el}\times 3})$.*

Proof. We will show that the operators $\mathcal{A} = -\frac{1}{2}\Delta$ and $\mathcal{B} = V(r)$ satisfy the conditions of the Kato-Rellich Theorem 2.17.

We will use the Fourier transform $\mathcal{F}$ on $\mathbb{R}^{n_{\text{el}}\times 3}$ and assume that the reader is familiar with its basic properties. First note that the Fourier transform of the Laplace operator is simply given by multiplication with $|\omega|^2$. Using this fact one can easily show that Δ is self-adjoint with domain given by H^2.

Hardy's inequality (Proposition 2.18) implies that there exists a constant C (depending on $(\mathbf{R}, \mathbf{Z})$) with

$$\|\mathcal{B}\Psi\| \leqslant C \cdot \|\nabla\Psi\|. \tag{22}$$

Therefore, by Plancherel's theorem and the fact that the Fourier transform of the gradient is a multiplication operator we have that

$$\|\mathcal{B}\Psi\|^2 \leqslant C \cdot \int_{\mathbb{R}^{n_{\text{el}}\times 3}} |\omega|^2 |\mathcal{F}\Psi(\omega)|^2 d\omega.$$

Let $\varepsilon > 0$ arbitrary and $D_\varepsilon > 0$ so that

$$|\omega|^2 \leqslant \varepsilon|\omega|^4 + D_\varepsilon \quad \forall \omega \in \mathbb{R}^{n_{\text{el}}\times 3}.$$

Then, we have that

$$\|\mathcal{B}\Psi\|^2 \leqslant C \cdot \varepsilon \int_{\mathbb{R}^{n_{\text{el}}\times 3}} |\omega|^4 |\mathcal{F}\Psi(\omega)|^2 d\omega + C \cdot D_\varepsilon \int_{\mathbb{R}^{n_{\text{el}}\times 3}} |\mathcal{F}\Psi(\omega)|^2 d\omega.$$

Using Plancherel's theorem again with the fact that the Laplace operator turns into multiplication with $|\omega|^2$ in Fourier space yields that

$$\|\mathcal{B}\Psi\|^2 \leqslant C \cdot \varepsilon \|\Delta\Psi\|^2 + C \cdot D_\varepsilon \|\Psi\|^2.$$

By choosing ε sufficiently small we can now satisfy the conditions of Theorem 2.17, which proves the desired statement. □

Having now established the self-adjointness of $\mathcal{H}$ we can now speak of the associated Schrödinger equation and its spectrum in a rigorous way.

First we establish that the spectrum of $\mathcal{H}$ is bounded from below.

Theorem 2.20. *It holds that*

$$\inf \sigma(\mathcal{H}) > -\infty.$$

Proof. The proof goes by establishing a Gårding inequality for $\mathcal{H}$. First note that by Lemma 2.15 it suffices to show that

$$\inf\{\langle \mathcal{H}\Psi, \Psi\rangle : \|\Psi\| = 1\} > -\infty.$$

Next we observe that the Hardy inequality Proposition 2.18 and the Cauchy-Schwartz Inequality imply that there is a constant $C \in (0, \infty)$ with

$$|\langle V\Psi, \Psi\rangle| \leqslant C\|\nabla\Psi\|\|\Psi\|. \tag{23}$$

This, together with the fact that $\langle \Delta\Psi, \Psi\rangle = \|\nabla\Psi\|^2$ implies that

$$\begin{aligned}
&2\langle(\mathcal{H}\Psi, \Psi) + \left(2C^2 + \frac{1}{2}\right)\langle\Psi, \Psi\rangle \\
&\geqslant \|\nabla\Psi\|^2 - 2C\|\Psi\|\|\nabla\Psi\| + \left(2C^2 + \frac{1}{2}\right)\|\Psi\|^2 \\
&= \frac{1}{2}\left(\|\nabla\Psi\|^2 + \|\Psi\|^2\right) + \frac{1}{2}\left(\|\nabla\Psi\|^2 - 4C\|\Psi\|\|\nabla\Psi\| + 4C^2\|\Psi\|^2\right) \\
&= \frac{1}{2}\left(\|\nabla\Psi\|^2 + \|\Psi\|^2\right) + \frac{1}{2}\left(\|\nabla\Psi\| - 2C\|\Psi\|\right)^2 \\
&\geqslant \frac{1}{2}(\|\Psi\| + \|\nabla\Psi\|^2) \geqslant 0.
\end{aligned}$$

In particular, it follows that

$$\langle \mathcal{H}\Psi, \Psi\rangle \geqslant -C^2 - \frac{1}{4} > -\infty$$

for every Ψ with $\|\Psi\| = 1$, which is what we wanted to show □

Remark 2.21. A careful examination of our proof of Theorem 2.20 yields a stronger statement. There exist constants $c_1, c_2 \in (0, \infty)$ such that

$$\langle \mathcal{H}\Psi, \Psi \rangle \geqslant c_1 \|\Psi\|_{H^1}^2 - c_2 \|\Psi\|^2 \quad \forall \Psi \in H^1. \tag{24}$$

Such an inequality is commonly referred to as *Gårding inequality* and shows that a shifted version of $\mathcal{H}$ is coercive on H^1. Furthermore, using (23), it is easy to see that there is another constant $c_3 \in (0, \infty)$ with

$$|\langle \mathcal{H}\Psi, \Phi \rangle| \leqslant c_3 \|\Psi\|_{H^1} \|\Phi\|_{H^1} \quad \forall \Psi, \Phi \in H^1. \tag{25}$$

This shows that the Schrödinger equation can be formulated in a weak form on H^1 and studied using familiar tools from the theory of elliptic operators. This route is taken in Yserentant (2010).

Since $\mathcal{H}$ is an unbounded operator, its spectrum must be unbounded. In order to find out more about its precise structure, we follow Yserentant (2010, Section 5) and define for $T \in (0, \infty)$ the quantity

$$\Sigma(T) := \tag{26}$$
$$\inf \left\{ \langle \mathcal{H}\Psi, \Psi \rangle : \ \|\Psi\| = 1 \ \wedge \ \forall \mathbf{r} \in \mathbb{R}^{n_{\mathrm{el}} \times 3} \setminus \{\mathbf{r} \in \mathbb{R}^{n_{\mathrm{el}} \times 3} : \ |\mathbf{r}| \leqslant T\} : \ \Psi(\mathbf{r}) = 0 \right\}$$

Lemma 2.22. *For all $T > 0$ it holds that $\Sigma(T) \leqslant 0$.*

Proof. Observe that the Hardy inequality (Proposition 2.18), the Cauchy-Schwartz Inequality, and the fact that $\langle \Delta\Psi, \Psi \rangle = \|\nabla\Psi\|^2$ imply that there is a constant $C \in (0, \infty)$ with

$$\langle \mathcal{H}\Psi, \Psi \rangle \leqslant C \cdot \left(\|\nabla\Psi\|^2 + \|\nabla\Phi\| \|\Phi\| \right). \tag{27}$$

Let $\Phi \in \mathcal{S}$ be a function vanishing on the unit ball of $\mathbb{R}^{n_{\mathrm{el}} \times 3}$ with $\|\Phi\| = 1$. Then for $t \in (0, \infty)$ the function $\Psi_t := \frac{1}{t^{\frac{n_{\mathrm{el}} \times 3}{2}}} \Psi\left(\frac{\cdot}{t}\right)$ has unit norm and vanishes on a ball of radius t. It is not hard to see (for example by a scaling argument) that $\lim_{t \to \infty} \|\nabla\Phi_t\| = 0$ and therefore, since for $t > T$ the function Ψ_t can be considered in the infimum of (26), the estimate (27) implies that $\Sigma(T) \leqslant 0$. □

Definition 2.23. Define the *ionization threshold*

$$\Sigma := \lim_{T \to \infty} \Sigma(T) \leqslant 0. \tag{28}$$

Remark 2.24. By Lemma 2.22, the limit in (28) exists and is bounded above by 0.

We now come to the main result about the structure of $\sigma(\mathcal{H})$.

Theorem 2.25. *Assume that*

$$\inf \sigma(\mathcal{H}) < \Sigma. \tag{29}$$

Then it holds that

$$\inf \sigma_{ess}(\mathcal{H}) = \Sigma.$$

In particular, the essential spectrum is nonempty and it holds that

$$\sigma(\mathcal{H}) \cap [\inf \sigma(\mathcal{H}), \Sigma) \subset \sigma_{disc}(\mathcal{H}).$$

Proof. We follow the arguments in Yserentant (2010, Section 5, Theorem 5.6).

1. First we show that for all $T > 0$ and all $\lambda \in \sigma_{\text{ess}}(\mathcal{H})$ it holds that

$$\lambda \geqslant \Sigma(R). \tag{30}$$

To this end, for any $T > 0$ let Ω_1, Ω_2 be functions with $\Omega_1^2 + \Omega_2^2 = 1$ and $\text{supp}(\Omega_1) \subset \{\mathbf{r} : \ |\mathbf{r}| \geqslant T\}$. Since this implies that $|\nabla\Psi|^2 = |\nabla(\Omega_1\Psi)|^2 + |\nabla(\Omega_2\Psi)|^2 - \left(|\nabla\Omega_1|^2 + |\nabla\Omega_2|^2\right)\Psi^2$, we get that

$$\langle \mathcal{H}\Psi, \Psi\rangle = \langle \mathcal{H}\Omega_1\Psi, \Omega_1\Psi\rangle + \langle \mathcal{H}\Omega_2\Psi, \Omega_2\Psi\rangle - \int \left(|\nabla\Omega_1|^2 + |\nabla\Omega_2|^2\right)\Psi^2.$$

Due to (24) there is $\mu > 0$ with $\langle \mathcal{H}\Omega_2\Psi, \Omega_2\Psi\rangle \geqslant -\mu\|\Omega_2\Psi\|^2$ and hence we get that

$$\langle \mathcal{H}\Psi, \Psi\rangle \geqslant \langle \mathcal{H}\Omega_1\Psi, \Omega_1\Psi\rangle - \int \left(\mu\Omega_2^2 + |\nabla\Omega_1|^2 + |\nabla\Omega_2|^2\right)\Psi^2.$$

Since $\Omega_1\Psi$ is supported in a ball of radius T we furthermore get that $\langle \mathcal{H}\Omega_1\Psi, \Omega_1\Psi\rangle \geqslant \Sigma(T)\|\Omega_1\Psi\|^2$ which implies that

$$\langle \mathcal{H}\Psi, \Psi\rangle \geqslant \Sigma(T)\|\Omega_1\Psi\|^2 - \int \left(\mu\Omega_2^2 + |\nabla\Omega_1|^2 + |\nabla\Omega_2|^2\right)\Psi^2.$$

Since $\|\Omega_1\Psi\|^2 = \|\Psi\|^2 - \|\Omega_2\Psi\|^2$ it follows that

$$\langle \mathcal{H}\Psi, \Psi\rangle \geqslant \Sigma(T)\|\Psi\|^2 - \int \left((\Sigma(T) + \mu)\Omega_2^2 + |\nabla\Omega_1|^2 + |\nabla\Omega_2|^2\right)\Psi^2.$$

Define $\Xi := (\Sigma(T) + \mu)\Omega_2^2 + |\nabla\Omega_1|^2 + |\nabla\Omega_2|^2$. Then Ξ is compactly supported and for all $\Psi \in \mathcal{D}(\mathcal{H})$ it holds that

$$\langle \mathcal{H}\Psi, \Psi\rangle + \langle \Xi\Psi, \Psi\rangle \geqslant \Sigma(T)\|\Psi\|^2. \tag{31}$$

By the assumption that $\lambda \in \sigma_{\text{ess}}(\mathcal{H})$ and Theorem 2.6 there is a sequence Ψ^n with $\|\Psi^n\| = 1$, $\|(\mathcal{H} - \lambda)\Psi^n\| \leqslant \frac{1}{n}$ and Ψ^n converges weakly to 0.

First of all this implies that there is a constant C with $\|\Psi\|_{H^1} \leqslant C$ for all n. To see this we observe that the condition $\|(\mathcal{H} - \lambda)\Psi^n\| \leqslant \frac{1}{n}$ implies that there is a constant $C > 0$ with $\langle \mathcal{H}\Psi^n, \Psi^n\rangle \leqslant C$ for all n. By (24) this implies that the norms $\|\Psi^n\|_{H^1}$ must be uniformly bounded. Take the function Ξ from (31) and consider $D := \mathrm{supp}(\Xi)$, which is a bounded subset of $\mathbb{R}^{n_{\mathrm{el}}\times 3}$. Since $H^1(D)$ is compactly embedded into $L^2(D)$ by the Rellich-Kondrachov Theorem, the uniform H^1 boundedness of the Ψ^n's implies that there is a subsequence of the Ψ^n's that converges *strongly* in $L^2(D)$. Since we also know that the Ψ^n's converge weakly to 0 this implies (possibly after passing to a subsequence) that $\lim_{n\to\infty} \|\Psi^n\|_{L^2(D)} = 0$. This implies that $\lim_{n\to\infty}\langle \Xi\Psi^n, \Psi^n\rangle = 0$. This, together with the fact that $\lim_{n\to\infty}\langle \mathcal{H}\Psi^n, \Psi^n\rangle = \lambda$, the fact that $\|\Psi^n\| = 1$, as well as (31) implies that $\lambda \geqslant \Sigma(T)$. This proves (30).

2. In the other direction we show that any S with $(-\infty, S] \cap \sigma_{\mathrm{ess}}(\mathcal{H}) = \varnothing$ satisfies that for every $\varepsilon > 0$ there exists T_ε with

$$S - \varepsilon \leqslant \Sigma(T_\varepsilon). \tag{32}$$

In particular, this implies that $\inf \sigma_{\mathrm{ess}}(\mathcal{H}) \leqslant \Sigma$.
To prove (32) we assume that S satisfies $(-\infty, S]\cap\sigma_{\mathrm{ess}}(\mathcal{H}) = \varnothing$. This means that the set $(-\infty, S] \cap \sigma(\mathcal{H})$ only contains isolated eigenvalues of finite multiplicity. By Theorem 2.20 there can only be finitely many such Eigenvalues. Let $\Psi^1, \ldots, \Psi^N$ be an orthonormal basis of the union of the finitely many corresponding Eigenspaces of finite dimension and let $\lambda_1, \ldots, \lambda_N$ denote the corresponding Eigenvalues. If $N = 0$, Lemma 2.15 implies that $S \leqslant \inf\{\langle \mathcal{H}\Psi, \Psi\rangle : \ \Psi \in \mathcal{D}(\mathcal{H})\} \leqslant \Sigma$ which implies (32). If $N > 0$ we define by $\mathcal{P}$ the orthogonal projection onto the span of the Ψ^n's, i.e. $\mathcal{P}\Psi = \sum_{i=1}^N \langle \Psi, \Psi^i\rangle\Psi^i$. The spectral Theorem 2.11 together with the fact that for every Ψ the spectral measure $\mu^{\mathcal{H}}_{\Psi-\mathcal{P}\Psi}$ is supported on $[S, \infty)$ implies that

$$\langle \mathcal{H}(\Psi - \mathcal{P}\Psi), \Psi - \mathcal{P}\Psi\rangle \geqslant S\|\Psi - \mathcal{P}\Psi\|^2$$

for all $\Psi \in \mathcal{D}(\mathcal{H})$. Furthermore, the fact that the span of the Ψ^n's is $\mathcal{H}$-invariant yields that

$$\langle \mathcal{H}\Psi, \Psi\rangle = \langle \mathcal{H}(\Psi - \mathcal{P}\Psi), \Psi - \mathcal{P}\Psi\rangle + \sum_{i=1}^N \lambda_i \langle \Psi, \Psi^i\rangle^2$$

Taken together the previous two inequalities yield that

$$\langle \mathcal{H}\Psi, \Psi\rangle \geqslant S\|\Psi - \mathcal{P}\Psi\|^2 + \sum_{i=1}^N \lambda_i \langle \Psi, \Psi^i\rangle^2 = S\|\Psi\|^2 - \sum_{i=1}^N (S - \lambda_i)\langle \Psi, \Psi^i\rangle^2$$

Now let Ω_T be the characteristic function of the set $\{\mathbf{r} \in \mathbb{R}^{n_{\mathrm{el}}\times 3} : \ |\mathbf{r}| \geqslant T\}$ and $\Psi \in \mathcal{D}(\mathcal{H})$ with $\|\Psi\| = 1$ and supported on $\{\mathbf{r} \in \mathbb{R}^{n_{\mathrm{el}}\times 3} : \ |\mathbf{r}| \geqslant T\}$. Then

$\Psi = \Omega_T \Psi$ and taking the infimum of the previous inequality yields

$$\Sigma(T) \geqslant S - \sum_{i=1}^{N} (S - \lambda_i) \langle \Psi, \Omega_T \Psi^i \rangle^2$$

Since $\lim_{T\to\infty} \|\Omega_T \Psi^i\| = 0$, $(S - \lambda_i) \geqslant 0$, and $|\langle \Psi, \Omega_T \Psi^i \rangle| \leqslant \|\Omega_T \Psi^i\|$ for all $i = 1, \ldots, N$ it follows that for every $\varepsilon > 0$ there is T_ε with $\sum_{i=1}^{N} (S - \lambda_i) \langle \Psi, \Omega_T \Psi^i \rangle^2 \leqslant \varepsilon$. For this T_ε it then holds that

$$\Sigma(T_\varepsilon) \geqslant S - \varepsilon,$$

as claimed. □

Remark 2.26. We comment on the assumption (29). Looking at the definition of the ionization threshold we notice that Σ represents precisely the infimum of energies that can be assumed by states whose electron positions are arbitrarily far removed from the nuclei. This is called ionization. Assumption (29) posits that it is energetically advantageous for electrons to stay bounded to the nuclei, which is certainly a natural thing to assume.

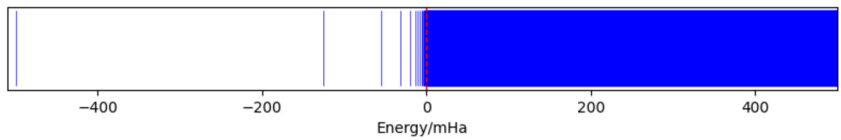

FIGURE 4 Spectrum of a Hydrogen Atom. Negative energies correspond to isolated Eigenvalues which cluster around the ionization threshold 0. The essential spectrum is made up by the positive reals.

Fig. 4 plots the spectrum of the hydrogen atom. Theorem 2.25 shows that for every molecule (that is capable of binding its electrons to the nuclei) the spectrum qualitatively looks the same.

