DETERMINING INITIAL VALUES FOR STIFF SYSTEMS
WITH INCOMPLETE INITIAL DATA
by
Yech-Kuang Ning

A thesis submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
Mathematics

Approved:

UTAH STATE UNIVERSITY
Logan, Utah
1980
ACKNOWLEDGMENTS

Words are not adequate to express my gratitude and my deep appreciation to my major professor, Dr. David S. Watkins, for his help and patient guidance.

Also, I would like to thank Dr. Chris Coray and Dr. Michael Windham for their time and help on my committee.

Yeck-Kuang Ning
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>I. STIFF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS</td>
<td>1</td>
</tr>
<tr>
<td>II. ELEMENTARY REFLECTORS</td>
<td>7</td>
</tr>
<tr>
<td>III. INTRODUCTION TO THE INITIALIZATION PROBLEM</td>
<td>10</td>
</tr>
<tr>
<td>IV. CALCULATION OF INITIAL VALUES FOR A LINEAR SYSTEM</td>
<td>13</td>
</tr>
<tr>
<td>V. CALCULATION OF THE TRANSFORMING MATRIX</td>
<td>20</td>
</tr>
<tr>
<td>VI. THE NONLINEAR PROBLEM</td>
<td>30</td>
</tr>
<tr>
<td>VII. NUMERICAL RESULTS</td>
<td>34</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>39</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.</td>
<td>$y_2$ vs time $t$ for the nonhomogeneous case</td>
</tr>
<tr>
<td>2.</td>
<td>$y_2$ vs time $t$ for the homogeneous case</td>
</tr>
</tbody>
</table>
ABSTRACT
Determining Initial Values for Stiff Systems with Incomplete Initial Data
by
Yech-Kuang Ning, Master of Science
Utah State University, 1980

Major Professor: Dr. David S. Watkins
Department: Mathematics

Stiff differential equations frequently arise in physical problems due to the existence of greatly differing time constants. One way of solving stiff differential equations is taking the step size very small, another way is using an A-stable or stiffly stable implicit method. The purpose of this thesis is to discuss the method of determining initial values for stiff systems. The algorithm was motivated by a problem from ionospheric physics in which some of the initial values are unknown. If we give those unknown initial values an arbitrary number and integrated with those values, we have an initial transient. However, if we integrate with the correct initial values then there is no initial transient.

(44 pages)
CHAPTER I
STIFF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS

Definition

For any single linear differential equation \( y' = \lambda y + b \), we say this is a **stiff** differential equation, if \( \text{Re}(\lambda) \) is negative and large in absolute value. The nonlinear single differential equation \( y' = f(y) \) is **stiff** at \( y_0 \) if \( \text{Re}(\frac{\partial f}{\partial y}(y_0)) \) is negative and large.

The linear system \( y' = Ay + B(x) \) is **stiff**, if at least one of the eigenvalues of \( A \) has negative real part and is large in absolute value. The nonlinear system \( y' = f(y) \) is **stiff**, if the Jacobian matrix \( \frac{\partial f}{\partial y}(\hat{x},\hat{y}) \) has at least one eigenvalue having large negative real part.

For example the differential equation \( y' = -100y + 99e^{-x} \), which has \( \lambda = -100 \) is stiff. The general solution \( y = e^{-x} + ce^{-100x} \) is the sum of a slowly varying term and rapidly decaying term.

If we solve the problem by a numerical method, say Euler's method, this scheme advances from \( y_n \) to an approximation at \( x_n + h = x_{n+1} \) by \( y_{n+1} = y_n + h f(x_n, y_n) = y_n + hy'_n \). Since these are the linear terms of a Taylor series expansion at \( x_n \), the local truncation error of the method is
\[ h^2 y''(x_n)/2 + O(h^3) \], which makes the local truncation error approximately
\[ \varepsilon = |h^2 y''(x_n)/2| \quad \text{or} \quad h = \left(\frac{2\varepsilon}{|y''(x_n)|}\right)^{1/2}. \]

If the numerical solution is close to the true solution, we can deduce the behavior of \( h \) from

\[ y'' = e^{-x} + (100)^2 e^{-100x} \quad (1) \]

when \( x \) is small then

\[ h = \left(\frac{2\varepsilon}{|y''(x_n)|}\right)^{1/2} = \left(\frac{2\varepsilon}{100^2 e^{-100x}}\right)^{1/2} \]

\[ = \left(\frac{2\varepsilon}{100^2 e^{-x}}\right)^{1/2} \]

which make the step size \( h \) to be very small. When \( x \) is large the second exponential term of (1) disappears so that

\[ h = \left(\frac{2\varepsilon}{e^{-x}}\right)^{1/2} \]

the step size become large and independent of \( \lambda \).

Thus we have quantified the step size needed for accuracy. It must initially be small to resolve the rapid change of the transient, but after the transient phase it becomes large and independent of \( \lambda \). The interval of rapid change is called an initial transient [1].

There are two main factors affecting the size of the step--accuracy and stability. Accuracy refers to smallness of the local error, that is the error introduced in a single
step. Stability refers to errors not growing in subsequent steps. We have seen from the above example the accuracy is easily handled. Now we want to examine the stability. Before examining it, we need to define a few terms.

**Definition**

A method is **absolutely stable** for the given $h$ and $\lambda$, if a perturbation in the initial value $y_0$ leads to errors which tend to zero in subsequent steps.

The above definition means that if $y_0, y_1, y_2, \ldots$ is the numerical solution of $y' = \lambda y + B(x)$ with some method with fixed step size $h$ and $y(x_0) = y_0$, and we solve it again with $y(x_0) = z_0 = y_0 + \delta$, with the solution $z_0, z_1, z_2, \ldots$, then $|y_n - z_n| \to 0$ as $n \to \infty$, if we replace $y_n - z_n$ by $y_n'$, then we can simplify the above definition to:

**Definition**

A method is **absolutely stable** for a given $h$ and $\lambda$ if the numerical solution of $y' = \lambda y$, $y(x_0) = \delta$ with step size $h$ satisfies $y_n \to 0$ and $n \to \infty$ (for any $\delta$).

**Definition**

The **region of absolute stability** of a numerical method is the set of all points $h_\lambda C$ such that the method is absolutely stable.

