

The Relationship between the Ritz Variational Method and Frobenius's Method of Solving Schrödinger's Equation

G. HUNTER

Centre for Research in Experimental Space Science, York University, Toronto, Canada

Received September 15, 1969/February 1, 1970

The solution of a linear homogeneous differential equation (in particular the Schrödinger equation) by expansion of the solutions (wave functions) in a discrete complete set of function is considered. The coefficients of the expansion are determinable by either the Ritz variational (integral) method, or by a generalisation of Frobenius's (non-integral) method. Each method leads to an infinite matrix eigenvalue equation. It is shown that the integral and non-integral matrix equations are related by the overlap matrix of the set of basis functions. The effects of truncating the infinite matrices to finite order are described. A hybrid method of transformation to a matrix representation is proposed, which employs some techniques from each of the original methods.

Die Lösung einer linearen homogenen Differentialgleichung (besonders der Schrödinger-Gleichung) durch Entwicklung der Lösungen (Wellenfunktionen) nach einem diskreten vollständigen Satz von Funktionen wird untersucht. Die Entwicklungskoeffizienten sind entweder durch die Variations-Methode (Integral-Methode) oder durch Verallgemeinerung der Methode von Frobenius (Methode ohne Integrale) bestimmbar. Beide Methoden führen zur einer unendlichen Matrix-Eigenwert-Gleichung. Es wird gezeigt, daß die Matrixgleichungen der beiden Verfahren durch die Überlappungsmatrix des Satzes von Basisfunktionen in Beziehung stehen. Es werden die Effekte beschrieben, die sich ergeben, wenn die Matrizen von unendlicher auf endliche Ordnung reduziert werden. Eine Hybridmethode zur Transformation in eine Matrixdarstellung wird vorgeschlagen, die einige Rechenoperationen aus jeder der Originalmethoden anwendet.

On considère la résolution d'une équation différentielle linéaire homogène (en particulier l'équation Schrödinger) en développant les solutions (fonctions d'ondes) dans un groupe discret complet de fonctions. Les coefficients du développement peuvent être déterminés par la méthode de la variation de Ritz (intégral) ou par une généralisation de la méthode de Frobenius (non-intégral). Chaque méthode mène à une équation infinie matrice eigenvalue. On montre que les équations matrice, intégrales et non-intégrales, sont reliées par la matrice à recouvrement du groupe des fonctions fondamentales. On décrit les effets de tronquer les matrices infinies en ordre limité. On présente une méthode hybride de transformation à une représentation de matrice, qui utilise des techniques de chacune des méthodes originales.

1. Introduction

The expansion of the eigenfunctions of linear operators as linear combinations of a known set of functions is a very basic idea arising from the linear algebra of the eigenfunctions as vectors in a Hilbert space [1]. The coefficients of the expansion are usually determined by forming the scalar product with each of the basis functions. As used within the Schrödinger representation of quantum mechanics, this technique is known as the Ritz variational method [2]. In the Ritz method the formation of the scalar product involves an integration over the whole of the

configuration space, so that the original differential operator representation is transformed into a matrix representation. This integral procedure is the basic technique used for almost all current numerical calculations on atomic and molecular systems.

By contrast many well known linear operator problems of wave mechanics [3] (hydrogen atom, harmonic oscillator, rigid rotator, etc.), as well as problems of classical physics [4], and mathematical analysis [5], are solved analytically or semi-analytically by the method of Frobenius [6]. This method also involves a linear combination of a set of basis functions (specifically a power series), but the coefficients are determined by using the linear independency of the basis functions, rather than by formation of scalar products, so that no integrals are involved as in the Ritz variational method. Frobenius's method is essentially a transformation to a matrix representation, although the matrix notation is not used explicitly in classical applications of the method [3, 7, 8].

For a given basis set, the two matrix equations resulting from the Ritz method, and from Frobenius's method, are equivalent through a transformation involving the overlap matrix of the set of basis functions, as shown in 5. This transformation provides an insight into the reasons why particular problems are susceptible to solution by either Ritz's method, or Frobenius's method, but not by both. The suitability of one method or the other for a particular problem is discussed further in the light of the effects of using an incomplete basis (represented by a truncated complete set), in which case the two methods are no longer equivalent. Features of both methods are combined in a proposed hybrid method of arriving at a matrix representation.

