The Computation of Consistent Initial Values for Nonlinear Index–2 Differential–Algebraic Equations

Diana Estévez Schwarz and René Lamour
Humboldt-University Berlin
Unter den Linden 6, D-10099 Berlin, Germany

Abstract

The computation of consistent initial values for differential–algebraic equations (DAEs) is essential for staring a numerical integration. Based on the tractability index concept a method is proposed to filter those equations of a system of index–2 DAEs, whose differentiation leads to an index reduction. The considered equation class covers Hessenberg–systems and the equations arising from the simulation of electrical networks by means of Modified Nodal Analysis (MNA). The index reduction provides a method for the computation of the consistent initial values. The realized algorithm is described and illustrated by examples.

Key words: Differential–Algebraic Equation, DAE, Index, Consistent Initial Values, Consistent Initialization, Circuit Simulation, Modified Nodal Analysis, MNA, Algorithm.

AMS Subject Classification: 65L05, 34A12.

1 Introduction

Differential–algebraic equations (DAEs) are systems of the form

$$f(x', x, t) = 0 \tag{1.1}$$

with a singular matrix $f'_{x'}$. The singularity of $f'_{x'}$ implies that (1.1) contains some derivative-free equations called constraints. Such systems arise in numerous applications as for instance multibody systems, electric circuit simulation and chemical kinetics.

To start up the numerical integration of DAEs consistent initial values are required. In the index-1 case, this means that we need to start from a point that lies in the manifold M_0 defined by the given constraints. For the higher-index cases, the so-called hidden constraints that result by differentiation define a sub-manifold of M_0 on which all solutions must lie. Thus, in these cases the consistent initial value has to lie in that manifold. To this end, a proper description of the hidden constraints becomes necessary.

In the literature different approaches have been presented. Among others, Pantelides [1] constructed an algorithm using graph theory methods to differentiate subsets of the system. Leimkuhler [2] used the global index definition combined with a finite difference approximation of the derivatives. Hansen [3] proposed a method based on the tractability index with time dependent projectors only, which applies formula manipulation methods and index reduction. Lamour [4] used the properties of the projectors related to the tractability-index to describe the part of the solution which we have to differentiate, while the differentiated part was replaced by its finite differences. Amodio and Mazzia [5] considered Hessenberg systems and realized differentiation by special finite differences.

In this article we consider index-2 DAEs fulfilling some structural properties, which are more general than the ones of the mentioned papers. We will describe the hidden constraints by making use of the projectors related to the tractability–index. Therefore, in Section 2 we briefly introduce this index definition. To prove that the expression we will define corresponds to the hidden constraints, we show in Section 3 that substituting some of them for a part of the original equations gives place to an index reduction.
This index reduction method permits us to establish a relation between the hidden constraints of two modeling techniques in circuit simulation, the conventional Modified Nodal Analysis and the charge-oriented Modified Nodal Analysis. This relation is outlined in Section 4. In Section 5 we present a possible Ansatz to fix values for a subset of variables whose cardinality is the so–called degree of freedom in order to set up a nonlinear system the solution of which provides a consistent initial value.
Finally, in Section 6 we describe the numerical realization of the presented results and some examples are given in Section 7. The programs are available at http://www-iam.mathematik.hu-berlin.de/∼lamour.

2 Spaces, Projectors, and Manifolds

Let us consider DAEs with an index at most 2 and a quasi-linear structure

$$f(x', x, t) := A(x, t)x' + b(x, t) = 0. \tag{2.1}$$

In the following we assume that all the appearing derivatives exist and that the partial derivatives with respect to x' and x are continuous.
If the coefficient matrix $A(x, t)$ is nonsingular, (2.1) represents an implicitly regular ODE. But we are interested in the case when $A(x, t)$ remains singular and assume that

$$\mathbf{A1}: \quad N := \ker A(x, t) = const, \quad \operatorname{im} A(x, t) = const.$$

For a proper analysis of these systems we define the projectors Q onto N, $P := I - Q$, and W_0 along im $A(x,t)$.

We apply the tractability index introduced by [6],[7], which is defined by considering a matrix chain based on the pencil matrices, i.e., on $f'_{x'}$, f'_x. Because of (2.1) $f'_{x'} = A(x,t)$ holds and for $B = f'_x$ we have

$$B(x', x, t) = [A(x,t)x']'_x + b'_x(x,t).$$

Notice now that all solutions of (2.1) lie in

$$M_0(t) := \{z \in I\!R^n : W_0 b(z,t) = 0\}. \tag{2.2}$$

The space S, which is closely related to the tangent space of $M_0(t)$, is given by

$$S(x,t) := \{z \in I\!R^n : W_0 B(x', x, t)z = 0\} = \{z \in I\!R^n : W_0 b'_x(x,t)z = 0\}.$$

Definition 2.1 *[6] If $A(x,t)$ is singular, then (2.1) has index 1*

$$\begin{aligned} &\Longleftrightarrow N \cap S(x,t) = \{0\} \\ &\Longleftrightarrow G_1(x', x, t) := A(x,t) + B(x', x, t)Q \ \textit{ is nonsingular.} \end{aligned}$$

In the index-1 case, there exists a solution through x_0 for each point $x_0 \in M_0(t)$. In this article we focus on the index–2 case. Therefore we consider the next matrix chain elements, which are given by $G_1(x', x, t)$ and

$$B_1(x', x, t) := B(x', x, t)P.$$

Assume that

$$\textbf{A2}: \quad \text{im}\, G_1(x', x, t) \text{ and } \ker G_1(x', x, t) \text{ do not depend on } x' \tag{2.3}$$

and let $W_1(x,t)$ be a projector along $\text{im}\, G_1(x', x, t)$. The relevant spaces on this level are

$$S_1(x', x, t) := \{z \in I\!R^n : W_1(x,t)B_1(x', x, t)z = 0\}$$

and

$$N_1(x,t) := \ker G_1(x', x, t)$$

and we denote by $Q_1(x,t)$ a projector onto $N_1(x,t)$ and $P_1(x,t) := I - Q_1(x,t)$.

Definition 2.2 *[6], [7] If (2.1) has not index 1 and $\dim N \cap S(x,t)$ is constant, then (2.1) has index 2*

$$\begin{aligned} \Longleftrightarrow &\quad N_1(x,t) \cap S_1(x', x, t) = \{0\} \\ \Longleftrightarrow &\quad G_2(x', x, t) := G_1(x', x, t) + B_1(x', x, t)Q_1(x,t) \ \textit{ is nonsingular.} \end{aligned}$$

It seems to be important to note that the index definition introduced above does not depend on the special choice of the different projectors.

For simplicity, in the following we will drop the arguments of A, B, G_1, S_1, N_1, Q_1, P_1, G_2 if they are clear from the context.

In the index–2 case we choose the so–called canonical projector onto N_1 **along** S_1, which fulfils $Q_1 = Q_1 G_2^{-1} B_1$, [7]. Furthermore, it can be shown (cf.[8]) that

$$N \cap S(x,t) = \text{im } QQ_1(x,t) \neq \{0\}.$$

In this article we further suppose that there exists a constant space L such that

$$\text{im } G_1(x', x, t) \oplus L = I\!R^n.$$

Thus it is possible to choose a projector $W_1(x,t)$ with im $W_1(x,t)$ is constant. Indeed this assumption is given for Hessenberg systems, because W_1 is constant itself (see Remark 3.3), and for the equations arising from Modified Nodal Analysis (cf. [9]). Note that, locally, this can always be assumed. Since $\text{im} A \subset \text{im} G_1$ and thus $L \cap \text{im } A = \{0\}$, we can define a constant projector $\hat{W}_1$ fulfilling:

$$\text{im } \hat{W}_1 = \text{im } W_1(x,t) \quad \text{and} \quad \ker \hat{W}_1 \supset \text{im } A, \tag{2.4}$$

which will become important later on.

In contrast to the index–1 case, where $M_0(t)$ is filled by solutions, for the index-2 case the so–called hidden constraints define the manifold

$$M_1(t) \subset M_0(t),$$

which fulfils the requirement that for each point $x_0 \in M_1(t)$ there exists a solution through x_0. These hidden constraints arise when differentiating a suitable part of (2.1). We will see that this part can be described properly with the aid of the projector W_1.
For later considerations we need the following properties.