Remark 2.27. There is much more to say about the spectral properties of $\mathcal{H}$. For example, for $\lambda \in \sigma_{\text{disc}}(\mathcal{H})$ it holds that $\|e^{\sqrt{2(\Sigma-\lambda)}|\cdot|}\Psi(\cdot)\|| < \infty$ for any corresponding Eigenfunction. This means that any Eigenfunction corresponding to an Eigenvalue below the ionization threshold decays exponentially. A proof can be found in Yserentant (2010, Theorem 5.17). These estimates are however not sharp. Since the potential does not decay equally fast in every direction, one can get improved anisotropic decay estimates that precisely characterize the decay of Eigenfunctions (Agmon, 2014).

Furthermore, one can show that the discrete spectrum is infinite (Hunziker and Sigal, 2000) and clusters at the ionization threshold.

Corollary 2.28. *Under the assumptions of Theorem 2.25 the quantity* $\mathcal{E}_{\mathbf{Z}}(\mathbf{R}) := \inf \sigma(\mathcal{H})$ *is called the* ground state energy. *It is an isolated Eigenvalue of* $\mathcal{H}$ *of finite multiplicity and corresponding Eigenvectors are called* ground states.

Furthermore it holds that

$$\mathcal{E}_{\mathbf{Z}}(\mathbf{R}) = \min_{\Psi \in H^1} \frac{\langle \mathcal{H}\Psi, \Psi \rangle}{\langle \Psi, \Psi \rangle} \tag{33}$$

and the minimizers are precisely the ground states.

Proof. This follows from Lemma 2.15 and by noting that $\mathcal{W}(\mathcal{H}) = H^1(\mathbb{R}^{n_{\mathrm{el}} \times 3})$. The fact that the infimum is a minimum that is attained by the ground states follows from the fact that the infimum of $\sigma(\mathcal{H})$ is an isolated Eigenvalue with finite multiplicity by Theorem 2.25. □

Our main interest will thus be to efficiently minimize the Raleigh quotient (33).

2.3 Spin and the Pauli exclusion principle

We will again write $\mathcal{H} = \mathcal{H}^{\mathrm{BO}}_{(\mathbf{R},\mathbf{Z})}$ for notational convenience.

There is another serious complication that we have not addressed thus far. The electrons $r_1, \dots, r_{n_{\mathrm{el}}}$ are *indistinguishable*. This means that for every permutation π of the set $\{1, \dots, n_{\mathrm{el}}\}$ the state $\Psi(\pi \circ \mathbf{r}) := \Psi(r_{\pi(1)}, \dots, r_{\pi(n_{\mathrm{el}})})$ must be identical to the state Ψ. In other words, there must be $\tau(\pi) \in \{z \in \mathbb{C} : |z| = 1\}$ with $\Psi(\pi \circ \mathbf{r}) = \tau(\pi) \cdot \Psi(\mathbf{r})$. For two permutations π, η it clearly holds that $\tau(\pi \circ \eta) = \tau(\pi) \cdot \tau(\eta)$. Furthermore, if π is a transposition (e.g., a permutation that exchanges two elements and leaves all other elements fixed) it holds that $\pi \circ \pi = \mathrm{id}$, and therefore it holds that $\tau(\pi)^2 = \tau(\mathrm{id}) = 1$ which implies that $\tau(\pi) \in \{\pm 1\}$ if π is a transposition. Since for any two transpositions π, π' there exists a permutation η with $\pi = \eta^{-1} \circ \pi' \circ \eta$ it follows that the phase factor must be equal for all transpositions. Since all permutations arise as products of transpositions, this implies that either $\tau(\pi) = 1$ for all permutations or $\tau(\pi) = \mathrm{sign}(\pi)$ for all permutations. In the first case the particles are called *bosonic*, in the latter case *fermionic*. Electrons are fermionic which thus implies that the electronic wave function Ψ must be antisymmetric.

Furthermore, each electron has an additional property called *spin* which can assume values in $\left\{-\frac{1}{2}, \frac{1}{2}\right\}$. Therefore the electronic wave function really depends on *spin coordinates*

$$\big((r_1, s_1), \dots, (r_{n_{\mathrm{el}}}, s_{n_{\mathrm{el}}})\big) \in \left(\mathbb{R}^3 \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)^{n_{\mathrm{el}}}.$$

Definition 2.29. The *Pauli exclusion principle* states that any admissible electronic wave function $\Psi : \big((r_1, s_1), \dots, (r_{n_{\mathrm{el}}}, s_{n_{\mathrm{el}}})\big) \in \left(\mathbb{R}^3 \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)^{n_{\mathrm{el}}} \to \mathbb{C}$ must be antisymmetric with respect to permutations of the spin coordinates. In other words, for every permutation π and every admissible wavefunction Ψ it

must hold that

$$\Psi\left((r_{\pi(1)}, s_{\pi(1)}), \ldots, (r_{\pi(n_{\mathrm{el}})}, s_{\pi(n_{\mathrm{el}})})\right) = \mathrm{sign}(\pi) \cdot \Psi\left((r_1, s_1), \ldots, (r_{n_{\mathrm{el}}}, s_{n_{\mathrm{el}}})\right). \tag{34}$$

Writing $\mathbf{s} = (s_1, \ldots, s_{n_{\mathrm{el}}})$ we succinctly write

$$\Psi(\pi \circ (\mathbf{r}, \mathbf{s})) = \mathrm{sign}(\pi) \cdot \Psi(\mathbf{r}, \mathbf{s}).$$

For each $\mathbf{s}$ and each $\Psi : \left(\mathbb{R}^3 \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)^{n_{\mathrm{el}}} \to \mathbb{C}$ we denote by $\Psi_{\mathbf{s}} : \mathbb{R}^{n_{\mathrm{el}} \times 3} \to \mathbb{C}$ the function defined by $\Psi_{\mathbf{s}}(\mathbf{r}) = \Psi(\mathbf{r}, \mathbf{s})$. The correct quantum Hilbert space for electronic systems is thus given by

$$H^{\mathrm{fermi}} := \left\{\Psi : \left(\mathbb{R}^3 \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)^{n_{\mathrm{el}}} \to \mathbb{C} : \ \Psi \text{ satisfies (34) and } \forall \mathbf{s} \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}^{n_{\mathrm{el}}} : \ \Psi_{\mathbf{s}} \in L^2(\mathbb{R}^{n_{\mathrm{el}} \times 3})\right\}$$

with inner product

$$\langle \Psi, \Phi \rangle_{H^{\mathrm{fermi}}} := \sum_{\mathbf{s} \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}^{n_{\mathrm{el}}}} \langle \Psi_s, \Phi_s \rangle_{L^2(\mathbb{R}^{n_{\mathrm{el}} \times 3})}. \tag{35}$$

The electronic Hamiltonian on H^{fermi} is defined componentwise as

$$(\mathcal{H}^{\mathrm{fermi}} \Psi)_s = \mathcal{H} \Psi_s, \tag{36}$$

for $\Psi \in H^{\mathrm{fermi}}$.

We still need to establish self-adjointness. This is done in the next result.

Theorem 2.30. *The operator $\mathcal{H}^{fermi}$ is self-adjoint on H^{fermi} with $\mathcal{D}(\mathcal{H}^{\mathrm{fermi}}) = \left[H^2(\mathbb{R}^{n_{el} \times 3})\right]^{2^{n_{el}}} \cap H^{\mathrm{fermi}}$.*

Proof. Let

$$H^{\mathrm{full}} := \left\{\Psi : \left(\mathbb{R}^3 \times \left\{-\frac{1}{2}, \frac{1}{2}\right\}\right)^{n_{\mathrm{el}}} \to \mathbb{C} : \forall \mathbf{s} \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}^{n_{\mathrm{el}}} : \ \Psi_{\mathbf{s}} \in L^2(\mathbb{R}^{n_{\mathrm{el}} \times 3})\right\}.$$

We first show that $\mathcal{H}^{\mathrm{full}}$, which is simply the extension of $\mathcal{H}^{\mathrm{fermi}}$ to H^{full} is self-adjoint on H^{full}. Not that the operators in this notation do not change, but only their domain of definition.

Self-adjointness of $\mathcal{H}^{\mathrm{full}}$ follows from the fact that H^{full} is a direct sum of $2^{n_{\mathrm{el}}}$ copies of $L^2(\mathbb{R}^{n_{\mathrm{el}}})$ and $\mathcal{H}^{\mathrm{full}}$ is the direct sum of $2^{n_{\mathrm{el}}}$ copies of a self-adjoint

operator (namely the Hamiltonian $\mathcal{H}$ on $\mathbb{R}^{n_{\text{el}}\times 3}$). By Reed and Simon (1978, Theorem XIII.85) a direct sum of self-adjoint operators is self-adjoint and therefore $\mathcal{H}^{\text{full}}$ is self-adjoint on H^{full}.

We want to show that the restriction of $\mathcal{H}^{\text{full}}$ to the antisymmetric subspace H^{fermi} is still self-adjoint. We will make use of the following fact that is proven in Hall (2013, Proposition 9.23): For any Hilbert space J and any symmetric operator $\mathcal{B}$ on J it holds that

$$\mathcal{B} \text{ self-adjoint } \Leftrightarrow \text{range}(\mathcal{B} \pm i) = J. \tag{37}$$

Consider the antisymmetrization operator

$$\mathcal{A}\Psi(\mathbf{r},\mathbf{s}) := \frac{1}{n_{\text{el}}!} \sum_{\pi\in\mathfrak{S}_{n_{\text{el}}}} \text{sign}(\pi)\cdot\Psi(\pi\circ(\mathbf{r},\mathbf{s})), \tag{38}$$

where $\mathfrak{S}_{n_{\text{el}}}$ denotes the group of permutations of n_{el} elements. Since $\mathcal{A}$ is the identity on H^{fermi}, $\mathcal{A}^2 = \mathcal{A}$, and $\mathcal{A}$ is symmetric (and therefore self-adjoint since it is bounded) it follows that $\mathcal{A}$ is the orthogonal projection from H^{full} onto H^{fermi}.

Observe that

$$\mathcal{A}\left(\mathcal{D}(\mathcal{H}^{\text{full}})\right) \subset \mathcal{D}(\mathcal{H}^{\text{full}}) \tag{39}$$

and

$$\forall\Phi\in\mathcal{D}(\mathcal{H}^{\text{full}}):\ \mathcal{A}\mathcal{H}^{\text{full}}\Phi = \mathcal{H}^{\text{full}}\mathcal{A}\Phi. \tag{40}$$

In view of (37) we need to prove that $\text{range}(\mathcal{H}^{\text{fermi}} \pm i) = H^{\text{fermi}}$. To this end let $\Psi \in H^{\text{fermi}}$ be arbitrary. Since $\mathcal{H}^{\text{full}}$ is self-adjoint, Eq. (37) implies that there exists $\Phi \in \mathcal{D}(\mathcal{H}^{\text{full}})$ with $(\mathcal{H}^{\text{full}} \pm i)\Phi = \Psi$. By (39) it holds that $\mathcal{A}\Phi \in \mathcal{D}(\mathcal{H}^{\text{full}})$ and that $(\mathcal{H}^{\text{full}} \pm i)\mathcal{A}\Phi = \mathcal{A}(\mathcal{H}^{\text{full}} \pm i)\Phi = \mathcal{A}\Psi = \Psi$. Thus, $\mathcal{H}^{\text{fermi}} \pm i$ are surjective on H^{fermi} which proves our desired claim. □

Remark 2.31. While the original spectral problem for the operator $\mathcal{H}$ is generally already very high-dimensional and extremely challenging, the Pauli exclusion principle adds another layer of intractability. Even evaluating the inner product according to the formula (35) requires the evaluation of $2^{n_{\text{el}}}$ terms, each of which is given as an integral of a $3\cdot n_{\text{el}}$ dimensional function! This sounds daunting. Fortunately, we will see that the spectral problem of $\mathcal{H}^{\text{fermi}}$ decouples into only n_{el} subproblems analogous to the spectral problem of $\mathcal{H}$ – one for each number of electrons with spin equal to $\frac{1}{2}$ – but with additional antisymmetry constraints.

In order to clarify the connection between functions $\Psi = (\Psi_s)_{s\in\{\pm 1/2\}^{n_{\text{el}}}} \in H^{\text{fermi}}$ and their components Ψ_s we need the following definition.

Definition 2.32. For $\mathbf{s} \in \left\{\pm\frac{1}{2}\right\}^{n_{\text{el}}}$ define $L^2(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}}$ as the subspace of $L^2(\mathbb{R}^{n_{\text{el}}\times 3})$ consisting of all functions Ψ which are antisymmetric with respect to all permutations that leave $\mathbf{s}$ invariant, i.e.,

$$\Psi(\pi \circ \mathbf{r}) = \operatorname{sign}(\pi) \cdot \Psi(\mathbf{r}) \quad \forall \pi \in \mathfrak{S}_{n_{\text{el}}} : \pi \circ \mathbf{s} = \mathbf{s}.$$

Analogously we define the spaces $H^k(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}}$. Finally we denote by $\mathcal{H}_{\mathbf{s}}$ the restriction of $\mathcal{H}$ to $L^2(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}}$ and its ionization threshold $\Sigma_{\mathbf{s}}$ in the same way as (28) but with the infimum (26) taken only over $L^2(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}}$.

Remark 2.33. It is important to keep in mind that the operators $\mathcal{H}_{\mathbf{s}}$ are really independent of $\mathbf{s}$ and all equal to $\mathcal{H}$. The only thing that distinguishes them is their domain, which does depend on $\mathbf{s}$ and is given by $H^2(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}}$ (see Theorem 2.34 below).

The operators $\mathcal{H}_{\mathbf{s}}$ behave similarly to the operators $\mathcal{H}$ in terms of their spectral properties.

Theorem 2.34. 1. *For each* $\mathbf{s} \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}^{n_{el}}$ *the operator* $\mathcal{H}_{\mathbf{s}}$ *is self adjoint with* $\mathcal{D}(\mathcal{H}_{\mathbf{s}}) = H^2(\mathbb{R}^{n_{el}\times 3})_{\mathbf{s}}$.

2. *For every permutation* $\pi \in \mathfrak{S}_{n_{el}}$ *and every* $\mathbf{s} \in \left\{-\frac{1}{2}, \frac{1}{2}\right\}$, *the operators* $\mathcal{H}_{\mathbf{s}}$ *and* $\mathcal{H}_{\pi\circ\mathbf{s}}$ *are unitarily equivalent.*

3. *Assume that*

$$\inf\sigma(\mathcal{H}_{\mathbf{s}}) < \Sigma_{\mathbf{s}}.$$

Then it holds that

$$\inf\sigma_{ess}(\mathcal{H}_{\mathbf{s}}) = \Sigma_{\mathbf{s}}.$$

In particular, the essential spectrum is nonempty and it holds that

$$\sigma(\mathcal{H}_{\mathbf{s}}) \cap [\inf\sigma(\mathcal{H}_{\mathbf{s}}), \Sigma_{\mathbf{s}}) \subset \sigma_{disc}(\mathcal{H}_{\mathbf{s}}).$$

Proof. The fact that the restriction $\mathcal{H}_{\mathbf{s}}$ of the self-adjoint operator $\mathcal{H}$ to $L^2_{\mathbf{s}}$ is self-adjoint can be argued in exactly the same way as the proof of Theorem 2.30. The second point follows from the fact that $\mathcal{H}_{\mathbf{s}}$ and $\mathcal{H}_{\pi\circ\mathbf{s}}$ are unitarily equivalent via the operator that permutes the electronic coordinates $\mathbf{r}$ via π. The proof of the third point is identical to the proof of Theorem 2.25. □

We now want to relate the spectral properties of the operators $\mathcal{H}_{\mathbf{s}}$ to properties of the full Hamiltonian $\mathcal{H}^{\text{fermi}}$.

Lemma 2.35. *A function* $\Phi \in L^2(\mathbb{R}^{n_{el}\times 3})$ *is a component* $\Psi_{\mathbf{s}}$ *of a function* $\Psi = (\Psi_{\mathbf{s}})_{\mathbf{s}\in\{\pm 1/2\}^{n_{el}}} \in H^{\text{fermi}}$ *if and only if* $\Phi \in L^2(\mathbb{R}^{n_{el}\times 3})_{\mathbf{s}}$.

Proof. First consider $\Psi = (\Psi_{\mathbf{s}})_{\mathbf{s}\in\{\pm 1/2\}^{n_{\mathrm{el}}}} \in H^{\mathrm{fermi}}$ and fix $\mathbf{s}$. Let π be a permutation with $\pi \circ \mathbf{s} = \mathbf{s}$.

Then

$$\Psi_{\mathbf{s}}(\pi \circ \mathbf{r}) = \operatorname{sign}(\pi) \cdot \Psi_{\pi\circ\mathbf{s}}(\pi \circ \mathbf{s}) = \operatorname{sign}(\pi) \cdot \Psi_{\mathbf{s}}(\pi \circ \mathbf{r}).$$

Therefore, $\Psi_{\mathbf{s}} \in L^2(\mathbb{R}^{n_{\mathrm{el}}\times 3})_{\mathbf{s}}$ which proves the "only if" part.

For the other direction we assume that for a fixed $\mathbf{s}$ it holds that $\Phi \in L^2(\mathbb{R}^{n_{\mathrm{el}}\times 3})_{\mathbf{s}}$. We need to find $\Psi = (\Psi_{\mathbf{t}})_{\mathbf{t}\in\{\pm 1/2\}^{n_{\mathrm{el}}}} \in H^{\mathrm{fermi}}$ with $\Psi_{\mathbf{s}} = \Phi$. Such a Ψ can be explicitly constructed as

$$\Psi_{\mathbf{t}}(\mathbf{r}) := \frac{\sum_{\pi\in\mathfrak{S}_{n_{\mathrm{el}}}} \operatorname{sign}(\pi)\Phi(\pi \circ \mathbf{r}) \cdot \delta_{\pi\circ\mathbf{t},s}}{\sum_{\pi\in\mathfrak{S}_{n_{\mathrm{el}}}} \delta_{\pi\circ\mathbf{t},\mathbf{t}}}. \tag{41}$$

□

Theorem 2.36. *It holds that*

$$\sigma(\mathcal{H}^{\mathrm{fermi}}) = \bigcup_{\mathbf{s}\in\left\{-\frac{1}{2},\frac{1}{2}\right\}^{n_{el}}} \sigma(\mathcal{H}_{\mathbf{s}}). \tag{42}$$

Proof. Let $\lambda \in \sigma(\mathcal{H}^{\mathrm{fermi}})$, then by Lemma 2.4 for every $n \in \mathbb{N}$ there is $\Psi^n \in H^{\mathrm{fermi}}$ with $\|(\mathcal{H}^{\mathrm{fermi}} - \lambda)\Psi^n\| \leqslant \frac{1}{n}\|\Psi^n\|$.

