An improved procedure for solving asymptotic equations

1997 J. Phys. A: Math. Gen. 30 1731

(http://iopscience.iop.org/0305-4470/30/5/033)

J. Phys. A: Math. Gen. **30** (1997) 1731–1737. Printed in the UK PII: S0305-4470(97)77501-2

An improved procedure for solving asymptotic equations

M R H Rudge† and D M Tiernan‡§

Department of Applied Mathematics and Theoretical Physics, The Queens University of Belfast, Belfast BT7 1NN, UK

Received 28 August 1996, in final form 28 November 1996

Abstract. In a previous publication a method was described for solving the close-coupling equations that arise in the non-relativistic scattering theory in the asymptotic region where the scattered particle is far removed from the residual atom, ion or molecule. In this paper an improved numerical procedure, based on an algorithm for solving the Sylvester problem, is described whereby a large set of such equations may be solved without incurring storage problems. Some typical results are presented that indicate the accuracy of the method and make comparisons with the results obtained from two other codes.

1. Introduction

In a previous publication (Rudge and Tiernan 1994) we described a numerical procedure for solving the coupled differential equations that arise in scattering excitation calculations in the asymptotic region defined as that region wherein only the long range r^{-n} coupling between the scattering channels is significant. The form of these equations, below break-up thresholds, is the same for the collision of a charged particle with atomic, ionic or molecular species and obtaining accurate numerical solutions to them is a significant part of the corresponding full scattering calculation. The method previously described was a simple one in which the problem was reduced to solving linear equations and was shown to be capable of high accuracy. It also generated an analytic form for the solutions which may be of value, for example, in calculating transition probabilities. The principal disadvantage of the method however is that the amount of computer storage required becomes a limiting feature as the number of channels becomes large. It is the purpose of the present paper to address this problem. We show that the equations can be reformulated in such a way that storage requirements are greatly reduced and then solved using the QZ algorithm of Moler and Stewart (1973). We thereby achieve two complementary schemes. In the first, where the number of channels is small enough to pose no storage problem, the method is formulated through a non-iterative solution of linear equations. The second scheme described here is used for large problems and uses an iterative numerical solution. Some typical problems addressed by other authors have been used to test the procedures.

† E-mail address: m.rudge@qub.ac.uk
‡ E-mail address: d.tiernan@nortel.co.uk
§ Current address: Nortel Technology, London Road, Harlow, Essex CM17 9NA, UK.

2. The scattering equations

Let

$$\mathcal{L}_0 = \mathrm{diag}\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + k_j^2\right) \tag{1}$$

and

$$\mathbf{V} = \sum_{\lambda=1}^{N_\lambda} r^{-\lambda}\mathbf{v}_\lambda \tag{2}$$

where if N is the number of channels then $\mathbf{v}_\lambda$ are $N \times N$ matrices. The asymptotic equations to be solved are

$$\mathcal{L}\mathcal{F} = (\mathcal{L}_0 + \mathbf{V})\mathcal{F} = \mathbf{0}. \tag{3}$$

For excitation problems the matrix $\mathbf{v}_1$ is simply a multiple of the unit matrix but we have chosen to generalize the equations so that $\mathbf{v}_1 = \mathrm{diag}(z_j)$ and define $\eta_j = z_j k_j^{-1}$. On writing

$$\theta_j = k_j r - \frac{\ell_j \pi}{2} + \eta_j \ln(2k_j r) + \arg\Gamma(\ell_j + 1 - \mathrm{i}\eta_j) \tag{4}$$

we can determine blocks of solution vectors, $\mathcal{F}_j$, such that

$$[\mathcal{F}_2 + \mathrm{i}\mathcal{F}_1]_{jk} \underset{r\to\infty}{\sim} \delta_{jk}\left[k_j^{-\frac{1}{2}} \exp(\mathrm{i}\theta_j)\right] \tag{5}$$

for $1 \leqslant j \leqslant N$ and $1 \leqslant k \leqslant N_o$, where N_o is the number of open channels (for which $k_j^2 > 0$). We can also obtain $N_c = N - N_o$ closed channel solution vectors specified by

$$[\mathcal{F}_3]_{jk} \underset{r\to\infty}{\sim} \delta_{jk}[\exp(-k_j r + \eta_j \ln(r))] \tag{6}$$

for $1 \leqslant j \leqslant N$ and $1 \leqslant k \leqslant N_c$ where $k_j = \left|k_j^2\right|^{\frac{1}{2}}$. In total there are thus $N_s = 2N_o + N_c$ solution vectors

$$\mathcal{F} = [\mathcal{F}_1 \mathcal{F}_2 \mathcal{F}_3] \tag{7}$$

that satisfy (3), where the first N_o columns are 'sine-like', the next N_o columns are 'cosine-like' and the last N_c channels are exponentially decreasing. It can be shown from the symmetry of $\mathbf{V}$ that the Wronskian matrix is

$$\mathbf{W} = \tilde{\mathcal{F}}\mathcal{F}' - \tilde{\mathcal{F}}'\mathcal{F} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{8}$$

