Variational Monte Carlo Study of Hydrogen Atom and Hydrogen Molecule

Gavin Xie 02410400

***Abstract*—A concise summary of aims, methods, results, and conclusions. Explain that the project investigates ground state wavefunctions of the harmonic oscillator, hydrogen atom, and hydrogen molecule using Variational Monte Carlo (VMC), along with optimisation and fitting to Morse potentials.**

I. Introduction

THIS project investigates the ground state energy for hydrogen molecules using Variational Monte Carlo (VMC). The project starts from choosing a suitable numerical differentiation method and sampling method. Those methods are then used to investigate 1D simple harmonic oscillator by calculating the local energy. A parametrised ansatz is then used to find the ground state wavefunction by minimising the Hamiltonian expectation with respect to the parameter θ.

Those methods are then generalized in to multi-dimension forms to determine the ground state energy of the Hydrogen atom,in which methods of differentiation, random sampling and minimization are verified by comparing the minimun local energy of the 1D simple harmonic system to the value provided from the note. The final stage is to determine the bond length and the dissociation energy of the Hydrogen molecules using the optimized ground state energy from different nuclie separations. A chi-square minimization method is used in the curve fitting for the Morse potential.

II. Theory

A. Calculating the Expectation Value $\langle H \rangle$

The expectation value of the Hamiltonian

$$\langle H \rangle = \frac{\int d\mathbf{R}\, \Psi_T^*(\mathbf{R})\, \hat{H}\, \Psi_T(\mathbf{R})}{\int d\mathbf{R}\, |\Psi_T(\mathbf{R})|^2} \tag{1}$$

is investigated using Monte Carlo integration. This is reasonable choice because the integration is performed over a $3N$-dimensional configuration space where N denotes the number of electrons. Deterministic quadrature schemes such as Newton-Cotes rules suffer from the curse of dimensionality as the number of grid points required grows exponentially with dimension, introducing computationally infeasibility for quantum many-body systems. Furthermore, Newton–Cotes methods require a structured grid and smooth integrands, whereas the quantum-mechanical wavefunction may contain sharp features, cusps near nuclei, that would require extremely fine grids to resolve.

Monte Carlo integration offers several important advantages. First, with N_s sampling points, Monte Carlo integration converges as $\mathcal{O}(N_s^{-1/2})$, a rate that is independent of the dimensionality of the integral. This dimensional independence makes Monte Carlo methods especially effective for high-dimensional quantum systems. Moreover, sampling from the probability density

$$\rho(\mathbf{R}) = \frac{|\Psi_T(\mathbf{R})|^2}{\int d\mathbf{R}\, |\Psi_T(\mathbf{R})|^2} \tag{2}$$

naturally concentrates computational effort in the physically relevant regions of configuration space where the wavefunction has significant weight. Further more, Monte Carlo methods do not require a uniform grid and therefore handle cusps and other non-smooth features of Ψ_T without difficulty, unlike grid-based quadrature schemes that would require prohibitively fine resolution.

For these reasons, Monte Carlo integration is the standard and practical method for evaluating expectation values in variational quantum calculations.

To make further progress, (1) can be rewritten as

$$\langle H \rangle = \int d\mathbf{R}\, \rho(\mathbf{R})\, \frac{\hat{H}\, \Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})} = \mathbb{E}_\rho[\, E_L(\mathbf{R})\,] \tag{3}$$

where the local energy is defined by

$$E_L(\mathbf{R}) = \frac{1}{\Psi_T(\mathbf{R})}\,\hat{H}\Psi_T(\mathbf{R}). \tag{4}$$

Using N_s independent samples $\mathbf{R}^{(i)} \sim \rho(\mathbf{R})$, the Monte Carlo estimator of the expectation value becomes

$$\langle H\rangle \approx \frac{1}{N_s}\sum_{i=1}^{N_s} E_L\big(\mathbf{R}^{(i)}\big)\,, \tag{5}$$

which converges to the exact variational energy as $N_s \to \infty$ by the central limit theorem.

