Abstract

In a previous paper [1] we have shown that information about possible tearing variables added to a Bond Graph (BG) with dependent storage elements can be exploited by a program like Dymola [6], if the stores are replaced by a so-called resistive companion model originally introduced in integrated circuit analysis [13].

In this paper it is shown how tearing information added to a BG with causal paths between stores of the same type can be exploited directly by a program like Dymola allowing for a mixed symbolic, numerical solution of the underlying Differential Algebraic Equation (DAE) system. For didactic reasons, the method is explained by means of some fairly small examples. Nevertheless, it is quite general.

Keywords: Bond Graphs, dependent storage elements, DAE systems, residual sinks, tearing.

1 Introduction

For many engineering systems, e.g. for multibody systems (MBS) reasonable modeling assumptions lead to algebraic dependencies between some of the state variables such that only a smaller number of them is independent. Consequently, the mathematical model has the form of a Differential Algebraic Equations (DAE) system. Its numerical solution requires an appropriate package like the DASSL code [3]. Before DAE solvers have become popular, a commonly used method was to turn the DAE system into an explicit Ordinary Differential Equations (ODE) system by adding small resistors in appropriate locations such that causal paths between storage ports of the same type vanish. The disadvantage of that approach is that equations may become stiff. A different option we advocate in this paper is to add information to the Bond Graph (BG) such that a program like Dymola [6] can apply its formulae manipulation capabilities to the model equations to allow for a subsequent efficient numerical solution. The key issue is tearing that is, the partitioning of an overall large system of equations into a number of smaller systems. The method was introduced by G. Kron [11] already in the early 60's and is not limited to linear equations. In order to be able split a system into smaller subsystems, so-called tearing variables must be determined somehow. Once these tearing variables are known, all other non-state variables can be computed by means of the tearing variables and the independent state variables. Of course, for general non-linear systems a (modified) Newton iteration will be needed, while for systems that are linear, at least in regard to the tearing variables, the original DAEs can be transformed into an explicit ODE system.

The objective we are aiming at is to give information to a program at the BG level such that tearing based formulae manipulation can take place automatically. For a system of hundreds of equations and variables derived automatically from a BG it is almost impossible to inspect the generated equations and to decide which variables can be used to tear the overall set of equations into a number of smaller sets. For fairly small systems the modeler might represent variables and their functional relations by a so-called dependency graph in order to see which variables are involved in which algebraic loops. However, since it is an NP complete problem to find a minimal set of tearing variables that cut all loops [12], the modeler might be seeking not for a minimal but small number of tearing variables that reduce the computational amount, at least to some extent and a method that is generally applicable. That is the issue we are addressing in the sequel. First, for a better understanding, the principle of tearing is briefly recalled. More information on tearing may be found in [11], [8].

2 Tearing

If an index-one DAE system is to be reduced symbolically to an explicit ODE system, then after analytical differentiation of the constraint equations a system must be solved that may be of considerable size. For simplicity reasons we assume a set of linear equations, although tearing is not limited to the linear case. Suppose that it has been decided which components of the vector $\mathbf{z}$ of unknowns including the dependent state variables are the tearing variables. If they are grouped into a subvector $\mathbf{z}_2$, then
permutation matrices $P$ and $Q$ can be determined such that the system can be written in the following way:

$$
\begin{pmatrix}
L & A_{12} \\
A_{21} & A_{22}
\end{pmatrix}
\begin{pmatrix}
z_1 \\
z_2
\end{pmatrix}
= 
\begin{pmatrix}
b_1 \\
b_2
\end{pmatrix}
$$

(1)

If $L$ is non-singular, the symbolic solution of the equation

$$
Lz_1 + A_{12}z_2 = b_1
$$

(2)

with respect to $z_1$ is not expensive, since the submatrix $L$ is least lower block triangular. If its solution is substituted into

$$
A_{21}z_1 + A_{22}z_2 = b_2,
$$

(3)

then we obtain an equation that determines the subvector $z_2$ of tearing variables. Once these are known, the other subvector $z_1$ can be computed at low cost. Obviously, if we are lucky to find a small number of tearing variables, then only a small subsystem must be solved, and all other unknowns $z_1$ are obtained essentially by evaluation of the expressions in eq. (2). The question is, how such a set of tearing variables can be found. In the following we will show that information about tearing can be expressed already in the BG.

