Numerical Method for Unitary Systems

EDWARD J. SHIPSEY

Department of Physics, University of Texas, Austin, Austin, Texas 78712

Received July 8, 1985; revised January 15, 1986

A method of constructing probability amplitudes is presented for cases in which the corresponding unitary propagator can be expressed in terms of the Cayley representation. The method is most appropriate when the Hermitian matrix in the Cayley representation is sparse. The algorithm is organized for minimum storage. © 1986 Academic Press, Inc.

I. INTRODUCTION

A unitary matrix may be expressed in the Cayley representation as

$$\mathbf{U} = (\mathbf{1} + i\mathbf{H})^{-1}(\mathbf{1} - i\mathbf{H}), \tag{1}$$

where **1** is the identity matrix, i the square root of minus one, and **H** is Hermitian. If this form is to have any real utility an efficient method of inversion is required. The object of the present work is to present the algorithm, CAYLLU (Cayley–Lanczos–LU decomposition), which accomplishes this in the special situation in which the action of **H** (which can just as well be considered to be an abstract operator) on a vector is simple to describe. **H**, for example, might be represented as a sparse matrix, in which case storage is required only for the nonzero elements.

The time-dependent Schrodinger equation

$$-\frac{h}{2\pi i}\frac{\partial}{\partial t}\Psi = \mathscr{H}\Psi \tag{2}$$

can be approximated in this manner. The equation is integrated from t_1 to t_2, and $\Psi(t)$ is replaced by $1/2[\Psi(t_1) + \Psi(t_2)]$ in the integrand. Defining $x = \Psi(t_2)$, $f = \Psi(t_1)$, and

$$H = \int_{t_1}^{t_2} \frac{\pi}{h}\mathscr{H}\, dt \tag{3}$$

results in

$$(\mathbf{1} + i\mathbf{H})\mathbf{x} = (\mathbf{1} - i\mathbf{H})\mathbf{f}. \tag{4}$$

0021-9991/86 $3.00

Usually, a transformation reducing the size of the diagonal elements is performed on $\mathscr{H}$, so that $\mathbf{H}$ may depend on time. This approximation is closely related to the Crank–Nicolson scheme which has enjoyed great success in many other applications. More elaborate expressions for H are possible [1].

The procedure starts with the Lanczos reduction to tridigiagonal form [2]. Subsequent developments follow closely the work of Paige and Saunders [3] and Widlund [4]. There is really nothing special in the present application about the Lanczos reduction so that previous work can be consulted with regard to the numerical aspects. The Lanczos reduction in exact arithmetic terminates in N steps where N is the dimension of the space. In many cases many less steps will suffice. A detailed analysis of this aspect of the Lanczos reduction is beyond the scope of the present work. Some numerical examples will be given, however, to give some indication of this behavior. Refinements in the Lanczos reduction are possible [5, 6, 7], particularly in applications to eigen value problems [8].

An important feature of the recent developments in the Lanczos method is that the algorithms (again provided that $\mathbf{H}$ can be conveniently applied to a vector) are organized in such a manner that storage requirements have been greatly reduced over the original method. The procedures are also organized to provide a means of stopping the process, when this is reasonable, before the full N steps have been carried out. The procedure is monitored by means of the magnitude of the residual error vector

$$11\mathbf{r}_a 11 = (\mathbf{r}_a^*, \mathbf{r}_a)^{1/2}, \tag{5}$$

where

$$\mathbf{r}_a = (\mathbf{I} - i\mathbf{H})\mathbf{f} - (\mathbf{I} + i\mathbf{H})\mathbf{x}_a, \tag{6}$$

$\mathbf{x}_a$ is some approximate solution to Eq. (4) and the (,) indicates the usual scalar dot product. If $\mathbf{x}$ is the true solution to Eq. (4) the error in the approximation is

$$\mathbf{e}_a = \mathbf{x} - \mathbf{x}_a \tag{7}$$

or

$$\mathbf{e}_a = (\mathbf{I} + i\mathbf{H})\mathbf{r}_a. \tag{8}$$

From this it readily follows that

$$11\mathbf{e}_a 11 < 11\mathbf{r}_a 11. \tag{9}$$

This condition, of course, is only true for the present system. It is a special property of Eq. (4) which makes a stopping procedure based on the norm of the residual error vector particularly useful.

