Linear index-1 DAEs: regular and singular problems

Ricardo Riaza*
Departamento de Matemática Aplicada a las Tecnologías de la Información
Escuela Técnica Superior de Ingenieros de Telecomunicación
Universidad Politécnica de Madrid - 28040 Madrid, Spain
rrr@mat.upm.es

Roswitha Mähr**
Institut für Mathematik, Humboldt-Universität zu Berlin
Unter den Linden, 6 - 10099 Berlin, Germany
maerz@mathematik.hu-berlin.de

Abstract

Several features and interrelations of projector methods and reduction techniques for the analysis of linear time-varying differential-algebraic equations (DAEs) are addressed in this work. The application of both methodologies to regular index-1 problems is reviewed, leading to some new results which extend the scope of reduction techniques through a projector approach. Certain singular points are well accommodated by reduction methods; the projector framework is adapted in this paper to handle (not necessarily isolated) singularities in an index-1 context. The inherent problem can be described in terms of a scalarly implicit ODE with continuous operators, in which the leading coefficient function does not depend on the choice of projectors. The nice properties of projectors concerning smoothness assumptions are carried over to the singular setting. In analytic problems, the kind of singularity arising in the scalarly implicit inherent ODE is also proved independent of the choice of projectors. The discussion is driven by a simple example coming from electrical circuit theory. Higher index cases and index transitions are the scope of future research.

Keywords: Differential-algebraic equation, index, singularity.

AMS subject classification: 34A09, 34A30.

*The work of this author was performed during a research stay at the Humboldt-Universität zu Berlin, granted by Secretaría de Estado de Educación y Universidades, Ministerio de Educación, Cultura y Deporte, Spain.
**Supported by the DFG research center ”Mathematics for key technologies” (FZT 86) in Berlin
1 Motivation

Consider the linear electrical circuit depicted in Figure 1. The resistor in this figure is assumed to be voltage-controlled, possibly in a time-varying way. Similarly, we are assuming that the magnetic flux \( \phi \) in the inductor has a time-dependent, current-controlled description. The current in the source \( i_s(t) \) is a known function of time.

![RL circuit](image)

Figure 1: RL circuit.

Modified Nodal Analysis (MNA) \([4, 6]\) uses the node voltage \( e \) and the current \( i \) in the inductor as circuit variables, and sets up the network equations in differential-algebraic form:

\[
\begin{align*}
    L(t)i'' + L'(t)i - e &= 0 \quad (1a) \\
    i + g(t)e &= i_s(t), \quad (1b)
\end{align*}
\]

where \( l(t) \) is assumed to be differentiable, and \( t \) takes values in some interval \( I \) which may well in this case be the entire \( \mathbb{R} \). This is a particularly simple instance of linear MNA equations, which would read, for a linear time-varying circuit containing an arbitrary number of resistors, inductors, capacitors, and both current and voltage sources,

\[
\begin{align*}
    L(t)i'' + L'(t)i - A_L^T e &= 0 \quad (2a) \\
    A_C C(t) A_C^T e' + A_L i + [A_C C'(t) A_C^T + A_R G(t) A_R^T] e + A_V i_V &= -A_R i_s(t) \quad (2b) \\
    A_V^T e &= v_s(t), \quad (2c)
\end{align*}
\]

the constant matrix \((A_R \ A_L \ A_C \ A_V \ A_I)\) describing the incidence between branches and nodes in the circuit.

Example (1) is intended to illustrate some of the features which may be expected in larger-scale problems (2) or, more generally, in linear time-varying (LTV) DAEs

\[
A(t)x' + B(t)x = b(t), \quad (3)
\]

where \( A, B : I \to \mathbb{R}^{n \times n} \), \( \text{rk} A < n \) for all \( t \), and \( b : I \to \mathbb{R}^n \); \( I \) is a real interval, all concepts being meant laterally if it is closed or half-closed. Smoothness assumptions will be extensively addressed later. The reader is referred to \([1, 8, 15, 18, 20]\) for background on linear time-varying DAEs, and specially to \([2, 3, 12]\) for canonical forms not addressed here.
The first aim of this paper, tackled in Section 2, is to discuss some features and interrelations of reduction methods [18, 19, 20] and projector techniques [8, 15, 17] for the analysis of (3). Driven by example (1), the specific goals of this study are introduced below.

Roughly speaking, the index of (1) would rely on $g(t)$ and, in a slightly more subtle way, on $l(t)$. If $g(t) \neq 0 \ \forall t \in \mathcal{I}$, we may insert $e = (i_s - i)/g$ into (1a) to get an ODE, which may be written in explicit form if $l(t) \neq 0 \ \forall t \in \mathcal{I}$. This would be a reduction of the DAE, which would have index 1 under these assumptions. Note that the construction of such a simple reduction has been based on the availability of an explicit constraint in (1b), and also on the non-vanishing of $g$ in the whole interval of study. A generalization of this reduction procedure to analytic linear time-varying DAEs (3) were introduced by Rabier and Rheinboldt in [18]. Non-analytic problems are considered briefly there and in detail in [20, Chapter 2], whereas discontinuous inputs and inconsistent initial conditions are discussed within a distributional framework in [19] and [20, Chapter 3]. This reduction procedure will be reviewed in 2.2 for index-1 problems. Concerning (1), $g$ and (in principle) the excitation $i_s$ would need to be $C^1$, whereas $l$ should be required to be $C^2$ to yield $C^1$ coefficients.

Projector methods based upon the tractability index [8, 15, 17], in contrast to the above-mentioned reduction technique and also to the differential index approach [1], are aimed at minimal smoothness requirements on both the operators $A$, $B$ and the excitation $b$ in (3). In the present example, both $g$ and $i_s$ would need only be continuous in the projector framework. The function $l$ would only be assumed $C^1$ in order to guarantee $C^0$ coefficients in the problem. These techniques unveil the “essential” ODE in a different manner, using a decoupling based on an inherent ODE without dimensional reduction. Projector methods are reviewed for index-1 problems in 2.3, and new results concerning their relation with the above-mentioned reduction technique are discussed in 2.4.

In a certain sense, things become more complicated if $l(t)$ or $g(t)$ vanish at some or all points. If $l(t) \neq 0$, $\forall t \in \mathcal{I}$, and $g(t)$ vanishes everywhere (that is, the resistor is open-circuited), we would be led to an index-2 DAE. In circuit-theoretic terms, this index-2 configuration owes to the presence of a cut-set formed exclusively by a current source and an inductor [6]. In contrast, cases in which $g(t)$ vanishes only at “some” (in a below-specified sense) points yield singularities or critical points. With additional complexity, this is also the case for points where $l(t)$ vanish. These singular problems are addressed in Section 3.

Singularities of both types are well accommodated by the reduction method, at least in analytic problems. In contrast, projector techniques require some modifications in order to handle singularities. This is the ultimate goal of Section 3, where singular points will be shown to lead to a scalarly implicit decoupling under certain hypotheses. These comprise the assumption that both $g$ and $l$ do not vanish in an open dense subset of $\mathcal{I}$; this holds in particular for analytic problems, which may only yield isolated singularities, but it also allows for non-isolated singularities (consider e.g. $l(t) = 1$, $g(t) = t \sin(1/t)$ or vice-versa), which seem to rule out in practice an interval-wise application of methods as defined for regular problems. The main structural properties at singular points will be shown to hold independently of the choice of projectors, and the benefits of projector techniques concerning smoothness will be carried over to the singular setting.
2 Regular problems

2.1 Introduction

It is well-known that, given a real interval $I$, $t_0 \in I$, and $x_0 \in \mathbb{R}^n$, a linear time-varying ODE $x' + B(t)x = b(t)$, with $B(t) \in C^k(I, \mathbb{R}^{n \times n})$, $k \geq 0$, behaves as a smoothing operator

$$C^k(I, \mathbb{R}^n) \to C^{k+1}(I, \mathbb{R}^n)$$

$$b(t) \to x(t)$$

(4)

where $x(t)$ is the unique solution of the ODE excited by $b(t)$ which satisfies $x(t_0) = x_0$. This map actually yields a bijection between $C^k \times \mathbb{R}^n$ and $C^{k+1}$, defined as $(b, x_0) \to x$, with inverse given by $x \to (x' + Bx, x(t_0))$. This functional analysis point of view is important for instance in electrical engineering, where $b(t)$ and $x(t)$ are viewed as input and output signals of a system defined by $B(t)$. The assumptions on the system supporting these transformations, as well as the function spaces in which these operators act (including where they behave bijectively), should be precisely characterized.

Now consider a linear, constant coefficient (time-invariant) DAE

$$Ax' + Bx = b(t)$$

(5)

with Kronecker index 1 [1, 7]. It is well-known (see e.g. [1]) that, if $G$ and $H$ are non-singular matrices such that \{GAH, GBH\} defines the Kronecker normal form of the pencil $\lambda A + B$, then premultiplying by $G$ and with the coordinate change $x = Hz$, (5) is transformed into

$$\dot{z}_1' + Kz_1 = f_1(t)$$

(6a)

$$z_2 = f_2(t),$$

(6b)

for some $r \times r$ matrix $K$, $r = \text{rk} A$. Since (6a) is a linear ODE, given $t_0$ and $z_1(t_0) \in \mathbb{R}^r$, every continuous $b$ yields a continuous $f = GB$ and hence defines a unique solution $z$ of (6). We may (tentatively) conclude that this leads to a solution $x = Hz$ of (5). But, since $x = Hz$ needs not (and will actually not, if $f_2$ is not differentiable) be differentiable, it would not make sense to differentiate $x$ in (5).

