Consistent initialization for higher-index DAEs using a projector based minimum-norm specification

Diana Estévez Schwarz
Beuth Hochschule für Technik Berlin

René Lamour
Humboldt Universität zu Berlin, Institut für Mathematik

January 21, 2016

Abstract

Higher index differential-algebraic equations (DAEs) present explicit and hidden constraints that restrict the prescription of initial values. In general, these constrains form an underdetermined system of equations, which can be solved in different ways. In this contribution we present a new method that, by construction, realizes a prescription for some differentiated components considering a special minimum-norm specification. The constraints yield adequate values for the undifferentiated components and for the remaining differentiated components. The orthogonal projectors that describe these different types of components are calculated considering the derivative array obtained by automatic differentiation.

Keywords: DAE, differential-algebraic equation, index, derivative array, projector based analysis, consistent initial value, Taylor series, automatic differentiation, underdetermined systems of equations

MSC-Classification: 65L05, 65L80, 34A09, 34A34, 65D25, 41A58

1 Introduction

In the following we will mainly focus on linear DAEs of the form

$$A(t)x' + B(t)x = q(t), \tag{1}$$

where A is singular. This means that (1) consists of coupled systems of differential equations and constraints. Moreover, in the so-called higher-index case the differentiation of (1) leads to further constraints, referred to as hidden constraints. All these constrains restrict the choice of initial values. For our purposes, there are two major difficulties in this context:

- On the one hand, a description of all explicit and hidden constraints is required. In this paper, we will compute them using automatic differentiation and the characterization from [11].

- On the other hand, starting from the expressions for the constraints and some initial guesses, so-called consistent initial values have to be computed. Since they are not unique in general, suitable criteria become necessary. For this aspect we develop a new approach.

Since the explicit and hidden constraints in general form an underdetermined system of equations, we summarize in Section 2 some notations and well-known results from linear algebra, which we will use later on.

In Section 3 we present a particular algorithm to solve possibly underdetermined systems of linear equations with specific properties. These properties are fulfilled for DAEs, such that in Section 4 our approach to compute consistent initial values can be introduced, focusing on the characteristics we have for DAEs.

In practice, the user of a simulation package often wants to specify exactly the values of some particular variables or components. In Section 5 we describe how this can be realized in combination with our new approach. Once it is clear how to compute consistent initial values, the computation of consistent Taylor series results to be a straight-forward generalization, which is described in Section 6. A generalization for nonlinear DAEs will be discussed in Section 7 and some important aspects concerning our implementation in Python are summarized in Section 8.

Since there is a large body of literature concerning the computation of initial values of DAEs, in Section 9 we compare the new approach with some other methods from literature. Finally, the algorithm is illustrated and discussed for the nonlinear examples presented in Section 10.

2 Orthogonal projectors and minimum-norm least squares solutions

For a better understanding of the forthcoming section we summarize some well-known results concerning the solution of possibly underdetermined systems of linear equations.

For $N \in \mathbb{R}^{m\times n}$ a least squares solution x_{LS} of

$$Nx = b, \tag{2}$$

is a solution of the problem

$$\text{minimize } \|Nx-b\|_2 . \tag{3}$$

- For $\operatorname{rank} N = n$ the least squares solution x_{LS} is unique. Of course, if $b \in \operatorname{im} N$, then the least squares solution x_{LS} is the unique solution of $Nx = b$.
- For a rank deficit N, i.e., $\operatorname{rank} N = r < n$, there are an infinite number of solutions of (3).

  However, among all the least squares solutions, there is a unique so-called minimum-norm least squares solution x_{MNLS} such that

  $$\|x_{MNLS}\|_2 < \|x_{LS}\|_2$$

  for any other least squares solution x_{LS} of (2). In this case, if $b \in \operatorname{im} N$, the minimum-norm least squares solution x_{MNLS} coincides with the minimum-norm solution x_{MN} of $Nx = b$.

The minimum-norm least squares (MNLS) solution can be described and computed be means of the following matrix decomposition.

Definition 1. *(cf. [12]) For $G \in \mathbb{R}^{m\times n}$, $r = \operatorname{rank} G$, the singular value decomposition (SVD) reads*

$$G = U\Sigma V^T$$

for orthogonal matrices $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, and the diagonal matrix

$$\Sigma = \operatorname{diag}(\sigma_1,\ldots,\sigma_r,0,\ldots,0) =: \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}$$

for the positive singular values $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$, $r := \operatorname{rank} G \leq \min\{m,n\}$.

For later considerations we introduce the square diagonal matrices

$$E_m := \begin{pmatrix} 0 & \\ & I_{m-r} \end{pmatrix} \in \mathbb{R}^{m\times m}, \quad E_n := \begin{pmatrix} 0 & \\ & I_{n-r} \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{4}$$

$$\Sigma^+ := \operatorname{diag}(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_r},0,\ldots,0) = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m\times n} \tag{5}$$

and recall that:

- The unique orthogonal projector onto $\ker G$, i.e., a square matrix $Q \in \mathbb{R}^{n\times n}$ with the properties

$$GQ = 0, \quad Q^2 = Q, \quad \operatorname{rank} Q = n - \operatorname{rank} G = n - r, \quad Q = Q^T,$$

can be described by

$$Q = VE_nV^T = V(:, r+1:n)\cdot(V(:, r+1:n))^T$$

using MATLAB-notation for the last expression. Note, for $P := I - Q$

$$\ker Q = \ker\left(V(:, r+1:n)\right)^T, \quad \ker P = \ker\left(V(:, 1:r)\right)^T \tag{6}$$

is given by definition.

- The unique orthogonal projector along $\operatorname{im} G$, i.e., a square matrix $W \in \mathbb{R}^{m\times m}$ with the properties

$$WG = 0, \quad W^2 = W, \quad \operatorname{rank} W = m - \operatorname{rank} G = m - r, \quad W = W^T,$$

can be described by

$$W = UE_mU^T = U(:, r+1:m)\cdot(U(:, r+1:m))^T,$$

where

$$\ker W = \ker\left(U(:, r+1:m)\right)^T. \tag{7}$$

Theorem 1. *(cf. [12]) For $N \in \mathbb{R}^{m\times n}$ and the SVD $N = U\Sigma V^T$ the minimum-norm least squares solution to the system of linear equations*

$$Nx = b$$

is unique and given by

$$x_{MNLS} = V\Sigma^+U^Tb. \tag{8}$$

Note that for the orthogonal projector Q_N onto $\ker N$ we have by construction

$$\ker\begin{pmatrix} Q_N \\ N \end{pmatrix} = \{0\}.$$

Consequently, the minimum-norm least squares solution x_{MNLS} is the least squares solution of the system

$$\begin{pmatrix} Q_N \\ N \end{pmatrix} x = \begin{pmatrix} 0 \\ b \end{pmatrix}.$$

This follows directly from

$$Q_N x_{MNSQ} = V \begin{pmatrix} 0 & 0 \\ 0 & I_{n-r} \end{pmatrix} \underbrace{V^T V}_{=I} \Sigma^+ U^T b = V \begin{pmatrix} 0 & 0 \\ 0 & I_{n-r} \end{pmatrix} \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^T b = 0.$$

We want to emphasize that setting $Q_N x = 0$ is only one possibility we have to fix d free components if d corresponds to the so-called degree of freedom $d := n - r =$ nullity $N =$ rank Q_N.