Lemma 2.3 *: Let (A, B) be a given matrix pencil, Q a projector onto* $\ker A$*, W_0 a projector along* $\text{im} A$ *and W_1 a projector along* $\text{im} G_1$ *with $G_1 := A + BQ$. The following conditions are valid*
a.) $W_1 BQ = 0$,
b.) $W_1 = W_1 W_0$.

Proof:
a.) With $0 = W_1 G_1 = W_1(A + BQ)$ we obtain

$$W_1 G_1 P = W_1 A = 0 \quad \text{and} \quad W_1 G_1 Q = W_1 BQ = 0.$$

b.) Denote by A^- the reflexive generalized inverse of A with $W_0 = I - AA^-$ and $Q = I - A^- A$ (A^- is uniquely determined by these assumptions). From $W_1 A = 0$ it follows that $0 = W_1 AA^- = W_1(I - W_0)$ or $W_1 = W_1 W_0$.

q.e.d.

Since all the above matrices depend continuously on (x, t), it holds that if Lemma 2.3 is valid at fixed $(x_\star, t_\star)$, then it remains valid in a sufficiently small neighborhood of $(x_\star, t_\star)$.

3 Index Reduction by Differentiation

3.1 Motivation

It is well known that the differentiation of a DAE or of parts of it sometimes reduces its index. For a better understanding of this principle we give some academic examples. Let us consider the linear index–2 DAE

$$f(x', x, t) = Ax' + Bx - q := \begin{pmatrix} 1 & 0 & 0 & 0 \\ & 0 & 0 & 0 \\ & & 0 & 0 \\ & & & 0 \end{pmatrix} x' + \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} x - q = 0$$

or, as single equations,

$$\begin{aligned} x_1' + x_4 &= q_1, \\ x_1 + x_2 &= q_2, \\ x_2 &= q_3, \\ x_3 &= q_4. \end{aligned}$$

Obviously, we do not require the differentiation e.g. of the fourth equation to obtain an explicit expression for the solution x_1, x_2, x_3, x_4. But the general application of the differentiation index (see e.g. [10],[11]) requires the computation of $\frac{d}{dt} f(x', x, t)$. Using the given semi-explicit structure we would only differentiate all the algebraic equations. With the projector $W_0 = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$ along im A we could write this in the form $\frac{d}{dt}(W_0 f(x', x, t))$[1]. However, if for $Q = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$ we use a projector W_1 along im G_1 with $G_1 = A + BQ = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, which is given by $W_1 = \begin{pmatrix} 0 & & & \\ & 1 & -1 & \\ & & 0 & \\ & & & 0 \end{pmatrix}$, we actually differentiate only the necessary constrains by considering $\frac{d}{dt}(W_1 f(x', x, t))$.
Our aim is to generalize the above Ansatz for some nonlinear DAEs. We want to show that the approach can be adapted to obtain an index–reduction for

[1] Another approach to select suitable equations can be found in [12]

more general equations. At a first glance, the above example may suggest that an index reduction is always obtained by considering the system

$$(I - W_1)f(x', x, t) + W_1 \frac{d}{dt}(W_1 f(x', x, t)) = 0. \tag{3.1}$$

If the projector W_1 is constant or depends only on t, then (3.1) certainly has index one (cf. [13], [14]). In [14] it was shown how to handle with the case that the projector depends on the part of the solution that appears together with its derivatives, i.e., $W_1(Px, t)$ is allowed.

In practice, we have noticed that W_1 may also depend on the other parts of the solution. For instance (cf. [9]), the charge-oriented Modified Nodal Analysis presents this property. For these systems, we have observed that the way to obtain a reasonable index reduction consists in considering the system[2]

$$(I - \hat{W}_1)f(x', x, t) + W_1(x, t)\frac{d}{dt}(f(x', x, t)) = 0. \tag{3.2}$$

Observe that the term $(I - \hat{W}_1)f(x', x, t)$ describes the equations that are not replaced by derived ones. The choice of such a constant projector $\hat{W}_1$ becomes important in the nonlinear case.

The following example illustrates why the index reduction described in (3.1) is not appropriate for nonlinear DAEs in general. For simplicity, we consider the index-2 DAE:

$$\begin{aligned} x_1' + x_4 &= q_1, \\ x_1 + x_2 x_3 &= q_2, \\ x_2 &= q_3, \\ x_3 &= q_4, \end{aligned}$$

$x_i(t) \in \mathbb{R}$. For the projector Q chosen as before, the projectors W_1 and $\hat{W}_1$ are given by

$$W_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -x_3 & -x_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \hat{W}_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Let us consider the expression corresponding to (3.1):

$$\begin{aligned} x_1' + x_4 &= q_1, \\ x_1' - (x_3 - q_4)x_2' - (x_2 - q_3)x_3' + q_3' x_3 + q_4' x_2 + 2x_2 x_3 - q_3 x_3 - q_4 x_2 &= q_2', \\ x_2 &= q_3, \\ x_3 &= q_4. \end{aligned}$$

This equation has the differential index 1, but:

[2] Note that the proper smoothness assumption required for $W_1(\cdot)\frac{d}{dt}f(x', x, t)$ is discussed later on.

1. The tractability index is not well applicable, because $\ker \frac{\partial f}{\partial x'}$ depends on x.

2. The perturbation index (cf. [15]) of this system is 2, as can be easily seen if we consider $q_1 \equiv q_2 \equiv q_3 \equiv q_4 \equiv 0$ and the following perturbation (cf. [16]):

$$\begin{aligned} x_1' + x_4 &= 0, \\ x_1' - x_2' x_3 - x_2 x_3' + 2x_2 x_3 &= 0, \\ x_2 &= \epsilon \sin t^2, \\ x_3 &= \epsilon \cos t^2. \end{aligned}$$

Straightforward computation leads to $x_4 := \epsilon^2 2t \cos(2t^2) + 2\epsilon^2 (\sin t^2)(\cos t^2)$, which implies that x_4 grows with the derivative of the perturbation.

Let us now consider the expression corresponding to (3.2):

$$\begin{aligned} x_1' + x_4 &= q_1, \\ x_1' + q_3' x_3 + q_4' x_2 &= q_2', \\ x_2 &= q_3, \\ x_3 &= q_4, \end{aligned}$$

For this system, all indices are defined and coincide, they are 1.

Therefore, the projector W_1 should not be differentiated itself. This is due to the fact that W_1 was defined considering the partial derivatives, not the equations themselves. Indeed, W_1 provides information on how to combine the equations we need to differentiate.
This fact motivated the introduction of the diagonal matrix I_{W_1} defined by

$$I_{W_1,i,i} = \begin{cases} 1 & if \quad \exists j \in [1,n] : W_{1\,i,j} \neq 0, \\ 0 & else. \end{cases}$$

Note that I_{W_1} is a projector and that $W_1 I_{W_1} = W_1$.
Hence, the system we consider for the index reduction is

$$(I - \hat{W}_1) f(x', x, t) + W_1(x, t) \frac{d}{dt} (I_{W_1} W_0 f(x', x, t)) = 0.$$

3.2 The Index–1 Formulation

Let us assume that the DAE (1.1) has index–2 and that it has the quasilinear structure (2.1) and that its solutions are continuously differentiable. Motivated by our discussion in Section 3.1 we assume that

$$\mathbf{A3} : \frac{d}{dt} \left\{ I_{W_1} W_0 f(x'(t), x(t), t) \right\} \quad \text{exists}$$

and consider the DAE

$$(I-\hat{W}_1)f(x',x,t)+W_1(x,t)\frac{d}{dt}(I_{W_1}W_0f(x',x,t))=0. \tag{3.3}$$

Moreover, to guarantee the equivalence with (2.1) we need the additional condition that the replaced equations are fulfilled at least in one point

$$\hat{W}_1f(x'(t_0),x(t_0),t_0)=0. \tag{3.4}$$

Remark 3.1 *: Using the quasilinear structure and* $\ker\hat{W}_1\supset\operatorname{im}A$ *(see (2.4)) we have the identities:*
1. $(I-\hat{W}_1)f(x',x,t)=A(x,t)x'+(I-\hat{W}_1)b(x,t)$,
2. $W_0f(x',x,t)=W_0b(x,t)$.

