SPECIAL METHODS FOR PROBLEMS WHOSE OSCILLATORY SOLUTION IS DAMPED

Beny Neta
Naval Postgraduate School
Department of Mathematics
Code 538d
Monterey, CA 93943

ABSTRACT

This paper introduces methods tailored especially for problems whose solution behaves like $e^{\lambda x}$, where $\lambda$ is complex. The shallow water equations with topography admit such solution.

This paper complements the results of Pratt and others on exponential-fitted methods and those of Gautschi, Neta, van der Houwen and others on trigonometrically-fitted methods.

1. Introduction

In this paper we consider linear multistep methods

$$
\sum_{\ell=0}^{k} a_{k} y_{n+1-\ell} = h \sum_{\ell=0}^{k} b_{\ell} f(x_{n+1-\ell}, y_{n+1-\ell}), \quad k \geq 1, \quad n \geq k-1
$$

(1)

for integrating the initial value problem

$$
y'(x) = f(x, y(x)), \quad y(x_0) = y_0.
$$

(2)

This linear multistep method is characterized by the polynomials

$$
\rho(\zeta) = \sum_{k=0}^{k} a_k \zeta^{k-l}, \quad \sigma(\zeta) = \sum_{k=0}^{k} b_k \zeta^{k-l}.
$$

(3)

The main assumption of this paper is that it is a priori known that the solution is approximately of the form

$$
y(x) \sim c_{0} + \sum_{j=1}^{m} c_{j} e^{i\lambda_{j} x},
$$

(4)

where $\lambda_{j} = \omega_{j} + i\psi_{j}$, and the frequencies $\omega_{j}$ are in a given interval $[\omega_{l}, \omega_{u}]$.

The special case where $j = j_{0}$ with $\omega_{0}$ given was considered first by Gautschi [8]. His approach was the following. Let:

$$
\phi(z) = \rho(z) - z\sigma(z)
$$

(5)

then the local truncation error of (1) is given by Lambert [11]

$$
T_{n+k} = \phi(h \frac{d}{dt}) y(t_{n}').
$$

(6)

Inserting (4) in (6) yields

$$
T_{n+k} \sim \phi(0)c_{0} + \sum_{j=1}^{m} c_{j} \phi(ih\lambda_{j}) e^{i\lambda_{j} t}, \quad \lambda_{j} = j\omega_{0}.
$$

(7)

The coefficients $b_{\ell}$ are chosen in such a way that

$$
\phi(ih\omega_{0}) = 0, \quad j = 0, 1, ..., q.
$$

(8)
for the largest value of \( q \) possible. \( q \) is then called the trigonometric order of the method. Gautschi has chosen \( a_q \) such that the methods are those of Adams and Stormer type. However, these methods are sensitive to changes in the frequency \( w_0 \). Neta and Ford [13] developed Nyström and generalized Milne-Simpson type methods. These methods showed less sensitivity to perturbation in \( w_0 \) but require the eigenvalues of the Jacobian to be purely imaginary. Neta [14] has developed families of backward differentiation methods that overcome the above-mentioned restriction. Salzer [17] has developed predictor-corrector methods based on trigonometric polynomials. See also Steifel and Bettis [18] and Bettis [3]. Van der Houwen and Sommeijer [10] have developed an alternative approach. The conditions (8) were replaced by

\[
\begin{align*}
\phi(0) &= 0, \\
\phi(1/k)(j) &= 0, \quad j = 1, 2, \ldots, q,
\end{align*}
\]

where the \( \lambda(j) \) are appropriately chosen points in the interval \([w_L, w_U]\).

An advantage of this so-called minimax approach over the fitting approach is the increased accuracy in cases where no accurate estimate of \( w_0 \) is available or when the frequency is varying in time.

The other special case considered in the literature is where \( \lambda_1 = i\psi \). Probably the first article on the subject is due to Brock and Murray [5]. They discuss the use of exponential sums in the integration of a system of first order ordinary differential equations. Dennis [7] also suggested special methods for problems whose solution is exponential. He suggested a transformation of variables. More recently, Carroll [6] has developed exponentially fitted one-step methods for the scalar Riccati equation. For the general first order system of equations, Pratt [16] suggests methods based on the three parameter exponential function

\[ I(x) = A + Be^{x^2} \]

The parameters \( A, B \) are given in terms of values of \( y \) and \( f \). Several possibilities for \( z \) are given based on results of Brandon [2] and Babcock et al. [1].

Lyche [12] analyses multistep methods which exactly integrate the set \([x^n w_0 w_1]\), where \( w_0 \) is real or imaginary.