In some cases we are not interested in a minimum-norm least squares solution x_{MNLS} with

$$\|x_{MNLS}\|_2 < \|x_{LS}\|_2$$

for any other least squares solution x_{LS}, but in a least squares solution $x_{MNLS\alpha}$ with the property

$$\|x_{MNLS\alpha}\|_2 < \|x_{LS} - \alpha\|_2$$

for a given α. In order to obtain a solution of this problem we can simply perform the substitution

$$\hat{x} := x - \alpha, \quad \hat{b} := b - N\alpha,$$

compute the unique minimum-norm least squares solution $\hat{x}_{MNLS}$ of

$$N\hat{x} = \hat{b}$$

and set

$$x_{MNLS\alpha} = \alpha + \hat{x}_{MNLS}.$$

In fact, $\hat{x}_{MNLS}$ can be interpreted as a minimum-norm correction of α. In terms of Q_N the solution $x_{MNLS\alpha} := \alpha + \hat{x}_{MNLS}$ can also be described as the unique least squares solution of

$$\begin{pmatrix} Q_N \\ N \end{pmatrix} x = \begin{pmatrix} Q_N \alpha \\ b \end{pmatrix}.$$

Note that for $b \in \operatorname{im} N$ the solution $x_{MNLS\alpha}$ corresponds to the solution of (2) that is closest to α in terms of the quadratic norm. For this closest solution to α we will use the notation $x_{MN\alpha}$.

3 Projector based specification of components

If we do not fix free components with $Q_N x = Q_N \alpha$, then we have different possibilities. Here, we will focus on the case that for $N \in \mathbb{R}^{m \times n}$, rank $N < n$ and an arbitrary given pair of projectors $P, Q \in \mathbb{R}^{n \times n}$ with $I = P + Q$ we have

$$\ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\}. \tag{9}$$

In fact, we are interested in the case that only a part of Px should be fixed, while the other part and Qx should result appropriately. Notice that the notation P and Q is motivated by the application for DAEs from Section 4 and should not be confused with the projector Q_N from Section 2. For our purposes, we first provide some technical results.

Lemma 1. *For a pair of complementary orthogonal projectors $P,\ Q \in \mathbb{R}^{n\times n}$, $P+Q=I$, $P=P^T$, $Q=Q^T$ and an arbitrary matrix $N \in \mathbb{R}^{m\times n}$ it holds*

$$\ker \begin{pmatrix} P \\ NQ \end{pmatrix} = \ker \begin{pmatrix} P \\ N \end{pmatrix}.$$

Consequently, if

$$\ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\}, \tag{10}$$

is given, then $\operatorname{rank} NQ = \operatorname{rank} Q$ *and for a projector W along* $\operatorname{im} NQ$ *it holds that* $\operatorname{rank} W = \operatorname{rank} P$.

Proof. The first assertion follows from the definition

$$\begin{aligned} \ker \begin{pmatrix} P \\ NQ \end{pmatrix} &= \{x \in \mathbb{R}^n \,:\, Px = 0,\ NQx = 0\}, \\ &= \{x \in \mathbb{R}^n \,:\, x = Qx,\ Nx = 0\} = \ker \begin{pmatrix} P \\ N \end{pmatrix}. \end{aligned}$$

The second assertion follows from $\operatorname{rank} NQ \le \min\{\operatorname{rank} N, \operatorname{rank} Q\}$ and, together with (10), also $\operatorname{rank} P + \operatorname{rank} NQ \ge n$ and $\operatorname{rank} P + \operatorname{rank} Q = n$. The latter assertion follows from

$$\operatorname{rank} W = \operatorname{nullity} NQ = \operatorname{nullity} Q = \operatorname{rank} P.$$

□

We will now address the fact that, instead of $Nx = b$, we may also consider

$$y_1 = Px, \quad y_2 = Qx, \quad NPy_1 + NQy_2 = b$$

or, equivalently,

$$\begin{pmatrix} Q & \\ NP & NQ \\ & P \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix},$$

and $x = y_1 + y_2$.

Lemma 2. *For an arbitrary matrix $N \in \mathbb{R}^{m\times n}$ and complementary projectors Q, $P := I - Q \in \mathbb{R}^{n\times n}$*

$$\text{nullity}\, N = \text{nullity} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}$$

is given.

Proof.

$$\text{rank} \begin{pmatrix} Q & P \\ P & Q \end{pmatrix} = 2n \quad \text{and} \quad \ker \begin{pmatrix} Q & P \\ NP & NQ \end{pmatrix} = \ker \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix},$$

implies

$$\begin{aligned} \text{nullity}\, N &= \text{nullity} \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix} = \text{nullity} \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} Q & P \\ P & Q \end{pmatrix} \\ &= \text{nullity} \begin{pmatrix} Q & P \\ NP & NQ \end{pmatrix} = \text{nullity} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}. \end{aligned}$$

□

These two lemmata permit now the characterization of the projector Π with the properties we will prove in the Theorems 2 and 3.

Theorem 2. *Consider an arbitrary matrix $N \in \mathbb{R}^{m\times n}$, complementary projectors Q, $P := I - Q \in \mathbb{R}^{n\times n}$, a projector W along* im NQ*, and the matrix*

$$M = \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix} \in \mathbb{R}^{(2n+m)\times 2n}.$$

For all projectors $Q_M \in \mathbb{R}^{2n\times 2n}$ onto ker M *and all projectors $\Pi \in \mathbb{R}^{n\times n}$ onto*

$$\ker \begin{pmatrix} Q \\ WNP \end{pmatrix} = \text{im}\, \Pi$$

the first n rows of Q_M, i.e., the matrix $H \in \mathbb{R}^{n\times 2n}$ defined by

$$Q_M =: \begin{pmatrix} H \\ * \end{pmatrix},$$

fulfill

$$\text{im}\, \Pi = \text{im}\, H.$$

Moreover, if P and Π are orthogonal, then $P\Pi = \Pi P = \Pi$.

This theorem is a special case of Theorem 1 from [11]. As it appears in a different context and also with a different notation there, we include an adapted proof for completeness here.

Proof. 1. "$\subseteq$". From the definition of W it follows that

$$\operatorname{im}(I-W)NP \subset \operatorname{im} NQ.$$

Consequently, for all

$$x \in \ker \begin{pmatrix} Q \\ WNP \end{pmatrix} = \operatorname{im} \Pi$$

there is a $z \in \mathbb{R}^n$, $z = Qz \in \ker Q$ such that

$$-(I-W)NPx = NQz \quad \text{and} \quad Pz = 0,$$

and with $WNPx = 0$

$$\begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = 0, \quad \text{i.e.,} \quad \begin{pmatrix} x \\ z \end{pmatrix} = Q_M \begin{pmatrix} x \\ z \end{pmatrix}.$$

Therefore $x \in \operatorname{im} H$.

2. "$\supseteq$". For $x \in \operatorname{im} H$ there is a $y \in \mathbb{R}^{2n}$ such that $x = Hy$. Consequently,

$$\begin{aligned} \begin{pmatrix} Q \\ WNP \end{pmatrix} x &= \begin{pmatrix} Q \\ WNP \end{pmatrix} Hy = \begin{pmatrix} I & 0 & 0 \\ 0 & W & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ NP & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} Hy \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} I & 0 & 0 \\ 0 & W & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ NP & 0 \\ 0 & 0 \end{pmatrix} Q_M y \end{aligned}$$

is given and, taking advantage of $WNQ = 0$ and $0 \cdot P = 0$, also

$$\begin{pmatrix} Q \\ WNP \end{pmatrix} x = \begin{pmatrix} I & 0 & 0 \\ 0 & W & 0 \end{pmatrix} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix} Q_M y = 0,$$

i.e., $x \in \operatorname{im} \Pi$.

