A new approach for computing consistent initial values and Taylor coefficients for DAEs using projector based constrained optimization

Diana Estévez Schwarz and René Lamour

April 19, 2017

Abstract

This paper describes a new algorithm for the computation of consistent initial values for Differential-Algebraic Equations (DAEs). The main idea is to formulate the task as a constrained optimization problem in which, for the differentiated components, the computed consistent values are as close as possible to user-given guesses.
The generalization to compute Taylor coefficients results immediately, whereas the amount of consistent coefficients will depend on the size of the derivative array and the index of the DAE.
The algorithm can be realized using Automatic Differentiation (AD) and sequential quadratic programming (SQP). The implementation in Python using AlgoPy and SLSQP has been tested successfully for several higher index problems.

Keywords: DAE, differential-algebraic equation, consistent initial value, index, derivative array, projector based analysis, nonlinear constrained optimization, SQP, automatic differentiation

MSC-Classification: 65L05, 65L80, 34A09, 34A34, 65D25, 90C30, 90C55

1 Introduction

We consider DAEs of the form

$$f(x',x,t) = 0, \quad f : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n. \tag{1}$$

In comparison with ODEs, the numerical treatment of DAEs needs additional computations. For ODEs initial values can be prescribed by $x(t_0) = \alpha$, for arbitrary $\alpha \in \mathbb{R}^n$. In contrast, for DAEs of index μ greater than 0, $x(t_0)$ has to

be *consistent*. This means that some constraints have to be fulfilled and that we cannot prescribe values α for the complete vector $x(t_0)$. Consistent initial values have to fulfill obvious constraints and, in the case of higher index DAEs, also hidden constraints that define the solution manifold. Consequently, arbitrary initial guesses are not consistent in general. In terms of the notation we introduce below, the formulation of suitable initial values will be described by

$$\Pi(x(t_0) - \alpha) = 0 \tag{2}$$

for a suitable matrix Π with rank $\Pi = d$, whereas d corresponds to the so-called *degree of freedom*.

The computation of consistent initial values is a difficult task that has to be tackled for solving DAEs and is closely related to the determination of the index μ. In this context, in [12] we introduced a specific orthogonal projector Π. Here, we present an equivalent approach formulated as a plausible constrained optimization problem. This new formulation is specially convenient because it does not require the computation of Π, which depends in the nonlinear case, in general, on the solution.

For our purposes we will not consider consistent initial values only, but consistent Taylor coefficients $x(t_0), x'(t_0), \frac{x''(t_0)}{2}, \ldots$ of the solution at t_0, which additionally have to fulfill equations obtained by the differentiation of the constraints. These Taylor coefficients are of interest for the numerical approximation of $x(t_0 + h)$ for a step-size h.

The article is organized as follows. In Section 2 we briefly recall the definition of the differentiation index and reconsider it according to [11], which leads to a formulation with a 1-full matrix. Section 3 illustrates the main idea of our new approach for a simple linear example and emphasizes the differences to other methods from the literature. Our new algorithm is then presented in Section 4, where the initialization problem is formulated as a constrained optimization problem. The objective function is quadratic and the constraints are precisely the derivative array. The general properties of this optimization problem are outlined in Section 5, whereupon, in Section 6, the Lagrange formulation is considered for the linear and the nonlinear case. Finally, in Section 7 we report the numerical results we obtained for several higher index examples. Some essential results from linear algebra are summarized in the Appendix.

2 The Derivative Array and the Index

Defining

$$F_j(x^{(j+1)}, x^{(j)}, \ldots, x', x, t) := \frac{d^j}{dt^j} f(x', x, t)$$

for $k \in \mathbb{N}$ we consider

$$f(x',x,t) = 0, \tag{3}$$
$$F_1(x'',x',x,t) = 0, \tag{4}$$
$$\vdots$$
$$F_k(x^{(k+1)},\ldots,x,t) = 0. \tag{5}$$

The differentiation index μ is defined as the smallest integer k for which the so-called derivative array (3)-(5) determines x' as a function of (x,t) (cf. [3, 2]). If this condition is given, the derivative array permits locally the formulation of the underlying ODE.

To characterize the index by a rank check, for $z_i \in \mathbb{R}^n$, $i = 0,\ldots,k$ we define

$$g^{[k+1]}(z_0,z_1,\ldots,z_{k+1},t) := \begin{pmatrix} f(z_1,z_0,t) \\ F_1(z_2,z_1,z_0,t) \\ \vdots \\ F_k(z_{k+1},\ldots,z_0,t), \end{pmatrix} \tag{6}$$

for $k \in \mathbb{N}$. If we denote by

$$\mathscr{A}^{[k+1]}(z_0,z_1,\ldots,z^{k+1},t) \in \mathbb{R}^{n\cdot(k+1)\times n\cdot(k+1)}$$

the Jacobian matrix of $g^{[k+1]}(z_0,z_1,\ldots,z_{k+1},t)$ with respect to $(z_1,\ldots,z_{k+1})$, then, in practice, the differentiation index μ is determined by verifying whether

$$\ker \mathscr{A}^{[k+1]} \subseteq \left\{ \begin{pmatrix} z_1 \\ \vdots \\ z_{k+1} \end{pmatrix},\ z_i \in \mathbb{R}^n \ : \quad z_1 = 0 \right\}$$

is given for $k \geq \mu$ and $z_i \in \mathbb{R}^n$ in the region of interest. This means that the matrix $\mathscr{A}^{[k+1]}$ is 1-full, cf. [3], [16] and Appendix.

In [11] we presented a different point of view that turns out to be very convenient for consistent initialization. Let

$$G^{[k]}(z_0,z_1,\ldots,z_k,t) \in \mathbb{R}^{n\cdot k\times n\cdot(k+1)}$$

denote the Jacobian matrix of $g^{[k]}(z_0,z_1,\ldots,z_k,t)$ with respect to $(z_0,z_1,\ldots,z_k)$. For our purposes we will focus on the first columns of $G^{[k]}$ that correspond to z_0, instead of the first columns of $\mathscr{A}^{[k+1]}$ that correspond to z_1.

Since the equations of the DAE should not be redundant, we assume that $G^{[k]}(z_0,z_1,\ldots,x^{(k)},t)$ has full row rank, i.e., rank $G^{[k]} = n\cdot k$ for all $z_i \in \mathbb{R}^n$ and

$t \in \mathbb{R}$ in the region of interest. Recall that for linear DAEs with constant coefficients this is equivalent to the regularity of the matrix pencil.

