Improving Accuracy at Low Resolution For Diffusion Monte Carlo

Robbie Morrison : 32965303

1

June 7, 2024

ABSTRACT

Aims. To develop a method of approximating systematic errors present in the quantum Diffusion Monte Carlo method. This would present a method of achieving high accuracy for a computational intensity lower than traditional methods.
Methods. The first systematic error investigated is caused by the initialisation of the walker. We investigate its magnitude through the 1D infinite square well by increasing the number of simulations and calculating the average deviation from the true value. The second systematic error is caused by the spatial approximations from representing the potential as a grid. We investigate its magnitude by developing a functional form between the approximated area and true area of the 2D circular well via Monte Carlo Integration.
Results. The initialisation relative error is $\approx 1\%$ whilst the spatial relative error is inversely proportional to the number of grid points between the center and edge of the well within two standard deviations.
Conclusions. The method is effective at reducing the error at low lying energy levels, but if high accuracy results are needed there are more suitable methods such as the boundary element method.

Key words. Diffusion Monte Carlo – Systematic Error – Random Walks

Contents

1. Introduction

Diffusion Monte Carlo (DMC) is a stochastic method that approximates the solution of the diffusion equation through random walks. Random walks are characterized by a sequence of random steps in any given number of directions, with the probability distribution of each step dependent on the specific problem. The similarities between the Schrödinger equation and the diffusion equation led to quantum DMC, a method through which low-lying eigenstates and eigenenergies may be numerically found.

In the context of quantum DMC, the repeated simulations of the random walk represent a Feynman path integral as the number of simulations and spatial resolution approach infinity. This means that traditionally the only way to achieve very small errors would be to use a high resolution and many simulations. This can be very computationally intensive. At lower levels of accuracy, both random error and systematic errors are present. The aim of this paper is to reduce the magnitude of these errors such that our estimations of the energy are accurate to an upper bound of 1.5%. We then apply the developed method to a "squashed circle" which has no analytical solution and compare our results to the boundary element method (BEM).

The first systematic error is a result of where we initialise the random walk. We choose to initialise it at antinodes of the wavefunction, where a walker is most likely to be found, but a full analysis requires the probability of a walker to obey Born's rule. This systematic error is explored in Section 4. The second systematic error considered is a result of approximating a continuous variable (e.g. curvature) via a set of grid points. We investigate this error by assuming it is directly proportional to the relative error between the area of the circular approximation and the true circle in the 2D circular potential well. We then find this error's dependence, if any, on the "radius" of the grid I. The full method is explained in Section 5. Lastly, we apply our systematic error removal method to a "squashed circle" infinite potential: a shape similar to both the circle and the square but without an analytical solution. This is done in section 6.

2. Theory

2.1. 1D Theory

For one dimension, the motion of a particle according to a random walk with a probability of being arrested at a given point is defined by

$$p(i, n + 1) = \frac{1}{2}(1 - a(i))(p(i - 1, n) + p(i + 1, n)) \qquad (1)$$

where i is a discretized spatial coordinate, n is a discretized time coordinate, $a(i)$ is the arrest probability and p is the probability of a particle reaching the specified point in discretized space-time. More specifically, the equation describes the probability of surviving to a point i after $n+1$ time steps based on the combined probabilities of the particle at n being at either $i \pm 1$ all multiplied by the probability of survival at i given by $1 - a(i)$.

The 1D time independent Schrödinger equation is written as

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + V(x)\psi = E\psi \qquad (2)$$

where x is the spatial axis, $\hbar$ is the reduced Planck's constant, m is the particle mass, ψ is the particle wavefunction, $V(x)$ is the

potential as a function of x, and E is the particle's energy. The goal is to recast Equation (2) to be in the same form as (1).

We now consider a specific example. The 1D infinite square well has a potential $V(x) = 0$ in the range $-b < x < b$ where b is a boundary constant and x is the coordinate of our one-dimensional spatial axis. Outside this range, $V(x) = \infty$. Recasting Equation (2) for this potential setup, we find

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = E\psi \qquad (3)$$

as for $V(x) = \infty$ the only way the equation could hold is if $\psi = 0$, and due to the Born rule the probability of finding the particle at x would be zero. We therefore only consider x for $V(x) = 0$. We will now derive this equation from (1).

We can Taylor expand each RHS probability term to second-order accuracy and the LHS probability term to first-order accuracy to find :

$$p(i, n + 1) = p(i, n) + \frac{\partial p}{\partial n}... \qquad (4)$$

$$p(i + 1, n) = p(i, n) + \frac{\partial p}{\partial i} + \frac{\partial^2 p}{\partial i^2}... \qquad (5)$$

$$p(i - 1, n) = p(i, n) - \frac{\partial p}{\partial i} + \frac{\partial^2 p}{\partial i^2}... \qquad (6)$$

Substituting these terms into Equation (1) and cancelling any common terms we find

$$p(i, n) + \frac{\partial p}{\partial n} = (1 - a(i))(p(i, n) + \frac{1}{2}\frac{\partial^2 p}{\partial i^2}) \qquad (7)$$

We assume the probability of arrest ,$a(i)$, is small as well as the second-order derivative so the product is ignored. With this approximation, we can rewrite Equation (7) as

$$\frac{\partial p}{\partial n} = \frac{1}{2}\frac{\partial^2 p}{\partial i^2} - a(i)p \qquad (8)$$

This resembles the time-dependent Schrödinger equation. To derive equation(2) in this form, we write the solution as a separable solution ansatz of the form $p(i, n) = X(i)T(n)$. We make another ansatz that $T(n) = e^{-\lambda n}$ where λ is the arrest rate. $T(n)$ is of this form because we assume the survival rate will exponentially decrease as n increases. We will investigate the accuracy of this assumption. Substituting our ansatz into equation(8) and rearranging we find

$$-\frac{1}{2}\frac{d^2X}{di^2} + a(i)X = \lambda X \qquad (9)$$

which clearly resembles Equation (2) (with $a(i) = 0$). The energy E is thus directly dependent on the value of λ. We can compare Equation (9) to Equation (3) by substituting into (3) the approximation $x = bi/I$ where I is the total number of grid points in the range $[0, b]$. This substitution leads (with slight rearranging) to the equation

$$-\frac{1}{2}\frac{d^2\psi}{di^2} = \frac{mb^2E}{\hbar^2 I^2}\psi \qquad (10)$$

and comparing this with Equation (9) we finally reach

$$E = (\frac{\hbar^2}{mb^2})(\lambda I^2) \tag{11}$$

The first bracketed term represents constants which we ignore. The second bracketed term can be determined from our computer program and so we take this term as the relative energy. It is unitless.

The eigenenergies of the 1D infinite square well are derived by Griffiths and Schroeter (2018) to be

$$E_n = \frac{n^2\pi^2\hbar^2}{8mb^2} \tag{12}$$

where $n = 1, 2, 3...$ and refers to the energy level. For the ground state, $n = 1$ and so the true value of $\lambda I^2 = n^2\frac{\pi^2}{8}$.

To solve for λ we take the natural logarithm of the equation $p(i, n) = X(i)e^{-\lambda n}$ to reach the equation

$$ln(p) = -\lambda n + ln(X(i)) \tag{13}$$

This equation is not useful if we only ran the simulation once, but if we take a Monte Carlo approach and simulate the trajectory for many particles and instead plot $ln(k)$ (the logarithm of the number of particles surviving to a time n) against n the magnitude of the gradient would still be $|\lambda|$. This method only works under the assumption that the number of steps n is large and so the distribution can be taken as continuous.