This may be overcome by rewriting $Ax' = (Ax)'$, leading to the reformulation

$$(Ax)' + Bx = b(t),$$

(7)

which encompasses all $C^1$ solutions of (5) but allows for the search of additional solutions in the larger space $C^1_A(I, \mathbb{R}^n) = \{x \in C^0(I, \mathbb{R}^n) / Ax \in C^1(I, \mathbb{R}^n)\}$. Continuous excitations $b$ are now well accommodated, since $GAH = \text{blockdiag}\{I_r, 0\}$ yields $Ax = G^{-1} \ddot{z}_1$, where $\ddot{z}_1 = (z_1 \cdot 0_{n-r})^T$, and therefore smoothness of $Ax$ relies only on that of $z_1$.

If $b$ is $C^1$, or actually $C^k$ for $k \geq 1$, then there is no need for the reformulation (7) in order to guarantee that solutions are $C^k$. But, more accurately, they can be shown to lie on

$$C^{k+1}_A(I, \mathbb{R}^n) = \{x \in C^k(I, \mathbb{R}^n) / Ax \in C^{k+1}(I, \mathbb{R}^n)\}.$$
Hence, fixing \( t_0 \) and assuming, for notational simplicity below, that a consistent initial value \( x_0 \in M_b(t_0) = \{ x \in \mathbb{R}^n / Bx - b(t_0) \in \text{Im}A \} \) is given for every \( b \), we may claim that the reformulation (7) of the index 1 LTI DAE (5) yields, for any \( k \geq 0 \), an injection
\[
C^k(\mathcal{I}, \mathbb{R}^n) \rightarrow C^k\mathcal{A}^{-1}(\mathcal{I}, \mathbb{R}^n) \subset C^k(\mathcal{I}, \mathbb{R}^n)
\]
\[
b(t) \rightarrow x(t).
\] (8)

As for ODEs, but with some additional technicalities, a bijection from \( \bigcup_{b \in C^k} \{ b \} \times M_b(t_0) \) onto a subspace of \( C^{k+1}_A \) can now be naturally constructed, using the fact that consistent initial values force solutions to lie pointwise on \( \{ x \in \mathbb{R}^n / Bx - b(t) \in \text{Im}A \} \). Now the inverse relies upon \( x \rightarrow (Ax)' + Bx \), which is naturally \( C^k \) if \( x \in C^{k+1}_A \).

Analysis methods for linear time-varying DAEs (3) under an index-1 assumption should in some sense extend the injection depicted in (8), including the case \( k = 0 \). In this direction, the reduction method of [18, 20] for time-varying DAEs is surveyed in 2.2 for index-1 cases. This nice, technically simple method yields an injection \( C^k \rightarrow C^k \) for \( k \geq 1 \), under \( C^k \) assumptions on the matrix pair defining the DAE.

Projector methods [8, 15, 17] are technically more cumbersome but allow us to inject \( C^0 \rightarrow C^k_P \subset C^0 \), for certain operator \( P(t) \), making it possible to handle \( C^0 \) excitations. If \( A(t), B(t) \) are \( C^k \), with \( k \geq 1 \), the injection \( C^k \rightarrow C^k \) is also achieved, and can be improved to \( C^k \rightarrow C^{k+1}_P \subset C^k \) under an additional assumption; this preserves the nice inversion property sketched above. The projector approach for index-1 cases is reviewed in 2.3.

Some new results concerning interrelations between both approaches are discussed in 2.4. Specifically, we show that the reduced ODE of [18, 20] can be obtained, in index-1 cases, from the restriction of the (projector) inherent ODE to an invariant subspace. This result makes it possible to relax some smoothness assumptions in the reduction approach, and provides a simple, explicit criterion for the choice of the “reduced” excitation which only requires \( b(t) \) to be \( C^0 \). In the above-presented terms, the scope of the reduction technique would be extended up to define the same injection \( C^0 \rightarrow C^1_P \subset C^0 \) obtained with projectors, relying on slightly stronger assumptions than those of the projector framework.

Other techniques requiring more smoothness on the operators (see [1, 2, 3]) will not be addressed here.

2.2 The reduction method of Rabier and Rheinboldt

The reduction technique presented in [20, Chapter 2] for \( C^\infty \) linear time-varying DAEs (3), as acknowledged by the authors, may be applied to \( C^k \) problems, \( k \) being the (eventually defined) index of the problem. Although the method can be phrased in a coordinate-free form (see [18, 20]), let us discuss it here assuming that \( A \) and \( B \) are time-dependent \( n \times n \) matrices. We restrict the attention to index-1 cases.

The reduction is supported on the assumption that \( \{ A(t), B(t) \} \) is a regular pair of matrices in a given open interval \( \mathcal{I} \) in the sense that, for some fixed \( r \leq n \),
\[
\text{rk} A(t) = r, \forall t \in \mathcal{I},
\] (9a)
\[
\text{rk} (A(t) B(t)) = n, \forall t \in \mathcal{I},
\] (9b)
where \((A, B)\) stands for the matrix in \(\mathbb{R}^{n \times 2n}\) constructed by joining the columns of \(A\) and \(B\).

Assume that both \(A(t), B(t)\) are \(C^1(\mathcal{I}, \mathbb{R}^{n \times n})\), and that \(r < n\). From the constant rank assumptions (9), it follows that the spaces \(\text{Im}A(t)\) and

\[
S_0(t) = \{x \in \mathbb{R}^n : B(t)x \in \text{Im}A(t)\}
\]

are also \(C^1\), that is, they are spanned by \(C^1\) bases. This follows from the fact that the constant rank assumption in \(A(t)\) guarantees the existence of a \(C^1\), constant (maximal) rank map \(W(t) : \mathcal{I} \to \mathbb{R}^{(n-r) \times n}\) such that \(v \in \text{Im}A(t) \iff W(t)v = 0\). In turn, \(W(t)B(t)\) has constant rank \(n-r\) due to (9b) and the property \(WA = 0\). Hence, the set \(S_0(t)\) may be described as the \(C^1\) space \(\text{Ker} W(t)B(t)\), which is \(r\)-dimensional for every fixed \(t\).

From the smoothness of \(\text{Im}A(t)\) and \(S_0(t)\), we are allowed to choose \(C(t) \in C^1(\mathcal{I}, \mathbb{R}^{n \times r}), D(t) \in C^1(\mathcal{I}, \mathbb{R}^{r \times n})\) such that, for every \(t \in \mathcal{I}\), \(C(t)\) and the restriction of \(D(t)\) to \(\text{Im}A(t)\) yield isomorphisms \(\mathbb{R}^r \leftrightarrow S_0(t)\) and \(\text{Im}A(t) \leftrightarrow \mathbb{R}^r\), respectively.

**Definition 1.** Let \(\{A(t), B(t)\}\) be a regular, \(C^1\) pair in \(\mathcal{I}\), and assume that \(C\) and \(D\) are defined as indicated above. The pair

\[
E_1 = DAC, \quad F_1 = D(BC + AC'),
\]

is called a reduction of \(\{A, B\}\) in \(\mathcal{I}\).

The pair \(\{A, B\}\) is said to have index one if \(E_1\) is non-singular in \(\mathcal{I}\).

The index-1 notion is proved in [18, 20] to be independent of the choice of \(C\) and \(D\). If \(E_1\) is singular, under additional regularity and smoothness assumptions the same procedure can be applied to \(\{E_1, F_1\}\). If the procedure can be continued indefinitely, it will eventually become stationary, and the number of reduction steps defines the index: see [18, 20].

A reduction of (3) requires also the construction of a “reduced” excitation

\[
b_1 = D(b - Bu_0 - Au_0'),
\]

constructed from some \(u_0\) verifying pointwise

\[
Bu_0 - b \in \text{Im}A.
\]

The existence of a continuously differentiable \(u_0\) satisfying this is guaranteed from the fact that (13) holds in particular for \(u_0 = B^T(AAT + BB^T)^{-1}b\). In this setting, classical solutions of a DAE (3) with an index-1, \(C^1\) pair \(\{A(t), B(t)\}\) and a \(C^1\) excitation \(b(t)\) can be recovered from those of the reduced ODE

\[
E_1(t)y' + F_1(t)y = b_1(t),
\]

through the relation

\[
x = Cy + u_0.
\]
This approach provides a nice framework for the analysis of index-1 DAEs under $C^k$ assumptions, $k \geq 1$. As stated above, and in the terms presented in 2.1, it would provide a $C^k \to C^k$ injection between excitations and solutions, for $k \geq 1$. It is technologically simple, and it is possible to extend it recursively to higher index cases without additional complexity. Besides this, these techniques have been extended to deal with rectangular systems, as well as with generalized solutions. It is also worth remarking that this approach has found successful extensions regarding nonlinear and singular phenomena in implicit DAEs, addressed by the same authors within a more general geometric framework.

