

Chapter 4

The Variational Method

4.1 Introduction

In previous lectures, we have discussed methods of solving the Schrödinger equation based upon integrating a (partial) differential equation. However, it is well known that the Schrödinger equation is an example of an *eigenvalue problem*, and so in this lecture we will explore an alternative method of solution, based upon a *variational inequality* that can be derived from the eigenvalue problem.

This new method, known as the variational method, will allow us to find an approximate value for the ground state energy without knowing the exact ground state wavefunction! It can also, in certain cases, be extended to finding the energy of excited states. We will also show how the estimate of the energy can be systematically improved.

4.2 The eigenvalue problem

We start by expressing the time-independent one-dimensional Schrödinger equation in terms of the Hamiltonian operator \hat{H} :

$$\hat{H}\phi_i(x) = E_i\phi_i(x) \tag{4.1}$$

where E_i is the energy eigenvalue associated with the eigenfunction $\phi_i(x)$ of the Hamiltonian \hat{H} . Typically, we know \hat{H} and want to solve this equation to find the associated eigenvalues and eigenfunctions. In principle, we can always solve this equation by diagonalizing \hat{H} which therefore gives *all* the eigenvalues and eigenfunctions. However, this is often computationally very expensive, and we are often only interested in finding the ground state solution, i.e. the lowest energy state corresponding to the smallest eigenvalue. We therefore seek a method for finding just the lowest eigenvalue, or at least, a reasonable approximation.

Let us consider an arbitrary wavefunction $\psi(x)$. Using the complete set of eigenfunctions $\{\phi_i(x)\}$ as a basis, we may expand the arbitrary function as follows:

$$\psi(x) = \sum_i c_i \phi_i(x) \quad (4.2)$$

where the $\{c_i\}$ are the complex expansion coefficients. Note that $\psi(x)$ may not be properly normalized.

We can now calculate the expectation value of the energy for this state:

$$\langle E \rangle = \frac{\int \psi^*(x) \hat{H} \psi(x) dx}{\int \psi^*(x) \psi(x) dx} \quad (4.3)$$

$$= \frac{\sum_{i,j} c_i^* c_j \int \phi_i^*(x) \hat{H} \phi_j(x) dx}{\sum_{i,j} c_i^* c_j \int \phi_i^*(x) \phi_j(x) dx} \quad (4.4)$$

$$= \frac{\sum_{i,j} c_i^* c_j E_j \int \phi_i^*(x) \phi_j(x) dx}{\sum_{i,j} c_i^* c_j \int \phi_i^*(x) \phi_j(x) dx} \quad (4.5)$$

$$= \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} \quad (4.6)$$

as the basis functions $\{\phi_i(x)\}$ form an ortho-normal set.

Finally, we are now ready to turn to the problem of estimating the ground state energy, E_0 (lowest eigenvalue) *without* solving the full eigenvalue problem. By definition, all excited states have $E_i \geq E_0$ and therefore we have

$$\sum_i |c_i|^2 E_i \geq E_0 \sum_i |c_i|^2 \quad (4.7)$$

and so we have the following *variational inequality*:

$$\langle E \rangle = \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} \geq E_0 \quad (4.8)$$

which shows that any trial wavefunction can be used to give an *upper bound* on the ground state energy. Obviously, the closer the trial wavefunction is to the true ground state wavefunction, the lower this upper bound becomes.

So given a trial wavefunction we can find an approximate value for the energy, and if we can then vary this trial wavefunction in some manner, we can generate a better estimate by seeking to minimize this energy.

4.2.1 Aside - a different view of the Schrödinger equation

In fact, equation 4.3 can actually be recast as a *functional*:

$$E[\psi] = \frac{\int \psi^*(x) \hat{H} \psi(x) dx}{\int \psi^*(x) \psi(x) dx} \quad (4.9)$$

that is, as a “function of a function”. We may then ask what function $\psi(x)$ minimizes this integral? If this sounds a bit unusual, consider asking the question: what function minimizes the distance between two points on a plane? We can express the distance between two points as an integral which gives the length of a curve that passes through both points, and what we want to find is the functional form of the curve which has the shortest length. Another example, is that starting from the principle of least action in mechanics, we can derive the Euler-Lagrange equations which can be reduced to Newton’s Second Law. These are all examples of the *calculus of variations*.

