Computing the regularization of a linear differential-algebraic system

Thomas Berger[a], Paul Van Dooren[b]

[a] *Fachbereich Mathematik, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany*
[b] *Department of Mathematical Engineering, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

Abstract

We study the regularization problem for linear differential-algebraic systems. As an improvement of former results we show that any system can be regularized by a combination of state-space and input-space transformations, behavioral equivalence transformations and a reorganization of variables. The additional state feedback which is needed in earlier publications is shown to be superfluous. We provide an algorithmic procedure for the construction of the regularization and discuss computational aspects.

Keywords: differential-algebraic systems; descriptor systems; regularization; behavioral approach.

1. Introduction

We study linear descriptor systems given by differential-algebraic equations (DAEs) of the form

$$\tfrac{\mathrm{d}}{\mathrm{d}t}Ex(t) = Ax(t) + Bu(t) \tag{1}$$

where $E, A \in \mathbb{R}^{l\times n}$, $B \in \mathbb{R}^{l\times m}$. The set of systems (1) is denoted by $\Sigma_{l,n,m}$ and we write $[E,A,B] \in \Sigma_{l,n,m}$. DAE systems of the form (1) naturally occur when modeling dynamical systems subject to algebraic constraints; for a further motivation we refer to [4, 8, 11, 12, 14] and the references therein. The system $[E,A,B]$ is called *regular*, if the matrix pencil $sE-A$ is regular, that is, $l = n$ and $\det(sE-A) \in \mathbb{R}[s]\setminus\{0\}$.

The functions $x : \mathbb{R} \to \mathbb{R}^n$ and $u : \mathbb{R} \to \mathbb{R}^m$ are usually called *state* and *input* of the system, resp. However, in the general case, u might be constrained and some of the state variables can play the role of an input. In the present paper we will take the viewpoint of the behavioral approach due to Willems [16], see also [17, 18]. Within this framework, the variables of the system do not have the interpretation of states and inputs until an analysis of the system reveals the free variables. These free variables should then be interpreted as inputs, since "they can be viewed as unexplained by the model and imposed on the system by the environment" [13]. This approach obeys the physical meaning of the system variables and it may turn out that in the original model the choice of states and inputs was inappropriate.

The *behavior* of the DAE system (1) is introduced as the following set of solutions of (1):

$$\mathfrak{B}_{[E,A,B]} := \{(x,u) \in \mathscr{L}^1_{\mathrm{loc}}(\mathbb{R};\mathbb{R}^n\times\mathbb{R}^m) \mid Ex \in \mathscr{AC}(\mathbb{R};\mathbb{R}^l), \\ (x,u) \text{ satisfies (1) for almost all } t \in \mathbb{R}\},$$

where $\mathscr{L}^1_{\mathrm{loc}}$ and $\mathscr{AC}$ denote the space of locally (Lebesgue) integrable and absolutely continuous functions, resp. DAE control systems based on the above behavior have been studied in detail e.g. in [4].

Nowadays, the modeling of huge industrial problems and complex physical systems is often performed using automatic modeling tools such as Modelica (https://www.modelica.org/). This approach naturally leads to differential-algebraic systems of the form (1). Since in the automatically generated models it is quite common that redundant equations appear and state and input variables are chosen inappropriately, the system (1) is not regular in general, while the physical background tells that a regular model must exist. Therefore, a remodeling, or a regularization, is often required, see [9].

In the present paper we study the regularization of DAE systems, which relies on a procedure developed in [9] and revisited in [3]. In [9] it is shown that, given any DAE system $[E,A,B] \in \Sigma_{l,n,m}$, by a combination of behavioral equivalence transformation, proportional state feedback and reorganization of variables (due to a possibly inappropriate initial choice of states and inputs) a new system $[E_{\mathrm{reg}}, A_{\mathrm{reg}}, B_{\mathrm{reg}}]$ can be obtained where $sE_{\mathrm{reg}} - A_{\mathrm{reg}}$ is regular and has index at most one. In the linear case, explicit transformations and a characterization of the regularized system have been obtained in [7]. In the present paper, we improve the results of [9, 7] by showing that an application of state feedback is not necessary. Furthermore, we derive a numerically stable algorithm of cubic complexity which establishes the regularization of the system.

The paper is organized as follows: In Section 2 we introduce some preliminary concepts and notation and give a precise problem formulation. The regularization algorithm, which is the main result of the paper, is presented in Section 3 and proved to be feasible for any given system. Numerical reliability and the computational speed of the regularization algorithm is discussed in Section 4. In Section 5 provide a detailed comparison of our algorithm with the method proposed in [9] and in Section 6 we demonstrate the regularization algorithm

Email addresses: thomas.berger[at]uni-hamburg.de (Thomas Berger), paul.vandooren[at]uclouvain.be (Paul Van Dooren)

by means of a numerical example. Conclusions are given in Section 7.