Nevertheless, this approach presents some limitations regarding its scope of application. As it happens with the differential index approach, the $C^1$ assumption on $A$, $B$ rules out the use of these techniques e.g. in semiexplicit systems with a right-hand side depending only continuously in $t$, which often arise in electrical circuit analysis or optimal control problems. At least a $C^3$ structure in $S_0(t)$ seems to be an essential smoothness requirement for the reduction method, in the light of the derivative $C'$ in (11).

From our point of view, another important limitation of the reduction procedure as stated in [18, 20] is the $C^1$ requirement on $b$ and, specially, the explicit appearance of the derivative $u'_t$ in the expression for $b_1$ depicted in (12). This might lead to the erroneous interpretation that solutions of LTV index-1 problems actually depend on the first derivative of the excitation, against common understanding in the framework of the perturbation index [9]. As shown in Theorem 1 in 2.4, this reduced excitation can be written without the need for any derivation of $b$; this result will extend the scope of the reduction technique to index-1 DAEs excited by just continuous functions.

**Reduction of the RL circuit equations.** In order to apply the reduction as stated above to (1), $g(t)$ and $l(t)$ must be assumed $C^1$ and $C^2$, and we also assume for the moment that $i_s(t)$ is $C^1$. Write (1) in the form (3) with

$$x = \begin{pmatrix} i \\ e \end{pmatrix}, \quad A(t) = \begin{pmatrix} l(t) & 0 \\ 0 & 0 \end{pmatrix}, \quad B(t) = \begin{pmatrix} l'(t) & -1 \\ 1 & g(t) \end{pmatrix}, \quad b(t) = \begin{pmatrix} 0 \\ i_s(t) \end{pmatrix}. \quad (16)$$

If $l(t) \neq 0$, $\forall t \in \mathcal{I}$, the matrix pair is regular, and $\text{Im} A(t)$ and $S_0(t)$ are spanned by $(1 \ 0)^T$ and $(g(t) - 1)^T$, respectively, allowing us to choose

$$C(t) = \begin{pmatrix} g(t) \\ -1 \end{pmatrix}, \quad D(t) = \begin{pmatrix} 1 & 0 \end{pmatrix}. \quad (17)$$

The reduction of $\{A, B\}$ arising from these choices may be easily checked to be

$$E_1 = \begin{pmatrix} l(t)g(t) \end{pmatrix}, \quad F_1 = \begin{pmatrix} 1 + (l(t)g(t))^T \end{pmatrix}.$$  

Assuming additionally that $g(t) \neq 0$ in $\mathcal{I}$, the non-singularity of $E_1$ indicates that the problem has index 1.

A simple way to satisfy (13) is to choose $u_0 = (i_s \ 0)^T$, since then $Bu_0 - b = (l_i \ 0)^T \in \text{Im} A(t)$. With this choice we get, after simple computations, $b_1 = D(b - Bu_0 - Au_0) = \ldots$
\((-l(t)s)'\). The reduced ODE (14) then reads
\[
 l(t)g(t)y' + (1 + (l(t)g(t))')y = -(l(t)i_s(t))',
\]
solutions of the DAE being recovered from those of (18) through
\[
 \begin{pmatrix}
 i \\
 e \\
 \end{pmatrix} = \begin{pmatrix}
 g(t) \\
 -1 \\
 \end{pmatrix} y + \begin{pmatrix}
 i_s(t) \\
 0 \\
 \end{pmatrix}.
\]
Remark that the second derivative of \(l(t)\) is not needed at all, and that the first derivative \(l'(t)\) owes to electrical reasons and already appears in the formulation of the problem (1).

From the remarks concerning \(S_0(t)\) in the discussion following (15), problems with just continuous \(g(t)\) are unavoidably excluded from the scope of the method, in agreement with the appearance of \(g'(t)\) in (18). Related issues apply to the excitation \(i_s(t)\): in this case, a particular choice of \(u_0\) may rule out the need for a \(C^1\) assumption on the excitation, as shown in 2.4.

Cases in which either \(g\) or \(l\) vanish at some points, and hypothetical implications of the rank deficiencies arising in \(A(t)\), are postponed until Section 3.

### 2.3 The tractability index and projector methods

Projector techniques (see [8, 15, 17] and references therein) are different in spirit to the above-presented reduction method. They are aimed at minimal smoothness requirements in both the operators and the solutions, and provide an index criterion through the construction of a matrix chain without the need to unveil any essential ODE. This essential ODE can be realized through a decoupling procedure which does not lead to a reduced ODE but to a so-called inherent ODE, with the same dimensionality as the original problem. The solutions can be recovered from the restriction of the inherent ODE to an invariant space.

Let \(A, B\) and \(b\) in (3) be continuous functions of time. As the unique smoothness assumption, let the space \(\text{Ker}A(t)\) be \(C^1\), that is, assume the existence of \(n - r\) \(C^1\) functions spanning \(\text{Ker}A(t)\) in \(\mathcal{I}\). This implies, in particular, the constant rank condition (9a). Conversely, constant rank in a \(C^1\) matrix \(A(t)\) yields a \(C^1\) Kernel. The \(C^1\) structure on \(\text{Ker}A(t)\) is equivalent to the existence of a \(C^1\) projector \(P(t)\) along \(\text{Ker}A(t)\), that is, a matrix depending smoothly on \(t\) and verifying \(P^2(t) = P(t), \text{Ker}P(t) = \text{Ker}A(t)\) for all \(t\) in \(\mathcal{I}\). Then \(Q(t) = I - P(t)\) is a projector onto \(\text{Ker}A(t)\).

Removing dependences with \(t\), and from the property \(AP = A\), (3) can be rewritten as
\[
 A(Px)' + B_0x = b,
\]
with \(B_0 = B - AP'\), allowing for the search of solutions in the space
\[
 C^1_p(\mathcal{I}, \mathbb{R}^n) = \{ x \in C^0(\mathcal{I}, \mathbb{R}^n) / Px \in C^1(\mathcal{I}, \mathbb{R}^n) \},
\]
which comprises \(C^1(\mathcal{I}, \mathbb{R}^n)\) as a proper subset. Note that \(C^1_p = C^1_R\) for any other smooth projector \(R\) along \(\text{Ker}A\), due to the property \(PR = P, RP = R\) of projectors along the same space (Lemma 2, item 4 in the Appendix).
\textbf{Definition 2.} Let $\{A, B\}$ be a $C^0$ pair. Assume that \text{Ker}A(t) is $C^1$ in $\mathcal{I}$, and that $P$, $Q$, $B_0$ are defined as indicated above. The pair $\{A, B\}$ is regular with tractability index-1 if

$$A_1(t) = A(t) + B_0(t)Q(t)$$

is non-singular in $\mathcal{I}$.

Several issues concerning higher-index extensions, independence of the index on the choice of projectors, and improved formulations, can be found in [8, 15, 17] and the bibliography therein. Let us just remark here that the index-1 notion is equivalent, under the same assumptions, to the non-singularity of the “simpler” matrix $A(t) + B(t)Q(t)$, and also to the property $\mathbb{R}^n = \text{Ker}A(t) \oplus S_0(t)$, the latter being defined in (10). Equivalence with the index-1 notion of Definition 1 in $C^1$ problems with constant rank in $A$ follows from [18, Theorem 7.1].

The decoupling of the DAE can be performed by premultiplying the equation by $A_1^{-1}$ and decomposing along the spaces $\text{Im}P$ and \text{Ker}A. Details are skipped, since this derivation can be seen as a particular case of the one detailed for Theorem 3 in 3.2. The decoupling procedure ends up with an inherent ODE

$$u' + (-P' + PA_1^{-1}B)u = PA_1^{-1}b,$$

with $u \in \mathbb{R}^n$, for which the space $\text{Im}P(t)$ is invariant. Solutions of the reformulated DAE (20) define an $r$-dimensional affine subspace of $C_P^1(\mathcal{I}, \mathbb{R}^n)$, and $x$ is a solution of (20) in $C_P^1$ if and only if it can be written as

$$x = P_cu + QA_1^{-1}b,$$

where $u$ is a $C^1$ solution of (23) lying on $\text{Im}P(t)$, whereas $P_c = (I - QA_1^{-1}B)$ is the so-called \textit{canonical projector} along $\text{Ker}A(t)$ onto $S_0(t)$ (see e.g. [16]). This solution can be also written as $x = u + v$, where $v$ is $C^0$ and satisfies

$$v = -QA_1^{-1}Bu + QA_1^{-1}b.$$  \hspace{1cm} (25)

This approach yields a $C^0 \rightarrow C_P^1 \subset C^0$ injection between excitations $b$ and solutions $x$, as explained in 2.1, provided that consistent initial values are given.

Furthermore, and always under the index-1 assumption, if $A$, $B$ and $b$ are $C^k$, for $k \geq 1$, then taking $P \in C^k$ yields $C^{k-1}$ operators in the inherent ODE (23), so that $u$ is $C^k$. Also, $\text{Ker}A$ and $S_0$ are $C^k$ spaces, and accordingly so it is $P_c$. From Lemma 4, $QA_1^{-1}$ is also $C^k$. It follows from all these properties that solutions $x$ are $C^k$, yielding an injection $C^k \rightarrow C^k$.