B. Verification of Minimization Methods

With a learning rate of 0.01, I verified the one-dimensional gradient-descent implementation using the parabola $f(x) = x^2$, which correctly converges to the analytical minimum at $x = 0$. The choice of learning rate is critical to the stability and efficiency of gradient–descent optimisation. Excessive learning rates cause the updates to overshoot the minimum, resulting in oscillatory behaviour; in our three-dimensional Monte Carlo calculations, a value as large as 0.5 already produces such instability. Conversely, overly small learning rates yield slow convergence and substantial computational overhead. Therefore, appropriate selection of both the initial parameter and the learning rate is essential for achieving robust and efficient minimisation.

In three dimensions, however, gradient descent becomes sensitive to Monte Carlo sampling noise and requires many iterations to reach the minimum. Therefore, a Quasi-Newton method was adopted for improved stability and convergence. To validate this method, it was applied to

$$f(x,y,z) = (x-1)^2 + 2\,(y-1)^2 + 3\,(z-1)^2, \tag{6}$$

which has a known minimum at $(1,1,1)$. The algorithm correctly recovered $x = 1$, $y = 1$, and $z = 1$, with a minimum value of 0, in agreement with the analytical result. The stopping condition used in the deterministic case is based on the relative change in the function value,

$$\Delta = \left|\frac{f_n - f_{n-1}}{f_n}\right| < T_{\text{tol}}. \tag{7}$$

However, when Monte Carlo integration is involved, the function evaluations contain statistical noise, and the optimizer may oscillate around the minimum even after effective convergence. In this case, a more robust stopping criterion should account for such fluctuations. I therefore use a windowed standard–deviation test. For a window of length M, the local mean and standard deviation are computed as

$$s_n = \sqrt{\frac{1}{M-1}\sum_{k=0}^{M-1}(f_{n-k} - \bar{f}_n)^2}. \tag{8}$$

where

$$\bar{f}_n = \frac{1}{M}\sum_{k=0}^{M-1} f_{n-k}, \tag{9}$$

The Monte Carlo noise level of a single function evaluation is estimated as

$$\sigma_{\text{MC}} \approx \frac{\sigma_E}{\sqrt{N_{\text{eff}}}},$$

where σ_E is the standard deviation of the sampled local energies and N_{eff} is the effective sampling number. Convergence is declared when

$$s_n \;\lesssim\; c\,\sigma_{\text{MC}},$$

with c a constant of order unity, indicating that the optimizer's fluctuations are no larger than the intrinsic statistical noise of the Monte Carlo estimator. This criterion effectively detects when the iterates are merely oscillating within the sampling uncertainty rather than making further progress toward the minimum.

III. 1D Harmonic Oscillator

A. The Dimensionless Schrödinger Equation

The Schrodinger equation for a simple harmonic oscillator in 1D, and a single electron is given by

$$-\frac{\hbar^2}{2m_e}\frac{d^2\psi(x)}{dx^2} \;+\; \frac{1}{2}m_e\omega^2x^2\psi(x) \;=\; \mathcal{E}\,\psi(x) \tag{10}$$

where ψ is the wavefunction, $\mathcal{E}$ is the energy, m_e is the electron mass, $\hbar$ is the reduced Planck constant, and ω is the harmonic oscillator strength.