For example: The region of absolute stability of Euler's method is

$$\{h_\lambda \in \mathbb{C} \mid |1 + h_\lambda| < 1\}$$

i.e., the disc of radius 1 center at (-1,0).
Definition

A method is called A-stable if its region of absolute stability contains the left half plane.

Now let us see how stability relates to stiffness.

We consider the simple case $y' = \lambda y + b$, $y(0) = y_0$. The exact solution is $y = (y_0 + b/\lambda) e^{\lambda x} - b/\lambda$. Now solve the differential equation numerically using Euler's method.

$$Y_{n+1} = y_n + h(\lambda y_n + b)$$

Solve this $\Rightarrow$

$$Y_{n+1} = (1 + h\lambda) y_n + hb$$

$$Y_1 = (1 + h\lambda)y_0 + hb$$

$$Y_2 = (1 + h\lambda)Y_1 + hb$$

$$= (1 + h\lambda)^2 y_0 + (1 + h\lambda) hb + hb$$

$$= (1 + h\lambda)^2 y_0 + hb(1 + (1 + h\lambda))$$

$$Y_3 = (1 + h\lambda)^3 y_0 + hb(1+(1+h\lambda) + (1+h\lambda)^2)$$

$$\vdots$$

$$Y_n = (1+h\lambda)^n y_0 + hb\left(\sum_{k=0}^{n-1} (1+h\lambda)^k\right)$$

$$= (1+h\lambda)^n y_0 + hb\left(\frac{(1+h\lambda)^n - 1}{h\lambda}\right)$$

$$Y_n = (1+h\lambda)^n (y_0+b/\lambda) - b/\lambda$$

Let $z_n = (y_0 + b/\lambda)(1 + h\lambda)^n$, $w_n = -b/\lambda$. Then

$y_n = z_n + w_n$ and $z_n$ is the solution of $z_{n+1} = (1+h\lambda) z_n$. 
\[ z_0 = (y_0 + b/\lambda), \text{ } w_n \text{ satisfies } w_{n+1} = (1+h\lambda) w_n + hb, \]
\[ w_0 = -b/\lambda. \]

This shows that the solutions of the difference equation behave analogously to the solutions of the differential equation, \( z_n \) is the numerical solution of the homogeneous differential equation \( y' = \lambda y \) by Euler's method, so

\[ z_n = (y_0 + b/\lambda)(1 + h\lambda)^n \]

is a good approximation to \((y_0 + b/\lambda)e^{\lambda x}\), if \(|h\lambda|\) is very small. For a stiff problem we must take \( h \) unreasonably small if we wish to make \( h\lambda \) small. However this is not always necessary. If we have passed the transient phase the term \((y_0 + b/\lambda)e^{\lambda x}\) is insignificant, so \( z_n \) need not approximate \((y_0 + b/\lambda)e^{\lambda x}\) well, but it should at least be insignificant. This will be guaranteed if \( z_n \to 0 \). But \( z_n \to 0 \) if and only if the method is absolutely stable for the given \( h \) and \( \lambda \). Thus, in order to get a good solution to a stiff problem \( h \) must be small enough that \( h\lambda \) lies in the region of absolute stability. Since the region of absolute stability of Euler's method is the disc of radius 1 centered at \((-1,0)\), \( h \) must be very small in order to be stable. For example, if \( \lambda = -100 \), we must take \( h < 1/50 \).

However there is some method suitable for stiff equations, for example the backward Euler method is A-stable, that is its region of absolute stability includes the whole left half plane. We do not need to have a very small step.
size, as $h \lambda$ always lie in the region of absolute stability. The backward Euler method $y_{n+1} = y_n + hf(x_{n+1}, y_{n+1})$ is implicit, so when solving the problem each step will be more expensive. But since the stability requirement is so much more stringent than the accuracy requirement, we can pick an $A$-stable or stiffly stable (implicit) method for a stiff system even though each step is much more expensive. More extensive information about stiff systems is given by Gear [2], and Shampine and Gear [1].
CHAPTER II
ELEMENTARY REFLECTORS

The elementary reflector or Householder transformation is a type of orthogonal transformation used frequently in numerical analytic calculations. Since we will be using them later, we include an introductory section here. A more extensive treatment of elementary reflectors is given by Stewart [3].

**Definition**

Let \( u \in \mathbb{R}^n \) with \( \|u\|_2 = 1 \) then \( U = I - 2uu^T \) is the Elementary Reflector corresponding to \( u \).

In applications it frequently happens that the vector \( u \) is given in an unnormalized form. We can extend the above definition by letting \( u' = u/\|u\|_2 \) then \( \|u'\|_2 = 1 \) and \( U = I - 2u'u'^T \) is the elementary reflector corresponding to \( u \).

**Theorem 2.1**

Let \( U \) be an elementary reflector, then \( U \) is symmetric \( (U = U^T) \), orthogonal \( (U^TU = I) \) and involutory \( (U^2 = I) \).

**proof:** Clearly it suffices to prove any two of those properties.
\[ u^T = (I - 2uu^T)^T \]
\[ = I^T - (2uu^T)^T \]
\[ = I - 2uu^T = U. \]

\[ U^T U = UU \]
\[ = (I - 2uu^T)(I - 2uu^T) \]
\[ = I - 2uu^T - 2uu^T + 4uu^T uu^T = \|u\|_2^2 = u^T u = 1 \]
\[ = I - 4uu^T + 4uu^T = I. \]

**Theorem 2.2**

Let \( x \in \mathbb{R}^n \), \( \sigma = \pm \|x\|_2 \) and suppose that \( x \neq -\sigma e_n \) where \( e_n = (0,0, \ldots, 0,1)^T \). Let \( u = x + \sigma e_n \) and \( \pi = \frac{1}{2}\|x\|_2^2 \).

Then \( U = I - \pi^{-1} uu^T \) is an elementary reflector and \( Ux = -\sigma e_n \).

**proof:** Let \( u' = u/\|u\|_2 \) then \( \|u'\|_2 = 1 \) and

\[ U = I - \pi^{-1} uu^T = I - 2 \frac{uu^T}{\|u\|_2^2} = I - 2u'u^T \]

so \( U \) is an elementary reflector.