2. Expansion of the Eigenfunctions

The linear homogeneous differential equation:

$$\mathcal{D}\psi_i = \lambda_i\psi_i \quad (1)$$

may be solved by a series expansion:

$$\psi_i = \sum_{j=1}^{\infty} c_j^i \Phi_j. \quad (2)$$

\mathcal{D} is a linear differential operator with respect to variables r , ψ_i is one of its eigenfunctions, and λ_i is the corresponding eigenvalue, the index i specifying a particular eigensolution. The Φ_j are known functions of r , and the coefficients c_j^i are to be determined.

For the expansion (2) to be an exact solution of the differential Eq. (1), the set $\{\Phi_j; j = 1 \text{ to } \infty\}$ must be a complete set of functions in the space spanned by the variables r . The functions Φ_j are required to be linearly independent, so that none of the coefficients c_j^i are redundant. Two methods of determining the coefficients c_j^i are widely used: in this discussion, they will be referred to as the integral and the non-integral methods. The integral method is usually called the Ritz variational method [2] in the case that \mathcal{D} is a Hermitian operator, as will henceforth be assumed herein. The non-integral method is known as the power

series method, or as the method of Frobenius [6], in the particular case where \mathcal{D} is an ordinary differential operator, and the Φ_j are powers of the independent variable. Both methods have as their first step, the substitution of the expansion (2) into the differential Eq. (1), to produce:

$$\sum_{j=1}^{\infty} c_j^i (\mathcal{D} - \lambda_i) \Phi_j = 0. \quad (3)$$

3. The Integral Method

The integral method proceeds by multiplication of (3) from the left by each member of the set $\{\Phi_k^*; k = 1 \text{ to } \infty\}$ (Φ_k^* denotes the complex conjugate of Φ_k), followed by integration over r . The resulting set of equations may be written compactly as the matrix eigenvalue equation [9, 10]:

$$H c^i = \lambda_i S c^i. \quad (4)$$

H and S are infinite square matrices with elements:

$$H_{kj} = \int \Phi_k^* \mathcal{D} \Phi_j \, dr, \quad S_{kj} = \int \Phi_k^* \Phi_j \, dr \quad (5)$$

for k and $j = 1 \text{ to } \infty$. c^i is the infinite column vector of the coefficients c_j^i .

4. The Non-Integral Method

The second step of the non-integral method is the elimination of all functions of r (other than members of the set $\{\Phi_k; k = 1 \text{ to } \infty\}$) from (3). That is $\mathcal{D} \Phi_j$ is written as a linear sum of the Φ_k :

$$\mathcal{D} \Phi_j = \sum_{k=j-l}^{j+m} A_{kj} \Phi_k. \quad (6)$$

The coefficients A_{kj} are determined by algebraic properties of the functions Φ_k , such as recurrence relations [11]. The operators \mathcal{D} occurring in quantum mechanics are often simple rational functions of the variables r , so that the expansion (6) of $\mathcal{D} \Phi_j$ may contain only a small number of terms: that is the integers l and m in (6) are usually finite and small.

Substitution of (6) into (3) produces:

$$\sum_{j=1}^{\infty} c_j^i \left[\sum_{k=j-l}^{j+m} (A_{kj} - \lambda_i \delta_{kj}) \Phi_k \right] = 0. \quad (7)$$

Since the c_j^i , A_{kj} , and λ_i , are independent of the variables r , and since the Φ_k are a linearly independent set of functions, (7) can only be satisfied for general values of the variables r , if the coefficient of each Φ_k in (7) is identically zero. Interchanging the order of the summations in (7), this condition produces the matrix equation:

$$A c^i = \lambda_i I c^i. \quad (8)$$

The elements of the infinite square matrix A are the coefficients A_{kj} of (6) and (7) in the range $j-l \leq k \leq j+m$, and are otherwise zero. I is the infinite identity matrix: $I_{kj} = \delta_{kj}$. The infinite column vector c^i contains the coefficients c_j^i as in (4).