If, additionally, $\text{Ker}A$ is $C^{k+1}$ (and $P$ is accordingly taken from $C^{k+1}$), denoting

$$C_{P}^{k+1}(\mathcal{I}, \mathbb{R}^n) = \{x \in C^k(\mathcal{I}, \mathbb{R}^n) / Px \in C^{k+1}(\mathcal{I}, \mathbb{R}^n)\},$$

an injection $C^k \rightarrow C_P^{k+1} \subset C^k$ naturally follows from the $C^k$ nature of operators in (23) and the property $u = Px$, generalizing the injection mentioned above for the case $k = 0$. 

8
One drawback of projector methods is that the derivation of the decoupling involves some technical computations, which become increasingly intricate in higher-index cases. Despite this, the projector approach has been successfully applied to different problems concerning e.g. circuit theory, numerical methods, optimal control, stability theory, or stochastic DAEs (see references in [6, 11, 13, 14, 22], respectively). Actually, posing the smoothness requirement only on Ker $A(t)$ makes it possible to apply this approach to systems defined by a constant $A$ (in particular, semiexplicit systems) and continuous, not necessarily differentiable $B(t), b(t)$. Problems in this category are Euler-Lagrange equations coming from linear-quadratic time-varying $C^0$ optimal control, which are semiexplicit, or charge-oriented linear MNA models. Systems with actually time-varying, continuous $A(t)$ displaying a smooth Kernel often arise in applications such as conventional MNA analysis of LTV circuits. Within the reduction framework, all these problems would require $C^1$ assumptions on $A(t)$ and $B(t)$, or at least on the space $S_0(t)$, and theoretically on the excitation $b(t)$, although this could be removed according to the results presented in 2.4.

**Projector-based analysis of the RL circuit equations.** Smoothness of Ker $A(t)$ in (16) relies only on the non-vanishing of $l(t)$. Hence, in the projector framework we just need to assume that $g(t)$ and $i_s(t)$ are continuous, $l(t)$ being $C^1$ for electrical reasons supporting the model (1). The Kernel of $A(t)$ is spanned by $(0\ 1)^T$, so that a natural choice for the projectors is
\[
P(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad Q(t) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.
\]

The constant nature of $P$ yields $B_0 = B$, and then
\[
A_1(t) = \begin{pmatrix} l(t) & -1 \\ 0 & g(t) \end{pmatrix}.
\]

Again, the assumptions $l(t) \neq 0, \ g(t) \neq 0$ in $I$ yield a (tractability) index-1 DAE. The inherent equation (23) reads, after some computations,
\[
\begin{pmatrix} u'_1 \\ u'_2 \end{pmatrix} + \frac{1}{l(t)g(t)} \begin{pmatrix} 1 + l'(t)g(t) & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \frac{1}{l(t)g(t)} \begin{pmatrix} i_s(t) \\ 0 \end{pmatrix}.
\]

The space Im$P(t)$ spanned by $(1\ 0)^T$ is obviously invariant for the inherent ODE, since $u'_2 = 0$. The solutions of the DAE can be then obtained from $u = (u_1\ 0)^T, u_1$ satisfying
\[
u'_1 + \frac{1 + l'(t)g(t)}{l(t)g(t)} u_1 = \frac{i_s(t)}{l(t)g(t)},
\]
through the relation (24), which reads in this case
\[
\begin{pmatrix} i \\ e \end{pmatrix} = \frac{1}{g(t)} \begin{pmatrix} g(t) & 0 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ 0 \end{pmatrix} + \frac{1}{g(t)} \begin{pmatrix} 0 \\ i_s(t) \end{pmatrix}.
\]

Note that, in contrast to (18) within the above-discussed reduction, the present approach and, in particular, equation (28) do not involve any derivative of $g$ or $i_s$. As indicated above, the first derivative $l'$ appears exclusively for electrical reasons.
2.4 A hybrid approach

As suggested in 2.2, the reduction procedure for index-1 LTV DAEs should be expected to somehow remain valid without (explicit or implicit) recourse to derivatives of \( b \), holding for just continuous excitations. A combination of projection and reduction techniques, relying on a \( C^1 \) structure in both \( \text{Ker} A(t) \) and \( S_0(t) \), proves that this is actually the case, as shown in Theorem 1 below. It is based on particular choices of \( u_0, b_1 \), and allows for solutions in the space \( C^1_p(I, \mathbb{R}^n) = \{ x \in C^0(I, \mathbb{R}^n) : P x \in C^1(I, \mathbb{R}^n) \} \supset C^1(I, \mathbb{R}^n) \), in the light of the reformulation (20). In particular, following Remark 1, the reduction procedure “works” for \( C^1 \) pairs with just continuous excitations.

Note that the smoothness of \( \text{Ker} A(t) \) in assumption A1 comprises a constant rank (denoted as \( r \)) condition on \( A(t) \), and that it supports both the tractability index-1 notion in A2 and the existence of the smooth projector \( P \) in b).

**Theorem 1.** Consider the linear time-varying DAE (3) in a given interval \( I \). Assume that:

A0. \( A(t), B(t) \in C^0(I, \mathbb{R}^{n \times n}), b(t) \in C^0(I, \mathbb{R}^n) \).

A1. Both \( \text{Ker} A(t) \) and \( S_0(t) = \{ x \in \mathbb{R}^n : B(t) x \in \text{Im} A(t) \} \) are spanned by \( C^1 \) bases in \( I \).

A2. The DAE has tractability index 1 in \( I \).

Let the following operators be given:

a) \( C(t) \in C^1(I, \mathbb{R}^{n \times r}) \) and \( D(t) \in C^0(I, \mathbb{R}^{n \times n}) \), defining isomorphisms \( \mathbb{R}^r \leftrightarrow S_0(t) \), \( \text{Im} A(t) \leftrightarrow \mathbb{R}^r \), respectively. Denote as \( \{ E_1, F_1 \} \) the reduction (11) defined by \( C, D \).

b) \( P(t) \in C^1(I, \mathbb{R}^{n \times n}) \) being a projector along \( \text{Ker} A(t) \), leading to the tractability matrix \( A_1 \) defined in (22). Denote \( Q(t) = I - P(t) \).

Define \( b_1 = D A A_1^{-1} b \), and consider the reduced ODE

\[
E_1(t) y' + F_1(t) y = b_1(t). \tag{30}
\]

Then, \( x \) is a solution of the reformulation (20) in \( C^1_p(I, \mathbb{R}^n) \) if and only if

\[
x = C y + u_0, \tag{31}
\]

where \( y \) is a \( C^1 \) solution of (30), and \( u_0 = Q A_1^{-1} b \).

Additionally, \( x \) is a \( C^1 \) solution of the original formulation (3) if and only if \( u_0 = Q A_1^{-1} b \) is \( C^1 \) and, as before, (31) holds for some \( C^1 \) solution \( y \) of (30).

**Remark 1.** Conditions A0, A1 and A2 in Theorem 1 hold in particular if

B0. \( A(t), B(t) \in C^1(I, \mathbb{R}^{n \times n}), b(t) \in C^0(I, \mathbb{R}^n) \).

B1. \( A(t) \) has constant rank in \( I \).

B2. The DAE has (tractability or reduction, equivalently) index 1 in \( I \).
Condition A0 holds trivially in the stronger setting defined by B0. In $C^1$ problems with constant rank in $A$, the reduction index-1 notion can be applied without explicitly requiring (9b), and is equivalent to the tractability one, as shown in [18, Theorem 7.1]. Hence, B2 makes sense for both tractability or reduction notions, and condition A2 of Theorem 1 holds under these assumptions. Indeed, (9b) follows from the tractability index-1 notion since $A + BQ$ is non-singular, so that smoothness of $S_0$ holds and, noting that $\text{Ker}A(t)$ is $C^1$ if $A$ is $C^1$ with constant rank, so it does condition A1 in Theorem 1. □

The proof of Theorem 1 relies upon the following Lemma, of interest in its own right. Some technical details can be found in the Appendix.

**Lemma 1.** Under the assumptions of Theorem 1, the reduced ODE (30) can be obtained from the restriction of the inherent ODE (23) to $\text{Im}P(t)$ through the $C^1$ non-singular transformation $u(t) = P(t)C(t)y(t)$ and premultiplication by $D(t)A_1(t)$.

**Proof:** $P(t)C(t)$ is non-singular in the sense that it defines an isomorphism $\mathbb{R}^r \to \text{Im}P$ for all $t$, owing to the property that $P(t)$ restricted to $\text{Im}C(t) = S_0(t)$ is itself an isomorphism onto $\text{Im}P$. This is in turn due to the identity $\text{Ker}P(t) = \text{Ker}A(t)$, and the property $\mathbb{R}^n = \text{Ker}A(t) \oplus S_0(t)$ characterizing tractability index 1.