2. Preliminaries and problem formulation

In the present paper we use the following notation: $\mathbb{R}$ and $\mathbb{C}$ denote the fields of real and complex numbers, resp.; $\mathbb{R}[s]$ is the ring of polynomials with coefficients in $\mathbb{R}$; $R^{n\times m}$ is the set of $n\times m$ matrices with entries in a ring R; $\mathscr{O}_n$ denotes the set of orthogonal real $n\times n$ matrices. A polynomial matrix $U(s)\in\mathbb{R}[s]^{n\times n}$ is called *unimodular*, if it is invertible over $\mathbb{R}[s]$ or, equivalently, if $\det U(s)$ is a nonzero constant.

The rank of a matrix $M\in\mathbb{K}^{n\times m}$, where $\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbb{C}$, is denoted by $\operatorname{rk} M$. If $M\in\mathbb{R}^{n\times m}$ with $\operatorname{rk} M=r$, then, using QR factorization with pivoting [10], there exists $T\in\mathscr{O}_n$ such that

$$TM=\begin{bmatrix}\Sigma_r\\0\end{bmatrix},$$

where $\Sigma_r\in\mathbb{R}^{r\times m}$ with $\operatorname{rk}\Sigma_r=r$, see also [1]. We will call T a *row compression* of the matrix M. Similarly, we call $S\in\mathscr{O}_m$ a *column compression*, if

$$MS=[\hat{\Sigma}_r,0],$$

where $\hat{\Sigma}_r\in\mathbb{R}^{n\times r}$ with $\operatorname{rk}\hat{\Sigma}_r=r$.

The index $\nu\in\mathbb{N}_0$ of a regular matrix pencil $sE-A\in\mathbb{R}[s]^{n\times n}$ is defined via its (quasi-)Weierstraß form [5, 11, 12]: if for some invertible $S,T\in\mathbb{R}^{n\times n}$

$$S(sE-A)T=\begin{bmatrix}sI_r-J & 0\\0 & sN-I_{n-r}\end{bmatrix},\quad N \text{ nilpotent},$$

$$\text{then}\quad \nu:=\begin{cases}0, & \text{if } r=n,\\ \min\{\,k\in\mathbb{N}_0\mid N^k=0\,\}, & \text{if } r<n.\end{cases}$$

The index is independent of the choice of S,T.

Finally, we recall the concept of behavioral equivalence which has been introduced for general behaviors in [13]. Roughly speaking, two systems are behaviorally equivalent, if their behaviors coincide.

Definition 2.1. Two systems $[E_i,A_i,B_i]\in\Sigma_{l,n,m}$, $i=1,2$, are called *behaviorally equivalent*, if

$$\mathfrak{B}_{[E_1,A_1,B_1]}\cap\mathscr{C}^\infty(\mathbb{R};\mathbb{R}^n\times\mathbb{R}^m)=\mathfrak{B}_{[E_2,A_2,B_2]}\cap\mathscr{C}^\infty(\mathbb{R};\mathbb{R}^n\times\mathbb{R}^m),$$

where $\mathscr{C}^\infty$ denotes the space of infinitely times differentiable functions; we write

$$[E_1,A_1,B_1]\ \simeq_{\mathfrak{B}}\ [E_2,A_2,B_2].$$

In order to obtain a behaviorally equivalent system, it is allowed that some of the equations in (1) are differentiated (and hence we require smooth solutions). This leads to a transformation of the form $U(\frac{\mathrm{d}}{\mathrm{d}t})(\frac{\mathrm{d}}{\mathrm{d}t}E-A)x(t)-U(\frac{\mathrm{d}}{\mathrm{d}t})Bu(t)=0$ with some $U(s)\in\mathbb{R}[s]^{l\times l}$. Furthermore, since the behaviors must coincide (on $\mathscr{C}^\infty$) the transformation $U(s)$ must be reversible, i.e., $U(s)$ must be unimodular. As shown in [13, Thms. 2.5.4 & 3.6.2] this is exactly the set of transformations that characterizes behavioral equivalence; this is summarized in the following lemma.

Lemma 2.2. *Let $[E_i,A_i,B_i]\in\Sigma_{l,n,m}$, $i=1,2$. Then $[E_1,A_1,B_1]\simeq_{\mathfrak{B}}[E_2,A_2,B_2]$ if, and only if, there exists a unimodular $U(s)\in\mathbb{R}[s]^{l\times l}$ such that*

$$[sE_1-A_1,-B_1]=U(s)[sE_2-A_2,-B_2].$$

Note that in initial value problems (1), $x(0)=x^0$, where $u\in\mathscr{C}^\infty(\mathbb{R};\mathbb{R}^m)$ is given, the consistency of the initial value $x^0\in\mathbb{R}^n$, i.e., existence of $x\in\mathscr{C}^\infty(\mathbb{R};\mathbb{R}^n)$ such that $(x,u)\in\mathfrak{B}_{[E,A,B]}$ and $x(0)=x^0$, is preserved under behavioral equivalence.

In the present paper we consider the following regularization problem.