This Ansatz suggests the following definition for the manifold

$$M_1(t):=\Big\{z\in M_0(t)\quad:\quad W_1(z,t)\Big[(I_{W_1}W_0b)'_x(z,t)y+(I_{W_1}W_0b)'_t(z,t)\Big]=0,\\ y\quad=\quad -A^-(z,t)b(z,t)\Big\}.$$

Let us investigate the index of (3.3). More detailed, (3.3) looks like

$$\begin{aligned} A(x,t)x' \quad &+\quad (I-\hat{W}_1)b(x,t)\\ &+\quad W_1(x,t)\Big[(I_{W_1}W_0b)'_x(x,t)x'+(I_{W_1}W_0b)'_t(x,t)\Big]=0. \end{aligned} \tag{3.5}$$

The pencil matrices of (3.5) are given by

$$\begin{aligned} \tilde{A}(x,t) \quad &:= \quad A(x,t)+W_1(x,t)(I_{W_1}W_0b)'_x(x,t)\\ \tilde{B}(x',x,t) \quad &:= \quad \Big\{\Big(A(x,t)+W_1(x,t)(I_{W_1}W_0b)'_x(x,t)\Big)x'\Big\}'_x+\\ &\qquad \{(I-\hat{W}_1)b(x,t)+W_1(x,t)(I_{W_1}W_0b)'_t(x,t)\}'_x. \end{aligned}$$

Because of assumption **A1** it holds that $W_1(x,t)[A(x,t)x']'_x=0$, and it follows

$$W_1(x,t)(I_{W_1}W_0b)'_x(x,t)=W_1(x,t)B(x',x,t).$$

By definition of W_1 we thus obtain by Lemma 2.3 a.:

$$\tilde{A}(x,t)=(A(x,t)+W_1(x,t)B(x',x,t)))P,$$

and from $\tilde{A}(x,t)=(I-W_1(x,t))A(x,t)+W_1(x,t)B(x',x,t)$ we conclude

$$\ker\tilde{A}(x,t)=\ker A(x,t)\cap\ker W_1(x,t)B(x',x,t)=\operatorname{im}Q.$$

At this point we want to emphasize that this implies that the space N corresponding to the original index–2 DAE and the space $\tilde{N}$ corresponding to the reduced index–1 DAE coincide. This means that in both DAEs there appear the same derivatives, which was our objective.

According to Definition 2.1, to prove that (3.5) has index 1, we have to check the nonsingularity of

$$\begin{aligned}\tilde{G}_1(x',x,t) &:= \tilde{A}(x,t)+\tilde{B}(x',x,t)Q\\ &= A(x,t)+\underbrace{W_1(x,t)(I_{W_1}W_0b)'_x(x,t)}_{3}\\ &+ \left[\left\{\left(\underbrace{A(x,t)}_{1}+W_1(x,t)(I_{W_1}W_0b)'_x(x,t)\right)x'\right\}'_x\right.\\ &+ \left.\left\{\underbrace{b(x,t)}_{2}-\hat{W}_1b(x,t)+W_1(x,t)(I_{W_1}W_0b)'_t(x,t)\right\}'_x\right]Q.\end{aligned}$$

To this aim we consider an arbitrary z fulfilling $\tilde{G}_1(x',x,t)z=0$, i.e.,

$$\begin{aligned}0 = \quad & \tilde{G}_1(x',x,t)z=\left(A(x,t)+(\underbrace{\{A(x,t)Px'\}'_x}_{1}+\underbrace{b'_x(x,t)}_{2})Q\right)z\\ &+\underbrace{W_1(x,t)(I_{W_1}W_0b)'_x(x,t)P}_{3}z-\left\{\hat{W}_1b(x,t)\right\}'_xQz \\ &+\left\{W_1(x,t)[(I_{W_1}W_0b)'_x(x,t)x'+(I_{W_1}W_0b)'_t(x,t)]\right\}'_xQz.\end{aligned} \tag{3.6}$$

We split (3.6) by multiplying it by $(I-W_1(x,t))$. From $(I-W_1(x,t))\hat{W}_1=0$ and $W_1=\hat{W}_1W_1$ we have

$$0=(I-W_1(x,t))\tilde{G}_1(x',x,t)z=(A(x,t)+B(x',x,t)Q)z.$$

Hence, it follows that $z=Q_1(x,t)z$ and with

$$\hat{W}_1b'_x(x,t)QQ_1(x,t)=\hat{W}_1(A(x,t)+B(x',x,t)Q)Q_1(x,t)=\hat{W}_1G_1Q_1=0$$

we have

$$\begin{aligned}0=\tilde{G}_1(x',x,t)z &= W_1(x,t)(I_{W_1}W_0b)'_x(x,t)PQ_1z\\ &+\underbrace{\Big[[W_1(x,t)[(I_{W_1}W_0b)'_x(x,t)x'+(I_{W_1}W_0b)'_t(x,t)]\Big]'_xQQ_1z}_{4}.\end{aligned}$$

If we assume that

$$\mathbf{A4}:\ker\left\{W_1(x,t)[(I_{W_1}W_0b)'_x(x,t)x'+(I_{W_1}W_0b)'_t(x,t)]\right\}'_x\subset N\cap S(x,t),$$

then expression 4 remains identical zero and we find

$$0 = W_1(x,t)(I_{W_1}W_0b)'_x(x,t)PQ_1z = W_1(x,t)B(x',x,t)PQ_1z. \tag{3.7}$$

For the canonical projector, i.e., for $Q_1 = Q_1G_2^{-1}BPQ_1$, $G_2Q_1G_2^{-1}$ projects along $imG_1 = kerW_1$. Hence, (3.7) implies $0 = G_2Q_1G_2^{-1}B(x',x,t)PQ_1z = G_2Q_1z = 0$, which leads to $Q_1z = 0$. Thus we have $z = 0$. This means that the matrix $\tilde{G}_1(x',x,t)$ is nonsingular, i.e., the DAE (3.5) has index 1.

What about the equivalence of the equations (2.1) and (3.5)? It seems to be clear that every solution of (2.1) remains also a solution of (3.5). Conversely, we have to show that if we start on M_0, then the whole solution of (3.5) lies there, too. Let $x_\star \in C^1$ be a solution of (3.5) with $x_\star(t_0) \in \tilde{M}_0$ fulfilling (3.4), whereas $\tilde{M}_0$ is the suitable manifold of this index-1 problem. Therefore, (3.5) is fulfilled particularly for $x_\star(t)$. Multiplying (3.5) by $\hat{W}_1$ provides

$$W_1(x_\star(t),t)\frac{d}{dt}(I_{W_1}W_0b(x_\star(t),t)) = 0. \tag{3.8}$$

Using this result and multiplying (3.5) by W_0 we obtain

$$W_0(I-\hat{W}_1)b(x_\star(t),t) = 0, \tag{3.9}$$

i.e., $W_0b(x_\star(t),t) = W_0\hat{W}_1b(x_\star(t),t)$. Further, (3.9) implies with (3.4) $x_\star(t_0) \in M_0$. With $\hat{W}_1 = W_1(x,t)\hat{W}_1$ this implies

$$\begin{aligned}
\frac{d}{dt}(\hat{W}_1b(x_\star(t),t)) &= \hat{W}_1\frac{d}{dt}(\hat{W}_1b(x_\star(t),t)) \\
&= W_1(x_\star(t),t)\hat{W}_1\frac{d}{dt}(\hat{W}_1b(x_\star(t),t)) \\
&= W_1(x_\star(t),t)\frac{d}{dt}(\hat{W}_1b(x_\star(t),t)) \\
&= W_1(x_\star(t),t)\frac{d}{dt}(I_{W_1}W_0\hat{W}_1b(x_\star(t),t)) \\
&\underset{(3.9)}{=} W_1(x_\star(t),t)\frac{d}{dt}(I_{W_1}W_0b(x_\star(t),t)) \underset{(3.8)}{\equiv} 0.
\end{aligned}$$