3 The underlying idea

Let us start with an example by considering a Bond Graph with a tree structure with five capacitances as shown in Fig. 1. Application of the standard Sequential Causality Assignment Procedure (SCAP) [10] reveals that two of the five capacitors must have derivative causality. Consequently, their co-energy variables are not state variables. This is indicated in the program Dymola by the command

- **variable unknown** C3.e C5.e

(C3.e denotes the effort of capacitance C3.) Since there are more unknown variables than equations, the right hand side of the ODE’s for the three independent state variables cannot be computed. For that reason, by means of the command

- **differentiate**

we invoke an implementation of Pantelides’ algorithm [14], which provides the missing equations. The equations derived from the BG can be written in the DAE form

$$
\begin{pmatrix}
C_{11} & C_{12} \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
e_1 \\
e_d
\end{pmatrix}
+ 
\begin{pmatrix}
G_{11} & 0 \\
G_{21} & G_{22}
\end{pmatrix}
\begin{pmatrix}
e_i \\
e_d
\end{pmatrix}
= 
\begin{pmatrix}
F \\
0
\end{pmatrix}
$$

(4)

with $e_d = (e_1, e_2, e_4)^T$, $e_d = (e_3, e_5)^T$ and $F = (E/R, 0, 0)^T$. As the matrix

$$
C = C_{11} - C_{12}G_{22}^T G_{21}
$$

(5)

is non-singular, the index of that DAE system is one. Hence, the sum of all efforts at the second and the third one junction

$$
e_1 = e_2 + e_3$$

(6a)

$$
e_3 = e_4 + e_5$$

(6b)

is differentiated only once.

Now, after causality assignment Dymola comes up with a system of nine simultaneous equations and solves them symbolically without exploiting the system structure in order to produce the state equations for the independent capacitances $C_1, C_2, C_4$. The corresponding Dymola output file including the incidence matrix of that system of simultaneous equations is given partly in Fig. 2. In that figure a variable in square brackets means that the variable is considered the unknown in that equation.

By looking at the BG it can be seen immediately that if the flows $f_3$ and $f_5$ would be known, all other algebraic variables could be computed from these flow variables and the state variables. Of course, this is expressed by the incidence matrix as well (cf. Fig. 2), but it is less ap-
In regard to the BG the condition becomes even more evident, if we add residual flow sinks \( rS_f \) to the BG (cf. Fig. 3). Residual sinks have been introduced in [1]. They simply represent a Lagrange multiplier \( \lambda \). There are two types of residual sinks. Residual flow sinks provide a flow such that their corresponding effort vanishes, while residual effort sinks are the dual elements. Apparently, by adding such sinks to the BG we do not modify the model. Such elements have been used as well by Gawthrop and Smith [9] in order to make algebraic variables associated with algebraic loops explicit on the BG. They have termed them sensing sources and label them SS.

In Dymola the role of a residual flow sink is expressed by means of a residue() operator:

\[
e = \text{residue}(f),
\]

which means that \( f \) is a possible tearing variable, while \( \text{residue}(f) \) is made to be zero. By following an idea of H. Elmqvist [7] we modify eq. (7) for a reason that will become obvious shortly:

\[
e = \text{Time} \ast \text{residue}(f)
\]

From the BG in Fig. 3 we see that the sum of efforts in equations (6.a), (6.b) become:

\[
\begin{align*}
e_1 &= e_3 + e_2 + \text{Time} \ast \text{residue}(f_3) \\
e_3 &= e_5 + e_4 + \text{Time} \ast \text{residue}(f_5)
\end{align*}
\]

Apparently, since \( \text{residue}(\ldots) \) vanishes, two capacitances are still dependent. When Pantelides’ algorithm differentiates these equations, we obtain

\[
\begin{align*}
der(e_1) &= \ der(e_3) + \ der(e_2) + \text{residue}(f_3) \\
der(e_3) &= \ der(e_5) + \ der(e_4) + \text{residue}(f_5)
\end{align*}
\]

The last term in the differentiated equations is the reason why we slightly changed the constitutive equation of the residual flow sink. Apparently, the algebraic constraints have just been differentiated. However, for Dymola the meaning is that \( f_3 \) and \( f_5 \) are possible tearing variables and equations (10.a), (10.b) are the associated residual equations that determine the tearing variables. That is, after solving the residual equations for \( f_3 \) and \( f_5 \), the right
hand side of the ODE system for \( e_1, e_2, e_4 \) can be computed. As can be seen from the corresponding Dymola output (cf. Fig. 4), adding the residual sinks to the BG, enables Dymola to identify tearing variables and their associated residual equations. With that information it solves symbolically for the tearing variables, and after that step determines all remaining non-state variables by means of the tearing variables and the state variables \((C_1.e, C_2.e, C_4.e)\).