By definition this implies that

$$\sum_{\mathbf{s}\in\{\pm 1/2\}^{n_{\mathrm{el}}}} \|(\mathcal{H}_{\mathbf{s}} - \lambda)\Psi^n_{\mathbf{s}}\|^2 \leqslant \frac{1}{n^2} \sum_{\mathbf{s}\in\{\pm 1/2\}^{n_{\mathrm{el}}}} \|\Psi_{\mathbf{s}}\|^2$$

and therefore, for each $n \in \mathbb{N}$ there must exist $\mathbf{s}_n$ such that

$$\|(\mathcal{H}_{\mathbf{s}_n} - \lambda)\Psi^n_{\mathbf{s}_n}\| \leqslant \frac{1}{n}\|\Psi^n_{\mathbf{s}_n}\|.$$

Since there are only finitely many values that $\mathbf{s}_n$ can assume, we conclude that there exists $\mathbf{s}$ such that for infinitely many $n \in \mathbb{N}$ there is $\Psi^n_{\mathbf{s}} \in L^2(\mathbb{R}^{n_{\mathrm{el}}\times 3})_{\mathbf{s}}$ with

$$\|(\mathcal{H}_{\mathbf{s}} - \lambda)\Psi^n_{\mathbf{s}}\| \leqslant \frac{1}{n}\|\Psi^n_{\mathbf{s}}\|.$$

By Lemma 2.4 this implies that $\lambda \in \sigma(\mathcal{H}_{\mathbf{s}})$.

For the other direction suppose that $\lambda \in \sigma(\mathcal{H}_{\mathbf{s}})$ and for each $n \in \mathbb{N}$ let $\Phi^n \in L^2(\mathbb{R}^{n_{\mathrm{el}}\times 3})_{\mathbf{s}}$ satisfy that

$$\|(\mathcal{H}_s - \lambda)\Phi^n\| \leqslant \frac{1}{n}\|\Phi^n\|.$$

The existence of Φ^n follows from Lemma 2.4. Now construct $\Psi^n \in H^{\mathrm{fermi}}$ according to (41), which only contains components for spin vectors $\mathbf{t}$ that arise

from permuting the fixed spin vector $\mathbf{s}$ in the sense that $s = \pi \circ \mathbf{t}$. By Item 2 of Theorem 2.34 the operators $\mathcal{H}_\mathbf{t}$ for such spin vectors are unitarily equivalent to $\mathcal{H}_\mathbf{s}$ and therefore it holds that

$$\|(\mathcal{H}_\mathbf{t} - \lambda)\Phi^n(\pi \circ \mathbf{r})\| = \|(\mathcal{H}_\mathbf{s} - \lambda)\Phi^n(\mathbf{r})\| \leqslant \frac{1}{n}\|\Psi^n_\mathbf{s}\| = \frac{1}{n}\|\Psi^n_\mathbf{t}\|$$

This implies that

$$\|(\mathcal{H}^{\text{fermi}} - \lambda)\Psi^n\| \leqslant \frac{1}{n}\|\Psi^n\|,$$

which by Lemma 2.4 implies that $\lambda \in \sigma(\mathcal{H}^{\text{fermi}})$ □

By Theorem 2.36 the spectrum of the full fermionic system $\mathcal{H}$ can be determined from solving the spectral problems of all operators $\mathcal{H}_\mathbf{s}$ for $\mathbf{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^{n_{\text{el}}}$. Since operators $\mathcal{H}_\mathbf{s}$ and $\mathcal{H}_\mathbf{t}$ are unitarily equivalent if $\mathbf{s}$ is a permutation of $\mathbf{t}$, it suffices to only consider spin assignments of the form

$$\mathbf{s}_{n_\uparrow} := \left(\underbrace{\frac{1}{2}, \ldots, \frac{1}{2}}_{n_\uparrow \text{ times}}, \underbrace{-\frac{1}{2}, \ldots, -\frac{1}{2}}_{n_{\text{el}} - n_\uparrow \text{ times}}\right), \quad n_\uparrow \in \{0, \ldots, n_{\text{el}}\}, \tag{43}$$

where $n_\uparrow$ corresponds to the number of electrons with positive spin. Since the spectral problem does not change if positive spin electrons are changed into negative spin electrons and vice versa, it suffices to solve the spectral problems of the operators $\mathcal{H}_{\mathbf{s}_{n_\uparrow}}$ for $n_\uparrow \in \left\{0, \ldots, \lfloor \frac{n_{\text{el}}}{2} \rfloor\right\}$ denoting the number of spin-up electrons.

By Theorem 2.34 the lowest possible energy is given as an isolated eigenvalue (the ground state energy) and corresponding eigenvectors are the ground states. By Lemma 2.15, the ground states and the ground state energies (and thus the PES) can be determined as solutions of the minimization problems

$$\mathcal{E}^{n_\uparrow}_\mathbf{Z}(\mathbf{R}) = \min_{\Psi \in H^1_{\mathbf{s}_{n_\uparrow}}} \frac{\langle \mathcal{H}\Psi, \Psi\rangle}{\langle \Psi, \Psi\rangle}, \quad n_\uparrow \in \left\{0, \ldots, \left\lfloor \frac{n_{\text{el}}}{2} \right\rfloor\right\}. \tag{44}$$

Finally we note that, although the space $H^1_{\mathbf{s}_{n_\uparrow}}$ is a complex Hilbert space, we can without loss of generality minimize (44) over the space of real-valued antisymmetric H^1 functions. This is because the Hamiltonian acts separately on the real and imaginary part which implies that the real (or imaginary) part of an Eigenvector is still an Eigenvector.

Our goal is thus to

Find efficient and accurate algorithms for solving the minimization problems (44).

3 Introduction to variational Monte Carlo (VMC)

Given the problems (44) we would like to devise numerical algorithms that provide an accurate solution. The most straightforward idea to achieve this is to start with a parametrized class of functions

$$\mathcal{F} := \left\{ \Psi_\theta : \theta \in \mathbb{R}^{N_{\text{param}}} \right\} \subset H^1(\mathbb{R}^{n_{\text{el}} \times 3})_{\mathbf{s}_{n_\uparrow}} \tag{45}$$

and (try to) solve the restricted minimization problem

$$\min_{\theta \in \mathbb{R}^{N_{\text{param}}}} \frac{\langle \mathcal{H}\Psi_\theta, \Psi_\theta \rangle}{\langle \Psi_\theta, \Psi_\theta \rangle}, \quad n_\uparrow \in \left\{ 0, \dots, \left\lfloor \frac{n_{\text{el}}}{2} \right\rfloor \right\} \tag{46}$$

in the parameter θ.

This approach has several obvious benefits:

1. It is conceptually simple
2. It is *variational* in the sense that for any approximate solution Ψ_{θ^*} of (46) (or rather any $\Psi_{\theta^*} \in H^1(\mathbb{R}^{n_{\text{el}} \times 3})_{\mathbf{s}_{n_\uparrow}}$) we always have the upper bound

$$\mathcal{E}^{n_\uparrow}_{\mathbf{Z}}(\mathbf{R}) \leqslant \frac{\langle \mathcal{H}\Psi_{\theta^*}, \Psi_{\theta^*} \rangle}{\langle \Psi_{\theta^*}, \Psi_{\theta^*} \rangle}, \quad n_\uparrow \in \left\{ 0, \dots, \left\lfloor \frac{n_{\text{el}}}{2} \right\rfloor \right\}. \tag{47}$$

   This means that we can compare the quality of different algorithms a posteriori – the smaller the computed energy, the better.
3. Since by (24) and (25) the Hamiltonian is bounded and coercive on H^1, one can in principle treat the corresponding Eigenvalue problem within the well-developed mathematical framework of elliptic eigenvalue problems, see Yserentant (2010) for results in this direction.

In reality, it is however extremely challenging to come up with efficient and accurate algorithms. This is at least due to the following issues.

1. The problem is high-dimensional and potentially carries the curse of dimension (it may well be NP-hard (Troyer and Wiese, 2005)).
2. Enforcing the antisymmetry condition of Definition 2.32 is nonstandard and challenging.
3. The minimization problem (47) is nonconvex and therefore algorithms can get stuck on local minima.
4. The electronic ground state is not globally smooth, has cusps near the nuclei and for electron coordinates approaching each other (Kato, 1957) and complicated long-range interactions. It is not at all clear how to choose $\mathcal{F}$ so that such functions can be well approximated with a reasonable number of parameters.

3.1 Slater determinants

We address the question of how to enforce the antisymmetry condition of Definition 2.32.

Slater determinants

Suppose that there are $n_\uparrow$ electrons with spin $\frac{1}{2}$ and $n_{\text{el}} - n_\uparrow$ electrons with spin $-\frac{1}{2}$. Then for $\mathbf{s}_{n_\uparrow}$ as in (43) a permutation π leaves $\mathbf{s}_{n_\uparrow}$ invariant if and only if π can be decomposed into $\pi = \pi_\uparrow \circ \pi_\downarrow$ where $\pi_\uparrow$ is any permutation of the first $n_\uparrow$ electron coordinates and $\pi_\downarrow$ is any permutation of the last $n - n_\uparrow$ electron coordinates.

Now take any set $\phi_\uparrow^i : \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, \ldots, n_\uparrow$ and $\phi_\downarrow^i : \mathbb{R}^3 \to \mathbb{R}$ for $i = 1, \ldots, n_{\text{el}} - n_\uparrow$ of $H^1(\mathbb{R}^3)$ functions. Then it is easy to see that the function

$$\mathfrak{D}\left(\phi_\uparrow^1, \ldots, \phi_\uparrow^{n_\uparrow}; \phi_\downarrow^1, \ldots, \phi_\downarrow^{n_{\text{el}}-n_\uparrow}\right) := \frac{1}{\sqrt{n_\uparrow!}} \begin{vmatrix} \phi_\uparrow^1(r_1) & \ldots & \phi_\uparrow^{n_\uparrow}(r_1) \\ \vdots & \ddots & \vdots \\ \phi_\uparrow^1(r_{n_\uparrow}) & \ldots & \phi_\uparrow^{n_\uparrow}(r_{n_\uparrow}) \end{vmatrix} \cdot \frac{1}{\sqrt{(n_{\text{el}} - n_\uparrow)!}} \begin{vmatrix} \phi_\downarrow^1(r_{n_\uparrow+1}) & \ldots & \phi_\downarrow^{n-n_\uparrow}(r_{n_\uparrow+1}) \\ \vdots & \ddots & \vdots \\ \phi_\downarrow^1(r_{n_{\text{el}}}) & \ldots & \phi_\downarrow^{n-n_{\text{el}}}(r_{n_{\text{el}}}) \end{vmatrix} \tag{48}$$

is in $H^1_{\mathbf{s}_{n_\uparrow}}$. Here we denote by $\begin{vmatrix} * & \ldots & * \\ \vdots & \ddots & \vdots \\ * & \ldots & * \end{vmatrix}$ denotes the determinant of a matrix. We also note that the determinant can be evaluated in cubic complexity using Gaussian elimination.

Functions of the Form (48) are called *Slater determinants*. A suitable approximation set $\mathcal{F}$ can now be constructed by choosing a *basis set* $\mathcal{B} = \chi_1, \ldots, \chi_{n_{\text{basis}}} \in H^1(\mathbb{R}^3)$, writing

$$\phi(\mathcal{B}, \mathbf{b}) := \sum_{j=1}^{n_{\text{basis}}} (\mathbf{b}^i)_j \chi_j .$$

Then one can define

$$\mathcal{F}_\mathcal{B} := \left\{ \mathfrak{D}\left(\phi(\mathcal{B}, \mathbf{b}_\uparrow^1), \ldots, \phi(\mathcal{B}, \mathbf{b}_\uparrow^{n_\uparrow}); \phi(\mathcal{B}, \mathbf{b}_\downarrow^1), \ldots, \phi(\mathcal{B}, \mathbf{b}_\downarrow^{n_{\text{el}}-n_\uparrow})\right) : \mathbf{b}_\uparrow^i,\ \mathbf{b}_\downarrow^i \in \mathbb{R}^{n_{\text{basis}}} \right\}$$

and solve the minimization problem (46). The resulting method is called the *Hartree Fock method* and it corresponds to an antisymmetrized rank-1 approximation. The Hartree Fock method is computationally relatively cheap. However,

it disregards electron correlations (beyond the correlations caused by the antisymmetry constraint) and is therefore of very limited accuracy.

To improve accuracy one can try to use linear combinations of several different slater determinants which corresponds to an antisymmetrized low-rank approximation. The corresponding approximation set is then given as

$$\begin{aligned}&\mathcal{F}_{n_{\text{det}},\mathcal{B}}\\&:=\left\{\sum_{l=1}^{n_{\text{det}}}\mathfrak{D}\left(\phi(\mathcal{B},\mathbf{b}_{\uparrow}^{1,l}),\dots,\phi(\mathcal{B},\mathbf{b}_{\uparrow}^{n_{\uparrow},l});\phi(\mathcal{B},\mathbf{b}_{\downarrow}^{1,l}),\dots,\phi(\mathcal{B},\mathbf{b}_{\downarrow}^{n_{\text{el}}-n_{\uparrow},l})\right):\right.\\&\qquad\left.\mathbf{b}_{\uparrow}^{i,l},\ \mathbf{b}_{\downarrow}^{i,l}\in\mathbb{R}^{n_{\text{basis}}}\right\}.\end{aligned}$$

Remark 3.1. These constructions can be augmented by multiplying each Slater determinant with any function that leaves the antisymmetry property intact. This holds for instance for symmetric functions. Such multiplicative corrections are called *Jastrow factor*. Such Jastrow factors are important to capture the correct cusp behavior of the ground state.

Approximation results for sets of the form $\mathcal{F}_{n_{\text{det}},\mathcal{B}}$ are derived in Yserentant (2010) where it is shown that the H^1 approximation error for the electronic ground state decays as $C_{N_{\text{det}}}\cdot n_{\text{det}}^{-1}$. Unfortunately, a close inspection of the proofs in Yserentant (2010) reveals that the constant $C_{N_{\text{det}}}$ grows exponentially in n_{det} which makes approximation by $\mathcal{F}_{n_{\text{det}},\mathcal{B}}$ intractable. Overall it seems that accurate and tractable approximations of electronic ground states using Slater determinants are limited to quite small systems of up to around 15 electrons.

Generalized slater determinants

A key drawback of Slater determinants is their inability to efficiently model electron correlations. To a certain extent this can be remedied by considering *generalized Slater determinants* of the following form.

For $n\in\mathbb{N}$, $l\in\{1,\dots,n\}$ denote the classes

$$\begin{aligned}&\mathcal{O}_{\uparrow}^{l,n}:=\\&\left\{\phi\in H^1(\mathbb{R}^{n\times 3}):\phi(r_1,\dots,r_n)\text{ is symmetric in the variables }(r_2,\dots,r_l)\right.\\&\qquad\left.\text{and }(r_{l+1},\dots,r_n)\right\}\end{aligned}\tag{49}$$

and

$$\begin{aligned}&\mathcal{O}_{\downarrow}^{l,n}:=\\&\left\{\phi\in H^1(\mathbb{R}^{n\times 3}):\phi(r_1,\dots,r_n)\text{ is symmetric in the variables }(r_{l+2},\dots,r_n)\right.\\&\qquad\left.\text{and }(r_2,\dots,r_l)\right\}.\end{aligned}\tag{50}$$

We also introduce the following notation

$$\{\mathbf{r}^{\uparrow}\} := \{r_1, \dots, r_{n_\uparrow}\}, \quad \{\mathbf{r}^{\downarrow}\} := \{r_{n_\uparrow+1}, \dots, r_{n_{\mathrm{el}}}\}, \quad \{\mathbf{r}^{\uparrow\downarrow}_{\setminus j}\} := \{\mathbf{r}^{\uparrow\downarrow}\} \setminus \{r_j\}. \tag{51}$$

Here, the symbol $\{\cdot\}$ should be understood as describing a *multiset*, meaning that it can contain equal elements several times. For example, if $r_1 = r_2 = 1$, $r_3 = 2$ and $n_\uparrow = 3$ it holds that

$$\{\mathbf{r}^{\uparrow}\} = \{1,1,2\} = \{1,2,1\} = \{2,1,1\} \neq \{1,2\} = \{\mathbf{r}^{\uparrow}_{\setminus 1}\} = \{\mathbf{r}^{\uparrow}_{\setminus 2}\}.$$

Suppose that $\phi^1_\uparrow, \dots, \phi^{n_\uparrow}_\uparrow \in \mathcal{O}^{n_\uparrow, n_{\mathrm{el}}}_\uparrow$ and $\phi^{n_\uparrow+1}_\downarrow, \dots, \phi^{n_{\mathrm{el}}}_\downarrow \in \mathcal{O}^{n_\uparrow, n_{\mathrm{el}}}_\downarrow$.

Then, we can define

$$\Phi^{i,j}_\uparrow(\mathbf{r}) := \phi^i_\uparrow(r_j; \{\mathbf{r}^{\uparrow}_{\setminus j}\}; \{\mathbf{r}^{\downarrow}\}), \quad i,j = 1, \dots, n_\uparrow, \tag{52}$$

and

$$\Phi^{i,j}_\downarrow(\mathbf{r}) := \phi^i_\downarrow(r_j; \{\mathbf{r}^{\uparrow}\}; \{\mathbf{r}^{\downarrow}_{\setminus j}\}), \quad i,j = n_\uparrow + 1, \dots, n_{\mathrm{el}}. \tag{53}$$

Observe that due to the symmetry requirements of the functions $\phi^i_{\uparrow\downarrow}$ the set notation (51) in (52) and (53) is justified since the respective coordinates are independent of their order.

It is again easy to see that the function

$$\begin{aligned}\mathfrak{D}^{\mathrm{fermi}}\left(\varphi^1_\uparrow, \dots, \varphi^{n_\uparrow}_\uparrow; \varphi^1_\downarrow, \dots, \varphi^{n_{\mathrm{el}}-n_\uparrow}_\downarrow\right) := \\ \frac{1}{\sqrt{n_\uparrow!}} \begin{vmatrix} \Phi^{1,1}_\uparrow(\mathbf{r}) & \dots & \Phi^{n_\uparrow,1}_\uparrow(\mathbf{r}) \\ \vdots & \ddots & \vdots \\ \Phi^{1,n_\uparrow}_\uparrow(\mathbf{r}) & \dots & \Phi^{n_\uparrow,n_\uparrow}_\uparrow(\mathbf{r}) \end{vmatrix} \\ \cdot \frac{1}{\sqrt{(n_{\mathrm{el}} - n_\uparrow)!}} \begin{vmatrix} \Phi^{n_\uparrow+1,n_\uparrow+1}_\downarrow(\mathbf{r}) & \dots & \Phi^{n_{\mathrm{el}},n_\uparrow+1}_\downarrow(\mathbf{r}) \\ \vdots & \ddots & \vdots \\ \Phi^{n_\uparrow+1,n_{\mathrm{el}}}_\downarrow(\mathbf{r}) & \dots & \Phi^{n_{\mathrm{el}},n_{\mathrm{el}}}_\downarrow(\mathbf{r}) \end{vmatrix}\end{aligned} \tag{54}$$

is in $H^1_{\mathbf{s}_{n_\uparrow}}$. In Pfau et al. (2020) such determinant functions are called *generalized Slater determinants*.