Furthermore, we assume that $\ker f_{z_1}(z_1, z_0, t)$ does not depend on z_1, z_0 and that there is a continuously differentiable projector valued function Q such that

$$\ker f_{z_1}(z_1, z_0, t) = \operatorname{im} Q(t)$$

with the complementary projector $P(t) = I - Q(t)$. For convenience we omit the argument t in the following and consider the decoupling $x' = (Px)' + (Qx)'$ for an alternative definition of the differentiation index. Reformulating the DAE (1) according to [18],

$$f(x', x, t) = f(Px', x, t) = f((Px)' - P'x, x, t) = 0 \tag{7}$$

we see that (cf. [11])

- $(Px)'$ is determined by (7). In this sense, Px describes the differentiated component and Qx the undifferentiated component of x.
- $(Qx)'$ has to be determined by $g^{[\mu+1]}(x, x', \ldots, x^{(\mu+1)}, t)$.

Therefore, to find out the index $\mu \geq 1$, it is in fact sufficient to check whether

$$g^{[\mu]}(x, x', \ldots, x^{(\mu)}, t)$$

uniquely determines Qx as a function of (Px, t), i.e., $Qx = \varphi(Px, t)$. The differentiation of this function φ leads then to the missing expression for $(Qx)'$. Accordingly, in [11] we defined the DAE-index μ as the smallest integer k for which

$$\ker \begin{pmatrix} P & 0 & \ldots & 0 \\ G^{[k]}(z_0, z_1, \ldots, z_k, t) \end{pmatrix} \subseteq \left\{ \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_k \end{pmatrix}, \; z_i \in \mathbb{R}^n \; : \quad z_0 = 0 \right\} \tag{8}$$

is given for $x^{(i)}(t_0) = z_i \in \mathbb{R}^n$ in the region of interest, $i = 0, \ldots, k$, i.e., for which the matrix

$$\mathscr{B}^{[k]} := \begin{pmatrix} P & 0 & \ldots & 0 \\ G^{[k]}(z_0, z_1, \ldots, z_k, t) \end{pmatrix} \in \mathbb{R}^{n \cdot (k+1) \times n \cdot (k+1)} \tag{9}$$

is 1-full. In the present article, we take advantage of this property with regard to consistent initialization, realizing that (9) is precisely the Jacobian of

$$P(z_0 - \alpha) = 0, \tag{10}$$

$$g^{[k]}(z_0, z_1, \ldots, z_k, t) = 0. \tag{11}$$

Note that, although (9) is singular, for $\mu > 1$ the system (10)-(11) is overdetermined with respect to z_0, because, indeed, only $\Pi(z_0 - \alpha) = 0$ can be prescribed for a matrix Π fulfilling rank $\Pi = d <$ rank P, cf. (2).

Formulating a suitable constrained optimization problem based on property (8), we compute uniquely defined initial values without an explicit computation of a matrix Π, cf. Section 4. In the next section, we will illustrate the main new idea of this approach before introducing it formally.

3 An Introductory Example

Although the computation of a consistent initialization has been identified as a crucial task when dealing with DAEs, in practice the state of the art is still unsatisfactory, in particular when considering that for the same initial guess different solvers may deliver different consistent values and, correspondingly, different numerical solutions of the DAE are computed. Even worse, the same solver may deliver different results depending on the formulation of the problem. In our opinion, the demands on the properties of the computed initial values are often not taken into account sufficiently.

The large body on literature on the initialization problem can be classified, very roughly speaking, into three categories:

(a) methods based on a structural analysis like [19], [20] and the references therein, which trace back to [7] and [21].

(b) methods that compute a minimal deviation from a given guess, cf. [14], [17], [8], among others,

(c) methods that use projectors to describe the different components, cf. [10], [18], and the related work.

Here, we aim at a combination of (b) and (c). In contrast to (a), our projector based approach obtains the information about the properties of matrices using the singular value decomposition (see Appendix). We illustrate the main differences to other well-established methods by means of the following example:

Example 1.

$$\begin{aligned} x_1' + x_2' + x_1 + x_3 &= 2, && (12)\\ x_1' + 2x_2' + x_1 + x_2 + x_3 &= 3, && (13)\\ x_1 + 2x_2 &= 4. && (14) \end{aligned}$$

Obviously, the differentiation index is 2 and the hidden constraint reads:

$$x_1 + x_2 + x_3 = 3.$$

We suppose that an (inconsistent) initial guess

$$\alpha = \begin{pmatrix} 1 \\ 2 \\ 9 \end{pmatrix}, \tag{15}$$

for the initial value x_0 is given. A summary of some results for the different consistent values we compute below is presented in Table 1.

3.1 A Structural Method

Let us focus on the results we obtain for Example 1 with Dymola[1], cf. [19]. For

```
model InitializationDAE
  Real x1(start=1),x2(start=2),x3(start=9);
equation
  der(x1) + der(x2) + x1+x3 = 2;
  der(x1) +2*der(x2)+ x1+x2+x3 = 3;
  x1+2*x2 = 4;
  annotation (uses(Modelica(version="3.2.1")));
end InitializationDAE;
```

we obtain the consistent value

$$x_0 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}.$$

Obviously, the program prescribes the initial value for x_2. Afterwards x_1, x_3 are computed accordingly, using the constraints. In contrast, rewriting the latter equation as

```
  2*x2+x1 = 4;
```

provides

$$x_0 = \begin{pmatrix} 1 \\ 1.5 \\ 0.5 \end{pmatrix}.$$

[1]Dymola – Dynamic Modeling Laboratory, Dynasim AB, Lund, Sweden. Homepage: http://www.Dynasim.se

In this case, the program prescribes the initial value for x_1. Afterwards, x_2, x_3 are computed, accordingly, using the constraints.

Evidently, the choice made by the structural algorithms depends on the order in which we write down the variables in the equations. As a consequence, although this is completely against our expectation, the obtained numerical results may depend on the order of the variables in the DAE formulation.

3.2 A Minimal Deviation Approach

An approach that does not exhibit the above described dependence computes the solution of minimizing

$$\|x_0 - \alpha\|_2$$

subject to the constraints

$$\underbrace{\begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}}_{=:N} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 4 \\ 3 \end{pmatrix}}_{=:b}. \tag{16}$$

The unique solution can be represented using the Moore-Penrose inverse $N^+$of N, i.e.,

$$x_0 = \alpha - N^+(N\alpha - b) = \begin{pmatrix} -2 \\ 3 \\ 2 \end{pmatrix}$$

with $\|x_0 - \alpha\|_2 = 7.6811$ and α according to (15). This result was also obtained using GELDA, cf. [17].