Since \( x^T x = \sigma^2 \), let \( x = (x_1, x_2, \ldots, x_n)^T \)

\[ \pi = \frac{1}{2}(x + \sigma e_n)^T(x + \sigma e_n) \]
\[ = \frac{1}{2}(x^T x + \sigma x^T e_n + \sigma x e_n^T + \sigma^2 e_n^T e_n) \]
\[ = \frac{1}{2}(\sigma^2 + 2\sigma x_n + \sigma^2) \]
\[ = \sigma^2 + \sigma x_n. \]
Hence \( Ux = x - \frac{uu^T x}{\pi} \) = \( x - \frac{(x + \sigma e_n)(x + \sigma e_n)^T x}{\sigma^2 + \sigma x_n} \)

= \( x - \frac{(x + \sigma e_n)(x^T x + \sigma^2 x_n)}{\sigma^2 + \sigma x_n} \)

= \( x - (x + \sigma e_n) \)

= \(-\sigma e_n\).

The statement of Theorem 2.2 nearly tells us how to compute \( u \) and \( \pi \). We need only decide on the sign of \( \sigma \).

Then

\[ u_i = x_i \quad i = 1, 2, \ldots, n-1 \]

\[ u_n = x_n + \sigma \]

and

\[ \pi = \sigma (\sigma + x_n) = \sigma^2 x_n + \sigma^2. \]

If \( \sigma \) and \( x_n \) have different signs, then cancellation can occur in the computation of \( u_n \). Hence we take \( \sigma \) to have the same sign as \( x_n \). In applications it is often the case that having formed \( u \) we no longer need \( x \). Hence the components of \( u \) may be stored over the components of \( x \). But the first \( n-1 \) components of \( u \) are the same as the first \( n-1 \) components of \( x \); hence only the last component of \( x \) need be altered in computing \( u \).
CHAPTER III

INTRODUCTION TO THE INITIALIZATION PROBLEM

Consider a stiff system of nonlinear ordinary differential equations

\[ y' = f(x,y) \quad y(x_0) = W. \]

Let

\[ W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \]

where \( W_1 \) is a vector of \( k \) known values, and \( W_2 \) is a vector of \( n-k \) unknown values. As we know, the solution to a stiff system usually has an initial transient. In the rest of the thesis, we present an algorithm which determines the unknown initial values \( W_2 \) in such a way that the solution does not have an initial transient.

The algorithm was developed to meet a need which arose when Schunk and Watkins [4] studied a system of transport equations for protons and electrons in the extreme upper ionosphere. The system consists of first order ordinary differential equations for the eight unknowns electron density, electron drift velocity, electron and proton temperature, heat flow and stress. The equations can be solved numerically by starting at a given altitude, at which values for all eight unknowns are given, and marching radially.
outward a distance of some ten thousand kilometers using a standard numerical method. Schunk and Watkins have solved this system for a wide range of initial conditions. The system is stiff, so they used a stiffly stable method. Reasonable starting values for density, velocity and temperatures are known from satellite measurements, and the heat flows can be estimated from measurements of temperature gradients. However, the stress values are unknown. In the present context stress is a measure of the difference in temperature in directions parallel and perpendicular to the earth's magnetic field. The variable which they have called temperature is a weighted average of the parallel and perpendicular temperatures. In all satellite measurements to date the measurements have been made on the assumption that the temperature is the same in all directions; that is, the stresses have been assumed to be zero. They believed that the stresses should be negligible at the initial altitude, so they set them to zero. When they integrated the equations they found that the solution has a large initial transient during which the stresses and temperatures change at rates on the order of 100K(km)$^{-1}$. Space craft have not observed such large changes in this altitude region. Furthermore, the starting altitude was chosen somewhat arbitrarily; it could have been lower or higher. There is no reason why large changes should take
place near this artificial boundary and not elsewhere. The transient is evidently caused by the imposition of the "wrong" conditions at the artificial boundary. Therefore, we need a procedure for calculating the "right" initial stress values, namely those which give rise to a solution which does not have an initial transient.
CHAPTER IV
CALCULATION OF INITIAL VALUES FOR
A LINEAR SYSTEM

In this section, we will develop an algorithm for linear systems which calculates the vector \( \mathbf{w}_2 \) by a direct procedure.

Consider the linear problem

\[
y' = A \mathbf{y} + \mathbf{b} \quad \mathbf{y}(x_0) = \mathbf{w} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}. \tag{2}
\]

This is a system of \( n \) linear ordinary differential equations in \( n \) unknowns, \( \mathbf{w}_1 \) is a vector of \( k \) known initial values, \( \mathbf{w}_2 \) is a vector of \( i \) unknown initial values, and \( k+i = n \).

Assume the system is stiff and the eigenvalues of \( A \) can be broken into two sets:

1. A set of "stiff" eigenvalues (having large negative real part).
2. A set of "small" eigenvalues (which are small in absolute value in comparison with the stiff eigenvalues).

We will see that in order to determine \( \mathbf{w}_2 \) in such a way that the solution does not have an initial transient, we must partition the eigenvalues so that the number of stiff eigenvalues is equal to the number of components of \( \mathbf{w}_2 \).
However, because of the vagueness of such terms as "small" and "large", a set of eigenvalues can sometimes be partitioned in several ways. For example, suppose that $A$ has eigenvalues $-10^7$, $-10^5$, $-10^2$, $-2$, 0. Then we can declare either one, two or three eigenvalues to be stiff, depending on whether there are one, two or three unknown initial values.

One way to determine $W_2$ is by transforming $A$ to a diagonal form. This works whenever $A$ is simple (i.e., $A$ has $n$ linearly independent eigenvectors). But it does not work if $A$ is defective (i.e., nonsimple). More importantly, the method does not work well when $A$ is close to being defective.

If $A$ is simple, there is a nonsingular matrix $V$ such that

$$A = VDV^{-1}$$

and $D$ is diagonal, with main diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ (the eigenvalues). Let $v^{(j)}$ denote the $j$th column of $V$. Then $v^{(j)}$ is the eigenvector of $A$ corresponding to $\lambda_j$.

Let $\hat{y} = V^{-1}y$, $\hat{b} = V^{-1}b$ and $\hat{w} = V^{-1}w$.