By writing the simultaneous equations in Fig. 4 as a matrix equation, it can be seen that the system indeed, has the form of equation (1)

4 The general case

In the previous section we considered a small and fairly simple example. The method however, can be well applied to more general cases. In the above example a dependent capacitance has been replaced by a C-element and a residual flow sink, both attached to a one junction. Correspondingly, if dependent inerterces occur as they do for instance, in BGs of rigid multibody systems, they are replaced by an inerter and a residual effort sink, both connected to a zero junction (cf. Fig. 5).

First, for simplicity let us exclude the case of kinematic loops in mechanical systems. Now, if \( x_i \) denotes the vector of state variables and \( x_d \) the vector of output variables of the dependent stores, then assuming that the state variables depend linearly on the outputs of the residual sinks, equations can be derived straightforward from a BG in the following form:

\[
\begin{align*}
D_i x_i &= G \cdot \lambda + g(x_i, x_d, \textbf{u}) \quad (11.a) \\
D_d x_d &= \lambda \\
x_d &= B x_i + Time \cdot \text{residue}(\lambda) \quad (11.c)
\end{align*}
\]

In many cases the storage elements are not directly coupled such that the matrices \( D_i \) and \( D_d \) are diagonal. Equation (11.a) represents just the state equations of all independent storage elements, in which the Lagrange multiplier \( \lambda \) is treated like a known input. True external system inputs are denoted by the vector \( \textbf{u} \). Equation (11.b) is due to the above mentioned replacement of dependent storage elements (cf. Fig. 5) and eq. (11.c) is just a summation of efforts at one junctions, to which residual flow sinks are attached, or a summation of flows at zero junctions respectively, to which a residual effort sink is connected. Since the last term vanishes, the equation expresses the algebraic dependencies between storage elements. Differentiation of that constraint equation gives the residual equation for the tearing variable \( \lambda \)

\[
\dot{x}_d = \dot{B} x_i + B \dot{x}_i + \text{residue}(\lambda) \quad (12)
\]

Substitution of \( \dot{x}_i \) and \( \dot{x}_d \) by means of equations (11.a), (11.b) yields the linear system determining the Lagrange
Then equations \(1\) and \(6\) are the residual equations with multipliers  
\[
[I - D_d BD_i^{-1} G] \lambda = D_d \dot{B} x_i + D_d BD_i^{-1} g \tag{13}
\]
Once \(\lambda\) is known, the derivatives of the state variables \(x_i\) can be easily computed.

If there are kinematic loops in a mechanical system, they introduce an algebraic dependency between the components of the vector \(x_i\). If we add residual sinks representing Lagrange multipliers \(\lambda_{kl}\) to the kinematic loops, and if we partition the vector \(x\) into a subvector \(x_d\) of state variables and a dependent subvector \(x_d\), then a weighted sum of both subvectors vanishes at those junctions, to which the additional residual sinks have been attached:

\[
C_1 x_{id} + C_2 x_{id} = Time \ast residue(\lambda_{kl}) , \tag{14}
\]
which can be written in the form  
\[
C x_i = residue(\lambda_{kl}) \tag{15}
\]
with \(C = (C_1 \mid C_2)\). In order to distinguish between those residual sinks added to the dependent storage elements and those inserted into the kinematic loops, we replace \(\lambda\) in eq. (12) by \(\lambda_{d}\) and partition \(\lambda\) in eq. (11) into subvectors \(\lambda_{d}\) and \(\lambda_{kl}\). Correspondingly, the matrix \(G\) can be partitioned into two submatrices: \(G = (G_1 \mid G_2)\). Now, if the above eq. (15) is differentiated

\[
\dot{C} x_i + C \dot{x}_i = residue(\lambda_{kl}) , \tag{16}
\]
then equations (12) and (16) are the residual equations for the tearing variables \(\lambda_{d}\) and \(\lambda_{kl}\). Substituting \(\dot{x}_i\) and \(x_d\) gives lengthy expressions for the unknown tearing variables \(\lambda_{d}\) and \(\lambda_{kl}\) similar to that in eq. (13). As a result, systems with kinematic loops can be treated by the tearing approach in the same manner as those without kinematic loops. In case of kinematic loops only additional residual sinks are needed.