5. Relationship between the two Methods

What has been said above is well known: it has been reproduced for the sake of clarity, and to establish notation. The relationship between the integral and non-integral methods may be established through their corresponding matrix Eqs. (4) and (8) respectively. Both of these equations determine the complete spectrum of eigenvalues of the original differential operator \mathcal{D} . Since the set $\{\Phi_j; j=1 \text{ to } \infty\}$ has been assumed to be complete, the eigenvalues λ_i of the infinite matrix Eqs. (4) and (8) are exact, and are therefore necessarily identical. The identification of the eigenvectors in Eqs. (4) and (8) is justified since the arbitrariness associated with phase, normalisation, and possible degeneracies, can be removed by auxiliary conditions (at our disposal) on the original eigenfunctions ψ_i , so that the coefficients c_j^i are then uniquely determined in the expansion (2).

Thus the infinite matrix Eqs. (4) and (8) determine the same eigenvalues λ_i and eigenvectors c^i , and yet they differ somewhat in form, and contain different matrices. Eq. (4) can be transformed into standard eigenvalue form by multiplication from the left by S^{-1} , which is possible since S is necessarily non-singular if the Φ_k are linearly independent as has been assumed. The result is:

$$S^{-1} H c^i = \lambda_i I c^i. \quad (9)$$

Eqs. (8) and (9) have the same form, and since they determine the same eigenvalues λ_i and eigenvectors c^i , it follows that they must be identical in every respect. This realisation leads to the identification:

$$A \equiv S^{-1} H \quad \text{or} \quad H \equiv S A. \quad (10)$$

This equation is the fundamental relationship between the integral operator matrix H , the overlap matrix S , and the non-integral operator matrix A , obtained by expanding the eigenfunctions ψ_i of the operator \mathcal{D} in the same basis set $\{\Phi_k; k=1 \text{ to } \infty\}$. It has been derived independently in a more abstract context by Löwdin [12].

6. Comparison of the Methods in General

The relationship (10) provides an insight into the merits of each method. Firstly it is noted that the operator matrices H and A , and also the corresponding eigenvalue Eqs. (4) and (8), are identical if the basis set $\{\Phi_k\}$ is orthonormal ($S_{kj} = \delta_{kj}; S \equiv I$): otherwise they are necessarily different. Clearly for an orthonormal basis set $\{\Phi_k\}$ the two methods are entirely equivalent. In this case the algebraic manipulations involved in deriving the elements of A are entirely equivalent to the task of evaluating the integrals H_{kj} (5) analytically. Henceforth only the case of a non-orthonormal basis is considered.

The Hermitian (or self-adjoint) character of \mathcal{D} is essentially an integral property [9] defined by the relation:

$$H_{kj}^* = H_{jk}. \quad (11)$$

In this case H is a Hermitian matrix by definition, and since S is always Hermitian (as is apparent from (5)), it follows from (4) that all the eigenvalues λ_i are necessarily real. However even though the operator \mathcal{D} is Hermitian, it follows from (10) that the non-integral operator matrix A is not Hermitian for a non-orthonormal basis. Thus A is generally non-Hermitian so that it is not apparent from (8) that the eigenvalues λ_i are real.

It is this non-Hermitian character (or non-symmetric character in the real case) of the A matrix, which is associated with the sparse structure (that is many identically zero elements) of A as implied by (6). A is described in terms of the integers l and m in (6), as a band matrix with $l+m$ bands parallel to the principal diagonal. The band structure of A is generally not symmetrical: that is the integers l and m in (6) are generally not equal. If A is a triangular matrix (l or $m=0$ in (6)) an analytical solution for the eigenvalues λ_i and eigenvectors c^i is obtained. In particular if A is an upper triangular matrix ($m=0$) the expansion (2) of the eigenfunction ψ_i contains a finite number of non-zero coefficients c_j^i .

Most of the exactly soluble problems of classical and quantum mechanics [13, 14], and of mathematical analysis [5], correspond to upper triangular A matrices with a single band of non-zero elements ($l=1$ in (6)) above the principal diagonal. These very sparse A matrices arise in one-dimensional problems by expansion of the eigenfunctions ψ_i in a power series of the independent variable. Generally all the elements of the overlap matrix S of a power series are non-zero, so that from (10) all the elements of H are also non-zero even though A is a band matrix of width two. Thus the integral (Ritz) method would not be suitable for the solution of these exactly soluble problems, since the simplicity of the eigenvectors would not be apparent in the resulting matrix Eq. (4).

