Numerical methods for complex ODE / DAE Systems

Introduction

We will further elaborate on numerical methods used for simulation of engineering systems described by ordinary differential equations, either explicit (ODE) or implicit (DAE).

It is crucial to consider that numerical integration is an approximation of the real integration process. Besides errors due to the time discretization, errors are generated because numerical integration formulas are approximations of the real integration function.

To compute the variables in the model as a time function on a digital computer, time will be discretized. Otherwise, we have an infinite amount of values of time between $t_0$ and $t_e$, since $t$ is a real variable. Only at discrete points $t_k$ the model is computed. The result is that the model variables are available as a discrete series of values, and not as a continuous function.

The time distance between two adjacent points is called the time step, $h_k$, or integration interval. This $h_k$ need not necessarily be constant. integration interval

The model suitable for simulation consists of a set of Ordinary Differential Equations, and is called a simulation model. These differential equations describe the system as a set of state equations, where the initial values of the state variables (also called initial conditions) are known. In numerical mathematics, this is called an Initial Value Problem (IVP):

$$\dot{x} = f(x, u, t)$$
$$y = g(x, u, t)$$

where $x$ is the array of state variables, also called the state vector; $x(0)$, the initial values, which are known. $u$ is the array of input variables. $f$ are the state equations. $y$ is the array of output variables, computed from the states and inputs using the functions $g$. 
Numerical integration methods for ODE

A numerical integration method specifies how \( x_{k+1} \) is approximated in terms of \( x \) itself and its derivatives, both at previous, current or future moments of time. Practical numerical methods consist of a linear combination of these terms \( x_{k+1-j}^{(i)} \) using parameters \( \alpha_{ij} \), \( n \) and \( r \) (\( \dot{x} = x^{(i)} \)):

\[
\sum_{i=0}^{n} \sum_{j=0}^{r} \alpha_{ij} h_{k+1-j}^i x_{k+1-j}^{(i)} = 0
\]  

(3)

The method is explicit if all \( \alpha_{i0}=0 \) for \( i=1...n \), otherwise the method is implicit.

Equation (3) is an approximation, so errors are made:

- **Truncation errors**, due to the non-idealness of the approximation formula.
- **Round-off errors**, due to the limited machine precision.

Round-off errors

At smaller timesteps, the truncation error usually becomes smaller, whereas the round-off errors become larger.

For a useful approximation, the time step \( h_k \) depends on both the model and the integration method.

The form of the numerical integration method, either explicit or implicit, the number of steps and the order of derivatives involved, significantly influence numerical stability and effective applicability of the method.

Explicit numerical integration methods have less computational load than implicit methods, but unfortunately they also have a small stability region, and thus relatively small time steps \( h_k \) are necessary. The disadvantage of implicit numerical integration methods is that iterative calculations involving matrix inversion

Most numerical integration methods have been developed and evaluated for linear equations or for nonlinear equations only in the vicinity of the point of linearization
Basically there are 3 general classes of integration methods, categorized by the number of steps and order of derivatives:

a) Multistep, high-order derivative methods
the full version of equation (3).

b) Multistep, first-order derivatives methods
n=1, only \( x \) and its derivative are used. Higher order methods take more values of the past, until \( x_{k+1-n} \).

c) One-step higher-order derivatives methods
r=1, only values of \( x_k \) and \( x^{(i)}_k \)
the, the \( i \)th derivative of \( x_k \), are used. These methods are Runge-Kutta methods. No values of previous moments are used.

**Multistep, first-order derivative methods**

For multistep, first-order derivative methods, the parameter \( n \) is equal to one Equation (3) can be written as:

\[
x_{r+1} = a_{10} \dot{x}_{k+1} + \sum_{j=1}^{r} a_{0j} x_{k+1-j} + \sum_{j=1}^{r} a_{1j} h_{k+1-j} \dot{x}_{k+1-j}
\]

(4)

If \( a_{10} = 0 \), the method is explicit, otherwise the method is implicit, as \( \dot{x}_{k+1} \) is the only term at \( x_{k+1} \) at the right-hand side. The first summation represents \( x \) at previous moments of time; the second summation represents the derivative of \( x \) at previous moments of time. If \( r = 1 \), then it is a one-step method: only \( x^{(i)}_k \) and its derivative are used. At \( r > 1 \), it is a multistep method, which is not self-starting: at the beginning, old values of \( \dot{x}_{k-1} \) etc. are