The algorithm used to compute these values is given as follows:

- Generate a walker at an antinode (explained shortly)
- Allow walker to take a maximum of n_{max} steps a fixed distance left or right at every step
- If the walker reaches a node then the walker is arrested and the length of time it survived for is recorded
- If the walker does not touch a node after n_{max} steps it is arrested and recorded at n_{max}
- Run this simulation for many walkers

Antinodes and nodes refer to the points on the grid space that correspond to the maximum and zero values of $|\psi(i)^2|$ respectively where $\psi(i)$ is the wavefunction along the grid points i. An example of this for the first excited state of the 1D infinite square well is shown in Figure 1. We use this method because our algorithm relies on the probability $a(i)$ of finding a particle at a given grid point i. Particles are most likely to be found at antinodes and so it is reasonable to begin their walk there. A full analysis would begin each walk at all non-nodal points with a probability of a particle starting at $i \propto |\psi(i)^2|$. This is beyond the scope of this paper. We arrest at all nodal points because the probability of finding the particle there is zero.

If I is an even number, it will always take the particle an even number of steps to reach a nodal line, and if I is odd it will always take an odd number of steps. We can hence deduce that our algorithm is a binomial distribution. When plotting the number of particles k arrested at a given time step n, we will only plot every other data point (n being even or odd depends on I, as previously explained).
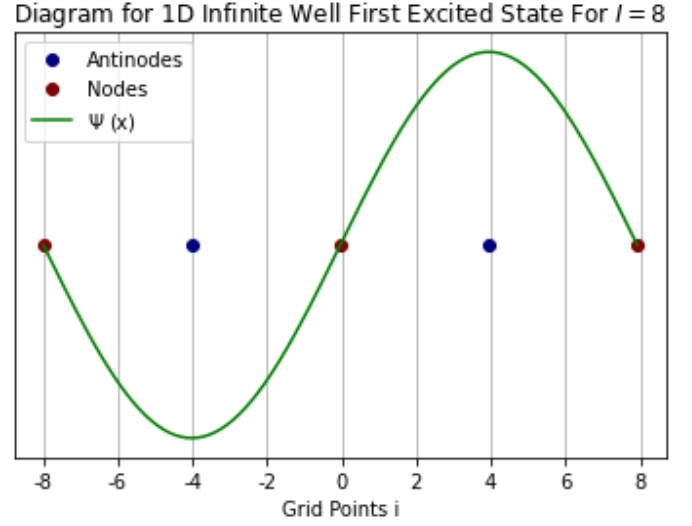

Fig. 1. Diagram representing how we choose starting and arrest points for different wells and levels of excitation. This particular diagram represents energy level $n = 2$ for the 1D infinite well evaluated over grid spacing $I = 8$.

2.2. 2D Wells

2.2.1. Random Walk 2D Energies

We now extend our use of diffusion Monte Carlo to two dimensions. Consider the random walk process defined by

$$p(i, j, n+1) = \frac{1}{4}(1-a(i, j))(p(i+1, j, n)+p(i-1, j, n)+p(i, j+1, n)+p(i, j-1, n)) \tag{14}$$

Where all definitions are the same as in Equation (1) except that we now have two discrete spatial positions: (i, j). Thus by making the same Taylor series expansions as in Equation (7) we reach

$$p + \frac{\partial p}{\partial n} = \frac{1}{4}(1 - a(i, j))(4p + \frac{\partial^2 p}{\partial i^2} + \frac{\partial^2 p}{\partial j^2}) \tag{15}$$

and making the assumption again that $a(i, j)$ and any second-order derivative are sufficiently small that their products are equal, we find

$$\frac{\partial p}{\partial n} = \frac{1}{4}(\frac{\partial^2 p}{\partial i^2} + \frac{\partial^2 p}{\partial j^2}) - a(i, j)p \tag{16}$$

Substituting our ansatz $p(i, j, n) = X(i, j)e^{-\lambda n}$ similar to what we did in Section 2.1 (with the exception being that the spatial solution now depends on two dimensions) we find

$$-\frac{1}{4}(\frac{\partial^2 X}{\partial i^2} + \frac{\partial^2 X}{\partial j^2}) + a(i, j)X = \lambda X \tag{17}$$

We take $a(i, j) = 0$ for the interior of the well. The 2D time-independent Schrödinger equation is

$$-\frac{\hbar^2}{2m}(\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2}) = E\psi \tag{18}$$

where ψ is an eigenstate and x, y are coordinates in two dimensions. Writing $x = \frac{bi}{I}$ and $y = \frac{bj}{I}$, we can rewrite Equation (18) as

$$\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} = -\frac{2Emb^2}{I^2\hbar^2} \tag{19}$$

Comparing Equation (19) with (17) (remembering we are taking $a(i, j) = 0$), we can write the expression

$$E = \frac{2\hbar^2\lambda I^2}{mb^2} \tag{20}$$

There are two important features of this equation. The first is that other than the well's maximum distance from the center b and the number of dimensions, we have not specified anything about the shape of the well. This generality makes it useful for solving the eigenenergies of wells without analytical solutions, such as in Section 6. The second is the similarity to Equation (11), with the only difference being a factor of two. If we were to rewrite Equation (15) with d number of walking directions, the division factor would be written as $\frac{1}{2d}$. This would result in the equation

$$E = \frac{d\hbar^2\lambda I^2}{mb^2} \tag{21}$$

where d is the number of dimensions. This method is thus extremely general; analytically unsolvable infinite wells of arbitrary dimension can have their eigenenergies estimated by this equation. We do not use this equation in the rest of the paper. We derived it only to emphasize the scope of quantum DMC's applicability.

2.2.2. Infinite Circular Well Energies

The time-independent Schrödinger equation for an N-dimensional well is the Poisson equation:

$$\nabla^2\psi = -k^2\psi \tag{22}$$

Where $k = \sqrt{\frac{2mE}{\hbar^2}}$ with E being the eigenenergy. Assume that $\psi = \psi(x, y)$ as we are working in two dimensions. Convert the Cartesian coordinates (x, y) to polar coordinates (r, θ). Strang (1991) shows that the polar Laplace operator can be written as

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \tag{23}$$

Hence Equation (22) can be rewritten in polar coordinates as

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial\psi(r,\theta)}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2\psi(r,\theta)}{\partial\theta^2} = -k^2\psi(r,\theta) \tag{24}$$

Assuming $\psi(r, \theta)$ is a separable solution, we rewrite $\psi(r, \theta) = R(\theta)\Theta(\theta)$ to show

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial(R\Theta)}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2(R\Theta)}{\partial\Theta^2} = -k^2 R\Theta \tag{25}$$

By expanding out the derivatives, as well as converting the partial differentials to ordinary differentials (as they now only depend on one variable), we can write this equation as

$$\frac{d^2R}{dr^2}\Theta + \frac{\Theta}{r}\frac{dR}{dr} + \frac{R}{r^2}\frac{d^2\Theta}{d\theta^2} = -k^2 R\Theta \tag{26}$$

We can multiply through by $\frac{r^2}{R\Theta}$ and separate our equation to have all R terms on LHS and all Θ terms on the RHS to find

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{r}{R}\frac{dR}{dr} + k^2r^2 = -\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} \tag{27}$$

For the LHS of Equation (27) to equal the RHS, both sides must be equal to the same constant. We denote this by the constant l^2 to find the two second order coupled ODEs

$$\frac{r^2}{R}\frac{d^2R}{dr^2} + \frac{r}{R}\frac{dR}{dr} + k^2r^2 = l^2 \tag{28}$$

and

$$\frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} = -l^2 \tag{29}$$

where Equation (28) is known as the Radial Equation, and Equation (29) is known as the Angular Equation. Solving the Angular Equation, we find:

$$\Theta(\theta) = Ae^{il\Theta} \tag{30}$$

Where A is a normalisation constant to be determined. Using the normalisation condition

$$\int_0^{2\pi}|Ae^{il\theta}|^2d\theta = 1 \tag{31}$$

we find that $A = \frac{1}{\sqrt{2\pi}}$. This leads us to the normalized form of Equation (30)

$$\Theta(\theta) = \frac{1}{\sqrt{2\pi}}e^{il\theta} \tag{32}$$

we now solve for $R(r)$ in Equation (28). Rearranging, we may write it as

$$\frac{r^2}{k^2}\frac{d^2R}{dr^2} + \frac{r}{k^2}\frac{dR}{dr} + \left(r^2 - \frac{l^2}{k^2}\right)R = 0 \tag{33}$$