3. The property $P\Pi = \Pi$ is obvious. If P and Π are orthogonal, then

$$\Pi P = \Pi^T \cdot P^T = (P\Pi)^T = \Pi^T = \Pi.$$

□

Theorem 3. *Suppose that an arbitrary matrix $N \in \mathbb{R}^{m\times n}$ and complementary projectors Q, $P := I - Q \in \mathbb{R}^{n\times n}$ fulfilling*

$$\ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\}$$

are given, and that W is an arbitrary projector along $\operatorname{im} NQ$*. Then all projectors* Π *onto*

$$\ker \begin{pmatrix} Q \\ WNP \end{pmatrix} = \ker Q \cap \ker WNP$$

fulfill

$$\operatorname{nullity} N = \operatorname{nullity} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix} = \operatorname{rank} \Pi$$

and

$$\ker \begin{pmatrix} \Pi \\ N \end{pmatrix} = \{0\}.$$

Proof. 1. On the one hand, from Theorem 2 it follows that for a projector

$$\tilde{Q} = \begin{pmatrix} H \\ * \end{pmatrix} \quad \text{with} \quad \operatorname{im} \begin{pmatrix} H \\ * \end{pmatrix} = \ker \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}$$

it holds that $\operatorname{im} \Pi = \operatorname{im} H$, and consequently

$$\operatorname{rank} \Pi = \operatorname{rank} H \leq \operatorname{rank} \begin{pmatrix} H \\ * \end{pmatrix} = \operatorname{nullity} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}.$$

2. On the other hand, for

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \ker \begin{pmatrix} \Pi & 0 \\ Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}$$

we obviously obtain

$$\begin{pmatrix} \Pi \\ Q \\ WNP \end{pmatrix} y_1 = 0, \text{ i.e., } y_1 = 0,$$

and consequently

$$\begin{pmatrix} NQ \\ P \end{pmatrix} y_2 = 0.$$

From

$$\ker \begin{pmatrix} NQ \\ P \end{pmatrix} = \ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\},$$

we derive $y_2 = 0$. This particularly implies

$$\operatorname{rank} \Pi \geq \text{nullity} \begin{pmatrix} Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix}.$$

3. The relation nullity $N = \operatorname{rank} \Pi$ now follows from Lemma 2.

4. Let us finally consider $x \in \ker \Pi \cap \ker N$ and define $y_1 := Px$, $y_2 = Qx$. By definition we have

$$Qy_1 = 0 \quad , Py_2 = 0 \quad \text{and} \quad NPy_1 + NQy_2 = 0.$$

Consequently,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \ker \begin{pmatrix} \Pi & 0 \\ Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix},$$

such that $y_1 = y_2 = 0$ results as shown above in step 2, and $x = y_1 + y_2 = 0$ has to be given.

□

After we have clarified the properties of Π, we can use this projector to fix the solution of underdetermined systems of equations we are interested in.

Theorem 4. *Consider $N \in \mathbb{R}^{m \times n}$ and a pair of orthogonal projectors Q, P with $P = I - Q$ and*

$$\ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\},$$

an arbitrary projector W along $\operatorname{im} NQ$ *and the orthogonal projector* Π *onto*

$$\ker \begin{pmatrix} Q \\ WNP \end{pmatrix} = \ker Q \cap \ker WNP.$$

(a) For $b \in \operatorname{im} N$, the unique solution $x_{\Pi\alpha}$ of the system

$$\begin{pmatrix} \Pi \\ N \end{pmatrix} x = \begin{pmatrix} \Pi\alpha \\ b \end{pmatrix} \tag{11}$$

fulfills

$$\|P(x_{\Pi\alpha} - \alpha)\|_2 < \|P(x - \alpha)\|_2$$

for any other solution x of $Nx = b$.

(b) For $b \notin \operatorname{im} N$, the unique least squares solution $x_{LS\Pi\alpha}$ of the system

$$\begin{pmatrix} \Pi \\ N \end{pmatrix} x = \begin{pmatrix} \Pi\alpha \\ b \end{pmatrix} \tag{12}$$

fulfills

$$\|P(x_{LS\Pi\alpha} - \alpha)\|_2 < \|P(x_{LS} - \alpha)\|_2 \tag{13}$$

for any other least squares solution x_{LS} of $Nx = b$.

Proof. (a) Let us first consider the case $b \in \operatorname{im} N$. According to Theorem 3, the system (11) has a unique solution. With the unique solution of the system

$$\begin{pmatrix} \Pi & 0 \\ Q & 0 \\ NP & NQ \\ 0 & P \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \Pi\alpha \\ 0 \\ b \\ 0 \end{pmatrix}$$

we can represent the unique solution of (11) by $x_{\Pi\alpha} = y_1 + y_2$. Moreover, y_1 can be interpreted as the closest solution to α of

$$\begin{pmatrix} Q \\ WNP \end{pmatrix} z = \begin{pmatrix} 0 \\ Wb \end{pmatrix},$$

i.e., $y_1 = z_{MN\alpha}$, and thus

$$Px_{\Pi\alpha} = Py_1 = y_1 = z_{MN\alpha}.$$

Thus, with $\Pi P = \Pi$ it follows

$$\Pi(x_{\Pi\alpha} - \alpha) = \Pi(y_1 - \alpha) = \Pi(z_{MN\alpha} - \alpha) = 0.$$

Any other solution x of (11) with $\Pi(x - \alpha) \neq 0$ would then fulfill

$$\|P(x - \alpha)\|_2 > \|P(z_{MN\alpha} - \alpha)\|_2 = \|P(x_{LS\Pi\alpha} - \alpha)\|_2 .$$

(b) For $b \neq \operatorname{im} N$ and the system (12) the argumentation is analogous, considering the unique least squares solution in each step. Moreover, for $N = U\Sigma V^T$ the minimum-norm least squares solution of $Nx = b$ fulfills

$$x_{MNLS} = V\Sigma^+ U^T b.$$

Note that for the orthogonal projector Q_N onto $\ker N$ and $P_N := I - Q_N$ this x_{MNLS} can also be interpreted as the unique minimum-norm solution of

$$P_N x = P_N V\Sigma^+ U^T b.$$

Since $P_N V\Sigma^+ U^T b \in \operatorname{im} P_N$, we can consider

$$\begin{pmatrix} \Pi \\ P_N \end{pmatrix} \tilde{x} = \begin{pmatrix} \Pi\alpha \\ P_N V\Sigma^+ U^T b \end{pmatrix} \tag{14}$$

and apply the result from above. The resulting unique solution $\tilde{x}_{\Pi\alpha}$ of (14) is a least squares solution of $Nx = q$ such that $x_{LS\Pi\alpha} := \tilde{x}_{\Pi\alpha}$ fulfills (13) by construction. □

We illustrate the difference between the solutions we obtain fixing the free components by Q_N and Π by a simple example.