In Eq. 10 each term must have dimensions of an energy multiplied by the wavefunction. Dimensions of each variable and constant are shown in the Tab. I

By using narural units, the dimensionless Schrodinger equation can be written as

$$-\frac{1}{2}\frac{d^2\psi(x)}{dx^2} \;+\; \frac{1}{2}\,x^2\psi(x) \;=\; \epsilon\,\psi(x) \tag{11}$$

Quantity	Symbol	Dimension
Wavefunction	$\psi(x)$	$L^{-1/2}$
Position	x	L
Mass	m_e	M
Reduced Planck constant	$\hbar$	ML^2T^{-1}
Angular frequency	ω	T^{-1}
Energy	$\mathcal{E}$	ML^2T^{-2}
Kinetic term	$-\frac{\hbar^2}{2m_e}\frac{d^2}{dx^2}\psi$	$ML^2T^{-2}\,\psi$
Potential term	$\frac{1}{2}m_e\omega^2x^2\psi$	$ML^2T^{-2}\,\psi$
Right-hand side	$\mathcal{E}\psi$	$ML^2T^{-2}\,\psi$

TABLE I: Dimensions of all quantities and terms in the 1D harmonic oscillator Schrödinger equation.

where ϵ is a dimensionless constant. The dimensionless Hamiltonian is given by

$$H = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2. \tag{12}$$

B. Ground State Eigenfunction and Energy

Using the ground-state wavefunction $\psi_0(x) = e^{-x^2/2}$, I evaluate the local energy through Eq. 4. The second derivative

$$\psi_0''(x) = (x^2 - 1)e^{-x^2/2} \tag{13}$$

allows me to apply the Hamiltonian directly:

$$\hat{H}\psi_0(x) = -\tfrac{1}{2}(x^2-1)e^{-x^2/2} + \tfrac{1}{2}x^2e^{-x^2/2} = \tfrac{1}{2}e^{-x^2/2}. \tag{14}$$

Dividing by $\psi_0(x)$ yields

$$E_L(x) = \tfrac{1}{2}, \tag{15}$$

which is the analytical ground state energy for the 1D harmonic oscillator.

C. 1D Finite Difference Method

To investigate the effect of the finite-difference step size h on numerical accuracy, the local error was evaluated as a function of h for several approximation orders. As shown in Fig. 1, the error initially decreases with h but subsequently increases once h becomes too small. This trend reflects the competition between truncation and roundoff errors. For a central finite-difference stencil of order $2m$, a Taylor expansion shows that the leading neglected term scales as $O(h^{2m})$, giving a truncation error

$$\varepsilon_{\text{trunc}} \sim h^{2m}.$$

Higher-order schemes converge more rapidly as h decreases.

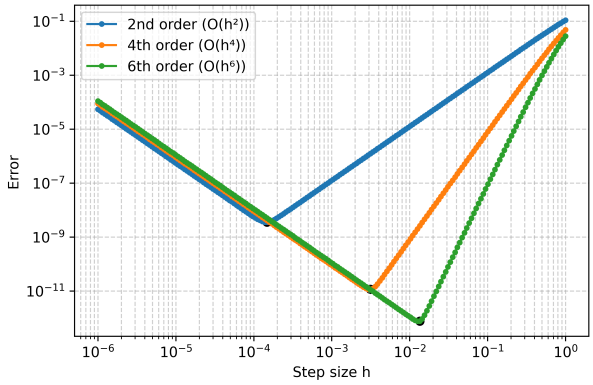

Fig. 1: Enter Caption

When h becomes small, however, roundoff error dominates due to cancellation among nearly equal floating-point quantities. For example, the fourth-order stencil requires multiple subtractions before the final division by h^2, producing a roundoff contribution

$$\varepsilon_{\text{round}} \sim \frac{\epsilon_{\text{mach}}}{h^2},$$

where ϵ_{mach} denotes machine precision. The total numerical error may therefore be approximated as

$$\varepsilon(h) \sim h^{2m} + \frac{\epsilon_{\text{mach}}}{h^2},$$

indicating that an optimal step size exists where the competing errors balance. Figure 1 shows that the second-, fourth-, and sixth-order schemes each exhibit a minimum. The corresponding optimal step sizes, determined via a gradient-descent search, are listed in Table II.