This form of a 2nd order ODE is known as Bessel's Equation. Bowman (2012) shows that the solution can be written as

$$R(r) = c_1 J_l(kr) + c_2 Y_l(kr) \tag{34}$$

where J_l and Y_l are the Bessel functions of the first and second kind, while $c_{1,2}$ are constants. As $r \to 0$, $Y_l(kr) \to \infty$. As $r = 0$ is a real state in our system, $c_2 = 0$ for the solution $R(r)$ to not diverge. Solving for $R(r) = c_1 J_l(kr)$ at $R(b) = 0$ (where b is the boundary of the infinite circular well), we find:

$$J_l(kb) = 0 \tag{35}$$

J_l is a periodic function which has an infinite number of roots (similar to the sine or cosine function). Denoting the nth root of J_l as x_{nl} we find

$$k = \frac{x_{nl}}{b} \tag{36}$$

Comparing this to our original k in Equation (22) we finally reach

$$E_{nl} = \frac{\hbar^2 x_{nl}^2}{2mb^2} \tag{37}$$

which is the set of eigenenergies for the infinite circular well. Comparing Equation (37) to (20), we find that $\lambda I^2 = \frac{x_{nl}^2}{4}$.

3. Linear Fitting

3.1. General Fitting Methods

In this section of the paper, we discuss the reasoning and methods behind the linear models used. All relevant data in this paper can be modeled using a linear fit, and the least squares method produces only global minima for our fitting parameters and no local minima. Hence, we always use the least squares method (which minimises residuals) for determining our fitting parameters. The method is explained fully by Björck (1990).

To evaluate the quality of a model we use the reduced chi-squared metric χ_ν^2, given by

$$\chi_\nu^2 = \frac{1}{N_{obs} - 2} \sum^{N_{obs}} \frac{(y_{model} - y_{obs})^2}{y_{error}} \tag{38}$$

where N_{obs} is the number of data points used in the model, y_{model} is the dependent variable's predicted value from the model, y_{obs} is the observed dependent variable value, and y_{error} is the error between the observed value and the true value. We subtract by 2 in the scaling factor as the linear fit has two fitting parameters: the gradient and the y-intercept. $\chi_\nu^2 < 1$ suggests overfitting in the model (residuals too small), whilst $\chi_\nu^2 > 1$ suggests underfitting (residuals too large). We estimate the values of y_{error} by running a given simulation multiple times, averaging over the dependent variable values and taking their standard deviation as the error.

To estimate the uncertainty of the fitting parameters, we use two different methods. If we have multiple fitting parameters γ_i representing the same quantity γ, we average over all γ_i and take their mean as our calculated parameter. The standard deviation on the mean is taken as the uncertainty on γ_{mean}. This method is preferred as it also reduces the effect of random error. We use this method in Sections 5 and 6 to determine the uncertainties on the relative energies. If we only have a single calculated parameter γ_{single}, we estimate the square root of its diagonal entry in the covariance matrix as its uncertainty. The diagonal entries broadly describe how well a given parameter of the model describes the data, with its quantitative value being the parameter's variance. We take its square root to give a standard deviation. This is explained in more detail by Chew (1970).

3.2. Determining Energies

In this paper, the least squares method is frequently used to determine the decay constant λ in Equation (13). To determine the
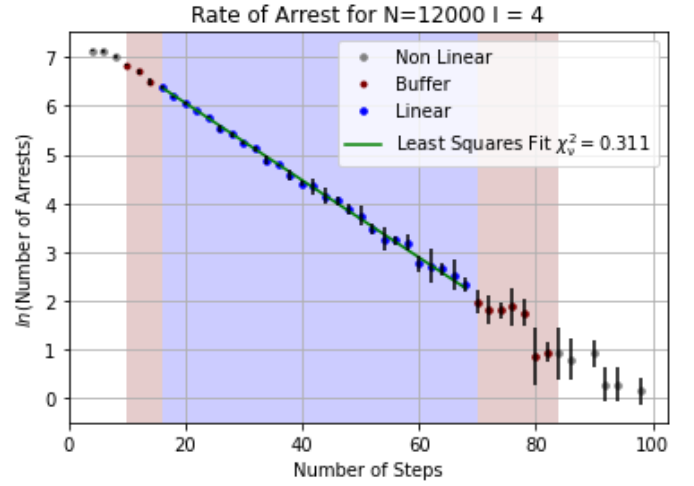

Fig. 2. An example of how we define non-linear, buffer, and linear regions on the graph of $ln(k)$ against n. The values of N and I are chosen purely for display purposes: graphs of higher I will have a larger domain so it is harder to see the data points.

value of λ we plot the number of steps n against the logarithm of the number of walkers arrested $ln(k)$, as explained in Section 2.1.

For small values of n the trend will be non-linear as it does not obey the assumption that n is large so we can model the distribution as continuous. Values of n close to n_{max} will be non-linear as we cannot achieve negative values of $ln(k)$ so the distribution begins to flatten. This motivates us to restrict the fit to a linear region which we determine empirically. As we average over multiple simulations the linear region's boundaries may vary slightly, and so we define small buffers (also determined empirically) to allow for the simulations to be evaluated over a constant linear region. This removes the risk of any non-linear points being included in the model. This is significant as for very large n, there is a possibility that zero walkers are arrested at that step and so its logarithmic value $ln(0)$ is undefined. Fitting an undefined value is not possible with the least squares method. An example of this is shown in Figure 2.

4. 1D Infinite Square Well: Initialisation Systematic Error

4.1. Methods

As our method represents a Feynman path integral, the starting point of the random walk depends on the value of $|\psi(x)|^2$, where $\psi(x)$ is the eigenstate of the particle. The most basic analysis would be to allow the particle to start any spatial step with equal probability. An improved method would be to start the random walk at the antinodes of the eigenstate i.e. where the particle has the highest probability of being found. A full analysis would initialise the particle at a point i with the probability of initialisation $\propto |\psi(i)|^2$. We use the second method in this paper so a systematic error will always be present. The aim of this section is to determine the magnitude of this systematic error.

For any given potential in any number of dimensions, as the number of walks $N \rightarrow \infty$ and the number of grid points the boundary is from the centre $I \rightarrow \infty$, the predicted eigenenergy from the method should converge to the true value (with the added systematic error discussed earlier). The reason that we need $I \rightarrow \infty$ is that for continuous changes (e.g. circular well or

Table 1. The linear and buffer regions used for different values of *I*. $L_{begin/end}$ represent the boundaries of the region used to fit the linear model. $\varepsilon_{begin/end}$ represent the widths of the buffers used, not the boundaries themselves.

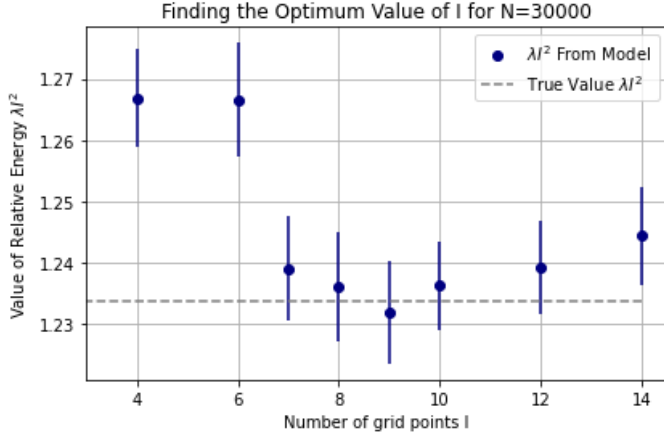| I	ε_{begin}	L_{begin}	L_{end}	ε_{end}
4/5	6	16	70	15
6/7	20	30	130	20
8/9	20	50	260	40
10/11	40	70	380	40
12/13	50	100	550	50
14/15	50	130	650	50

Fig. 3. How the predicted eigenenergy for 1D infinite well's ground state varies with the number of grid points *I*. The minimum appears to lie between $I = 7$ and $I = 10$, so we take the value as $I = 8$ ($I = 9$ would also be valid ; non-integer *I* cannot be evaluated).

the Coulomb potential), the grid we evaluate with can only approximate the potential setup unless we have an "infinite" resolution. The 1D ground state infinite potential well has no continuous changes and so we can test convergence simply by letting $N \rightarrow \infty$.