Example 1. *If we consider*

$$x_1 + x_2 = b_1, \tag{15}$$
$$x_1 + 2x_3 = b_2, \tag{16}$$

and

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we directly obtain

$$N = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}, \quad Q_N = \begin{pmatrix} \frac{4}{9} & -\frac{4}{9} & -\frac{2}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{2}{9} \\ -\frac{2}{9} & \frac{2}{9} & \frac{1}{9} \end{pmatrix},$$

$$NQ = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} Q \\ WNP \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Pi = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The properties $\Pi P = P\Pi = \Pi$, $\operatorname{rank}\Pi = \operatorname{rank} Q_N$*=1 are easy to verify. Let us consider* $b_1 = 4$, $b_2 = 5$ *and* $\alpha = (1,2,3)$ *for instance. The unique solution of*

$$\begin{aligned} Nx &= \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}, \\ Q_N(x-\alpha) &= 0, \end{aligned} \tag{17}$$

reads

$$\begin{pmatrix} \frac{11}{9} \\ \frac{25}{9} \\ \frac{17}{9} \end{pmatrix}.$$

In contrast, the solution of

$$\begin{aligned} Nx &= \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \\ \Pi(x-\alpha) &= 0, \end{aligned} \tag{18}$$

reads

$$\begin{pmatrix} \frac{3}{2} \\ \frac{5}{2} \\ \frac{7}{4} \end{pmatrix}.$$

For $b_1 = 4$, $b_2 = 5$ *and* $\alpha = (0,0,0)$ *we obtain*

$$\begin{pmatrix} \frac{7}{3} \\ \frac{5}{3} \\ \frac{4}{3} \end{pmatrix} \quad \textit{and} \quad \begin{pmatrix} 2 \\ 2 \\ \frac{3}{2} \end{pmatrix}$$

for (17) *and* (18)*, respectively.*

We emphasize that the prescription $\Pi(x-\alpha) = 0$ does not involve Qx due to $\Pi P = P\Pi = P$. With regard to DAEs, where Qx stands for the undifferentiated components, this is the aspect we are interested in.

4 Consistent initial values for linear DAEs

All index concepts from literature give a kind of measure of the deviation of a DAE from regular ODEs. Here, we will focus on the differentiation index only.

Definition 2. *(see e.g., [1]) The differentiation index μ of the linear DAE*

$$A(t)x' + B(t)x = q(t) \tag{19}$$

with sufficiently smooth coefficients and right-hand side is the smallest integer μ such that the derivative array

$$\begin{aligned} Ax' + Bx - q &= 0, && (20)\\ \frac{d}{dt}(Ax' + Bx - q) &= 0, && (21)\\ &\vdots\\ \frac{d^j}{dt^j}(Ax' + Bx - q) &= 0, && (22)\\ &\vdots\\ \frac{d^\mu}{dt^\mu}(Ax' + Bx - q) &= 0, && (23) \end{aligned}$$

uniquely determines x' as a continuous linear function of (x,t).

For a projector Q_0 onto $\ker A$, $P_0 := I - Q_0$ we will denote the differentiated component by P_0x and the undifferentiated component of x by Q_0x . In [11] the following characterization of the differentiation index was introduced.

Definition 3 ([11])**.** *Let us define*

$$G_{Lj} := \begin{pmatrix} B \\ B' \\ \vdots \\ B^{(j)} \end{pmatrix}, \quad G_{Rj} := \begin{pmatrix} A & & & \\ B + A' & A & & \\ \vdots & \ddots & \ddots & \\ jB^{(j-1)} + A^{(j)} & & B + jA' & A \end{pmatrix},$$

$G_{Lj} \in \mathbb{R}^{(j+1)\cdot n \times n}$, $G_{Rj} \in \mathbb{R}^{(j+1)\cdot n \times (j+1)\cdot n}$ *(note that L and R stand for left and right) as well as a projector Q_0 onto* $\ker A$*, $P_0 := I - Q_0$ and a projector W_{Rj} along* $\operatorname{im} G_{Rj}$*.*

If there is an integer μ such that

$$\ker \begin{pmatrix} P_0(t) \\ W_{R\mu-1}(t)G_{L\mu-1}(t) \end{pmatrix} = \{0\}, \tag{24}$$

$\operatorname{rank} G_{Rj}(t)$ *as well as*

$$r_j(t) := \operatorname{rank} \begin{pmatrix} P_0 \\ W_{Rj}G_{Lj} \end{pmatrix}$$

are constant for $j = 0, \ldots, \mu - 1$, and $r_{j-1}(t) \le r_j(t) < r_{\mu-1}(t) = n$ for $j = 1, \ldots, \mu - 2$ on an open interval $\mathscr{I}$, we will use the notation

$$\operatorname{ind}_{\mathrm{D}}(A(t), B(t)) = \mu, \quad \textit{for} \quad t \in \mathscr{I}.$$

The subscript D has been chosen to emphasize the relationship to the derivatives of Definition 2. If the index $\mathrm{ind}_{\mathrm{D}}(A(t),B(t))$ is μ, then all hidden and explicit constraints can be described by means of

$$W_{R\mu-1}(t)G_{L\mu-1}(t)x = W_{R\mu-1}(t)\begin{pmatrix} q(t) \\ q'(t) \\ \vdots \\ q^{\mu-1} \end{pmatrix} =: W_{R\mu-1}(t)\tilde{q}(t). \tag{25}$$

In fact, if the nullspace (24) is trivial for $j = \mu - 1$, then these constraints uniquely determine Q_0x as a function of P_0x and t, and an explicit representation of Q_0x can be obtained from

$$\begin{pmatrix} P_0 \\ W_{R\mu-1}G_{L\mu-1} \end{pmatrix} Q_0x = \begin{pmatrix} 0 \\ W_{R\mu-1}\tilde{q} - W_{R\mu-1}G_{L\mu-1}P_0x \end{pmatrix}.$$

Differentiating the resulting expression we can represent $(Q_0x)'$ as a function of (x,t) (substituting $(P_0x)'$, cf. [11]) such that it becomes clear that the differentiation index is μ.

Definition 4. *A vector $x_0 \in \mathbb{R}^n$ is a consistent initial value of* (19) *if there exists a solution of* (19) *that fulfills $x(t_0) = x_0$.*

A consistent initial value has to fulfill the constraints (25). With respect to the results presented in Section 3, we emphasize that for $P = P_0$, $N = W_{R\mu-1}G_{L\mu-1}$ condition (24) precisely corresponds to condition (9) such that particularly Theorem 4 can be applied for orthogonal projectors.

Consequently, we now consider a decoupling of the constraints (25). Using a projector W_{RLQ} along $\operatorname{im} W_{R\mu-1}G_{L\mu-1}Q_0$ we obtain the constraints that restrict P_0x:

$$W_{RLQ}W_{R\mu-1}G_{L\mu-1}x = W_{CP}W_{R\mu-1}G_{L\mu-1}P_0x = W_{CP}W_{R\mu-1}\tilde{q}.$$

For an orthogonal Q_0 and the orthogonal projector Π fulfilling

$$\ker \begin{pmatrix} Q_0 \\ W_{RLQ}W_{R\mu-1}G_{L\mu-1} \end{pmatrix} = \operatorname{im} \Pi, \tag{26}$$

Πx describes components of x for which we can prescribe initial values. In fact, it corresponds to differentiated components that are not determined by inherent differentiation. Of course, on principle, such prescription is not unique and different alternatives may be chosen. Nevertheless, Π results to be an adequate orthogonal description.