Remark 3.2. The paper by Pfau et al. (2020, Appendix B) proves a universal approximation result that shows that one can approximate any function $L^\infty(\mathbb{R}^{n_{\mathrm{el}}\times 3})_{\mathbf{s}_{n_\uparrow}}$ arbitrarily well in the L^∞ norm if one allows orbitals Φ^i that are discontinuous (and not in H^1). While this result is reassuring, an application to solving the Schrödinger equation would require such results in the H^1 norm.

It is an open question if such an approximation result from a single generalized Slater determinant holds true, see however Ye et al. (2024) for recent progress in this direction.

Given a pair of mappings

$$\phi_\uparrow : \mathbb{R}^{n_{\text{oparams}}} \to \mathcal{O}_\uparrow^{n_\uparrow, n_{\text{el}}} \quad \text{and} \quad \phi_\downarrow : \mathbb{R}^{n_{\text{oparams}}} \to \mathcal{O}_\downarrow^{n_\uparrow, n_{\text{el}}} \tag{55}$$

we can now again construct an approximation set by using parametrized orbitals $\phi_\uparrow^i$, $\phi_\downarrow^i$, e.g.,

$$\phi_\uparrow^i = \phi_\uparrow(\theta_\uparrow^i) \quad \text{for some } \theta_\uparrow^i \in \mathbb{R}^{n_{\text{oparams}}} \tag{56}$$

and

$$\phi_\downarrow^i = \phi_\downarrow(\theta_\downarrow^i) \quad \text{for some } \theta_\downarrow^i \in \mathbb{R}^{n_{\text{oparams}}}. \tag{57}$$

The corresponding approximation set using N_{det} generalized Slater determinants is then given as

$$\begin{aligned} &\mathcal{F}^{\text{fermi}}_{n_{\text{det}},\Phi_\uparrow,\Phi_\downarrow} := \\ &\left\{ \sum_{l=1}^{n_{\text{det}}} \mathfrak{D}^{\text{fermi}}\left(\phi_\uparrow(\theta_\uparrow^{1,l}), \dots, \phi_\uparrow(\theta_\uparrow^{n_\uparrow,l}); \phi_\downarrow(\theta_\downarrow^{1,l}), \dots, \phi_\downarrow(\theta_\downarrow^{n_{\text{el}}-n_\uparrow,l})\right) : \right. \\ &\qquad \left. \theta_\uparrow^{i,l},\ \theta_\downarrow^{i,l} \in \mathbb{R}^{n_{\text{oparams}}} \right\}. \end{aligned} \tag{58}$$

Remark 3.3. One can further generalize (54) by considering n_{el} spin-up orbitals of the form (56) and n_{el} spin-down orbitals of the form (57) and assembling the $n_{\text{el}} \times n_{\text{el}}$ matrix $(\Xi_{i,j})_{i,j=1}^{n_{\text{el}}}$ with

$$\Xi_{i,j} = \begin{cases} \Phi_\uparrow^{i,j} & j \in \{1, \dots, n_\uparrow\} \\ \Phi_\uparrow^{i,j} & j \in \{n_\uparrow + 1, \dots, n_{\text{el}}\}, \end{cases}$$

and taking its determinant. It is easy to see that this approach yields functions with the correct antisymmetry properties. Furthermore, the construction (54) arises as a special case by letting $\Phi_\uparrow^i = 0$ for all $i = n_\uparrow + 1, \dots, n_{\text{el}}$ and $\Phi_\downarrow^i = 0$ for all $i = 1, \dots, n_\uparrow$. This construction is sometimes referred to as *full Determinant* (Lin et al., 2023).

Furthermore, the construction (58) is typically extended by including a Jastrow factor, i.e., a symmetric function as multiplicative correction term, see Remark 3.1.

Having mappings (55) at hand we can thus construct corresponding parametrized classes of functions given by (58) and try to solve the minimization problem (44) in terms of the parameter $\theta = (\theta_\uparrow^{i,l})_{l=1,\dots,n_{\text{det}},\ i=1,\dots,n_\uparrow} \times$

$(\theta_{\downarrow}^{i,l})_{l=1,\dots,N_{\text{det}},\ i=1,\dots,n_{\text{el}}-n_{\uparrow}} \cong \mathbb{R}^{n_{\text{det}}\times n_{\text{el}}\times n_{\text{oparams}}}$ over this set. In this case we have that $n_{\text{params}} = n_{\text{det}} \times n_{\text{el}} \times n_{\text{oparams}}$ and for $\theta = (\theta_{\uparrow\downarrow}^{i,l})_{i,l,\uparrow\downarrow} \in \mathbb{R}^{n_{\text{params}}}$ it then holds that

$$\Psi_\theta := \sum_{l=1}^{n_{\text{det}}} \mathfrak{D}^{\text{fermi}}\left(\phi_\uparrow(\theta_\uparrow^{1,l}), \dots, \phi_\uparrow(\theta_\uparrow^{n_\uparrow,l}); \phi_\downarrow(\theta_\downarrow^{1,l}), \dots, \phi_\downarrow(\theta_\downarrow^{n_{\text{el}}-n_\uparrow,l})\right) \in H^1(\mathbb{R}^{n_{\text{el}}\times 3})_{\mathbf{s}_{n_\uparrow}}. \tag{59}$$

The key challenge is then to

Find parametrized orbitals (55) such that the sets (58) are maximally expressive.

3.2 Sampling using the Metropolis-Hastings algorithm

For solving the optimization problem (44) we need to be able to efficiently evaluate the $3n_{\text{el}}$-dimensional integrals $\frac{\langle \mathcal{H}\Psi_\theta, \Psi_\theta\rangle}{\langle \Psi_\theta, \Psi_\theta\rangle}$. Therefore, during optimization of the parameters θ and to get a final prediction of the ground-state energy, it is essential to have an efficient method for calculating the high-dimensional integrals. In Variational Monte Carlo, one commonly employs *Markov chain Monte Carlo* (MCMC), utilizing algorithms like *Metropolis-Hastings*. To perform Monte Carlo integration, one needs to rewrite the Rayleigh-Ritz Quotient

$$\frac{\langle \mathcal{H}\Psi_\theta, \Psi_\theta\rangle}{\langle \Psi_\theta, \Psi_\theta\rangle} = \int \frac{|\Psi_\theta(\mathbf{r})|^2}{\langle \Psi_\theta, \Psi_\theta\rangle} \frac{\mathcal{H}\Psi_\theta(\mathbf{r})}{\Psi_\theta(\mathbf{r})} d\mathbf{r} \approx \frac{1}{n_{\text{s}}} \sum_{a=1}^{n_{\text{s}}} \frac{\mathcal{H}\Psi_\theta(\mathbf{r}_a)}{\Psi_\theta(\mathbf{r}_a)} \tag{60}$$

with n_{s} electron samples $\mathbf{r}_a \sim p_\theta(\mathbf{r}) := \frac{|\Psi_\theta(\mathbf{r})|^2}{\langle \Psi_\theta, \Psi_\theta\rangle}$.

The density p_θ we sample from with MCMC gets updated during optimization due to the dependence on θ (cf. Sec. 3.3). To ensure the sampled electron positions follow the correct distribution, we have to perform multiple consecutive steps of MCMC, whereas a single step with the Metropolis-Hastings algorithm can be broken down into the following stages, as outlined in Algorithm 1. Firstly, starting with an initial set of electron positions, either from a previous step or randomly initialized, represented as $\mathbf{r}_n \in \mathbb{R}^{n_{\text{el}}\times 3}$ a proposed state $\mathbf{r}_{\text{p}} \in \mathbb{R}^{n_{\text{el}}\times 3}$ is generated by sampling from a proposal function q. Subsequently, the acceptance probability is computed, and the proposed state is accepted if this probability exceeds a uniformly distributed random value. Thus, given a suitable proposal function q, the algorithm produces a sample $\mathbf{r}_n$ that is distributed according to p_θ in the limit of $n \to \infty$. Intuitively, over multiple steps n the sample tends to move towards high-probability regions because proposals towards higher probability are always accepted, while proposals towards lower-probability regions are often rejected. Because there is some chance to accept proposals towards low-probability regions, the algorithm does not only return

the value with highest probability, but a distribution of samples. Fig. 5a depicts the convergence of the distribution of samples towards the target distribution p_θ for a simple 1D example.

Algorithm 1 Metropolis-Hastings sampling

Require: Probability density $p_\theta(\mathbf{r})$, proposal distribution $q(\mathbf{r}_p|\mathbf{r}_n)$, initial configurations $\mathbf{r}_0$, number of steps N
  for $n = 0$ to $N-1$ **do**
    $\mathbf{r}_p \sim q(\mathbf{r}_p|\mathbf{r}_n)$ ▷ Propose new configuration $\mathbf{r}_p$
    $a = \min\left(1, \frac{p_\theta(\mathbf{r}_p)q(\mathbf{r}_n|\mathbf{r}_p)}{p_\theta(\mathbf{r}_n)q(\mathbf{r}_p|\mathbf{r}_n)}\right)$ ▷ Compute acceptance probability a
    if $a \geqslant$ RandomUniform$(0, 1)$ **then**
      $\mathbf{r}_{n+1} \leftarrow \mathbf{r}_p$ ▷ Accept the proposal with probability a
    else
      $\mathbf{r}_{n+1} \leftarrow \mathbf{r}_n$ ▷ Reject the proposal with probability $1-a$
    end if
  end for
  return $\mathbf{r}_N$ ▷ In the limit of $N \to \infty$, $\mathbf{r}_N$ is distributed according to $p_\theta(\mathbf{r})$

A common choice for the proposal function $q(\mathbf{r}_p|\mathbf{r}_n)$ is a multivariate Gaussian distribution centered around $\mathbf{r}_n$ and variance s^2

$$q(\mathbf{r}_p|\mathbf{r}_n) \propto \exp\left(-\frac{1}{2s^2}|\mathbf{r}_p - \mathbf{r}_n|^2\right) \tag{61}$$

where s is a tuneable parameter known as the stepsize. It is a valid choice, because in principle any configuration can be reached from any other configuration in a single step (since the Gaussian distribution has support on the whole domain) and thus the proposal satisfies ergodicity. Furthermore the fact that the Gaussian is symmetric in $\mathbf{r}_p$ and $\mathbf{r}_n$ allows to omit the q-ratio on the calculation of the acceptance rate since it is always 1. A slight modification to Algorithm 1 is to divide the number of electrons into multiple blocks of electrons and separately accepting / rejecting a subset of electron positions. This potentially reduces the number of steps one needs to perform to reach more decorrelated electron positions (von Glehn et al., 2023).

One important aspect of the Metropolis-Hastings algorithm is that its acceptance criterion only depends on the ratio of probability densities but not on p_θ directly. Therefore, one can consider instead the unnormalized density $|\Psi_\theta|^2$.

An alternative to a Gaussian Proposal distribution is to bias proposals towards increasing probability density to increase the probability of acceptance. This is known as Metropolis Adjusted Langevin Algorithm (MALA) and increases sampling efficiency (Schätzle et al., 2023) at the expensive of higher computational cost to evaluate ∇p_θ.

$$\mathbf{r}_p = \mathbf{r}_n + \tau \nabla_{\mathbf{r}} \log p_\theta(\mathbf{r}_n) + s\,\mathcal{N}(\mathbf{0}, 1) \tag{62}$$

The choice of the stepsize s is important to achieve fast convergence and mixing of the Markov Chain: Choosing a very small stepsize only allows very small changes in $\mathbf{r}$, leading to slow convergence. Choosing a very large stepsize leads to proposed configurations $\mathbf{r}_p$ that are far from the original configuration $\mathbf{r}_n$ and are very often rejected, thus not moving at all. To address this issue, one can set a target acceptance rate of around 50% and automatically adjust s to approximately reach this acceptance rate.

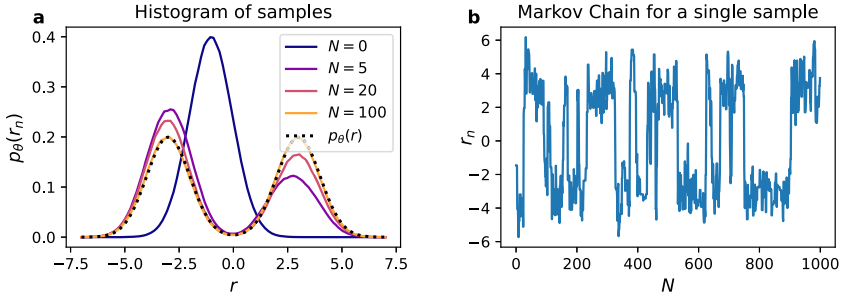

FIGURE 5 1D example of MCMC on a 1D density $p_\theta(\mathbf{r})$ consisting of two Gaussians. The initial configurations $\mathbf{r}_0$ are drawn from a single Gaussian distribution. **a**: Histogram of samples after different number of MCMC steps N. After $N \approx 100$ steps the distribution of $\mathbf{r}_N$ aligns with the target distribution p_θ. **b**: Path of a single sample. Subsequent samples are strongly correlated, depicting two distinct time-scales: A short time-scale corresponding to moves within a density peak and a long time scale corresponding to moves between the two peaks.

In principle any initial distribution can be used for the samples $\mathbf{r}_0$, but choosing an initial distribution that resembles the target distribution is obviously advantageous. Therefore the typical approach is to first take a large number of steps $N_{\text{burn-in}}$ for the samples to converge to the target distribution. Then, to obtain more samples one does not start again from the initial distribution, but uses the latest sample as starting point for the next N_{intermed} steps to obtain a new sample. In practice $N_{\text{burn-in}} \gg N_{\text{intermed}}$, for sampling from a wavefunction of a small molecule $N_{\text{burn-in}} \approx 10^3$, while $N_{\text{intermed}} \approx 10^1$. A disadvantage of this approach is that subsequent samples are not fully independent of each other, but can still be correlated if N_{intermed} is too small. Fig. 5b shows the trajectory of a single sample as a function of Metropolis-Hastings steps, clearly showing correlations between subsequent samples. This issue is particularly pronounced when p_θ has multiple maxima that are separated by regions of low probability, because it takes many steps to transition between these maxima.

Besides using Metropolis-Hastings to sample from $|\Psi_\theta|^2$, one can in principle also design models which allow direct sampling from the probability distribution.

One option are Normalizing Flows, a type of model that maps an easy to sample probability distribution (e.g. a Gaussian) to the target probability distribution p_θ. It has been applied to the Schrödinger Equation, but only in the substantially simplified 1D case (Thiede et al., 2023).

Another option are autoregressive models, which generate a full configuration of electrons one electron at a time, by conditioning the probability distribution on all previously added electrons:

$$p_\theta(\mathbf{r}_1, \dots \mathbf{r}_{n_{\text{el}}}) = p_\theta(\mathbf{r}_1)\, p_\theta(\mathbf{r}_2|\mathbf{r}_1)\, p_\theta(\mathbf{r}_3|\mathbf{r}_1, \mathbf{r}_2)\, \dots\, p_\theta(\mathbf{r}_{n_{\text{el}}}|\mathbf{r}_1, \dots \mathbf{r}_{n_{\text{el}}-1}) \tag{63}$$

Instead of sampling from the probability distribution of the left hand side of Eq. (63), which is $3 \times n_{\text{el}}$-dimensional, all at once, one samples n_{el} times from a 3-dimensional probability distribution (each term on the right hand side of Eq. (63)). This structure is the currently dominant paradigm in large language models, which autoregressively sample one token / word at a time, from a probability distribution which is conditioned on the previously generated tokens (Radford et al., 2021). This approach has also been applied to wavefunctions, but so far only for model Hamiltonians (Hibat-Allah et al., 2020) and molecules in second quantization (Barrett et al., 2022). In both cases the state space is discretized, simplifying the sampling from the low-dimensional conditional probability distributions.

3.3 Optimization

After discussing the variational principle and the sampling technique for the high-dimensional integral, we can now delve into the problem of optimizing for the parameters θ. Applying the variational principle, we can utilize the formulation provided in Eq. (44) as our loss function:

$$L(\theta) = \langle E \rangle_{\mathbf{r}\sim\Psi_\theta^2} := \mathbb{E}_{\mathbf{r}\sim\Psi_\theta^2}\left[\frac{\mathcal{H}\Psi_\theta(\mathbf{r})}{\Psi_\theta(\mathbf{r})}\right]. \tag{64}$$

To perform gradient descent, the computation of the gradient with respect to the parameters would in general require third derivatives: second derivatives with respect to $\mathbf{r}$ for the kinetic energy and first derivatives with respect to θ. Furthermore, stochastic gradient descent potentially leads to a biased estimator because the sampling process depends on the parameterized wave function Ψ_θ. Fortunately, one can exploit the hermiticity of the Hamiltonian to rewrite the gradient with respect to the parameters of the loss function as:

$$\nabla_\theta L(\theta) = 2\left\langle \left(\frac{\mathcal{H}\Psi_\theta(\mathbf{r})}{\Psi_\theta(\mathbf{r})} - L(\theta)\right)\nabla_\theta \log|\Psi_\theta| \right\rangle_{\mathbf{r}\sim p_\theta(\mathbf{r})}. \tag{65}$$

Eq. (65) allows the computation of an unbiased estimator for stochastic gradient descent with, at most, second derivatives. Additionally, in the case that Ψ_θ represents the true wave function, the Monte Carlo estimator has zero variance due to the local energy $\frac{H\Psi_\theta(\mathbf{r})}{\Psi_\theta(\mathbf{r})}$ being spatially constant and equal to $L(\theta)$. A full derivation of the gradient can be found for example in Inui et al. (2021). The energy can be minimized using gradient based optimizers, such as Stochastic

Gradient Descent (SGD) or the Adam optimizer (Kingma and Ba, 2017), used in many deep-learning applications. The update rule for SGD with learning rate λ is given by

$$\theta_{t+1} = \theta_t - \lambda \nabla_\theta L. \tag{66}$$

Convergence of optimization can be substantially accelerated by not using the energy gradient directly as in (66), but rather preconditioning it with the following matrix $\mathbf{S} \in \mathbb{R}^{n_{\text{param}} \times n_{\text{param}}}$:

$$S_{\mu\nu} := \left\langle \frac{\partial \log|\Psi|}{\partial \theta_\mu} \frac{\partial \log|\Psi|}{\partial \theta_\nu} \right\rangle - \left\langle \frac{\partial \log|\Psi|}{\partial \theta_\mu} \right\rangle \left\langle \frac{\partial \log|\Psi|}{\partial \theta_\nu} \right\rangle \tag{67}$$

and then using this preconditioned gradient for stochastic gradient descent

$$\theta_{t+1} = \theta_t - \lambda \mathbf{S}^{-1} \nabla_\theta L. \tag{68}$$

Here, and in what follows, we use the notation

$$\langle \Phi \rangle := \mathbb{E}_{\mathbf{r} \sim \Psi_\theta^2}[\Phi]$$

for any $\Phi : \mathbb{R}^{n_{\text{el}} \times 3} \to \mathbb{R}$.