<table>
<thead>
<tr>
<th>AB-1</th>
<th>Adams-Bashford = explicit Euler</th>
<th>( x_{k+1} = x_k + h_k \dot{x}_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB-2</td>
<td>Adams-Bashford = trapezium rule</td>
<td>( x_{k+1} = x_k + \frac{3}{2} h_k \dot{x}<em>k - \frac{1}{2} h_k \dot{x}</em>{k-1} )</td>
</tr>
<tr>
<td>AM-1</td>
<td>Adams-Moulton = implicit Euler</td>
<td>( x_{k+1} = x_k + h_{k+1} \dot{x}_{k+1} )</td>
</tr>
<tr>
<td>AM-2</td>
<td>Adams-Moulton = trapezium rule</td>
<td>( x_{k+1} = x_k + \frac{1}{2} h_{k+1} \dot{x}_{k+1} + \frac{1}{2} h_k \dot{x}_k )</td>
</tr>
</tbody>
</table>
One-step higher-order derivative methods

Writing \( x_{k+1} \) as a Taylor’s series, the coefficients of the numerical method, \( a_{ij} \), can be determined. This way is useful for theoretical research, but for a practical algorithm, the \( a_{ij} \)’s are approximated via extrapolation. 

Extrapolation is calculated at a number of points \( t_k + h_i h_k \) where \( 0 \leq h_i \leq 1 \). The values \( x(t_k + h_i h_k) \) and \( x(t_k) \) are used to match terms in the Taylor’s series, which result in a specification of a numerical integration method.

These methods are called \( r \)-stage Runge-Kutta methods. They need \( r \) model computations per time step. The general \( r \)-stage Runge-Kutta method is given by:

\[
x_0 = x_k + \sum_{j=1}^{r} b_{qj} z_j \quad q = 1,2,...,r
\]

\[
t_0 = t_k + h_k \sum_{j=1}^{r} b_{qj}
\]

\[
z_q = h_k f(x_{0}, t_0)
\]

\[
x_{k+1} = x_k + \sum_{q=1}^{r} \gamma_q z_q
\]

Equations for the parameters \( \gamma_q \) and \( \beta_{qj} \) are derived by matching terms with the Taylor’s series. If \( \beta_{qj} = 0 \) for \( j \geq q \), the method is explicit, otherwise implicit. As indicated above, there are more parameters to determine than equations available, so that a family of methods results. For explicit methods, the order is equal to the stage \( r \) for \( r \leq 4 \). Otherwise, the order is less than the stage. The implicit methods can have a maximal order of \( 2r \).
Variable time step Runge-Kutta’s do exist. Their time step is controlled by calculating each time step twice by either using two Runge-Kutta. For the first method, computations are minimized by choosing the coefficients such, that some intermediate results are common to both algorithms (Gear, 1971). Two algorithms which are base on the multiple use of intermediate results, are often used: the Runge-Kutta-Merson algorithm and the Runge-Kutta-Fehlberg algorithm.

**General remarks on integration formulas**

Some remarks on the order of derivatives, the number of steps, and the coefficients of the numerical integration methods:

1. Order of derivatives
The higher the order of derivatives used in the formula, the more accurate and more flexible the formula is. But at each time step, an accompanying heavy computation load exists, i.e. the computation of the high order derivatives of \( f(x,t) \) must be done.
2. Number of steps
The larger the number of steps, the more accurate and more flexible the formula is. But estimates of the unknown initial values of $x_k$ are needed.

3. Implicit formula versus explicit formula
Implicit formulas always require the computation of an inverse of a matrix or its equivalent. Explicit formulas are easy to apply to complicated nonlinear differential equations, and do not involve the calculation of a matrix inverse, but have a smaller stability region.

**Stiffness**

Stiffness is a subtle, difficult, and important concept in the numerical solution of ordinary differential equations. It depends on the differential equation, the initial conditions, and the numerical method.