Problem 2.3. For a given system $[E,A,B]\in\Sigma_{l,n,m}$, find a unimodular matrix $U(s)\in\mathbb{R}[s]^{l\times l}$, orthogonal state space and input space transformations $T\in\mathscr{O}_n$, $V\in\mathscr{O}_m$ and a permutation matrix $P\in\mathscr{O}_{n+m}$ such that

$$[sE-A,-B]\begin{bmatrix}T & 0\\0 & V\end{bmatrix}P=U(s)\begin{bmatrix}0 & 0\\sE_{\mathrm{reg}}-A_{\mathrm{reg}} & -B_{\mathrm{reg}}\end{bmatrix},\quad(2)$$

where $sE_{\mathrm{reg}}-A_{\mathrm{reg}}\in\mathbb{R}[s]^{\hat{n}\times\hat{n}}$ is regular and has index at most one.

Each kind of the transformations in Problem 2.3 have an interpretation in terms of their physical meaning:

(i) T and V represent coordinate changes in state space and input space respectively,

(ii) $U(s)$ represents an equivalence transformation which does not change the behavior of the system,

(iii) P represents a permutation of state and input variables. Here, we seek a permutation of free state variables with constraint input variables, so that in the resulting system the free variables are exactly the input variables. This may be viewed as a reinterpretation of certain states as inputs and vice versa.

At first glance it may be surprising that (2) in Problem 2.3 does not read

$$W(s)[sE-A,-B]\begin{bmatrix}T & 0\\0 & V\end{bmatrix}P=\begin{bmatrix}0 & 0\\sE_{\mathrm{reg}}-A_{\mathrm{reg}} & -B_{\mathrm{reg}}\end{bmatrix},\quad(3)$$

where $W(s)\in\mathbb{R}[s]^{l\times l}$ is unimodular. The reason is that $U(s)$ in (2) may be easier to compute than $W(s)$ in (3). In fact, we show in Section 3 that $U(s)$ has degree 1, i.e., it is a matrix pencil, and it is obtained with cubic complexity. On the other hand, the inverse $W(s)=U(s)^{-1}$ may have higher degree and can only be obtained with quartic complexity in general, see Section 4.

3. Regularization algorithm

In this section we provide a step by step procedure for the derivation of the regularization of a descriptor system as in (2).

Initialization. Let $[E,A,B] \in \Sigma_{l,n,m}$ be given.

Step 1. Compute a row compression $S_1 \in \mathscr{O}_l$ such that $S_1 B = \left[\begin{smallmatrix} 0 \\ B_2 \end{smallmatrix}\right]$, where B_2 has full row rank r. Consider

$$S_1[sE-A,-B] = \begin{bmatrix} sE_1 - A_1 & 0 \\ sE_2 - A_2 & -B_2 \end{bmatrix},$$

where $sE_1 - A_1 \in \mathbb{R}[s]^{(l-r)\times n}, sE_2 - A_2 \in \mathbb{R}[s]^{r\times n}$.

Step 2. Compute orthogonal $S_2 \in \mathscr{O}_{l-r}, T_2 \in \mathscr{O}_n$ that take $sE_1 - A_1$ into staircase form

$$S_2(sE_1 - A_1)T_2 = \begin{bmatrix} sE_\eta - A_\eta & 0 & 0 & 0 \\ * & sE_\infty - A_\infty & 0 & 0 \\ * & * & sE_f - A_f & 0 \\ * & * & * & sE_\varepsilon - A_\varepsilon \end{bmatrix},$$

where

(i) $E_\eta, A_\eta \in \mathbb{R}^{l_\eta \times n_\eta}, l_\eta > n_\eta$, are such that $\operatorname{rk}(\lambda E_\eta - A_\eta) = n_\eta$ and $\operatorname{rk} E_\eta = n_\eta$;

(ii) $E_\infty, A_\infty \in \mathbb{R}^{n_\infty \times n_\infty}$, A_∞ is invertible and $A_\infty^{-1}E_\infty$ is nilpotent;

(iii) $E_f, A_f \in \mathbb{R}^{n_f \times n_f}$ and E_f is invertible;

(iv) $E_\varepsilon, A_\varepsilon \in \mathbb{R}^{l_\varepsilon \times n_\varepsilon}, l_\varepsilon < n_\varepsilon$, are such that $\operatorname{rk}(\lambda E_\varepsilon - A_\varepsilon) = l_\varepsilon$ and $\operatorname{rk} E_\varepsilon = l_\varepsilon$.

This form can be computed by a numerically stable algorithm, see [15, 1].

Step 3. Compute an embedding of the pencil $sE_\eta - A_\eta$, i.e., $K \in \mathbb{R}^{l_\eta \times (l_\eta - n_\eta)}$ such that $[K, sE_\eta - A_\eta]$ is unimodular. A numerically stable algorithm for the solution of this embedding problem using the staircase form is given in [2]. Define the unimodular matrix

$$U_1(s) := -\begin{bmatrix} K & sE_\eta - A_\eta & 0 \\ 0 & * & sE_\infty - A_\infty \end{bmatrix} \in \mathbb{R}[s]^{(l_\eta + n_\infty)\times(l_\eta + n_\infty)}$$

and consider

$$\left[\begin{array}{c|c} sE-A & -B \end{array}\right] \begin{bmatrix} T_2 & 0 \\ 0 & I_m \end{bmatrix} = S_1^\top \begin{bmatrix} S_2^\top & 0 \\ 0 & I_r \end{bmatrix} \left[\begin{array}{c|c} U_1(s) & 0 \\ \hline 0 & I_{l-l_\eta-n_\infty} \end{array}\right] \times \underbrace{\left[\begin{array}{c|ccc|c} 0 & 0 & 0 & 0 & 0 \\ -I_{n_\eta+n_\infty} & 0 & 0 & 0 & 0 \\ \hline * & sE_f - A_f & 0 & 0 & 0 \\ * & * & sE_\varepsilon - A_\varepsilon & 0 & 0 \\ * & * & * & -B_2 & 0 \end{array}\right]}_{=:\left[\begin{array}{c|c} s\overline{E}-\overline{A} & -\overline{B} \end{array}\right]}$$