Therefore, $\hat{W}_1b(x_\star(t),t)$ is constant and because of $x_\star(t_0) \in M_0$ it holds that $\hat{W}_1b(x_\star(t),t) \equiv 0$. This proves the following:

Theorem 3.2 *Let the assumptions* **A1**–**A4** *be fulfilled. Then equation (3.5) has index–1 and the C^1-solutions of the index–2 equation (2.1) and the index–1 equation (3.5) fulfilling (3.4) are equivalent.*

Remark 3.3 *The assumptions* **A1**–**A3** *are dependence and smoothness conditions only. The most interesting condition is given by* **A4**. *When does* **A4** *become valid? Roughly speaking, we can say that* **A4** *is fulfilled if the equation defining M_1 in M_0 does not depend on variables lying in $N \cap S(x,t) = \operatorname{im} QQ_1$.*

1. *It is clear that for linear (time–dependent) DAEs* **A4** *is fulfilled, but this is also true for the case that* $W_1(x,t) = W_1(Px,t)$*, which was investigated by [14]. For Hessenberg systems*

$$\begin{aligned} x_1' + b_1(x_1,x_2,t) &= 0 \\ b_2(x_1,t) &= 0 \end{aligned}$$

it easily can be seen that

$$A = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11}(x_1,x_2,t) & B_{12}(x_1,x_2,t) \\ B_{21}(x_1,t) & 0 \end{pmatrix}, \qquad Q = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}$$

and therefore

$$G_1 = \begin{pmatrix} I & B_{12}(x_1,x_2,t) \\ 0 & 0 \end{pmatrix}, \qquad W_1 = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

Thus, **A4** *is fulfilled.*

2. *If* $N \cap S(x,t) = \operatorname{im} QQ_1 = const$ *and the equation has the structure*

$$A(Ux,t)x' + \tilde{b}(Ux,t) + \mathcal{B}(t)Tx = 0, \tag{3.10}$$

where T *denotes a projector onto the space* $N \cap S(x,t)$ *and* $U := I - T$*. Thus* $W_1 \equiv W_1(Ux,t)$ *holds and* **A4** *is also valid. This case covers a broad class of systems arising from the Modified Nodal Analysis (cf. [9]).*

4 Application to Electrical Networks

We analyze in detail the systems resulting in circuit simulation by means of Modified Nodal Analysis (MNA) in order to show how they fulfil our assumptions. We consider the systems generated by two commonly used modelling techniques: the conventional approach and the charge-oriented approach of the MNA (cf. [17], [18], [9]).

The conventional MNA leads to systems of the form:

$$A(x,t)x' + f(x,t) = 0, \tag{4.1}$$

where the vector of unknowns x consists of

1. the nodal potentials
2. the currents of the voltage-controlled elements.

These systems arise from Kirchhoff's nodal law for each node but the datum node and the characteristic equations of the voltage-controlled elements.

The charge-oriented MNA provides systems of the form:

$$\begin{aligned} \tilde{A}q' + \tilde{f}(x,t) &= 0, && (4.2)\\ q - g(x,t) &= 0, && (4.3) \end{aligned}$$

whereas $\tilde{A}$ is constant.
In this case, the vector of unknowns (q, x) consists of

1. the vector q, that is introduced additionally and contains
   (a) the charge of the capacitors
   (b) the flux of the inductors.
2. the vector x, which remains the same it was for the conventional MNA.

The equations (4.3) correspond to the characteristic equations for charge and flux.

Observe that both modelling techniques are closely related, because

$$\begin{aligned} A(x,t) &= \tilde{A}g'_x(x,t) \quad \text{and} && (4.4)\\ \tilde{f}(x,t) &= f(x,t) + \tilde{A}g'_t(x,t). && (4.5) \end{aligned}$$

The structural properties of these systems have been discussed in detail in [9] for a large class of electric networks. We restrict our further consideration to the class of networks described in [19]. For these networks, it follows from the results of [9] that the assumptions **A1, A2** are always satisfied. Furthermore, assumption **A4** is fulfilled because of Remark 3.3. Therefore, in the following we only have to assume additionally that the smoothness conditions are also given. Condition **A3** will be discussed later on.

Further, it is shown in [9] that the following relations are fulfilled for a special choice of projectors:

1. There exists a projector W_1 along to image of the matrix G_1 corresponding to the conventional MNA fulfilling

$$W_1 \quad \text{is constant.}$$

2. There exists a projector $\tilde{W}_1$ along to image of the matrix $\tilde{G}_1$ corresponding to the charge-oriented MNA that fulfils

$$\tilde{W}_1 = \tilde{W}_1(x,t) = \begin{pmatrix} W_1 & W_1 \cdot H(x,t) \\ 0 & 0 \end{pmatrix}, \tag{4.6}$$

whereas the matrix H is defined in a way that particularly

$$W_1 \cdot H(x,t) g'_x(x,t) = W_1 \tilde{f}'_x(x,t) = W_1 f'_x(x,t) \tag{4.7}$$

is fulfilled (cf. [9]). Furthermore,

$$\operatorname{im} \tilde{W}_1(x,t) = \operatorname{im} \begin{pmatrix} W_1 & 0 \\ 0 & 0 \end{pmatrix}$$

holds, whereas the second column corresponds to the equations (4.3).

Remark: For the systems arising from MNA this implies that for both formulations the reduced index-1 systems are again closely related, because we have:

1. For the conventional MNA

$$A(x,t)x' + W_1 f'_x(x,t)x' + (I - W_1)f(x,t) + W_1 f'_t(x,t) = 0. \tag{4.8}$$

2. For the charge-oriented MNA, if we set the projector fulfilling (2.4)

$$\hat{\tilde{W}}_1 := \begin{pmatrix} W_1 & 0 \\ 0 & 0 \end{pmatrix} \tag{4.9}$$

and make use of the relations

$$\begin{aligned} \tilde{W}_1 \cdot \begin{pmatrix} 0 & \tilde{f}'_x(x,t) \\ I & -g'_x(x,t) \end{pmatrix} &= \begin{pmatrix} W_1 & W_1 H(x,t) \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \tilde{f}'_x(x,t) \\ I & -g'_x(x,t) \end{pmatrix} \\ &\underset{(4.7)}{=} \begin{pmatrix} W_1 H(x,t) & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

then we obtain

$$\begin{aligned} \tilde{A}q' + W_1 H(x,t)q' + (I - W_1)\tilde{f}(x,t) & & \\ +W_1 \tilde{f}'_t(x,t) - W_1 H(x,t) g'_t(x,t) &= 0, & (4.10) \\ q - g(x,t) &= 0. & (4.11) \end{aligned}$$

Observe now that, by (4.5),

$$\begin{aligned} (I - W_1)\tilde{f}(x,t) &= (I - W_1)\big(f(x,t) + \tilde{A}g'_t(x,t)\big) \\ &= \tilde{A}g'_t(x,t) + (I - W_1)f(x,t), \\ W_1 \tilde{f}'_t(x,t) &= W_1\big(f(x,t) + \tilde{A}g'_t(x,t)\big)'_t = W_1 f'_t(x,t) \end{aligned}$$

are fulfilled and therefore (4.10) is equivalent to

$$\begin{aligned} \tilde{A}q' + \tilde{A}g'_t(x,t) + W_1 H(x,t)(q' - g'_t(x,t)) & \\ +(I - W_1)f(x,t) + W_1 f'_t(x,t) &= 0. \end{aligned}$$

Finally, taking into account that, due to the derivation of (4.11) and property (4.7), the relation

$$W_1 H(x,t)(q' - g'_t(x,t)) = W_1 H(x,t) g'_x(x,t)x' = W_1 f'_x(x,t)x'$$

is fulfilled, we recognize that the system (4.10)-(4.11) is again closely related to (4.8) because we may write it in the form

$$\begin{aligned} \tilde{A}q' + \tilde{A}g_t'(x,t) + W_1 f_x'(x,t)x' + (I - W_1)f(x,t) + W_1 f_t'(x,t) &= 0, \\ q - g(x,t) &= 0. \end{aligned}$$

Let us finally focus on the smoothness assumptions required for the DAEs arising from MNA. Since W_1 is a constant projector, for the conventional MNA we only have to require that $\frac{d}{dt}(W_1 f(x(t),t))$ exists, instead of **A3**. Taking into account that for the charge-oriented MNA the projector $\tilde{W}_1$ can be chosen as described in (4.6), in this case it suffices to assume the existence of $\frac{d}{dt}(W_1 \tilde{f}(x(t),t))$ and $\frac{d}{dt}(q - g(x,t))$. Observe that the smoothness conditions coincide.