If we now define

$$P = \int \psi^*(x) \hat{H} \psi(x) dx \quad (4.10)$$

$$Q = \int \psi^*(x) \psi(x) dx \quad (4.11)$$

we can write the change in energy δE due to a first-order change in the trial wavefunction $\delta\psi$ as:

$$\begin{aligned} \delta E &= \frac{\int (\psi + \delta\psi)^* \hat{H} (\psi + \delta\psi) dx}{\int (\psi + \delta\psi)^* (\psi + \delta\psi) dx} - \frac{\int \psi^* \hat{H} \psi dx}{\int \psi^* \psi dx} \quad (4.12) \\ &\approx \frac{\int \delta\psi^* \hat{H} \psi dx - \frac{P}{Q} \int \delta\psi^* \psi dx}{Q} + \frac{\int \psi^* \hat{H} \delta\psi dx - \frac{P}{Q} \int \psi^* \delta\psi dx}{Q} \quad (4.13) \end{aligned}$$

If we now consider the stationary states of this energy functional, we require $\delta E = 0$ which, when substituted into equation 4.13, leads to:

$$\begin{aligned} \hat{H}\psi &= \frac{P}{Q}\psi \\ &= E\psi \end{aligned} \quad (4.14)$$

i.e. the Schrödinger equation!

So, this is saying that we can derive the Schrödinger equation as being the necessary condition on a function for it to minimize the expectation value of the energy. That is, the variational principle is just as good a starting place for QM as the Schrödinger equation!

4.3 Rayleigh-Ritz method

How then do we use this variational inequality? One way is to guess a function $\psi_\alpha(x)$ which contains some free parameter α and which satisfies the appropriate boundary conditions. We then substitute this function into equation 4.3 which then gives the energy as a function of α for all values of α . We want the value of α which minimizes the energy, and so can use this to solve for α . We can then improve the functional form of the trial guess, incorporating more free parameters, and so successively reduce the upper bound on the ground state energy.

4.3.1 Simple Harmonic Oscillator example

Here we write the Hamiltonian as:

$$\hat{H} = -\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + \frac{1}{2}m\omega^2 x^2 \quad (4.15)$$

As a simple trial wavefunction that satisfies the appropriate boundary conditions ($\phi(x) \rightarrow 0$ as $x \rightarrow \pm\infty$) we choose:

$$\psi_\alpha(x) = \begin{cases} (\alpha^2 - x^2) & \text{if } |x| < \alpha \\ 0 & \text{otherwise} \end{cases} \quad (4.16)$$

With this trial form the integrals in equation 4.3 can now be done simply, resulting in:

$$E(\alpha) = \frac{5}{4} \left(\frac{\hbar^2}{m} \frac{1}{\alpha^2} + \frac{2}{35} m\omega^2 \alpha^2 \right) \quad (4.17)$$

which when minimized w.r.t. α gives:

$$E_0 \leq 0.598\hbar\omega \quad (4.18)$$

which is quite remarkable given that this problem can, of course, be solved analytically to give an exact ground state energy of

$$E_0^{exact} = 0.5\hbar\omega \quad (4.19)$$

If desired, our estimate of the ground state could be further improved by changing the trial wavefunction, e.g.

$$\psi_{\alpha,\beta}(x) = \begin{cases} (\alpha^2 - x^2)^\beta & \text{if } |x| < \alpha \\ 0 & \text{otherwise} \end{cases} \quad (4.20)$$

However, given the simplicity of the initial trial wavefunction, the estimate we have obtained is already remarkably good. Why is this method so accurate? The answer is that as we are looking for an energy which minimizes

equation 4.3, we have a stationary state w.r.t. first-order variations in $\psi(x)$ and so the answer is accurate to second-order!

Note that this method only finds an upper bound on the lowest state. It may also in certain circumstances be extended to excited states, e.g. if there is a symmetry that can be exploited. For example, with the simple harmonic oscillator, we know that the potential $V(x)$ has a definite *parity*, and so therefore the solutions have a definite *symmetry*. The ground state solution has even symmetry (as does our successful trial wavefunction), the first excited state has odd symmetry, etc. So if we choose an odd function as our trial wavefunction it cannot converge to the ground state, and so will instead converge to the first excited state. In general, this method will work for an excited state as long as our trial function is *orthogonal* to all lower states. Be aware, however, that excited states can have stationary points that are not minima!

4.4 Linear (Matrix) method

Whilst the Rayleigh-Ritz method is very useful for deriving an analytic approximation to the exact ground state wavefunction, its accuracy is dependent upon making a good guess for the functional form, which can be very difficult for more complex problems. Therefore, we now consider a more powerful method, which exploits techniques from matrix algebra, and which can be used successfully in general-purpose computer codes.

We again expand our unknown wavefunction $\psi(x)$ in terms of some basis set, but unlike equation 4.2 where we used the set of eigenfunctions as our basis, this time we use a general set of known functions $\{\chi_i(x)\}$ as a basis:

$$\psi(x) = \sum_{i=1}^N c_i \chi_i(x) \quad (4.21)$$

where the $\{c_i\}$ are the new complex expansion coefficients. Note that in general the set $\{\chi_i(x)\}$ must be linearly independent in order to form a basis set, but they do not need to be complete, nor orthogonal, nor normalized!