We choose an initial number of particles *N* and evaluate λI^2 (the relative energy) for the different values of *I*. As $N \neq \infty$, there will be an error initially caused by our value of *I* being too small, and then by our value of *N* being too small for the larger *I* due to underfitting. Thus, for our initial starting *N*, the optimum *I* is the value of the point of zero gradient for the graph of λI^2 against *I*. We then increase *N* for this value of *I* and observe where λI^2 converges , compared to the true value of $\frac{\pi^2}{8}$. For all calculated values of λ, we evaluate the simulation five times before fitting a model to reduce the effect of random error. Errors on λI^2 are taken from the covariance matrix.

4.2. Results

Table 4.2 represents the regions we use to fit our model in this section. This is the only table that includes the values of the buffer regions as they are wider than usual. We define thick buffer regions as we use the same region for two values e.g. $I = 4, 5$.

Figure 4 represents λI^2 against *I* for $N = 30000$. All values of *I* in the range $[7, 12]$ predict the true value within one standard deviation. The optimum value is taken as $I = 8$, although $I = 9$ is equivalently acceptable as the range $[7, 10]$ is where the minimum appears to occur. We take $I = 8$ rather than $I = 9$ as it requires shorter computational times.
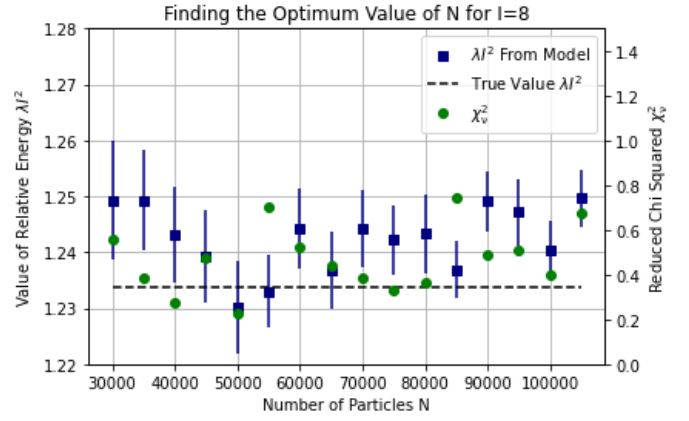
Fig. 4. How λI^2 varies with *N* for $I = 8$. The values of the reduced χ_ν^2 are included to briefly analyse the effect of overfitting.

Figure 4 shows how λI^2 varies with *N* for $I = 8$. The majority of the values of λI^2 rest above the true value. The mean of these values is 0.00867 ± 0.0003 above the true value suggesting a systematic error of $+0.975\%$. The average of the χ_ν^2 values is 0.46 ± 0.1 with no strong correlation to the error on λI^2.

4.3. Discussion

The only value that consistently changes with *N* is the uncertainty on λI^2. The residuals on the model become smaller as *N* increases and so the covariance matrix entry related to the gradient is reduced. Additionally, the systematic error seems to be the only source of error other than random error. Even the χ_ν^2 (quality of the model) seems to have negligible impact on the predicted relative energy, possibly because overfitting does not worsen the predicted λ. We can therefore assume that the systematic error from the initialisation of the walker does not depend on *N* or *I*, and from now on we take it as a constant overestimate of $\approx 1\%$.

There is a possibility that this systematic error changes as the number of dimensions increases. We assume in this paper it does not, but a method of analysing this would be to repeat this section's method using the nD infinite square well, where $n > 1$ and an integer. These wells do not suffer from systematic errors due to $I < \infty$, and so are higher dimensional equivalents of the 1D infinite square well in this context.

5. 2D Circular Well: Spatial Systematic Errors

5.1. Deriving A Formula for the Boundary Systematic Error

The 2D circular well has curved boundaries, and so using a value of $I < \infty$ will always result in a piecewise approximation of the potential boundary. An example of this is shown in Figure 5. The goal of this section is to determine a method of approximating this systematic error's magnitude. To do this, we use Monte Carlo integration to approximate the area of the potential well governed by $V = 0$ and calculate its relative error ε compared to πI^2, a circle's exact area of radius *I*.

We assume that the relative error's relation to the *I* is of the form

$$\varepsilon = AI^p \tag{39}$$

where *A* and *p* are constants to be determined. We can determine these constants by taking logarithms of both sides and using the
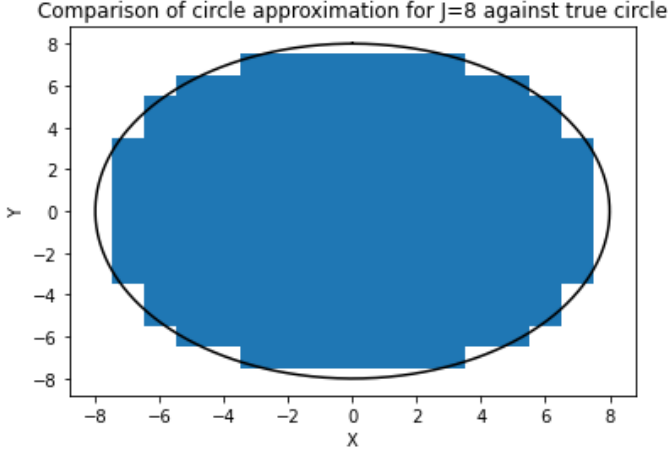
Fig. 5. An approximation of a circle using a spatial resolution of $I = 8$. Note that the circle appearing compressed is a limitation of the graphics, not the analysis itself. Also note that the corners jutting out of the circle boundary are errors in the visualisation but do not effect any data.
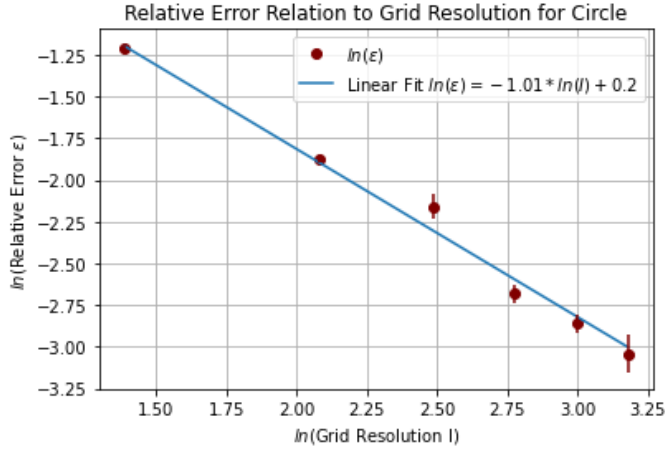

Fig. 6. Plot showing how $ln(I)$ relates to the logarithm of the relative error $ln(\varepsilon)$ on the circular potential well approximation. Within one standard deviation, ε is inversely proportional to I.

least squares method to fit a linear model to $ln(\varepsilon)$ against $ln(I)$. We calculate ε three times for each I to obtain errors and reduce the effect of systematic error. The plot is shown in Figure 6.