II. The Algorithm CAYLLU

The Lanczos reduction to trigidiagonal form is accomplished by constructing a sequence of orthonormal vectors, $\mathbf{v}_j$, by means of the recurrence relation

$$\beta_{j+1}\mathbf{v}_{j+1} = \mathbf{H}\mathbf{v}_j - \alpha_j\mathbf{v}_j - \beta_j\mathbf{v}_{j-1} \tag{10}$$

with

$$\mathbf{v}_0 = 0 \tag{11}$$

and

$$(\mathbf{v}_k^*, \mathbf{v}_j) = \delta_{jk}. \tag{12}$$

Since the phase of $\mathbf{v}_j$ is arbitrary the β_j's can be taken to be real and greater than or equal to zero. Equation (12) gives

$$\alpha_j = (\mathbf{v}_j^*, \mathbf{H}\mathbf{v}_j) \tag{13}$$

which shows that the α_j's are real since $\mathbf{H}$ is Hermitian. Besides being matrix elements, the β_j's can also be regarded as being normalization constants. Taking the latter viewpoint, the vector $\mathbf{u}_{j+1}$ is defined by

$$\mathbf{u}_{j+1} = \mathbf{H}\mathbf{v}_j - \alpha_j\mathbf{v}_j - \beta_j\mathbf{v}_{j-1} \tag{14}$$

so that

$$\beta_{j+1}^2 = (\mathbf{u}_{j+1}^*, \mathbf{u}_{j+1}). \tag{15}$$

The positive root is chosen for β_{j+1}. The procedure can be written compactly as

$$\mathbf{HV} = \mathbf{VT}, \tag{16}$$

where $\mathbf{V}$ is an $N \times N$ matrix whose columns consist of the N orthonormal vectors $\mathbf{v}_k$ and $\mathbf{T}$ is the symmetric tridigiagonal matrix whose diagonal elements are the α_j's and whose off-diagonals are the $\beta_{j's}$. That is,

$$T_{jj} = \alpha_j \tag{17}$$

and

$$T_{j,j+1} = \beta_j. \tag{18}$$

The unknown vector $\mathbf{x}$ in Eq. (4) is written in terms of a new unknown vector $\mathbf{c}$ by

$$\mathbf{x} = \mathbf{Vc}. \tag{19}$$

The known vector, **f**, in Eq. (4) is taken to define β_1, $\mathbf{v}_1$, that is

$$\mathbf{f} = \beta_1 \mathbf{v}_1 \tag{20}$$

with

$$\beta_1^2 = (\mathbf{f}^*, \mathbf{f}). \tag{21}$$

In most applications **f** will be already normalized so that β_1 can usually be set equal to one. Substituting these expressions into Eq. (4) gives

$$(\mathbf{I} + i\mathbf{T})\mathbf{c} = \beta_1(1 - i\alpha_1)\mathbf{e}_1 - i\beta_1\beta_2\mathbf{e}_2, \tag{22}$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are N-dimensional vectors whose only non-zero elements are in the first and second positions, respectively.

The second part of the CAYLLU algorithm is the LU decomposition of $\mathbf{I} + i\mathbf{T}$. Writing

$$\mathbf{LU} = \mathbf{I} + i\mathbf{T} \tag{23}$$

with

$$\begin{aligned} L_{jj} &= \rho_j, \\ L_{j,j-1} &= \sigma_j, \\ U_{jj} &= 1, \\ U_{j,j+1} &= \tau_j, \end{aligned} \tag{24}$$

and all other elements zero gives

$$\rho_1 = 1 + i\alpha_1, \tag{25}$$

$$\rho_j = 1 + i\alpha_j + \beta_j^2/\rho_{j-1}, \qquad j > 1, \tag{26}$$

$$\sigma_j = i\beta_j, \tag{27}$$

$$\tau_j = i\beta_{j+1}/\rho_j. \tag{28}$$

These relations give for the real part of ρ_j

$$\begin{aligned} \operatorname{Re} \rho_1 &= 1 \\ \operatorname{Re} \rho_j &= 1 + \left| \frac{\beta_j}{\rho_{j-1}} \right|^2 \operatorname{Re} \rho_{j-1} \end{aligned} \tag{29}$$