The drawback of this approach is that, by construction, the results obtained for x_1 and x_2 depend on the value for the third component of α, i.e. the initial guess for the undifferentiated component x_3. For e.g. $\hat{\alpha} = \begin{pmatrix} 1 & 2 & 0 \end{pmatrix}^T$ we obtain the value $\hat{x}_0 = \begin{pmatrix} 1 & 1.5 & 0.5 \end{pmatrix}^T$.

We want to mention that according to the documentation of GELDA, when setting the option INFO(12)=1, only the "differential variables" are prescribed. However, since the strangeness-free formulation is used, these "differential variables" do not coincide with Px in the higher index case. The option INFO(12)=1 is implemented, although it has to be activated. For (12)-(14) and α according to (15), GELDA yields

$$x_0 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix},$$

provided that INFO(12)=1 is forced. A comparison of this result and our new approach from below can be found in Table 1.

3.3 New Approach

Our purpose is to combine both advantages:

- on the one hand, unique solvability should be given at least for linear DAEs,
- on the other hand, the specification of consistent initial values should focus on differentiated components only.

For the simple example presented above, the differentiated component may be described by Px for the orthogonal projector

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{17}$$

and the approach we present in Section 4 consists in minimizing $\|P(x-\alpha)\|_2$, i.e.,

$$\min \quad \sqrt{(x_1-\alpha_1)^2+(x_2-\alpha_2)^2}, \tag{18}$$

subject to

$$\begin{aligned} x_1+2x_2 &= 4, & (19)\\ x_1+x_2+x_3 &= 3. & (20) \end{aligned}$$

For this problem we obtain the unique solution:

$$x_0 = \begin{pmatrix} 0.8 \\ 1.6 \\ 0.6 \end{pmatrix}, \quad \|P(x_0-\alpha)\|_2 = \sqrt{0.2} \approx 0.45. \tag{21}$$

Notice that for this x_0 we have $\|x_0-\alpha\|_2 \approx 8.42$, but this is not the norm we want to minimize.

In our previous work [12], we described a completely different algorithm to compute the same minimal norm solution. There we used a projector based formulation with a uniquely defined orthogonal projector Π. With the notation from (16), the orthogonal projector P from (17), $Q = I - P$, and an arbitrary projector W along $\operatorname{im} NQ$, for Example 1 Π is defined as the unique orthogonal projector onto

$$\ker \begin{pmatrix} Q \\ WNP \end{pmatrix} = \ker Q \cap \ker WNP = \ker \begin{pmatrix} 0 & 0 & 1 \\ 1 & 2 & 0 \end{pmatrix}.$$

$\alpha = \begin{pmatrix}1\\2\\9\end{pmatrix}$	Dymola		GELDA		New
	$x_0 = \begin{pmatrix}0\\2\\1\end{pmatrix}$	$x_0 = \begin{pmatrix}1\\1.5\\0.5\end{pmatrix}$	$x_0 = \begin{pmatrix}-2\\3\\2\end{pmatrix}$	$x_0 = \begin{pmatrix}2\\1\\0\end{pmatrix}$	$x_0 = \begin{pmatrix}0.8\\1.6\\0.6\end{pmatrix}$
$\|x_0-\alpha\|_2$	8.0623	8.5147	$\boxed{7.6811}$	9.1104	8.4119
$\|P(x_0-\alpha)\|_2$	1	0.5000	3.1623	1.4142	$\boxed{0.4472}$

Table 1: Comparison of the values obtained for Example 1 by the different approaches. The framed values are minimal.

Consequently, we obtain

$$\Pi = \frac{1}{5}\begin{pmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

According to [12], with this projector, the consistent value x_0 from (21) can be determined as the unique solution of

$$\begin{aligned} \Pi x_0 &= \Pi\alpha, \\ Nx_0 &= b. \end{aligned}$$

Using our new formulation from Section 4 as a constrained optimization problem, we avoid the explicit computation of Π. This is of special interest for the nonlinear case since Π is not constant in general. Let us also emphasize that, in the nonlinear case, a formulation of the constraints like (19)-(20) may not be possible without algebraic manipulations. Therefore, the complete derivative array will be considered below, cf. (23)-(25) and Example 2.

4 Consistent Initialization by Constrained Optimization

For given $\alpha \in \mathbb{R}^n$, $t_0 \in \mathbb{R}$, $P(t_0) \in \mathbb{R}^{n\times n}$ we minimize

$$\|P(x_0-\alpha)\|_2 \tag{22}$$

subject to

$$f(x_0', x_0, t_0) = 0, \tag{23}$$
$$F_1(x_0'', x_0', x_0, t_0) = 0, \tag{24}$$
$$\vdots$$
$$F_k(x_0^{(k+1)}, \ldots, x_0, t_0) = 0, \tag{25}$$

whereas $X := (x_0, x_0', \ldots, x_0^{(k+1)})$ represent initial values for $x(t)$ and its derivatives at t_0.

As will be shown below, the crucial aspect of this elegant formulation is that P has to be the unique orthogonal projector fulfilling

$$\ker P = \ker f_{x'}.$$

In fact, the requirement (22) is a user-friendly condition that (for linear DAEs) uniquely determines consistent initial Taylor coefficients if k is sufficiently large, depending on the index. Moreover, it means that for the differentiated component the consistent initial value is computed as close as possible to an initial guess, whereas initial guesses for the undifferentiated component are ignored.

Example 2. *For the example from Section 3 and $k = 1$ the approach* (22)-(25) *minimizes*

$$\sqrt{(x_1 - \alpha_1)^2 + (x_2 - \alpha_2)^2} \tag{18}$$

subject to

$$\begin{aligned} x_1' + x_2' + x_1 + x_3 &= 2, && (26)\\ x_1' + 2x_2' + x_1 + x_2 + x_3 &= 3, && (27)\\ x_1 + 2x_2 &= 4, && (28) \end{aligned}$$

$$\begin{aligned} x_1'' + x_2'' + x_1' + x_3' &= 0, && (29)\\ x_1'' + 2x_2'' + x_1' + x_2' + x_3' &= 0, && (30)\\ x_1' + 2x_2' &= 0. && (31) \end{aligned}$$

Recall that the computed values for x_3', x_1'', x_2'' will not be consistent in general, and the optimization problem is underdetermined for

$$X = (\underbrace{x_1, x_2, x_3}_{z_0}, \underbrace{x_1', x_2', x_3'}_{z_1}, \underbrace{x_1'', x_2'', x_3''}_{z_2}).$$

In fact, x_3'' does not even appear in the above equations. However, $z_0 = (x_1, x_2, x_3)$ result to be uniquely determined since the constraints (19)-(20) are contained in (26)-(31). Increasing k, we would analogously obtain consistent values for higher derivatives.

