IMA Journal of Numerical Analysis (1989) **9**, 1–14

Analysis of the Enright–Kamel Partitioning Method for Stiff Ordinary Differential Equations

DESMOND J. HIGHAM

Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

[Received 25 April 1987 and in revised form 11 April 1988]

The use of implicit formulae in the solution of stiff ODEs gives rise to systems of nonlinear equations which are usually solved iteratively by a modified Newton scheme. The linear algebra costs associated with such schemes may form a substantial part of the overall cost of the solution. The work of W. H. Enright and M. S. Kamel attempts to reduce the cost of the iteration by automatically transforming and partitioning the system. We provide new theoretical justification for this method in the case where the stiff eigenvalues of the Jacobian matrix used in the modified Newton iteration are small in number and well separated from the other eigenvalues. The theory of Y. Saad is introduced and adapted to show that the method uses the projection of the Jacobian onto a Krylov subspace which virtually contains the dominant subspace. This is shown to have favourable consequences. Numerical evidence is provided to support the theory.

1. Introduction

GIVEN a stiff system of ordinary differential equations (ODEs)

$$\frac{dy}{dt}=f(t,y)\quad (y\in\mathbb{R}^n),\qquad y(t_0)=y_0,$$

where $f:\mathbb{R}\times\mathbb{R}^n\to\mathbb{R}^n$, many stiff solvers generate approximations $y_i \simeq y(t_i)$ using an implicit linear multistep formula. At each step, it is necessary to solve a nonlinear equation in the unknown vector y_{i+1}. This equation has the form

$$G(y_{i+1}) \equiv y_{i+1}-h\beta f(t_{i+1},y_{i+1})-\gamma_i=0, \tag{1.1}$$

where γ_i is a vector of known quantities, β is a method-dependent parameter which may vary between steps, and $h\ (=h_i)=t_{i+1}-t_i$ is the current stepsize. Equation (1.1) is usually solved by an iteration of the form

$$W(y_{i+1}^{(j+1)}-y_{i+1}^{(j)})=-G(y_{i+1}^{(j)})\quad (j=1,2,\ldots), \tag{1.2}$$

where $y_{i+1}^{(0)}$ is given and W is a fixed iteration matrix. Ostrowski's theorem [12: p. 300] shows that a sufficient condition for the local convergence of this scheme to a locally unique solution y^* of (1.1) is

$$\rho\big(I-W^{-1}(I-h\beta J(y^*))\big)<1, \tag{1.3}$$

where f is assumed differentiable at y^*. Here, $J(y)$ denotes the Jacobian of $f(t_{i+1},y)$ and $\rho(B)$ denotes the spectral radius of B.

The simple choice $W=I$ in (1.2) gives rise to functional iteration. The convergence condition (1.3) then becomes $\rho(h\beta J(y^*)) < 1$, which, in the case of stiff systems, may represent a severe constraint on h. Following [17] and [1], an eigenvalue λ of $J(y^*)$ for which the condition $|h\beta\lambda| < 1$ is more restrictive on h than the local accuracy requirement will be called a *stiff eigenvalue.* We let k denote the number of stiff eigenvalues of $J(y^*)$. The subspace of $\mathbb{C}^n$ spanned by the corresponding stiff eigenvectors will be called the *dominant subspace* and denoted $D(J(y^*))$.

Stiff solvers usually employ a modified Newton (MN) iteration scheme. Here, $W = I - h\beta A$ in (1.2), with $A \simeq J(y^*)$. Typically, A is a Jacobian matrix (or a finite difference approximation to one) from a previous step. The linear algebra costs associated with the MN iteration can form a substantial part of the overall cost of the solution, particularly for large systems. Recently, considerable attention has been paid to the problem of reducing these costs [1–4, 6–8, 10, 14, 19]. In this paper, we present an analysis of the partitioning method of Enright and Kamel [10; for an earlier version see 7]. A short discussion on the use of partitioning methods is given below. In the next section, we describe the Enright–Kamel method and compare it briefly with that of Bjorck [1, 2]. The effectiveness of the method on a certain class of stiff problems is investigated in Sections 3 and 4. Numerical evidence to support the theory is given in Section 5.

In the three partitioning methods described in [1], [10], and [19], the aim is to replace A by an approximation $\tilde{A}$ which reduces the cost of solving the linear systems in (1.2) while maintaining a reasonable convergence rate for the iteration. To isolate the effect of this approximation, we assume that A is the exact Jacobian $J(y^*)$. In each of the three methods, $\tilde{A}$ is the orthogonal projection of A onto some subspace. The ideal subspace to use is the dominant subspace $D(A)$. Suppose that $D(A)$ has full dimension k and that the orthogonal matrix $[Q_\mu Q_\nu]$, with $Q_\mu \in \mathbb{R}^{n\times k}$, is such that span $\{Q_\mu\} = D(A)$, where span $\{Q_\mu\}$ denotes the space of all complex linear combinations of columns of Q_μ. Then we can write

$$[Q_\mu Q_\nu]^{\mathrm{T}} A [Q_\mu Q_\nu] = \begin{array}{c c} & \begin{array}{cc} k & n-k \end{array} \\ \begin{array}{c} k \\ n-k \end{array} & \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \end{array}. \tag{1.4}$$

and the orthogonal projection of A onto $D(A)$ may be expressed as

$$\tilde{A} = [Q_\mu Q_\nu] \begin{bmatrix} T_{11} & T_{12} \\ 0 & 0 \end{bmatrix} [Q_\mu Q_\nu]^{\mathrm{T}}. \tag{1.5}$$