“A problem is stiff if the solution being sought is varying slowly, but there are nearby solutions that vary rapidly, so the numerical method must take small steps to obtain satisfactory results”

This systems are given by a differential equation in the general form

$$M(t) \frac{dy}{dt} = f(t,y) \quad y(t_0) = y_0$$

with a mass matrix $M(t)$ that is non-singular and usually sparse

A model of flame propagation provides an example. If you light a match, the ball of flame grows rapidly until it reaches a critical size. Then it remains at that size because the amount of oxygen being consumed by the combustion in the interior of the ball balances the amount available through the surface.

The simple model is

$$\begin{align*}
\dot{y} &= y^2 - y^3 \\
y(0) &= \delta \\
0 \leq t \leq 2/\delta
\end{align*}$$

The scalar variable $y(t)$ represents the radius of the ball. The $y^2$ and $y^3$ terms come from the surface area and the volume. The critical parameter is the initial radius, $\delta$, which is “small.” We seek the solution over a length of time that is inversely proportional to $\delta$. 
delta = 0.0001;
ode45(F,[0 2/delta],delta,opts);

Figure 7.4. Stiff behavior of ode45

What can be done about stiff problems? You don't want to change the differential equation or the initial conditions, so you have to change the numerical method. Methods intended to solve stiff problems efficiently do more work per step, but can take much bigger steps. Stiff methods are *implicit*. At each step they use Matlab matrix operations to solve a system of simultaneous linear equations that helps predict the evolution of the solution.
Figure 7.5 shows the computed solution and the zoom detail. You can see that ode23s takes many fewer steps than ode45. This is actually an easy problem for a stiff solver. In fact, ode23s takes only 99 steps and uses just 412 function evaluations, while ode45 takes 3040 steps and uses 20179 function evaluations. Stiffness even affects graphical output. The print files for the ode45 figures are much larger than those for the ode23s figures.

**Systems of Equations**

Many mathematical models involve more than one unknown function, and second and higher order derivatives. These models can be handled by making \( y(t) \) a vector-valued function of \( t \). Each component is either one of the unknown functions or one of its derivatives. The MATLAB vector notation is particularly convenient here.

A slightly more complicated example, the *two-body problem*, describes the orbit of one body under the gravitational attraction of a much heavier body. Using Cartesian coordinates, \( u(t) \) and \( v(t) \), centered in the heavy body, the equations are
\[\ddot{u}(t) = -u(t)/r(t)^3\]
\[\ddot{v}(t) = -v(t)/r(t)^3\]

where
\[r(t) = \sqrt{a(t)^2 + v(t)^2}\]

The vector \(\mathbf{y}(t)\) has four components,
\[
\mathbf{y}(t) = \begin{bmatrix} u(t) \\ v(t) \\ \dot{u}(t) \\ \dot{v}(t) \end{bmatrix}
\]

The differential equation is
\[
\dot{\mathbf{y}}(t) = \begin{bmatrix} \dot{u}(t) \\ \dot{v}(t) \\ -u(t)/r(t)^3 \\ -v(t)/r(t)^3 \end{bmatrix}
\]

**Numerical integration methods for DAE**

Differential algebraic equations (DAEs) are systems of differential equations

\[F(x', x, t) = 0\]

with \(\partial F/\partial x'\) identically singular. The name arises since often (1) is a mix of differential and algebraic equations. DAEs arise naturally in many areas [1], [15]. The algebraic equations can arise because of physical constraints, desired behavior, or when discretizing spatial operators during the method of lines solution of a PDE. Depending on the area of application, DAEs are also called implicit, descriptor or singular.

2.1 **Analytical**

Consider the following simple example of (1):

\[(2a)\]
\[x_2' = x_1\]
\[(2b)\]
\[x_2 = t + \alpha(t)x_3\]
\[(2c)\]
\[x_3' = x_3 + 1\]

Here \(\alpha(t)\) is a nonzero coefficient. The solution of (2) is

\[x = \begin{bmatrix} 1 + \alpha'(t)(-1 + ce^t) + c\alpha(t)e^t \\ t + \alpha(t)(-1 + ce^t) \\ -1 + ce^t \end{bmatrix}, \quad c \text{ an arbitrary constant}\]

From this example we can make several observations about how DAEs differ from ordinary differential equations.
1. The solution \( x \) of (1) can depend on derivatives of the defining equations \( F \). Note the \( \alpha'(t) \) term that appears in the solution (3).