We now proceed in a similar manner to equation 4.3 but this time:

$$\langle E \rangle = \frac{\int \psi^*(x) \hat{H} \psi(x) dx}{\int \psi^*(x) \psi(x) dx} \quad (4.22)$$

$$= \frac{\sum_{i,j=1}^N c_i^* c_j \int \chi_i^*(x) \hat{H} \chi_j(x) dx}{\sum_{i,j=1}^N c_i^* c_j \int \chi_i^*(x) \chi_j(x) dx} \quad (4.23)$$

$$= \frac{\sum_{i,j=1}^N c_i^* c_j H_{i,j}}{\sum_{i,j=1}^N c_i^* c_j S_{i,j}} \quad (4.24)$$

where $H_{i,j}$ are the *matrix elements* of the Hamiltonian, and $S_{i,j}$ are the matrix elements of the *overlap matrix*. Note that if the χ_i basis functions are orthogonal, then the overlap matrix reduces to (a scalar times) the identity matrix and life is a bit simpler.

Minimizing equation 4.24 w.r.t. any particular expansion coefficient c_k gives:

$$\delta E = \left(\sum_{i,j=1}^N c_i^* c_j S_{i,j} \right) \left(\sum_{i=1}^N c_i^* H_{i,k} \right) - \left(\sum_{i,j=1}^N c_i^* c_j H_{i,j} \right) \left(\sum_{i=1}^N c_i^* S_{i,k} \right) = 0 \quad (4.25)$$

which can then be re-arranged and substituted into equation 4.24 to give all the eigenenergies $\{E\}$ of the Hamiltonian:

$$\sum_{i=1}^N c_i^* (H_{i,k} - ES_{i,k}) = 0 \quad (4.26)$$

This is just a set of N simultaneous linear equations for the $\{c_i^*\}$ values, which can be written in matrix form as:

$$\mathbf{HC} = \mathbf{ESC} \quad (4.27)$$

where \mathbf{C} is the vector of expansion coefficients. Non-trivial solutions to this matrix equation are only possible if the equation is singular, i.e. the determinant:

$$|\mathbf{H} - \mathbf{ES}| = 0 \quad (4.28)$$

which can then be solved for the eigenenergies $\{E\}$ in the usual way.

4.4.1 Incomplete basis sets

For many practical applications of equation 4.27 we shall not be using a complete basis set - this may be infinitely large! For example, common basis sets that are used in calculations of atoms in molecules and solids include plane-waves, Gaussians and atomic orbitals - these will be discussed in more detail in a later lecture. Obviously, we cannot cope with an infinite basis set in any finite computer, and so we must truncate the basis in some way, resulting in an *incomplete basis set*.

What is the effect of using an incomplete basis set? We will be restricting our search for the lowest eigenvalue to a *subspace* of the full space we should be using. Therefore, the minimum value we find must always be equal to, or greater than, the “true” value. This is the same as in the Rayleigh-Ritz method, where we tried to minimize the upper bound on the target expectation value. If we increase the size of the basis set we use, then the subspace

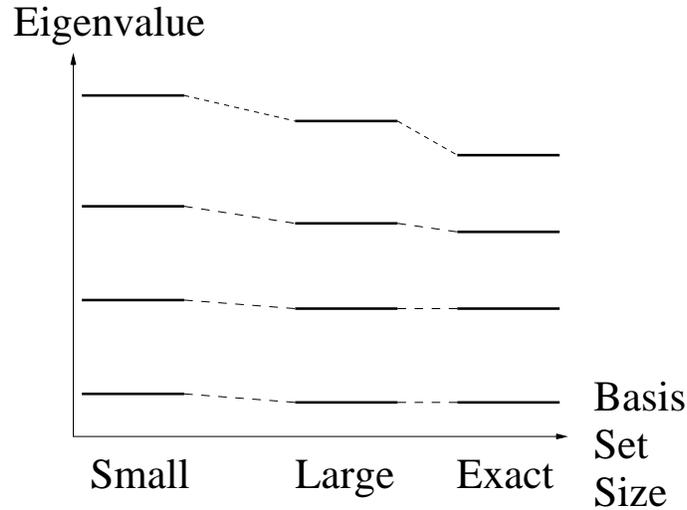


Figure 4.1: Schematic showing the convergence of the eigenvalue spectrum with basis set size. It can be seen that the lower eigenvalues converge more rapidly towards the exact limit as the basis set size is increased than the higher eigenvalues.

becomes larger, and the minimum value we find will decrease until we reach *convergence*. At convergence, we have a sufficiently large basis set such that adding any more basis states will not change the answer. Obviously, we must use larger basis sets to calculate excited states. In general, we will get a spectrum of eigenvalues as the solutions to equation 4.28, with the higher eigenvalues converging more slowly than the lower ones with basis set size. This is shown schematically in figure 4.1 and illustrated quantitatively in the following example.