In practice it is convenient to develop $2N_o$ solutions in the range $R_{cl} \leqslant r \leqslant R_\infty$ where R_{cl} is a radius at which the exponentially decreasing terms become significant. At $r = R_{cl}$ the Wronskian condition implies that

$$\begin{bmatrix} -\tilde{\mathcal{F}}_1' & \tilde{\mathcal{F}}_1 \\ -\tilde{\mathcal{F}}_2' & \tilde{\mathcal{F}}_2 \end{bmatrix} \begin{bmatrix} \mathcal{F}_3 \\ \mathcal{F}_3' \end{bmatrix} = [\,\mathbf{0}\,] \tag{9}$$

and that

$$\tilde{\mathcal{F}}_3\mathcal{F}_3' - \tilde{\mathcal{F}}_3'\mathcal{F}_3 = \mathbf{0} \tag{10}$$

which allows us to select N_c exponentially decreasing starting solutions.

3. The computational procedure

We can represent any one of the solution vectors, $\boldsymbol{f}$, in the range $R_1 \leqslant r \leqslant R_2$ by interpolation in a basis φ of size N_f as

$$\boldsymbol{f} = \mathbf{X}\varphi \tag{11}$$

where $\boldsymbol{f}$ is $N \times 1$ and φ is $N_f \times 1$ and $\mathbf{X}$ is an $N \times N_f$ matrix of coefficients. Equation (11) is the point of departure from our previous representation (equation (18) of Rudge and Tiernan, 1994)). The N_s solutions of the form (11) then comprise the columns of $\mathcal{F}$. In the collocation procedure the equation $\mathcal{L}\,\mathcal{F} = \mathbf{0}$ is satisfied over a discrete set of points r_j, $1 \leqslant j \leqslant N_p$ that lie in the range. Whilst defining

$$\Phi = [\varphi(r_1) \cdots \varphi(r_{N_p})] \tag{12}$$

we see that the collocation equations are

$$\mathbf{X}\Phi'' + \mathbf{K}\mathbf{X}\Phi + \sum_{\lambda=1}^{N_\lambda} \mathbf{v}_\lambda \mathbf{X}\Phi\mathbf{D}_\lambda = \mathbf{0} \tag{13}$$

where $\mathbf{K} = \mathrm{diag}(k_j^2)$ and $\mathbf{D}_\lambda = [\mathrm{diag}(r_j^{-\lambda})]$. The boundary conditions are

$$\begin{aligned} \mathbf{X}\Phi(R_2) &= \boldsymbol{f}(R_2) \\ \mathbf{X}\Phi'(R_2) &= \boldsymbol{f}'(R_2) \end{aligned} \tag{14}$$

where $\boldsymbol{f}(R_2)$ and $\boldsymbol{f}'(R_2)$ are known. Now if $N_p = N_f - 2$ and we redefine

$$\boldsymbol{\Phi}'' = [\boldsymbol{\Phi}''|\varphi(R_2)\varphi'(R_2)] \tag{15}$$

$$\boldsymbol{\Phi} = [\boldsymbol{\Phi}|\mathbf{0}\ \mathbf{0}] \tag{16}$$

$$\boldsymbol{\Phi}_\lambda = \boldsymbol{\Phi}[\mathbf{D}_\lambda|\mathbf{0}\ \mathbf{0}] \tag{17}$$

then (13) and the boundary conditions (14) can be rewritten in the form

$$\mathbf{A}_1\mathbf{X}\tilde{\mathbf{B}}_1 + \mathbf{A}_2\mathbf{X}\tilde{\mathbf{B}}_2 = \mathbf{C}(\mathbf{X}) \tag{18}$$

where

$$\mathbf{A}_1 = \mathbf{I} \qquad \mathbf{B}_1 = \tilde{\boldsymbol{\Phi}}'' \qquad \mathbf{A}_2 = \mathbf{K} \qquad \mathbf{B}_2 = \tilde{\boldsymbol{\Phi}} \tag{19}$$

and

$$\mathbf{C}(\mathbf{X}) = \left[-\sum_{\lambda=1}^{N_\lambda} \mathbf{v}_\lambda \mathbf{X}\Phi\mathbf{D}_\lambda \middle| \boldsymbol{f}(R_2)\boldsymbol{f}'(R_2) \right]. \tag{20}$$