Order	Optimal step size	Average error
2nd	$(1.55 \pm 0.2) \times 10^{-4}$	$(3.6 \pm 0.2) \times 10^{-9}$
4th	$(3.078 \pm 0.001) \times 10^{-3}$	$(1.148 \pm 0.6) \times 10^{-11}$
6th	$(1.174 \pm 0.001) \times 10^{-2}$	$(8.43 \pm 0.08) \times 10^{-13}$

TABLE II: Optimal step sizes h_{opt} and corresponding average errors for the 2nd-, 4th-, and 6th-order finite-difference approximations of the second derivative. Uncertainties represent the standard deviation obtained from plateau sampling of the minimisation algorithm.

Table II shows that the fourth–order finite-difference scheme achieves its optimal accuracy at $h = (3.078 \pm 0.001) \times 10^{-3}$ and produces a numerical error of $(1.148 \pm 0.6) \times 10^{-11}$. This error is roughly two orders of magnitude smaller

than that of the second–order method. Although the sixth–order approximation reduces the error by an additional factor of ten, it causes a substantially higher computational cost.

In Variational Monte Carlo (VMC) calculations, a numerical error of 10^{-11} already lies far below the intrinsic statistical uncertainty of the Monte Carlo estimator. Reducing the truncation error further offers little practical benefit while increasing the computational effort per iteration. I therefore chose the fourth–order method in the finite-difference evaluation of the 3D Laplacian in the calculations.

D. Metropolis sampling

The Metropolis–Hastings algorithm constructs a Markov chain whose stationary distribution is the target density $\rho_{prob}(x) = |\psi_n(x)|^2$. Starting from an initial point $x^{(0)}$, trial moves of the form $x' = x^{(i)}+\eta$ are proposed using a symmetric distribution $q(\eta)$, typically a Gaussian of adjustable width δ. Each proposal is accepted with probability

$$A(x \to x') = \min\left\{1,\ \frac{|\psi_n(x')|^2}{|\psi_n(x)|^2}\right\}, \tag{16}$$

and rejected moves leave the walker at $x^{(i)}$. After discarding an initial burn-in period, the subsequent samples follow $\rho(x)$, despite with correlations.

At each accepted configuration the local energy is computed using the same fourth order finite-difference approximation for the kinetic term, and the expectation value is estimated via an average over the Markov chain.

The motivation for using Metropolis sampling becomes clear when contrasted with the alternatives. Both transformation and rejection sampling generate independent samples, but transformation sampling is only applicable when the cumulative distribution of $\rho_{prob}(x)$ can be inverted analytically, which is a rare situation for realistic quantum systems. Rejection sampling requires a known envelope function that efficiently bounds $\rho_{prob}(x)$; for structured wave-functions such as that of H_2, constructing such an envelope is difficult, and naive choices lead to extremely low acceptance rates.

Metropolis sampling avoids these limitations entirely because it requires only the ability to evaluate $|\psi_n(x)|^2$ up to a multiplicative constant, making it applicable to the VMC problem. Its main drawback is the presence of correlations and the need to tune the proposal width δ. However, with appropriate diagnostics, the algorithm remains efficient and robust.

In Metropolis sampling, the proposal step size δ controls the statistical efficiency of the Markov chain through its impact on the integrated autocorrelation time τ_{int} and thus on the effective sample size $N_{\text{eff}} \approx N/(2\tau_{\text{int}})$. Small steps produce high acceptance but advance the chain only gradually, whereas large steps are rarely accepted and cause the chain to stall. The autocorrelation time τ_{int} can be calculate from

$$\tau_{\text{int}} = \frac{1}{2} + \sum_{k=1}^{\infty} \rho_{auto}(k). \tag{17}$$