2. Only some initial conditions will admit smooth solutions. These are called consistent initial conditions.

3. There can be hidden constraints. The solutions of (2) satisfy not only the constraint (2b) but also the constraint \( \alpha'(t)x_3 + \alpha(t)(1 + x_3) - x_1 = 0 \).

4. The best that can be hoped for is that the solutions form a smooth manifold, called the solution manifold. It is parameterized by \( t, c \) in the example above.

Suppose the DAE (1) is a system of \( n \) equations in the \( n \) dimensional state variable \( x \) and that \( F \) is sufficiently differentiable in the variables \((x', x, t)\) so that all needed differentiations can be carried out. Suppose also that the solutions form a smooth manifold and that a (consistent) initial value on the manifold uniquely determines a solution. Such a DAE is sometimes called solvable. As noted, the solution \( x \) of (1) will depend, in general, on derivatives of \( F \). If the \( i \)th equation of (1) is differentiated \( r_i \) times with respect to \( t \) for \( i = 1, \ldots, n \), we get the \( m = n + \sum_{i=1}^{n} r_i \) derivative array equations

\[
G(x', w, x, t) = \begin{bmatrix}
F(x', x, t) \\
\frac{d}{dt}F(x', x, t) \\
\vdots \\
\frac{d^r}{dt^r}F(x', x, t)
\end{bmatrix} = \begin{bmatrix}
F(x', x, t) \\
F_{x'}x' + F_x \cdot + F_t \\
\vdots \\
F_{x'}x^{(r+1)} + \ldots
\end{bmatrix} = 0
\]

where \( w = [x^{(2)}, \ldots, x^{(r+1)}] \)

There are several versions of the index of a DAE and they are not equivalent for general nonlinear DAEs [6]. For our purposes, we shall define the index \( \nu \) to be the least value of \( r \) for which the derivative array equations (4) uniquely determine \( x' \) given a consistent \((x, t)\). This index is sometimes called the differentiation index. It should be noted that in a physical model the index is not determined just by the equations but also by which variables are considered known, such as inputs, and which are considered unknown.

In order to integrate a DAE, or for that matter an ODE, we need an estimate of \( x' \). The index measures, in some sense, how hard it is to get \( x' \). The index also measures the loss of smoothness in going from the coefficients and forcing functions to the solutions. Finally the index also affects the conditioning of the matrices which occur in some methods such as BDF (Backward differentiation formulas). Several numerical methods have been proposed for solving DAEs including BDF or implicit Runge-Kutta (IRK) methods [1].

For index one DAEs we have codes such as DASSL. There is no general code available for even index two problems.
To explain why it is difficult to solve high-index problems, let us briefly review the methods to solve ODEs. When solving an ODE on state space form the main task is to integrate. The derivatives are approximated by finite differences. A simple approximation is forward Euler, \( \dot{x}_n \approx \frac{x_{n+1} - x_n}{h} \) where \( h \) is the step size. Applied to \( \dot{z} = f(t, x) \), we obtain the explicit recursion \( x_{n+1} = x_n + hf(t_n, x_n) \). We can also use backward Euler, \( \dot{x}_{n+1} \approx \frac{x_{n+1} - x_n}{h} \) to obtain \( x_{n+1} = x_n + hf(t_{n+1}, x_{n+1}) \), which is an implicit recursion.

For example, backward Euler discretization transforms the differential algebraic problem, \( F(t, x, \dot{x}) = 0 \) into the recursion \( F(t_n, x_n, (x_n - x_{n-1})/h) = 0 \), which is readily solved when the index is 0 or 1. Backward differences of order up to five are used in the differential-algebraic solver DASSL developed by Petzold (see Brenan et al., 1989).

Backward differences for order two is given by the recursive equation \( F(t_n, x_n, (3x_n - 4x_{n-1} + x_{n-2})/h) = 0 \). In general BDF formulas are applied as follows

\[
F(t_n, x_n, \sum_{i=0}^k \alpha_i x_{n-i} / h) = 0
\]

The index can be reduced by differentiating the equations a number of times, see e.g., Brenan et al. (1989, p. 33). The new problem obtained in this way is called the underlying ODE, or UODE. However, it is often less satisfactory to solve the UODE, because its set of solutions is larger.