The update rule (68) is known as *Stochastic Reconfiguration* in the physics community (Becca and Sorella, 2017) (where $\mathbf{S}$ is then referred to as the *Quantum Geometric Tensor*) and is very closely related to *Natural Gradient Descent* in the machine learning community (Martens and Grosse, 2015) (where an object closely related to $\mathbf{S}$ is referred to as the Fisher information matrix). The following argument, adapted from Becca and Sorella (2017), should give some perspective on why using $\mathbf{S}$ as a preconditioner is a sensible choice.

Stochastic reconfiguration as a local metric

When performing SGD, a crucial choice is the stepsize λ. One way of formulating this is to consider at every step as loss $\mathcal{L}$ the original loss $L(\theta)$ plus an additional regularization term, which penalizes large changes δ in parameter space.

$$\mathcal{L}^{\text{SGD}}(\delta) = L(\theta + \delta) + \frac{\lambda}{2} \delta^T \delta \tag{69}$$

$$\delta := \theta_{t+1} - \theta_t \tag{70}$$

When minimizing Eq. (69) with respect to all parameter updates δ_μ the classical SGD update rule is recovered:

$$\frac{\partial \mathcal{L}^{\text{SGD}}}{\partial \delta_\mu} = \frac{\partial L}{\partial \delta_\mu} + \lambda \delta_\mu \stackrel{!}{=} 0 \tag{71}$$

$$\delta = -\lambda \nabla_\theta L. \tag{72}$$

SGD with a given learning-rate λ therefore minimizes the energy, while at the same time minimizing the Euclidean norm of the parameter update. While this is not an unreasonable choice per se, it would be better to minimize the energy, while making *minimal changes to the wavefunction*. After all, the wavefunction might be very sensitive to some parameters and insensitive to others. We would therefore like to make small steps for sensitive parameters and larger steps for insensitive parameters.

The following metric can be used to assess the distance between two unnormalized real-valued wavefunctions Ψ and Φ:

$$s(\Phi, \Psi)^2 = 1 - \frac{\langle \Psi, \Phi \rangle^2}{\langle \Psi, \Psi \rangle \langle \Phi, \Phi \rangle} \tag{73}$$

Eq. (73) corresponds to 1 minus the squared overlap of the normalized wavefunctions and is thus 0 for $\Phi \equiv \Psi$ and 1 for $\Phi \perp \Psi$. Using Eq. (73) as a metric to regularize the loss yields

$$\mathcal{L}^{\mathrm{SR}}(\delta) = L(\theta + \delta) + \lambda \left(1 - \frac{\langle \Psi_\theta, \Psi_{\theta+\delta} \rangle^2}{\langle \Psi_\theta, \Psi_\theta \rangle \langle \Psi_{\theta+\delta}, \Psi_{\theta+\delta} \rangle} \right). \tag{74}$$

The updated wavefunction $\Psi_{\theta+\delta}$ can be expressed as Taylor expansion up to first order,

$$\Psi_{\theta+\delta} \approx \Psi_\theta + \delta^T \nabla_\theta \Psi_\theta, \tag{75}$$

yielding (subscripts θ omitted for clarity):

$$\mathcal{L}^{\mathrm{SR}}(\delta) = \langle E \rangle + \lambda \left(1 - \frac{\left\langle \Psi, \Psi + \delta^T \nabla_\theta \Psi \right\rangle^2}{\langle \Psi, \Psi \rangle \left\langle \Psi + \delta^T \nabla_\theta \Psi, \Psi + \delta^T \nabla_\theta \Psi \right\rangle} \right). \tag{76}$$

Expanding the regularization term, dividing the denominator and enumerator by $\langle \Psi, \Psi \rangle^2$, and introducing O yields

$$O := \frac{\nabla_\theta \Psi}{\Psi} = \nabla_\theta \log |\Psi| \tag{77}$$

$$\mathcal{L} = \langle E \rangle + \lambda \left(1 - \frac{\left(1 + \delta^T \frac{\langle \Psi, \nabla_\theta \Psi \rangle}{\langle \Psi, \Psi \rangle} \right)^2}{1 + 2\delta^T \frac{\langle \Psi, \nabla_\theta \Psi \rangle}{\langle \Psi, \Psi \rangle} + \delta^T \frac{\langle \nabla_\theta \Psi, \nabla_\theta \Psi \rangle}{\langle \Psi, \Psi \rangle} \delta} \right) \tag{78}$$

$$= \langle E \rangle + \lambda \left(1 - \frac{\left(1 + \left\langle \delta^T O \right\rangle \right)^2}{1 + 2\delta^T \langle O \rangle + \delta^T \left\langle O O^T \right\rangle \delta} \right) + \mathcal{O}(|\delta|^3). \tag{79}$$

Expanding the denominator up to second order in δ (using $(1+x)^{-1} \approx 1 - x + x^2$) and multiplying all terms finally yields

$$\mathcal{L}^{\mathrm{SR}}(\delta) = \langle E \rangle + \lambda \delta^T \mathbf{S} \delta + \mathcal{O}(|\delta|^3) \tag{80}$$

$$\mathbf{S} = \left\langle O O^T \right\rangle - \langle O \rangle \langle O \rangle^T \tag{81}$$

The regularized loss in Eq. (80) has the same structure as (69): The original loss + a quadratic regularization term – the only difference being that this time the metric is given by $\mathbf{S}$ instead of the Euclidean-norm. When minimizing this regularized loss one obtains the stochastic reconfiguration update rule (up to a factor of 2 in the learning rate)

$$\delta = -2\lambda \mathbf{S}^{-1} \nabla_\theta L. \tag{82}$$

Note that $\mathbf{S}$ has been motivated here using (73) as distance metric. The same result (up to a constant factor) can be obtained by expanding the Kullback-Leibler divergence – a well known divergence to measure the distance between probability distribution – between the probability distributions $|\Psi_\theta^2|$ and $|\Psi_{\theta+\delta}|^2$. If the probability distributions are normalized then $\langle O \rangle \equiv 0$, simplifying (81) to the first term and the corresponding update rule of *natural gradient descent*.

Toy example: SR for 2-parameter system

Fig. 6 demonstrates the effect of this preconditioning on a 1D-example (a single particle in a parabolical potential) with a wavefunction that has only two parameters:

$$\Psi(x) = e^{-\frac{1}{2}\left(\frac{x-\theta_2}{2\sigma(\theta_1)}\right)^2}, \tag{83}$$

with the sigmoid function $\sigma(\theta) = \frac{1}{1+e^{-\theta}}$. This system has its ground-state at $\theta_1 = \theta_2 = 0$, depicted as Ψ_{GS} in Fig. 6a and the corresponding point in parameter-space θ_{GS} in Fig. 6b.

Starting from an arbitrary initial wavefunction Ψ_0 (and corresponding parameters θ_0), two distinct new wavefunctions (and corresponding parameters) are depicted. The parameters θ_{SGD} (and its wavefunction Ψ_{SGD}) are obtained from the gradient descent update rule $\theta_{\text{SGD}} = \theta_0 - \lambda_{\text{SGD}} \nabla_\theta E$. The parameters θ_{SR} (and its wavefunction Ψ_{SR}) are obtained from the stochastic-reconfiguration update rule $\theta_{\text{SR}} = \theta_0 - \lambda_{\text{SR}} \mathbf{S}^{-1} \nabla_\theta E$. The learning rates λ_{SGD} and λ_{SR} are chosen such that the Euclidean distance in parameter space is identical in both case (as depicted in Fig. 6c). However, the change of the wavefunction is markedly different: The SR-update rule leads to a much smaller change in the wavefunction (compared to the SGD update rule). This can be seen from the smaller S-distance in Fig. 6d, as well as visually when comparing Ψ_{SR} and Ψ_{SGD} in Fig. 6a. Overall this leads to a lower energy after the update step (as can be seen in Fig. 6b) and will lead to overall faster convergence towards the ground-state when iterating. The effect of the preconditioning is that the SR-update skews the update step towards larger updates along the θ_1 parameter, which is less sensitive in this point of the parameter space.

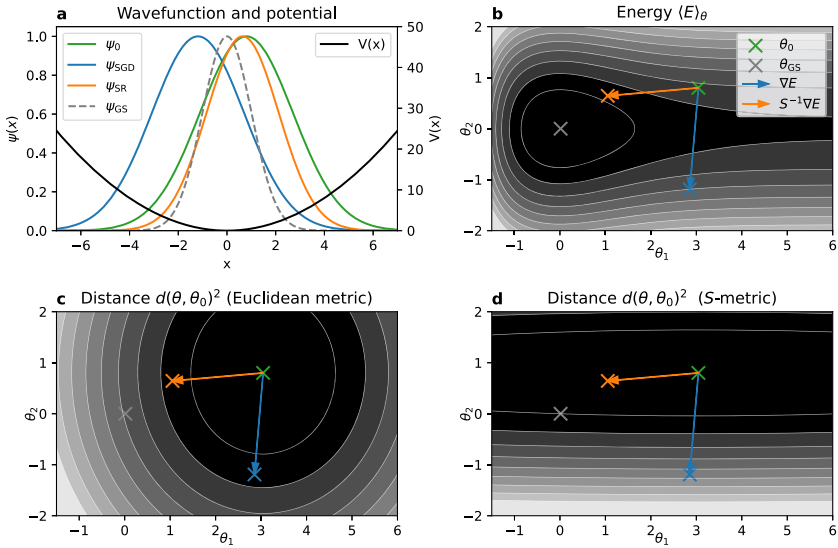

FIGURE 6 1D toy example for 2-parameter wavefunction: a) Plot of ground-state wavefunction Ψ_{GS}, initial wavefunction Ψ_0, and the resulting wavefunctions after update steps according to stochastic gradient descent (SGD) and stochastic reconfiguration (SR). b-d) Contour-plots as a function of wavefunction parameters θ_1 and θ_2 (darker colors correspond to lower values). b) Energy expectation value of corresponding wavefunction. c,d) Distance from initial parameter vector θ_0 measured in Euclidean metric and the metric induced by the preconditioner S.

Practical considerations for stochastic reconfiguration

Computing S^{-1} can be practically challenging. The first complication is that when $\mathbf{S}$ is being estimated from N_{s} samples it is at most of rank N_{s}:

$$S_{\mu\nu} = \langle O_\mu O_\nu \rangle - \langle O_\mu \rangle \langle O_\nu \rangle \tag{84}$$

$$= \Big\langle \Big(O_\mu - \langle O_\mu \rangle \Big) \Big(O_\nu - \langle O_\nu \rangle \Big) \Big\rangle \tag{85}$$

$$\approx \sum_{n=1}^{N_{\mathrm{s}}} \Big(O_\mu(\mathbf{r}_n) - \bar{O}_\mu \Big) \Big(O_\nu(\mathbf{r}_n) - \bar{O}_\nu \Big) \tag{86}$$

Therefore for typical values of $N_{\mathrm{s}} \approx 10^3$ and $n_{\mathrm{param}} \approx 10^6$, $\mathbf{S}$ is rank deficient and cannot be inverted. The latter problem is typically addressed via Tikhonov regularization with a small damping constant ϵ:

$$\mathbf{S}_{\mathrm{reg}} = \mathbf{S} + \epsilon \, \mathbb{1}_{N_{\mathrm{s}}} \tag{87}$$

Another approach is to estimate $\mathbf{S}$ not only from the current batch, but as a moving average of the estimates from past batches, thus increasing the rank of the estimator. This helps to reduce Monte Carlo noise in the estimation but typically

still requires regularization to avoid a singular matrix $\mathbf{S}$. The second complication arises due to the size of $\mathbf{S}$, which is of dimension $n_{\text{param}} \times n_{\text{param}}$. Therefore, for a neural network wavefunction with $\approx 10^6$ parameters, even storing this matrix with $\approx 10^{12}$ elements becomes impossible. Even worse, this large matrix must be inverted, an operation that has computational cost $\mathcal{O}({n_{\text{param}}}^3)$ using Gaussian elimination. There are two viable routes in practice: Find a (sparse) approximation of $\mathbf{S}$ and invert it exactly, or find a way to approximately invert $\mathbf{S}$ without fully materializing $\mathbf{S}$.

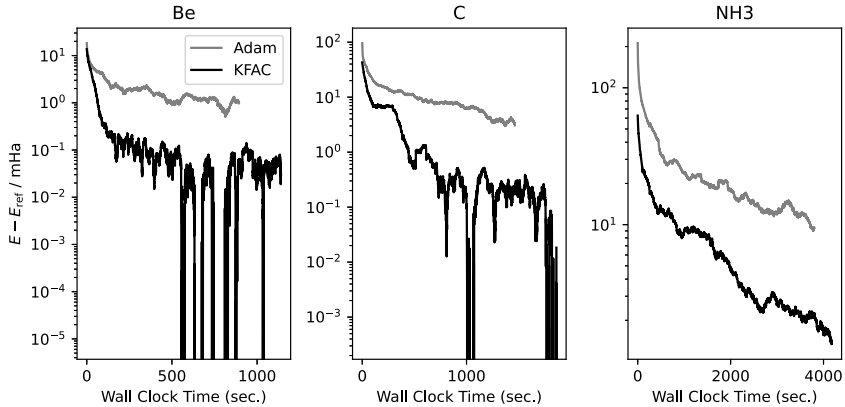

FIGURE 7 Comparing the optimizers Adam (grey) and KFAC (black) for three molecules (Beryllium, Carbon, and Ammonia). The x-axis represents the wall clock time in seconds, and the y-axis (log-scale) the energy error with respect to a highly accurate reference calculation in millihartree (mHa).

KFAC (Kronecker-Factored Approximation of Curvature) (Martens and Grosse, 2015) is of the first type, making two approximations to $\mathbf{S}$. First it assumes that there are no dependencies between parameters belonging to different layers of the neural network, effectively assuming $\mathbf{S}$ to be block-diagonal. Second it assumes that each remaining block can be expressed as an eponymous Kronecker product of two smaller matrices. This allows inversion of the approximated $\mathbf{S}$ via inversion of many small matrices, which is computationally feasible even for networks involving millions of parameters. KFAC has first been applied to Neural Network wavefunctions by Pfau et al. (2020) and since been used widely throughout the neural wavefunction community. Compared to nonpreconditioned methods it yields substantially faster rates of convergence as depicted in Fig. 7. While it is computationally efficient and can yield good results, it has two downsides in practice. First, it involves approximations that cannot be systematically improved upon. Second, the optimizer does not only require access to the wavefunction, energies and their gradients, but also requires access to intermediate activations and gradients of the model. This can introduce substantial complexity for practical implementation and leads to some unwanted coupling between the wavefunction model and the optimizer.

An alternative approach is to make no approximations to $\mathbf{S}$, but to only invert it approximately. One common approach is to use the Conjugate Gradient (CG) method to compute $\mathbf{S}^{-1}\nabla E$ without materializing $\mathbf{S}$. CG only requires the repeated evaluation of the matrix-vector product $\mathbf{Sx}$ for arbitrary vectors $\mathbf{x}$. This can be obtained by a vector-jacobian-product (VJP) followed by a jacobian-vector-product (VJP), which are implemented using back-propagation and forward-mode differentiation respectively and don't require materializing the full jacobian.

Another approach is to use the fact that the regularized $\mathbf{S}_{\text{reg}}$ is a sum of a full-rank, but easy to invert diagonal matrix ($\epsilon\,\mathbb{1}$), and a rank-N_s matrix $\mathbf{S}$. Inversion of $\mathbf{S}_{\text{reg}}$ can therefore be done using the Sherman-Morrison-Woodbury formula (Woodbury, 1950), which only requires the inversion of a $N_\text{s} \times N_\text{s}$ matrix. This forms the basis for the MinSR (Rende et al., 2023) and SPRING optimizer (Goldshlager et al., 2024).

Supervised pretraining

In addition to the variational optimization, Pfau et al. (2020) proposed to perform a supervised optimization phase with respect to a reference method. This step, referred to as *supervised pretraining* in the following, minimizes the difference between the neural network orbitals and reference orbitals (we omit the spin dependence in the notation)

$$L^{\text{pre}}(\theta) = \mathbb{E}_{\mathbf{r}\sim p_\theta(\mathbf{r})}\left[\sum_{k=1}^{n_\text{el}}\sum_{i=1}^{n_\text{el}}\left(\phi_k^{\text{ref}}(r_i) - \phi(\theta_i)(r_i)\right)^2\right], \tag{88}$$

whereas the neural network-based orbitals are calculated as described in (119) and (120). As reference calculation, a mean field solution such as Hartree Fock, is usually used, where each orbital depends on a single electron. Pfau et al. (2020) argued that supervised pretraining can improve convergence and numerical stability of the subsequent variational optimization. Compared to variational optimization, supervised pretraining is computationally cheaper because it does not require evaluation of the Hamiltonian; in particular it avoids the derivatives associated with the kinetic energy. The downside of supervised pretraining is that it unlike variational optimization it requires a reference method. Furthermore excessive pretraining can bias the network initialization, leading to less accurate results after subsequent variational optimization (Gerard et al., 2022).

4 Deep learning VMC

In this section we introduce the DL-VMC method. In particular, we review how neural networks can be used to construct suitable parametrized orbitals (55) that can serve as a numerical ansatz in the VMC problem (44).