Step 4. Compute column compressions $T_3 \in \mathscr{O}_{n_\varepsilon}, V_3 \in \mathscr{O}_m$ such that

$$E_\varepsilon T_3 = [\Sigma_1, 0], \quad B_2 V_3 = [\Sigma_2, 0],$$

where $\Sigma_1 \in \mathbb{R}^{l_\varepsilon \times l_\varepsilon}$ and $\Sigma_2 \in \mathbb{R}^{r\times r}$ are invertible. Consider

$$\left[\begin{array}{c|c} s\overline{E}-\overline{A} & -\overline{B} \end{array}\right] \left[\begin{array}{c|cc} I_{n-n_\varepsilon} & 0 & 0 \\ \hline 0 & T_3 & 0 \\ 0 & 0 & V_3 \end{array}\right] = \underbrace{\left[\begin{array}{cc|cc|cc} 0 & 0 & 0 & 0 & 0 & 0 \\ -I_{n_\eta+n_\infty} & 0 & 0 & 0 & 0 & 0 \\ sE_{31} - A_{31} & sE_f - A_f & 0 & 0 & 0 & 0 \\ \hline sE_{41} - A_{41} & * & s\Sigma_1 - A_{43} & -A_{44} & 0 & 0 \\ sE_{51} - A_{51} & * & * & sE_{54} - A_{54} & -\Sigma_2 & 0 \end{array}\right]}_{=:\left[\begin{array}{c|c} s\hat{E}-\hat{A} & -\hat{B} \end{array}\right]}$$

Step 5. Define the unimodular matrix

$$U_2(s) := \begin{bmatrix} I_{l_\eta - n_\eta} & 0 & 0 & 0 & 0 \\ 0 & I_{n_\eta + n_\infty} & 0 & 0 & 0 \\ 0 & -sE_{31} + A_{31} & I_{n_f} & 0 & 0 \\ 0 & -sE_{41} + A_{41} & 0 & I_{l_\varepsilon} & 0 \\ 0 & -sE_{51} + A_{51} & 0 & 0 & I_r \end{bmatrix} \in \mathbb{R}[s]^{l\times l}.$$

Step 6. Compute a singular value decomposition of $E_{54} \in \mathbb{R}^{r\times(n_\varepsilon - l_\varepsilon)}$, i.e., $S_4 \in \mathscr{O}_r$, $T_4 \in \mathscr{O}_{n_\varepsilon - l_\varepsilon}$ such that

$$S_4 E_{54} T_4 = \begin{bmatrix} \Sigma_3 & 0 \\ 0 & 0 \end{bmatrix},$$

where $\Sigma_3 \in \mathbb{R}^{q\times q}$ is invertible. Compute, using QR factorization (without pivoting), a column operation $V_4 \in \mathscr{O}_r$ such that

$$S_4 \Sigma_2 V_4 = \begin{bmatrix} \Sigma_{21} & 0 \\ * & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{21} \in \mathbb{R}^{q\times q}, \Sigma_{22} \in \mathbb{R}^{(r-q)\times(r-q)}$ are invertible. Then

$$\left[\begin{array}{c|c} s\hat{E}-\hat{A} & -\hat{B} \end{array}\right] \left[\begin{array}{c|cc|c} I_{n+l_\varepsilon-n_\varepsilon} & 0 & 0 & 0 \\ \hline 0 & T_4 & 0 & 0 \\ 0 & 0 & V_4 & 0 \\ \hline 0 & 0 & 0 & I_{m-r} \end{array}\right] = U_2(s) \begin{bmatrix} I_{l-r} & 0 \\ 0 & S_4^\top \end{bmatrix} \times \underbrace{\left[\begin{array}{ccc|cc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -I_{n_\eta+n_\infty} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & sE_f - A_f & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & s\Sigma_1 - A_{43} & -\tilde{A}_{44} & -\tilde{A}_{45} & 0 & 0 & 0 \\ \hline 0 & * & * & s\Sigma_3 - \tilde{A}_{54} & -\tilde{A}_{55} & -\Sigma_{21} & 0 & 0 \\ 0 & * & * & -\tilde{A}_{64} & -\tilde{A}_{65} & * & -\Sigma_{22} & 0 \end{array}\right]}_{=:\left[\begin{array}{c|c} s\tilde{E}-\tilde{A} & -\tilde{B} \end{array}\right]}.$$