Using $W = I - h\beta\tilde{A}$ in (1.2), the sufficient condition for local convergence (1.3) becomes

$$\rho(h\beta T_{22}) < 1. \tag{1.6}$$

The eigenvalues of T_{22} are precisely the nonstiff eigenvalues of A, and so (1.6) is the condition which arises when functional iteration is applied to the nonstiff subsystem. In practice, a condition such as

$$\|h\beta T_{22}\| < 1 \tag{1.7}$$

would be desirable to ensure a reasonable rate of convergence. Here, and throughout this paper, we use the Frobenius matrix norm and the Euclidean vector norm,

$$\|A\| = \Big(\sum_{i,j} |a_{ij}|^2\Big)^{\frac{1}{2}} \quad\text{and}\quad \|x\| = \Big(\sum_{i} |x_i|^2\Big)^{\frac{1}{2}},$$

respectively.

Intuitively one would expect the partitioning methods to be most effective when $k \ll n$ and the stiff eiginvalues are well separated from the nonstiff eigenvalues. A general term for such systems is 'separably stiff' (see [2] and [19]). Separably stiff systems arise, for example, in circuit simulations [11]. The problems discussed by Robertson [14] are also separable. However, this structure is not shared by all practical stiff problems; Curtis [5] cites examples from the field of mass action kinetics where the stiff eigenvalues are neither small in number nor well separated from the others. For the remainder of this paper, we will assume that the system is separable. We also assume that A can be stored in high-speed memory.

2. The Enright-Kamel partitioning method

In the Enright-Kamel method [10], the matrix A is reduced to partial upper Hessenberg (UH) form by an orthogonal similarity transformation. After an initial permutation and $m-1$ Householder stages, we have

$$Q^{\mathrm{T}}AQ = \begin{array}{cc} \begin{array}{cc} m & n-m \end{array} & \\ \begin{bmatrix} H & S_{12} \\ 0\,b & S_{22} \end{bmatrix} & \begin{array}{l} m \\ n-m \end{array} \end{array}, \tag{2.1}$$

where $H \in \mathbb{R}^{m\times m}$ is UH and $b \in \mathbb{R}^{n-m}$. The approximation $\tilde{A}$ takes the form

$$\tilde{A} = Q\begin{bmatrix} H & S_{12} \\ 0 & 0 \end{bmatrix}Q^{\mathrm{T}}. \tag{2.2}$$

Using $W = I - h\beta\tilde{A}$ in (1.2), the convergence condition (1.3) becomes

$$\rho(h\beta B) < 1, \tag{2.3}$$

where

$$B = \begin{array}{cc} \begin{array}{cc} m-1 & n-m+1 \end{array} & \\ \begin{bmatrix} 0 & -h\beta\bar{H}^{-1}S_{12}\,[b\;S_{22}] \\ 0 & b\;S_{22} \end{bmatrix} & \begin{array}{l} m \\ n-m \end{array} \end{array},$$

with $\bar{H} = I - h\beta H$. Now $\rho(B)$ is the same as the spectral radius of the lower $(n-m+1)\times(n-m+1)$ principle submatrix of B. Hence, on taking norms, we see that (2.3) will certainly hold if

$$|h\beta|\,(\|S_{22}\|^2 + \|b\|^2 + \|x\|^2)^{\frac{1}{2}} < 1, \tag{2.4}$$

where x^{T} is the mth row of the matrix $-h\beta\bar{H}^{-1}S_{12}[b\;S_{22}]$. Note that (2.4) can be tested at each stage of the reduction to UH form. In practice, a smaller bound than 1 may be desirable to ensure that the scheme converges in a reasonable number of iterations. Once the condition becomes true, the approximation $\tilde{A}$ in (2.2) may be considered suitable. This allows an approximation, and in particular a value of m, to be determined dynamically. (The test used by Kamel [10] is slightly different to (2.4) and includes the assumption $\|\bar{H}^{-1}S_{12}\| \leq \|S_{12}\|$, which is difficult to justify in general.) When $m \ll n$ the resulting iteration scheme can be considerably cheaper than the standard MN method [7, 10].

Enright and Kamel combined the Householder transformations with row and column interchanges in an attempt to force m to be small. However, owing to the essential uniqueness of the reduction, this extra effort will almost always be fruitless [2, 9]. Hence, the only freedom in the method is the choice of initial permutation. Bjorck [2] shows that attention may be restricted to a single row and column interchange. For example, the row which is largest in norm could be swapped with the first row.