4.1 Multilayer perceptrons

The multi layer perceptron (MLP) forms the basic building block of all deep-learning based architectures. It consists of L layers, alternating between affine transformations and elementwise nonlinear functions. Given an input $x^0 \in \mathbb{R}^{d_0}$, the output of each subsequent layer $l = 1 \dots L$ is computed as

$$y_n^l = \sum_m w_{nm}^l x_m^{l-1} + b_m^l \qquad l = 1 \dots L \tag{89}$$

$$x_n^l = \sigma(y_n^l) \qquad l = 1 \dots L-1 \tag{90}$$

$$\mathrm{MLP}\left(x^0\right) := y^L, \tag{91}$$

with trainable weights $w^l \in \mathbb{R}^{d_l \times d_{l-1}}$, $b^l \in d_l$ for every layer l and a nonlinear function σ, referred to as *activation function*. Common choices for this activation function include

$$\sigma(x) = \tanh(x) \tag{92}$$

$$\sigma(x) = \mathrm{SiLU}(x) = \frac{x}{1+e^{-x}} \tag{93}$$

Note that $\mathrm{ReLU}(x) := \max(x, 0)$, which is used extensively in many deep-learning applications, is not used for neural wavefunctions, because its first derivative is discontinuous and its second derivative (required for the kinetic energy of $\mathcal{H}^{\mathrm{BO}}$) is zero everywhere.

4.2 Overall structure of neural network wavefunctions

To model the functions $\Phi_{\uparrow}^{k,i}(\mathbf{r})$ (52) and $\Phi_{\downarrow}^{k,i}(\mathbf{r})$ (53), several different architectures have been proposed, which all follow the following structure. All functions in this section can depend on parameters θ, but we suppress this index for clarity.

1. **Input features**: Compute features $x_i \in \mathbb{R}^{n_{\mathrm{feat}}^{\mathrm{el}}}$ for single-electrons and $p_{ij} \in \mathbb{R}^{n_{\mathrm{feat}}^{\mathrm{el\text{-}el}}}$ for pairs of electrons i, j from the coordinates $\mathbf{r}$ and spins σ.

$$x_i = f^{\mathrm{el}}(r_i, \sigma_i, (\mathbf{R}, \mathbf{Z})) \qquad i = 1 \dots n_{\mathrm{el}} \tag{94}$$

$$p_{ij} = f^{\mathrm{el\text{-}el}}(r_i, \sigma_i, r_j, \sigma_j) \qquad i, j = 1 \dots n_{\mathrm{el}} \tag{95}$$

Analogous to the multisets $\{\mathbf{r}^{\uparrow}\}$, $\{\mathbf{r}^{\downarrow}\}$ defined in (51) we define the following multisets of features:

$$\begin{aligned} \{p_i^{\uparrow}\} &:= \{(x_1, p_{i,1}), \dots, (x_{n_{\uparrow}}, p_{i,n_{\uparrow}})\}, \\ \{p_i^{\downarrow}\} &:= \{(x_{n_{\uparrow}+1}, p_{i,n_{\uparrow}+1}), \dots, (x_{n_{\mathrm{el}}}, p_{i,n_{\mathrm{el}}})\} \end{aligned} \tag{96}$$

$$\begin{aligned} \{p_{i,\setminus i}^{\uparrow}\} &:= \{p_i^{\uparrow}\} \setminus \{(x_i, p_{i,i})\}, \\ \{p_{i,\setminus i}^{\downarrow}\} &:= \{p_i^{\downarrow}\} \setminus \{(x_i, p_{i,i})\} \end{aligned} \tag{97}$$

2. **Embedding**: Compute a high-dimensional embedding $h_i \in \mathbb{R}^{d_{\text{emb}}}$ for each electron i. These embeddings not only depend on the input features of electron i but also on the multisets of input features of all other electrons.

$$h_i = h\left(x_i,\ \{p^{\uparrow}_{i,\setminus i}\},\ \{p^{\downarrow}_{i,\setminus i}\}\right), \qquad i = 1 \dots n_{\text{el}} \tag{98}$$

3. **Orbitals**: Compute entries $\phi^{d,k,i}_{\uparrow\downarrow}$ from the embeddings h_i for each orbital k and determinant d.

$$\phi^{d,k,i}_{\uparrow} = \phi^{\uparrow}_{dk}(h_i) \qquad i = 1 \dots n_{\uparrow}, \quad k = 1 \dots n_{\text{el}}, \quad d = 1 \dots n_{\text{det}} \tag{99}$$

$$\phi^{d,k,i}_{\downarrow} = \phi^{\downarrow}_{dk}(h_i) \quad i = n_{\uparrow} + 1 \dots n_{\text{el}}, \quad k = 1 \dots n_{\text{el}}, \quad d = 1 \dots n_{\text{det}} \tag{100}$$

4. **Slater determinant**: Compute Ψ as a sum of Slater determinants of these orbitals $\phi^{d,k,i}_{\uparrow}$ (cf. (54)) and optionally multiply it with a Jastrow factor J that is invariant under permutation of electrons with the same spin.

$$\begin{aligned}\Psi = J\left(\{h_i\}_{i=1\dots n_{\uparrow}}, \{h_i\}_{i=n_{\uparrow}+1\dots n_{\text{el}}}\right) \\ \times \sum_d \det\left(\phi^{d,1}_{\uparrow}, \dots, \phi^{d,n_{\uparrow}}_{\uparrow}, \phi^{d,n_{\uparrow}+1}_{\downarrow}, \dots, \phi^{d,n_{\text{el}}}_{\downarrow}\right)\end{aligned} \tag{101}$$

Note that by construction the embeddings h_i are equivariant under permutation of electrons of the same spin and therefore the whole ansatz satisfies antisymmetry. In the following we discuss a few common choices for each of these four steps.

4.3 Input features

Pairwise features

For the pairwise features p_{ij}, typically 3D difference vectors $r_i - r_j$ and the pairwise distances $|r_i - r_j|$ are being used. Some ansätze (Pfau et al., 2020; Gerard et al., 2022; Gao and Günnemann, 2022) use a simple concatenation (denoted by $[\cdot, \cdot]$) of these features

$$p^{\text{concat}}_{ij} = \left[r_i - r_j, |r_i - r_j|\right], \qquad p^{\text{concat}}_{ij} \in \mathbb{R}^4. \tag{102}$$

Others additionally include trainable functions of these distances and difference vectors. For example, Gao and Günnemann (2023a) proposed

$$\begin{aligned}p^{\text{MOON}}_{ij,\nu} = \text{MLP}_{\nu}(r_i - r_j) \sum_{\mu} W_{\nu\mu} \exp\left(-\frac{|r_i - r_j|^2}{\zeta_{\mu}}\right) \\ \nu = 1 \dots n^{\text{el-el}}_{\text{feat}},\ \mu = 1 \dots n_{\text{filters}}\end{aligned} \tag{103}$$

It may seem redundant to include both $r_i - r_j$ and $|r_i - r_j|$ as an input feature, because the norm is just a function of $r_i - r_j$. But including $|r_i - r_j|$ which has discontinuous derivatives at $r_i = r_j$ allows the network to model functions that are not smooth at $r_i = r_j$, which is required to satisfy the Kato cusp conditions (Kato, 1957). Early approaches (Hermann et al., 2020) used only the distances $|r_i - r_j|$ as input features, but this has been recognized to be insufficiently expressive (Gerard et al., 2022).

For large molecules the interparticle distances $|r_i - r_j|$ can become very large, making training of neural networks using them numerically challenging. von Glehn et al. (2023) proposed to logarithmically scale the interparticle distances and differences to alleviate this problem:

$$\widetilde{r}_{ij} = \log\left(1 + |r_i - r_j|\right) \tag{104}$$

$$p_{ij}^{\log} = \left[\frac{\widetilde{r}_{ij}}{|r_i - r_j|}\left(r_i - r_j\right), \widetilde{r}_{ij}\right], \qquad p_{ij}^{\log} \in \mathbb{R}^4. \tag{105}$$

Single electron features

The single electron features x_i are typically computed as functions of the electron-nuclei distances and differences. FermiNet (Pfau et al., 2020) proposed a simple concatenation of all electron-nucleus pairs

$$x_i^{\text{concat}} = \left[r_i - R_1, |r_i - R_1|, \ldots, r_i - R_{n_{\text{nuc}}}, |r_i - R_{n_{\text{nuc}}}|\right], \qquad p_i^{\text{concat}} \in \mathbb{R}^{4n_{\text{nuc}}}, \tag{106}$$

whereas Gao and Günnemann (2022); Gerard et al. (2022) use sums of trainable functions of the differences and distances.

$$x_i^{\text{MLP}} = \sum_{J=1}^{n_{\text{nuc}}} \text{MLP}\Big([r_i - R_J, |r_i - R_J|]\Big) \tag{107}$$

While some architectures (von Glehn et al., 2023) encode spin explicitly as a feature $x_i = \sigma_i$, many others do not encode spin explicitly, but instead opt for different subsequent embedding functions depending on spin.

4.4 Embedding

The role of the embedding network is to take simple input features x_i and p_{ij} and compute embeddings h_i that form expressive basis functions for the subsequent many-body orbitals. To do this, the embedding network must on the one hand be able to incorporate information from all other electrons $i \neq j$, and on the other hand be able to represent arbitrary functions of a single electron i. These two requirements are typically addressed by interleaving two kinds of computation over multiple rounds l: A function f that gathers information from other electrons and function that acts only on a single electron (typically implemented as MLP or a single affine transformation). Most embedding networks

thus follow the following structure:

$$h_i^0 = x_i \qquad \text{Initialization} \tag{108}$$

$$m_i^{\uparrow,l} = \sum_{j \in \{\uparrow \setminus i\}} f^{\uparrow}(h_i^{l-1}, h_j^{l-1}, p_{ij}) \qquad \text{Gather information across electrons} \tag{109}$$

$$m_i^{\downarrow,l} = \sum_{j \in \{\downarrow \setminus i\}} f^{\downarrow}(h_i^{l-1}, h_j^{l-1}, p_{ij})$$

$$h_i^l = \text{MLP}\left(\left[h_i^{l-1},\ m_i^{\uparrow,l},\ m_i^{\downarrow,l}\right]\right) \qquad \text{Single-electron computation} \tag{110}$$

The sets $\{\uparrow \setminus i\}$ and $\{\downarrow \setminus i\}$ correspond to the electron-indices of all electrons of a given spin excluding i:

$$\{\uparrow \setminus i\} := \{1 \dots n_{\uparrow}\} \setminus \{i\}, \qquad \{\downarrow \setminus i\} := \{n_{\uparrow} + 1 \dots n_{\text{el}}\} \setminus \{i\} \tag{111}$$

The initial embeddings are $h^0 \in \mathbb{R}^{n_{\text{feat}}^{\text{el}}}$, and $h^l \in \mathbb{R}^{d_{\text{emb}}}$ for $l > 0$.

After iterating Eq. (109) and (110) for $l = 1 \dots L$, the final embeddings are given by the output of the last layer, i.e. $h_i = h_i^L$. A few design considerations are worth discussing:

- The *messages* $m_i^{\uparrow\downarrow,l}$ in Eq. (109) are a sum over all particles of a given spin. Since the sum is a permutation invariant operation, the resulting message is invariant under permutation of two electrons of the same spin. This ensures that h_i has the structure defined in (98), ultimately ensuring wavefunction antisymmetry.
- Not all parts of the network have the same impact on computational cost. While the functions $f(h_i, h_j, p_{ij})$ in Eq. (109) are evaluated for every pair of electrons – and thus have computational cost scaling as $\mathcal{O}({n_{\text{el}}}^2)$ – the MLP in Eq. (110) is only evaluated for every electron, thus scaling as $\mathcal{O}(n_{\text{el}})$. This difference in scaling is typically reflected by the fact that most architectures use wide (and thus costly) MLPs for the one-electron computations (Eq. (110)), and computationally cheaper functions for $f^{\uparrow}$ and $f^{\downarrow}$ (Eq. (109)).
- Some architectures differentiate between messages from up- and down-electrons (as denoted in Eq. (109)), while others differentiate between messages from spin-parallel or spin-antiparallel electrons. The latter choice enforces invariance w.r.t. exchanging all spin-up particles with spin-down particles and has been shown to be advantageous for closed-shell systems (Gao and Günnemann, 2023b).

Given this very general framework, the key difference between the various embedding architectures lies therefore in the message functions $f^{\uparrow}$, $f^{\downarrow}$, with a few common choices outlined below.

Hartree-Fock

If no information from other electrons is gathered (e.g. $f^{\uparrow} \equiv f^{\downarrow} \equiv 0$), the final embedding h_i only depends on the input features of that electron x_i. The network cannot capture any correlation effects and thus the best possible accuracy is Hartree-Fock.

FermiNet

In FermiNet (Pfau et al., 2020; Spencer et al., 2020) the message-function f simply concatenates the feature vectors of embedding h_j and an MLP of p_{ij}.

$$f^{\uparrow}(h_j, p_{ij}) = f^{\downarrow}(h_j, p_{ij}) = \left[h_j,\ \mathrm{MLP}\left(p_{ij}\right)\right] \tag{112}$$

FermiNet clearly improves upon a simple noninteracting embedding, by including in every layer information about all other electron embeddings as well as their relative positions. Note however that in FermiNet all electron embeddings h_j contribute equally to the message m_i, irrespective of the distance between electron i and j. This runs against physical intuition, which suggests that electrons at large separations would have a smaller influence.

Graph convolutional neural networks

In a Graph Convolutional Neural Network (GCN) (Zhou et al., 2020), the contributions of each electron j to the message m_i are weighted by their relative geometric positions encoded in p_{ij}. This weighting is commonly achieved using an elementwise product along the feature dimension, denoted by $\odot$:

$$f^{\uparrow}(h_j, p_{ij}) = \mathrm{MLP}_{\mathrm{h}}\left(h_j\right) \odot \mathrm{MLP}_{\mathrm{p}}^{\uparrow}\left(p_{ij}\right), \tag{113}$$

$$f^{\downarrow}(h_j, p_{ij}) = \mathrm{MLP}_{\mathrm{h}}\left(h_j\right) \odot \mathrm{MLP}_{\mathrm{p}}^{\downarrow}\left(p_{ij}\right). \tag{114}$$

Empirically, including graph convolutions can improve the accuracy and convergence of the ansatz compared to the purely MLP-based FermiNet (Gerard et al., 2022). One potential reason for the improved performance is that it enables the network to put larger weight on closer neighboring electrons than electrons which are far apart. Some approaches (Gao and Günnemann, 2023a) enforce this prior knowledge explicitly, by multiplying the $\mathrm{MLP}\left(p_{ij}\right)$ with functions that explicitly decay as a function of the distance between electrons i and j.

Self-attention based

Neither in FermiNet nor the GCN embedding does the message m_i explicitly depend on the message receiver h_i, but instead only depends on the message sender h_j and the pairwise features p_{ij}. Self-attention is an approach where the weighting of each message j is computed as an inner product between a query vectors q_i (derived from the receiving embedding h_i) and a key vector k_j (derived from the sending embedding h_j). For embeddings $h \in \mathbb{R}^{n_{\mathrm{el}} \times d_{\mathrm{emb}}}$ and

trainable matrices $W^{\mathrm{q}}, W^{\mathrm{k}}, W^{\mathrm{v}} \in \mathbb{R}^{d_{\mathrm{emb}} \times d_{\mathrm{attn}}}$ the weights $w \in \mathbb{R}^{n_{\mathrm{el}} \times n_{\mathrm{el}}}$ and the corresponding message function are given as

$$q = hW^{\mathrm{q}}, \qquad k_j = hW^{\mathrm{k}}, \qquad v_j = hW^{\mathrm{q}}, \qquad q, k, v \in \mathbb{R}^{n_{\mathrm{el}} \times d_{\mathrm{attn}}} \tag{115}$$

$$\hat{w} = \exp\left(\frac{qk^T}{\sqrt{d_{\mathrm{attn}}}}\right), \qquad \hat{w} \in \mathbb{R}^{n_{\mathrm{el}} \times n_{\mathrm{el}}} \tag{116}$$

$$w_{ij} = \frac{\hat{w}_{ij}}{\sum_j' \hat{w}_{ij'}} \tag{117}$$

$$f^{\uparrow}(h_i, h_j) = f^{\downarrow}(h_i, h_j) = w_{ij} v_j. \tag{118}$$

The message m_i explicitly depends on the embedding for electron i and j, but no longer explicitly depends on their pairwise features p_{ij}. This geometric information must be inferred from the inner product of q_i and k_j, and thus requires that the single-electron input features contain information about their absolute positions. This architecture was implemented by von Glehn et al. (2023) and has been empirically shown to be among the most expressive ansätze.

4.5 Orbitals

Given permutation equivariant embeddings h_i, the elements $\phi_{\uparrow\downarrow}^{d,k,i}$ of the slater matrix are typically computed as

$$\phi_{\uparrow}^{d,k,i} = \left(W_{dk}^{\uparrow} \cdot h_i\right) \widetilde{\varphi}_{dk}^{\uparrow}(r_i), \qquad i = 1 \ldots n_{\uparrow} \tag{119}$$

$$\phi_{\downarrow}^{d,k,i} = \left(W_{dk}^{\downarrow} \cdot h_i\right) \widetilde{\varphi}_{dk}^{\downarrow}(r_i), \qquad i = n_{\uparrow} + 1 \ldots n_{\mathrm{el}} \tag{120}$$

where $W^{\uparrow\downarrow}$ are trainable matrices $W^{\uparrow}, W^{\downarrow} \in \mathbb{R}^{n_{\mathrm{det}} \times n_{\mathrm{el}} \times d_{\mathrm{emb}}}$, and $\widetilde{\varphi}$ is an envelope function enforcing that $\phi \to 0$ as $r_i \to \infty$.

For molecules the envelope function is typically expressed as a sum over nuclei, leading to

$$\widetilde{\varphi}_{dk}(r_i) = \sum_{J=1}^{n_{\mathrm{nuc}}} \varphi_{dkJ}(r_i). \tag{121}$$

The most common choice for the envelope function, originally proposed by Pfau et al. (2020) and simplified by Spencer et al. (2020) is *exponential envelopes*

$$\varphi_{dkJ}(r_i) = \pi_{dkJ}\, e^{-\alpha_{dkJ}|r_i - R_J|}, \tag{122}$$

with trainable parameters $\pi, \alpha \in \mathbb{R}^{n_{\mathrm{el}} \times n_{\mathrm{det}} \times n_{\mathrm{nuc}}}$.

An alternative proposed by Hermann et al. (2020) is using the single-particle orbitals from a Hartree-Fock calculation

$$\varphi_{dkJ}(r_i) = \varphi_{kJ}^{\mathrm{HF}}(r_i) = \sum_{\mu=1}^{n_{\mathrm{basis}}} c_{kJ\mu}\, b_\mu(r_i - R_J), \tag{123}$$

where $b_\mu : \mathbb{R}^3 \to \mathbb{R}$ are atom-centered basis functions and $c_{kJ\mu} \in \mathbb{R}$ are the expansion coefficients of orbital φ_k^{HF} in this basis.