Then the system of equations can be rewritten as

$$\hat{y}' = D\hat{y} + \hat{b} \quad \hat{y}(x_0) = \hat{w}.$$  

The $j$th equation in this uncoupled system is
\[ \dot{y}_j = \lambda_j y_j + \hat{b}_j \]

which has general solution

\[
y_j(x) = \begin{cases} 
  c_j \exp(\lambda_j (x-x_0)) - \hat{b}_j \lambda_j^{-1} & \text{if } \lambda_j \neq 0 \\
  c_j + \hat{b}_j (x-x_0) & \text{if } \lambda_j = 0 
\end{cases}
\]  

(3)

where \( \hat{b}_j \) is the \( j \)th component of \( b \), and \( c_j \) is a constant of integration.

The general solution to \( y' = Ay + b \) is then given by

\[ y = V\dot{y} \text{ or } \]

\[ y = \sum_{j=1}^{n} v^{(j)} \hat{y}_j(x). \]  

(4)

The initial transient is caused by the rapid decay of those terms \( c_j \exp(\lambda_j (x-x_0)) \) which correspond to stiff eigenvalues. We can eliminate the initial transient by adjusting the initial conditions in such a way that the resulting solution has \( c_j = 0 \) for all stiff eigenvalues; that is \( c_{k+1} = c_{k+2} = \ldots = c_n = 0 \). From (3) we see that \( \hat{y}_j(x_0) = c_j - u_j \lambda_j^{-1} \hat{b}_j \) if \( \lambda_j \neq 0 \) and \( u_j = 0 \) if \( \lambda_j = 0 \).

These equations can be expressed in matrix form as \( \dot{y}(x_0) = C - Mb \) where \( M \) is the diagonal matrix whose \( j \)th main diagonal entry is \( u_j \). Thus \( y(x_0) = V(C-Mb) \) we can write this matrix equation in partitioned form.
where $d = \begin{bmatrix} d \\ -e \end{bmatrix}$, $e$ consists of those components of $C$ which are to be set to zero and $\hat{M} = \begin{bmatrix} r \\ -s \end{bmatrix}$. Setting $e = 0$ in (5) we get

$$W_1 = V_{11}(d-r) - V_{12}s.$$  \hspace{1cm} (6)

$$W_2 = V_{21}(d-r) - V_{22}s.$$

If $V_{11}$ is nonsingular we can solve the first equation of (6) for $d-r$, then substitute $d-r$ into the second equation to calculate $W_2$. This gives the unique $W_2$ for which $e = 0$ and

$$W_2 = V_{21}V_{11}^{-1}(W_1 + V_{12}s) - V_{22}s.$$

If $V_{11}$ is ill-conditioned then most choices of $W_1$ will lead to a large and inaccurate $W_2$, unless $V_{21}$ is unusually small. The ill-conditioning is because of the transformation matrix $V$ is ill-conditioned. The ill condition of $V$ is caused by $A$ being nearly defective.

However we can avoid ill-conditioning by using an orthogonal similarity-transformation $Q$ such that

$$A = QAQ^{-1} = QAQ^T \quad \text{where} \quad \hat{A} = \begin{bmatrix} A_{11} & A_{12} \\ \\ 0 & A_{22} \end{bmatrix}.$$
In this case we cannot hope to make $A$ diagonal in general. However Schur's theorem (Stewart [1]) tells us that it is possible to make $A$ block triangular.

Suppose we have a routine which can find an orthogonal matrix $Q$ and a block triangular matrix

$$B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

such that $A = QBQ^{-1} = QBQ^T$

$B_{11}$ is $k$ by $k$ and has the small eigenvalues of $A$ and $B_{22}$ is $i$ by $i$ and has the stiff eigenvalues of $A$.

Let $\hat{y} = Q^T y$, $\hat{b} = Q^T b$ and $\hat{W} = Q^T W$ then the system (2) can be written $\hat{y}' = B\hat{y} + \hat{b}$ $\hat{y}(x_0) = \hat{W}$, or in the partly decoupled form

$$\begin{align*}
\hat{y}_1' &= B_{11}\hat{y}_1 + B_{12}\hat{y}_2 + \hat{b}_2 \\
\hat{y}_1(x_0) &= \hat{W}_1 & (7) \\
\hat{y}_2' &= B_{22}\hat{y}_2 + \hat{b}_2 \\
\hat{y}_2(x_0) &= \hat{W}_2 & (8)
\end{align*}$$

where all the stiffness is concentrated in the second system (8). We can eliminate the transient by making appropriate adjustments involving the second system. Since $B_{22}$ has only stiff eigenvalues, the solutions of (8) have the form

$$\text{constant term + rapidly decaying terms}.$$
The rapidly decaying terms cause the initial transient, so in order to get a solution which does not have an initial transient, we should find that initial value $\hat{W}_2$ for which the rapidly decaying terms are all zero. That is, we seek the value of $W_2$ for which the solution to (8) is constant. Such a solution will satisfy

$$\hat{y}_2(x) = \hat{W}_2 \text{ and } \hat{y}_2'(x) = 0 \quad (\forall x)$$

and we get such a solution if and only if

$$0 = B_{22} \hat{W}_2 + \hat{b}_2. \quad (9)$$

Since $B_{22}$ is nonsingular, we can always solve this system for a unique $\hat{W}_2$. (In fact $B_{22}$ is generally well conditioned because it does not contain any small eigenvalues.)

Now that we know the value of $\hat{W}_2$ which gives a solution with no initial transient, we must find the corresponding value of $\hat{W}_2$, we have $W = Q\hat{W}$

$$W_1 = Q_{11} W_1 + Q_{12} W_2 \quad (10)$$

$$W_2 = Q_{21} W_1 + Q_{22} W_2 \quad (11)$$

$W_1$ is fixed, as is $W_2$. If $Q_{11}$ is nonsingular then we can solve (10) for $\hat{W}_1$ and substitute $\hat{W}_1$ into (11) to find $\hat{W}_2$. Combine (9), (10), and (11) and let

$$Q^T = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix},$$
then $\hat{b}_2 = U_2 b$. We get

$$W_2 = Q_{21} Q_{11}^{-1} W_1 + (Q_{21} Q_{11}^{-1} Q_{12} - Q_{22}) B_{22}^{-1} U_2 b.$$  (12)

We will use this explicit expression for $W_2$ in the section when we discuss the convergence of the algorithm for nonlinear problem.