We mention briefly the partitioning method of Bjorck [1]. A similar method was proposed independently by Watkins & HansonSmith [19]. In Bjorck's method, the aim is to use the ideal approximation $\tilde{A}$ given by (1.5). The matrix A is first reduced by Householder transformations to the partial UH form (2.1) with $m = k$. A block QR iteration is then performed which, under mild conditions, leads to a partitioning of the form (1.4). The convergence of the QR iteration is linear with rate

$$\mu = |\lambda/\hat{\lambda}|, \tag{2.5}$$

where λ is the largest nonstiff eigenvalue in modulus and $\hat{\lambda}$ is the smallest stiff eigenvalue in modulus. Note that μ is a measure of the separation between the stiff and nonstiff eigenvalues. A cost comparison in [9] shows that, in order for the Enright–Kamel method to be competitive with Bjorck's method, the number $m - k$ of extra Householder stages must not be much greater than the number of QR iterations—typically two or three for a separable system. However, a direct comparison between the two methods is not strictly valid since Bjorck's method requires the number k of stiff eigenvalues (or an upper bound for k) to be known, while the Enright–Kamel partitioning is obtained dynamically.

In the following two sections, we investigate whether it is reasonable to expect (2.4) to hold for a small value of $m - k$.

3. Krylov subspace theory

The first $r - 1$ Householder stages of the reduction (2.1) produce the matrix

$$Q^{rT}AQ^r = \begin{bmatrix} H^r & S_{12}^r \\ 0\, b^r & S_{22}^r \end{bmatrix}, \tag{3.1}$$

where H^r is $r \times r$ UH. We write $q_1, q_2, \ldots, q_r$ for the first r columns of Q^r and h_{ij} for the (i, j)th element of H^r. The superscript r is omitted since these quantities are unchanged at later stages. Equating the first $r - 1$ columns in the equation

$$AQ^r = Q^r \begin{bmatrix} H^r & S_{12}^r \\ 0\, b^r & S_{22}^r \end{bmatrix} \tag{3.2}$$

gives

$$\begin{aligned} Aq_1 &= h_{11}q_1 + h_{21}q_2 \\ Aq_2 &= h_{12}q_1 + h_{22}q_2 + h_{32}q_3 \\ &\vdots \\ Aq_{r-1} &= h_{1,r-1}q_1 + \cdots + h_{r-1,r-1}q_{r-1} + h_{r,r-1}q_r. \end{aligned} \tag{3.3}$$

We will assume for the moment that $h_{i+1,i} \neq 0$ for $i = 1, \ldots, r-1$. It then follows from (3.3) that $\{q_1, \ldots, q_r\}$ is a unitary basis for the Krylov subspace $K_r := \text{span}\,\{q_1, Aq_1, \ldots, A^{r-1}q_1\}$. If A has a dominant eigenvalue, then, since $A^{r-1}q_1 \in K_r$, we see that K_r contains a power method type approximation to the dominant eigenvector. It is mentioned by Gear & Saad [8] that K_r can contain good approximations to all the stiff eigenvectors for a relatively small value of r. We give below a careful analysis of this phenomenon for the case where A has a separable spectrum. To do this, we introduce and adapt the relevant theory of Saad [15, 16]. In the next section, we show that these results help to justify the Enright-Kamel partitioning method.

In the following analysis, the eigenvalues of A will be denoted $\{\lambda_i\}_{i=1}^n$ and assumed distinct. This implies the existence of n linearly independent eigenvectors $\{x_i\}_{i=1}^n$, which we normalize so that $\|x_i\| = 1$. Let $\pi_r : \mathbb{C}^n \to K_r$ denote the orthogonal projector onto the subspace K_r. Note that the distance from a vector x to K_r is given by $\|(I - \pi_r)x\|$. Our aim is to bound this distance when x is a stiff eigenvector. Any vector in the space K_r may be written $p(A)q_1$, where p is a member of P_{r-1}, the set of polynomials of degree not exceeding $r-1$. Since the eigenvectors $\{x_i\}_{i=1}^n$ form a basis for $\mathbb{C}^n$, we may write

$$q_1 = \sum_{j=1}^{n} \alpha_j x_j. \tag{3.4}$$

The following result appears in Saad [15].

Lemma 1 *For each x_i, if $\alpha_i \neq 0$ in* (3.4), *then*

$$\|(I - \pi_r)x_i\| \leq \xi_i \varepsilon_i^{(r)},$$

where

$$\xi_i = \frac{1}{|\alpha_i|} \sum_{\substack{j=1 \\ j \neq i}}^{n} |\alpha_j| \quad \textit{and} \quad \varepsilon_i^{(r)} = \min_{\{p \in P_{r-1} : p(\lambda_i) = 1\}} \max_{\substack{1 \leq j \leq n \\ j \neq i}} |p(\lambda_j)|. \tag{3.5}$$

Remarks

1. Examination of a proof of Lemma 1 shows that the bound is unlikely to be sharp (see [9]).

2. In our case, q_1 is determined by the choice of initial permutation in (2.1), and we are interested in applying Lemma 1 when x_i is one of the relatively few stiff eigenvectors. The assumption $\alpha_i \neq 0$ appears reasonable. (A similar situation arises in the analysis of the power method, where the starting vector is usually assumed not to be deficient in the component of the dominant eigenvector.) Furthermore, while ξ_i cannot be bounded *a priori*, one would imagine that, in general, its value would not be excessively large.