The value of p is -1.01 ± 0.03. We therefore estimate the relation between ε and I as inversely proportional. The relative error on the area approximation is not necessarily the same as the systematic error on the predicted value of λI^2, the relative energy. We assume the two are directly proportional to each other to obtain the relation

$$\varepsilon_{energy} = \frac{\kappa}{I} \tag{40}$$

where ε_{energy} is the systematic error on λ, and κ is a constant not necessarily equivalent to A in Equation 39 but still a constant. We discuss the error on λ and on λI^2 interchangeably as the only difference is a factor of I^2. Noting that ε_λ can be written as

$$\varepsilon_{energy} = 1 - \frac{\lambda_{true}}{\lambda_{pred}} \tag{41}$$

where λ_{true} is the theoretical value of λ associated with the energy and λ_{pred} is the value predicted from our model, we can
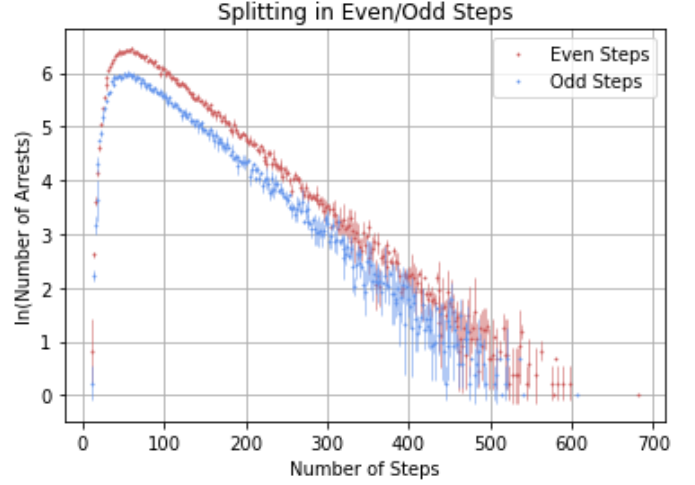

Fig. 7. A plot of $ln(k)$ against n. The details of the fitting parameters are unimportant; the graph is only present to illustrate the splitting in even vs odd steps.

write a relation between two different values of λI^2 calculated at different I to predict the value of λI^2 without this systematic error. We calculate it the relative error on the λ_{pred}, not λ_{true}, as we often do not know the exact value of λ_{true} . Equating both equations' κ, we find

$$I_1(1 - \frac{\lambda_{true}}{\lambda_1}) = I_2(1 - \frac{\lambda_{true}}{\lambda_2}) \tag{42}$$

Rearranging for λ_{true}, we reach the formula

$$\lambda_{true} = \frac{I_1 - I_2}{\frac{I_1}{\lambda_1} - \frac{I_2}{\lambda_2}} \tag{43}$$

5.2. Methods

We can use Equation (43) to estimate the relative energy without the systematic error produced by a finite value of I. We expect this value to be an overestimate of $\approx 1\%$, using the systematic error outlined in Section 4. We thus remove 1% from the predicted value. Determining the value of λ is almost identical to the method described in Section 3.2. The only difference is that if I is even (odd), there will be more boundary grid points an even (odd) number of steps away. This means that there will be a systematically larger number of particles arrested at even (odd) steps. We thus fit two separate models to both the even and odd arrested walkers, and average over the two values of λ to reduce the effect of random error. An example of this is shown in Figure 7.

As discussed in Section 4, for a given value of N there will be errors both due to the size of I being too small (note that this error is different to the systematic error in this section, it is always present unless $N \rightarrow \infty$) and the size of N being too small for a given I. To choose our values of I_1 and I_2 in Equation 43, we look for the point of zero gradient on a plot of λI^2 against I and choose its I value as well as the smaller adjacent I value to it. We choose the first value to minimise the previously discussed errors. We choose the second value because we expect the systematic error's magnitude to decrease as I increases, so an adjacent value to the right would have a larger error produced

Table 2. Table giving linearity beginnings and endings for all λI^2 calculated for 2D circular well. Note that L_{begin} and L_{end} are given in terms of step number, not the value of n as we use the same range for both even and odd steps. An example is $L_{begin} = 25$ would be $n = 50$ for even steps.

I	Energy Level	L_{begin}	L_{end}
8	0	25	182
10	0	31	219
12	0	36	272
14	0	49	347
16	0	67	431
8	1	18	116
10	1	46	178
12	1	50	225
14	1	58	318
16	1	78	400
7	2	7	59
9	2	12	88
12	2	26	148
14	2	34	180
16	2	50	250

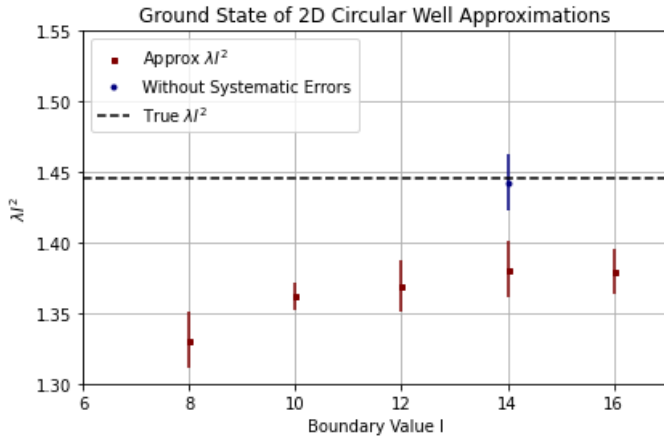**Fig. 8.** λI^2 against I for N=60000 and averaging over five different repeats. The removal of the two discussed systematic errors is shown for the maximum on this graph.

not from the systematic error but from the value of N being too small.

The antinodal positions are determined using the appropriate Bessel functions and rounding the antinode's position to the nearest integer. This works well for the first excited state as they are given by exact integers for even I. For the second excited state we choose values of I that minimise the round-off error.

5.3. Results

Figure 8 shows our estimate of the true value of energy without the discussed systematic errors. We take $I = 14$ as the maximum and $I = 12$ as the second value used to determine the systematic error. Removing the systematic error from the initialisation of the walker and the spatial systematic error gives a value of $\lambda I^2 = 1.443 \pm 0.02$. The relative error between this and the true value is -0.23%.

Figure 9 shows how λI^2 varies with I with $N = 60000$ for the first excited state $n = 1$ and $L = 1$ (Bessel function root indices) of the 2D circular well. We use $I = 12$ and $I = 10$ to determine a
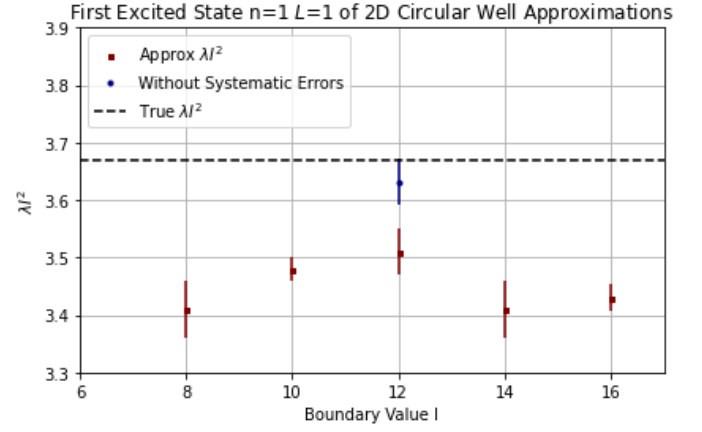
Fig. 9. Approximation of excited state $n = 1$ $L = 1$ for 2D circular well with and without discussed systematic errors. We use $N = 60000$ and average over five simulations.
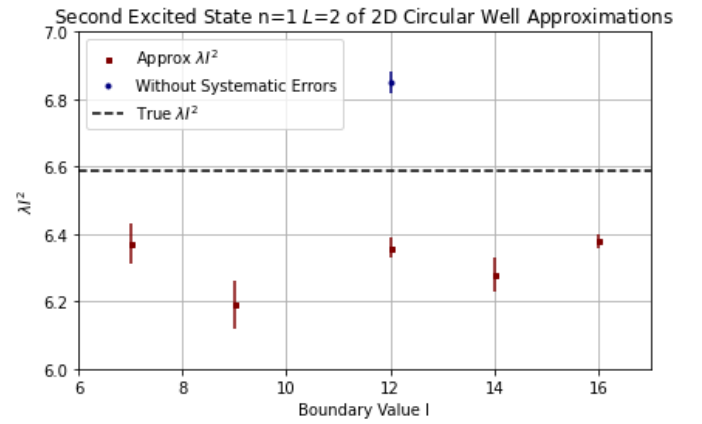

Fig. 10. λI^2 against I for second excited state $n = 1$ $l = 2$ 2D circular well. We use $N = 60000$ and average over five simulations.

value without either systematic errors as 3.631 ± 0.03, producing an error of -1.05%.