In this article we developed various methods fitting exponentials and methods obtained via the minimax approach.

2. Construction of Methods

2.1 Fitting Methods

In this subsection we discuss various fitting methods. To this end, we separate \( \phi(i\lambda j) = 0 \), \( j = 1, 2, \ldots, q \), into real and imaginary parts. This yields the following equations relating the coefficients \( a_k, b_k \),

\[
\begin{align*}
\frac{1}{k} \sum_{k=0}^{k} a_k e^{j\mu(k-\nu)} \cos jv(k-\nu) - \frac{1}{k} \sum_{k=0}^{k} b_k e^{j\mu(k-\nu)} [j\mu \cos jv(k-\nu) - jv \sin jv(k-\nu)] &= 0, \\
\frac{1}{k} \sum_{k=0}^{k} a_k e^{j\mu(k-\nu)} \sin jv(k-\nu) - \frac{1}{k} \sum_{k=0}^{k} b_k e^{j\mu(k-\nu)} [j\mu \sin jv(k-\nu) + jv \cos jv(k-\nu)] &= 0,
\end{align*}
\]

where \( \lambda = \omega + i\psi \), \( \mu = -h\psi \), \( v = hv, j = 1, 2, \ldots, q \).
For explicit methods, \( b_0 = 0 \). For Adams type methods \( a_0 = 1, a_1 = -1, a_4 = 0 \) for \( i = 2, \ldots, k \). For Nyström or generalized Milne-Simpson methods \( a_0 = 1, a_2 = -1 \) and other \( a_i = 0 \).

### \( k = 1 \) Implicit

**Adams**

\[
\begin{align*}
b_0 &= \frac{\omega e^{-\mu} + \mu \sin \nu - \nu \cos \nu}{(\mu^2 + \nu^2) \sin \nu}, \\
b_1 &= \frac{\omega e^{-\mu} - \mu \sin \nu - \nu \cos \nu}{(\mu^2 + \nu^2) \sin \nu}.
\end{align*}
\]

For \( \nu = 0 \), the coefficients become \( b_0 = \frac{1 - \cos \nu}{\sin \nu} = b_1 \), which agree with Gautschi [8] if the coefficients are expanded in Taylor series with respect to \( \nu \).

### \( k = 2 \) Explicit

**Adams**

\[
\begin{align*}
b_1 &= \frac{(\mu \sin 2\nu - \nu \cos 2\nu) e^{\mu} + \nu \cos \nu - \mu \sin \nu}{(\mu^2 + \nu^2) \sin \nu}, \\
b_2 &= \frac{(\nu \cos \nu - \mu \sin \nu) e^{2\mu} - \nu e^{\mu}}{(\mu^2 + \nu^2) \sin \nu}.
\end{align*}
\]

**Nystrom**

\[
\begin{align*}
b_1 &= \frac{(\mu \sin 2\nu - \nu \cos 2\nu) e^{\mu} + \nu e^{-\mu}}{(\mu^2 + \nu^2) \sin \nu}, \\
b_2 &= \frac{(\nu \cos \nu - \mu \sin \nu) e^{2\mu} + \nu e^{-\mu}}{(\mu^2 + \nu^2) \sin \nu}.
\end{align*}
\]

For \( \nu = 0 \), the coefficients become \( b_1 = 2 \sin \nu \), \( b_2 = 0 \) which agree with Neta [14].

### \( k = 2 \) Implicit

In this case, one obtains a one-parameter family of (Adams, generalized Milne-Simpson) methods of trigonometric order 1. The free parameter can be used to increase the algebraic order of the method as in [13].

**Backward Differentiation**

\[
\begin{align*}
e^{2\mu} \cos 2\nu + a_1 e^{\mu} \cos \nu + a_2 &= -b_0 e^{2\mu} (\mu \cos 2\nu - \nu \sin 2\nu) = 0, \\
e^{2\mu} \sin 2\nu + a_1 e^{\mu} \sin \nu &= -b_0 e^{2\mu} (\mu \sin 2\nu + \nu \cos 2\nu) = 0, \\
1 + a_1 + a_2 &= 0.
\end{align*}
\]

This system can be solved by MACSYMA (Project MAC’s Symbolic Manipulation system written in LISP and used for performing symbolic as well as numerical mathematical manipulation [4]) or by REDUCE [9]. The solution is
\[
\begin{align*}
    a_1 &= \frac{ve^{2\mu} - \mu \sin 2\nu - \nu \cos 2\nu}{e^{\mu}(\nu \cos \nu + \mu \sin \nu) + \nu \sin 2\nu + \nu \cos 2\nu}, \\
    a_2 &= -1 - a_1, \\
    b_0 &= \frac{-e^{2\mu} \sin \nu}{e^{\mu}(\nu \cos \nu + \mu \sin \nu) + \nu \sin 2\nu + \nu \cos 2\nu}.
\end{align*}
\]

For \( \psi = 0 \), the coefficients agree with those given in [14].