The next task is to find an upper bound for $\varepsilon_i^{(r)}$. We consider first the case where A has a real spectrum. In the following theorem, the technique used by Saad [15: Theorem 2.2; see also 8] is adapted to exploit the separable nature of the spectrum.

THEOREM 1 Suppose that $r \geq k$, and A has real eigenvalues ordered

$$\lambda_n < \lambda_{n-1} < \cdots < \lambda_{n-k+1} \ll \lambda_{n-k} < \cdots < \lambda_1,$$

where the (stiff) eigenvalues $\lambda_{n-k+1}, \ldots, \lambda_n$ are negative, having much larger moduli than the other (nonstiff) eigenvalues. Then, for each stiff eigenvalue λ_i, we have

$$\varepsilon_i^{(r)} \leq \Bigg(\prod_{\substack{j=n-k+1 \\ j \neq i}}^{n} \left| \frac{\lambda_1 - \lambda_j}{\lambda_i - \lambda_j} \right| \Bigg) [T_{r-k}(\gamma_i)]^{-1},$$

where

$$\gamma_i = 1 + 2\frac{\lambda_{n-k} - \lambda_i}{\lambda_1 - \lambda_{n-k}}$$

and T_{r-k} is the $(r-k)$th degree Chebyshev polynomial of the first kind.

Proof. Given λ_i, the first step is to restrict the set over which the minimization in (3.5) takes place. We constrain p to be zero at the other stiff eigenvalues; that is, letting

$$l(y) = \prod_{\substack{s=n-k+1 \\ s \neq i}}^{n} \frac{y - \lambda_s}{\lambda_i - \lambda_s},$$

so that $l \in P_{k-1}$ and $l(\lambda_i) = 1$, we consider polynomials of the form $l(y)h(y)$, where $h \in P_{r-k}$ and $h(\lambda_i) = 1$. This gives

$$\varepsilon_i^{(r)} \leq \min_{\{h \in P_{r-k} : h(\lambda_i) = 1\}} \max_{\substack{1 \leq j \leq n \\ j \neq i}} |l(\lambda_j)h(\lambda_j)|.$$

By construction, $l(y)$ is zero at the points $\{\lambda_j\}_{\substack{j=n-k+1 \\ j \neq i}}^{n}$, so the inequality above reduces to

$$\varepsilon_i^{(r)} \leq \min_{\{h \in P_{r-k} : h(\lambda_i) = 1\}} \max_{1 \leq j \leq n-k} |l(\lambda_j)h(\lambda_j)|.$$

Also, since the maximum in $\max_{1 \leq j \leq n-k} |l(\lambda_j)|$ is attained at $j = 1$, it follows that

$$\varepsilon_i^{(r)} \leq |l(\lambda_1)| \min_{\{h \in P_{r-k} : h(\lambda_i) = 1\}} \max_{1 \leq j \leq n-k} |h(\lambda_j)|,$$

which we weaken to

$$\varepsilon_i^{(r)} \leq |l(\lambda_1)| \min_{\{h \in P_{r-k} : h(\lambda_i) = 1\}} \max_{y \in [\lambda_{n-k}, \lambda_1]} |h(y)|.$$

The minimax term above is known to equal $[T_{r-k}(\gamma_i)]^{-1}$ [13: Appendix B], and the result follows immediately. □

Remarks

1. When $y \geq 1$, we may write $T_s(y)$ as

$$T_s(y) = \tfrac{1}{2}\{[y + (y^2 - 1)^{\frac{1}{2}}]^s + [y + (y^2 - 1)^{\frac{1}{2}}]^{-s}\}.$$

Hence, $T_s(y)$ takes large values when $y \gg 1$ and $s \geq 1$. Moreover, these values increase rapidly with s. In Theorem 1, we have $\gamma_i \gg 1$, so the term $[T_{r-k}(\gamma_i)]^{-1}$ will be small for $r > k$ and will quickly become negligible as r increases.

2. The term

$$\prod_{\substack{j=n-k+1 \\ j \neq i}}^{n} \left| \frac{\lambda_1 - \lambda_j}{\lambda_i - \lambda_j} \right|$$

in Theorem 1 may be large in certain cases. Each factor $|\lambda_1 - \lambda_j|$ in the numerator is of the same order as a stiff eigenvalue. The factors $|\lambda_i - \lambda_j|$ in the denominator depend upon the (unknown) separation of the stiff eigenvalues, the worst case being when other stiff eigenvalues are close to λ_i. However, the whole term is independent of r and, as r increases, the $[T_{r-k}(\gamma_i)]^{-1}$ term should dominate the bound.

