Teaching Quantum Chemistry to a Deep Learning Model

Transformers for the many body Schrodinger Equation

Jorge Munoz Laredo, Angel Flores, Daniel Paredes

January 5, 2026

Outline

The Schrödinger equation

On 1926 Schrodinger derived his equatin:

$$\hat{H}\,\Psi = E\Psi \tag{1}$$

- Ψ is a complex value function called **wave function**.
- $\hat{H}$ is called the **Hamiltonian Operator**.

Hamiltonian

In the position basis:

$$\hat{H} = \frac{\hat{\vec{P}}^2}{2m} + \hat{V} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V} \tag{2}$$

- Find the electrostatic potential V of the system.

Many-Body System

When considering n bodies, we have:

$$\hat{H}\psi(\mathbf{x}_1, \ldots, \mathbf{x}_n) = E\psi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \tag{3}$$

With $\mathbf{x}_i = \{\mathbf{r}_i, \sigma\}$, where $\mathbf{r}_i \in \mathbb{R}^3$ is the position of each particle and $\sigma \in \{\uparrow . \downarrow\}$ is the so called spin.

Considerations

- Each particle interact with all the another particles.
- For atoms, consider all the protons, electrons and neutrons.
- Solution obey physical laws.

Setting up the Hamiltonian

The first step is obtain a practical form of the **Hamiltonian**.

- Kinetic energy: $T = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2$.
- Electron-electron repulsion: $V_{ee} = \sum_{i<j} \frac{1}{r_{ij}}$.

$$\hat{H} = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N} \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{1\leq i<j\leq N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{4}$$

Fermi-Dirac statistics

All the fermions follow the Fermi-Dirac Statistics, this is.

- Exchanging two electrons flips the wavefunction's sign: $\Psi(\dots i, j \dots) = -\Psi(\dots j, i \dots)$.

Slater Determinant

Enforce it using a determinant.

$$\psi = \begin{vmatrix} \phi_1^k(\mathbf{x}_1) & \dots & \phi_1^k(\mathbf{x}_n) \\ \vdots & & \vdots \\ \phi_n^k(\mathbf{x}_1) & \dots & \phi_n^k(\mathbf{x}_n) \end{vmatrix} \tag{5}$$

Where ϕ are called spin orbitals

Kato cusp conditions, Jastrow Factor

The potential are:

$$\sum \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

- Coulomb potentials cause a sharp cusp in Ψ when particles overlaps.

Jastrow Factor $\exp(\mathcal{J})$

In this work we are going to use this specific form:

$$\mathcal{J}_\theta(x) = \sum_{i<j;\sigma_i=\sigma_j} -\frac{1}{4}\frac{\alpha_{par}^2}{\alpha_{par} + |\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i,j;\sigma_i\neq\sigma_j} -\frac{1}{2}\frac{\alpha_{anti}^2}{\alpha_{anti} + |\mathbf{r}_i - \mathbf{r}_j|} \quad (6)$$

Loss: Variational Principle

Variational principle states:

$$E[\Psi] = \frac{\langle \Psi \mid H \mid \Psi \rangle}{\langle \Psi \mid \Psi \rangle} \geq E_0$$

Minimizing $E[\Psi]$ drives the ansatz toward the ground state.

$$E[\Psi] = \mathcal{L}(\Psi_\theta) = \frac{\langle \Psi_\theta | \, \hat{H} \, | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\int d\mathbf{r} \Psi^*(\mathbf{r}) \hat{H} \Psi(\mathbf{r})}{\int d\mathbf{r} \Psi^*(\mathbf{r}) \Psi(\mathbf{r})}$$

Define:

$$p_\theta(x) = \Psi_\theta^2(x) \frac{1}{\int dx' \Psi_\theta^2(x')} \wedge E_L(x) = \Psi_\theta^{-1}(x) \hat{H} \Psi_\theta(x)$$