The respective merits of Hermiticity for the integral method, and of the sparseness for the non-integral method, apply only with regard to the determination of the eigensolutions of the resulting matrix Eqs. (4) and (8): that is the actual convergence of the expansion (2) is the same for both methods, since it is determined by the choice of the basis functions Φ_k . However the non-integral method may be a guide to the choice of basis, since a basis producing an approximately upper triangular A matrix (in the sense that the non-zero elements below the diagonal are few and/or small in magnitude), is one approximating to an analytical solution with a finite number of coefficients c_j^i . The integral method cannot yield such an insight into the convergence of c^i , since it only gives an analytical solution in the trivial case where the basis is the set of eigenfunctions of $\mathcal{D}\{\psi_i\}$.

7. Truncation

The most common method of obtaining an approximate eigensolution in numerical calculations is to truncate the expansion (2) at a finite number of terms, the presumption being that the discarded terms make a contribution to ψ_i ,

which is negligible compared with the desired accuracy of the approximate solution. Although the infinite matrix Eqs. (4) and (8) are equivalent, this is no longer true when they are truncated at the same finite number of terms n . This non-equivalence is demonstrated by comparing the equations expressing the matrix multiplications in (10) for the non-truncated expansion:

$$H_{kj} = \sum_{i=j-1}^{j+m} S_{ki} A_{ij}, \quad A_{kj} = \sum_{i=1}^{\infty} S_{ki}^{-1} H_{ij} \quad (12)$$

and for the truncated expansion:

$$H_{kj} = \sum_{\substack{i=j-1 \\ \text{but } i \leq n}}^{j+m} S_{ki} A_{ij}, \quad A_{kj} = \sum_{i=1}^n S_{ki}^{-1} H_{ij}. \quad (13)$$

Thus in general the matrix Eqs. (4) and (8) are transformable into each other only by infinite matrix multiplications. The two truncated matrix Eqs. (4) and (8) are only equivalent if the overlap matrix S is diagonal, corresponding to the basis set $\{\Phi_k\}$ being orthogonal.

In the case where \mathcal{D} is a Hermitian operator, the approximate eigenvalues $\lambda_i^{(n)}$ obtained by truncation of (4) to order n are real, and are variational (upper bounds to the exact eigenvalues $\lambda_i^{(\infty)}$ with respect to the set of coefficients $\{c_j^i; j = 1 \text{ to } n\}$ retained) [15]. There is no variational principle for the eigenvalues of Eq. (8) for a finite order of truncation n , since the non-integral operator matrix A is in general non-Hermitian. Because of this non-Hermiticity, the truncated A matrix (formed by the non-integral method, and then truncated in (8)) may have some complex eigenvalues, even though the exact eigenvalues $\lambda_i^{(\infty)}$ of \mathcal{D} (and of the infinite A matrix) are necessarily real. Comparing (12) and (13), it is apparent that if the truncated non-integral operator matrix $A^{(n)}$ was formed from the second of Eqs. (13) from $S^{(n)}$ and $H^{(n)}$, rather than from (12) (which is equivalent to the derivation of $A^{(n)}$ by the non-integral method), that its eigenvalues would be identical with those of (4) truncated to the same order, so that they would be real and variational even though $A^{(n)}$ generated in this way would still not be Hermitian. Thus it is apparent, that it is the terms excluded in (13), but retained in (12) ($i = n + 1$ to ∞), which make $A^{(n)}$ essentially non-Hermitian in the sense of possibly having some complex eigenvalues.

8. Comparison of the Methods in Computations

The Hermitian form of the eigenvalue Eq. (4) in the integral method is an advantage in numerical determinations of the eigenvalues, since the available eigenvalue algorithms [20] are more efficient for Hermitian matrices than for general matrices. Two demerits of the integral method are that a considerable amount of work may be involved in the evaluation of the matrix elements (5) (especially the H_{kj}), and that the resulting matrix Eq. (4) is not in standard eigenvalue form. Although computational methods for simultaneously diagonalising the H and S matrices of (4) are being developed [21, 22], the usual numerical methods (Jacobi, Givens, Householder) [20] operate on a matrix

eigenvalue equation in standard form. Usually (4) is transformed into standard form by the Choleski decomposition [20], before computing the eigenvalues.