Example - infinitely deep potential well

This has a potential given by (in 1D):

$$V(x) = \begin{cases} \infty & \text{for } |x| > |a| \\ 0 & \text{for } |x| < |a| \end{cases} \quad (4.29)$$

which forces the wavefunction to vanish on the boundaries, i.e. $\psi(x) \rightarrow 0$ at $x = \pm a$. Of course, this is a problem that can be solved analytically. Here, we demonstrate how to solve this problem using the variational method. We choose units such that $a = \frac{\hbar^2}{2m} = 1$ and for basis functions we choose simple polynomials that vanish at the well boundary:

$$\chi_n(x) = x^n(x-a)(x+a), \quad n = 1, 2, \dots \quad (4.30)$$

Eigenvalue	N=5	N=8	N=12	N=16	Exact
1	2.4674	2.4674	2.4674	2.4674	2.4674
2	9.8754	9.8696	9.8696	9.8696	9.8696
3	22.2934	22.2074	22.2066	22.2066	22.2066
4	50.1246	39.4892	39.4784	39.4784	39.4784
5	87.7392	63.6045	61.6862	61.6850	61.6850

Table 4.1: Lowest part of the eigenvalue spectrum of the infinite square well. The effect of increasing the basis set size N can be clearly seen, in the convergence (from left to right) of the eigenvalue towards the exact value.

where this simple form is chosen so that the matrix elements of the overlap matrix (also known as the *overlap integrals*) can be calculated simply and analytically. The matrix elements of the Hamiltonian can also be found analytically, and so all the elements in equation 4.28 are known and the eigenvalue spectrum can be simply calculated. The results are given in table 4.1 and the resulting spectrum is similar to that shown in figure 4.1 as a function of basis set size. We see that the lowest eigenvalue converges most rapidly with basis set size, whilst considerably more basis functions are required to converge the excited states.

4.4.2 Efficiency

With an orthonormal basis the overlap matrix becomes the identity matrix, and equation 4.27 becomes the standard eigenvalue problem, and may be simply solved using standard linear algebra techniques. This is commonly referred to as *diagonalizing the Hamiltonian*: we seek to find the set of eigenvalues and eigenvectors of the Hamiltonian matrix. To do this, we exploit the *Hermitian* properties of \mathbf{H} . If we write the eigenvectors of \mathbf{H} as the columns of a new matrix \mathbf{U} , then this forms a *unitary* matrix, i.e. $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$. The diagonal form of \mathbf{H} then follows:

$$\mathbf{H}_{diag} = \mathbf{U}^\dagger\mathbf{H}\mathbf{U} \quad (4.31)$$

There exist many high quality library packages (such as BLAS or LAPACK) to perform this task, often with variants to exploit particular properties of the problem where these exist (such as *sparseness* of the matrix to be diagonalized).

In the general case, we have to deal with a non-orthogonal basis, with non-trivial overlap integrals, and the more complicated *generalized eigenvalue problem* must be solved. This requires some sophisticated matrix manipulations and will be covered in a later lecture.

The practical computational limitations on the effectiveness of this method, are the storage requirements of these matrices which is $O(N^2)$ and the com-

puter time required to diagonalize each matrix, which is $O(N^3)$ (where N is the size of the basis set). For this reason, the basis set should be kept as small as possible, whilst giving converged results for the quantities of interest. Note that diagonalizing the Hamiltonian will give *all* the eigenvalues, whereas it is often only a small number of the lowest eigenvalues, corresponding to occupied states, that are of interest. For this reason, various techniques such as the *Lanczos algorithm* have been developed to only return a limited number of eigenvalues, with corresponding increases in efficiency.

For most problems, the linear variational method is more efficient than the direct-solution methods discussed in earlier lectures. In general, the basis set sizes can be chosen to be far smaller than the number of integration points required to achieve comparable accuracy. For an integration grid containing M points however, the finite difference equations can often be formulated as sparse matrices which can then be solved using specialised $O(M)$ methods. As the problem size increases therefore, grid based methods become more efficient and so are starting to become more widely used in research.

4.5 Final comments

A few final points to highlight:

- We can express a wavefunction in terms of basis functions, which then turns the Schrödinger equation into an eigenvalue equation.
- An approximate analytic form for the ground state solution can be found from the Rayleigh-Ritz variational method.
- A numerical solution can be found using the linear variational method.
- If the basis set is incomplete, then the eigenvalues obtained will be upper bounds on the true eigenvalues.
- If the basis set is non-orthogonal, then the generalized eigenvalue problem has to be solved.

4.6 Further Reading

- “Computational Physics” by J.M. Thijssen, Chapter 3