5 The Computation of Consistent Initial Values

Consider the initial value problem

$$\begin{aligned} f(x',x,t) &= 0, && (5.1) \\ PP_1(x(t_0),t_0)(x(t_0) - \alpha) &= 0 && (5.2) \end{aligned}$$

with a given α, which is adequate for an index–2 DAE [20].
Note that the assumption $\ker f_{x'}'(x',x,t) = N$ leads to $f(x',x,t) = f(Px',x,t)$. A vector $x_0 \in R^n$ is a consistent initial value of (1.1) if there exists a solution of (1.1) that satisfies $x(t_0) = x_0$. To compute a consistent initial value we therefore determine the unknown values $(I - PP_1(x(t_0),t_0))x(t_0)$ and $(Px)'(t_0)$.

For the calculation of the consistent initial values we use the equations arising from the reduced index–1 representation and the conditions concerning the initial values. They are given by:

$$\begin{aligned} (I - \hat{W}_1)f(x',x,t) + W_1(x,t)\frac{d}{dt}(I_{W_1}W_0 f(x',x,t)) &= 0, && (5.3) \\ \hat{W}_1 f(x'(t_0),x(t_0),t_0) &= 0, && (5.4) \\ PP_1(x(t_0),t_0)(x(t_0) - \alpha) &= 0. && (5.5) \end{aligned}$$

If we consider this set of equations in the point $t = t_0$ with the aim to calculate values $x = x(t_0)$, $y = Px'(t_0)$, and rearrange them, we obtain the system

$$\left.\begin{aligned} f(y,x,t) &= 0, \\ PP_1(x,t)(x - \alpha) &= 0, \\ Qy &= 0, \\ W_1(x,t)[(I_{W_1}W_0 b)_x'(x,t)y + (I_{W_1}W_0 b)_t'(x,t)] &= 0, \end{aligned}\right. \qquad (5.6)$$

to determine the unknowns (y,x), whereas $Qy = 0$ is introduced to guarantee $y = Py$.

Theorem 5.1 *Let the assumptions* **A1–A4** *be valid and suppose additionally that the implication*

$$\textbf{A5}: \quad \{PP_1(x,t)(x-\alpha)\}'_x(I - PQ_1(x,t))z = 0 \Rightarrow PP_1(x,t)z = 0 \qquad (5.7)$$

holds. Then the system (5.6) has a full rank Jacobian matrix in a neighborhood of a solution.

For a better understanding of **A5** observe that for a constant or a time-dependent projector $PP_1(t)$ the assumption is trivially fulfilled. Some more general cases are discussed in Remark 5.3.

Proof: For $A := f'_y(y,x,t)$ and $B := f'_x(y,x,t)$ the Jacobian matrix of (5.6) reads

$$J = \begin{pmatrix} A & B \\ 0 & \{PP_1(x,t)(x-\alpha)\}'_x \\ Q & 0 \\ W_1B & \{W_1((I_{W_1}W_0b)'_x y + (I_{W_1}W_0b)'_t)\}'_x \end{pmatrix}.$$

To prove its nonsingularity we consider a z fulfilling $Jz = 0$. For $z = (z_y, z_x)^T$ we obtain the first equation:

$$Az_y + Bz_x = 0.$$

Multiplying it by G_2^{-1} and PP_1, PQ_1 and Q yields

$$\begin{aligned} PP_1z_y + PP_1G_2^{-1}BPP_1z_x &= 0, &(5.8)\\ PQ_1z_x &= 0, &(5.9)\\ -QQ_1z_y + (QG_2^{-1}BPP_1 + QQ_1 + Q)z_x &= 0. &(5.10) \end{aligned}$$

The other equations provide

$$\begin{aligned} \{PP_1(x-\alpha)\}'_x z_x &= 0, &(5.11)\\ Qz_y &= 0, &(5.12)\\ W_1Bz_y + \{W_1((I_{W_1}W_0b)'_x y + (I_{W_1}W_0b)'_t)\}'_x z_x &= 0. &(5.13) \end{aligned}$$

With $PQ_1z_x = 0$ from (5.9) we derive from (5.11) the expression of assumption **A5** and it follows that $PP_1z_x = 0$. With (5.8) it follows that $PP_1z_y = 0$. From (5.12) we obtain that $Qz_y = 0$ and thus we have $z_y = PQ_1z_y$ and $z_x = Qz_x$. With (5.10) this leads to $QQ_1z_y = Qz_x$ and, finally, **A4** and (5.13) imply $W_1BPQ_1z_y = 0$, i.e., analogously as it was concluded from (3.7), $PQ_1z_y = 0$, which means that $Qz_x = 0$.

Remark 5.2 *Let us have a look at system (5.6). Is it really necessary to use the second equation* $PP_1(x-\alpha) = 0$ *in this form, which does not make the theoretical considerations easier ? In fact, if we know the projector* $PP_1(x_*(t),t) = PP_1(t)$ *on the solution, which depends only on* t*, we can always fix the free parameters* *corresponding to the degree of freedom correctly. This motivated the consideration of* $PP_1(x(t),t)$*. In [2], [5] a nonlinear initial condition* $B(x(t_0)) = 0$ *is required, assuming that* B *is chosen in such a way that the initial value problem has a unique solution. In contrast to our condition, which has, indeed, the structure* $B(x) = PP_1(x,t)(x-\alpha)$*, we already know that the initial value problem has a unique solution (see [20]). Nevertheless if we know an easier (e.g. not depending on* x*) condition, we can replace our condition by a similar one with the same degree of freedom if the obtained Jacobian matrix becomes nonsingular. For instance, if* $\ker PQ_1(x,t) = const$ *(valid for the conventional Modified Nodal Analysis), there exists a constant projector* PV *with* $\ker PV = \ker PQ_1$*. As both projectors project along the same subspace, it holds that*

$$PQ_1PV = PQ_1, \quad PVPQ_1 = PV.$$

Therefore we better describe the fixing of the free parameters corresponding to the degree of freedom by considering

$$(P - PV)(x-\alpha) = 0$$

instead of equation (5.5)(cf. [21]). When analyzing the Jacobian then, we obtain analogously as above $PQ_1z = 0$ *and also* $(P-PV)z = 0$*, i.e.* $Pz = 0$*. Of course, this leads to* $PP_1z = 0$*.*

Remark 5.3 *If we make use of the fact that* $PP_1 = P - PQ_1$*, then the left-hand side of the implication* **A5** *reads*

$$(P - PQ_1(x,t))z - \{PQ_1(x,t)(x-\alpha)\}'_x(I - PQ_1(x,t))z = 0. \tag{5.14}$$

Therefore, for the following cases assumption **A5** *is fulfilled:*

1. $PP_1 = const$ *or only* t*–dependent.*

2. $\operatorname{im} PQ_1(x,t) = const$ *(valid for the charge–oriented Modified Nodal Analysis). This means that a constant projector* $P\bar{V}$ *exists with the same image as* PQ_1*. Both projectors project onto the same subspace, it holds that*

   $$P\bar{V}PQ_1 = PQ_1, \quad PQ_1P\bar{V} = P\bar{V} \text{ or } P(I - PQ_1)P\bar{V} = PP_1P\bar{V} = 0.$$

   Therefore the equation (5.14) is equal to

   $$(P - PQ_1(x,t))z - P\bar{V}\{PQ_1(x,t)(x-\alpha)\}'_x(I - PQ_1(x,t))z = 0.$$

   Multiplying this by PP_1 *we obtain that the assumption* **A5** *is fulfilled because of* $PP_1P\bar{V} = 0$*.*

3. *In case of* $PQ_1(x,t) = PQ_1(Px,t)$ *(valid for the mechanical systems described in [22]), equation (5.14) becomes*

   $$PP_1(x,t)z - \{PQ_1(x,t)(x-\alpha)\}'_xPP_1(x,t)z = 0.$$

If we multiply this expression by $PP_1(x,t)$, we see that the nonsingularity of $(I - PP_1(x,t)\{PQ_1(x,t)(x-\alpha)\}'_x)$ implies $PP_1(x,t)z = 0$. This is valid e.g. for the pendulum, but up to now it was not possible to prove this for general mechanical systems.