If $Q_{11}$ is singular the algorithm fails, but this is only because the problem itself has no solution. In this case, for most choices of $W_1$ the equation (10) has no solution, for these $W_1$ there is no value of $W_2$ for which the solution does not have an initial transient.

However $Q_{11}$ will almost never be singular, but it may be ill-conditioned. In this case, most choices of $W_1$ will lead to a large and inaccurate $\hat{W}_1$, which in turn will lead to a large and inaccurate $W_2$, unless $Q_{21}$ is unusually small. The ill-conditioning is not an artifact of the algorithm; it is inherent in the actual problem. It can generally be detected by the unreasonably large size of the solution $W_2$. Also, in the algorithm for nonlinear problems given below, ill-conditioning of $Q_{11}$ may cause nonconvergence of the iterations.
CHAPTER V
CALCULATION OF THE TRANSFORMING MATRIX

From the last section we wish to find an orthogonal transformation \( Q \) such that

\[
A = QBQ^T
\]

where

\[
B = \begin{bmatrix}
B_{11} & B_{12} \\
0 & B_{22}
\end{bmatrix}
\]

Watkins [6] has considered the case in which the given system only has one stiff eigenvalue. Since there is only one stiff eigenvalue and it is much larger in absolute value than all other eigenvalues, the power method is the obvious choice here for calculating the eigenvalue and an associated eigenvector. For the power method, we need to have a starting vector. One possibility is to inspect the main diagonal entries of \( A \), find the largest (in absolute value) negative one, say \( a_{jj} \), and take as a starting vector \( \mathbf{e}_j = (0,0,\ldots,0,1,0,\ldots,0)^T \). This choice is motivated by the Gerschgorin Disk Theorem (Stewart [3]).

Having chosen a starting value we apply the power method to \( A^T \). If we normalize the vector at each step, the method
will converge to an eigenvector $V$ with $A^TV = \lambda V$ and $\|V\|_2 = 1$.
Now determine the elementary reflector which maps $e_n$ to $\pm V$
and call it $Q$. The procedure for constructing an elementary
reflector is given in Chapter II.

Now I claim that $Q$ is the desired transforming matrix.
Clearly the last column of $Q$ is $\pm V$, and $Q^{-1} = Q^T$ implies
the other columns of $Q$ are orthogonal to $V$.

$$B = Q^TAQ = \begin{bmatrix}
\hat{Q}^T \\
\pm V^T
\end{bmatrix}
A \begin{bmatrix}
\hat{Q} \\
\pm V
\end{bmatrix} =
\begin{bmatrix}
\hat{Q}^TA\hat{Q} & Q^TAV \\
\pm V^TAQ & V^TAV
\end{bmatrix}$$

Since $A^TV = \lambda V$ we have $V^TA = \lambda V^T$. Thus $V^TA\hat{Q} = \lambda V^T\hat{Q} = 0$
because the columns of $Q$ are orthogonal to $V$. Also $V^TAV = \lambda V^TV = \lambda$, thus

$$Q^TAQ = \begin{bmatrix}
B_{11} & B_{12} \\
0 & \lambda
\end{bmatrix} \quad \text{where} \quad B_{11} = \hat{Q}^TA\hat{Q} \quad B_{12} = \pm \hat{Q}^TAV$$

If we had applied the method to $A$ instead of $A^T$, we would
have gotten a block lower triangular matrix rather than a
block upper triangular one, so we apply the power method to
$A^T$ instead of $A$.

Now how do we know that this is the only stiff eigenvalue?
This can be done by checking the rate of convergence
of the power method iterations. Fast convergence indicates
only one stiff eigenvalue. Because suppose $A$ has a complete
system of eigenvectors $x_1, x_2, \ldots, x_n$ corresponding to the
eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ which satisfy $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|$. Since the vectors $x_1, x_2, \ldots, x_n$ form a basis for $\mathbb{C}^n$, any vector $q_0$ may be expressed in the form

$$q_0 = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n$$

It follows that if we define

$$q_k = \frac{A^k q_0}{\lambda_1^k}$$

$$q_k = c_1 x_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k c_2 x_2 + \ldots + \left(\frac{\lambda_n}{\lambda_1}\right)^k c_n x_n \quad (13)$$

Since the largest of the ratios $|\lambda_i/\lambda_1|$ is $|\lambda_2/\lambda_1|$, if $\lambda_2$ is another stiff eigenvalue then $|\lambda_2/\lambda_1|$ will close to 1 so the second term in (13) will converge slowly to zero. If the system only has one stiff eigenvalue $|\lambda_i/\lambda_1| \ll 1$ for $i = 2, 3, \ldots, n$, then we only need to take few steps because $|\lambda_i/\lambda_1|^k \to 0$ rapidly for $i = 2, 3, \ldots, n$.

If there is another stiff eigenvalue we can get it by applying the same algorithm to $B_{11}$. Then a nonsingular elementary reflector $D$ of dimension $n-1$ by $n-1$ is found so that

$$D^T B_{11} D = \begin{bmatrix} C_{11} & C_{12} \\ 0 & \lambda_2 \end{bmatrix}$$
and then with $Q_2 = \begin{bmatrix} D & 0 \\ 0 & 1 \end{bmatrix}$, we have

$$\begin{bmatrix} & D & 0 \\ 0 & 1 & \end{bmatrix} Q^T AQ \begin{bmatrix} & D & 0 \\ 0 & 1 & \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & \\ 0 & \lambda_2 & D_{12} \\ \end{bmatrix}.$$  

This process is carried out until all the stiff eigenvalues come out, and the transformation matrix is $Q_1 Q_2 Q_3 \cdots Q_k$.

**Cost of one iteration**

For finding the eigenvector $V$ each iteration costs $n^2$ multiplications. The cost of calculating $Q$ is about $n^2 + 3n + 1$.

Now consider the case when the number of stiff eigenvalues is not equal one, say $i$. We will try to isolate the $i$ stiff eigenvalues simultaneously.