5 Properties of the Optimization Problem

For convenience we suppose that $k \in \mathbb{N}$ is fixed and define

$$\tilde{P} := \begin{pmatrix} P & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{n\cdot(k+1)\times n\cdot(k+1)}, \tilde{\alpha} := \begin{pmatrix} \alpha \\ * \end{pmatrix} \in \mathbb{R}^{n\cdot(k+1)}, \tag{32}$$

and

$$X := (x_0, x_0', \ldots, x_0^{(k)})$$

such that $\tilde{P}\tilde{X} = \tilde{P}\tilde{\alpha}$ is equivalent to $Px_0 = P\alpha$. Moreover, we skip the index $[k]$ for g and G.

With this notation and (6), the above constrained problem is equivalent to minimizing

$$\frac{1}{2}\left\|\tilde{P}(X-\tilde{\alpha})\right\|_2^2 \tag{33}$$

subject to

$$g(X,t_0) \;=\; 0. \tag{34}$$

Obviously, we consider a constrained optimization problem with only equality constraints. Moreover, for a fixed $\alpha \in \mathbb{R}^n$, the objective function

$$\begin{aligned} f(X) &= \frac{1}{2}(X-\tilde{\alpha})^T\tilde{P}(X-\tilde{\alpha}) \\ &= \frac{1}{2}(X^T\tilde{P}X - 2\tilde{\alpha}^T\tilde{P}X + \tilde{\alpha}^T\tilde{P}\tilde{\alpha}) \\ &= \frac{1}{2}X^T\tilde{P}X + q^TX + c, \end{aligned} \tag{35}$$

for $q^T = -\tilde{\alpha}^T\tilde{P}$, $c = \frac{1}{2}\tilde{\alpha}^T\tilde{P}\tilde{\alpha}$ is quadratic, whereas $\tilde{P}$ is symmetric and positive semidefinite, due to the fact that it is an orthogonal projector. Consequently, f is convex but not strictly convex, since $\tilde{P}$ is singular, cf. e.g. [1].

Our constraints $g(X,t_0)$ are nonlinear and not convex in general, and the set of feasible solutions of the optimization problem

$$\mathscr{G}(t_0) := \left\{ X \in \mathbb{R}^{k+1} \quad : g(X,t_0) = 0 \right\}$$

contains all possible consistent initial values for $[x(t_0), x'(t_0), \ldots, x^{(k)}(t_0)]$. Moreover, the Jacobian matrix G of g is supposed to have full row rank.

For our optimization problem the components of interest are uniquely determined, even though some other components are not, as will be pointed out in the

next section. In particular, the values for some initial values for higher Taylor coefficients are not uniquely determined by the equations, such that in practice we compute the minimum norm solutions for them. We illustrate this discussing the Lagrange approach and using the 1-full property from $\mathscr{B}^{[k]}$.

6 The Lagrange Approach

In this section we focus on the Lagrange equations resulting from (22)-(25). For linear DAEs a closer look allows us to show that the initial value x_0 is uniquely determined if k is sufficiently large.

The Lagrange approach leads to

$$L(X,\lambda) \;=\; \frac{1}{2}(X-\tilde{\alpha})^T\tilde{P}(X-\tilde{\alpha})+\lambda^T g(X,t_0)$$

with

$$\frac{\partial L}{\partial X} = (\tilde{P}(X-\tilde{\alpha}))^T+\lambda^T G(X,t_0) = 0 \tag{36}$$

and

$$\frac{\partial L}{\partial \lambda} = g(X,t_0) = 0. \tag{37}$$

Therefore, we consider the system

$$\begin{aligned} \tilde{P}(X-\tilde{\alpha})+G^T(X,t_0)\lambda &= 0, && (38)\\ g(X,t_0) &= 0. && (39) \end{aligned}$$

The Jacobian of (38)-(39) reads

$$\begin{pmatrix} \tilde{P}(t_0)+\Gamma(X,\lambda,t_0) & G^T(X,t_0) \\ G(X,t_0) & 0 \end{pmatrix} \tag{40}$$

for $\Gamma(X,\lambda,t_0) := \frac{\partial}{\partial X}(G^T(X,t_0)\lambda)$. Recall that, e.g. in [1], conditions for the local quadratic convergence of the Lagrange-Newton method are established, but they are not fulfilled in our case since

$$\frac{\partial^2 L}{\partial X \partial X} = \tilde{P}(t_0)+\Gamma(X,\lambda,t_0)$$

is singular in general. In fact, if g is linear, then $\nabla^2_{XX}L = \tilde{P}$ holds.

It will be shown below that this system is underdetermined in general while the first components of X are uniquely determined. This will also be illustrated in the Example 3 of Section 7.

6.1 Linear DAEs

We suppose that $k \geq \mu - 1$ is given and drop the arguments of the matrices. With regard to the Lagrange approach, in the linear case we suppose that

$$g(X,t_0) = G(t_0)X + r(t_0), \quad r(t) = \begin{pmatrix} q(t) \\ \dots \\ q^{(\mu-1)}(t) \end{pmatrix},$$

and, consequently, that the Jacobian matrix corresponding to (38)-(39) reads

$$\begin{pmatrix} \tilde{P} & G^T \\ G & 0 \end{pmatrix} \in \mathbb{R}^{n(2k+1)\times n(2k+1)}. \tag{41}$$

Let us have a closer look on matrices presenting this structure.