The normalized autocorrelation function ρ_{auto} at lag k is defined as

$$\rho_{auto}(k) = \frac{C(k)}{C(0)}, \tag{18}$$

where the autocovariance is given by

$$C(k) = \frac{1}{N_s - k} \sum_{t=1}^{N_s-k} \left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right), \tag{19}$$

and $C(0)$ denotes the zero-lag variance. The efficiency of the Metropolis sampling can be given by the efficiency rate

$$r = \frac{Neff}{N}. \tag{20}$$

I use gradient descent minimisation to find the highest efficient rate $r = 0.217$ at $\delta = 1.744$ for 1D oscillator. The acceptance rate is 0.435, which agrees with the general strategy of tuning. In Monte Carlo integral, the error shares the order with $1/\sqrt{N_s}$, here I use $N_s = 10^6$ sampling points, indicating the error has an order to 3 decimal place. Therefore, the δ and r value is truncated to 3 decimal place. Same truncation is also applied to calculation in Hydrogen atom and molecule.

Hence, the standard deviation of the Monte Carlo integral can be calculated as:

$$\sigma_{\langle H\rangle} = \sqrt{\frac{1}{rN_s(N_s-1)} \sum_{i=1}^{N_s} \left(E_L^{(i)} - \langle H\rangle\right)^2}. \tag{21}$$

Overall, while transformation and rejection methods are preferable when they are analytically feasible and computationally efficient, the Metropolis

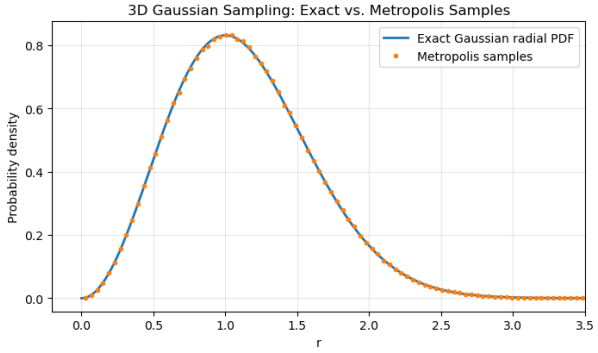

Fig. 2: Enter Caption

algorithm provides a flexible and widely applicable sampling strategy for cases where the structure of ψ_n makes direct sampling difficult or impossible.

To validate the correctness of the Metropolis sampling procedure, a three-dimensional Gaussian probability density was adopted as the target distribution. The empirical distribution obtained from the Metropolis samples closely reproduces the analytical Gaussian PDF, as illustrated in Fig. 2. Further verification was performed in one dimension by selecting a fixed direction in the 3D space and comparing the resulting marginal distribution along that axis with the corresponding one-dimensional Gaussian density. The excellent agreement in both the 3D and reduced 1D cases confirms that the Metropolis sampler accurately reproduces the intended probability distribution.

E. Monte Carlo Verification

Using a Metropolis sample of $N_s = 10^6$ points, I evaluated the numerical local energy $E_L(x)$ and compared it with the analytical ground-state value of 0.5 given in Eq. 15. The results in Fig. 3 show that the point-wise deviation $E_L(x) - E_0$ remains below 10^{-10} across the sampled domain, demonstrating the high accuracy of the finite-difference evaluation of $\psi''(x)$ and the resulting local energy. The sampled mean local energy is $\langle H \rangle = 0.5$. The statistical uncertainty is on the order of 10^{-14}, which arises not from Monte Carlo fluctuations—since the ground-state local energy is constant—but from numerical roundoff errors in the floating-point evaluation.

F. Error Analysis

Discuss statistical vs numerical error and justify the optimal step size used later in 3D problems.