If $A^T$ is simple, then $R^n$ has a basis $V_1, \ldots, V_n$ consisting of eigenvectors of $A^T$. If $A^T$ is not simple, then we have to include generalized eigenvectors in order to get a basis $V_1, \ldots, V_n$. Let $V_1, \ldots, V_k$ be those which correspond to small eigenvalues, and let $V_{k+1}, \ldots, V_n$ be those corresponding to stiff eigenvalues. Let $U = \langle V_1, \ldots, V_k \rangle$, $T = \langle V_{k+1}, \ldots, V_n \rangle$ if we can determine the subspace $T$, then we can make the desired similarity transformation.

Let $S$ be any $i$ dimensional subspace of $R^n$ whose intersection with $U$ is $\{0\}$. Since $U$ contains the generalized null space of $A^T$, $S$ contains no generalized null vectors.
(The generalized null space is the set of all \( V \) such that \((A^T)^jV = 0\) for some \( j \), equivalently it is the linear span of the generalized eigenvectors corresponding to the eigenvalue 0. If 0 is not an eigenvalue, then the generalized null space is \( \{0\} \).) Since \( S \) contains no generalized null vectors, the spaces in the sequences

\[
S, A^T S, (A^T)^2 S, (A^T)^3 S, (A^T)^4 S, \ldots
\]

all have dimension \( i \). In fact, this sequence of subspaces approaches \( T \), since for each element of \( S \), as \( A^T \) is repeatedly applied, the components in the \( U \) direction become small in relation to those in the \( T \) direction. However we need to find \( S \) first. One possibility is apply \( A^T \) to the space \( \langle e_1, \ldots, e_n \rangle \), then pick those \( i \) columns which have the biggest norm (i.e., pick \( i \) terms from \( A^T e_1, \ldots, A^T e_n \)), let \( w_1, \ldots, w_i \) be the columns we picked, but which will not be orthonormal in general. We apply the Gram-Schmidt procedure to this basis to produce an orthonormal basis. We remind the reader that the Gram-Schmidt procedure is as follows:

For \( k = 1, \ldots, i \)

\[
\hat{w}_k = w_k - \sum_{j=1}^{k-1} \frac{(w_k, z_j) z_j}{\|w_k\|^2}
\]

\[
z_k = \frac{\hat{w}_k}{\|\hat{w}_k\|^2}
\]
then \( z_1', \ldots, z_i \) is an orthonormal basis of \( S \). After we find the basis of \( S \), apply \( A^T \) to the subspace \( S \), then \( A^T z_1', \ldots, A^T z_i \) is a basis of \( A^T S \) which are not orthonormal. We apply Gram-Schmidt procedure again. Then get \( w_1', \ldots, w_i \) an orthonormal basis of \( A^T S \). i.e.,

\[
\hat{z}_k = A^T z_k - \sum_{j=1}^{k-1} (A^T z_k, w_j) w_j
\]

\[
w_k = \frac{\hat{z}_k}{\|\hat{z}_k\|_2}
\]

If \( \|\hat{z}_k\|_2 \) is very small for some \( k \), then \( A^T z_k \) is nearly contained in \( \langle A^T z_1', \ldots, A^T z_{k-1} \rangle = \langle w_1', \ldots, w_{k-1} \rangle \). That is \( A^T z_1', \ldots, A^T z_k \) are nearly linearly dependent. What this implies is that some of the eigenvalues which were thought to be stiff are in fact much smaller than the other stiff eigenvalues. In this case it is better to find the stiffest ones first, then do a separate iteration to find the less stiff ones. Thus, if \( \|z_k\| \) is small, we continue the computation with only the \( k-1 \) dimensional space \( \langle A^T z_1', \ldots, A^T z_{k-1} \rangle = \langle w_1', \ldots, w_{k-1} \rangle \).

It follows from the previous paragraph that we may save computer time by doing the computations in the following order:
For $k = 1, \ldots, i$

compute $A^T z_k$

compute $\hat{z}_k$.

If $\|\hat{z}_k\|_2$ is small relative to $\|z_1\|_2$ stop,
otherwise calculate $w_k$ and continue.

Once we have $w_1, \ldots, w_i$, we can apply the same procedure repeatedly to get $(A^T)^2 S, (A^T)^3 S$, and so on.

We need a stopping criterion. Let

$$(A^T)^n S = \langle z_1', \ldots, z_i' \rangle$$

$$(A^T)^{n+1} S = \langle w_1', \ldots, w_i' \rangle$$

where $z_1, \ldots, z_i$ and $w_1, \ldots, w_i$ are orthonormal bases.

To measure the closeness of these two spaces we measure the distance of each of the vectors $w_j$ to $\langle z_1', \ldots, z_i' \rangle$ as follows.

The best approximation to $w_j$ from the space $\langle z_1', \ldots, z_i' \rangle$ is

$$\hat{w}_j = \sum_{k=1}^i (w_j, z_k) z_k.$$

Let $u_j = \|w_j - \hat{w}_j\|_2$ if $\max \{u_1, \ldots, u_i\}$ is sufficiently small we accept $\langle w_1', \ldots, w_i' \rangle$ as $T$. This gives us not only an approximation to $T$ but also an orthonormal basis $w_1', w_2', \ldots, w_i'$. 
Cost of one iteration

The main cost is that of multiplying \( A^T \) by \( i \) vectors. This costs \( n^2i \) multiplications. The cost of the Gram-Schmidt orthonormalization is about \( ni^2 \) multiplications, and the cost of the convergence check is about \( ni^2 \) multiplications.

Now that we have \( T = \langle w_1, \ldots, w_i \rangle \), let us consider how to construct a transforming matrix. Let \( W \) be the \( n \) by \( i \) matrix whose columns are \( w_1, \ldots, w_i \), and let \( R \) be an \( n \) by \( k \) matrix whose columns are orthonormal and orthogonal to \( w_1, \ldots, w_i \). Then \( A = [R | W] \) is an \( nxn \) orthogonal matrix.

\[
Q^T A Q = \begin{bmatrix}
R^T & \quad -
\end{bmatrix} A \begin{bmatrix}
R^T A R \\
R^T A W \\
W^T A R \\
W^T A W \\
\end{bmatrix}
\]

First we will show that \( W^T A R = 0 \). Since \((W^T A R)^T = R^T A^T W\), the columns of \( W \) lie in the space \( T \) and \( T \) is invariant under \( A^T \). Thus the columns of \( A^T W \) lie in \( T \). But the rows of \( R^T \) are orthogonal to \( T \), so \( R^T (A^T W) = 0 \).