5 Applications of the method

The presented tearing approach has been applied to a number of examples ranging from BGs of rather small systems to those of medium size, including mechanical systems with kinematic loops like the well known slider crank mechanism. Due to the lack of space not all BGs of those examples along with explanations can be given here. Nevertheless, statistics of the program Dymola for a number of examples are listed in table 1. The input to the program has been a description of the BG in the language Dymola. By means of a library containing model classes for the BG elements Dymola generates the model equations. These are differentiated by Pantelides’ algorithm. After indicating to Dymola which outputs of storage elements are no state variables, and after assigning variables to equations, Dymola can produce an input file for a number of simulation programs. In our experiments we produced ACSL code. If Dymola’s tearing capability is activated by the command

- set tear on

the equations are automatically torn. Of course, for convenience all necessary commands and settings of switches can be collected in a command input file.

Table 1 clearly shows two results. If a symbolic reduction of an index-one DAE system is wanted, then after differentiation, tearing of the system of simultaneous equations can significantly reduce the computational amount in terms of arithmetic operations. Even in the case of a small system with two capacitances connected in parallel the number of arithmetic operations is reduced by 40%.

The second result is that for larger linear systems the number of algebraic operations increases considerably. As has been experienced by Bos [2] and others as well, the symbolic solution of larger linear systems indeed, requires much memory. Consequently, for some systems a symbolic reduction without tearing turns out to be impossible. In that case a possible remedy is to calculate the coefficients of the right hand side of the ODE system numerically. One disadvantage of course, is that the impact of system parameters cannot be directly analyzed any longer.

For modeling the 2D mechanical systems the basic building block of a free rigid body given by Bos [2] has been used accounting for simplifications due to planar motion. The ring structure listed in table 1 is a simple BG that involves a causal path between two stores of the same type, as well as an algebraic loop between resistors \(R_1\) and \(R_2\) (cf. Fig. 6). In each causal path a residual sink has been added as it is shown in Fig. 7. Of course, different types of dependent storage elements can be treated as well.

6 Conclusions

In this paper we have shown that by adding residual sinks to a BG with dependent stores, information specifying tearing variables and corresponding residual equations can be given to a program like Dymola. As a result, the formulae manipulation capabilities of that program enable it to use that information directly to reduce without any further help an initial DAE system (of index-one) sym-
Table 1: Numbers of arithmetic operations for some examples

<table>
<thead>
<tr>
<th>System</th>
<th>No. of storage ports</th>
<th>No. of simultaneous equations</th>
<th>No. of arithmetic operations without tearing</th>
<th>No. of arithmetic operations with tearing</th>
<th>Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>two parallel Cs</td>
<td>2</td>
<td>4</td>
<td>20</td>
<td>12</td>
<td>40 %</td>
</tr>
<tr>
<td>ring structure (cf. Fig. 7)</td>
<td>2</td>
<td>$6 + 6$</td>
<td>99</td>
<td>29</td>
<td>70.7 %</td>
</tr>
<tr>
<td>C network (cf. Fig. 3)</td>
<td>5</td>
<td>9</td>
<td>166</td>
<td>32</td>
<td>80.7 %</td>
</tr>
<tr>
<td>mechanical lever</td>
<td>4</td>
<td>12</td>
<td>233</td>
<td>58</td>
<td>75.1 %</td>
</tr>
<tr>
<td>2D pendulum</td>
<td>3</td>
<td>21</td>
<td>1636</td>
<td>109</td>
<td>93.3 %</td>
</tr>
<tr>
<td>2D double pendulum</td>
<td>6</td>
<td>53</td>
<td>—</td>
<td>604</td>
<td></td>
</tr>
<tr>
<td>slider crank</td>
<td>8</td>
<td>$9 + 61$</td>
<td>—</td>
<td>2253</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6: BG with dependent capacitances and an algebraic loop

Figure 6: BG with dependent capacitances and an algebraic loop

bolically to an explicit ODE system. That is, the modeller can support symbolic processing of the equations generated from a BG description without having to inspect them, which obviously, becomes impossible for larger sets of equations. The method presented is quite general and has been applied to a number of BGs of electrical and mechanical systems, including those with kinematic loops. Considering the general case, we have only assumed that the equations are linear in the tearing variables.

With our approach smaller non torn subsystems may occur in the reduction phase for instance, if the equations of a coordinate transformation need to be inverted, or in case kinematic loops are present. Since the symbolic solution of larger linear systems needs much memory, it is important to exploit tearing even more comprehensively. One possible way is to exploit information given by the modeler at the BG level and in addition to tear emerging smaller subsystems formally at an embedded level. Whatever additional approaches are used, the major goal should be that once equations have been generated, the modeler should not be concerned with their inspection.

Acknowledgment

This paper is based on research work carried out during the first author’s sabbatical leave at the University of Arizona in Tucson, in 1995, who gratefully acknowledges
a grant from the German Research Council “Deutsche Forschungsgemeinschaft” (DFG) in Bonn, Germany.

References