To avoid such difficulties, one may try to obtain a low-index formulation, with a solution set identical to that of the original problem. This can be achieved by augmenting the system as the index reduction proceeds: all original equations, and their successive derivatives are retained in the process. The result is an overdetermined but consistent index 1 DAE.
For each differentiated equation appended to the original system, we introduced one "new" dependent variable. The introduced variables represent derivatives and we call them *dummy derivatives*. Dummy derivatives are purely algebraic variables and are not subject to discretization.

To find the differentiated index 1 problem we can use Pantelides’s algorithm, cf. Pantelides (1988), intended for finding consistent initial values for a DAE. Pantelides’s algorithm establishes the minimum number $m_i$ each equation has to be differentiated to make the differentiated problem $G\dot{x} = F^T x = 0$ structurally non-singular with respect to its

The MATLAB ODE and DAE Solvers

This section is derived from the Algorithms portion of the MATLAB Reference Manual page for the ODE solvers.

`ode45` is based on an explicit Runge-Kutta (4,5) formula, the Dormand-Prince pair. It is a one-step solver. In computing $y(t_{n+1})$, it needs only the solution at the immediately preceding time point, $y(t_n)$. In general, `ode45` is the first function to try for most problems.

`ode23` is an implementation of an explicit Runge-Kutta (2,3) pair of Bogacki and Shampine. It is often more efficient than `ode45` at crude tolerances and in the presence of moderate stiffness. Like `ode45`, `ode23` is a one-step solver.

`ode113` is a variable-order Adams-Bashforth-Moulton PECE solver. It is often more efficient than `ode45` at stringent tolerances and if the ODE file function is particularly expensive to evaluate. `ode113` is a multistep solver — it normally needs the solutions at several preceding time points to compute the current solution.

The above algorithms are intended to solve nonstiff systems. If they appear to be unduly slow, try using one of the stiff solvers below.

`ode15s` is a variable-order solver based on the numerical differentiation formulas (NDFs). Optionally, it uses the backward differentiation formulas (BDFs, also known as Gear’s method) that are usually less efficient. Like `ode113`, `ode15s` is a multistep solver. Try `ode15s` if `ode45` fails or is very inefficient and you suspect that the problem is stiff, or if solving a differential-algebraic problem.

`ode23s` is based on a modified Rosenbrock formula of order 2. Because it is a one-step solver, it is often more efficient than `ode15s` at crude tolerances. It can solve some kinds of stiff problems for which `ode15s` is not effective.
The simplest system of differential algebraic equations (DAEs) has the following semiexplicit form:

\[ u' = f(t,u,v) \quad (I^a) \]
\[ 0 = g(t,u,v) \quad (1b) \]

In this notation, \( t \) is the independent variable (time); \( u \) stands for the differential variables, and \( v \) stands for the algebraic variables.

One idea for solving (1) could be to solve (1a) as an ODE. Evaluating the derivative \( u' \) for a given \( u \) would require solving the algebraic equation (1b) for the corresponding value of \( v \). This is easy enough in principle, but can lead to initial value problems for which the evaluation of \( f \) is rather expensive.

MATLAB solvers use a different approach. To solve DAEs in MATLAB, first you need to combine the differential and algebraic part. Note that any semiexplicit system (1) can be written as

\[ M * y' = F(t,y) \]

where \( M = [I \ 0 \ 0 \ 0] \) and \( y = [u;v] \), \( F=[f;g] \).
For differential algebraic equations, the mass matrix $M$ is singular, but such systems can still be solved with ODE15S and ODE23T. In fact, the solvers can handle more general systems, with time- and state-dependent singular mass matrices

$$M(t,y) \cdot y' = F(t,y)$$

The only restriction on DAEs solved in MATLAB is that they must be of index 1. Semi explicit DAEs are of index 1 when their matrix of partial derivatives $dg/dv$ is non-singular.