Furthermore the eigenvalues of \( W^T A W \) are just the stiff eigenvalues of \( A \). To verify this we again look at the transposed matrix \((W^T A W)^T = W^T A^T W\). \( W^T \) is a left inverse for \( W \); that is \( W^T W = I_{ixi} \). Thus \( \lambda I - W^T A^T W = \lambda W^T W - W^T A^T W = W^T (\lambda I - A^T) W \). Furthermore, the range of \( W \) is just its column space, which is \( T \). \( A^T \) and \( \lambda I - A^T \) leave \( T \) invariant and \( W^T |_T \) is one-to-one. Thus \( \lambda \) is an eigenvalue of \( W^T A^T W \).
if and only if \((\lambda I - A^T)_T\) is singular, i.e., if and only if 
\(\lambda\) is an eigenvalue of \((A^T)_T\). The eigenvalues of \((A^T)_T\) are 
just the stiff eigenvalues of \(A^T\) (and of \(A\)). Thus if we 
let \(B_{11} = R^T AR\), \(B_{12} = R^T AW\), and \(B_{22} = W^T AW\) then 
\[
A = QBQ^T \quad \text{Where} \quad B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}
\]

\(B_{22}\) contains the stiff eigenvalues. \(B_{11}\) contains the small 
eigenvalues.

**Cost of computation**

From Chapter IV we see that we do not need to know \(B_{11}\) 
or \(B_{12}\), but we do need to know \(B_{22}\) and \(Q\). The cost of cal­
culating \(B_{22} = W^T AW\) is \(n^2 i + ni^2\) multiplications. In order 
to know \(A = [R W]\). We must calculate \(R\). One way to do 
this is to start with the orthonormal vectors \(w_1, \ldots, w_i\), 
which comprise \(W\). Augment this basis with vectors from the 
set \(e_1, e_2, \ldots, e_n\). Then use Gram-Schmidt to calculate 
orthonormal \(w_{i+1}, \ldots, w_n\) which are also orthogonal to 
\(w_1, \ldots, w_i\). Define \(R\) to be the matrix whose columns are 
\(w_{i+1}, \ldots, w_n\). The cost is \(n^3 - ni^2\) multiplications.

If we choose this method, then there is a remote possi­
bility that one of the \(e_1, \ldots, e_n\) which we choose will be 
a linear combination of previously computed \(w\)'s. Suppose we 
have computed \(w_1, \ldots, w_k\) and we are going to use \(e_j\) to com­
pute \(w_{k+1}\), we then have
\[ w_{k+1} = e_j - \sum_{i=1}^{k} (e_j, w_i) w_i \]

\[ w_{k+1} = \frac{\hat{w}_{k+1}}{\|w_{k+1}\|_2} \]

If \( e_j \) is in the space spanned by \( w_1, \ldots, w_k \), then \( w_{k+1} = 0 \) and the procedure breaks down. In this case we must discard \( e_j \) and try, say \( e_{j+1} \) instead. In practice we should reject \( e_j \) not only when \( w_{k+1} = 0 \), but also whenever \( w_{k+1} \) is very small. This will be the case whenever \( e_j \) is close to being a linear combination of \( w_1, \ldots, w_k \).

Finally, if there is another stiff eigenvalue we can get it by applying the same algorithm to \( B_{11} \) the transforming matrix will be

\[ Q_1 \cdot \begin{pmatrix} Q_2 & 0 \\ 0 & I \end{pmatrix} \]

where \( Q_1 \) is the elementary reflector associated with \( A \), and \( Q_2 \) is the elementary reflector corresponding to \( B_{11} \).
CHAPTER VI
THE NONLINEAR PROBLEM

Consider a nonlinear system

\[ y' = f(y) \quad y(x_0) = W. \quad (14) \]

Assume that \( f \) is twice continuously differentiable. This guarantees that the initial value problem has a unique solution in a neighborhood of \( W \). It also guarantees that \( f \) is well approximated by its first degree Taylor polynomial near \( W \). We may consider the linear problem

\[ y' = f(W) + \frac{\partial f}{\partial y}(W) (y - W) \quad y(x_0) = W \quad (15) \]

in place of (14). The differential equation in (15) has the form \( y' = Ay + b \) where \( A = \frac{\partial f}{\partial y}(W) \) and \( b = f(W) - AW \).

We cannot apply the procedure developed in Chapter IV to this linear problem because \( A \) and \( b \) depend on \( W \) and some of the components of \( W \) are unknown. Instead we must adopt an iteration procedure: We make an initial guess

\[
W(0) = \begin{bmatrix} W_1 \\ W(0)_2 \end{bmatrix}.
\]
then given an \( m^{th} \) iterate
\[
\begin{bmatrix}
W_1^{(m)} \\
W_2^{(m)}
\end{bmatrix}
\]
we calculate \( A^{(m)} = \frac{\partial f}{\partial y}(W^{(m)}) \) and \( b^{(m)} = f(W^{(m)}) - A^{(m)}W^{(m)} \), and use the procedure of Chapter IV to find \( W_2^{(m+1)} \) such that the solution of \( y' = A^{(m)}y + b^{(m)}, y(x_0) = W^{(m+1)} \) does not have an initial transient. Hopefully the iterates will converge to a set of initial conditions \( W \) such that the solution to (15) does not have an initial transient.