Let us remove all dependences with $t$ and insert $u = PCy$ in (23), getting

$$(PCy)' + (-P' + PA_1^{-1}B)PCy = PA_1^{-1}b. \quad (32)$$

Now $(PCy)' = PCy' + (P'C + PC')y$. Since $P = PP$ implies $P' = P'P + PP'$, we may write $P'C - P'PC = PP'C$, and then (32) reads (reordering terms)

$$PCy' + ((P'P + PA_1^{-1}BP)C + PC')y = PA_1^{-1}b.$$ 

In the light of Lemma 3 (item 3), this can be rewritten as

$$PCy' + (PA_1^{-1}BC + PC')y = PA_1^{-1}b.$$ 

The canonical projector $Q_c = QA_1^{-1}B$ onto $\text{Ker}A$ along $S_0$ verifies $Q_cC = 0$, since $C$ is onto $S_0$. This means that $PA_1^{-1}BC = (I - Q)A_1^{-1}BC = A_1^{-1}BC$, and then we get

$$PCy' + (A_1^{-1}BC + PC')y = PA_1^{-1}b.$$ 

Premultiplying by $DA_1$ and using $A_1P = A$ (Lemma 3, item 1), we are led to

$$DACy' + D(BC + AC')y = DAA_1^{-1}b.$$ 

This is (30), where $E_1 = DAC$, $F_1 = D(BC + AC')$ and $b_1 = DAA_1^{-1}b$. □
Proof of Theorem 1: Assume that \( x \) is a solution of (20) within \( C^1_p \). Then, the projector approach implies that \( u = Px \) is a \( C^1 \) solution of the inherent ODE (23). According to the proof of Lemma 1, \( y = (PC)^{-1}u \) is a \( C^1 \) solution of (34), that is, of (30).

On the other hand, using the fact that \( P_c \) is onto \( S_0 = \text{Im}C \), we get that \( P_cC = C \). Also, \( P_cP = P_c \) owing to Lemma 2 (item 4), so that \( P_cu = P_cPCy = P_cCy = Cy \), and (24) can be written simply as

\[
x = Cy + QA_1^{-1}b.
\]

This is (31), where \( u_0 = QA_1^{-1}b \).

Conversely, let \( y \) be a \( C^1 \) solution of (30) or, equivalently, of (34), and assume that \( x \) is given by (31), that is, \( x = Cy + QA_1^{-1}b \). We only need to check that \( u = PCy \) is a solution of the inherent ODE (23), because then solutions would be recovered as \( x = P_cu + QA_1^{-1}b \), according to (24), and this would hold in virtue of the identities \( P_cPCy = P_cCy = Cy \), and \( x = Cy + QA_1^{-1}b \). Note also that \( P_x \) satisfies \( PCy \) is \( C^1 \), so that \( x \) would actually belong to \( C^1_p \).

In order to see that \( u = PCy \) satisfies the inherent ODE (23), rewrite (34) as

\[
D(AC'y' + (BC + AC')y - AA_1^{-1}b) = 0. 
\]

The matrix \( D \) can be removed from this identity since it yields an isomorphism \( \text{Im}A \leftrightarrow \mathbb{R}^r \), and in (35) \( D \) is acting on a vector within \( \text{Im}A \): note that \( BCy \in \text{Im}A \) due to the fact that \( C \) is onto \( S_0 \), which by definition is \( \{x \in \mathbb{R}^r / Bx \in \text{Im}A \} \). Hence, premultiplying by \( A_1^{-1} \) and using \( A_1^{-1}A = P \) (Lemma 3, item 1) we get (33), which is equivalent to (32) as detailed in the proof of Lemma 1. Equation (32) is exactly the inherent ODE (23) for \( u = PCy \).

Finally, the simultaneous \( C^1 \) nature of \( x \) and \( u_0 \) holds trivially in the light of (31), since \( Cy \) is \( C^1 \).

\[ \square \]

Remark 2. The choices \( b_1 = DAA_1^{-1}b \), \( u_0 = QA_1^{-1}b \) satisfy (12)-(13) if \( u_0 = QA_1^{-1}b \) is \( C^1 \).

In this regard, note that \( Bu_0 - b = (BQA_1^{-1} - I)b = (BQ - A_1)A_1^{-1}b \) and \( BQ - A_1 = AP'Q - A \), by definition of \( A_1 \). Hence,

\[
Bu_0 - b = (AP'Q - A)A_1^{-1}b \in \text{Im}A,
\]

showing that (13) is satisfied.

On the other hand, note that \( Au_0' = APu_0' = A(Pu_0)' - AP'u_0 = -AP'u_0 \) since \( Pu_0 = 0 \). This expression makes it easy to show that (12) holds:

\[
D(b - Bu_0 - Au_0') = D((A - AP'Q)A_1^{-1}b + AP'QA_1^{-1}b) = DAA_1^{-1}b = b_1.
\]

\[ \square \]

Note that Theorem 1 would lead, with little additional effort, to a simple proof of the already mentioned fact that \( C^k \)-smoothness of the excitation is transferred to the solutions by index-1 \( C^k \) pairs, when \( k \geq 1 \), yielding an injection \( C^k \rightarrow C^k \). The proof would be a trivial consequence of Lemma 4 and the fact that the operators \( C, D, P \) can be taken in a way such that \( E_1, F_1 \) and \( b_1 \) in the reduced ODE (30) are \( C^{k-1} \). Since \( E_1 \) is non-singular due to the index-1 assumption, solutions \( y \) of this ODE would belong to \( C^k(\mathbb{I}, \mathbb{R}^r) \). Additionally, \( u_0 = QA_1^{-1}b \) would be \( C^k \), the result following from (31).
The RL circuit revisited. In the light of Theorem 1, a reduction of the RL circuit equations can be performed just assuming that \( l(t), \ g(t) \neq 0, \ \forall t \in \mathcal{I} \), and that \( g \) (as well as \( l \), for electrical reasons) is a \( C^1 \) function. In this setting, smoothness of Ker\( A \) simply follows from the non-vanishing of \( l \), whereas smoothness of \( S_0 \) is given by the \( C^1 \) nature of \( g \). This acknowledges that the \( C^2 \) assumption on \( l \) can be removed, since \( C^1 \) smoothness on the operators can be relaxed to \( A_0 \) and \( A_1 \) in Theorem 1, and also that there is no need to use any derivative of \( i_s \) or to assume that this function is differentiable.

With the criteria defined in Theorem 1, and the choices (17) and (27) for \( C, D \), and the projectors \( P, Q \), respectively, we get

\[
u_0 = QA_1^{-1}b = \begin{pmatrix} 0 \\ i_s/g \end{pmatrix}
\]

and

\[
b_1 = DA_1^{-1}b = (i_s/g),
\]

so that the reduced ODE (30) now reads

\[
\dot{l}(t)g(t)y' + (1 + (l(t)g(t))^t) y = i_s(t)/g(t), \tag{36}
\]

and solution are recovered through (31), that is,

\[
\begin{pmatrix} i \\ e \end{pmatrix} = \begin{pmatrix} g(t) \\ -1 \end{pmatrix} y + \begin{pmatrix} 0 \\ i_s(t)/g(t) \end{pmatrix}. \tag{37}
\]

Comparing the new reduction (36) with (18), note that no derivative of \( i_s \) is involved in the former. Finally, let us emphasize that a derivative of \( g \) is still present in (36), in contrast to the inherent ODE (28) coming from the projector approach.

3 Singular problems

A rough introduction to singularities comes again from the circuit example (1). Suppose for the moment that \( l(t) \neq 0 \), \( \forall t \in \mathcal{I} \). Assuming that \( g(t) \) vanishes only at isolated points, or even imposing the milder condition that \( g(t) \) does not vanish on an open dense subset of \( \mathcal{I} \), we may define \( t \) as a singular point if \( g(t) = 0 \). Note that these are the points where both \( E_1 \) in the reduction procedure of 2.2 and \( A_1 \) in the projector approach discussed in 2.3 fail to be invertible. But remark also that it is possible to carry out the first (and only, in index-1 cases) step of the matrix chain construction in order to define \( E_1 \) or \( A_1 \), respectively. These singularities will be therefore called of type 1, and are considered in subsection 3.1.

In contrast, if \( l(t) \) vanishes at some points or, more generally, if \( A(t) \) does not have constant rank in \( \mathcal{I} \), even the first step of the reduction procedure or the tractability chain construction might find problems, at least in non-analytic cases. Note that in this example actually both processes can still be performed, although this fact is supported on more subtle (implicit) assumptions. These singular points of type 0 are addressed in subsection 3.2.
Both kind of zeroes are nicely accommodated by the reduced ODE (18), which in these situations becomes a singular ODE. This is not the case of the inherent ODE in the form depicted in (28). The goal of this Section is to show that projector methods for the DAE (3) can be reformulated in order to reflect such singularities also through a singular inherent ODE, which will actually display a scalarly implicit form. This can be done with little effort for singularities of type 1, but more care is required for those of type 0. The latter can also meet severe difficulties within the reduction approach, due to the failing of constant rank assumptions at the beginning of the process. This reformulation will be shown to preserve the nice properties of projector methods regarding smoothness, and certain important properties of singular points will be proved in 3.3 to be independent of the choice of projectors.

Singular points in higher index cases, for which the projector approach becomes more cumbersome, are the scope of future research. Note that this would also comprise index transitions such as the one defined by $g(t) > 0$ for $t < t^*$, $g(t) = 0$ for $t \geq t^*$, $l(t) \neq 0 \forall t$, modeling a persistent hard fault due to a (possibly unexpected) open-circuit in Figure 1.