**The MODELICA ODE and DAE Solvers**

Hay dos formas de resolver las ecuaciones DAE en Modelica

1. **Por linealización de la DAE:**

   $$\dot{Y} = F(Y, Z), \quad 0 = G(Y, Z)$$

   Forma matricial:
   $$\begin{pmatrix} \dot{Y} \\ R \end{pmatrix} = \begin{pmatrix} F(Y, Z) \\ G(Y, Z) \end{pmatrix}$$

   Linealizando:
   $$\begin{pmatrix} \dot{Y} \\ R \end{pmatrix} = \begin{pmatrix} F_y & F_z \\ G_y & G_z \end{pmatrix} \begin{pmatrix} Y \\ Z \end{pmatrix}$$

   Nominalmente $R \equiv 0$.

   Entonces:
   $$\dot{Y} = (F_y - F_z G_z^{-1} G_y) Y$$

   • Estimamos los jacobianos $(F_y, F_z, G_y$ y $G_z)$.
   • Resolvemos la ecuación diferencial obtenida $(Y)$.
   • Resolvemos la ecuación algebraica $G(Y, Z) = 0$, y obtenemos $Z$.

   Es necesario que $G_z$ sea no singular.

2. **Por diferenciación de la DAE:**

   • Si el índice diferencial es 1, se realizará la integración por aplicación de métodos de integración implícitos, tales como Runge-Kutta Implicito (RADAU5) o por aplicación de formulas BFD de diferente orden (DASSL) sobre la nueva ecuación

   $$M(t, x_d) \cdot x_d' = G(x_d, t)$$

   siendo $x_d$ el nuevo vector de derivadas.
• Si el índice diferencial es superior a 1, no es posible obtener un sistema de ecuaciones en $x_d$ de primer orden, por lo que no es posible aplicar los métodos de integración implícitos.

Por ello se realizará la reducción de índice diferencial hasta obtener una sistema de índice diferencial 1 añadiendo variables mudas (dummy), resultando un sistema de mayor dimension, siendo resuelto como el caso anterior.

Para la generación del sistema de ecuaciones en $x_d$ así como la obtención de variables mudas se utilizará el algoritmo de Pantelides.

Modelica permite la integración tanto de ODE’s como de DAE’s a través de varios métodos de integración (ver tabla).

<table>
<thead>
<tr>
<th>Method</th>
<th>Model Type</th>
<th>Order</th>
<th>Stiff</th>
<th>Root Find</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEABM</td>
<td>ODE</td>
<td>1-12</td>
<td>No</td>
<td>No</td>
<td>Adams/Bashforth/Moulton; reliable, maybe slow</td>
</tr>
<tr>
<td>LSODE1</td>
<td>ODE</td>
<td>1-12</td>
<td>No</td>
<td>No</td>
<td>Adams/Bashforth/Moulton, faster than DEABM</td>
</tr>
<tr>
<td>LSODE2</td>
<td>ODE</td>
<td>1-5</td>
<td>Yes</td>
<td>No</td>
<td>Backward Difference Formulae; Gear method</td>
</tr>
<tr>
<td>LSODAR</td>
<td>ODE</td>
<td>1-12, 1-5</td>
<td>Both</td>
<td>Yes</td>
<td>Adams/Bashforth/Moulton</td>
</tr>
<tr>
<td>DOPRI5</td>
<td>ODE</td>
<td>5</td>
<td>No</td>
<td>No</td>
<td>Runge-Kutta by Dormand and Prince</td>
</tr>
<tr>
<td>DOPRI8</td>
<td>ODE</td>
<td>8</td>
<td>No</td>
<td>No</td>
<td>Runge-Kutta by Dormand and Prince</td>
</tr>
<tr>
<td>GRK4T</td>
<td>ODE</td>
<td>4</td>
<td>Yes</td>
<td>No</td>
<td>Linearly-implicit Rosenbrock</td>
</tr>
<tr>
<td>DASSL</td>
<td>DAE</td>
<td>1-5</td>
<td>Yes</td>
<td>Yes</td>
<td>BDF; default choice in Dymola</td>
</tr>
<tr>
<td>ODASSL</td>
<td>ODAE</td>
<td>1-5</td>
<td>Yes</td>
<td>Yes</td>
<td>Modified DASSL for overdetermined DAEs</td>
</tr>
<tr>
<td>MEXX</td>
<td>ODAE</td>
<td>2-24</td>
<td>No</td>
<td>No</td>
<td>Special index-2 DAE, not for ODE</td>
</tr>
</tbody>
</table>