Figure 10 shows how λI^2 varies with I for the second excited state of the 2D circular well with $n = 1$ and $L = 2$. We use $I = 9$ and $I = 12$ The removal of systematic errors gives us an estimate of the relative energy as 6.85 ± 0.03 which has a relative error compared to the true value of 3.9%.

5.4. Discussion

The systematic error removal works well for the ground state and first excited state of the 2D circular well, producing relative errors smaller than without the removal. The second excited state has a significantly larger relative error with the removal of the systematic errors compared to the previous states. This is likely because the maximum is less obvious- we take it as $I = 12$ as the points in range $I = [9, 12]$ mimic the behaviour of a maximum, but $I = 7$ and $I = 16$ have values almost identical to the maximum at $I = 12$. Additionally, the difference between the relative energy at $I = 9$ and $I = 12$ is much larger than in the previous states, suggesting that the error from the low value of I for a given N (not the curved systematic error) is much larger than before. This results in a significant overestimate of the curved systematic error's magnitude.
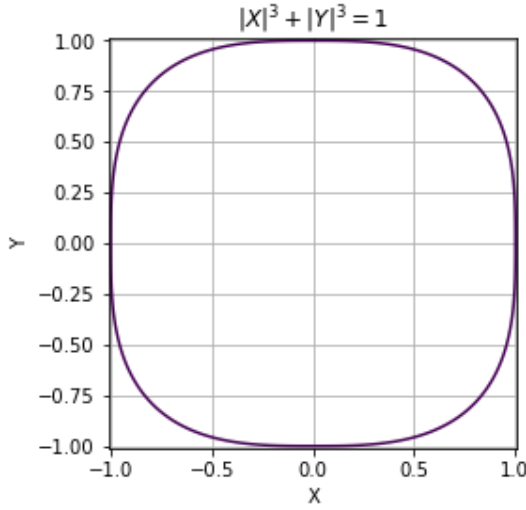
Fig. 11. Diagram showing shape $|x|^3 + |y|^3 = 1$

An additional source of error associated with the excited states is the nodal lines' widths. The nodal lines should be infinitesimally small, but in our grids they have a relative width of $1/I$. It is more difficult to estimate the magnitude of this error, but given that it reduces the area the walker can survive in (which is also the curvature systematic error's source) it will result in an underestimate of the relative energy. This is a possible explanation of the underestimate produced in the first excited state.

6. The Squashed Circle

6.1. Methods

The circle and the square both belong to a family of shapes of the form

$$|x|^n + |y|^n = b^n \tag{44}$$

where $n > 0$ and b is a constant denoting the "radius" of the shape. The circle exists at $n = 2$ and the square occurs as $n \to \infty$. As 2D infinite potential wells, the eigenstates for both the square and circle have analytical solutions (2D square well solution derived by Aronstein and Stroud (1997)), but the shapes in the range $n = (2, \infty)$ do not. We investigate the eigenenergies of one particular well, $n = 3$ and $b = 1$, whose shape is shown in Figure 11.

The method for determining eigenenergies in this well is identical to the method used in Section 5. We can argue that as the antinode positions for the ground state and first excited state are a result of symmetry, the relative locations remain the same $(-\frac{I}{2}, \frac{I}{2})$. We do not investigate the second excited state as we could not achieve a relative error $< 1.5\%$ with the systematic error removal. However, through the same symmetry arguments, the nodal lines should remain the same for second excited state.

Although analytical solutions to the area of the well exist, we use double integral Gaussian quadrature to numerically calculate the area. This method provides a higher accuracy than other numerical integration methods e.g. Simpson's rule and does not noticeably increase computational time. This area is necessary to find the relative error's dependence on I, explained in more detail in Section 5.1. We achieve a dependence of $p = -0.96 \pm 0.02$. The relative error is inversely proportional to I within 2 standard deviations, which is a larger uncertainty than in the case of the

Table 3. Table giving linearity beginnings and endings for all λI^2 calculated for squashed circle. Note that L_{begin} and L_{end} are given in terms of step number, not the value of n as we use the same range for both even and odd steps. An example is $L_{begin} = 25$ would be $n = 50$ for even steps.

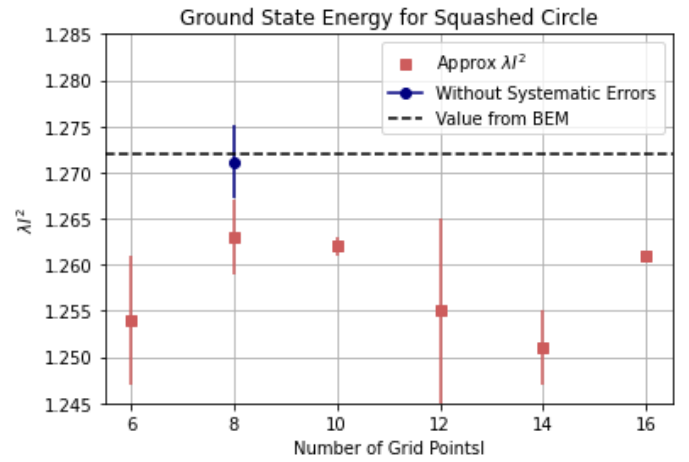| I	Energy Level	L_{begin}	L_{end}
8	0	30	135
10	0	50	200
12	0	55	240
14	0	80	325
16	0	100	400
6	1	9	30
8	1	15	62
10	1	20	80
12	1	30	120
14	1	38	160

Fig. 12. Ground state energy estimate for the curved square potential. $N = 60000$ and values averaged over five repeats.

circle. We note that this may introduce a small error when approximating the curvature systematic error.

The benchmark value we use comes from the boundary element method (BEM) used by Lee (2017). Each value has an uncertainty of 0.1% with no known systematic errors. Note that the values are given in terms of the wavenumber, which is related to our relative energy in the form $k = 2\sqrt{\lambda}I$.

6.2. Results

Figure 12 shows our estimate of the relative ground state energy of the curved square potential. We take $I = 8$ as the maximum and $I = 6$ as our second value to approximate the curved systematic error. Our estimate is 1.2701 ± 0.004, which has a relative error -0.1% compared to the BEM's result of 1.272 ± 0.001.

Figure 13 shows our estimate of the first excited state energy of the curved square potential. We take $I = 8$ as the maximum (it is higher than $I = 10$ by 0.003) and use $I = 6$ as the second value to approximate the curved systematic error. Removing systematic errors gives an estimate of 3.616 ± 0.01, which has a relative error 1.48% compared to the BEM solution of 3.561.
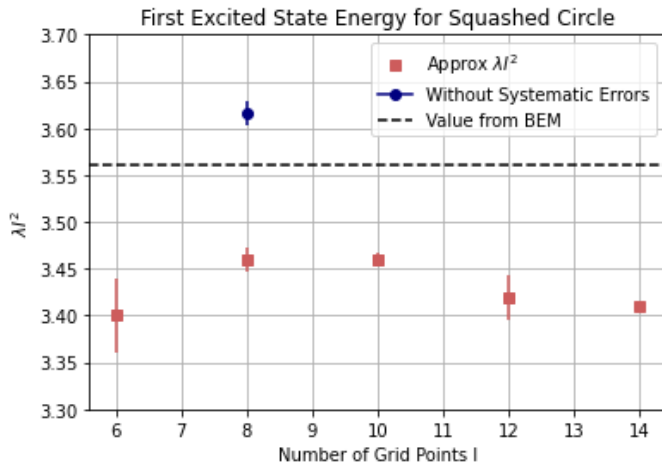
Fig. 13. First excited state energy estimate for the curved square potential. $N = 6000$ and values averaged over five repeats.