Given that Hartree-Fock yields a good approximation of the groundstate wavefunction, one might think that using Hartree-Fock orbitals as envelopes provides a useful prior and good starting point for optimization. However, in practice the exponential envelopes are not only simpler to implement but also lead to substantially more accurate results (Gerard et al., 2022), potentially due to a bias introduced by the Hartree-Fock orbitals.

Even though using the HF-envelopes directly can decrease accuracy – and several groups that originally used them (Hermann et al., 2020; Scherbela et al., 2022), replaced them in later work with exponential envelopes (Gerard et al., 2022; Schätzle et al., 2023) – there is still information in the HF-orbitals which can be used: First, different HF-orbitals typically have different length-scales: Some orbitals (known by chemists as core orbitals) are tightly localized on an atom, whereas other orbitals (known by chemists as valence orbitals) are somewhat delocalized. This can be quantified and used to initialize the exponents α of the exponential envelopes, using large values for α to initialize strongly localized core orbitals and small values to initialize delocalized valence electrons. Numerical experiments show that this initialization accelerates wavefunction optimization (Gerard et al., 2022), in particular for heavy atoms where the length-scale of core orbitals differs by an order of magnitude from the length-scale of the valence electrons. Second, one can use the expansion coefficients of an HF-orbital as a descriptor of that orbital. This can be useful when designing a transferable wavefunction (cf. Sec. 4.7).

4.6 Jastrow factor

The wavefunction can be multiplied with a function $J(\mathbf{r}) : \mathbb{R}^{n_{\mathrm{el}} \times 3} \to \mathbb{R}$, which is invariant under permutations of electrons with the same spin, without affecting the total antisymmetry of the wavefunction. It is common to use a Jastrow-factor that does not alter the sign of Ψ, by enforcing $J > 0$ via $J = \exp(\hat{J})$ with an arbitrary permutation invariant function $\hat{J}$.

The Jastrow factor generally serves two purposes: Increasing expressivity of the wavefunction ansatz and enforcing the Kato cusp conditions (Kato, 1957). The first can be achieved by a permutation invariant pooling of the embeddings h_i, e.g. as

$$J = \exp\left(\sum_{i=1}^{n_\uparrow} \mathrm{MLP}^{\uparrow}(h_i) + \sum_{i=n_\uparrow+1}^{n_{\mathrm{el}}} \mathrm{MLP}^{\downarrow}(h_i)\right). \tag{124}$$

The latter refers to cusps that are present in the groundstate wavefunction whenever the positions of two particles coincide. The local energy $\frac{H\Psi}{\Psi}$ of the groundstate wavefunction is constant, but the individual terms in the Hamiltonian are not. In particular the potential energy terms in Eq. (3) diverge whenever the distance between two particles approaches zero. To obtain a constant local energy, the kinetic energy – given by the curvature of the wavefunction – must diverge with opposite sign, leading to discontinuous first derivatives of the wavefunction Ψ whenever two particles coincide. These cusps of high electron density (when an electron approaches a nucleus) or low electron density (when an electron approaches another electron) can be represented by an ansatz that has discontinuous derivatives at distance $|r_i - r_j| = 0$. A choice used by Hermann et al. (2020); von Glehn et al. (2023) is:

$$J^{\text{cusp}} = \exp\left(\sum_{i=1}^{n_{\text{el}}} \sum_{j=i+1}^{n_{\text{el}}} \frac{a}{b + |r_i - r_j|}\right) \tag{125}$$

with parameters a, b.

The full wavefunction Ψ is then given as

$$\Psi = J^{\text{cusp}} J \sum_{d=1}^{n_{\text{det}}} \det[\Phi_d] \tag{126}$$

4.7 Architectures for transferable wavefunctions

In many instances it is advantageous to have an ansatz which not only accurately represents $\Psi_{(\mathbf{R},\mathbf{Z})}(\mathbf{r})$ for a fixed geometry $(\mathbf{R}, \mathbf{Z})$, but which explicitly depends on the molecule and yields accurate groundstate wavefunctions across molecules. For example, when computing a Potential Energy Surface $\mathcal{E}_{\mathbf{Z}}(\mathbf{R})$, which requires finding the minimum eigenvalue for many instances of $\mathcal{H}^{\text{BO}}$, it can be more efficient to train a single transferable ansatz for all geometries, rather than training a separate ansatz for every geometry.

The architectures described in this section so far parameterize a wavefunction $\Psi_{(\mathbf{R},\mathbf{Z})}(\mathbf{r})$ which explicitly depends on the electron coordinates, but only implicitly depends on the molecular geometry contained in $(\mathbf{R}, \mathbf{Z})$. For example the nuclear coordinates $\mathbf{R}$ are used explicitly to compute input features in most architectures, but in many architectures the wavefunction Ψ does not explicitly depend on the nuclear charges $\mathbf{Z}$. Any realization of a wavefunction with parameters θ obtained through variational optimization will still implicitly depend on $\mathbf{Z}$ – because $\mathcal{H}^{\text{BO}}$ used throughout optimization depends on $\mathbf{Z}$ – but evaluating it for a molecule with different $\mathbf{Z}$, will not yield the correct groundstate wavefunction for this new molecule.

Several approaches for such transferable wavefunctions have been proposed by Gao and Günnemann (2022, 2023b,a); Scherbela et al. (2024, 2023), which make the following changes to the architecture outline in Sec. 4.2:

Input features and embedding

The input features explicitly depend on $(\mathbf{R}, \mathbf{Z})$. To encode the nuclear charges $\mathbf{Z}$, one-hot-encodings are typically used. The embeddings h_i no longer depend only on x_i, $\{p^{\uparrow}_{i,\setminus i}\}$, $\{p^{\downarrow}_{i,\setminus i}\}$, but also explicitly depend on the multiset of all nuclear positions and their nuclear charges:

$$h_i = h\left(x_i,\ \{p^{\uparrow}_{i,\setminus i}\},\ \{p^{\downarrow}_{i,\setminus i}\},\ \{(\mathbf{R}, \mathbf{Z})\}\right) \tag{127}$$

$$\{(\mathbf{R}, \mathbf{Z})\} := \{(R_1, Z_1),\ \ldots,\ (R_{n_{\text{nuc}}}, Z_{n_{\text{nuc}}})\} \tag{128}$$

This extra dependence is typically implemented in a similar fashion as the dependence on the set of other electrons, for example using graph convolutional networks or self-attention.

Orbitals

The widely used computation of orbitals in (119), (120) poses a challenge in designing transferable architectures, which generalize not only across different geometries $\mathbf{R}$, but also across molecules with different number of electrons. Because the dimensions of the trainable matrices $W^{\uparrow}, W^{\downarrow} \in \mathbb{R}^{n_{\text{el}} \times n_{\text{det}} \times d_{\text{emb}}}$ explicitly depend on n_{el} (corresponding to the number of orbitals in the Slater determinant), they cannot be transferred across different molecules with varying n_{el}. This is in contrast to other parts of the architecture (e.g. the GCN embedding) where the number of parameters does not depend on n_{el}, because all parameterized functions are only applied to objects corresponding to single electrons or pairs of electrons. The computational cost grows with system size – because the functions must be evaluated for more electrons / pairs of electrons – but the number of parameters is independent of n_{el}.

A solution proposed by Gao and Günnemann (2023a); Scherbela et al. (2024) is to compute W_k as a function of some features $c_k \in \mathbb{R}^{n^{\text{orb}}_{\text{feat}}}$ of each orbital k. The features c_k in turn are evaluated using a conventional method that yields orbitals for a given molecule $(\mathbf{R}, \mathbf{Z})$. Gao and Günnemann (2023a) propose a heuristic based on chemical bonds to obtain orbital positions and ultimately features c_k. Scherbela et al. (2024) propose to use the Hartree-Fock orbitals – which are typically already computed for the purpose of supervised pretraining (cf. Eq. (88)) – to provide atom-wise orbital features c_{kJ}. Given Hartree-Fock orbitals $\varphi^{\text{HF}}_k(r_i)$, expanded in atom-centered basis functions $b_\mu : \mathbb{R}^3 \to \mathbb{R}$

$$\varphi^{\text{HF}}_k(r_i) = \sum_{J=1}^{n_{\text{nuc}}} \sum_{\mu=1}^{n_{\text{basis}}} c_{kJ\mu}\, b_\mu(r_i - R_J), \tag{129}$$

with corresponding expansion coefficients $c \in \mathbb{R}^{n_{\text{el}} \times n_{\text{nuc}} \times n_{\text{basis}}}$, these expansion coefficients can be used as orbital features with $n^{\text{orb}}_{\text{feat}} = n_{\text{basis}}$. To this end the expansion coefficients are first mapped to $\hat{W}_{dkJ} \in \mathbb{R}^{d_{\text{emb}}}$ and $\hat{a}_{dkJ} \in \mathbb{R}$ using MLPs

and then used to evaluate the orbitals analogous to (119) and (122). Omitting spins for clarity, this leads to

$$\hat{W}_{dkJ} = \mathrm{MLP}_{\mathrm{W}}(c_{kJ}) \qquad \mathrm{MLP}_{\mathrm{W}} : \mathbb{R}^{n_{\mathrm{basis}}} \to \mathbb{R}^{d_{\mathrm{emb}}} \tag{130}$$

$$\hat{a}_{dkJ} = \mathrm{MLP}_{\mathrm{a}}(c_{kJ}) \qquad \mathrm{MLP}_{\mathrm{a}} : \mathbb{R}^{n_{\mathrm{basis}}} \to \mathbb{R} \tag{131}$$

$$\phi^{d,k,i} = \sum_{J=1}^{n_{\mathrm{nuc}}} \left(\hat{W}_{dkJ} \cdot h_i \right) e^{-\hat{a}_{dkJ}|r_i - R_j|}, \tag{132}$$

which no longer requires a number of trainable parameters dependent on n_{el}.

5 Results

In this section we survey numerical results that have been achieved using DL-VMC.

5.1 Highly accurate variational energies

In general architectures like FermiNet introduced in Sec. 4.2 are capable of accurately representing the ground-state wavefunction and allow for highly accurate ground-state energy predictions. For example, Fig. 8 compares a neural network architecture proposed in 2022 against other computational approaches for a set of molecules up to 42 electrons (Gerard et al., 2022). The work finds that across a range of molecules, Deep Learning-based Variational Monte Carlo is able to reach lower energy predictions than conventional approaches and,

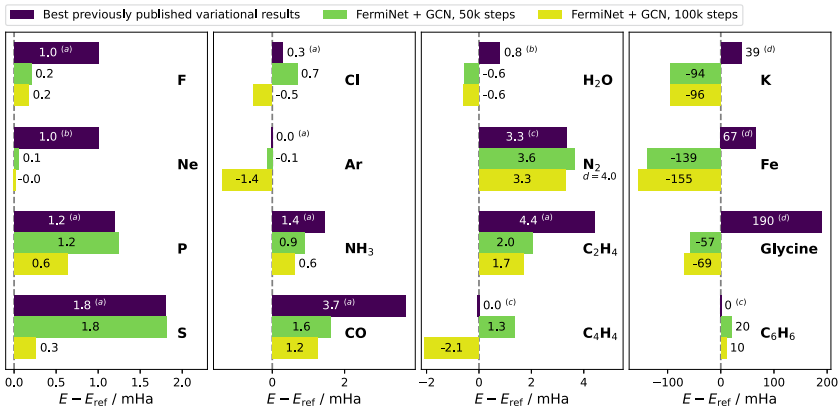

FIGURE 8 Energies relative to the previously known best estimate, (lower is better). Blue bars depict best published variational energies, footnotes mark the method: a: FermiNet VMC (Pfau et al., 2020; Spencer et al., 2020), b: Conventional DMC (Seth et al., 2011; Nemec et al., 2010; Clark et al., 2011), c: FermiNet DMC (Ren et al., 2023), d: MRCI-F12. Note that E_{ref} is not necessary variational and thus may underestimate the true energy. Figure and caption is taken from Gerard et al. (2022).

therefore, better estimates for the ground state, due to the variational principle. The figure distinguishes between the best estimate and other variational energies, whereas the best estimate can include methods such as CCSD(T), which is widely considered the gold standard for highly accurate ground-state energy predictions. However, a notable drawback of CCSD(T) lies in its nonvariational nature, potentially leading to an underestimation of the ground-state energy and offering no uncertainty guarantees. In contrast, DL-VMC exceeds all conventional variational approaches, including Diffusion Monte Carlo, except for the case of Benzene, the largest system tested, where another deep learning-based approach outperforms the proposed architecture. Since the architecture's initial publication, further improvements have been made in incorporating interparticle correlation, enhancing the presented results even further (Li et al., 2023; von Glehn et al., 2023; Gao and Günnemann, 2023a). These advancements underscore the potential of Variational Monte Carlo and its capabilities in the field.

5.2 Transfer learning for ground-state energy predictions

A potential direction to improve the method's efficiency and reduce training cost, is to use techniques such as deep transfer learning (Devlin et al., 2018; Alayrac et al., 2022). The idea is to pretrain a neural network ansatz on a specific set of molecules to predict the ground-state energy and subsequently transfer this pretrained model to novel, previously unencountered molecules. As discussed in the preceding Sec. 4.7, a common challenge with the proposed architectures lies in the inherent dependence of the ansatz's parameter count on the system size due to the unique construction of the orbital matrix. To address this issue, Scherbela et al. (2023) proposed an approach to map computationally cheap orbital descriptors from methods such as Hartree Fock to highly accurate deep-learning-based orbitals.

To assess the transfer capabilities of a pretrained model, Scherbela et al. pretrained a single neural network ansatz on a diverse set of molecules, comprising approximately 100 molecules, each containing up to four heavy atoms (counted as the number of nonhydrogen atoms). In Fig. 9, the model was evaluated on test sets, each containing four randomly perturbed molecules, grouped by the number of nonhydrogen atoms with up to 7 heavy atoms. To prevent any potential train/test leakage, none of the molecules in the test set were included in the training set. Utilizing a pretrained model, the authors achieved CCSD(T) accuracy with a 2Z basis set without the need for additional variational optimization steps (zero-shot) for molecules containing up to 6 heavy atoms. Additional optimization steps further improved accuracy (refer to Fig. 9 b). To benchmark the performance against other state-of-the-art Deep Learning-based Variational Monte Carlo methods, the energy error as a function of optimization steps, was compared for molecules with three heavy atoms. Specifically, an attention-based approach (von Glehn et al., 2023) without pretraining was used as a baseline. In summary, a pretrained model can generally achieve certain levels of accuracy orders of magnitude faster but may be outperformed after more prolonged

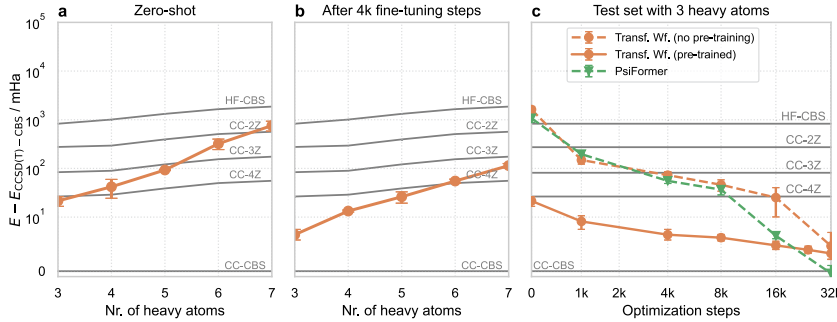

FIGURE 9 **Absolute energies**: Energies relative to CCSD(T)-CBS (complete basis set limit) when re-using the pretrained model on molecules of varying size without optimization (a) and after fine-tuning (b). (c) depicts energy for the test set containing 3 heavy atoms as a function of optimization steps and compares against SOTA method. Solid lines are with pretraining, dashed lines without. Gray lines correspond to conventional methods: Hartree-Fock in the complete basis set limit (HF-CBS), and CCSD(T) with correlation consistent basis sets of double to quadruple valence (CC-nZ). Figure and caption is taken from Scherbela et al. (2023).

optimization. This phenomenon could be related to the orbital construction, potentially suggesting that the proposed transferable ansatz might be inherently less expressive (Scherbela et al., 2023).

The more challenging task of relative energies was evaluated in a second experiment. By relative energies we denote the energy difference between different geometrical conformations of the same molecule. The authors show that by using only a few additional variational optimization steps qualitatively and quantitatively correct results can be achieved.

For instance, in Fig. 10a, the relative energy of five conformers to the equilibrium-state geometry of Bicyclobutane is illustrated. This system is of particular interest because CCSD(T) tends to inaccurately predict the relative energy for the "dis-TS" conformer, significantly underestimating the energy by approximately 60 mHa. While the pretrained deep learning model correctly predicts the sign of the relative energies without the need for additional optimization steps, it does yield quantitatively different relative energies. Therefore, additional optimization steps are necessary to bring the model into closer alignment with the Diffusion Monte Carlo reference method and FermiNet. With just 700 optimization steps per geometry, the pretrained model achieved a maximum deviation of 2.1 millihartree (mHa) to the reference method compared to FermiNet, which required 10,000 steps for a maximum deviation of 7.1 mHa.

Another frequently encountered test case involves the dissociation curve of the Nitrogen dimer. The dimer serves as another example wherein methods like CCSD(T) commonly struggle by notably overestimating the energy for conformations in the bond-breaking regime (i.e. a distance of 3-4 Bohr between the Nitrogen atoms). In this scenario, the pretrained model also tended to overesti-

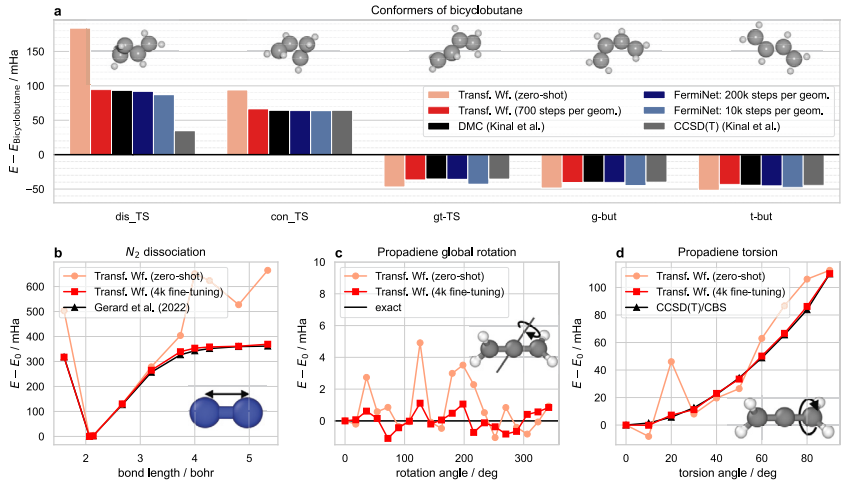

FIGURE 10 Challenging relative energies: Relative energies obtained with and without fine-tuning on 4 distinct, challenging systems, compared against high-accuracy reference methods. a) Relative energy of bicyclobutane conformers vs. the energy of bicyclobutane; b) Potential energy surface (PES) of N_2; c) global rotation of propadiene; d) relative energy of twisted vs. untwisted propadiene. Figure and caption is taken from Scherbela et al. (2023).

mate the energy significantly. However, it only required a few optimization steps to better align with the reference method (cf. Fig. 9 c).