Theorem 1. *Let $\tilde{P} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$ be an orthogonal projector. Then for an arbitrary matrix $G \in \mathbb{R}^{\tilde{m}\times\tilde{n}}$ with* rank $(G) = \tilde{m}$ *it holds*

$$\ker \begin{pmatrix} \tilde{P} & G^T \\ G & 0 \end{pmatrix} = \left\{ \begin{pmatrix} z \\ 0 \end{pmatrix} \in \mathbb{R}^{\tilde{n}+\tilde{m}} \quad : \quad z \in \ker \begin{pmatrix} \tilde{P} \\ G \end{pmatrix} \right\}.$$

Proof. For

$$\begin{pmatrix} z \\ \lambda \end{pmatrix} \in \ker \begin{pmatrix} \tilde{P} & G^T \\ G & 0 \end{pmatrix}$$

it holds for $\tilde{Q} := I - \tilde{P}$ that

$$\begin{aligned} \tilde{P}z + G^T\lambda &= 0, \\ G(\tilde{P}+\tilde{Q})z &= 0, \end{aligned}$$

and consequently

$$\begin{aligned} \tilde{Q}G^T\lambda &= 0, \\ G\tilde{Q}z &= GG^T\lambda, \end{aligned}$$

i.e.,

$$\lambda = (GG^T)^{-1}G\tilde{Q}z,$$

and therefore

$$0 = z^T \underbrace{\tilde{Q}G^T\lambda}_{=0} = z^T\tilde{Q}G^T(GG^T)^{-1}G\tilde{Q}z \underset{\tilde{Q}=\tilde{Q}^T}{=} (G\tilde{Q}z)^T(GG^T)^{-1}G\tilde{Q}z.$$

From the positive definiteness of $(GG^T)^{-1}$ it follows that $G\tilde{Q}z = 0$ and, hence, $\lambda = 0$, $\tilde{P}z = 0$. Summarizing, we obtain

$$z \in \ker \begin{pmatrix} \tilde{P} \\ G\tilde{Q} \end{pmatrix} = \ker \begin{pmatrix} \tilde{P} \\ G \end{pmatrix}, \quad \lambda = 0.$$

□

The following result, which, for our purposes, will be considered when N describes the matrix corresponding to all constraints, follows immediately.

Corollary 1. *For any orthogonal projector $P \in \mathbb{R}^{n\times n}$ and any matrix $N \in \mathbb{R}^{m\times n}$ fulfilling*

$$\ker \begin{pmatrix} P \\ N \end{pmatrix} = \{0\}, \quad \operatorname{rank} N = m,$$

the matrix

$$\begin{pmatrix} P & N^T \\ N & 0 \end{pmatrix}$$

is regular.

Hence, in order to obtain a uniquely solvable optimization problem, the explicit computation of all constraints $Nx = b$ can be realized for linear DAEs.

Corollary 2. *For $k \geq \mu - 1$ all solutions of the underdetermined system* (38)-(39) *for linear DAEs, i.e. of*

$$\begin{pmatrix} \tilde{P} & G^T \\ G & 0 \end{pmatrix} \begin{pmatrix} X \\ \lambda \end{pmatrix} = \begin{pmatrix} \tilde{P}\alpha \\ r(t) \end{pmatrix}$$

involve a uniquely determined first component x_0. If $k > \mu - 1$, then also unique values for higher derivatives up to order $k - \mu$ are obtained for the s-full matrix (cf. Definition 3 from the Appendix) with $s = k - \mu + 1$. Further, the unique value x_0 solves the minimization problem (22).

Proof. Since $\mathscr{B}^{[k]}$ is 1-full for $k \geq \mu$ (cf. (9)), Theorem 1 implies that x_0 is uniquely determined by the underdetermined system. Since any solution of the system is a solution of the Lagrange-formulation of the minimization problem, x_0 results to be the solution of (22). For the higher derivatives, the corresponding result follows from Definition 3 from the Appendix. □

6.2 Nonlinear DAEs

In the nonlinear case, some iterative approaches that construct quadratic programming (QP) subproblems replace

- the objective function f by its local quadratic approximation

$$f(X) \approx f(X^i) + \nabla f(X^i)(X - X^i) + \frac{1}{2}(X - X^i)^T \nabla^2 f(X^i)(X - X^i) \qquad (42)$$

- the constraint function g by the local affine approximation

$$g(X, t_0) = g(X^i, t_0) + \nabla_X g(X^i, t_0)(X - X^i).$$

Setting

$$d_X := X - X^i$$

and using

$$\nabla^2_{XX} f(X^i) = \tilde{P}, \quad \nabla_X g(X^i, t_0) =: G,$$

we obtain the following QP subproblem:

minimize

$$\nabla f(X^i)^T d_X + \frac{1}{2} d_X^T \tilde{P} d_X$$

subject to

$$g(X^i, t_0) + G d_X \quad = \quad 0.$$

For this QP subproblem the Lagrange approach leads to

$$\begin{pmatrix} \tilde{P} & G^T \\ G & 0 \end{pmatrix} \begin{pmatrix} d_X \\ d_\lambda \end{pmatrix} = - \begin{pmatrix} \nabla f(X^i) \\ g(X^i, t_0) \end{pmatrix}.$$

Consequently, Theorem 1 reveals why the linear system obtained in the iteration has a unique solution for the first components.

However, if instead of (42)

$$f(X) \approx f(X^i) + \nabla f(X^i)(X - X^i) + \frac{1}{2}(X - X^i)^T \nabla^2 L(X^i, \lambda^i)(X - X^i) \qquad (43)$$

is considered, then no structural information is given in general.

Nevertheless, comparing the formulation from Section 4 with the formulation from [12] for nonlinear DAEs we have several advantages:

- one crucial advantage is that there is plenty of theory and software for constraint optimization available.
- in important applications like Hessenberg DAEs, DAEs from network simulation, the orthogonal projector P results to be constant, whereas the projector Π depends on the solution in general. Using the formulation with P, we avoid the computation and handling of Π.

7 Examples

We implemented the algorithm for nonlinear DAEs in Python using:

- Automatic Differentiation (AlgoPy, cf. [22]) to compute all the derivatives of $f(x',x,t)$, considering Taylor coefficients for $D = k+1$:

$$\left[\underbrace{x(t) \quad x'(t) \quad \frac{x''(t)}{2} \quad \dots \quad \frac{x^{(k)}(t)}{k!}}_{(k+1)=D \text{ elements}}\right].$$

- Sequential Least Squares Programming (SLSQP) to solve the optimization problem (see scipy.optimize.minimize, and [15]).

By the first example, we illustrate for the higher-index case the consequences of the fact that $\mathscr{B}^{[k]}$ is 1-full, but does not have full column rank.