6 Algorithmic Realization

Our aim was to implement a general purpose code which is not based on the quasilinear structure. From (5.3) we obtain

$$\begin{aligned} W_1(x,t)[(I_{W_1}W_0 b)'_x(x,t)y &+ (I_{W_1}W_0 b)'_t(x,t)] = \\ &= W_1(x,t)[(A(x,t)x')'_x + b'_x)y + A'_t(x,t)x' + b'_t)] \\ &= W_1(x,t)(By + f'_t(y,x,t)). \end{aligned}$$

In our realization we choose the projector W_1 by $W_1 := G_2 PQ_1 G_2^{-1}$. This has the advantage that we are able to combine all equations in two parts only (see (6.1)), but the disadvantage is that for systems with a very simple (i.e. constant) projector W_1 we choose now a projector with difficult dependences on x.
It is easy to see that for this projector $W_1 G_1 = G_2 PQ_1 G_2^{-1} G_1 = G_2 PQ_1 P_1 = 0$, which leads to

$$\begin{aligned} W_1(x,t)(B(y,x,t)y + f'_t(y,x,t)) &= G_2 PQ_1 G_2^{-1}(B(y,x,t)y + f'_t(y,x,t)), \\ &= G_2 PQ_1(y + G_2^{-1} f'_t(y,x,t)). \end{aligned}$$

If we do so, we have to solve the following nonlinear systems of equations

$$\begin{aligned} f(y,x,t) &= 0, \\ PP_1(x,t)(x-\alpha) + PQ_1(y + G_2^{-1} f'_t(y,x,t)) + Qy &= 0. \end{aligned} \tag{6.1}$$

For the solution of (6.1) we use a Newton–like method, where the used "Jacobian" matrix does not take into consideration the dependence of the projectors PP_1, PQ_1 and of G_2, which leads to a "Jacobian" matrix

$$J = \begin{pmatrix} A & B \\ PQ_1(I + G_2^{-1}A') + I - PP_1 & PP_1 + PQ_1 G_2^{-1} B' \end{pmatrix},$$

where $A' := f''_{x't}$ and $B' := f''_{xt}$, with the explicit inverse

$$J^{-1} = \begin{pmatrix} PP_1 - PQ_1 Z & -PP_1 Y + I - PP_1 \\ I - PP_1 - QQ_1(I + Z) & PP_1 + QQ_1 - QY \end{pmatrix} \begin{pmatrix} G_2^{-1} & \\ & I \end{pmatrix},$$

where $Z := G_2^{-1}(A' + B'(I - PP_1))$ and $Y := G_2^{-1} BPP_1$. The solution of (6.1) is realized by the following principle algorithm:

1. $i = -1$

2. input: y_0^0, x_0^0, t_0
3. $\alpha := x_0^0$
4. *initialization*: i=i+1
5. computation of $A := f'_{x'}(y_i^0, x_i^0, t_0)$(check of singularity - ODE), $B := f'_x(y_i^0, x_i^0, t_0)$, where A and B are approximated by finite differences or computed by an user–written subroutine
6. computation of the matrix chain elements and projectors:
   Q, G_1 (check of singularity - index 1)
   if G_1 singular: Q_1, G_2 (check of singularity - not index 2)
   if G_2 nonsingular:$Q_{1s} := Q_1 G_2^{-1} BP, G_{2s}, G_{2s}^{-1}, PP_1, PQ_1$
7. computation of J^{-1}
8. computation of (6.1), where f'_t is approximated by the central difference quotient or computed by an user–written subroutine
9. if norm(6.1) $<$ accuracy **goto** *finish*
10. $j = 1$
11. *jump*:
12. Newton step, calculation of the correction $\Delta y, \Delta x$
13. calculation of y_i^j, x_i^j
14. if norm(6.1) $>$ accuracy
    then
        j=j+1
        goto *jump*
    else
        $y_{i+1}^0 := y_i^j, x_{i+1}^0 := x_i^j$
        goto *initialization*
15. *finish*:

6.1 The Computation of the Projectors

The computation of the projectors, which we need only at the point t_0, represents an important part of the algorithm. To this end we have carried out a decomposition of a matrix G and the determination of its rank. For this problem the following methods are available: The Householder decomposition or the SVD. We prefer the first option because of the computational costs.
We are looking for the nullspace projectors Q_i of the matrix chain element G_i

(the given representation assumes constant projectors). The matrix chain grows with (see [7])

$$G_{i+1} := G_i + B_i Q_i, \quad B_{i+1} := B_i P_i.$$

Let us assumes that we have a (e.g. Householder) decomposition of G_i of the form

$$U_i G_i = \begin{pmatrix} R_{i_{11}} & R_{i_{12}} \\ 0 & 0 \end{pmatrix} P_{c_i}^T$$

with U_i an orthogonal matrix, $R_{i_{11}}$ a nonsingular, triangular matrix of dimension r_i and P_{c_i} a column permutation matrix. Using U_i we define $U_i B_i =: (\bar{B_{i1}}, \bar{B_{i2}}) P_{c_i}^T$. Then a projector onto $\ker A$ is given (as it was used in former papers [23], [4]) by

$$Q_i = P_{c_i} \begin{pmatrix} 0 & R_{i_{11}}^{-1} R_{i_{12}} \\ 0 & I_{n-r_i} \end{pmatrix} P_{c_i}^T, \quad P_i := I - Q_i.$$

This gives the following representation of G_{i+1}:

$$U_i G_{i+1} = \begin{pmatrix} R_{i_{11}} & \vdots & \\ & \vdots & -\bar{B_{i1}} R_{i_{11}}^{-1} R_{i_{12}} + \bar{B_{i2}} \\ 0 & \vdots & \end{pmatrix} P_{c_i}^T \quad =: \begin{pmatrix} R_{i_{11}} & \vdots & \bar{R_{i_{12}}} \\ 0 & \vdots & \bar{R_{i_{22}}} \end{pmatrix} P_{c_i}^T.$$

Now we have to decompose $\bar{R_{i_{22}}}$ and obtain $\bar{U_i} \bar{R_{i_{22}}} = \begin{pmatrix} R_{i_{22,1}} & R_{i_{22,2}} \\ 0 & 0 \end{pmatrix}$. Updating the matrices $U_{i+1} := \begin{pmatrix} I & \\ & \bar{U_i} \end{pmatrix} U_i, \quad P_{c_{i+1}}^T := \begin{pmatrix} I & \\ & \bar{P_{c_i}^T} \end{pmatrix} P_{c_i}^T$ we obtain the relevant representation for G_{i+1}, and the adequate representation for B_{i+1} uses U_{i+1} and $P_{c_{i+1}}^T$.
If we start with $G_0 = A$ and $U_0 = I$ we attain, in the index–k case after $k+1$ steps, the nonsingular matrix G_k in a decomposed triangular form. This makes the computation of the so–called canonical projector, which is given by $Q_{ks} := Q_k A_k^{-1} B_k$, relatively simple.

7 Examples

We use a first example with constant coefficient matrices, which does not meet the assumptions from [1], to illustrate the mode of action of the method. The system is represented by

$$\begin{aligned} x_1' - (x_1 + 2x_2 + 3x_3) &= 0, \\ x_1 + x_2 + x_3 + 1 &= 0, \\ 2x_1 + x_2 + x_3 &= 0. \end{aligned}$$

The matrices A and B are given by

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} -1 & -2 & -3 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}.$$

The relevant projectors are $PP_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $PQ_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\hat{W}_1 = W_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 1 \end{pmatrix}$. Since PP_1 is identical zero, it is not possible to prescribe any initial values, all values are determined. The equations for the computation of the consistent initial values (5.6) are

$$\begin{aligned} Ay + Bx = q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} + \begin{pmatrix} -1 & -2 & -3 \\ 1 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix} \\ Qy = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= 0 \\ W_1 By = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= 0. \end{aligned}$$

The solution is directly visible as $y = 0$ from the second and the third equation and $x = \begin{pmatrix} 1 \\ -5 \\ 3 \end{pmatrix}$.