Step 7. Define the permutation matrix

$$P := \begin{bmatrix} I_{n_\eta + n_\infty + n_f + l_\varepsilon + q} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_{n_\varepsilon - l_\varepsilon - q} & 0 \\ 0 & 0 & I_q & 0 & 0 \\ 0 & I_{r-q} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{m-r} \end{bmatrix} \in \mathscr{O}_{n+m}.$$

Then

$$\left[s\tilde{E}-\tilde{A} \;\middle|\; -\tilde{B}\right]P = \left[\begin{array}{ccccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -I_{n_\eta+n_\infty} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & sE_f-A_f & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & * & s\Sigma_1-A_{43} & -\tilde{A}_{44} & 0 & 0 & -\tilde{A}_{45} & 0 \\ 0 & * & * & s\Sigma_3-\tilde{A}_{54} & 0 & -\Sigma_{21} & -\tilde{A}_{55} & 0 \\ 0 & * & * & -\tilde{A}_{64} & -\Sigma_{22} & * & -\tilde{A}_{65} & 0 \end{array}\right] =: \left[\begin{array}{c|c} 0_{(l_\eta-n_\eta)\times\hat{n}} & 0_{(l_\eta-n_\eta)\times m} \\ sE_{\mathrm{reg}}-A_{\mathrm{reg}} & -B_{\mathrm{reg}} \end{array}\right]$$

where it should be noted that the system on the right hand side has other dimensions of state space and input space than the system on the left hand side.

Theorem 3.1. *Let $[E,A,B] \in \Sigma_{l,n,m}$ and let $[E_{\mathrm{reg}},A_{\mathrm{reg}},B_{\mathrm{reg}}] \in \Sigma_{\hat{n},\hat{n},\hat{m}}$ be the result of the regularization algorithm. Then $sE_{\mathrm{reg}}-A_{\mathrm{reg}} \in \mathbb{R}[s]^{\hat{n}\times\hat{n}}$ is regular and has index at most one.*

Proof. Denote

$$sE_{\mathrm{reg}}-A_{\mathrm{reg}} = \begin{bmatrix} -I_{n_\eta+n_\infty} & 0 & 0 & 0 & 0 \\ 0 & sE_f-A_f & 0 & 0 & 0 \\ 0 & sE_{42}-A_{42} & s\Sigma_1-A_{43} & -\tilde{A}_{44} & 0 \\ 0 & sE_{52}-A_{52} & sE_{53}-A_{53} & s\Sigma_3-\tilde{A}_{54} & 0 \\ 0 & sE_{62}-A_{62} & sE_{63}-A_{63} & -\tilde{A}_{64} & -\Sigma_{22} \end{bmatrix}$$

and observe that

$$\det\left(sE_{\mathrm{reg}}-A_{\mathrm{reg}}\right) = (-1)^{n_\eta+n_\infty+r-q}\det(\Sigma_{22})\det(sE_f-A_f) \times \det\left(\begin{bmatrix} s\Sigma_1-A_{43} & -\tilde{A}_{44} \\ sE_{53}-A_{53} & s\Sigma_3-\tilde{A}_{54} \end{bmatrix}\right),$$

which is a nonzero polynomial since $\left[\begin{smallmatrix} \Sigma_1 & 0 \\ E_{53} & \Sigma_3 \end{smallmatrix}\right]$ is invertible. This shows regularity of $sE_{\mathrm{reg}}-A_{\mathrm{reg}}$. To show that the index does not exceed one, we use that by [6, Eq. (3.4)] the index of $sE_{\mathrm{reg}}-A_{\mathrm{reg}}$ is at most one if, and only if,

$$\operatorname{im}A_{\mathrm{reg}} \subseteq \operatorname{im}E_{\mathrm{reg}} + A_{\mathrm{reg}}\ker E_{\mathrm{reg}}.$$

It is a simple calculation that

$$\ker E_{\mathrm{reg}} = \operatorname{im}\begin{bmatrix} I_{n_\eta+n_\infty} & 0 \\ 0 & 0 \\ 0 & I_{r-q} \end{bmatrix}$$

and hence

$$\operatorname{im}E_{\mathrm{reg}} + A_{\mathrm{reg}}\ker E_{\mathrm{reg}} = \operatorname{im}\begin{bmatrix} 0 \\ I_{n_f+l_\varepsilon+q} \\ * \end{bmatrix} + \operatorname{im}\begin{bmatrix} I_{n_\eta+n_\infty} & 0 \\ 0 & 0 \\ 0 & \Sigma_{22} \end{bmatrix} = \mathbb{R}^l.$$

This shows that $sE_{\mathrm{reg}}-A_{\mathrm{reg}}$ has index at most one. □

Note that the outcome of the regularization algorithm in particular improves [7, Thm. 5.1], because here we show that the additional state feedback used in [7] is not necessary. In other words, we may always choose $F=0$ in [7, Thm. 5.1].

4. Computational aspects

In this section we discuss the numerical reliability and the computational speed of the regularization algorithm presented in Section 3.