Theorem 5. *Suppose that a vector $\alpha \in \mathbb{R}^n$ is a user-given guess for an initial value. If no further restrictions are given, then a consistent initial value may be computed solving the system*

$$\Pi(x-\alpha) = 0, \tag{27}$$

$$W_{R\mu-1}G_{L\mu-1}x = W_{R\mu-1}\begin{pmatrix} q \\ \vdots \\ q^{(\mu-1)} \end{pmatrix}. \tag{28}$$

The unique solution $x_{\Pi\alpha}$ fulfills

$$\|P_0(x_{\Pi\alpha}-\alpha)\|_2 < \|P_0(x_0-\alpha)\|_2$$

for any other consistent initial value x_0.

This particular consistent initial value $x_{\Pi\alpha}$ may also be computed solving the system

$$\Pi(z_0-\alpha) = 0, \tag{29}$$

$$\begin{pmatrix} G_{L\mu-1} & G_{R\mu-1} \end{pmatrix}\begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_\mu \end{pmatrix} = \begin{pmatrix} q \\ q' \\ \vdots \\ q^{(\mu-1)} \end{pmatrix} \tag{30}$$

in a minimum-norm sense for $z = (z_0, \ldots, z_\mu)$ and setting $x_{\Pi\alpha}$ equal to the first n components of $z_{MN\alpha}$.

Proof. The first result follows directly from Theorem 4. Note that we assumed $b \in \operatorname{im} N$ since this has to be given for solvable DAEs. In case that the expressions for the constraints are not accurately given, the minimum-norm least squares solution may be obtained.

For the second representation of $x_{\Pi\alpha}$ we merely have to recall that the constraints (28) are implicitly included in the system (30) such that the minimum-norm requirement does not apply to z_0. □

Although we will use the representation (28) for the constraints here, other descriptions may be also considered. In case that the constraints are correctly identified by, for example, a structural analysis (cf., e.g., [18], [19]), the projector based specification of initial values is also possible. In fact, condition (9) could also be used to check whether structural analysis has identified sufficient constraints.

To illustrate the approach we start with a simple index-2 DAE that leads to constraints with the same structure as Example 1.

Example 2. *Let us consider the DAE*

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

leading to

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

$P_0 = A$, $Q_0 = I - P_0$. *The index is not one since for* $G_{L0} = B$, $G_{R0} = A$ *and* $W_{R0} = Q_0$ *we obtain*

$$\ker \begin{pmatrix} P_0 \\ W_{R0} G_{L0} \end{pmatrix} = \ker \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hdashline 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 \\ 1 & 1 & 0 \end{array}\right) \neq \{0\}.$$

In fact, the index is $\mu = 2$ *since for*

$$\left(\begin{array}{c|c} G_{L1} & G_{R1} \end{array} \right) = \left(\begin{array}{ccc:cccccc} 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \end{array}\right)$$

we may choose the simple projector

$$W_{R1} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 \end{pmatrix},$$

leading to

$$\ker \begin{pmatrix} P_0 \\ W_{R1} G_{L1} \end{pmatrix} = \ker \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & -2 \end{array} \right) = \{0\}, \quad W_{R1} G_{L1} \tilde{q} = \begin{pmatrix} 0 \\ 0 \\ q_3 \\ 0 \\ 0 \\ -q_1 - q_2 + q_3' \end{pmatrix}.$$

Since the sign of the last row does not affect the nullspace, we obtain the same projector Π *as in Example 1. In fact, the constraints could be rewritten as*

$$Nx = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} x = \begin{pmatrix} q_3 \\ q_1 + q_2 - q_3' \end{pmatrix} =: b$$

such that for corresponding values of α *and* b *we would obtain the consistent values from Example 1.*

Note that with regard to an implementation it suffices to consider suitable basis instead of complete projectors P_0, Π, W_{Rj}, such that a compact formulation of the constraints is obtained directly, cf. Section 8.

5 Handling of user-given specifications

In case that the user wants to fix some specific values $\alpha_{k1}, \ldots, \alpha_{k\nu}$, we may proceed as follows. If e_{k1} denotes the k_1-th unit vector of the standard basis, we check whether

$$\text{nullity} \begin{pmatrix} W_{R\mu-1} G_{L\mu-1} \\ e_{k_1}^T \end{pmatrix} = \text{nullity} \left(W_{R\mu-1} G_{L\mu-1} \right) - 1.$$

If this condition is not fulfilled, the prescription of α_{k1} is not admissible and cannot be considered. Of course, user-given prescriptions may also be given in a more general form

$$U(x_0 - \alpha) = 0$$

for $U \in R^{\nu \times n}$, rank $U = \nu$, if

$$\text{nullity} \begin{pmatrix} W_{R\mu-1} G_{L\mu-1} \\ U \end{pmatrix} = \text{nullity} \left(W_{R\mu-1} G_{L\mu-1} \right) - \nu$$

is given. Supposing that such admissible prescriptions are given, then, for the orthogonal projector Π_U with

$$\ker \begin{pmatrix} Q_0 \\ W_{R\mu-1}G_{L\mu-1} \\ U \end{pmatrix} = \operatorname{im} \Pi_U$$

the system

$$\begin{aligned} U(x_0-\alpha) &= 0, && (31)\\ \Pi_U(x_0-\alpha) &= 0, && (32)\\ W_{R\mu-1}G_{L\mu-1}x_0 &= W_{R\mu-1}\begin{pmatrix} q \\ \vdots \\ q^{(\mu-1)} \end{pmatrix} && (33) \end{aligned}$$

can be solved to compute a consistent initial value.

6 The computation of consistent Taylor series

Solving

$$\begin{aligned} \Pi(z_0-\alpha) &= 0, && (34)\\ \begin{pmatrix} G_{Lj} & G_{Rj} \end{pmatrix}\begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_{j+1} \end{pmatrix} &= \begin{pmatrix} q \\ q' \\ \vdots \\ q^{(j)} \end{pmatrix} && (35) \end{aligned}$$

for $j \geq \mu - 1$ in a minimum-norm sense permits the computation of consistent values for $(x(t_0), x'(t_0), \ldots, x^{(j-\mu+1)}(t_0))$. This aspect is of special interest with regard to the integration of DAEs using Taylor series methods, cf. [10], [9].

7 Consistent initialization for nonlinear DAEs

For nonlinear DAEs our approach leads to an iterative method based on linearization. If we consider DAEs of the form

$$f(x', x, t) = 0 \tag{36}$$

and

$$\begin{aligned} A(z_1, z_0, t) &= f_{z_1}(z_1, z_0, t) \in \mathbb{R}^{n\times n}, && (37)\\ B(z_1, z_0, t) &= f_{z_0}(z_1, z_0, t) \in \mathbb{R}^{n\times n}, \end{aligned}$$

the matrices G_L, G_R and the orthogonal projector Q_0, P_0, W_R, Π may be computed considering the linearization. With regard to the computation of consistent initial values discussed above, one important difference with respect to linear DAEs is that now the matrices G_L, G_R and the projectors P_0, Q_0, W_R, Π themselves may depend on z_0 and z_1 and possibly z_j for corresponding higher j-th derivatives. Nevertheless, the following algorithm results to be successful in many relevant cases, cf. Section 10.

1. Suppose that α and an approximation $z := (z_0, z_1, \ldots)$ for $(x(t_0), x'(t_0), \ldots)$ is given. If no better guess is available, we may assume $z := (\alpha, 0, 0, \ldots)$.