6.3. Discussion

Similarly to section 5, our ground state value is very accurate compared to not removing the systematic error. The first excited state, although more accurate than without the systematic error removal, does not significantly reduce the error with the BEM solution. This could be due to other systematic errors present e.g. the approximation of the nodal line, but it is just as likely that the error comes from random fluctuations. The boundary systematic error removal requires the λI^2 estimates not to experience significant random error, otherwise, we would estimate the systematic errors as either an overestimate or underestimate. The method we use is not consistently useful: it is always subject to random fluctuations which are inherently unpredictable. For methods without analytical solutions we have no "true value" to base our λI^2 estimates on, so we cannot predict the form that the distribution of λI^2 against I should take.

7. Conclusions and Applications

7.1. Conclusions

The aim of this paper was to achieve relative errors $< 1.5\%$ on eigenenergies of quantum wells without using very large I or N. We did this by estimating the systematic error due to walker initialisation as a 1% overestimate and the curved boundary systematic error as an underestimate inversely proportional to the number of grid points from the center I. For the circle, this produced relative errors -0.23%, -1.1%, and 3.9% for the first three energy levels respectively. For the squashed circle, the relative errors compared to the BEM method are -0.1% and 1.5% for the first two energy levels respectively. The method evidently decreases in quality as the energy level increases. Although we attempted to reduce the effect of random error through repeated trials and averaging, the accuracy at this level still significantly depends on random fluctuations. This can be seen in the second excited state of the circle, where the removal of systematic errors produced a worse result than letting them remain. Although DMC can be used for the first two energy levels to achieve accuracies of $< 1.5\%$, it is better to use methods with much smaller random errors e.g. BEM. The strength of quantum DMC lies in its adjustability: for any number of dimensions and any given potential shape it can provide rough solutions with errors $\leq 10\%$.

7.2. Applications

In Section , we argued that the squashed circle should have the same first two nodal lines as the circle because it is symmetric with respect to the x and y axes. An equally valid argument is that the nodal lines also appear for the excited states of the square, and so as the circle converges to the square the nodal lines should remain. Another type of nodal line for the circular well that we did not investigate were the radial nodal lines. These occur as the n in Equation (36), which denotes the root present in the specific Bessel function, increases. They take the form of concentric circles inside the well. The radial lines are not present in the square well, so the squashed circles present an interesting method of analysing what happens to these radial nodal lines.

To find where the nodal lines are, we require the form of the wavefunction. Wavefunction determination was not explored for quantum DMC in this paper, though methods do exist as described by Needs et al. (2009). It requires a trial wavefunction, which could be iteratively approximated using analytical solutions. An example would be the trial wavefunction for $|x|^{2.1} + |y|^{2.1} = b^{2.1}$, which would be the eigenstates for $n = 2$. The trial wavefunction for $n = 2.2$ would be given from our obtained solution for $n = 2.1$. It is unclear how effective this method would be compared to non-stochastic methods such as the Numerov method (described by Johnson (1978)), which uses finite differencing to solve differential equations. Both would be worth investigating, as it is a direct parallel to the effectiveness of the non-stochastic (BEM) vs stochastic (quantum DMC) methods compared in this paper.

References

D. L. Aronstein and C. Stroud. Fractional wave-function revivals in the infinite square well. *Physical Review A*, 55(6):4526, 1997.

C. Batchelor-McAuley and R. G. Compton. Diffusion to a cube: A 3d implicit finite difference method. *Journal of Electroanalytical Chemistry*, 877:114607, 2020.

K. Bhaskaran, A. Gasparrini, S. Hajat, L. Smeeth, and B. Armstrong. Time series regression studies in environmental epidemiology. *International journal of epidemiology*, 42(4):1187–1195, 2013.

Å. Björck. Least squares methods. *Handbook of numerical analysis*, 1:465–652, 1990.

F. Bowman. *Introduction to Bessel functions*. Courier Corporation, 2012.

B. Chen, J. M. Chen, G. Mo, T. A. Black, and D. E. Worthy. Comparison of regional carbon flux estimates from co2 concentration measurements and remote sensing based footprint integration. *Global biogeochemical cycles*, 22 (2), 2008.

V. Chew. Covariance matrix estimation in linear models. *journal of the American Statistical Association*, 65(329):173–181, 1970.

Q. Din. Dynamics of a discrete lotka-volterra model. *Advances in Difference Equations*, 2013:1–13, 2013.

R. Ferretti. High-order approximations of linear control systems via runge-kutta schemes. *Computing*, 58(4):351–364, 1997.

C. F. Fischer. Self-consistent-field (scf) and multiconfiguration (mc) hartree-fock (hf) methods in atomic calculations: Numerical integration approaches. *Computer Physics Reports*, 3(5):274–325, 1986.

T. A. Gallagher, A. J. Nemeth, and L. Hacein-Bey. An introduction to the fourier transform: relationship to mri. *American journal of roentgenology*, 190(5): 1396–1405, 2008.

D. J. Griffiths and D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

S. Gulen, C. Popescu, and M. Sari. A new approach for the black–scholes model with linear and nonlinear volatilities. *Mathematics*, 7(8):760, 2019.

B. M. Irons. Engineering applications of numerical integration in stiffness methods. *AIAA journal*, 4(11):2035–2037, 1966.

B. R. Johnson. The renormalized numerov method applied to calculating bound states of the coupled-channel schroedinger equation. *The Journal of Chemical Physics*, 69(10):4678–4688, 1978.

M. S. Kaylen, P. V. Preckel, and E. T. Loehman. Risk modeling via direct utility maximization using numerical quadrature. *American Journal of Agricultural Economics*, 69(3):701–706, 1987.

M. Kocaleva, B. Zlatanovska, N. Stojkovic, and A. Stojanova. Application of runge-kutta and euler methods for ode through examples. 2018.

K. Kubota and E. Twizell. A nonlinear numerical model of percutaneous drug absorption. *Mathematical biosciences*, 108(2):157–178, 1992.

D. Lee. From the circle to the square: Symmetry and degeneracy in quantum mechanics. B.S. thesis, Oberlin College, 2017.

D. Marion. An introduction to biological nmr spectroscopy. *Molecular & Cellular Proteomics*, 12(11):3006–3025, 2013.

M. H. Mousavi and S. Ghavidel. Structural time series model for energy demand in iran's transportation sector. *Case Studies on Transport Policy*, 7(2):423–432, 2019.

R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos. Continuum variational and diffusion quantum monte carlo calculations. *Journal of Physics: Condensed Matter*, 22(2):023201, 2009.

G. Strang. *Calculus*, volume 1. SIAM, 1991.

A. Wang. The shazam music recognition service. *Communications of the ACM*, 49(8):44–48, 2006.

Z. Wang, Q. Wang, D. Klinke, et al. Simulation study on effects of order and step size of runge-kutta methods that solve contagious disease and tumor models. *Journal of computer science and systems biology*, 9(5):163, 2016.

B. E. Warren. *X-ray Diffraction*. Courier Corporation, 1990.

S. Xiao and X. Hu. Novel optic high voltage transducer based on gaussian quadrature method. In *2010 Asia-Pacific Power and Energy Engineering Conference*, pages 1–5. IEEE, 2010.

8. Addendum

8.1. Integration

In one dimension, numerical integration approximates the integral as a sum over the areas of shapes describing the area under the curve. The order of accuracy depends on the shape being used e.g. rectangles for first-order accuracy (Riemann sums), trapeziums for second-order etc. This principle can be extended to any number of dimensions with the corresponding shape increasing in dimension. For example, double integrals (two dimensions) sum over the volumes of shapes. As the number of dimensions increases the calculation becomes costly, and so often Monte Carlo integration is used for high dimension integrals. They can only be evaluated for definite (bounded) integrals.