3. Numerical examples of the upper bound for $\varepsilon_i^{(r)}$ given by Theorem 1 are presented in Tables 2 and 3 of Section 5.

For the more general case where A has a complex spectrum, we refer to the following theorem of Saad [16: p. 138].

Theorem 2 *Given an eigenvalue of A, which for convenience we label λ_1, there exist r other eigenvalues of A, which we label $\lambda_2, \ldots, \lambda_{r+1}$, such that*

$$\varepsilon_1^{(r)} = \left(\sum_{j=2}^{r+1} \prod_{\substack{s=2 \\ s \neq j}}^{r+1} \left| \frac{\lambda_1 - \lambda_s}{\lambda_j - \lambda_s} \right| \right)^{-1}. \tag{3.6}$$

We emphasize that this theorem is not constructive; the members of the set $\{\lambda_j\}_{j=2}^{r+1}$ are not determined. Unfortunately when λ_1 is a stiff eigenvalue and $r \simeq k$ the right-hand side of (3.6) depends critically on the set $\{\lambda_j\}_{j=2}^{r+1}$. To illustrate this, suppose λ_1 is stiff with $k = 4$ and $r = 4$. In order to show that ε_1^4 is small, it is sufficient to show that one of the terms

$$\prod_{\substack{s=2 \\ s \neq j}}^{5} \left| \frac{\lambda_1 - \lambda_s}{\lambda_j - \lambda_s} \right|$$

is large. Taking $j = 5$, we may write this term as

$$\left| \frac{\lambda_1 - \lambda_2}{\lambda_5 - \lambda_2} \right| \left| \frac{\lambda_1 - \lambda_3}{\lambda_5 - \lambda_3} \right| \left| \frac{\lambda_1 - \lambda_4}{\lambda_5 - \lambda_4} \right|. \tag{3.7}$$

Now, if λ_2, λ_3, λ_4, and λ_5 are nonstiff, then each of the three factors in (3.7) will be large. However, if λ_2 is replaced by a stiff eigenvalue, then only the second and third factors in (3.7) are guaranteed to be large; the first factor has a numerator which depends on the unknown separation $|\lambda_1 - \lambda_2|$. Similarly, with λ_2 and λ_3 stiff, only the third factor in (3.7) is guaranteed to be large, and, if λ_2, λ_3, and λ_4 are stiff, then the whole term depends upon the separation of the stiff eigenvalues.

For the complex spectra used in Section 5, the right-hand side of (3.6) was evaluated for every possible set $\{\lambda_j\}_{j=2}^{r+1}$, and the largest value was recorded (see Tables 4–6). This, of course, gives an upper bound for $\varepsilon_1^{(r)}$. In every case, the upper bound occurred when the maximum number of stiff eigenvalues appeared in $\{\lambda_j\}_{j=2}^{r+1}$. Also, as the above example indicates, for the upper bound to be small, it is desirable that the stiff eignenvalues be well separated. However, even if this is not the case, the number of nonstiff eigenvalues in the set will increase as r increases, and the corresponding large $|(\lambda_1 - \lambda_s)|/|(\lambda_j - \lambda_s)|$ factors in (3.6) should force the bound to become small (see Table 5).

In conclusion, Theorems 1 and 2 suggest that, for a stiff eigenvalue λ_i, the quantity $\varepsilon_i^{(r)}$ will quickly become small as r increases beyond k. Once this has happened, it follows from Lemma 1 that the corresponding stiff eigenvector will almost certainly be close to span $\{q_1, \ldots, q_r\}$.

Although we assumed that the $h_{i+1,i}$ values in (3.3) were all nonzero, the above conclusion remains valid in the pathological case when one of the values $h_{s+1,s}$ $(1 \leq s \leq r-1)$ is zero. The space span $\{q_1, \ldots, q_r\}$ may then be written as the union of two Krylov subspaces:

$$\begin{aligned}\text{span}\,\{q_1, \ldots, q_r\} &= \text{span}\,\{q_1, \ldots, q_s\} \cup \text{span}\,\{q_{s+1}, \ldots, q_r\}\\ &= \text{span}\,\{q_1, Aq_1, \ldots, A^{s-1}q_1\} \cup \text{span}\,\{q_{s+1}, Aq_{s+1}, \ldots, A^{r-s}q_{s+1}\}.\end{aligned}$$

The first of these subspaces is invariant under A, and hence contains s of the eigenvectors of A. If a stiff eigenvector x_i is not contained in this space, then the results of this section show that, as r increases, the distance between the second Krylov subspace and x_i should become small.