2. Compute Π for the linearized DAE of index μ at z .

3. For
$$F_j(x^{(j+1)}, x^{(j)}, \ldots, x', x, t) := \frac{d^j}{dt^j} f(x', x, t)$$
solve
$$\begin{aligned} \Pi(z_0 - \alpha) &= 0, \\ f(z_1, z_0, t_0) &= 0, \\ F_1(z_2, z_1, z_0, t_0) &= 0, \\ &\vdots \\ F_{\mu-1}(z_\mu, \ldots, z_0, t_0) &= 0, \end{aligned}$$
in a minimum-norm least squares sense. When you obtain a better approximation z, go to Step 2 until the residual is sufficiently small.

4. Set $x_0 = z_0$.

However, all difficulties related to nonlinear DAEs may also appear in this context, cf. [14].

8 Implementation

A prototype has been implemented in Python, using the AD-package AlgoPy [20]. In this regard, we briefly state some practical remarks:

- According to the structure of the Taylor series AlgoPy works with, the system of equations is not set up for approximation $z := (z_0, z_1, z_2, \ldots)$ for $(x(t_0), x'(t_0), x''(t_0), \ldots)$, but for approximations of the Taylor coefficients $(x(t_0), x'(t_0), \frac{1}{2}x''(t_0), \ldots)$. The number of Taylor coefficients D has to fulfill at least $D \geq \mu$.

- Although we compute the expressions using automatic differentiation, in the end we set up a real matrix. All singular value decompositions are performed with the implementation from the python library *numpy*.
- For a compact formulation, instead of the projectors P_0, Π we set up the matrices considering the corresponding orthogonal basis, cf. (6).
- Analogously, instead of considering the projectors $W_{R_{\mu-1}}$, W_{RLQ} we use the corresponding orthogonal bases (7) to obtain a compact formulation. We will verify this in Lemma 3.
- In our first implementation we directly perform the iteration described in Section 7 to solve the nonlinear systems, for simplicity and illustrative reasons. In future versions more sophisticated solver or standard software should be used.

Lemma 3. *Let $B_\square := U_\square(:, r_\square + 1 : m) \in \mathbb{R}^{m\times m-r_\square}$ denote the matrix built by the corresponding orthogonal basis from* (7)*. For the matrices from Section 4 we consequently have*

$$W_{R_{\mu-1}} = B_R \cdot B_R^T, \quad W_{RLQ} = B_{RLQ} \cdot B_{RLQ}^T$$

for $m = \mu \cdot n$, and for the unique orthogonal projector W_{BRLQ} along

$$\operatorname{im} B_R^T G_{L_{\mu-1}} Q_0,$$

we obtain

$$W_{BRLQ} =: B_{BRLQ} \cdot B_{BRLQ}^T.$$

Using these notations, the relation

$$\ker W_{RLQ} W_{R\mu-1} G_{L\mu-1} P_0 \quad = \quad \ker B_{BRLQ}^T B_R^T G_{L\mu-1} P_0$$

holds true, i.e., in the implementation we can use bases instead of projectors.

Proof. We obviously have

$$\operatorname{rank} B_{BRLQ} = \operatorname{nullity} B_R^T G_{L_{\mu-1}} Q_0 = \operatorname{nullity} W_{R_{\mu-1}} G_{L_{\mu-1}} Q_0 = \operatorname{rank} B_{RLQ}$$

and

$$\ker W_{BRLQ} \subseteq \ker W_{RLQ} B_R, \quad \text{i.e.,} \quad \ker B_{BRLQ}^T \subseteq \ker B_{RLQ}^T B_R.$$

Let $x \in \ker B_{RLQ}^T B_R$ and $z := B_R x$. Then, by definition, there exists a $\tilde{z}$ such that

$$z = W_{R_{\mu-1}} G_{L_{\mu-1}} Q_0 \tilde{z}$$

end thus, with $B_R^T B_R = I$,

$$x = B_R^T z = B_R^T G_{L_{\mu-1}} Q_0 \tilde{z} \in \ker W_{BRLQ} = \ker B_{BRLQ},$$

i.e.,

$$\ker B_{RLQ}^T B_R \subseteq \ker B_{BRLQ}^T.$$

Thus, we obtain the required property

$$\begin{aligned} \ker W_{RLQ} W_{R\mu-1} G_{L\mu-1} P_0 &= \ker B_{RLQ}\, B_{RLQ}^T B_R B_R^T G_{L\mu-1} P_0 \\ &= \ker B_{RLQ}^T B_R B_R^T G_{L\mu-1} P_0 \\ &= \ker B_{BRLQ}^T B_R^T G_{L\mu-1} P_0. \end{aligned}$$

□

9 Comparison with previous approaches

There is a large body of literature related to the computation of consistent initial values for DAEs. For an overview of classical approaches we refer to [6]. Here, we focus on some particular aspects only. First we specify the differences to methods that are based on structural analysis. Afterwards, we discuss other methods that compute consistent initial values that are closest (in some sense) to a given α. Finally, the differences to our previous work are pointed out.

9.1 Structural analysis

Many algorithms to investigate the structure of DAEs are based on a structural analysis, cf., e.g., [18], [19]. Nevertheless, all structural analysis methods can fail for simple, solvable DAEs and may give incorrect structural information including the index. A well-known example is the robotic arm presented in Table 4, Section 10. Although, for many examples (including the mentioned robotic arm), there are remedies to reformulate the equations such that the structural analysis is successful, for new examples the reliability is not given. Moreover, since no numerical values are involved, singularities might not be detected. This is a crucial difference to our projector based investigation.

Apart from the index and constraints determination, the methods from [15], [16], [17] also differ considerably from our approach with respect to consistent initialization. In fact, their algorithms compute Taylor coefficients in a stage-wise manner, whereas the structural analysis indicates at each stage which equations are solved for which variables.

Many simulation packages like Modelica and Mathematica use a structural analysis to reformulate the equation and solve them by well-established DAE-solvers, which are suitable for lower-index DAEs. Also in this case, so-called state variables are identified.

Our approach is also considerably different in this aspect, since it is based on projections that do not necessarily correspond to particular variables or equations.

Example 3. *For the index-2 DAE from Example 2 we have*

$$P_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \Pi = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and the explicit constraint

$$x_1 + x_2 = q_3. \tag{38}$$

From our point of view, if no specifications like those from Section 5 are given, it is more advisable to fix

$$\Pi(x-\alpha) = \begin{pmatrix} \frac{1}{2}(x_1 - x_2) \\ -\frac{1}{2}(x_1 - x_2) \\ 0 \end{pmatrix} - \begin{pmatrix} \frac{1}{2}(\alpha_1 - \alpha_2) \\ -\frac{1}{2}(\alpha_1 - \alpha_2) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

i.e., in fact, to claim

$$x_1 - x_2 = \alpha_1 - \alpha_2,$$

than to set $x_1 = \alpha_1$ and compute x_2 afterwards or vice versa. Note that for this example, in all these three cases, the explicit constraint (38) *is used to compute the missing part of the differentiated component $P_0 x$, and the hidden constraint permits to calculate the suitable value for the undifferentiated component Q_0, i.e., x_3.*

9.2 Considering the deviation from a given α

Our method results to be closely related to the coordinate projection method for DAEs from multibody dynamics, cf. [4]. Since these DAEs are semi-explicit, $P_0 x$ corresponds precisely to the position p and the velocity v. For these components, the method performs an orthogonal projection of numerically obtained approximations onto the manifold described by the constraints. However, the minimization is sequential and in a specific order (first p and then v). An analogous specific order is also stated in [8], were higher-index DAEs are considered.