Equations (18) are of the so-called Sylvester form (Sylvester 1884), and the Sylvester problem is to solve (18) for $\mathbf{X}$ given a fixed matrix $\mathbf{C}$. It can be seen that we need to store the matrices $\mathbf{A}_j$ which are $N \times N$ and the matrices $\mathbf{B}_j$ which are $N_f \times N_f$. The previous method, on the other hand, required the storage of a matrix of at least $(N_f \times N)^2$. For a 20 channel case in which typically $N_f \approx 200$, it can be seen that a major storage problem has been alleviated. Our previous method may be obtained if the elements of $\mathbf{X}$ are written as the components of a column vector. If storage permits there is an advantage in doing this because all N_s solutions can be obtained in one operation without iteration. On the other hand, since $\mathbf{C}$ depends on $\mathbf{X}$, it is necessary to solve (18) iteratively and a separate iteration must be performed for each channel. We have therefore sought to rewrite (18) in such a way that the iterations converge rapidly. We write

$$\mathbf{v}_\lambda = [\mathbf{v}_\lambda - \alpha_\lambda\mathbf{I}] + \alpha_\lambda\mathbf{I} \qquad \mathbf{D}_\lambda = [\mathbf{D}_\lambda - \beta_\lambda\mathbf{I}] + \beta_\lambda\mathbf{I} \tag{21}$$

and we again obtain equations (18) but now where

$$\mathbf{A}_1 = \mathbf{I} \qquad \tilde{\mathbf{B}}_1 = \Phi'' + \sum_{\lambda=1}^{N_\lambda} \alpha_\lambda \Phi \mathbf{D}_\lambda \tag{22}$$

$$\mathbf{A}_2 = \mathbf{K} + \sum_{\lambda=1}^{N_\lambda} \beta_\lambda (\mathbf{v}_\lambda - \alpha_\lambda \mathbf{I}) \qquad \tilde{\mathbf{B}}_2 = \Phi \tag{23}$$

and

$$\mathbf{C}(\mathbf{X}) = \left[-\sum_{\lambda=1}^{N_\lambda} [\mathbf{v}_\lambda - \alpha_\lambda \mathbf{I}] \mathbf{X} \Phi [\mathbf{D}_\lambda - \beta_\lambda \mathbf{I}] \middle| \boldsymbol{f}(R_2) \boldsymbol{f}'(R_2) \right]. \tag{24}$$

We have chosen the parameters β_λ as

$$\beta_\lambda = R^{-\lambda} \qquad R = \tfrac{1}{2}(R_1 + R_2). \tag{25}$$

The speed of convergence was not found to depend strongly on this particular choice of these parameters though there is a gain in efficiency through their use. In solving for solution $\boldsymbol{f}_j$ we chose

$$\alpha_\lambda = (\mathbf{v}_\lambda)_{jj}. \tag{26}$$

The equations may be simplified by diagonalizing the real symmetric matrix $\mathbf{A}_2$

$$\mathbf{A}_2 = \boldsymbol{S} \Lambda \tilde{\boldsymbol{S}} \qquad \boldsymbol{S} \tilde{\boldsymbol{S}} = \mathbf{I} \tag{27}$$

giving

$$\tilde{\boldsymbol{S}} \mathbf{X} \tilde{\mathbf{B}}_1 + \Lambda \tilde{\boldsymbol{S}} \mathbf{X} \tilde{\mathbf{B}}_2 = \tilde{\boldsymbol{S}} \mathbf{C}. \tag{28}$$