8.1.1. Integration in Industry

Gaussian quadrature is often used in engineering for finite element analysis (FEA). FEA can be used to approximate the stiffness matrix. This gives values for how a material compresses and behaves under various loads, which usually do not have analytical solutions. Numerical integration is thus crucial to evaluating the structural integrity of buildings, machinery and more. Explained in Irons (1966).

Numerical integration is very often used in finance. An example of this is utility and risk models (URM). The theory says that individuals weigh different paths forward under risky circumstances by assessing their different expected utilities / usefulness. To do this mathematically, probabiltiy densities can be calculated for different paths using numerical integration on the utility function. This is explained in more detail in Kaylen et al. (1987).

The field of medicine often employ numerical integration in the study of pharmacokinetics: the study of how pharmaceuticals interact with bodies. An example of this is the absorption rate of a pill being digested: to calculate the amount of the pill absorbed over a given time, Simpson's rule is often used. This is crucial for determining appropriate quantities of drugs to be given beforehand. Described more in Kubota and Twizell (1992).

8.1.2. Integration in Academia

In classical physics, the EM field can be given by integrating the charge density over a volume. Irregular densities and volumes can prove to be either difficult or impossible to evaluate over analytically. Gaussian quadrature allows high order accuracy approximations to be made of the potential field. Example given in Xiao and Hu (2010).

Numerical integration also appears in quantum mechanics. An example is The Hartree-Fock method, which is a numerical method used to approximate the wavefunction and energy of a many-body systems e.g. two electrons interacting. Within the method integrals appear, often with no anayltical solution. Here numerical integration proves useful yet again, particularly Gaussian quadrature. Empirical methods to remove these integrals exist, but often numerical integration is more efficient and accurate. Example given in Fischer (1986).

The carbon cycle is the process by which carbon atoms are recycled by Earth's natural processes e.g. absorption by plants and storage in stone. Describing it in real-time can be difficult so often it is easier to estimate carbon concentration of a given region e.g. a forest by integrating over its production rate. This is done via numerical integration and is called carbon flux integration. The power of this technique stems from it being used not just for forests, but ocean floors and many other regions too. Explained in more detail in Chen et al. (2008).

8.2. Runge-Kutta and Finite Differencing

Runge-Kutta and finite differencing are two methods used to numerically solve ODEs. Finite differencing discretizes the derivative/s involved and can either solve them explicitly (sequential equation) or implicitly (system of simultaneous equations). It is useful for solving simpler PDEs and ODEs at lower orders of accuracy. Runge-Kutta divides the domain into small steps and evaluates new gradients at each interval. These can be used to estimate new dependent variables (e.g. y in a y vs x graph). The gradients are then averaged over to obtain an estimate of the dependent value that we are solving for. The width of the interval determines the order of accuracy e.g. splitting a domain into 4 intervals generates RK4, the most widely used Runge-Kutta method due to its high accuracy but low computational time compared to other RKn methods.

8.2.1. RK and FD in Academia

The Lotka-Volterra equations are differential equations that describe population dynamics. RK can be employed to solve these to describe the distribution of prey vs predator density over time. The equations assume that the rate predators consume prey is directly proportional to the prey abundance, and so it is natural to model them using differential equations. Typically, the RK4 method is used, as with most numerically solved differential equations. Example given in Din (2013).

An application of FD in chemistry comes from modelling the diffusion of a cube into a medium. It uses a 3D implicit method to model the diffusion for a higher order accuracy in time. This is particularly useful for nanophysics, as often simple cubic structures will undergo diffusion processes. More generally, FD is often used to model the diffusion equation as it is the simplest time-dependent PDE possible. Example given in Batchelor-McAuley and Compton (2020).

Control systems are designed to generate a response to a sudden change e.g. car hitting a bump. Modeling a steep rate of change is best done by RKn methods, generally RK4. A response can be given to maintain the car's speed as constant despite the unexpected disturbances. They are employed mostly in kinematic and electric circuits. An example of this is given in Ferretti (1997).

8.2.2. RK and FD in Industry

The Black-Scholes equation can be used to predict option prices from starting variables. Option prices are the amount the buyer pays to the seller for an amount of stock. This can be predicted using the Black-Scholes equation, a PDE, by transforming it into an ODE using the method of lines. The method of lines converts the 2nd order part of the PDE into a set of discrete value and then allows the 1st order part to be evaluated using RKn methods. Example in Gulen et al. (2019).

Runge kutta methods are great for studying kinematics, as this requires the solving of time-dependent ODEs, but this approach can be extended to the kinematics of video games. Less physics reliant games typically use Euler methods as they are not concerned about the accuracy of the motion, but for triple A games modeling motion, particularly orbital mechanics in games like Stellaris requires reasonably high accuracy (e.g. 4th order) so RK4 is a very suitable method. Example given in Kocaleva et al. (2018).

RK methods are frequently used in medicine to model growth, absorption and more. A particular example is the growth of tumours: it is modeled by a 1st order time dependent differential equation. To determine the size of a tumour after a given amount of time, the RK method can be employed. Extending this to radiotherapy, the death rate of cancerous cells can also be modeled by a 1st order ODE and hence the number of deaths can be solved for using RK methods. Examples given in Wang et al. (2016).

8.3. Fourier Transforms and Time Series

Sounds and visuals are often made of a superposition of individual light / sound waves with different frequencies and amplitudes for each frequency. Fourier transforms are a method of taking a superposed sound and breaking it down into its respective frequencies and amplitudes. Time series analysis takes data points collected at regular time intervals and analyses the underlying relationships behind them.

8.3.1. FT and TS in Industry

Perhaps the most financially successful use of Fourier transforms in industry is Shazam. It is an app that listens to a sample of a song, breaks the sample down into its component frequencies, and compares the broken sample to a database. It then returns the most likely song (if any) based on the sample used. A remarkably simple use, but incredibly famous tool. Explained by the Shazam creators in Wang (2006).

Another use of FT is for MRI medical imaging. The magnetic properties of protons oscillating produce EM waves. The EM waves are in k-space, which is another way of saying frequency space. We therefore have the locations of the protons in frequency space. To convert from k-space to real space, we apply a Fourier Transform. This gives us a scan of the body in real space. Explained in Gallagher et al. (2008).

A very important use of time series analysis is in the energy sector. Energy companies attempt to forecast how much energy is going to be demanded in the coming days from customers and so they use time series analysis on regularly collected data about energy usage to forecast the expected demand. The analysis is not restricted to the energy suppliers: infrastructure planners can use more long term data to understand the expected number of transmission lines, power plants required etc. Example in Mousavi and Ghavidel (2019).

8.3.2. FT and TS In Academia

NMR spectroscopy is a technique used in chemistry to determine the different nuclei and elements present in a sample. It is similar to MRI imaging. Elements with a magnetic moment respond to an EM wave by becoming excited. When they drop back to lower energy levels they emit light of a specific wavelength. By measuring the dropped down energy signals over time we develop time domain signals. We then convert using a Fourier Transform to go from time to frequency. An important difference between this and MRI is that this uses temporal related domains while MRI uses spatial related domains. The peaks of the frequency graph show the elements present, and the intensity of the peak shows how common they are in the sample. Shown in Marion (2013).

As stated before, k space is a form of spatial frequency. X rays diffracting through a crystalline solid cannot easily be described in real space: they "see" the crystal in k space and the equations of diffraction e.g. for the structure factor do not depend on normal space. Thus data from X ray diffractometry must be converted from k space to real space using a Fourier Transform, which then gives information about the electronic structure of a material. Described in Warren (1990).

Epidemiology is the study of how infectious disease spreads (personally I think this next one would be a worthwhile addition to the projects). Time series analysis on the number of people infected, the number of deaths, vaccinations etc can all be used to predict possible trends in how the the disease will spread in the near future. For a pandemic like COVID, this is obviously extremely important as if there is an incoming wave we can in advance advise the public to remain at home. The general method is described in Bhaskaran et al. (2013).