4. Implications for the Enright–Kamel method

The conclusions drawn in the previous section can be used to give insight into the effectiveness of the Enright–Kamel method. First, we look at the size of b^r in (3.1). Partitioning

$$Q_r = \begin{array}{cc} r & n-r \\ {[Q_1} & {Q_2]}, \end{array}$$

it follows from (3.1) that

$$Q_2^{\mathrm{T}} A Q_1 = [0 \quad b^r]. \tag{4.1}$$

Recalling that q_r denotes the rth column of Q_1, we equate the rth columns in (4.1) to obtain

$$Q_2^{\mathrm{T}} A q_r = b^r. \tag{4.2}$$

Next we expand q_r in terms of the eigenvectors of A:

$$q_r = \sum_{j=1}^{n} \beta_j x_j, \tag{4.3}$$

where we assume that $x_1, \ldots, x_k$ are the stiff eigenvectors. Equation (4.2) then becomes

$$b^r = Q_2^T A \sum_{j=1}^{n} \beta_j x_j = \sum_{j=1}^{n} \beta_j \lambda_j Q_2^T \mathrm{x}_j,$$

and, on taking norms, we find

$$\|b^r\| \le \sum_{j=1}^{n} |\beta_j \lambda_j|\ \|Q_2^T x_j\|. \tag{4.4}$$

Now

$$\|Q_2^T x_j\| = \|Q_2 Q_2^T x_j\| = \|(I - \pi_r) x_j\|,$$

which, from Section 3, should become negligible for $j = 1, \ldots, k$. Hence, in (4.4),

$$\|b^r\| \lessapprox \sum_{j=k+1}^{n} |\beta_j \lambda_j|\ \|Q_2^T x_j\|,$$

and, since $\|Q_2^T x_j\| \le \|x_j\| = 1$, we have

$$\|b^r\| \lessapprox \sum_{j=k+1}^{n} |\beta_j \lambda_j|. \tag{4.5}$$

Although the size of the coefficients β_j cannot be bounded *a priori*, (4.5) does suggest that $\|b^r\|$ will generally be of roughly the same order as the nonstiff eigenvalues and consequently much less than $\|A\|$. A similar argument applies to the columns of S_{22}^r in (3.1), and, under the stronger assumption that the stiff eigenvectors are contained in span $\{Q_1\}$, we may use the following lemma.

LEMMA 2 *If*

$$\left[\begin{array}{cc} \overset{k}{Q_\mu} & \overset{n-k}{Q_\nu} \end{array}\right] \quad \text{and} \quad \left[\begin{array}{cc} \overset{r}{Q_1} & \overset{n-r}{Q_2} \end{array}\right] \qquad (r \ge k),$$

are orthogonal matrices such that

$$\text{span}\,\{Q_\mu\} \subseteq \text{span}\,\{Q_1\},$$

then

$$\|Q_2^T A Q_2\| \le \|Q_\nu^T A Q_\nu\|.$$

Proof. The proof is straightforward: see [9].

Now, with the ideal partitioning (1.4), the error in the approximate Jacobian $\tilde{A}$ of (1.5) is easily seen to be

$$\|A - \tilde{A}\| = \|Q_\nu^T A Q_\nu\|,$$

while the Enright–Kamel partitioning (3.1) would give

$$\|A - \tilde{A}\| = \|[b^r\ Q_2^T A Q_2]\|.$$

Hence, Lemma 2 and inequality (4.5), combined with the results of Section 3, indicate that, as the value of r in (3.1) begins to increase beyond k, the

Enright–Kamel method should be able to produce an approximate Jacobian $\tilde{A}$ such that $\|A - \tilde{A}\|$ is of the same order as the corresponding value given by the ideal partitioning. Although this does not guarantee that the convergence condition (2.4) will be satisfied, it suggests that the method will be successful in reducing the sizes of $\|S_{22}\|$ and $\|b\|$, which are likely to be the dominant terms in (2.4).

Finally, note that $\tilde{A}$ in (2.2) is the orthogonal projection of A onto span $\{q_1, \ldots, q_m\}$. Hence, we may regard the Enright–Kamel method as using the projection of A onto a space which virtually contains the dominant subspace.

5. Numerical tests

Kamel [10] incorporated the Enright–Kamel method into a stiff solver which originally used a MN scheme. His results show that, if a suitable partitioning (2.1) can be found with $m \ll n$, then a substantial saving in solution time can be achieved. Hence, in testing the method, our approach is to generate a matrix which models a separably stiff Jacobian, perform the reduction, and record the value of m produced.

The model Jacobians which we use have the general form

$$A = V^{\mathrm{T}} M V,$$

where V is a random orthogonal matrix generated by the method of Stewart [18: Theorem 3.3] and M is chosen so that A has a suitable set of eigenvalues. In all the tests, we have $n = 25$ with $k = 3$ stiff eigenvalues.