from which it is observed that

$$\operatorname{Re} \rho_j \geqslant 1 \tag{30}$$

so that the effective pivot in the LU decomposition cannot vanish. Substituting the LU decomposition into Eq. (22) gives

$$\mathbf{LUc} = \beta_1(1 - i\alpha_1)\mathbf{e}_1 - i\beta_1\beta_2\mathbf{e}_2. \tag{31}$$

The economy of storage is achieved by a device similar to that of Paige and Saunders [3]. Define

$$\mathbf{a} = \mathbf{Uc} \tag{32}$$

and the matrix $\mathbf{Y}$ by means of

$$\mathbf{YU} = \mathbf{V}. \tag{33}$$

The elements of $\mathbf{a}$ are easily obtained to give

$$a_1 = (1 - i\alpha_1)\,\beta_1/\rho_1, \tag{34}$$

$$a_2 = (-\sigma_2 a_1 - i\beta_1\beta_2)/\rho_2, \tag{35}$$

and

$$a_j = -\sigma_j a_{j-1}/\rho_j, \qquad j > 2. \tag{36}$$

Similarly the columns of $\mathbf{Y}$ are obtained to give

$$\mathbf{y}_1 = \mathbf{v}_1, \tag{37}$$

$$\mathbf{y}_j = \mathbf{v}_j - \tau_{j-1}\mathbf{y}_{j-1}, \qquad j > 1. \tag{38}$$

Finally the unknown $\mathbf{x}$ is given by

$$\mathbf{x}_n = \sum_{j=1}^{n} a_n \mathbf{y}_n. \tag{39}$$

In exact arithmetic $\mathbf{x}_n$ becomes exact when $n = N$, the dimension of the system. The algorithm can be truncated at some smaller value of n in which case, by methods similar to those of Paige and Saunders the residual is

$$11\mathbf{r}_n 11 = \beta_{n+1}\,|a_n|, \qquad n > 2. \tag{40}$$

This expression, in view of Eq. (9), provides a means of stopping the procedure at some preassigned upper limit on the error in $\mathbf{x}_n$.

The economy of storage comes about by accumulating $\mathbf{x}_n$ as the a_j's and $\mathbf{y}_j$'s become available. From Eqs. (10), (12), (13), (14), and (15) it is seen that only a few $\mathbf{v}_j$'s need be saved fo the Lanczos tridiagonalation and from Eqs. (37) and (39) that a similar situation holds for the $\mathbf{y}_j$'s. By means of replacing components of vectors no longer needed by components of new vectors and careful organization, $\mathbf{x}_n$

can be computed using, besides $\mathbf{x}_n$ itself, only four other vectors. Another interesting point is that Eq. (22) is actually of the same Cayley form as is Eq. (4) so that in the tridiagonalization the length of $\mathbf{x}_n$ will only be affected by errors in $\mathbf{V}$. This, in fact, was the reason for the choice of $\mathbf{v}_1$ made in Eq. (20).

III. Numerical Example

The unitary nature of the system provides some simple means of checking. The length of $\mathbf{x}_a$ should equal that of $\mathbf{f}$ in Eq. (4), but the algorithm has built in features that decrease the sensitivity of this test. Another approximation $\mathbf{x}'_a$ can be found to a similar system involving a vector $\mathbf{f}'$ orthogonal to $\mathbf{f}$ and the corresponding orthogonality of $\mathbf{x}_a$ and $\mathbf{x}'_a$ tested. Finally if $\mathbf{H}$ and $\mathbf{f}$ are real, replacing $\mathbf{f}$ by $\mathbf{x}_a^*$ in Eq. (4) and applying the algorithm a second time should return the original $\mathbf{f}$ vector. These tests have all been made with satisfactory results.