We take the next example from [11], [24]. The FORTRAN–subroutine for the description of the problem reads

```
      subroutine ftraj(y,x,t,fyxt,*)
c
c     x(1) = H
c     x(2) = xi
c     x(3) = lambda
c     x(4) = VR
c     x(5) = gamma
c     x(6) = A
c     x(7) = alpha
c     x(8) = beta
c
c-----+--+---------+---------+---------+---------+---------+
      implicit real*8 (a-h,o-z)
      real*8 mue,m,l
```

```
      common /round/ uround
      dimension y(1),x(1),fyxt(1)
      cl(p)=1d-2*p
      data ae,mue,ome,m,s,pi /0.209029d+8,0.1407653916d+17,
     *     0.72921159d-4,0.2890532728d+1,1d0,
     *     3.14159265358979323846264338327950288d0/
      if(abs(x(4)).lt.uround) return 1
      r=x(1)+ae
      if(abs(r).lt.uround) return 1
      g=mue/r/r
      hpi=pi/18d1
      hsing=sin(x(5)*hpi)
      hsina=sin(x(6)*hpi)
      hcosg=cos(x(5)*hpi)
      hcosa=cos(x(6)*hpi)
      hsinl=sin(x(3)*hpi)
      hcosl=cos(x(3)*hpi)
      hn1=r*hcosl
      if(abs(hn1).lt.uround) return 1
      hn2=x(4)*hcosg
      if(abs(hn2).lt.uround) return 1
      vr2=x(4)*x(4)
      hcl=cl(x(7))
      rho=0.002378d0*exp(-x(1)/238d2)
      l=0.5d0*rho*hcl*s*vr2
      cd=0.04d0+0.1d0*hcl*hcl
      d=0.5d0*rho*cd*s*vr2
      ome2=ome*ome
      sc=hsinl*hcosa
      fyxt(1)=y(1)-x(4)*hsing
      fyxt(2)=y(2)-x(4)*hcosg*hsina/hn1
      fyxt(3)=y(3)-x(4)*hcosg*hcosa/r
      fyxt(4)=y(4)+d/m+g*hsing+ome2*r*hcosl*(sc*hcosg-hcosl*hsing)
      fyxt(5)=y(5)-l*cos(x(8)*hpi)/m/x(4)-hcosg/x(4)*(vr2/r-g)-2d0*
     *      ome*hcosl*hsina-ome2*r*hcosl/x(4)*(sc*hsing+hcosl*hcosg)
      fyxt(6)=y(6)-l*sin(x(8)*hpi)/m/hn2-x(4)/r*hcosg*hsina*tan(x(3)*
     *   hpi)+2d0*ome*(hcosl*hcosa*tan(x(5)*hpi)-hsinl)-ome2*r*hcosl*
     *   hsinl*hsina/x(4)/hcosg
      fyxt(7)=x(5)+1d0+9d0*t*t/9d4
      fyxt(8)=x(6)-45d0-9d1*t*t/9d4
      return
      end
```

With

$$\hat{W}_1 = W_1 = \begin{pmatrix} 0 &&&&&&& \\ & 0 &&&&&& \\ && 0 &&&&& \\ &&& 0 &&&& \\ &&&& 0 &&& \\ &&&&& 0 && \\ &&&&&& 1 & \\ &&&&&&& 1 \end{pmatrix}$$

we discover that we have to differentiate the latter two equations to obtain the derivatives of the PQ_1–components of x, where

$$PQ_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{500}x_4x_7\cos(\frac{\pi x_8}{180}) & -\frac{1}{500}x_4x_7\cos(\frac{\pi x_5}{180})\sin(\frac{\pi x_8}{180}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Note that the kernel of PQ_1 is constant.

In [11] the following consistent initial value is ascertained by :

$$x = (100000 \quad 0 \quad 0 \quad 12000 \quad -1 \quad 45 \quad 2.6728700742 \quad -0.05220958616134),$$

where the first six components were given by the task and the two latter were computed analytically making use of the special structure of the problem.

For the computation of the initial values we start from the perturbed values

$$\alpha = x_0^0 = (100000 \quad 0 \quad 0 \quad 12000 \quad -1 \quad 45 \quad 2 \quad -1)$$

and $y_0^0 = 0$, $t_0 = 0$. The result with a defect of $7.4544861723 \cdot 10^{-15}$ is given by

$$y = \begin{pmatrix} -2.09428877247402130D+02 \\ 4.03943694670064170D-04 \\ 4.03943694670064170D-04 \\ -3.49782603527764870D+01 \\ 3.43346554416848270D-95 \\ 2.19741794826782840D-93 \\ 0 \\ 0 \end{pmatrix}, x = \begin{pmatrix} 100000 \\ 7.39557098644698560D-31 \\ 1.09476442525376340D-47 \\ 12000 \\ -1 \\ 45 \\ 2.67287004806323260D+00 \\ -5.22095857372783550D-02 \end{pmatrix}.$$

Since for the approximation of f_t' we used a step-size $h = 10^{-3}$, the accuracy is as expected.

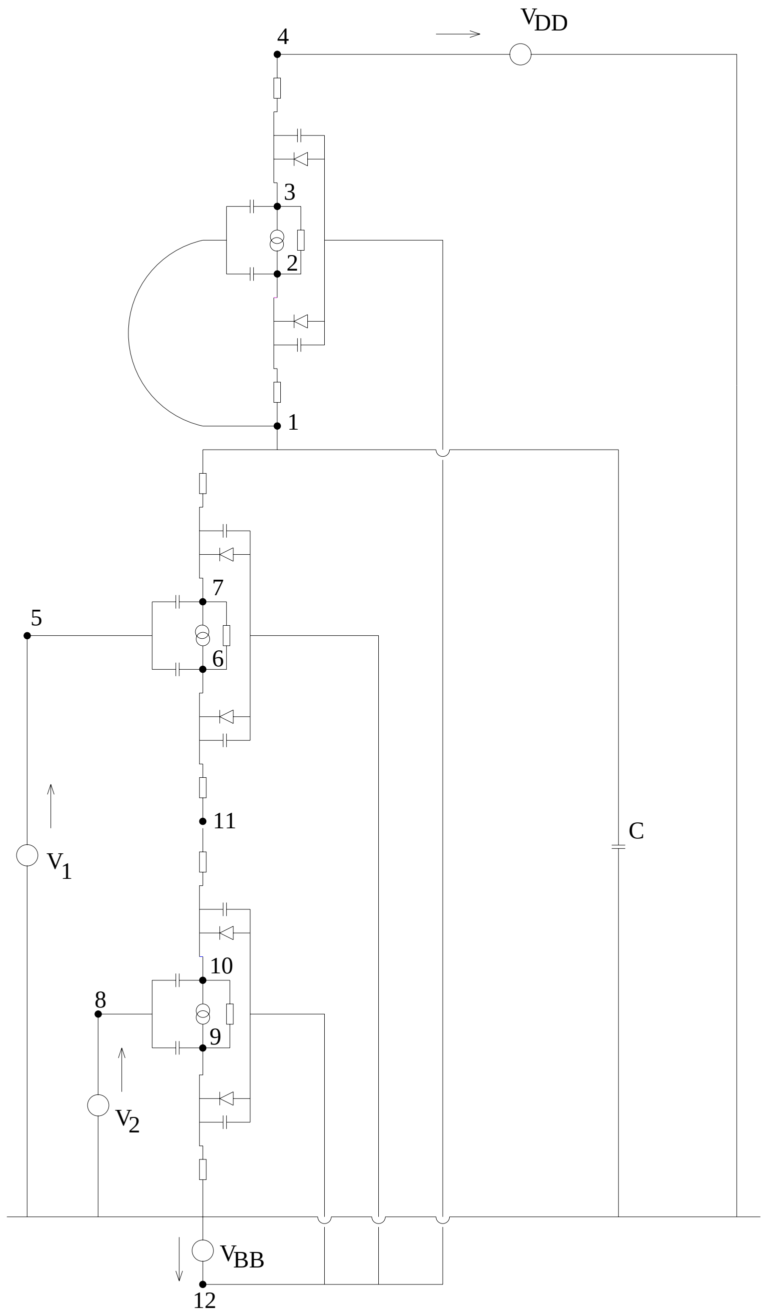

Figure 1: NAND–Gate

The last example comes from the electrical network simulation and was extensively discussed in [8]. Here we consider the NAND–gate model containing nonlinear capacitances. The NAND–gate circuit is represented in Figure 1.