The first test involves the real spectrum

$$\mathrm{R1} = \{-\tfrac{1}{2}j\}_{j=1}^{22} \cup \{-3 \times 10^3, -2 \times 10^3, -10^3\},$$

for which $\mu \simeq 10^{-2}$ in (2.5). M is upper triangular with the stiff eigenvalues forming the first three diagonal elements of M and the remaining eigenvalues placed along the diagonal in an arbitrary order. The strictly upper triangular part of M consists of random numbers (taken from a $N(0, 1)$ distribution) with the first three rows scaled by 10^2. The scaling is intended to represent coupling between the stiff components and between the stiff and nonstiff subsystems. Following (1.7), we choose $h\beta$ so that $h\beta\, \|M_{22}\| = 1$, where M_{22} is the lower 22×22 principle submatrix of M.

In the next test, M is constructed in the manner described above, but using the spectrum

$$\mathrm{R2} = \{-\tfrac{1}{2}j\}_{j=1}^{22} \cup \{-1{\cdot}02 \times 10^3, -1{\cdot}01 \times 10^3, -10^3\},$$

in which the stiff eigenvalues are clustered together.

To generate complex spectra, the only modification required is the introduction of 2×2 blocks along the diagonal of M. The sets

$$\mathrm{C1} = \{-\tfrac{1}{2}j\}_{j=1}^{18} \cup \{-2 \pm 2\mathrm{i}, -5 \pm 5\mathrm{i}\} \cup \{-10^3 \pm 10^3\mathrm{i}, -10^3\}$$

$$\mathrm{C2} = \{-\tfrac{1}{2}j\}_{j=1}^{18} \cup \{-2 \pm 2\mathrm{i}, -5 \pm 5\mathrm{i}\} \cup \{-10^3 \pm 10\mathrm{i}, -10^3\}$$

are used for whch $\mu \simeq 10^{-2}$. Note that C2 has stiff eigenvalues which are closer together than those of C1. The final spectrum considered is

$$C3 = \{-\tfrac{1}{2}j\}_{j=1}^{18} \cup \{-2 \pm 2i, -5 \pm 5i\} \cup \{-10^2 \pm 10^2 i, -10^2\},$$

where $\mu \simeq 10^{-1}$.

The values of m obtained are presented in Table 1. Tables 2–6 record upper bounds on $\varepsilon^{(r)}$ (see Lemma 1) for the stiff eigenvalues used. These were found by applying Theorem 1 in Tables 2 and 3 and Theorem 2 in Tables 4–6. (Note that bounds for a complex conjugate pair are equal by symmetry.)

TABLE 1
Test results

Spectrum:	R1	R2	C1	C2	C3
m:	5	6	5	5	5

TABLE 2
Spectrum = R1

	Stiff eigenvalue		
r	-3×10^3	-2×10^3	-10^3
4	2×10^{-3}	8×10^{-3}	2×10^{-2}
5	2×10^{-6}	1×10^{-5}	4×10^{-5}
6	1×10^{-9}	1×10^{-8}	1×10^{-7}
7	1×10^{-12}	2×10^{-11}	3×10^{-10}

TABLE 3
Spectrum = R2

	Stiff eigenvalue		
r	$-1{\cdot}02 \times 10^3$	$-1{\cdot}01 \times 10^3$	-10^3
4	3×10^1	5×10^1	3×10^1
5	7×10^{-2}	1×10^{-1}	7×10^{-2}
6	2×10^{-4}	4×10^{-4}	2×10^{-4}
7	5×10^{-7}	1×10^{-6}	5×10^{-7}

TABLE 4
Spectrum = C1

	Stiff eigenvalue	
r	$-10^3 \pm 10^3 i$	-10^3
4	2×10^{-3}	1×10^{-2}
5	7×10^{-6}	4×10^{-5}
6	2×10^{-8}	2×10^{-7}
7	5×10^{-11}	6×10^{-10}

TABLE 5
Spectrum = C2

r	Stiff eigenvalue $-10^3 \pm 10i$	 -10^3
4	3×10^{-1}	1×10^{0}
5	7×10^{-2}	2×10^{-1}
6	5×10^{-4}	1×10^{-3}
7	1×10^{-6}	3×10^{-6}

TABLE 6
Spectrum = C3

r	Stiff eigenvalue $-10^2 \pm 10^2i$	 -10^2
4	2×10^{-2}	1×10^{-1}
5	7×10^{-4}	4×10^{-3}
6	2×10^{-5}	2×10^{-4}
7	5×10^{-7}	7×10^{-6}

Comments

1. From Table 1, we see that the method was always able to produce a suitable partitioning with $m - k \leq 3$. Also, in each test, the reduction (3.1) halted at a stage where the $\varepsilon^{(r)}$ bounds had become small.

2. At each stage of the reduction, the norms of the submatrices were examined. Their behaviour followed the pattern predicted by the theory of Sections 3 and 4. For example, Table 7 gives the results for the test with C1.

We see that $\|S^r_{22}\|^2$ and $\|b^r\|^2$ decrease sharply as r increases from 3 to 5. A similar pattern emerged in all the testing. Table 7 also shows the size of $\|x\|^2$ (see (2.4)) at each stage. The small value of $\|x\|^2$ at the exit stage $r = 5$ is typical.