**k = 3 Explicit**

Again here, one obtains a one-parameter family of methods of trigonometric order 1. In order to get methods of trigonometric order 2, one has to construct a 3 step implicit method of Adams or generalized Milne-Simpson type. In order to increase the trigonometric order without going to a higher step number, one can construct linear multistep methods for which the coefficients \( a_3 \) are also functions of \( \psi, \nu \). Some examples are given in the next subsection.

### 2.2 Generalized Fitting Methods

In this section, we construct some linear multistep methods of the form

\[
\sum_{k=0}^{k} a_k(\lambda) y_{n+k-1} = h \sum_{k=0}^{k} b_k(\lambda) y_{n+k-1}, \quad k \geq 1, \quad n \geq k-1.
\]

Since \( a_3 \) are functions of \( \lambda \) one has more free parameters for his disposal which can be used to obtain higher trigonometric order methods with relatively lower step number.

**k = 2 Implicit**

In this case, one has to solve the following linear system of five equations for the parameters \( a_1, a_2, b_0, b_1, b_2 \) to obtain a method of trigonometric order 2.

\[
\begin{align*}
    1 + a_1 + a_2 &= 0, \\
    e^{2\mu} \cos 2\nu + a_1 e^{\mu} \cos \nu + a_2 &= b_0 e^{2\mu}(\nu \cos 2\nu - \nu \sin 2\nu) \\
    - b_1 e^{\mu}(\nu \cos \nu - \nu \sin \nu) - \nu b_2 &= 0, \\
    e^{2\mu} \sin 2\nu + a_1 e^{\mu} \sin \nu &= b_0 e^{2\mu}(\mu \sin 2\nu + \nu \cos 2\nu) \\
    - b_1 e^{\mu}(\nu \sin \nu + \nu \cos \nu) - \nu b_2 &= 0, \\
    e^{4\mu} \cos 4\nu + a_1 e^{2\mu} \cos 2\nu + a_2 &= b_0 e^{4\mu}(2\mu \cos 4\nu - 2\nu \sin 4\nu) \\
    - b_1 e^{2\mu}(2\mu \cos 2\nu - 2\nu \sin 2\nu) - 2\nu b_2 &= 0, \\
    e^{4\mu} \sin 4\nu + a_1 e^{2\mu} \sin 2\nu &= b_0 e^{4\mu}(2\mu \sin 4\nu + 2\nu \cos 4\nu) \\
    - b_1 e^{2\mu}(2\mu \sin 2\nu + 2\nu \cos 2\nu) - 2\nu b_2 &= 0.
\end{align*}
\]

The system was solved by REDUCE [9]. The expressions for the coefficients are complicated but REDUCE produces an output in the form of Fortran statements that can be incorporated in a computer program for numerical experiments with such a method.
2.3 Minimax Methods

In this section we discuss minimax methods, i.e., methods obtained by satisfying conditions (9). These conditions can be written in terms of $a_k, b_k$ as follows:

\[ \sum_{k=0}^{n} a_k e^{(k-\mu)}u^{(j)}(j) \cos(k-\nu) = \frac{k}{j=0} b_k e^{(k-\mu)}u^{(j)} \left[ -v(j) \sin(k-\nu)u^{(j)}(j) \right], \]

\[ j = 1, 2, \ldots, q, \]  

\[ \sum_{k=0}^{n} a_k e^{(k-\mu)}u^{(j)} \sin(k-\nu) = \frac{k}{j=0} b_k e^{(k-\mu)}u^{(j)} \left[ v(j) \cos(k-\nu)u^{(j)}(j) \right], \]

\[ j = 1, 2, \ldots, q, \]

where $\mu(j) = h\psi(j), \nu(j) = h\omega(j)$.

$h\lambda(j)$ are the zeros of the function $\phi(h\lambda)$ such that it has a small maximum norm in the rectangle $0 < \omega < \omega$, $\psi < \psi < \psi$.