Three advantages of the non-integral method are that: no integrals have to be evaluated in setting up the matrix A ; the matrix Eq. (8) is already in standard eigenvalue form; and the usual sparse (band) structure of the A matrix facilitates the determination of the eigenvalues, since every zero element corresponds to a partial reduction to diagonal form. The principal disadvantage of the non-integral method is that the resulting matrix A is not Hermitian, so that the eigenvalues are in general complex, which complicates the numerical algorithms.

9. A Hybrid Method

It is possible through the relationship (10), to combine some of the advantages of the two methods into a hybrid method, which overall may be simpler in some cases than either of the original methods. This hybrid method will result in a matrix equation, which is both Hermitian (for a Hermitian operator \mathcal{D}), and which is in standard eigenvalue form. In addition the evaluation of integrals will be avoided as far as possible.

An essential part of the hybrid method is the transformation of (4) to standard eigenvalue form by the Choleski decomposition [20] of S , which is alternatively known as the square root method [23]. This is a non-iterative numerical procedure, which is represented by the equations:

$$G = THT^\dagger, \quad (14)$$

$$d^i = U^\dagger c^i, \quad (15)$$

where $T = U^{-1}$, and \dagger denotes the Hermitian conjugate: that is complex conjugation plus transposition. The transformation matrix U is defined to be an upper triangular matrix satisfying:

$$S = UU^\dagger. \quad (16)$$

The transformation changes (4) into:

$$Gd^i = \lambda_i d^i. \quad (17)$$

It can be shown that if H is Hermitian, then so is G , so that the transformation produces a Hermitian eigenvalue Eq. (17) in standard form. The transformation is only unique [20] if U is restricted to be an upper triangular matrix satisfying (16).

In the usual numerical procedure [20, 23], the matrix U of (15) and (16) is defined to be a lower triangular, rather than an upper triangular, matrix, which is equivalent to reversing the order of the functions Φ_j in the set $\{\Phi_j; j = 1 \text{ to } n\}$. The transformation represented by Eqs. (14) to (17) is equivalent to the Schmidt orthonormalisation [24] of the basis set $\{\Phi_j\}$ taken in the original order, the elements d_j^i of d^i being the coefficients of ψ_i in the new orthonormal basis. However a practical consideration is that it may be more expeditious to do the orthonormalisation implicitly by the numerical square root method, than to orthonormalise the original basis algebraically by the Schmidt process.

In the hybrid method the operator matrix G is derived from the non-integral operator matrix A , rather than from H via (14). This is possible, since substitution of (10) into (14) using (16) produces:

$$G = U^\dagger A T^\dagger. \quad (18)$$

Thus the hybrid method involves the following steps:

- (i) evaluation of A algebraically from \mathcal{D} and the chosen basis $\{\Phi_j\}$ by the non-integral method (section 4);
- (ii) evaluation of S by computation of the overlap integrals S_{kj} of (5);
- (iii) numerical computation of U^\dagger and T^\dagger by the square root method;
- (iv) transformation of A into G by (18); and
- (v) determination of the eigenvalues λ_i and eigenvectors d^i of G by solution of (17).

An advantage of the hybrid method is that only the overlap integrals S_{kj} have to be evaluated: no operator integrals H_{kj} are involved. Another advantage is that only two of the four transformation matrices U , U^\dagger , T , T^\dagger , are involved in (15) and (18) compared with three of them in (14) and (15). This reduces somewhat the computing involved in the transformation [22]. A disadvantage is that the matrix multiplications implicit in (18) must be performed before truncation: that is the matrix multiplications must be continued beyond the truncation order n , until the matrix elements of G have converged to the required precision.

A special case worth noting is where the basis $\{\Phi_k; k = 1 \text{ to } \infty\}$ is orthogonal, but not normalised. Such a basis may be particularly suitable for the hybrid method, since the problem of convergence in the matrix multiplications implicit in (18) disappears (in fact the transformation from A to G is very simple, since the U matrices are all diagonal), and since an unnormalised orthogonal set of functions can usually be chosen to have simpler algebraic properties (utilised in setting up A), than the corresponding set of normalised functions (for example: the Legendre polynomials [11]).