3. More extensive testing, including the case where A has repeated eigenvalues, was performed in [9]. The results were equally encouraging. However, we mention that it is possible to construct examples where neither the Enright–

TABLE 7
Submatrix norms in the reduction (3.1) *using the spectrum* C1

r	$\|H^r\|^2$	$\|S^r_{12}\|^2$	$\|b^r\|^2$	$\|S^r_{22}\|^2$	$\|x^r\|^2$
1	4×10^{4}	2×10^{7}	2×10^{5}	9×10^{7}	1×10^{9}
2	2×10^{6}	9×10^{7}	3×10^{4}	2×10^{7}	7×10^{7}
3	6×10^{7}	3×10^{7}	5×10^{6}	7×10^{6}	4×10^{7}
4	1×10^{8}	1×10^{6}	1×10^{3}	1×10^{3}	1×10^{2}
5	1×10^{8}	8×10^{5}	1×10^{1}	8×10^{2}	3×10^{-1}

Kamel method nor the block QR method are effective (see, for example, [2]).

In summary, the new insight given by the analysis of Sections 3 and 4 and the promising numerical results of this section suggest that the Enright–Kamel method is a useful tool for dealing with separably stiff systems.

Acknowledgements

It is a pleasure to thank Professor W. H. Enright of the University of Toronto for guidance and encouragement during his stay at the University of Manchester. I am also grateful to Dr I. Gladwell and Dr N. J. Higham for carefully reading the original manuscript and suggesting improvements. This work was supported by a Science and Engineering Research Council Research Studentship.

REFERENCES

[1] BJORCK, A. 1983 A block QR algorithm for partitioning stiff differential systems. *BIT* **23**, 329–45.

[2] BJORCK, A. 1984 Some methods for separating stiff components in initial value problems. In: *Numerical Analysis, Dundee 1983* (D. F. Griffiths, Ed.), Lecture Notes in Mathematics 1066. Berlin: Springer-Verlag, pp. 30–43.

[3] BROWN, P. J., & HINDMARSH, A. C. 1986 Matrix-free methods for stiff systems of ODE's. *SIAM J. Numer. Anal.* **23**, 610–38.

[4] CHAN, T. F., & JACKSON, K. R. 1986 The use of iterative linear-equation solvers in codes for large systems of stiff IVPs for ODEs. *SIAM J. Sci. Stat. Comput.* **7**, 378–417.

[5] CURTIS, A. R. 1983 Jacobian matrix properties and their impact on the choice of software for stiff ODE systems. *IMA J. Numer. Anal.* **3**, 397–415.

[6] ENRIGHT, W. H. 1978 Improving the efficiency of matrix operations in the numerical solution of stiff ordinary differential equations. *ACM Trans. Math. Software* **4**, 127–36.

[7] ENRIGHT, W. H., & KAMEL, M. S. 1979 Automatic partitioning of stiff systems and exploiting the resulting structure. *ACM Trans. Math. Software* **5**, 374–85.

[8] GEAR, C. W., & SAAD, Y. 1983 Iterative solution of linear equations in ODE codes. *SIAM J. Sci. Stat. Comput.* **4**, 583–601.

[9] HIGHAM, D. J. 1986 Analysis of the partitioning method of Enright and Kamel in the numerical solution of stiff ODEs. M.Sc. Thesis, University of Manchester, England.

[10] KAMEL, M. S., 1981 Improving the efficiency of stiff ODE solvers by partitioning. Technical Report No. 149/81, Department of Computer Science, University of Toronto.

[11] LEE, H. B. 1967 Matrix filtering as an aid to numerical integration. *Proc. IEEE* **55**, 1826–31.

[12] ORTEGA, J. M., & RHEINBOLDT, W. C. 1970 *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic Press.

[13] PARLETT, B. N. 1980 *The Symmetric Eigenvalue Problem*. Englewood Cliffs, NJ: Prentice-Hall.

[14] ROBERTSON, H. H. 1976 Numerical integration of systems of stiff ODE's with special structure. *J. Inst. Math. Applic.* **18**, 249–63.

[15] SAAD, Y. 1980 Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices. *Linear Algebra Applic.* **34**, 269–95.

[16] SAAD, Y. 1983 Projection methods for solving large sparse eigenvalue problems. In: *Matrix Pencils, Proceedings, 1982* (B. Kagstrom & A. Ruhe, Eds), Lecture Notes in Mathematics 973. Berlin: Springer-Verlag, pp. 121–44.

[17] SODERLIND, G. 1981 On the efficient solution of nonlinear equations in numerical methods for stiff differential systems. Report TRITA-NA-8114, The Royal Institute of Technology, Stockholm, Sweden.

[18] STEWART, G. W. 1980 The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM J. Numer. Anal.* **17,** 403–9.

[19] WATKINS, D. S., & HANSONSMITH, R. W. 1983 The numerical solution of separably stiff systems by precise partitioning. *ACM Trans. Math. Software* **9,** 293–301.