Then:

$$\mathcal{L}_\theta = \mathbb{E}_{x \sim p_\theta}[E_L(x)] \tag{7}$$

Quantum Monte Carlo

Variational Monte Carlo

With the samples $\mathbf{R}_1, \ldots, \mathbf{R}_M \sim p_\theta(\mathbf{R})$ we can make:

$$\mathcal{L}_\theta = \mathbb{E}_{x \sim p_\theta}[E_L(x)] \approx \frac{1}{M} \sum_{i=1}^{m} E_L(\mathbf{R}_k) \tag{8}$$

With:

$$E_L(\mathbf{R}_k) = \frac{\hat{H}\psi(\mathbf{R}_k)}{\psi(\mathbf{R}_k)} = -\frac{1}{2}\frac{\nabla^2\psi(\mathbf{R_k})}{\psi(\mathbf{R}_k)} + V(\mathbf{R}_k)$$

Obtain $\mathbf{R}_k \rightarrow$ **Metropolis-Hastings Algorithm**.

Metropolis-Hastings Algorithm

Goal: Obtain ρ, Requirement: $C\rho$ 1. Take a initial configuration $\mathbf{X}_0 \in E$ arbitrary:
2. Propose $\mathbf{X}' = \mathbf{X}_0 + \eta$,where $\eta \sim q(\eta)$, q is a probability density on E called proposal kernel.(Normal Gaussian)
3. Compute the quantity:

$$A(\mathbf{X_0}, \mathbf{X}') = \min\left(1, \frac{\rho(\mathbf{X}')}{\rho(\mathbf{X}_0)} \frac{q(\mathbf{X}' - \mathbf{X}_0)}{q(\mathbf{X}_0 - \mathbf{X}')}\right)$$

Where ρ is the target distribution where we want sample, In the case where q is symmetric, this simplifies to:

$$A(\mathbf{X}_0, \mathbf{X}') = \min\left(1, \frac{\rho(\mathbf{X}^{'})}{\rho(\mathbf{X}_0)}\right)$$

4. Generate a uniform number $U \in [0, 1]$. If: $U < A(\mathbf{X}_0 \to \mathbf{X}'_l)$ then $\mathbf{X_1} = \mathbf{X}'$, otherwise try another $\mathbf{X}'$. Accept or decline.

Back propagation pitfall

Gradients of the Loss

Using calculus you obtain:

$$\nabla_\theta \mathcal{L} = 2\mathbb{E}_{x\sim\Psi^2}[(E_L(x) - \mathbf{E}_p(E_L))\log\psi] \tag{9}$$

This expectation is calculated in the same way.

Outline

Attention on the room

Multi Head Attention Let d be the embedding dimension, n_h be the number of attention heads, d_h be the dimension per head, and $\mathbf{h}_t \in \mathbb{R}^d$ be hidden dimension. The learn matrices are:

$$W^Q, W^K, W^V \in \mathbb{R}^{d_h n_h \times d}$$

We obtain the key, queries, and values vectors with:

$$\mathbf{k}_i = \mathbf{W}^k \mathbf{h}_i, \mathbf{q}_i = \mathbf{W}^q \mathbf{h}_i, \mathbf{v}_i = \mathbf{W}^v \mathbf{h}_i$$

We slice this vectors:

$$[\mathbf{q_1}, \mathbf{q_2}, \ldots, \mathbf{q_{n_h}}] = \mathbf{q}$$
$$[\mathbf{k_1}, \mathbf{k_2}, \ldots, \mathbf{k_{n_h}}] = \mathbf{k}$$
$$[\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_{n_h}}] = \mathbf{v}$$

In the $i - th$ head:

$$\mathbf{o}_{t,i} = \sum_{j=1}^{t} \text{Softmax}\left(\frac{\mathbf{q}_{t,i}^T \mathbf{k}_{j,i}}{\sqrt{d_h}}\right) \mathbf{v}_{j,i} \tag{10}$$

Attention!