We will make no attempt to give rigorous sufficient conditions for convergence. Instead we will point out those factors which help or hinder convergence. We have here a fixed point iteration, \( W_2^{(m+1)} = \phi(W_2^{(m)}) \), in which \( \phi \) can be gotten from (12)
\[
W_2^{(m+1)} = \phi(W_2^{(m)}) = (Q_{21}Q_{11}^{-1},(m)W_1
\]
\[
+ \left[(Q_{21}Q_{11}^{-1}Q_{12} - Q_{22})B_{22}^{-1}u_2b\right]^{(m)}
\]
where the superscript \( (m) \) marks those quantities which are functions of \( W_2^{(m)} \). Although it is possible to obtain an explicit upper bound on
\[
W_2^{(m+1)} - W_2^{(m)} = \phi(W_2^{(m)}) - \phi(W_2^{(m-1)})
\]
in terms of quantities such as \( Q_{22}^{(m)} - Q_{22}^{(m-1)}, b^{(m)} - b^{(m-1)} \) etc., we can understand \( \phi \) without doing that. Clearly \( \phi \) will be a contraction if \( b \) and the eigensystem of \( A \) do not vary too rapidly as functions of \( W_2^{(m)} \). Large variations in \( A \) do not necessarily force large variations in the eigenvectors. On the other hand, if \( A \) varies slowly and the eigensystem is well-conditioned, then the eigensystem will vary slowly. The rate of variation of \( b \) and \( A \) is a measure of nonlinearity of the system. Thus, if the system is not too highly nonlinear, and the eigensystem is well-conditioned, the iterations will converge.

It is instructive to examine the sensitivity of \( W_2 \) to changes in \( b \), assuming that the eigensystem is constant. From (12) we see that

\[
\| \delta W_2 \| \leq \left[ \| Q_{21} \| \| Q_{11}^{-1} \| \| Q_{12} \| + \| Q_{22} \| \right] \| B_{22}^{-1} \| \| U_2 \| \| \delta b \|
\]

\[
\leq \left[ \| Q_{21} \| \| Q_{11} \| + 1 \right] \kappa(Q) \| B_{22}^{-1} \| \| \delta b \|
\]

\[
= \left[ \| Q_{21} \| \| Q_{11} \| + 1 \right] \| B_{22}^{-1} \| \| \delta b \| \quad \therefore \kappa(Q) = 1
\]

where \( \kappa \) denotes the condition number. The most significant factor is \( \| B_{22}^{-1} \| \), which equals the maximum absolute value of the reciprocals of the stiff eigenvalues. This shows that
in this case the rate of convergence is proportional to the reciprocal of the stiff eigenvalues. The term

$$\|Q_{21}\| \|Q_{11}\|^{-1} \cdot \kappa(Q_{11})$$

shows that if $Q_{11}$ is ill-conditioned or if the entries of $Q_{11}$ are small relative to those of $Q_{21}$, convergence may be adversely affected. This term is a measure of the sensitivity of $W_2$ to changes in $W_1$; that is, it is a measure of the ill-conditioning of the actual problem.
Consider the following problem

\[
\begin{align*}
y_1' &= -50y_1 - 499y_2 + 2y_3 - 3y_4 + y_1y_3 - y_2y_4 + 1000 & y_1(0) &= 1 \\
y_2' &= -499y_1 - 50y_2 + 3y_3 - 2y_4 + y_1y_2 + y_3y_4 + 1000 & y_2(0) &= \,? \\
y_3' &= -3y_1 - 2y_2 - 50y_3 + 499y_4 + y_2y_3 + y_1y_3y_4 + 100 & y_3(0) &= \,? \\
y_4' &= 2y_1 + 3y_2 + 499y_3 - 50y_4 - y_1y_2 - y_2y_3y_4 + 100 & y_4(0) &= 1
\end{align*}
\]

which has two stiff eigenvalue. If we suppose that \(y_2(0)\) and \(y_3(0)\) are unknown, and we apply our algorithm with initial guess \(y_2(0) = 0\) and \(y_3(0) = 0\). We get the sequence

\[
\begin{array}{c|c}
y_2 & y_3 \\
0.00 & 0.00 \\
1.00075658024 & 0.99189014543 \\
1.00174984534 & 0.99188843012 \\
1.00174984999 & 0.99188853271 \\
1.00174984997 & 0.99188853272 \\
\end{array}
\]

which evidently converges. When the system is integrated with \(y_2(0) = 1.00174984997\) and \(y_3(0) = 0.99188853272\), there
is no initial transient. We can see that Figure 1 is using fixed \( y_1(0) = 1, y_3(0) = 0.99188853272, y_4(0) = 1 \) and varies \( y_2 \), and plot \( y_2 \) vs time \( t \). The curve with initial value \( 1.00174984997 = 1.0 \) has no initial transient, for the other graphs a small step size was needed through the transient phase, which can be observed in their graphs.

Let us consider the above problem without the constant term:

\[
\begin{align*}
y_1' &= -501y_1 - 499y_2 + 2y_3 - 3y_4 + y_1y_3 - y_2y_4 & y_1(0) &= 1 \\
y_2' &= -499y_1 - 501y_2 + 3y_3 - 2y_4 + y_1y_2y_3 + y_1y_4 & y_2(0) &= 0 \\
y_3' &= -3y_1 - 2y_2 - 501y_3 + 499y_4 + y_2y_3 + y_1y_3y_4 & y_3(0) &= 0 \\
y_4' &= 2y_1 + 3y_2 + 499y_3 - 501y_4 - y_1y_2 + y_2y_3y_4 & y_4(0) &= 1 
\end{align*}
\]

This problem still has two stiff eigenvalues. The same as above suppose \( y_2(0) \) and \( y_3(0) \) unknown, and apply the algorithm with initial guess \( y_2(0) = 0 \) and \( y_3(0) = 0 \). We get the sequence:

\[
\begin{array}{c|c}
Y_2 & Y_3 \\
0.00 & 0.00 \\
-0.99699410523 & 0.99998744352 \\
-0.9979919831 & 0.99999898096 \\
-0.9979919832 & 0.99999899764 \\
-0.9979919833 & 0.99999899764 \\
\end{array}
\]

Figure 1. $y_2$ vs time $t$ for the nonhomogeneous case.
and use fixed $y_1(0) = 1, y_3(0) = 0.99999899764, y_4(0) = 1$
and varies $y_2$, draw the graph $y_2$ vs time $t$ (see Figure 2).
The graph with initial value $-0.997991983 \div -1.0$ shows that there is no initial transient.

To solve the above two problems, we use the package of Dvoger from IMSL, this package can solve both stiff and nonstiff systems, and changes the step size automatically.
Figure 2. $y_2$ vs time $t$ for the homogeneous case.
REFERENCES