Using the QZ algorithm (cf Moler and Stewart 1973, Gardiner *et al* 1992a) we can write

$$\tilde{\mathbf{B}}_1 = \boldsymbol{Q} \boldsymbol{T} \tilde{\boldsymbol{Z}} \qquad \text{and} \qquad \tilde{\mathbf{B}}_2 = \boldsymbol{Q} \boldsymbol{V} \tilde{\boldsymbol{Z}} \tag{29}$$

where $\boldsymbol{Q}\tilde{\boldsymbol{Q}} = \boldsymbol{Z}\tilde{\boldsymbol{Z}} = \mathbf{I}$, $\boldsymbol{T}$ is upper-triangular and $\boldsymbol{V}$ is quasi upper-triangular. We have used the package of Gardiner *et al* (1992b) to perform this. We may note that step (29) is expedited by choosing to solve equations (18) in their transposed form which makes the dimension of the matrices involved as small as possible. The equations become

$$\mathbf{Y} \boldsymbol{T} + \Lambda \mathbf{Y} \boldsymbol{V} = \mathbf{G} \tag{30}$$

where $\mathbf{Y} = \tilde{\boldsymbol{S}} \mathbf{X} \boldsymbol{Q}$, $\mathbf{G} = \tilde{\boldsymbol{S}} \mathbf{C} \boldsymbol{Z}$ and are solved by the iteration

$$\mathbf{Y}_n \boldsymbol{T} + \Lambda \mathbf{Y}_n \boldsymbol{V} = \mathbf{G}(\mathbf{Y}_{n-1}). \tag{31}$$

The structure of $\boldsymbol{T}$ and $\boldsymbol{V}$ makes the rapid solution of (31) possible (Gardiner *et al* 1992a). In order to start the iteration we need

$$[\boldsymbol{f}\ \boldsymbol{f}'] = \mathbf{X}[\varphi \varphi'] \tag{32}$$

where again $\boldsymbol{f}$ is one of the $N_s = 2N_o + N_c$ linearly independent solution vectors of dimension N evaluated at the start radius. Hence

$$\begin{bmatrix} \tilde{\varphi} \\ \tilde{\varphi}' \end{bmatrix} \tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\boldsymbol{f}} \\ \tilde{\boldsymbol{f}}' \end{bmatrix} \tag{33}$$

where $\boldsymbol{f}$ and $\boldsymbol{f}'$ are known.

For each right-hand side column we have two equations in N_f unknowns and it follows that $(N_f - 2)$ of the corresponding column of $\tilde{\mathbf{X}}$ can be chosen at random. In practice we choose $(N_f - 2)$ entries to be zero and the two non-zero entries to be those that correspond to what we call the primary functions for column j. These primary functions for the open channels are the sine- and cosine-like JWKB solutions that correspond to channel j.

4. Illustrative calculations

To give an indication of the accuracy of the improved method, we consider three typical test cases which we will denote case A, case B and case C. Case A is a seven channel e^-–CIII scattering problem, which is the test case used by Burke and Noble (1995). Case B is a 19 channel e^-–He case used as a test case by Rudge (1984). This case is significantly more complex than others in the literature and it provides a stringent test in that there are many nearly degenerate channel groups due to the inclusion of fine structure levels, and many channels close to threshold. Case C is a 31 channel e^-–H test case that we have generated, which provides a demonstration of the ability of the improved method to handle cases containing a large number of channels.

The quantity ε, defined as

$$\varepsilon = \frac{1}{N_s N_p' N} \sum_{i=1}^{N} \sum_{j=1}^{N_s} \sum_{k=1}^{N_p'} |(\mathcal{L}_0 + \mathbf{V})\mathcal{F}_{ij}(r_k)| \tag{34}$$

can be used as a measure of the average error of the calculation. The number of points, N_p', in each subrange was chosen typically to be 500. The accuracy of the calculation is therefore illustrated in each case by plotting ε as a function of r.

In figures 1–3, $\log_{10}(\varepsilon)$ is plotted in the range $R_{in} \leqslant r \leqslant 500$ where R_{in} is the lowest value of R_1. This is the innermost part of the region where the scattering equations are

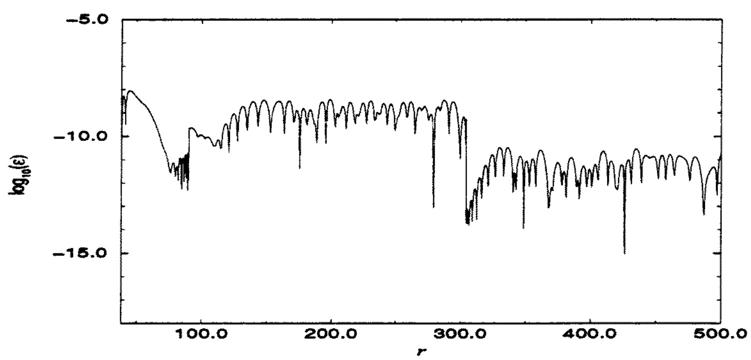

Figure 1. Plot of $\log_{10}(\varepsilon)$ as a function of r, case A.