3.1 Type 1

Consider the linear time-varying DAE (3), with continuous $A$, $B$, $b$, in a given interval $I$. Let us assume (just for this subsection) that the space $\text{Ker}A(t)$ has a $C^1$ structure in $I$. By definition, this means that there exist $n - r$ ($r$ being the constant rank of $A$ in $I$) $C^1$ functions spanning a basis of $\text{Ker}A(t)$, $\forall t \in I$. This is well-known to be equivalent to the existence of a $C^1$ projector $P(t)$ along $\text{Ker}A(t)$, and also to the $C^1$-smoothness of the orthogonal projector along $\text{Ker}A(t)$. The existence of such a $P$ allows us to define globally $Q = I - P$, $B_0 = B - AP'$, and the matrix $A_1 = A + B_0Q$ already depicted in (22).

Points with singular $A_1$ would be defined by the zeroes of the continuous function $\det A_1(t)$, and hence define a closed (in $I$) set. Let us also assume that the open set where $A_1$ is non-singular is also dense in $I$. This assumption makes it possible to rule out “regular” higher index cases from the present discussion. Analytic problems with not-everywhere singular $A_1$ satisfy this, so that this assumption seems to provide a natural extension of the setting of analytic DAEs with singularities, beyond cases with just isolated singularities.

In this context, we may define $t^* \in I$ as a singular point of type 1 if $A_1(t^*)$ is singular. This is consistent with Definition 4 below, aimed at a more general setting. Denote as $I_1$, the set of singularities of type 1.

The inverse matrix $A_1^{-1}$ appearing in (23)-(25) is well-defined in $I - \mathcal{I}_1$. If $A_1^T$ stands for the transpose of the matrix of cofactors, this inverse may be written there as $A_1^{-1} = (1/\det A_1)A_1^T$. For later use, note that $A_1A_1^{-1} = A_1A_1^{-1} = \det A_1I$ in $I$. From the assumed density of $I - \mathcal{I}_1$, it is natural to expect that, rewriting those equations in scalarly implicit form, solutions of the DAE could be recovered “bijectively” from the resulting decoupling. This is actually the case, as acknowledged in Theorem 2 below. Its proof and several details are delayed to the more general setting of Theorem 3. We believe the pure type-1 case to be worth of this independent statement since it comprises many cases arising in applications, e.g., all singularities of problems with constant $A$ and, in particular, semiexplicit systems.
Theorem 2. Consider the linear time-varying DAE (3) in a given interval \( \mathcal{I} \). Assume that:

C0. \( A(t), B(t) \in C^0(\mathcal{I}, \mathbb{R}^{n \times n}), b(t) \in C^0(\mathcal{I}, \mathbb{R}^n) \).

C1. \( \ker A(t) \) is \( C^1 \) in \( \mathcal{I} \).

Let \( P \) denote a smooth projector along \( \ker A(t) \), define \( A_1 \) as in (22), and let \( \mathcal{I}_s \) stand for the points where \( A_1 \) is singular. Assume also that:

C2. The set \( \mathcal{I} - \mathcal{I}_s \) is dense in \( \mathcal{I} \).

Then, \( x \in C^1_p(\mathcal{J}, \mathbb{R}^n) \) is a solution of the reformulation (20) in some interval \( \mathcal{J} \subseteq \mathcal{I} \) if and only if it can be written as

\[
x(t) = u(t) + v(t),
\]

where \( u \) is a \( C^1(\mathcal{J}, \mathbb{R}^n) \) solution of the scalarly implicit inherent ODE

\[
\omega_1(t) u' + L_1(t) u = H_1(t) b(t)
\]

lying on the (invariant) space \( \text{Im} P \), and \( v \in C^0(\mathcal{J}, \mathbb{R}^n) \) satisfies, together with \( u \),

\[
\omega_1(t) v + L_2(t) u = H_2(t) b(t).
\]

The operators in (39)-(40) are continuous in the whole interval \( \mathcal{I} \) and have the expressions

\[
\omega_1 = \det A_1, \quad L_1 = -\omega_1 P' + \overline{P A_1}, \quad H_1 = \overline{P A_1}, \quad L_2 = Q \overline{A_1}, \quad H_2 = Q \overline{A_1}.
\]

Remark 3. Conditions C0 and C1 in Theorem 2 hold if

B0. \( A(t), B(t) \in C^1(\mathcal{I}, \mathbb{R}^{n \times n}), b(t) \in C^0(\mathcal{I}, \mathbb{R}^n) \).

B1. \( A(t) \) has constant rank in \( \mathcal{I} \).

Under B0 and B1, \( \mathcal{I}_s \) is the set where \( E_1 \) in any reduction of the form (11) becomes singular.

This follows immediately from [18, Theorem 7.1]. \( \square \)

3.2 Types 0 and 1

Definition 3. Consider a continuous matrix function \( A(t) \) defined in some interval \( \mathcal{I} \). A point \( t^* \) such that \( \text{rk} A(t^*) \) is not constant in any neighborhood of \( t^* \) will be called a singular point of type 0. The set of singular points of type 0 will be denoted as \( \mathcal{I}_{s_0} \).

In order to allow for a global, projector-based treatment of singularities including those of type 0, it seems natural to replace assumption C1 above by the following.

Assumption D1: There exists a \( C^1 \) vector space \( N(t) \) defined in \( \mathcal{I} \) and verifying \( N(t) = \ker A(t) \) for all \( t \in \mathcal{I} - \mathcal{I}_{s_0} \).
This assumption always holds in analytic problems, due to the restricted nullspace notion in this setting (see [18]). Assumption D1 actually comprises several different properties: Ker$A(t)$ has a $C^1$ structure in $\mathcal{I} - \mathcal{I}_{s_0}$; this structure can be smoothly extended to the whole $\mathcal{I}$; and, additionally, non-smooth Kernels in constant rank problems (which could be only displayed in non-$C^1$ contexts) are excluded.

The previous notion of singularity of type 1 can be extended without difficulty to this setting, since there exists a globally defined operator $P$ yielding a projector along Ker$A$ in $\mathcal{I} - \mathcal{I}_{s_0}$. Again, the matrices $Q$, $B_0$ and $A_1$ can be defined in the whole interval, although some of their properties might be lost in this framework: see, in this sense, Proposition 1 and Lemma 5. Let us only remark here that the operators defined in (41) are still continuous.

**Definition 4.** A point $t^* \in \mathcal{I} - \mathcal{I}_{s_0}$ will be called a singular point of type 1 if $A_1(t^*)$ is singular. The set of singular points of type 1 will be denoted as $\mathcal{I}_{s_1}$. Points in $\mathcal{I}_r = \mathcal{I} - (\mathcal{I}_{s_0} \cup \mathcal{I}_{s_1})$ will be called regular.

The focus of this Section will be placed on problems with tractability index-1 almost everywhere, in the sense described by the following extension of C2:

**Assumption D2:** $\mathcal{I}_r = \mathcal{I} - (\mathcal{I}_{s_0} \cup \mathcal{I}_{s_1})$ is an open dense subset of $\mathcal{I}$.

Item 1 of Lemma 5 makes it possible to work with the reformulation (20) also in this setting. Now, solutions need not be defined in the whole interval $\mathcal{I}$ but in certain subintervals $\mathcal{J}$. Solutions would then be considered in $C^1_p(\mathcal{J}, \mathbb{R}^n)$.

**Theorem 3.** Consider the linear time-varying DAE (3) in a given interval $\mathcal{I}$. Let

D0. $A(t), B(t) \in C^0(\mathcal{I}, \mathbb{R}^{n \times n}), b(t) \in C^0(\mathcal{I}, \mathbb{R}^n)$.

Assume additionally that D1 and D2 hold. Let $P(t)$ be a smooth projector along the space $N(t)$ appearing in Assumption D1, and assume that $Q$, $B_0$, $A_1$ are defined from $P$.

Then the conclusions of Theorem 2 hold, that is, $x \in C^1_p(\mathcal{J}, \mathbb{R}^n)$ solves (20) if and only if it can be written as $x = u + v$, where $u \in C^1(\mathcal{J}, \mathbb{R}^n)$ is a solution of (39) lying on the (invariant) space $\text{Im} P(t)$, and $v \in C^0(\mathcal{J}, \mathbb{R}^n)$ (together with $u$) satisfies (40).

**Proof:** Let $x(t)$ be a solution of (20) defined in some interval $\mathcal{J} \subseteq \mathcal{I}$. Premultiplying (20) by $A_1$ and using Lemma 5 (item 3) we get

$$\det A_1 P(Px)' + A_1 Bx - \det A_1 PP'x = A_1 b(t).$$

Writing $x = Px + Qx$ after $A_1 B$ yields, in the light of Lemma 5 (item 4),

$$\det A_1 P(Px)' + A_1 BPx + \det A_1 (Q + PP')x - \det A_1 PP'x = A_1 b(t),$$

that is,

$$\det A_1 P(Px)' + A_1 BPx + \det A_1 Qx = A_1 b(t). \quad (42)$$

16
Now, premultiplication of (42) by $P$ leads to
\[ \det A_1 P (P x)' + P \overline{A}_1 B P x = P \overline{A}_1 b(t), \] (43)
which reads, using $(P x)'' = (P P x)' = P (P x)' + P' P x \Rightarrow P (P x)' = (P x)' - P'' P x,
\[ \det A_1 (P x)' + (-\det A_1 P' + P \overline{A}_1 B) P x = P \overline{A}_1 b(t). \] (44)
This is exactly (39) with $P x = u$, in the light of (41). Also, multiplying (42) by $Q$ yields
\[ Q \overline{A}_1 B P x + \det A_1 Q x = Q \overline{A}_1 b(t), \] (45)
which is (40) with $u = P x$, $v = Q x$.