10. Applications in Quantum Mechanics

A well known application of the non-integral method in quantum mechanics is Pekeris's solution of the Schrödinger equation for the helium atom [25]. This application involved the special case of an orthonormal basis set, in which the truncated non-integral and Ritz variational methods are entirely equivalent, as Pekeris realised. Nevertheless he chose to derive the matrix Eq. (4), or (8), by the non-integral method, rather than by the Ritz method.

The non-integral method has been used implicitly (as the analytical power series method) in the solution of the exactly soluble problems of quantum mechanics [14]. It has also been used implicitly to obtain semi-analytical solutions of the two-centre Kepler problem, the eigenvalues being obtained numerically through the theory of continued fractions [17, 18]. Recently the non-integral method has been formulated explicitly in matrix notation for this problem [19], and for the anharmonic oscillator problem [16], in which cases the truncated matrices have same spurious complex eigenvalues.

The ideas discussed in this paper, and in particular the hybrid method, are being applied by the author to the two-centre Kepler problem of wave mechanics in order to produce better (in the sense of more rapid convergence, and the non-occurrence of spurious complex eigenvalues) semianalytical solutions than those currently available [17–19].

References

1. Dirac, P. A. M.: The principles of quantum mechanics, 4th ed. Oxford: Clarendon Press 1958.
2. Margenau, H., Murphy, G. M.: The mathematics of physics and chemistry, Vol. 1, 2nd ed, p. 377. Princeton: D. Van Nostrand 1956.
3. Pauling, L., Wilson, E. B.: Introduction to quantum mechanics. New York: McGraw-Hill 1935.
4. Ref. [2], Chapter 7.
5. Whittaker, E. T., Watson, G. N.: A course of modern analysis, 4th ed. Cambridge: University Press 1927.
6. Piaggio, H. T. H.: An elementary treatise on differential equations and their applications, 4th ed, Chapter IX. London: G. Bell and Sons 1960.
7. Eyring, H., Walter, J., Kimball, G. E.: Quantum chemistry. New York: John Wiley and Sons 1944.
8. Messiah, A.: Quantum mechanics, Volumes 1 and 2. Amsterdam: North-Holland 1966.
9. Hameka, H. F.: Introduction to quantum theory, Chapter 11. New York: Harper and Row 1967.
10. Ref. [2], Chapter 8, Sections 7 and 8.
11. Ref. [2], Chapter 3.
12. Löwdin, P. O.: Linear algebra and the fundamentals of quantum theory. Preprint No. 125 (July 1, 1969). Uppsala Quantum Chemistry Group.
13. Ref. [2], Chapters 3 and 11.
14. Buckingham, R. A.: Exactly soluble bound state problems, Chapter 3 of: Quantum theory, I. Elements, edited by D. R. Bates. New York: Academic Press 1961.
15. Ref. [2], Chapter 11, Section 20.
16. Hunter, G.: M. Sc. Thesis, Victoria University of Manchester, England (1962).
17. Baber, W. G., Hasse, H. R.: Proc. Cambridge philos. Soc. **31**, 564 (1935).
18. Bates, D. R., Ledsham, Kathleen, Stewart, A. L.: Philos. Trans. Roy. Soc. (London) A **246**, 215 (1953).
19. Hunter, G., Pritchard, H. O.: J. chem. Physics **46**, 2146 (1967).
20. Wilkinson, J. H.: The algebraic eigenvalue problem. Oxford: Clarendon Press 1965.
21. Wallis, A.: Ph. D. Thesis, Victoria University of Manchester, England (1968).
22. McElwain, S., Wallis, A., Pritchard, H. O.: The variation method and the algebraic eigenvalue problem. Int. J. quant. Chem. III. 711 (1969).
23. Fadeev, D. K., Fadeeva, V. N.: Computational methods of linear algebra, p. 144. San Francisco: W. H. Freeman 1963.
24. Hall, G. G.: Matrices and tensors, p. 51. Oxford: Pergamon Press 1963.
25. Pekeris, C. L.: Physic. Rev. **112**, 1649 (1958).

Professor G. Hunter
Centre for Research in Experimental Space Science
York University
Toronto, Canada