To obtain the best approximation in this case is certainly not easy; but we will assume that one can write $\phi(h\lambda(j)) = \phi(h\psi(j), h\omega(j))$ as a product of 2 one variable functions. Thus $\psi(j)$ and $\omega(j)$ can be taken as Chebyshev's points on the corresponding interval, i.e.

\[ \psi(j) = \frac{1}{2}(\psi_k + \psi) + \frac{1}{2}(\psi - \psi) \cos \frac{2j-1}{2q} \pi, \]

\[ \omega(j) = \frac{1}{2}(\omega_k + \omega) + \frac{1}{2}(\omega - \omega) \cos \frac{2j-1}{2q} \pi, \]

\[ j = 1, 2, \ldots, q. \]

For this choice of points, one can evaluate the coefficients $a_k, b_k$ by solving

\[ \phi(0) = 0, \]

\[ \phi(h\psi(j), h\omega(j)) = 0, \quad j = 1, 2, \ldots, q. \]

We call such methods product minimax ($P^2$).

The number of free parameters for implicit methods $2k+1$ and the number of equations is $2q+1$, thus, the trigonometric order $q$ is equal to the step number $k$.

$k = 1$ Implicit

In this case

\[ \psi(1) = \frac{1}{2}(\psi_k + \psi) \]

\[ \omega(1) = \frac{1}{2}(\omega_k + \omega) \]

\[ a_0 = 1 = -a_1 \]

and the system of equations can be solved for $b_1, b_2$. This yields the coefficients given by (12) where
\[ \mu = -h\psi(1), \quad \nu = h\psi'(1). \]  

Thus, the product minimax method would suggest using the center of the rectangle \([\psi_0, \psi_2] \times [\omega_0, \omega_2]\) as \(\lambda_0\). To obtain a product minimax method of trigonometric order 2, one has to solve a system of 5 equations similar to equation (18) with the unknowns \(b_0, b_1, b_2, a_1, a_2\). The difference is that in the last 2 equations one should replace \(2u\) by \(\psi(2)\) and \(2v\) by \(\psi'(2)\). In the second and third equations of (18), the \(u, v\) should be replaced by \(\psi(1), \psi'(1)\) respectively. The resulting system can be solved by MACSYMA [4] or REDUCE [9].

In the next section we implement two methods of trigonometric order 1 and 2 and see how the product minimax methods compare with fitting methods.

3. Numerical Example

In this section we compare Adams fitting method of trigonometric order 1, the generalised fitting method of trigonometric order 2 obtained when solving system (18), the product minimax method of trigonometric order 1 given by (23) and of trigonometric order 2 obtained when solving the system (19)-(21).

Both systems (18) and (19)-(21) were solved by REDUCE which produced a FORTRAN subroutine for the evaluation of the coefficients. This subroutine is called only once during the integration.

The methods were compared for the solution of the initial value problem
\[
\begin{align*}
\dot{x} - \frac{\pi}{2} (1 + i)x &= 0, \quad 0 \leq t \leq 4, \\
x(0) &= 1
\end{align*}
\]  

whose exact solution
\[
x(t) = e^{\frac{\pi}{2}(1+i)t},
\]
thus
\[
\lambda = \frac{\pi}{2}(1 + i), \quad \psi = -\frac{\pi}{2}, \quad \omega = \frac{\pi}{2}.
\]

In order to avoid complex arithmetic, we rewrite the differential equation as a system of equations for the real and imaginary part of \(z = u + iv\).

\[
\begin{align*}
\dot{u} + \frac{\pi}{2}(u + v) &= 0, \quad 0 \leq t \leq 4, \\
\dot{v} - \frac{\pi}{2}(u - v) &= 0, \\
u(0) &= 1, \\
v(0) &= 0.
\end{align*}
\]

The system is solved by fitting methods of trigonometric order 1 and 2 with \(h = .01\) and various values of \(\psi\) and \(\omega\). In Table 1 we list the Euclidean norm of the error at \(t = 4\). It is clear that the method is not sensitive to perturbations in the values of \(\psi\) and \(\omega\).
\begin{tabular}{|c|c|c|c|}
\hline
$\Delta \psi$ & $\Delta \omega$ & \textbf{error} & \\
& & first order & second order \\
\hline
0 & 0 & .3678 (-12) & .4433 (-12) \\
0 & .1 & .4482 (-7) & .3508 (-11) \\
.1 & .1 & .4346 (-7) & .2544 (-11) \\
0 & .2 & .6342 (-7) & .4421 (-11) \\
.2 & 0 & .9241 (-7) & .8126 (-11) \\
\hline
\end{tabular}

Table 1

Using the product minimax methods of trigonometric order 1 and 2 with $h = .01$ and various squares centered at $\psi = -\pi/2$, $\omega = \pi/2$, the error is much larger but again is insensitive to small perturbations in the length of the sides of the squares. In Table 2 we list the Euclidean norm of the error at $\epsilon = .4$.