The ground-state energy in general is invariant under global rotation of the molecule. Although the energy is invariant, this is in general not the case for the wavefunction. Enforcing complete invariance of the wavefunction under global rotation would be overly restrictive, as discussed in the work by Gao and Günnemann (2022). To address this, Scherbela et al. incorporated data augmentation during pretraining, involving random rotations of the entire molecules. While this proved to be a suitable proxy, with energy variations of approximately 1 mHa for propadiene, it was not entirely sufficient, leading to outliers with variations reaching up to 5 mHa. A short variational optimization phase once again helped to mitigate larger energy errors, achieving chemical accuracy as depicted in Fig. 10c.

For the final evaluation, the transition barrier of propadiene twisted around the C=C bond was examined. The equilibrium and transition states are differentiated by an energy difference of approximately 110 mHa. The model was intentionally pretrained on twisted molecules, encompassing equilibrium conformations, transition conformations, and one intermediate twist. However, as depicted in Fig. 10, this proves insufficient to accurately predict the complete transition path without the incorporation of additional optimization steps. However, again only a short amount of fine-tuning allowed for an accurate prediction of the whole path.

5.3 Literature overview

Research into Deep Learning-based Variational Monte Carlo has expanded rapidly over the last years, rendering a comprehensive overview of the field impossible. For further reviews on this subject we refer the reader to Hermann et al. (2022); Zhang et al. (2023); Schätzle et al. (2023); Medvidović and Moreno (2024) and highlight several advances in key areas below.

Embedding architectures

FermiNet (Pfau et al., 2020; Spencer et al., 2020) and PauliNet (Hermann et al., 2020) have been the first two neural network architectures to successfully demonstrate the potential of neural network-based wavefunctions for molecules in first quantization. As discussed in Sec. 4.2, a significant portion of research has been invested in improving the originally proposed embedding architectures by incorporating attention-based techniques (von Glehn et al., 2023; Pescia et al., 2023; Li et al., 2023) and graph-based approaches (Gerard et al., 2022; Gao and Günnemann, 2023a). Currently, the two state-of-the-art architectures, LapNet (Li et al., 2023) and PsiFormer (von Glehn et al., 2023), in terms of accuracy are based on attention mechanisms.

Antisymmetrization

In terms of antisymmetrization, as previously discussed, a common technique is to employ Slater determinants. From a scaling perspective, the determinant is the leading factor with cubical scaling. Therefore, efforts have been made to reduce the theoretical scaling using sorting algorithms (Richter-Powell et al., 2023) or products of two-particle functions (Han et al., 2019). Although the results are partially promising, they are still at a proof-of-concept stage, and the Slater determinant remains the most commonly used antisymmetrization method. However, for certain systems, such as for the case of superfluids, it was shown that approaches like antisymmetric geminal powered wavefunctions (Lou et al., 2024) or Pfaffian wavefunctions (Kim et al., 2023) can be beneficial. A recent preprint (Ye et al., 2024) proposed antisymmetrization using Vandermonde-like determinants, and showed that any continuous antisymmetric function can be represented by $\mathcal{O}(n_{\text{el}})$ of these objects, potentially yielding another approach to antisymmetrization with favorable scaling of computational cost.

Generalization across molecules

By taking advantage of regularities within the geometrical conformation space of molecules, several approaches have been proposed to learn neural network-based wavefunctions simultaneously across a range of geometrical conformations. Either by only sharing parts of the neural network architecture (Scherbela et al., 2022) or by using a meta neural network to predict the linear mappings within the orbital construction (Gao and Günnemann, 2022, 2023b). Although

they achieve faster evaluation of the potential energy surface of a molecule by an order of magnitude, they don't allow the efficient transfer to new molecules. A key reason for this is the explicit dependence of the parameter count of the architecture on the system size due to the unique construction of the orbital matrix. Therefore, Gao and Günnemann (2023a) and Scherbela et al. (2024, 2023) generalized the existing approach to allow for efficient optimization of a single neural network across a diverse range of molecules. This again allowed for a significant reduction in optimization steps, as discussed in Sec. 5.2.

Optimization

Another active area of research involves enhancing neural network optimization techniques. In Deep Learning-based Variational Monte Carlo, it is common to employ second-order methods like Natural Gradient Descent. Consequently, there arises the need to invert a preconditioner matrix with dimensions $n_{\text{params}} \times n_{\text{params}}$. FermiNet has proposed the use of KFAC as an approximation technique, which relies on the assumption that the matrix is block-diagonal. Another very recent line of work showed that it is possible to convert the problem of inverting the matrix of shape $n_{\text{params}} \times n_{\text{params}}$ to a problem of inverting a matrix of shape $n_{\text{samples}} \times n_{\text{samples}}$, whereas $n_{\text{samples}} \ll n_{\text{params}}$ represent the number of Monte Carlo samples (Goldshlager et al., 2024; Rende et al., 2023). On the other hand, Neklyudov et al. (2023) interpreted the optimization in Variational Monte Carlo as a gradient flow and by improving the underlying metric of the distribution space the author reached empirically faster convergence.

Observables and applications

Besides the improvements to the neural network-based architecture and the computation of the ground-state energy for molecules in first quantization, the method was also applied to a plethora of other observables and systems. For example the method has been applied to the computation of forces (Qian et al., 2022; Scherbela et al., 2022) and excited states (Entwistle et al., 2023; Pfau et al., 2023), as well as to other systems such as solids including real-solids and the homogeneous electron gas (Pescia et al., 2023; Cassella et al., 2023; Li et al., 2022a), superfluids (Lou et al., 2024; Kim et al., 2023), positron-molecule complexes (Cassella et al., 2024) and discrete systems (Carleo and Troyer, 2017). Also, techniques like effective core potentials (Li et al., 2022b) or Diffusion Monte Carlo (Ren et al., 2023) have been explored in the context of Deep Learning-based methods for the electronic Schrödinger equation to improve the scaling and accuracy further.

References

Agmon, Shmuel, 2014. Lectures on Exponential Decay of Solutions of Second-Order Elliptic Equations: Bounds on Eigenfunctions of N-Body Schrodinger Operations. (MN-29), vol. 29. Princeton University Press.

Alayrac, Jean-Baptiste, et al., 2022. Flamingo: a visual language model for few-shot learning. In: Advances in Neural Information Processing Systems, vol. 35. Curran Associates, Inc., pp. 23716–23736.

Barone, Vincenzo, et al., 2013. Accurate structure, thermodynamic and spectroscopic parameters from CC and CC/DFT schemes: the challenge of the conformational equilibrium in glycine. Physical Chemistry Chemical Physics 15 (25), 10094–10111. https://doi.org/10.1039/C3CP50439E.

Barrett, Thomas D., Malyshev, Aleksei, Lvovsky, A.I., 2022. Autoregressive neural-network wavefunctions for ab initio quantum chemistry. Nature Machine Intelligence 4 (4), 351–358. https://doi.org/10.1038/s42256-022-00461-z.

Bayer, Christian, et al., 2023. Pricing high-dimensional Bermudan options with hierarchical tensor formats. SIAM Journal on Financial Mathematics 14 (2), 383–406.

Becca, Federico, Sorella, Sandro, 2017. Quantum Monte Carlo Approaches for Correlated Systems. Cambridge University Press.

Berner, Julius, Grohs, Philipp, Voigtlaender, Felix, 2023. Training ReLU networks to high uniform accuracy is intractable. In: Proceedings ICLR 2023.

Carleo, Giuseppe, Troyer, Matthias, 2017. Solving the quantum many-body problem with artificial neural networks. Science 355 (6325), 602–606. https://doi.org/10.1126/science.aag2302.

Cassella, G., et al., 2024. Neural network variational Monte Carlo for positronic chemistry. arXiv preprint. arXiv:2310.05607.

Cassella, Gino, et al., 2023. Discovering quantum phase transitions with fermionic neural networks. Physical Review Letters 130 (3), 036401. https://doi.org/10.1103/PhysRevLett.130.036401.

Chuang, Pi-Yueh, Barba, Lorena A., 2022. Experience report of physics-informed neural networks in fluid simulations: pitfalls and frustration. arXiv preprint. arXiv:2205.14249.

Clark, Bryan K., et al., 2011. Computing the energy of a water molecule using multideterminants: a simple, efficient algorithm. Journal of Chemical Physics 135 (24), 244105. https://doi.org/10.1063/1.3665391.

De Ryck, Tim, Mishra, Siddhartha, 2024. Numerical analysis of physics-informed neural networks and related models in physics-informed machine learning. arXiv preprint. arXiv:2402.10926.

Devlin, Jacob, et al., 2018. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805.

Dirac, Paul Adrien Maurice, 1929. Quantum mechanics of many-electron systems. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 123 (792), 714–733.

Elbrächter, Dennis, et al., 2021. Deep neural network approximation theory. IEEE Transactions on Information Theory 67 (5), 2581–2623.

Entwistle, Mike.T., et al., 2023. Electronic excited states in deep variational Monte Carlo. Nature Communications 14 (1), 274. https://doi.org/10.1038/s41467-022-35534-5.

Evans, Lawrence C., 2022. Partial Differential Equations, vol. 19. American Mathematical Society.

Gao, Nicholas, Günnemann, Stephan, 2022. Ab-initio potential energy surfaces by pairing GNNs with neural wave functions. In: International Conference on Learning Representations.

Gao, Nicholas, Günnemann, Stephan, 2023a. Generalizing neural wave functions. In: Proceedings of the 40th International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 202. PMLR, pp. 10708–10726.

Gao, Nicholas, Günnemann, Stephan, 2023b. Sampling-free inference for ab-initio potential energy surface networks. In: The Eleventh International Conference on Learning Representations.

Gerard, Leon, et al., 2022. Gold-standard solutions to the Schrödinger equation using deep learning: how much physics do we need? In: Advances in Neural Information Processing Systems, vol. 35. Curran Associates, Inc., pp. 10282–10294.

Goldshlager, Gil, Abrahamsen, Nilin, Lin, Lin, 2024. A Kaczmarz-inspired approach to accelerate the optimization of neural network wavefunctions. arXiv preprint. arXiv:2401.10190.

Hall, Brian C., 2013. Quantum Theory for Mathematicians. Graduate Texts in Mathematics, vol. 136.

Han, Jiequn, Zhang, Linfeng, E, Weinan, 2019. Solving many-electron Schrödinger equation using deep neural networks. Journal of Computational Physics 399, 108929. https://doi.org/10.1016/j.jcp.2019.108929.

Hermann, Jan, et al., 2022. Ab-initio quantum chemistry with neural-network wavefunctions. arXiv preprint. arXiv:2208.12590.

Hermann, Jan, Schätzle, Zeno, Noé, Frank, 2020. Deep-neural-network solution of the electronic Schrödinger equation. Nature Chemistry 12 (10), 891–897. https://doi.org/10.1038/s41557-020-0544-y.

Hibat-Allah, Mohamed, et al., 2020. Recurrent neural network wave functions. Physical Review Research 2 (2), 023358. https://doi.org/10.1103/PhysRevResearch.2.023358.

Hunziker, Walter, Sigal, Israel Michael, 2000. The quantum N-body problem. Journal of Mathematical Physics 41 (6), 3448–3510.

Inui, Koji, Kato, Yasuyuki, Motome, Yukitoshi, 2021. Determinant-free fermionic wave function using feed-forward neural networks. Physical Review Research 3 (4), 043126. https://doi.org/10.1103/PhysRevResearch.3.043126.

Jecko, Thierry, 2014. On the mathematical treatment of the Born-Oppenheimer approximation. Journal of Mathematical Physics 55, 5.

Kato, Tosio, 1957. On the eigenfunctions of many-particle systems in quantum mechanics. Communications on Pure and Applied Mathematics 10 (2), 151–177.

Kato, Tosio, 2013. Perturbation Theory for Linear Operators, vol. 132. Springer Science & Business Media.

Kim, Jane, et al., 2023. Neural-network quantum states for ultra-cold Fermi gases. arXiv preprint. arXiv:2305.08831.

Kingma, Diederik P., Ba, Jimmy, 2017. Adam: a method for stochastic optimization. arXiv:1412.6980.

Li, Ruichen, et al., 2023. Forward Laplacian: a new computational framework for neural network-based variational Monte Carlo. arXiv:2307.08214.

Li, Xiang, et al., 2022b. Fermionic neural network with effective core potential. Physical Review Research 4, 013021. https://doi.org/10.1103/PhysRevResearch.4.013021.

Li, Xiang, Li, Zhe, Chen, Ji, 2022a. Ab initio calculation of real solids via neural network ansatz. Nature Communications 13 (1), 7895. https://doi.org/10.1038/s41467-022-35627-1.

Lin, Jeffmin, Goldshlager, Gil, Lin, Lin, 2023. Explicitly antisymmetrized neural network layers for variational Monte Carlo simulation. Journal of Computational Physics 474, 111765.

Lou, Wan Tong, et al., 2024. Neural wave functions for superfluids. arXiv preprint. arXiv:2305.06989.

Martens, James, Grosse, Roger, 2015. Optimizing neural networks with Kronecker-factored approximate curvature. arXiv:1503.05671.

Medvidović, Matija, Moreno, Javier Robledo, 2024. Neural-network quantum states for many-body physics. arXiv preprint. arXiv:2402.11014.

Neklyudov, Kirill, et al., 2023. Wasserstein quantum Monte Carlo: a novel approach for solving the quantum many-body Schrödinger equation. arXiv preprint. arXiv:2307.07050.

Nemec, Norbert, Towler, Michael D., Needs, R.J., 2010. Benchmark all-electron ab initio quantum Monte Carlo calculations for small molecules. Journal of Chemical Physics 132 (3), 034111. https://doi.org/10.1063/1.3288054.

Pescia, Gabriel, et al., 2023. Message-passing neural quantum states for the homogeneous electron gas. arXiv:2305.07240.

Pfau, David, et al., 2020. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. Physical Review Research 2 (3), 033429. https://doi.org/10.1103/PhysRevResearch.2.033429.

Pfau, David, et al., 2023. Natural quantum Monte Carlo computation of excited states. arXiv:2308.16848.

Qian, Yubing, et al., 2022. Interatomic force from neural network based variational quantum Monte Carlo. Journal of Chemical Physics 157 (16), 164104. https://doi.org/10.1063/5.0112344.

Radford, Alec, et al., 2021. Improving Language Understanding by Generative Pre-Training.

Reed, M., Simon, B., 1978. Analysis of Operators. Methods of Modern Mathematical Physics, vol. 4. Academic, New York.

Ren, Weiluo, et al., 2023. Towards the ground state of molecules via diffusion Monte Carlo on neural networks. Nature Communications 14 (1), 1860. https://doi.org/10.1038/s41467-023-37609-3.

Rende, Riccardo, et al., 2023. A simple linear algebra identity to optimize large-scale neural network quantum states. arXiv preprint. arXiv:2310.05715.

Richter-Powell, Jack, et al., 2023. Sorting out quantum Monte Carlo. arXiv preprint. arXiv:2311.05598.

Schätzle, Zeno, et al., 2023. DeepQMC: an open-source software suite for variational optimization of deep-learning molecular wave functions. Journal of Chemical Physics 159 (9), 094108. https://doi.org/10.1063/5.0157512.

Scherbela, Michael, et al., 2022. Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks. Nature Computational Science 2 (5), 331–341. https://doi.org/10.1038/s43588-022-00228-x.

Scherbela, Michael, Gerard, Leon, Grohs, Philipp, 2023. Variational Monte Carlo on a budget - fine-tuning pre-trained neural wavefunctions. In: Thirty-Seventh Conference on Neural Information Processing Systems.

Scherbela, Michael, Gerard, Leon, Grohs, Philipp, 2024. Towards a transferable fermionic neural wavefunction for molecules. Nature Communications 15 (1), 120. https://doi.org/10.1038/s41467-023-44216-9.

Schütt, Kristof T., et al., 2020. Machine Learning Meets Quantum Physics. Lecture Notes in Physics.

Seth, P., López Ríos, P., Needs, R.J., 2011. Quantum Monte Carlo study of the first-row atoms and ions. Journal of Chemical Physics 134 (8), 084105. https://doi.org/10.1063/1.3554625.

Spencer, James S., et al., 2020. Better, faster fermionic neural networks. arXiv:2011.07125.

Stone, Marshall H., 1932. On one-parameter unitary groups in Hilbert space. Annals of Mathematics, 643–648.

Szabo, Attila, Ostlund, Neil S., 1996. Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. Dover Publications. ISBN 9780486691862.

Thiede, Luca, Sun, Chong, Aspuru-Guzik, Alán, 2023. Waveflow: enforcing boundary conditions in smooth normalizing flows with application to fermionic wave functions. arXiv preprint. arXiv:2211.14839.

Troyer, Matthias, Wiese, Uwe-Jens, 2005. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. Physical Review Letters 94 (17), 170201.

von Glehn, Ingrid, Spencer, James S., Pfau, David, 2023. A self-attention ansatz for ab-initio quantum chemistry. In: The Eleventh International Conference on Learning Representations.

von Neumann, John, 1932. Über einen Satz von Herrn MH Stone. Annals of Mathematics 33 (3), 567–573.

Woodbury, Max A., 1950. Inverting Modified Matrices. Department of Statistics, Princeton University.

Ye, Haotian, Li, Ruichen, Gu, Yuntian, Lu, Yiping, He, Di, Wang, Liwei, 2024. $\widetilde{O}(N^2)$ representation of general continuous anti-symmetric function. https://doi.org/10.48550/arXiv.2402.15167. arXiv:2402.15167 [quant-ph]. http://arxiv.org/abs/2402.15167 (visited on 03/01/2024).

Yserentant, Harry, 2010. Regularity and Approximability of Electronic Wave Functions. Springer.

Zhang, Xuan, et al., 2023. Artificial intelligence for science in quantum, atomistic, and continuum systems. arXiv:2307.08423.

Zhou, Jie, et al., 2020. Graph neural networks: a review of methods and applications. AI Open (ISSN 2666-6510) 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001. https://www.sciencedirect.com/science/article/pii/S2666651021000012 (visited on 02/13/2024).