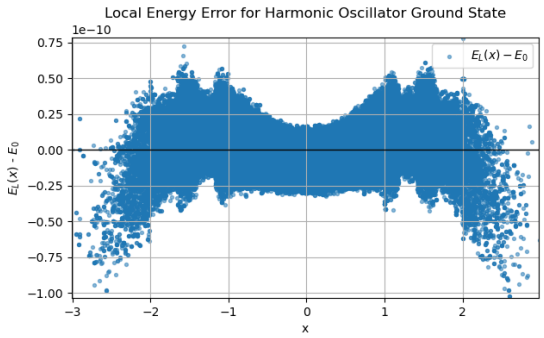

Fig. 3: Difference between the numerical local energy $E_L(x)$ and the exact ground-state value for the one-dimensional harmonic oscillator. Using $N_s = 10^6$ Metropolis samples, the deviation $E_L(x) - 0.5$ remains below 10^{-10}, verifying the accuracy of the finite-difference second-derivative approximation.

IV. The Hydrogen Atom

A. Hamiltonian and Ansatz

In this section, I use a parametrised ansatz

$$\psi(r;\theta) = e^{-\theta r}. \tag{22}$$

to find the ground state eigenfunction of the Hydrogen atom under the Born-Oppenheimer approximation. The form of the ansatz preserves spherical symmetry and resembles the exact ground state e^{-r}, allowing a single-parameter optimisation. The dimensionless Hamiltonian for this system is

$$H = -\frac{1}{2}\nabla^2 - \frac{1}{r}. \tag{23}$$

I verified the three-dimensional Laplacian by comparing the numerical results with the analytical values of the test function $f(x, y, z) = e^{x^2+y^2+z^2}$. The deviation along the x-axis is plotted in Fig. 4. The results show that the numerical error is on the order of 10^{-10}, confirming the accuracy of the finite-difference Laplacian.

Using the same gradient-descent procedure, I obtained an optimal proposal step size of $\delta = 1.000$, which corresponds to an efficiency ratio of $r = 0.053$. This choice of δ yields an acceptance rate of approximately 0.40. The verification of this 3D sampling is shown in the Fig. 4.

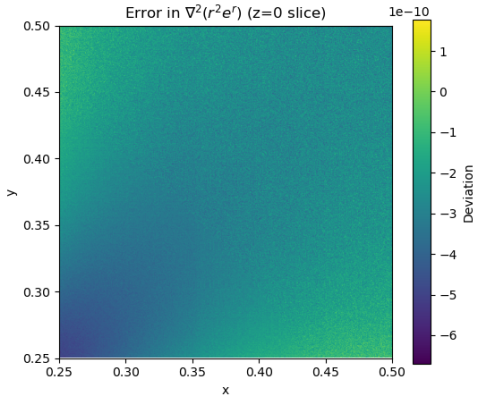

Fig. 4: Enter Caption

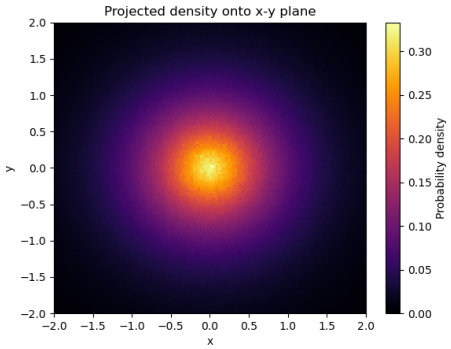

Fig. 5: Enter Caption

B. Gradient of Energy and Minimisation

The gradient of the variational energy with respect to θ is evaluated using

$$\partial_\theta \langle H \rangle \approx \frac{2}{N_s} \sum_{i=1}^{N_s} \left[\left(E_L(\mathbf{R}^{(i)}) - \langle H \rangle \right) \frac{\partial_\theta \psi(\mathbf{R}^{(i)}; \theta)}{\psi(\mathbf{R}^{(i)}; \theta)} \right] \tag{24}$$

where the configurations $\mathbf{R}^{(i)}$ are drawn from the Metropolis sampler according to $\rho(\mathbf{R}; \theta)$. In this situation, the variation Monte Carlo only has one parameter. Therefore I choose a 1D gradient descent minimization method to find the optimal value of $\theta = 0.9999946 \pm 0.0000001$. Fixing θ at 0.9999946 the ground state energy is found at $\langle H \rangle = -0.4997 \pm 0.0001$ The histogram plot of x-y plane is shown in Fig. 5.