Example 3. *For illustrative reasons we start with a simple example in Kronecker canonical form with index 4 and an obvious solution:*

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} x' + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \sin(t) \end{pmatrix}, \; x = \begin{pmatrix} Ce^{-t} \\ \cos(t) \\ -\sin(t) \\ -\cos(t) \\ \sin(t) \end{pmatrix}. \quad (44)$$

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 coefficients, we see in Table 2 that, corresponding to the description from [12], the last four coefficients are not correct for* x_2*, and so are the last three for* x_3 *and the last two for* x_4*. In contrast, for* x_1 *all coefficients are exact up to numerical accuracy and for* x_5 *only the last coefficient is not correct. 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. In terms of Corollary 2 this means that* $\mathscr{B}^{[k]}$ *for* $k = D-1 = 5$ *is s-full for* $s = 2$*, i.e., that at least two Taylor coefficients of the computed solution are correct.*

In the Tables 3 - 6, we summarize the results we obtained for some well-known meaningful examples from applications described in literature. We report the normalized pendulum, the trajectory prescribed path control from [2], the robotic arm from [5], and the catalyst mixing from [9].

The following details are provided:

	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	0.00000000e+00	3.53553391e-01	3.53553391e-01	-3.53553391e-01
$\frac{1}{3!}x'''_*(\frac{\pi}{4})$	-1.66666667e-01	1.07580963e-16	0.00000000e+00	-1.17851130e-01	-1.17851130e-01
$\frac{1}{4!}x^{(iv)}_*(\frac{\pi}{4})$	4.16666667e-02	0.00000000e+00	-2.68952407e-17	0.00000000e+00	2.94627825e-02
$\frac{1}{5!}x^{(v)}_*(\frac{\pi}{4})$	-8.33333333e-03	0.00000000e+00	0.00000000e+00	5.37904813e-18	0.00000000e+00

Table 2: Solution of the system (44) from Example 3 for $t_0 = \pi/4$ and $\alpha = [1,0,0,0,0]$ using Taylor coefficients with $D = 6$. The framed values are obviously not consistent, since the Taylor coefficients are correct up to the second coefficients ($\mu = 4$, $D - \mu = 2$).

- Number n of equations (and variables), index μ of the DAE. The index is known from the literature and confirmed using the approach from [11] considering the linearization at the consistent values.
- Some information for the matrices from Section 6.2:
  - $r_P := \operatorname{rank} P$
  - $r_\Pi := \operatorname{rank} \Pi$ corresponds to the so-called degree of freedom d.
- Number of Taylor coefficients D used in AlgoPy.
- $ftol$ precision goal for the value of f in the minimizer of SciPy.
- $||P(x_0 - \alpha)||_2$
- cond $\mathscr{B}^{[D-1]} := \frac{\sigma_1}{\sigma_r}$ for the singular values for $\mathscr{B}^{[D-1]}$, cf. Definition 1 from the Appendix.
- s according to Corollary 4 from the Appendix. For the computation of s we check the amount of zero-valued rows of the last $(D-1)\cdot n - r$ columns of the matrix V. An entry $v_{i,j}$ is considered to be zero if

  $$\left|v_{ij}\right| \leq eps \cdot \operatorname{cond} \mathscr{B}^{[D-1]},$$

  whereas eps corresponds to the relative machine precision.
- The used initial guess α and the obtained consistent initial value x_0. The Taylor coefficients computed simultaneously are reported for Example 3 only, cf. Table 2.

Pendulum

DAE properties			$(\alpha,0)$	(x_0,x_0')
n	5	x_1	1	7.07106781e-01
index	3	x_2	1	7.07106781e-01
r_P	4	x_3	0	-1.26938674e-16
r_Π	2	x_4	0	1.26938680e-16
Numerical information		x_5	0	7.07106781e-01
D	7	x_1'	0	-1.26938667e-16
ftol	1e-10	x_2'	0	1.26938667e-16
$\|\|P(x_0-\alpha)\|\|_2$	0.4142	x_3'	0	5.00000000e-01
cond $\mathscr{B}^{[D-1]}$	73.5249	x_4'	0	-5.00000000e-01
s	4	x_5'	0	3.80816078e-16

Table 3: Characteristics and results for the normalized pendulum, i.e., $g = 1$, $m = 1, l = 1$.

Trajectory Prescribed Path Control Example

DAE properties			α	x_0
n	8	x_1	1.0e+05	1.00000000e+05
index	2	x_2	0.0e+00	3.78720216e-22
r_P	6	x_3	0.0e+00	1.12338674e-06
r_Π	4	x_4	1.2e+04	1.20000000e+04
Numerical information		x_5	-2.0e+00	-1.00000000e+00
D	11	x_6	5.0e+01	4.50000000e+01
ftol	1e-10	x_7	2.6e+00	2.67287005e+00
$\|\|P(x_0-\alpha)\|\|_2$	5.0990	x_8	-5.0e-02	-5.22099022e-02
cond $\mathscr{B}^{[D-1]}$	1.2769e+8			
s	9			

Table 4: Results for the trajectory prescribed path control example from [2].

Robotic Arm

DAE properties			α	x_0
n	8	x_1	-1.71828183e+00	-1.71828183e+00
index	5	x_2	0	3.90881478e-01
r_P	6	x_3	1.71828183e+00	1.71828183e+00
r_Π	0	x_4	0	-2.71828183e+00
Numerical information		x_5	0	4.28789456e+00
D	11	x_6	0	1.71828183e+00
ftol	1e-10	x_7	0	1.35912606e+01
$\|\|P(x_0-\alpha)\|\|_2$	4.3057	x_8	0	1.93304288e+01
cond $\mathscr{B}^{[D-1]}$	2.1623e+7			
s	6			

Table 5: Results for the robotic arm from [5]. For a better comparison with the results from [4] we used the exact values of the solution for two components of α, namely $[x_1, x_3] = [1-\mathrm{e}^{\mathrm{t}}, \mathrm{e}^{\mathrm{t}}-\mathrm{t}]$ at $t = 1$.

Catalyst Mixing

DAE properties			α	x_0
n	7	x_1	9.40853360e-01	9.39924293e-01
index	3	x_2	5.91466397e-02	7.14104593e-02
r_P	4	x_3	-1.15356546e-02	3.39241956e-02
r_Π	2	x_4	-1.54643453e-01	-2.14479270e-01
Numerical information		x_5	1.0e+00	2.27142083e-01
D	7	x_6	1.0e+00	2.27142083e-01
ftol	1e-10	x_7	1.0e-01	7.72857917e-01
$\|\|P(x_0-\alpha)\|\|_2$	0.0761			
cond $\mathscr{B}^{[D-1]}$	72.0672			
s	4			

Table 6: Results for the catalyst mixing from [9].

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

In the latter tables, we included only the values obtained for x_0, even though the Taylor coefficients were computed for all examples. Notice that the main difference to [6], where some of these examples were already discussed, is that in [6] a consistent x_0 was supposed to be given.