The invariance of $\text{ImP}(t)$ for (39) follows from the identities $QL_1 = -\omega_1 Q P'$, $QH_1 = 0$: let $u(t)$ satisfy (39) in a given open interval $J$. Multiplying (39) by $Q$ yields $\omega_1 (Q u' - Q P' u) = 0$. Since $Q P = 0$ and, therefore, $-Q P' = Q' P$, we get $\omega_1 ((Q u)' - Q' u + Q' P u) = 0$, that is, $\omega_1 ((Q u)' - Q' Q u) = 0$. Now, using the fact that regular points in $J$ define a dense subset of $J$, we have $(Q u)' - Q' Q u = 0$ in the whole interval $J$. This is a homogeneous explicit ODE for $Q u$, and the condition $u(t_0) \in \text{ImP}(t_0)$ is equivalent to $Q(t_0) u(t_0) = 0$: hence, $Q(t) u(t) = 0$ for all $t \in J$, i.e., $u(t) \in \text{ImP}(t)$ for all $t \in J$.

Now, let $u(t) \in \text{ImP}(t)$ satisfy (39) and (together with $v(t)$) also (40) in some interval $J$. From the fact that $u \in \text{ImP}$ we have $u = P u$, $Q u = 0$. Premultiplying (40) by $P$ we get $\omega_1 P v = 0$, so that by density $P v = 0$, $Q v = v$. With $x = u + v$, we then get $P x = u$, $Q x = v$.

The addition of (39) and (40) reads
\[ \omega_1 u' - \omega_1 P' u + \overline{A}_1 B u + \omega_1 (t) v = \overline{A}_1 b(t). \] (46)
Using $u = P u$, $\omega_1 u' - \omega_1 P' u$ can be rewritten as $\omega_1 P u' = \overline{A}_1 A (P x)'$, from Lemma 5 (item 3) and the fact that $u = P x$. Also, $\overline{A}_1 B u = \overline{A}_1 B P x = \overline{A}_1 B_0 P x$ since $B P = (B_0 + A P') P = B_0 P$ (note that $A P' P = -A Q' P = A Q P' = 0$), and $\omega_1 v = \overline{A}_1 A_1 Q x = \overline{A}_1 (A + B_0 Q) Q x = \overline{A}_1 B_0 Q x$. This way we transform (46) into
\[ \overline{A}_1 A (P x)' + \overline{A}_1 B_0 x = \overline{A}_1 b(t). \] (47)
Since $\overline{A}_1$ is invertible in $J \cap \mathcal{L}_r$, this equation can be rewritten as (20) in $J \cap \mathcal{L}_r$. Due to the continuity and vanishing of $A (P x)' + B_0 x - b$ in $J \cap \mathcal{L}_r$, which is dense in $J$, it follows that it actually vanishes in the whole $J$, meaning that (20) is satisfied in $J$ with $x = u + v$.

\[ \square \]

In regions where $\omega_1 \neq 0$ (or in the whole $\mathcal{I}$ in regular problems) system (39)-(40) amounts to the decoupling obtained with classical projector methods. Theorem 2 holds in virtue of this proof as well. Note also that it is not necessary to assume that $J$ comprises only regular points. Well-defined solutions may be displayed through singular points, and this would be reflected by the decoupling (39)-(40). Solvability statements in this direction are beyond the scope of the present discussion, focused on how to derive an inherent ODE somehow reflecting the singularities of the DAE. The role of the choice of projectors in this derivation, together with some additional related properties, are discussed below.
3.3 Some properties of the scalarly implicit inherent ODE

The inherent equation (39) is a singular linear time-varying ODE in scalarly implicit form. Much of the features of this kind of problems rely upon the leading coefficient \( \omega_1(t) \).

**Proposition 1.** Let \( A(t), B(t) \) be continuous. Under Assumptions D1 and D2:

1. The function \( \omega_1 = \det A_1 \) is invariant w.r.t. the choice of the (smooth) projector \( P \).
2. \( \omega_1(t) = 0 \) if and only if \( t \) is a singular point.
3. If, additionally, \( A \) and \( B \) are \( C^1 \), then \( \omega_1 \) is \( C^1 \).

**Proof:** 1. Let \( R \) denote any other smooth projector defined in \( \mathcal{I} \), along \( \ker A \) in \( \mathcal{I} - \mathcal{I}_{s_0} \), and write \( T = I - R \). Denote \( B_{0,r} = B - AR', A_{1,r} = A + B_{0,r}T \). Then it holds that

\[
A_{1,r} = A_1(P + T).
\]

This follows from the identities \( A_1P = (A + B_0Q)P = AP = A \), \( A_1T = (A + B_0Q)T = B_0T \), the latter being supported on the fact that \( QT = T \) (Lemma 2, item 3). Hence, \( A_1(P + T) = A + B_0T \), and it suffices to show that \( B_0T = B_{0,r}T \), which owes to the identities \( B_0T = (B - AP')T = BT + APT' = BT + ART' = (B - AR')T = B_{0,r}T \), where we have used \( AP = A = AR \) and \( P'T = -PT', RT = -RT' \) since \( PT = RT = 0 \).

The proof is complete if we remark that \( \det(P + T) = 1 \). This follows from the fact that \( \sigma(P + T) = \{1\} \), which can be seen to hold by rewriting \( P + T = I - (Q - T) \) and noting that \( Q - T \) is a nilpotent index-2 matrix, since \( (Q - T)(Q - T) = Q - T - Q + T = 0 \).

2. The only non-trivial part is \( t^* \in \mathcal{I}_{s_0} \Rightarrow \omega_1(t^*) = 0 \). Due to item 2 of Lemma 5, \( \ker P(t^*) \) is strictly contained in \( \ker A(t^*) \), and therefore it is possible to take \( w \in \ker A(t^*) - \ker P(t^*) \). Then \( P(t^*)w \neq 0 \), \( A(t^*)w = 0 \) and \( A_1(t^*)P(t^*)w = A(t^*)P(t^*)w + B_0(t^*)Q(t^*)P(t^*)w = 0 \), since \( A(t^*)P(t^*)w = A(t^*)w = 0 \) and \( Q(t^*)P(t^*) = 0 \). This means that the Kernel of \( A_1(t^*) \) is not trivial, showing that \( \det A_1(t^*) = 0 \).

3. The matrix \( A = A + B_0Q = A + BQ - AP'Q \) can be written as \( (A + BQ)(I - P'Q) \), due to the fact that \( QP'Q = -QPQ' = 0 \). In virtue of the same property, it follows that \( P'Q \) is a nilpotent matrix with index 2, so that \( \sigma(I - P'Q) = \{1\} \) and then \( \det(I - P'Q) = 1 \). Therefore, \( \omega_1 = \det(A + BQ) \) is \( C^1 \) due to the \( C^1 \) nature of \( A, B \) and \( Q \).

\[ \Box \]

Property 2 characterizes the location of both type 0 and 1 singularities. It is worth remarking that, broadly speaking, the “natural” analog in higher index extensions does not hold: in the “index-2” pair

\[
A(t) = \begin{pmatrix} 0 & 0 \\ t & 0 \end{pmatrix}, \quad B(t) = I,
\]

18
$A(t)$ has a rank deficiency at the origin, but both $A_2$ in the tractability chain and $E_2$ in the second step of the reduction procedure would be easily checked to be non-singular. This means that $t^* = 0$ would be a singular (or, maybe better, critical) point of type 0 for which $\omega_2(t^*) = \det A_2(t^*) \neq 0$.

There exist an extensive theory for singular linear time-varying ODEs, at least in the analytic setting (see e.g. [5, 10, 21]). Let us just perform here a brief digression in order to show that the kind of singularity arising in the inherent ODE is not dependent on the choice of projectors. In this direction, focus on homogeneous analytic ODEs

$$E(t)x' + F(t)x = 0. \quad (49)$$

If $E$ becomes singular at an isolated point $t^*$, we say that $t^*$ has a singularity of the $k$-th kind, with $k$ a positive integer, if (49) can be rewritten as $(t - t^*)^k x' + G(t)x = 0$, in some neighborhood of $t^*$, with $G$ analytic and non-vanishing at $t^*$. Or, equivalently,

$$x' + H(t)x = 0, \quad (50)$$

$H(t)$ having a pole of order $k$ at $t^*$.

Let us now drive the attention to cases with a scalarly implicit analytic ODE

$$\omega_1(t)x' + L(t)x = 0. \quad (51)$$

Let $\omega_1(t^*) = 0$. If $\omega_1$ does not vanish in some connected neighborhood of $t^*$, there exists a minimum positive integer such that $\omega_1^{(k_1)}(t^*) \neq 0$. It follows from elementary properties of analytic functions that, if $L^{(k)}(t^*) = 0$ for all non-negative $k < k_1$, (51) could be rewritten in the form (50) with analytic $H$. Otherwise, it is easy to check that $t^*$ is a singularity of the $k$-th kind for (51) if and only if $k = k_1 - k_2$, with $k_1$ defined above and $k_2 < k_1$ being the minimum non-negative integer such that $L^{(k_2)}(t^*) \neq 0$.