Let $W^O \in \mathbb{R}^{d \times d_h n_h}$ be the output projection matrix.

$$\mathbf{u}_t = W^O[\mathbf{o}_{t,1}; \mathbf{o}_{t,2}; \ldots; \mathbf{o}_{t,n_h}]$$

Outline

FermiNet baseline

Orbital

To compute the orbitals we need to compute.

$$\phi_i^{k\alpha}(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = \left(\mathbf{w}_i^{k\alpha} \cdot \mathbf{h}_j^{L\alpha} + g_i^{k\alpha}\right) \sum_m \pi_{im}^{k\alpha} \exp\left(-|\boldsymbol{\Sigma}_{im}^{k\alpha}(\mathbf{r}_j^\alpha - \mathbf{R}_m)|\right) \tag{11}$$

Where the hidden states are:

$$\mathbf{h}_i^{\ell+1\alpha} = \tanh\left(\mathbf{V}^\ell \mathbf{f}_i^{\ell\alpha} + \mathbf{b}^\ell\right) + \mathbf{h}_i^{\ell\alpha}$$
$$\mathbf{h}_{ij}^{\ell+1\alpha\beta} = \tanh\left(\mathbf{W}^\ell \mathbf{h}_{ij}^{\ell\alpha\beta} + \mathbf{c}^\ell\right) + \mathbf{h}_{ij}^{\ell\alpha\beta}$$

Finally:

$$\psi(\mathbf{r}_1^\uparrow, \ldots, \mathbf{r}_{n^\downarrow}^\downarrow) = \sum_k \omega_k \Big(\det\left[\phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}; \{\mathbf{r}^\downarrow\})\right] \det\left[\phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}^\downarrow\}); \{\mathbf{r}^\uparrow\};\right]\Big). \tag{12}$$

Fermi Net Architecture

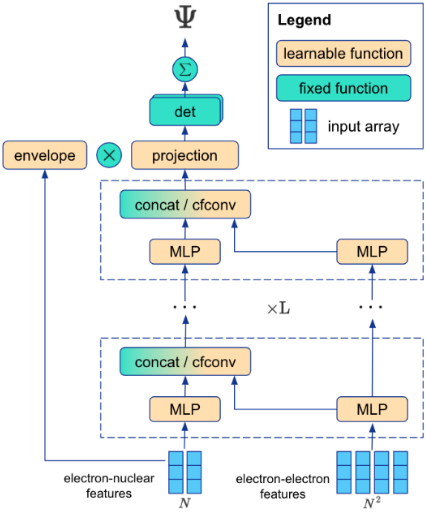

Figure 1: Fermi Net Architecture

The Psiformer Ansatz (overview)

The output of **Psiformer** is:

$$\Psi_\theta(\mathbf{x}) = \exp(\mathcal{J}_\theta(\mathbf{x})) \sum_{k=1}^{N_{\text{det}}} \det\left[\mathbf{\Phi}_\theta^k(x)\right] \tag{13}$$

We obtain the hidden states:

$$A_h^l = [\text{SelfAttn}(\mathbf{h}_1^l, \ldots, \mathbf{h}_N^\ell; \mathbf{W}_q^{\ell h}, \mathbf{W}_k^{\ell h}, \mathbf{W}_v^{\ell h})]$$

$$A^\ell = \text{concat}_h[A_h^\ell]$$

And then used it apply:

$$\mathbf{f}_i^{\ell+1} = \mathbf{h}_i^\ell + W_o^\ell A^\ell$$

And generate a new hidden state:

$$\mathbf{h}_i^{\ell+1} = \mathbf{f}_i^{\ell+1} + \tanh\left(\mathbf{W}^{\ell+1}\mathbf{f}_i^{\ell+1} + \mathbf{b}^{\ell+1}\right)$$

Psi Former Architecture

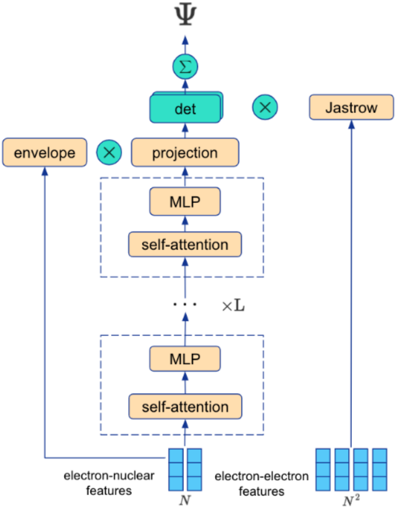

Figure 2: Psi Former Architecture

Thanks

Thanks.