In [12] we reported the results for these examples, computing the same consistent initial values by the approach based on the explicit computation of the orthogonal projector Π and setting up a suitable nonlinear system of equations, which is solved with Newton-like methods. The obtained numerical values coincide up to rounding errors.

8 Summary

In this article we present a new approach to compute consistent initial values and consistent Taylor coefficients 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. For the optimization problem we analyzed the consequences of the fact that the relevant matrix of the underdetermined system of equations is 1-full. The derivative array is computed using automatic differentiation and the computation is done using algorithms from constrained optimization. Due to its plausible formulation and the possibility to use advanced solvers from constraint optimization for the computation, the new approach is well-suited to be integrated into existing DAE-solvers. It is of special interest for the restart of integration methods after discontinuities and for the integration with Taylor-series methods. First numerical tests have confirmed our expectations.

Appendix: Toolbox from Linear Algebra

Definition 1. *(cf. [13])*
For $A \in \mathbb{R}^{m\times n}$, $r = \operatorname{rank} A$ the singular value decomposition (SVD) reads

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

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 $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_r > 0$, $r := \operatorname{rank} A \leq \min\{m,n\}$.

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

$$AQ = 0, \quad Q^2 = Q, \quad \text{rank}\, Q = n - \text{rank}\, A = n - r, \quad Q = Q^T,$$

can be described by

$$Q = V \begin{pmatrix} 0 & \\ & I_{n-r} \end{pmatrix} V^T = V(:, r+1:n) \cdot (V(:, r+1:n))^T \tag{45}$$

using MATLAB-notation for the last expression. For the complementary orthogonal projector $P := I - Q$

$$P = I - Q = V \begin{pmatrix} I_r & \\ & 0 \end{pmatrix} V^T = (V(:, 1:r))(V(:, 1:r))^T \tag{46}$$

is given by definition. The rectangular matrix $B := V(:, 1:r)$ contains an orthonormal basis.

Lemma 1. *For an orthogonal projector $P = BB^T$ it holds*

$$\|Px\|_2 = \left\|B^T x\right\|_2 \quad \textit{for all} \quad x \in \mathbb{R}^n.$$

Proof.

$$\|Px\|_2^2 = x^T P^T P x = x^T P x = x^T B^T B x = \left\|B^T x\right\|_2^2.$$

□

Hence, in practice (22) can be formulated with $\left\|B^T x\right\|_2$. For the theory we preferred the formulation using projectors.

It is known from linear algebra (the row echelon normal form) that the first components j of a consistent system of equations $Ax = b$ with $A \in \mathbb{R}^{m\times n}$ are uniquely defined iff there is a nonsingular matrix $R \in \mathbb{R}^{m\times m}$ and a matrix $H \in \mathbb{R}^{(m-j)\times(n-j)}$ such that

$$RA = \begin{pmatrix} I_j & 0 \\ 0 & H \end{pmatrix}.$$

This property has often been used to define the differentiation index, cf. [3], [2], [16]. Since we consider an alternative characterization, we prove the equivalency below for completeness. The proof provides deep insights into the information that can be obtained from the SVD.

Theorem 2. *For $A \in \mathbb{R}^{m\times n}$, there are a nonsingular matrix $R \in \mathbb{R}^{m\times m}$ and a matrix $H \in \mathbb{R}^{(m-j)\times(n-j)}$ such that*

$$RA = \begin{pmatrix} I_j & 0 \\ 0 & H \end{pmatrix},$$

iff the unique orthogonal projector $Q \in \mathbb{R}^{n\times n}$ onto $\ker A$ *has the structure*

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}$$

for an orthogonal projector $T \in \mathbb{R}^{(n-j)\times(n-j)}$.

Proof. ($\Rightarrow$):
Since $\ker A = \ker RA$, Q has the block structure

$$Q = \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}.$$

This follows immediately from $RAQ = 0$ and $Q_2 = Q_3^T$ for the orthogonal projector T onto $\ker H$.

($\Leftarrow$):

With the singular value decomposition $A = U\Sigma V^T$ and $j \le r \le \min\{m,n\}$, the block matrix representation

$$V = \begin{pmatrix} V_{11} & V_{12} & V_{13} \\ V_{21} & V_{22} & V_{23} \\ V_{31} & V_{32} & V_{33} \end{pmatrix}, \quad V_{11} \in \mathbb{R}^{j\times j}, \quad V_{22} \in \mathbb{R}^{(r-j)\times(r-j)}, \quad V_{33} \in \mathbb{R}^{(n-r)\times(n-r)}$$

and (45)– (46) lead to

$$Q = \begin{pmatrix} V_{13} \\ V_{23} \\ V_{33} \end{pmatrix} \begin{pmatrix} V_{13}^T & V_{23}^T & V_{33}^T \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix}, \tag{47}$$

$$P = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \\ V_{31} & V_{32} \end{pmatrix} \begin{pmatrix} V_{11}^T & V_{21}^T & V_{31}^T \\ V_{12}^T & V_{22}^T & V_{32}^T \end{pmatrix}^T = \begin{pmatrix} I_j & 0 \\ 0 & I_{n-j} - T \end{pmatrix}. \tag{48}$$

Comparing the block matrices we obtain

$$T = \begin{pmatrix} V_{23} \\ V_{33} \end{pmatrix} \begin{pmatrix} V_{23}^T & V_{33}^T \end{pmatrix}, \quad 0 = V_{13} \in \mathbb{R}^{j\times(n-r)}$$

and

$$\begin{aligned}
I_j &= V_{11}V_{11}^T + V_{12}V_{12}^T,\\
0 &= V_{11}V_{21}^T + V_{12}V_{22}^T,\\
0 &= V_{11}V_{31}^T + V_{12}V_{32}^T,\\
0 &= V_{21}V_{11}^T + V_{22}V_{12}^T,\\
0 &= V_{31}V_{11}^T + V_{32}V_{12}^T,\\
I_{n-j} - T &= I_{n-j} - \begin{pmatrix} V_{23} \\ V_{33} \end{pmatrix} \begin{pmatrix} V_{23}^T & V_{33}^T \end{pmatrix} = \begin{pmatrix} V_{21} & V_{22} \\ V_{31} & V_{32} \end{pmatrix} \begin{pmatrix} V_{21}^T & V_{31}^T \\ V_{22}^T & V_{32}^T \end{pmatrix}.
\end{aligned}$$