When the scalarly implicit ODE (51) comes from the decoupling of a homogeneous, analytic DAE around a singular point $t^*$, the kind of singularity arising at $t^*$ is not dependent on the choice of projectors:

**Proposition 2.** Consider an analytic DAE $A(t)x' + B(t)x = 0$ with tractability index 1 in $\mathcal{I}_\tau = \mathcal{I} - \{t^*\}$ and a singularity at $t^*$. Let

$$\omega_1(t)u' + \omega_1 P' + P\overline{A}_1 B u = 0 \quad (52)$$

be the inherent ODE (39) coming from the choice of an analytic operator $P(t)$ which yields a projector along $\text{Ker} A$ in $\mathcal{I}_\tau$. If $t^*$ is a singularity of the $k$-th kind for (52), then any other analytic projector $R(t)$ leads to a singularity of the same kind at $t^*$.

**Proof:** With the notation introduced above, $k_1$ is obviously independent of the choice of projectors due to item 1 of Proposition 1.

To see that $k_2$ is also independent of the choice of the projector, let us first remark that $(\omega_1 P')^{(j)}(t^*) = 0$ for $j < k_1$ and, in particular, for $j \leq k_2 < k_1$, owing to Leibniz's rule and
the fact that $\omega^{(j)}_1(t^*) = 0$ for all $j < k_1$. The same happens if $P$ is replaced by $R$. It then follows that

\begin{align}
(PA_1^*B)^{(j)}(t^*) &= 0, \text{ for } 0 \leq j < k_2, \quad (53a) \\
(PA_1B)^{(k_2)}(t^*) &\neq 0. \quad (53b)
\end{align}

We need to show that (53) holds if $P$ is replaced by $R$ and $A_1$ by $A_{1R} = A + (B - AR')T$, with $T = I - R$.

Recall that (48) states that $A_{1R} = A_1(P + T)$. From the property $\det(P + T) = 1$ already used in the proof of Proposition 1 (item 1), it follows that $P + T$ is invertible (actually, with inverse $R + Q$), and that the transposed cofactor matrix $P + T$ is exactly the inverse $R + Q$. Accordingly,

$$A_{1R} = (R + Q)A_1. \quad (54)$$

Premultiplying by $R$ we get $RA_{1R} = RA_1 = RPA_1$, since $RP = R$ (Lemma 2, item 4). Hence $RA_{1R}B = R(PA_1B)$ and, dually, $PA_1B = P(RA_{1R}B)$. Therefore,

$$RA_{1R}B = 0 \iff PA_1B = 0,$$

and the proof is complete if $k_2 = 0$.

If $k_2 \geq 1$, we use $RA_{1R}B(t^*) = 0 \iff PA_1B(t^*) = 0$ to show that $(RA_{1R}B)'(t^*) = R(PA_1B)'(t^*)$, and $(PA_1B)'(t^*) = P(RA_{1R}B)'(t^*)$, so that

$$(RA_{1R}B)'(t^*) = 0 \iff (PA_1B)'(t^*) = 0,$$

and, inductively,

$$(RA_{1R}B)^{(j)}(t^*) = 0 \iff (PA_1B)^{(j)}(t^*) = 0 \text{ for all } j \leq k_2.$$

The result then follows from the property $k = k_1 - k_2$.

\[\square\]

4 Concluding remarks

The interrelations between reduction techniques and projector methods addressed in Section 2 for regular, linear index-1 DAEs, might be extended to higher index cases. Also, in the singular setting of Section 3, further relations between the scalarly implicit decoupling (39)-(40) and the reduction (14)-(15) or (30)-(31) could be explored; in this direction, Remark 3 illustrates that singularities of type 1 are essentially the same in both approaches, and the failing of the rank condition defining those of type 0 is not dependent on the analysis method. Singularities in higher index problems, as well as index transitions, are the scope of future work.
Appendix

Lemma 2. Assume that $A(t) \in C^0(\mathcal{I}, \mathbb{R}^{n \times n})$, and that Ker$A(t)$ is spanned by a $C^1$ basis in $\mathcal{I}$. Let $P(t)$ be a $C^1$ projector along Ker$A(t)$, that is, assume $P^2(t) = P(t)$, Ker$P(t) = \text{Ker}A(t)$. Denote as $Q(t) = I - P(t)$ the corresponding projector onto Ker$A(t)$. The following elementary properties hold:

1. $PQ = QP = 0; \quad AQ = 0, \quad AP = A$.
2. $P' = -Q'; \quad P'Q = -PQ'; \quad Q'P = -QP'$.
3. If $T$ is another projector onto Ker$A$, then $TQ = Q, \quad QT = T$.
4. If $R$ is another projector along Ker$A$, then $PR = P, \quad RP = R$.

Lemma 3. Under the assumptions of Lemma 2 and given $B(t) \in C^0(\mathcal{I}, \mathbb{R}^{n \times n})$, define $B_0 = B - AP', \quad A_1 = A + B_0Q$. If $A_1$ is non-singular in $\mathcal{I}$, then the following properties hold:

1. $A_1P = A, \quad A_1^{-1}A = P$.
2. $A_1^{-1}BQ = Q + P'Q$.
3. $PP' + PA_1^{-1}BP = PA_1^{-1}B$.

Proof: Item 1 is simply due to $A_1P = AP + B_0QP = AP = A$.

Item 2 follows from the fact that $A_1(Q + P'Q) = (A + (B - AP')Q)(Q + P'Q) = AP'Q + (B - AP')Q + (B - AP')QP'Q = BQ + (B - AP')QP'Q$. Note that, since $PQ = QP = 0$, we have $QP'Q = -QPQ' = 0$, so that $A_1(Q + P'Q) = BQ$ and, therefore, $A_1^{-1}BQ = Q + P'Q$, as claimed.

Finally, item 3 is based on item 2 above, from which we have $PA_1^{-1}B = PA_1^{-1}BP + PA_1^{-1}BQ = PA_1^{-1}BP + PP'Q$. The result follows from the remark that $PP'Q = -PPQ' = -PQ' = PP'$.

Lemma 4. Assume that $A, B \in C^k(\mathcal{I}, \mathbb{R}^{n \times n})$, $k \geq 1$, and that $A$ has constant rank in $\mathcal{I}$. Let $P \in C^k(\mathcal{I}, \mathbb{R}^{n \times n})$ be a projector along Ker$A$, $Q = I - P$. If $A_1$ is non-singular in $\mathcal{I}$, then $QA_1^{-1} \in C^k(\mathcal{I}, \mathbb{R}^{n \times n})$.

Proof: Using $QP'Q = -QPQ' = 0$, write $A_1 = A + BQ - AP'Q = (A + BQ)(I - P'Q)$. Noting that $(I - P'Q)(I + P'Q) = (I + P'Q)(I - P'Q) = I$, the inverse $A_1^{-1}$ reads $(I + P'Q)(A + BQ)^{-1}$, so that $QA_1^{-1} = Q(A + BQ)^{-1}$. All operators in this expression are $C^k$ and the claim is proved.
Lemma 5. Let $A, B \in C^0(\mathcal{I}, \mathbb{R}^{m \times n})$, and suppose that Assumptions D1 and D2 hold. Let $P$ be a smooth projector along the space $N$ appearing in Assumption D1, and assume that $Q, B_0, A_1$ are defined from $P$. Then

1. $AQ = 0$ and $AP = A$ in $\mathcal{I}$.

2. $\text{Ker} P \subseteq \text{Ker} A$ in $\mathcal{I}$, $\text{Ker} P(t) = \text{Ker} A(t)$ if and only if $t \in \mathcal{I} - \mathcal{I}_0$.

3. $\overline{A_1} A = \det A_1 P$

4. $\overline{A_1} B Q = \det A_1 (Q + PP')$.

Proof: Item 1 follows from continuity and the corresponding properties in the dense set $\mathcal{I}_r$. In item 2, the inclusion $\text{Ker} P \subseteq \text{Ker} A$ is due to the property $AQ = 0$ in the whole interval $\mathcal{I}$, which reads, equivalently, $\text{Im} Q \subseteq \text{Ker} A$. The inclusion follows from the identity $\text{Ker} P = \text{Im} Q$. For the second claim in item 2, the “if” part holds by definition of $P$. The “only if” part follows from the remark that assumptions D1 and D2 together imply that $A(t)$ has maximal rank in $\mathcal{I} - \mathcal{I}_0$ and that the set $\mathcal{I}_0$ comprises the rank-deficient points.

Item 3 owes to $A_1 P = AP + B_0 Q P = A$, since then $\overline{A_1} A_1 P = \det A_1 P = \overline{A_1} A$. In turn, item 4 follows from $A_1 Q = B_0 Q$, which implies $\det A_1 Q = \overline{A_1} B_0 Q$, and then $\det A_1 Q = A_1 B Q - A_1 A P' Q = A_1 B Q - \det A_1 PP' Q$, that is, $\overline{A_1} B Q = \det A_1 (Q + PP' Q)$. Finally, $PP' Q = -PP Q' = -PQ' = PP'$.

\[\square\]

References