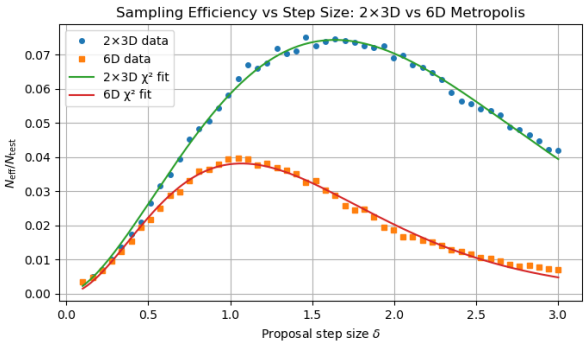

Fig. 6: should add chi2 to the legend

V. The Hydrogen Molecule H_2

A. Hamiltonian

The dimensionless Hamiltonian of Hydrogen molecule can be written as

$$H = -\frac{1}{2} \sum_{i=1}^{2} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) - \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{1}{|\mathbf{r}_i - \mathbf{q}_j|} + \frac{1}{|\mathbf{r}} \tag{25}$$

Therefore the local energy and the expectation of the Hamiltonian can be calculated by Eq. 4 and Eq. 5.

B. 6D Metropolis Sampling

In Metropolis sampling of the H_2 molecule, the electronic coordinates are typically updated using two independent three-dimensional Gaussian proposals, one for each electron, rather than a single six-dimensional displacement of the full configuration. This choice has a significant impact on sampling efficiency: although both approaches can be tuned to yield similar acceptance rates, their autocorrelation properties differ markedly. A 6D proposal perturbs all coordinates simultaneously, producing larger effective jumps and stronger coupling between degrees of freedom, which generally slows exploration of configuration space and increases the autocorrelation time. In contrast, 3D single-electron moves generate localized, physically reasonable displacements that decorrelate electronic positions more effectively, leading to a smaller integrated autocorrelation time τ_{int} and thus a larger effective sample size N_{eff}. The dimensionality of the proposal distribution therefore plays a central role in determining the practical efficiency of Metropolis sampling for many-body quantum systems. The results comparing the efficiency rate against δ is shown in Fig. 6.

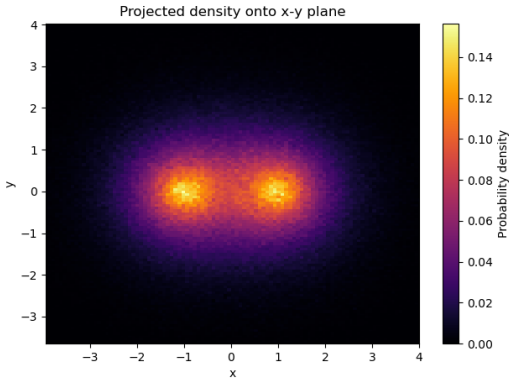

Fig. 7: Enter Caption

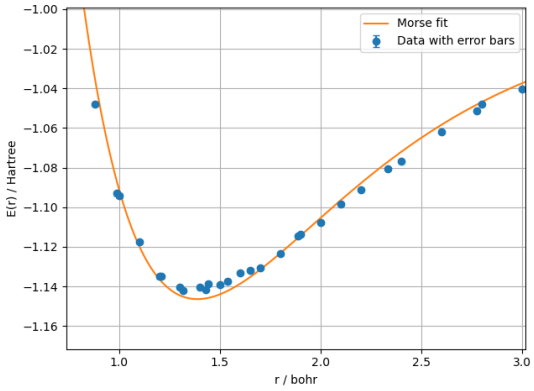

Fig. 8: Enter Caption

C. Two-Electron Ansatz

$$\psi(r_1, r_2; \theta) = \left[e^{-\theta_1(|r_1-q_1|+|r_2-q_2|)} + e^{-\theta_1(|r_1-q_2|+|r_2-q_1|)}\right] e^{-\frac{\theta_2}{1+\theta_3 r_{12}}}, \quad (26)$$