We can then construct a nonsingular matrix R

$$R := \begin{pmatrix} V_{11} & V_{12} & 0 \\ V_{21} & V_{22} & 0 \\ 0 & 0 & I_{m-r} \end{pmatrix} \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & I_{m-r} \end{pmatrix} U^T \in \mathbb{R}^{m\times m}.$$

For this particular R it holds

$$\begin{aligned}
RA &= \begin{pmatrix} V_{11} & V_{12} & 0 \\ V_{21} & V_{22} & 0 \\ 0 & 0 & I_{m-r} \end{pmatrix} \underbrace{\begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & I_{m-r} \end{pmatrix}}_{m\times m} \underbrace{U^T U}_{=I_m} \underbrace{\begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix}}_{m\times n} V^T\\
&= \begin{pmatrix} V_{11} & V_{12} & 0 \\ V_{21} & V_{22} & 0 \\ 0 & 0 & I_{m-r} \end{pmatrix} \underbrace{\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}}_{m\times n} V^T\\
&= \begin{pmatrix} V_{11} & V_{12} & 0 \\ V_{21} & V_{22} & 0 \\ 0 & 0 & I_{m-r} \end{pmatrix} \underbrace{\begin{pmatrix} V_{11}^T & V_{21}^T & V_{31}^T \\ V_{12}^T & V_{22}^T & V_{32}^T \\ 0 & 0 & 0 \end{pmatrix}}_{m\times n}\\
&= \begin{pmatrix} I_j & 0 & 0 \\ 0 & H_{11} & H_{12} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} I_j & 0 \\ 0 & H \end{pmatrix}
\end{aligned}$$

for $H_{11} = V_{21}V_{21}^T + V_{22}V_{22}^T$, $\quad H_{12} = V_{21}V_{31}^T + V_{22}V_{32}^T$. □

Corollary 3. *If, for a matrix $A \in \mathbb{R}^{m\times n}$ with $A = U\Sigma V^T$, we denote the largest integer fulfilling*

$$V_{13} = V(1:r, n-j+1:n) = 0, \tag{49}$$

by j, then exactly the first j components of any consistent system of equations $Ax = b$ with $b \in \mathbb{R}^m$ are uniquely determined.

Proof. With the notation from above we obtain, according to (47),

$$Q = \begin{pmatrix} V_{13} \\ V_{23} \\ V_{33} \end{pmatrix} \begin{pmatrix} V_{13}^T & V_{23}^T & V_{33}^T \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & T \end{pmatrix},$$

whereas $0 = V_{13} \in \mathbb{R}^{r \times j}$. □

In practice, the basis of the orthogonal projector Q may also be computed by the QR-decomposition of A^T.

For DAEs the following definition for block matrices is often considered.

Definition 2. *(see e.g. [3], [2], [16])*
A block matrix $A \in \mathbb{R}^{kn \times ln}$ is called 1-full (with respect to the block structure built from $n \times n$-matrices) if there exists a nonsingular matrix $R \in \mathbb{R}^{kn \times kn}$ such that

$$RA = \begin{pmatrix} I & 0 \\ 0 & H \end{pmatrix}$$

for the identity matrix $I \in \mathbb{R}^{n \times n}$ and a matrix $H \in \mathbb{R}^{k(n-1) \times j(n-1)}$.

For our purposes we prefer the following characterization. The equivalence follows directly from Theorem 2. We formulate it more generally for s-*full* matrices, cf. [3], [2].

Definition 3. *A block matrix $A \in \mathbb{R}^{kn \times ln}$ is s-full (with respect to the block structure built from $n \times n$-matrices) iff*

$$\ker A \subseteq \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_l \end{pmatrix}, \; x_i \in \mathbb{R}^n \quad i = 1, \ldots, l \; : \quad x_i = 0 \quad i = 1, \ldots, s \right\}.$$

From Theorem 2 we obtain an elegant criterion to check the s-fullness of a matrix.

Corollary 4. *If, for a block matrix $A \in \mathbb{R}^{kn \times ln}$, the index $j \geq 1$ from* (49) *is given, then A is s-full for all $s \in \mathbb{N}$ fulfilling $s \cdot n \leq j$.*

References

[1] W. Alt. *Nichtlineare Optimierung. Eine Einführung in Theorie, Verfahren und Anwendungen.* Wiesbaden: Vieweg+Teubner, 2011.

[2] 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.

[3] S. L. Campbell. The numerical solution of higher index linear time varying singular systems of differential equations. *SIAM J. Sci. Stat. Comput.*, 6:334–348, 1985.

[4] 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.

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

[6] S. L. Campbell and R. Hollenbeck. Automatic differentiation and implicit differential equations. M. Berz (ed.) et al., Computational differentiation: techniques, applications, and tools. Proceedings of the second international workshop on computational differentiation, February 12–14, 1996. Philadelphia, PA: SIAM. 215-227, 1996.

[7] I.S. Duff and C.W. Gear. Computing the structural index. *SIAM J. Algebraic Discrete Methods*, 7:594–603, 1986.

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

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

[10] D. Estévez Schwarz and R. Lamour. The computation of consistent initial values for nonlinear index-2 differential-algebraic equations. *Numer. Algorithms*, 26(1):49–75, 2001.

[11] D. Estévez Schwarz and R. Lamour. A new projector based decoupling of linear DAEs for monitoring singularities. *Numer. Algorithms*, 73(2):535–565, 2016.

[12] D. Estévez Schwarz and R. Lamour. Consistent initialization for higher-index DAEs using a projector based minimum-norm specification. Technical

Report 1, Institut für Mathematik, Humboldt-Universität zu Berlin, 2016. http://www2.mathematik.hu-berlin.de/publ/pre/2013/P-2016-01-2.pdf.

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

[14] 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.

[15] D. Kraft. A software package for sequential quadratic programming. tech. rep. DFVLR-FB 88-28,. Technical report, DLR German Aerospace Center Institute for Flight Mechanics, Köln, Germany, 1988.

[16] P. Kunkel and V. Mehrmann. *Differential-Algebraic Equations - Analysis and Numerical Solution*. EMS Publishing House, Zürich, Switzerland, 2006.

[17] P. Kunkel, V. Mehrmann, W. Rath, and J. Weickert. A new software package for linear differential-algebraic equations. *SIAM J. Sci. Comput.*, 18(1):115–138, 1997.

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

[19] S.E. Mattsson, H. Olsson, and H. Elmqvist. Dynamic selection of states in Dymola. In *Modelica Workshop 2000, Oct 23-24, Lund, Sweden*, pages 61–67, 2000.

[20] 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.

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

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