The computations in Steps 1 and 4–7 are certainly numerically stable, since they are based on the singular value decomposition and QR factorization (with pivoting) or they are mere definitions using the data at hand. The staircase form in Step 2 can also be computed by a numerically stable algorithm, see [15, 1]. For the computation of the unimodular embedding $U_1(s)$ in Step 3, we propose to use the numerically stable algorithm developed in [2].

We analyze the computational complexity for each step of the regularization algorithm separately:

Step 1. The computation of the row compression relies on a QR factorization with pivoting [10], which has a computational cost of $O(m(l^2+m^2))$ flops in the worst case according to [1]. Here, "flop" means *floating point operation*, which is a scalar addition or multiplication.

Step 2. By [1] the computation of the staircase form is possible with a cost of $O(l^2n)$ flops.

Step 3. According to [2] the computation of the embedding, which also uses the staircase form, has a computational cost of $O(l(l^2+n^2)+l^2n)$ flops.

Step 4. The computation of the column compressions again use QR factorization with pivoting and requires $O(n(l^2+n^2))$ and $O(m(l^2+m^2))$ flops in the worst case, resp.

Step 5. $U_2(s)$ is obtained at no cost.

Step 6. The singular value decomposition has a cost of $O(n(l^2+n^2))$ flops in the worst case according to [10], and the QR factorization of the invertible matrix $S_4\Sigma_2$ has a cost of $O(m^3)$ flops.

Step 7. P is obtained at no cost.

Summarizing, the computational cost of the regularization algorithm for a given system $[E,A,B] \in \Sigma_{l,n,m}$ is

$$O\left(l^2(l+n+m)+n^2(l+n)+m^3\right),$$

and hence the algorithm is cubic in the dimensions of the system.

Remark 4.1. If a relation of the form (3) is sought for the solution of the regularization problem, then $U(s)$ as in (2) computed by the regularization algorithm needs to be inverted. First recall that $U(s)=sU_1+U_2 \in \mathbb{R}[s]^{l\times l}$ is a matrix pencil. For the inversion of this pencil an algorithm is proposed in [2]. Again, the staircase form is used for the computation of $W(s)=U(s)^{-1}$, however the inversion of a triangular matrix is required as well. This cannot be avoided in general, see also the discussion in [2]. Hence, the algorithm is numerically stable up to the feasibility of this inversion problem.

Concerning computational complexity, the computation of $W(s)$ needs $O(ql^3)$ flops, where $q = \deg W(s)$. As discussed in [2] it is important to keep the degree q as small as possible. However, even if q is chosen minimal, in the worst case it may be as large as $l-1$ and hence the computation of $W(s)$ has quartic complexity in general. Note that q is also the index of the pencil $sU_1 + U_2$, which is regular and equivalent to a pencil of the form $sN - I$ for some nilpotent matrix N. This index is also revealed by the application of the staircase form.

Remark 4.2. We like to stress that rank decisions are an important issue in the computation of the regularization. The computation of the staircase form, which is used in Steps 2 and 3 of the regularization algorithm, involves a sequence of rank decisions, which in case of "bad" data with very small singular values close to the truncation tolerance, may lead to a wrong rank decision. This problem is unavoidable in general. However, it is desirable to keep the number of rank decisions as small as possible. Therefore, depending on the application, it may be recommendable to use condensed forms based on derivative arrays (see e.g. [3, 11] and the references therein) instead of the staircase form.

5. Comparison with [9]

In this section we provide a detailed comparison of our regularization algorithm with the method proposed in [9]. At first glance, a main difference is that we formulate the regularization in terms of explicit equivalence transformations performed on the pencil $[sE - A, -B]$ and additional column permutations, while in [9] a principle procedure is described. A detailed list of advantages and disadvantages of the method in [9] compared to our method is given below. After that we illustrate the difference by means of a short example.

Advantages of the method by Campbell et al. [9]:

+ It only uses variable transformations when unavoidable; the original variables are kept as long as possible.
+ It is made explicit where the physical background of the considered system can be exploited.
+ Regularization of the initial conditions can be done using the original state and input variables.
+ Fewer rank decisions are required in general.[1]

Disadvantages of the method by Campbell et al. [9]:

- The transformations leading to a strangeness free system are no equivalence transformations and not reversible in general.
- No explicit transformations for the reinterpretation of variables are provided. The decision for the choice of variables is left to the user and should "depend on the physical background of the system".
- During the reinterpretation it is possible that variables that are differentiated are selected as inputs. This requires the introduction of new variables, e.g. $\tilde{u} = \dot{u}$.
- The application of feedback is necessary in general.
- The result of the regularization method is not unique in general as it depends on the choice of variables performed by the user.
- Computational complexity of the method is not yet investigated.

[1] However, if condensed forms are used in our regularization algorithm instead of the staircase form, the number of rank decisions may be equally small, cf. Remark 4.2.