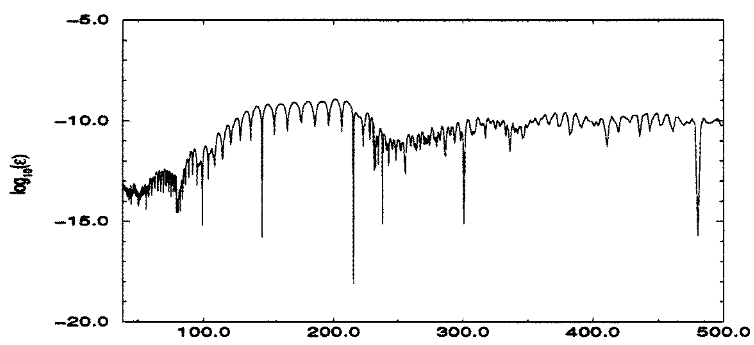

Figure 2. Plot of $\log_{10}(\varepsilon)$ as a function of r, case B.

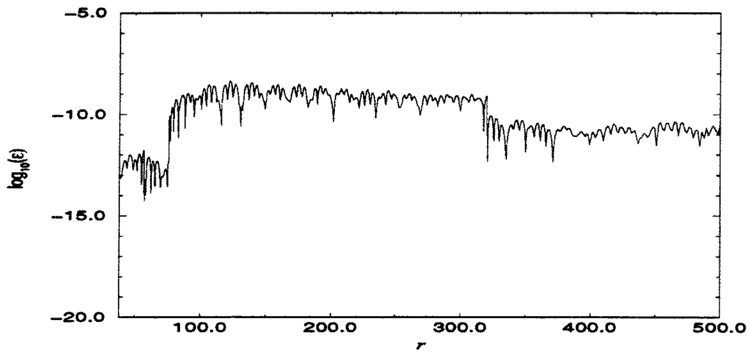

Figure 3. Plot of $\log_{10}(\varepsilon)$ as a function of r, case C.

to be solved. The number of Chebyshev polynomials used in the basis expansion was six for each channel in all cases. The appearance of very localized minima of $\log_{10}(\varepsilon)$ on the figures is due to the proximity of collocation points at these values of r. At the collocation points the linear equations are satisfied identically to machine accuracy, giving a negligible average error.

In order to compare our method with other calculations the K-matrix of atomic scattering theory (cf Rudge and Tiernan 1994) was generated for cases A and B using our code, the code of Rudge (1984) and the code of Burke and Noble (1995). In all cases the average value of $|\delta K_{ij}/K_{ij}|$ in the K-matrices generated by our code compared with those generated by the two other codes was less than 1%, indicating excellent agreement.

The speed of the iteration procedure depends on the values of r, but is in general rapid requiring about four or five iterations to achieve accurate solutions at the collocation points. The QZ algorithm works extremely well. Care has to be exercised, as in any calculation of this type, to ensure that the basis is in a numerical sense linearly independent over any particular range.

It can be seen from the figures that high accuracy can be gained in all cases with relatively small numbers of basis functions, even in the regions where the calculation is numerically more difficult and where the number of channels is large. Further details are given in Tiernan (1996).

5. Concluding remarks

We have written and tested a computer code for determining, in analytic form, the solutions of sets of coupled differential equations. We find that we can deal with many more equations than previously without storage problems and that the accuracy of the method is high. We find satisfactory agreement between the results generated by the present method and those used as test cases in the literature by other authors.

References

Burke V M and Noble C J 1995 *Comput. Phys. Comm.* **85** 471–500
Gardiner J D, Laub A J, Amato J J and Moler C B 1992a *ACM Trans. Math. Software* **18** 223–31
Gardiner J D, Wette M R, Laub A J, Amato J J and Moler C B 1992b *ACM Trans. Math. Software* **18** 232–8
Moler C B and Stewart G W 1973 *SIAM J. Numer. Anal.* **10** 241–56
Rudge M R H 1984 *Comput. Phys. Comm.* **34** 187–97

Rudge M R H and Tiernan D M 1994 *J. Phys. A: Math. Gen.* **27** 2545–52
Sylvester J J 1884 *Comptes Rendus Acad. Sci.* **99** 527–9
Tiernan D M 1996 *PhD Thesis* The Queens University of Belfast