The DAE obtained by the charge-oriented Modified Nodal Analysis from this circuit has dimension 29, where the dimensions of the various parts are as follows: $dim(\text{im } PP_1) = 7, dim(\text{im } PQ_1) = 3$ and $dim(\text{im } Q) = 19$. As usual in the simulation of electrical networks (see e.g. [19]) we start with the so–called DC-operating point (i.e. $y = 0$). The initial values and the solution after 2 iterations with a defect of $6.6994486890D - 16$ (initial defect $7.2053507612D - 10$) are given by

$$\alpha = \begin{pmatrix} 2.500D-13 \\ -2.423D-25 \\ -1.522D-27 \\ 5.679D-13 \\ 5.679D-13 \\ -3.000D-13 \\ -7.049D-14 \\ 5.679D-13 \\ 1.818D-13 \\ -7.049D-14 \\ -6.071D-29 \\ 1.818D-13 \\ 1.031D-13 \\ 5.000D+00 \\ 5.000D+00 \\ 5.000D+00 \\ 5.000D+00 \\ .000D+00 \\ 1.175D+00 \\ 5.000D+00 \\ .000D+00 \\ 1.012D-15 \\ 1.175D+00 \\ 1.175D+00 \\ -2.500D+00 \\ -5.740D-42 \\ .000D+00 \\ 1.113D-14 \\ -1.138D-14 \end{pmatrix}, \quad x = \begin{pmatrix} 2.49999999999795810D-13 \\ 7.20937266233088540D-25 \\ -9.64763467748850800D-25 \\ 5.67931034483715540D-13 \\ 5.67931034481278110D-13 \\ -2.99999999999753350D-13 \\ -7.07126793345685290D-14 \\ 5.67931034482234760D-13 \\ 1.81607077172520250D-13 \\ -7.07126793345671350D-14 \\ -1.26217744835361880D-29 \\ 1.81607077172518920D-13 \\ 1.03103448275862090D-13 \\ 4.99999999997230570D+00 \\ 4.99999999998838440D+00 \\ 5.00000000004383870D+00 \\ 5.00000000000000000D+00 \\ -5.43200633726060640D-52 \\ 1.17854465551473990D+00 \\ 4.99999999999589040D+00 \\ 6.06471004166420880D-59 \\ 3.49073780770501550D-11 \\ 1.17854465557611880D+00 \\ 1.17854465554542930D+00 \\ -2.50000000000000000D+00 \\ -7.40659401174118060D-05 \\ -7.83789956664430850D-12 \\ 7.40659263235377980D-05 \\ 1.09597363218609600D-11 \end{pmatrix}.$$

As we expected according to [19], only x_{26} and x_{28}, i.e. the currents through V_1 and V_{BB}, changed considerably.

8 Conclusion

Under weak assumptions a method has been proposed to choose suitable equations of an index–2 DAE, whose differentiation leads to an index reduction. Based on this result, a numerical algorithm to compute consistent initial values has been developed. Conditions that guarantee its successful application have been carefully discussed. Nevertheless, there are still a few open problems. What are the necessary assumptions to describe the hidden constraints with the projectors related to the tractability-index? Is there a possibility to describe them without these projectors? Which possibility to fix values for a subset of variables whose cardinality is the so-called degree of freedom is the "best" for a given problem? These questions will be handled in further investigations.

Acknowledgment
We would like to thank R. März for many helpful discussions and valuable comments on a first version of this paper.

References

[1] C. C. Pantelides, "The consistent initialization of differential-algebraic systems," *SIAM J. Sci. Statist.Comput.*, vol. 9, pp. 213–231, 1988.

[2] B. Leimkuhler, *Approximation Methods for the Consistent Initialization of Differential-Algebraic Equations.* PhD thesis, University of Illinois, 1988.

[3] B. Hansen, "Computing consistent initial values for nonlinear index–2 differential-algebraic equations," in *Seminarbericht 92–1* (E. Griepentrog, M. Hanke, and R. März, eds.), pp. 142–157, Berlin: Humboldt-Universität, 1992.

[4] R. Lamour, "A shooting method for fully implicit index–2 differential–algebraic equations," *SIAM J. Sci. Comput*, vol. 18, pp. 94–114, 1997.

[5] P. Amodio and F. Mazzia, "An algorithm for the computation of consistent initial values for differential–algebraic equations," *Numerical Algorithms*, vol. 19, pp. 13–23, 1998.

[6] E. Griepentrog and R. März, *Differential-Algebraic Equations and Their Numerical Treatment.* Teubner-Texte Math. 88, Leipzig: Teubner, 1986.

[7] R. März, "Numerical methods for differential-algebraic equations," *Acta Numerica*, pp. 141–198, 1992.

[8] C. Tischendorf, *Solution of index-2 differential algebraic equations and its application in circuit simulation.* PhD thesis, Humboldt-Universität zu Berlin, 1996.

[9] D. Estévez Schwarz and C. Tischendorf, "Structural analysis for electric circuits and consequences for MNA," Preprint 98–21, Humboldt-Universität, Berlin, 1998.

[10] C. W. Gear, "Differential-algebraic equation index transformations," *SIAM J. Sci. Stat. Comput.*, vol. 9, pp. 39–47, 1988.

[11] K. Brenan, S. Campell, and L. Petzold, *Numerical solution of initial-value problems in differential–algebraic equations.* New York: North-Holland, 1989.

[12] E. Griepentrog, "Index reduction methods for differential-algebraic equations," in *Seminarbericht 92–1* (E. Griepentrog, M. Hanke, and R. März, eds.), pp. 14–29, Berlin: Humboldt-Universität, 1992.

[13] R. März and A. R. Rodriguez S., "Analyzing the stability behaviour of DAE solutions and their approximations," Preprint 99–2, Humboldt-Universität, Berlin, 1999.

[14] R. März, "Analysis and numerics of DAEs." Special lecture, 1998.

[15] E. Hairer, C. Lubich, and M. Roche, *The numerical solution of differential-algebraic systems by Runge-Kutta methods.* Lecture Notes in Math. 1409, Springer Verlag, 1989.

[16] S. Campbell and C. Gear, "The index of general nonlinear DAEs," *Numer. Math.*, vol. 72, pp. 173–196, 1995.

[17] M. Günther and U. Feldmann, "CAD-based electric-circuit modeling in industry I. Mathematical structure and index of network equations," *Surv. Math. Ind.*, vol. 8, pp. 97–129, 1999.

[18] R. März and C. Tischendorf, "Recent results in solving index 2 differential-algebraic equations in circuit simulation," *SIAM J. Sci. Stat. Comput.*, vol. 18(1), pp. 139–159, 1997.

[19] D. Estévez Schwarz, "Consistent initial values for DAE systems in circuit simulation," Preprint 99–5, Humboldt-Universität, Berlin, 1999.

[20] R. März, "EXTRA-ordinary differential equations: Attempts to an analysis of differential-algebraic systems," *Progress in Mathematics*, vol. 168, pp. 313–334, 1998.

[21] D. Estévez Schwarz, "Topological analysis for consistent initializatioin in circuit simulation," Preprint 99–3, Humboldt-Universität, Berlin, 1999.

[22] C. Gear, G. Gupta, and B. Leimkuhler, "Automatic integration of Euler-Lagrange equations with constraints," *J. Comp. and Appl. Math.*, vol. 12/13, pp. 77–90, 1985.

[23] R. Lamour, "A well–posed shooting method for transferable DAEs," *Numerische Mathematik*, vol. 59, pp. 815–829, 1991.

[24] B. Leimkuhler, L. Petzold, and C. Gear, "Approximation methods for the consistent initialization of differential–algebraic equations," *SIAM J. Numer. Anal.*, vol. 28, pp. 205–226, 1991.