Example 5.1. We illustrate the different behavior of the methods by means of the system (1) with

$$E = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Method by Campbell et al. [9]: Since the system (in the variables x and u) is strangeness free (in the sense of [9], see also [11]) with $d = 1$ and $a = 1$, a reinterpretation of variables does not take place. In the last step, a feedback is applied to the system, i.e., with $F = [0, 1]$ the closed-loop system

$$E = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad A + BF = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

is constructed and clearly $sE - (A + BF)$ is regular with index at most one. We have

$$\begin{aligned} &\left[sE_{\mathrm{reg}} - A_{\mathrm{reg}} \;\middle|\; -B_{\mathrm{reg}} \right] \\ &\qquad = \left[sE - (A + BF) \;\middle|\; -B \right] = \left[\begin{array}{cc|c} s & s & 0 \\ 0 & -1 & -1 \end{array}\right]. \end{aligned}$$

Our method: In Step 1 we have $S_1 = I_2$ and $sE_1 - A_1 = [s, s]$. For the staircase form in Step 2 we find $S_2 = I_2$ and $T_2 = \left[\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right]$ and hence

$$S_2(sE_1 - A_1)T_2 = [s, 0],$$

which is in staircase form with $l_\eta = n_\eta = n_\infty = 0$, $n_f = 1$ and $E_f = [1], A_f = [0]$, and $l_\varepsilon = 0, n_\varepsilon = 1$, i.e., $E_\varepsilon, A_\varepsilon \in \mathbb{R}^{0\times 1}$. Steps 3–5 are not necessary and in Step 6 we find $q = 0$ because $sE_{54} - A_{54} = [0]$. Furthermore, $S_4 = T_4 = [1]$ and $\Sigma_{22} = [1]$, thus

$$\left[s\tilde{E} - \tilde{A} \;\middle|\; -\tilde{B} \right] = \left[\begin{array}{cc|c} s & 0 & 0 \\ 0 & 0 & -1 \end{array}\right].$$

In Step 7 we choose

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

and hence the regularization is

$$\begin{aligned} &\left[sE_{\mathrm{reg}} - A_{\mathrm{reg}} \;\middle|\; -B_{\mathrm{reg}} \right] \\ &\qquad = \left[sE - A \;\middle|\; -B \right] \begin{bmatrix} T_2 & 0 \\ 0 & 1 \end{bmatrix} P = \left[\begin{array}{cc|c} s & 0 & 0 \\ 0 & -1 & 0 \end{array}\right]. \end{aligned}$$

We clearly see that the results of the respective regularization procedure are different; there are several possibilities for the regularization in general, depending on the allowed transformations. Furthermore, it can be seen that the method by Campbell et al. [9] requires feedback, while we exclude the class of feedback transformations in our method. In the present paper we have shown that it is always possible to avoid feedback. In the above example, the number of differential variables ($d = 1$) and the number of algebraic variables ($a = 1$) sum up to the number of state variables ($n = 2$) and hence the method by Campbell et al. [9] does not recognize that a reinterpretation of variables would be appropriate.

Finally, we like to stress, and this is shown by the above example, that state space and input space transformations are unavoidable for the regularization in general, i.e., it is not possible to choose $T = I$ and $V = I$ in Problem 2.3. In [9] this is avoided, if possible, by augmenting the state space as explained in the list of disadvantages.

6. Numerical experiments

In this section we give numerical experiments illustrating the regularization algorithm presented in Section 3. For the implementation of the regularization algorithm in Matlab we used a simplified variant of the staircase algorithm described in [15].

For our example the original system $[E,A,B] \in \Sigma_{10,9,2}$ was based on the data

$$\left[sE_0 - A_0 \mid B_0\right] = \left[\begin{array}{ccc:c:ccccc:cc} -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & s-2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & s & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s & -3 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4 & -5 & 7 & 8 \end{array}\right]$$

to which we applied random orthogonal transformations

$$\left[sE - A \mid B\right] := \left[Q_l \cdot (sE_0 - A_0) \cdot Q_r \mid Q_l \cdot B_0 \cdot Q_b\right]$$

in order to make the system dense and "hide" the Kronecker structure.

Note that the pencil $\left[sE_0 - A_0 \mid B_0\right]$ is in staircase form and has a full column rank part (η-block) of dimension 5×3, an ODE part (f-block) of dimension 1×1, and a full row rank part (ε-block) of dimension 4×7 (including the columns of B). We used a tolerance of $100 * \varepsilon$ where ε is the machine accuracy ($\approx 10^{-16}$ for our machine running with IEEE double precision standard). We ran our regularization algorithm to see if we recover correctly the different substructures. The result of our algorithm is given in Figure 1, with the embedded constant columns in front in gray color. Whenever computed data were within tolerance level of an integer value, we rounded it to make the result more readable. Note that this was obtained by orthogonal transformations only.

The leading 5×5 submatrix in Figure 1 is clearly unimodular since it can be permuted to a block lower triangular matrix with constant invertible diagonal blocks:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0.963s & 0.268s & 0.188 & -0.982 & 0 \\ 0.268s & -0.963s & 0.982 & 0.188 & 0 \\ 0 & 0 & 0 & -s & 1 \end{bmatrix}$$

The new trailing 5×5 submatrix in Figure 1 is regular and of index at most 1 as is easily seen from its upper triangular form

$$\begin{bmatrix} s-2 & 0 & 0 & 0 & 0 \\ 0 & -s & -1 & 0 & 0 \\ 0 & 0 & s & -2 & 0 \\ 0 & 0 & 0 & s & 0 \\ 0 & 0 & 0 & 0 & -8 \end{bmatrix}.$$

The Matlab codes can be found in the supplementary material to the present article.