The method from [13] minimizes the deviation from a given α by a successive linear programming approach. However, this minimization does not consider the $\|.\|_2$-norm and is not formulated for $P_0 x$ in general. Here, we precisely exploit this property.

9.3 Differences to our previous work

The approaches from [6] and [7] suppose that α fulfills the explicit constraints and that the index is not greater than 2. In that case, $P_0(x-\alpha)=0$ results automatically such that the obtained values coincide with the ones we obtain here.

In contrast, for higher-index DAEs the comparison turned out to be more complicated. Although our projector Π and the projector $\Pi_{\mu-1}$ from [14] seem to be closely related to each other and fulfill $\operatorname{rank}\Pi = \operatorname{rank}\Pi_{\mu-1}$, $P_0\Pi = \Pi P_0 = \Pi$, $P_0\Pi_{\mu-1} = \Pi_{\mu-1}P_0 = \Pi_{\mu-1}$, we realized that they are not identical in general, not even for the simple pendulum with index 3.

With respect to the handling of Taylor series using AD we also want to emphasize a significant difference to our recent work [10], [9]. On the one hand, there we solved systems of Taylor equations using a Newton-Kantorowitsch method. Here, in contrast, we flattened the equations such that a standard nonlinear system of real equations was considered. On the other hand, in [10], [9] we used the decoupling related to the tractability index and solved sequentially the equations for the different components, starting with the inherent ODE. In the present article, we consider projections only to fix suitable free components and solve the equation for all components simultaneously.

10 Examples

Example 4. *For illustrative reasons, we start with a simple example in Kronecker canonical form with index 4 and an obvious solution:*

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} x' + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sin(t) \end{pmatrix}, \quad x = \begin{pmatrix} C\mathrm{e}^{-\mathrm{t}} \\ \cos(t) \\ -\sin(t) \\ -\cos(t) \\ \sin(t) \end{pmatrix}.$$

In this particular case, we obtain $\Pi = \operatorname{diag}(1,0,0,0,0)$*, since* $x_1' - x_1 = 0$ *corresponds to the inherent ODE. Concerning the values obtained for the Taylor series, we see in Table 1 that, corresponding to the description from [11], the last three coefficients are not correct for* x_2*, and so are the last two for* x_3 *and the last one for* x_4*. In contrast, for* x_1 *and* x_5 *all coefficients are exact up to numerical accuracy. For unstructured DAEs, these properties will apply to components that, in general, do not correspond to a particular variable* x_j*, but to a linear combination of several variables.*

For nonlinear DAEs, the projector Π depends on the solution in general.

	x_1	x_2	x_3	x_4	x_5
$x_*(\frac{\pi}{4})$	1.00000000e+00	7.07106781e-01	-7.07106781e-01	-7.07106781e-01	7.07106781e-01
$x'_*(\frac{\pi}{4})$	-1.00000000e+00	-7.07106781e-01	-7.07106781e-01	7.07106781e-01	7.07106781e-01
$\frac{1}{2}x''_*(\frac{\pi}{4})$	5.00000000e-01	-3.53553391e-01	3.53553391e-01	3.53553391e-01	-3.53553391e-01
$\frac{1}{3!}x'''_*(\frac{\pi}{4})$	-1.66666667e-01	-5.77954590e-16	1.17851130e-01	-1.17851130e-01	-1.17851130e-01
$\frac{1}{4!}x^{(iv)}_*(\frac{\pi}{4})$	4.16666667e-02	2.21294774e-15	1.44488648e-16	-2.94627825e-02	2.94627825e-02
$\frac{1}{5!}x^{(v)}_*(\frac{\pi}{4})$	-8.33333333e-03	1.32853064e-15	-4.42589549e-16	-2.88977295e-17	5.89255651e-03

Table 1: Solution of the system (29)-(30) from Example 4 for $t_0 = \pi/4$ and $\alpha = [1,0,0,0,0]$ using Taylor series with $D = 6$. The framed values are obviously not consistent, since the Taylor series are correct up to the third coefficients ($\mu = 4$, $D - \mu + 1 = 3$).

Example 5. *Considering the DAE corresponding to the (normalized) pendulum*

$$\begin{aligned} x'_1 &= x_3, \\ x'_2 &= x_4, \\ x'_3 &= x_1 x_5, \\ x'_4 &= x_2 x_5 - 1, \\ x_1^2 + x_2^2 &= 1, \end{aligned}$$

for the value x_0 from Table 2 we obtained the projector Π:

$$\begin{pmatrix} 5.00000000e-01 & -5.00000000e-01 & -3.60822483e-16 & -4.99600361e-16 & 6.06058204e-16 \\ -5.00000000e-01 & 5.00000000e-01 & 4.44089210e-16 & 4.16333634e-16 & -6.06058204e-16 \\ -3.60822483e-16 & 4.44089210e-16 & 5.00000000e-01 & -5.00000000e-01 & -9.04790914e-17 \\ -4.99600361e-16 & 4.16333634e-16 & -5.00000000e-01 & 5.00000000e-01 & 9.04790914e-17 \\ 6.06058204e-16 & -6.06058204e-16 & -9.04790914e-17 & 9.04790914e-17 & 7.50986026e-31 \end{pmatrix},$$

and, in contrast, for x_0

$$[0.4472136 \quad 0.89442719 \quad 0.4 \quad -0.2 \quad 0.69442719]$$

the projector Π

$$\begin{pmatrix} 6.66666667e-01 & -3.33333333e-01 & -1.49071198e-01 & -2.98142397e-01 & -3.33923253e-16 \\ -3.33333333e-01 & 1.66666667e-01 & 7.45355992e-02 & 1.49071198e-01 & 1.66961627e-16 \\ -1.49071198e-01 & 7.45355992e-02 & 8.33333333e-01 & -3.33333333e-01 & 1.65829285e-16 \\ -2.98142397e-01 & 1.49071198e-01 & -3.33333333e-01 & 3.33333333e-01 & 1.03754131e-16 \\ -3.33923253e-16 & 1.66961627e-16 & 1.65829285e-16 & 1.03754131e-16 & 1.77645195e-31 \end{pmatrix}.$$

In the tables we summarize the results we obtained for some well-known challenging examples from literature. The following details are provided:

- Number n of equations (and variables), index of the DAE. The index is known from the literature and confirmed using the approach from [11] considering the linearization.

	k	Residuum
n = 5	0	1.4142e+00
index = 3	1	3.0619e-01
$r_{P_0} = 4$	2	4.3400e-02
$r_\Pi = 2$	3	1.5344e-04
$r_N = 3$	4	3.9713e-10
$D = 6$	5	2.7448e-16

	α	x_0
x_1	1	7.07106781e-01
x_2	1	7.07106781e-01
x_3	0	2.83650820e-16
x_4	0	-3.02484358e-16
x_5	0	7.07106781e-01

Table 2: Results for the pendulum from Example 5.

	k	Residuum
n = 8	0	4.2024e+02
index = 2	1	6.1258e-01
$r_{P_0} = 6$	2	1.9341e-02
$r_\Pi = 4$	3	3.0144e-05
$r_N = 4$	4	7.1077e-15
$D = 10$		

	α	x_0
x_1	1.0e+05	1.00000000e+05
x_2	0.0e+00	-3.15415713e-14
x_3	0.0e+00	3.77140587e-14
x_4	1.2e+04	1.20000000e+04
x_5	-2.0e+00	-1.00000000e+00
x_6	5.0e+01	4.50000000e+01
x_7	2.6e+00	2.67287005e+00
x_8	-5.0e-02	-5.22095858e-02

Table 3: Results for the trajectory prescribed path control example from [1]

.