The most useful test, in view of Eq. (9) involves $11\mathbf{r}_a 11$. A reasonable tolerance, ε, is assigned and the procedure stopped when $11\mathbf{r}_a 11 < \varepsilon$. If the number of iterations, K, is significantly less than N, particularly for a system of the size to be discussed, reasonable assurance can be had that the process is working properly. A further check has been the good agreement between $11\mathbf{r}_a 11$ computed from Eqs. (6) and (40).

The test problem has been taken to be a very simple system with two degrees of freedom. For clarity of presentation the index on $\mathbf{x}$ or $\mathbf{f}$ will be made double. That is,

$$x_l \rightarrow x_{jk} \tag{41}$$

with

$$l = j + k(N-1), \tag{42}$$

where N^2 is the dimension of the system. In the example, N is 80. $\mathbf{H}$ is taken to be real and to have the simple form

$$H_{j,k;j+1,k} = H_{j+1,k;jk} = g(j)^{\eta}, \tag{42}$$

$$H_{jk;j,k+1} = H_{j,k+1,jk} = g(k)^{\eta}, \tag{43}$$

and all other elements of $\mathbf{H}$ are zero. For $\eta = 1/2$ the system is roughly similar to two coupled harmonic oscillators. The vector initially has all components equal to zero except $j = j_0$ and $k = k_0$.

Results are shown in Table I. These are calculations which from all appearances are entirely adequate and are intended to illustrate situations in which the algorithm works properly. In many cases the algorithm was applied successively, the number of times being indicated by N_{steps}. The average number of individual iterations (n in Eq. 39) required to reduce $11\mathbf{r}_n 11$ below ε is denoted by K_{ave}. This

TABLE I

Performance of CAYLLU

j_0	k_0	η	g	$-\log \varepsilon$	K_{ave}	N_{steps}	$P(j_0, k_0)^a$
1	2	1	0.01	7	12	100	4.1(−3)
80	80	1	0.01	7	49	100	2.5(−5)
1	2	1	0.02	7	26	50	4.1(−3)
80	80	1	0.02	7	94	50	5.7(−4)
1	2	1	0.04	7	73	25	4.1(−3)
80	80	1	0.04	7	183	25	1.2(−3)
1	2	1	0.1	7	97	1	0.80
80	80	1	0.1	7	443	1	0.64
1	2	1	0.1	5	34	1	0.80
80	80	1	0.1	5	329	1	0.64
1	2	0.5	0.1	7	24	10	3.5(−3)
80	80	0.5	0.1	7	58	10	5.8(−2)
1	2	0.5	0.5	7	167	1	5.9(−2)
80	80	0.5	0.5	7	271	1	0.44
1	2	0.5	0.5	5	115	1	5.9(−2)
80	80	0.5	0.5	5	204	1	0.44

[a] The number in parentheses is the power of 10 by which the number is to be multiplied.

number is seen to be well below N^2. The final column in Table I labelled $P(j_0, k_0)$ is the "probability" (i.e., $a_{jk}^* a_{jk}$) that the initial state will be occupied after N_{steps}. That $P(j_0, k_0)$ is small in many cases is indicative of "strong coupling." The converse, however, need not be true, since in Eq. (4) if **H** becomes very large in some sense compared to the unit matrix **I**, **x** becomes nearly equal to **f**. A few cases approaching this behavior can be seen in Table I.

Acknowledgments

The support and encouragement of Professor J. C. Browne is deeply appreciated. This research was supported by the Department of Energy under Grant DE-AS05-81ER10987.

References

1. E. J. Shipsey, *J. Chem. Phys.* **70** (1979), 5281.
2. C. Lanczos, *J. Res. Nat. Bur. Stand.* **45** (1950), 255.
3. C. C. Paige and M. A. Saunders, *SIAM J. Numer. Anal.* **12** (1975), 617.
4. O. Widlund, *SIAM J. Numer. Anal.* **15** (1978), 801.
5. B. N. Parlett and D. S. Scott, *Math. Comput.* **33** (1979), 217.
6. B. N. Parlett, *Linear Algebra Appl.* **29** (1980), 323.
7. H. D. Simon, *Math. Comput.* **42** (1984), 115.
8. B. N. Parlett and B. Nour-Omid, *Linear Algebra Appl.* **68** (1985), 179.