7. Conclusion

In the present paper we have presented a numerically stable algorithm for the computation of the regularization of a linear descriptor system by a combination of behavioral equivalence transformation, orthogonal state-space and input-space transformation and a permutation of variables. The latter is necessary since the initial choice of variables may not have been appropriate within the framework of the behavioral approach. A consequence of our algorithm is that the application of additional state feedback used in earlier publications [9, 7] is not necessary. We show that the regularization algorithm requires $O(p^3)$ operations, where p is the largest dimension of the descriptor system. A detailed comparison with the method proposed in [9] as well as a numerical example is provided.

Acknowledgement

We thank Volker Mehrmann (TU Berlin) for several constructive discussions.

References

[1] Beelen, T.G.J., Van Dooren, P.M., 1988a. An improved algorithm for the computation of Kronecker's canonical form of a singular pencil. Lin. Alg. Appl. 105, 9–65.

[2] Beelen, T.G.J., Van Dooren, P.M., 1988b. A pencil approach for embedding a polynomial matrix into a unimodular matrix. SIAM J. Matrix Anal. & Appl. 9, 77–89.

[3] Benner, P., Losse, P., Mehrmann, V., Voigt, M., 2015. Numerical linear algebra methods for linear differential-algebraic equations, in: Ilchmann, A., Reis, T. (Eds.), Surveys in Differential-Algebraic Equations III. Springer-Verlag, Berlin-Heidelberg. Differential-Algebraic Equations Forum. To appear.

$$
\left[\begin{array}{cc|ccc:c:cccc|ccc}
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.982 & 0 & 0.963s & 0.268s & 0.188 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.188 & 0 & 0.268s & -0.963s & 0.982 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & -s & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hdashline
0 & 0 & 0 & 0 & 0 & s-2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hdashline
0 & 0 & 0 & 0 & 0 & 0 & -s & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & s & -2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & s & 0 & -0.797 & -2.892 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -8 & 3.757 & -5.185 & 7
\end{array}\right]
$$

Figure 1: Computed transformed coefficient matrices with ε-rounding to the nearest integer. The first two columns (colored in gray) correspond to the embedding computed in Step 3 and do not belong to the regularization.

[4] Berger, T., 2014. On differential-algebraic control systems. Ph.D. thesis. Institut für Mathematik, Technische Universität Ilmenau. Universitätsverlag Ilmenau, Ilmenau, Germany.

[5] Berger, T., Ilchmann, A., Trenn, S., 2012. The quasi-Weierstraß form for regular matrix pencils. Lin. Alg. Appl. 436, 4052–4069. doi:10.1016/j.laa.2009.12.036.

[6] Berger, T., Reis, T., 2013. Controllability of linear differential-algebraic systems - a survey, in: Ilchmann, A., Reis, T. (Eds.), Surveys in Differential-Algebraic Equations I. Springer-Verlag, Berlin-Heidelberg. Differential-Algebraic Equations Forum, pp. 1–61.

[7] Berger, T., Reis, T., 2015. Regularization of linear time-invariant differential-algebraic systems. Syst. Control Lett. 78, 40–46.

[8] Brenan, K.E., Campbell, S.L., Petzold, L.R., 1989. Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. North-Holland, Amsterdam.

[9] Campbell, S.L., Kunkel, P., Mehrmann, V., 2012. Regularization of linear and nonlinear descriptor systems, in: Biegler, L.T., Campbell, S.L., Mehrmann, V. (Eds.), Control and Optimization with Differential-Algebraic Constraints. SIAM, Philadelphia. volume 23 of *Advances in Design and Control*, pp. 17–36.

[10] Golub, G.H., van Loan, C.F., 1996. Matrix Computations. Number 3 in Johns Hopkins series in the mathematical sciences. 3rd ed., Johns Hopkins University Press, Baltimore, MD.

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

[12] Lamour, R., März, R., Tischendorf, C., 2013. Differential Algebraic Equations: A Projector Based Analysis. volume 1 of *Differential-Algebraic Equations Forum*. Springer-Verlag, Heidelberg-Berlin.

[13] Polderman, J.W., Willems, J.C., 1998. Introduction to Mathematical Systems Theory. A Behavioral Approach. Springer-Verlag, New York.

[14] Riaza, R., 2008. Differential-Algebraic Systems. Analytical Aspects and Circuit Applications. World Scientific Publishing, Basel.

[15] Van Dooren, P.M., 1979. The computation of Kronecker's canonical form of a singular pencil. Lin. Alg. Appl. 27, 103–140.

[16] Willems, J.C., 1979. System theoretic models for the analysis of physical systems. Ricerche di Automatica 10, 71–106.

[17] Willems, J.C., 1991. Paradigms and puzzles in the theory of dynamical systems. IEEE Trans. Autom. Control AC-36, 259–294.

[18] Willems, J.C., 2007. The behavioral approach to open and interconnected systems. IEEE Control Systems Magazine 27, 46–99.