\begin{tabular}{|c|c|c|}
\hline
\textbf{length of side} & \textbf{error} & \\
& first order & second order \\
\hline
.4 & .3678 (-12) & .1469 (-6) \\
.8 & .3678 (-12) & .2348 (-5) \\
1.2 & .3678 (-12) & .1186 (-4) \\
1.6 & .3678 (-12) & .3728 (-4) \\
2 & .3678 (-12) & .9001 (-4) \\
\hline
\end{tabular}

Table 2

Note that the perturbations in the product minimax methods are larger than those allowed in the fitting methods. It is possible that the larger errors in the product minimax methods are due to the assumption that $\phi$ can be written as a product of 2 one variable functions. Also note that for the first order method, one always gets a good result since $\phi M^2$ always uses the center of the square.

Acknowledgment

The author would like to thank the NPS Foundation Research Program for its support of this research.

REFERENCES


APPENDIX

Here we show that the shallow water equations with topography have a solution of the form $e^{\lambda x}$, where $\lambda$ is complex. This system of equations consists of three equations with three forecast variables, $u$, $v$ and $\phi$.

The equations are:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0,$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0,$$

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x} [u(\phi-\phi_B)] + \frac{\partial}{\partial y} [v(\phi-\phi_B)] = 0,$$

where $\phi$ = $\Phi h$ is the geopotential height ($h$ = height of free surface), $\phi_B$ is the bottom topography (assumed to be independent of time), $u$, $v$ are the components of the wind velocity in the $x$, $y$ direction, respectively, and $f$ is the Coriolis parameter. Linearizing the equations by letting

$$u = U + u', \quad v = V + v', \quad \phi = \Phi + \phi'$$

where $U$, $V$ are the constant mean flow and $\Phi$ is independent of time. Assuming that $U$, $V$ are related to $\Phi$ via the geostrophic relations

$$U = -\frac{1}{f} \frac{\partial \Phi}{\partial y}, \quad V = \frac{1}{f} \frac{\partial \Phi}{\partial x}$$

(A.4)
one obtains the linear system (after dropping the primes):

\[
\frac{2u}{\partial t} + U \frac{2u}{\partial x} + V \frac{2u}{\partial y} - fV + \frac{2\Phi}{\partial x} = 0 ,
\]

(A.5)

\[
\frac{2v}{\partial t} + U \frac{2v}{\partial x} + V \frac{2v}{\partial y} + fU + \frac{2\Phi}{\partial y} = 0 ,
\]

(A.6)

\[
\frac{\partial \phi}{\partial t} + U \frac{\partial \phi}{\partial x} + V \frac{\partial \phi}{\partial y} + \frac{\partial (u\psi)}{\partial x} + \frac{\partial (v\psi)}{\partial y} = \frac{\partial \phi_B}{\partial x} + V \frac{\partial \phi_B}{\partial y} ,
\]

(A.7)

where \( \gamma = \phi - \phi_B \).

If the flow is assumed to be along the topography as in [15], then the right hand side of (A.7) is zero. In such a case, one can write the solution in the form

\[
u = u_0 e^{i(\xi x + \eta y - \sigma t)},
\]

\[
\psi = \psi_0 e^{i(\xi x + \eta y - \sigma t)},
\]

(A.8)

where

\[
\xi = \mu - 1\rho ,
\]

\[
\eta = \nu - 1\Theta .
\]

(A.9)

In order for (A.8) to be a solution for (A.5)-(A.7), one must have

\[
\lambda = -\sigma + \xi U + \eta V ,
\]

(A.10)

satisfying

\[
(i\lambda)^2 - i\eta(\xi\eta + \frac{\partial \psi}{\partial y}) - f[ -i\xi \psi - i\xi(\xi\eta + \frac{\partial \psi}{\partial y})]
\]

\[
+ (i\xi \psi + \frac{\partial \psi}{\partial x})(-i\psi + \xi \lambda) = 0 .
\]

(A.11)

The real and imaginary parts of

\[
\lambda = -\sigma + \mu U + \nu V ,
\]

(A.12)

\[
\lambda = -\rho U - \Theta V ,
\]

satisfy the following system of equations (after dropping nonlinear terms in \( \lambda \))

\[
\lambda_1 (f^2 + \gamma(\xi^2 + \eta^2)) + \lambda_1 (\xi \frac{\partial \psi}{\partial x} + \eta \frac{\partial \psi}{\partial y}) = f(\eta \frac{\partial \psi}{\partial x} - \xi \frac{\partial \psi}{\partial y}) ,
\]

(A.13)

In general, \( \lambda \) is complex and, thus, the shallow water equations have a solution in the class of problems to be discussed here.