- Some rank information for the matrices from Section 7:
  - $r_{P_0} := \operatorname{rank} P_0 = \operatorname{rank} A$
  - $r_\Pi := \operatorname{rank} \Pi$ corresponds to the so-called degree of freedom.
  - $r_N := \operatorname{rank} N$ for $N = W_{R\mu-1} G_{L\mu-1}$, i.e. the matrix that describes the constraints.
- Number of Taylor coefficients D used in AlgoPy.
- The residuum in the k-th iteration step.
- The used initial guess α and the obtained consistent initial value x_0. The Taylor coefficients computed simultaneously are reported for Example 4 only, cf. Table 1.

Although the results obtained for these particular examples are highly encouraging, the convergence of the iteration depends on the choice of α in general.

	k	Residuum
n = 8	0	1.4240e+01
index = 5	1	3.8659e+04
$r_{P_0} = 6$	2	5.2003e+04
$r_\Pi = 0$	3	2.0488e+04
$r_N = 8$	4	1.0186e-06
$D = 10$	5	1.3652e-11
	6	8.8896e-12

	α	x_0
x_1	-1.71828183e+00	-1.71828183e+00
x_2	0	3.90881478e-01
x_3	1.71828183e+00	1.71828183e+00
x_4	-2.71828183e+00	-2.71828183e+00
x_5	0	4.28789456e+00
x_6	1.71828183e+00	1.71828183e+00
x_7	0	1.35912606e+01
x_8	0	1.93304288e+01

Table 4: Results for the robotic arm from [3]. For a better comparison with the results from [2] we used the exact values of the true solution for four components of α, namely $[x_1, x_3, x_4, x_6] = [1 - \mathrm{e}^{\mathrm{t}}, \mathrm{e}^{\mathrm{t}} - \mathrm{t}, -\mathrm{e}^{\mathrm{t}}, \mathrm{e}^{\mathrm{t}} - 1]$ at $t = 1$.

	k	Residuum
n = 7	0	1.5240e+00
index = 3	1	1.9007e-01
$r_{P_0} = 4$	2	6.5268e-03
$r_\Pi = 2$	3	1.4421e-03
$r_N = 5$	4	1.5137e-06
$D = 6$	5	7.1761e-13

	α	x_0
x_1	9.40853360e-01	9.37204208e-01
x_2	5.91466397e-02	7.12038017e-02
x_3	-1.15356546e-02	3.63071783e-02
x_4	-1.54643453e-01	-2.12668806e-01
x_5	1.0e+00	2.27142083e-01
x_6	1.0e+00	2.27142083e-01
x_7	1.0e-01	7.72857917e-01

Table 5: Results for the catalyst mixing from [5].

11 Outlook

In this article we present a new approach to compute consistent initial values and consistent Taylor series for higher index DAEs. The consistent values result from the constraints and a specification that, for given values, minimizes the correction for the differentiated components and can be described using suitable orthogonal projections.

This new algorithm to compute consistent Taylor series is of special interest with regard to the integration of DAEs with the Taylor method. In fact, it is a promising alternative to [10] and will be investigated in further work.

References

[1] K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations. Unabridged, corr. republ.* Classics in Applied Mathematics. 14. Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics, 1996.

[2] S. L. Campbell. A general method for nonlinear descriptor systems: an example from robotic path control. Technical report, Department of Mathematics and Center for Research in Scientific Computing, North Carolina State University, CRSC Technical Report 090488-01, October 1988.

[3] S. L. Campbell and E. Griepentrog. Solvability of general differential algebraic equations. *SIAM J. Sci. Comput.*, 16(2):257–270, 1995.

[4] E. Eich-Soellner and C. Führer. *Numerical methods in multibody dynamics.* European Consortium for Mathematics in Industry. Stuttgart: B. G. Teubner., 1998.

[5] R. England, S. Gómez, and R. Lamour. The properties of differential-algebraic equations representing optimal control problems. *Appl. Numer. Math.*, 59, 2009.

[6] D. Estévez Schwarz. *Consistent initialization for index-2 differential algebraic equations and its application to circuit simulation.* PhD thesis, Berlin: Humboldt-Univ., Mathematisch-Naturwissenschaftliche Fakultät II, http://edoc.hu-berlin.de/docviews/abstract.php?id=10218 , 2000.

[7] D. Estévez Schwarz. A step-by-step approach to compute a consistent initialization for the MNA. *Int. J. Circuit Theory Appl.*, 30(1):1–6, 2002.

[8] D. Estévez Schwarz. Consistent initialization for DAEs in Hessenberg form. *Numer. Algorithms*, 52(4):629–648, 2009.

[9] D. Estévez Schwarz and R. Lamour. Monitoring singularities while integrating DAEs. In S. Schöps, A. Bartel, M. Günther, E.J.W. ter Maten, and P.C. Müller, editors, *Progress in Differential-Algebraic Equations, Deskriptor 2013*, Differential-Algebraic Equations Forum. Springer, 2014.

[10] D. Estévez Schwarz and R. Lamour. Projector based integration of DAEs with the Taylor series method using automatic differentiation. *J. Comput. Appl. Math.*, 262:62–72, 2014.

[11] D. Estévez Schwarz and R. Lamour. Indicators to monitor singularities of linear higher index DAEs. *Submitted*, 2015.

[12] G. H. Golub and C. F. van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore and London, 1996.

[13] V. Gopal and L. T. Biegler. A successive linear programming approach for initialization and reinitialization after discontinuities of differential-algebraic equations. *SIAM J. Sci. Comput.*, 20(2):447–467, 1999.

[14] R. Lamour, R. März, and C. Tischendorf. *Differential-algebraic equations: A projector based analysis*. Differential-Algebraic Equations Forum 1. Berlin: Springer, 2013.

[15] N.S. Nedialkov and J.D. Pryce. Solving Differential-Algebraic Equations by Taylor Series (I): Computing Taylor coefficients. *BIT Numerical Mathematics*, 45:561–591, 2005.

[16] N.S. Nedialkov and J.D. Pryce. Solving Differential-Algebraic Equations by Taylor Series (II): Computing the System Jacobian. *BIT Numerical Mathematics*, 47:121–135, 2007.

[17] N.S. Nedialkov and J.D. Pryce. Solving Differential-Algebraic Equations by Taylor Series (III): the DAETS Code. *Journal of Numerical Analysis, Industrial and Applied Mathematics*, 1(1):1–30, 2007.

[18] C.C. Pantelides. The consistent initialization of differential-algebraic systems. *SIAM J. Sci. Stat. Comput.*, 9(2):213–231, 1988.

[19] J. D. Pryce. A Simple Structural Analysis Method for DAEs. *BIT*, 41(2):364–294, 2001.

[20] S. F. Walter and L. Lehmann. Algorithmic differentiation in python with algopy. *Journal of Computational Science*, 4(5):334 – 344, 2013.