The variational parameters in the two-electron ansatz each play a distinct physical role. The parameter θ_1 governs the localisation of each electron around the nuclei by controlling the exponential decay of the single-particle terms, thereby setting the effective electron–nucleus binding strength. The parameter θ_2 appears in the Jastrow correlation factor and modulates the strength of the electron–electron repulsion, ensuring that configurations with small inter-electronic separation r_{12} are appropriately suppressed. Finally, θ_3 adjusts the correlation length of this Jastrow term; by altering how rapidly the denominator $1 + \theta_3 r_{12}$ grows with distance, it controls how strongly the correlation weakens as the electrons move apart. For atomic distance $|q1 - q2| = 2$, a 2D histogram of the probability distribution is shown in Fig. 7.

D. Sampling Strategy

- Metropolis sampling in 6D,
- Combined electron moves for efficiency,
- Burn-in stage justification,
- Step-size adjustment,
- Correlation analysis via autocorrelation time.

How those choices influence the error rate from MS sampling?

E. Energy Minimisation

Describe:

- gradient estimator,
- regularisation needed to maintain $\theta_1, \theta_3 > 0$,
- convergence diagnostics (energy trace, gradient norm, noise level).

Why gradient descent is no longer robust to MC noise?

Cannot converge even for more than 400 steps but only wandering around the minimun. Cannot achieve a small enough (only 1e-2 level) gradient or abs change in the norm of theta.

Use Quist-Newton and set the tolerance to grad tol 1e-3 (cannot set for even smaller for MC noise?) step size to 1e-4.

Also use mc noise from local energy to set for the converge condition. sigma delta (sq average of the std) comparing to the difference of the Energy. Meaning less if it not changing a lot.

How should I set for the tolerance? According to what information?

VI. Fitting to the Morse Potential

Given energies $E(r)$ at various nuclear separations r, fit

$$V_{\text{Morse}}(r) = D(1 - e^{-a(r-r_0)})^2 - D + 2E_{\text{single}}. \quad (27)$$

The fitting result is shown in Fig. 8

A. Justification of Fit Procedure

- nonlinear least squares using `curve_fit`;
- weighting by uncertainty in E;
- justification for choice of r-range based on convergence.

B. Results

The fitted Morse potential parameters are

$$r_0 = 1.40 \pm 0.001 \text{ bohr},$$
$$D = 0.1452 \pm 0.0002 \text{ Hartree},$$
$$a = 1.228 \pm 0.002 \text{ bohr}^{-1}.$$

The equilibrium bond length agrees closely with the expected value of approximately 1.4 bohr. The dissociation energy is slightly lower than the typical reference value of about 0.17 Hartree, but remains consistent with variational estimates. Overall, these fitted parameters provide a coherent and physically reasonable description of the molecular potential.

VII. Discussion

- Physical interpretation of parameters,
- Comparison to known exact solutions,
- Sources of error: statistical, numerical (finite differences), optimisation noise,
- Robustness of results across sample sizes.

VIII. Conclusion

Summarise:

- successes (correct hydrogen and H_2 energies),
- challenges (correlated sampling, multi-parameter optimisation),
- future improvements (importance sampling, correlated sampling, automatic differentiation).

A. Subsection Heading Here

Subsection text here.

1) Subsubsection Heading Here: Subsubsection text here.

IX. Conclusion

The conclusion goes here.

Appendix A
Proof of Einstein's Famous Equation

The famous equation

$$E = mc^2$$

can be derived.

Appendix B

[Appendix two text goes here.]

Acknowledgment

[acknowledgments go here]