Characterization and Computation of a "Good Control"

Stephen L. Campbell* **Roswitha März****

May 3, 2004

Abstract

Models for physical systems often take the form of implicit or behavioral models. One important problem is the identification of which combinations of variables are good candidtates for control variables. This paper first provides one solution to this problem for linear time varying systems. The solution is shown to be related to a general optimization problem. It is then shown how these same algorithms can be extended to a large and important class of nonlinear systems.

1 Introduction

For many types of problems the initial formulation consists of a number of equations relating various variables and some of their derivatives. The result is a differential aglebraic equation (DAE) [6]

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

System (1.1) is also sometimes referred to as a behavioral model [16]. Usually only some of the variables appear differentiated so that we actually have a system in the form

$$F(x_1', x_1, \widetilde{x}_2, t) = 0. \tag{1.2}$$

The variables x_1 are sometimes referred to as differential variables and the $\widetilde{x}_2$ variables as algebraic variables.

The next step in many design procedures is to choose some portion of the variables to be control variables. Several considerations drive this choice of control variables. One is that the variables need to correspond to quantitites that are physically reasonable to use as controls. Often this includes such quantities as orientation angles, torques, etc. A second consideration is that the response of the system to the control should be nice in some sense. This second criteria usually rules out choosing part of x_1 as a control since it becomes differentiated when entering the system.

If some portion of the algebraic variables $\widetilde{x}_2$ are choosen as control variables and denoted u, and the remaining portion of $\widetilde{x}_2$ denoted x_2, we wind up with the system

$$F(x_1', x_1, x_2, u, t) = 0. \tag{1.3}$$

The equation (1.3) is interpreted as a DAE in x_1, x_2 with a forcing or input function u. That is, it is a controlled DAE. Physical considerations may rule out some choices of control variables, but an important question is whether there is a choice of control for which the DAE in (1.3) is solvable index one. In this case x_1, x_2 will be at least as smooth as u and there will be no constraints on the control variable. That is, for every piecewise continuous u, there will exist x_1, x_2 satisfying (1.3) and x_1, x_2 are uniquely determined by there intial values. The control can be free to achieve desired control objectives. We shall refer to this problem of identifying a good control as the *control selection problem* and it is the problem that we consider here. For the sake of brevity we assume that the reader has some familiarity with optimal control and the basic theory of DAEs [6].

*North Carolina State University, Department of Mathematics, Raleigh, NC. 27695-8205 USA. e-mail: `slc@math.ncsu.edu`. Research supported in part by NSF Grants DMS-0101802, DMS-020695, and ECS-0114095.

**Humboldt-Universität Berlin, Institut für Mathematik, D-10099 Berlin, Germany, maerz@mathematik.hu-berlin.de, supported by the DFG Research Center " Mathematics for key technologies" (FZT86) in Berlin.

In the simplest case, x_2, u can be chosen by a partitioning of $\widetilde{x}_2$ which in turn comes from a partitioning of x. However, in the general case it may require nonlinear time varying maps

$$\begin{pmatrix} x_1 \\ \widetilde{x}_2 \end{pmatrix} = K_1(x,t), \quad \begin{pmatrix} x_2 \\ u \end{pmatrix} = K_2(x_1, \widetilde{x}_2, t).$$

This is especialliy likely in problems where different subsystems are defined in terms of local coordinate systems.

Another factor complicating the control selection problem is that the systems can be large with complicated equations and hidden structure. Thus it is desirable to have a computational algorithm for the control selection problem.

It is interesting to note that while we think of x_2 and u in (1.3) differently from a control, or equivalently an input-state, point of view, this distinction disappears when we consider optimal control. Consider, for example, the control problem

$$\min \int_0^T L(x_1, x_2, u, t)dt, \tag{1.4a}$$

$$F(x_1', x_1, x_2, u, t) = 0. \tag{1.4b}$$

From the point of view of this optimization problem, there is no difference between x_2 and u. This fact is exploited, for example, by the optimal control code SOCS [4, 5] developed at the Boeing Company. It can solve problems of the form (1.4) provided the DAE in (1.4b) is index one.

It has been noted by a number of authors that if one writes down the Euler Lagrange necessary conditions for (1.4), then one gets a DAE in $\{x_1, x_2, u, \lambda\}$ and this DAE can be index one when the original DAE (1.4b) has index above one. In fact, these Euler Lagrange equations can result in a solvable DAE even if the dynamics (1.4b) are not solvable. In general, the addition of a cost criteria (1.4a) can result in a Euler Lagrange DAE which has higher or lower index than the original DAE (1.4b). A recent analysis [12] provides conditions and analysis which amount to characterizing when the Euler Lagrange equations are a solvable index one DAE. In this note we show that this characterization of when the Euler Lagrange equations are index one provides a characterization of a control selection that results in an index one plant.

Before proceeding we give two simple, but informative, examples.

Example 1.1 *Consider the system*

$$x_1' - x_1 - x_2 - x_3 = 0, \tag{1.5a}$$

$$x_1 + x_2 = 0. \tag{1.5b}$$

If we choose the control variable to be x_2, then we get an index two DAE in x_1, x_3 and the response in x_3 involves derivatives of u. On the other hand, if we choose the control to be x_3, then we get a solvable index one DAE in x_1, x_2. Also, x_1, x_2 are one order somoother than x_3.

Example 1.2 *Consider the system*

$$x_1' - x_1 - x_2 + x_4 = 0, \tag{1.6a}$$

$$x_2' - x_1 - x_2 - x_3 = 0, \tag{1.6b}$$

$$x_1 + x_2 = 0. \tag{1.6c}$$

Here x_1, x_2 are differential variables. For this problem any choice for the control of the form $u = k_1 x_3 + k_2 x_4$, and remaining state of the form $\widehat{x}_3 = k_3 x_3 + k_4 x_4$, results in an index two DAE in $x_1, x_2, \widehat{x}_3$.

The control selection problem is not new. What is new here is relating it to an optimal control problem and giving an algorithm that can determine a projection of the state that can be used as a control and also determine the rest of the state. Also, our results are applied to both linear time varying and a large class of nonlinear systems.

It should be pointed out that the question examined here is different ¿from that examined in the literature on feedback regularization of descriptor (DAE) systems. That literature, see [8, 9, 11, 17] for example, starts with a system of the form (1.3) along with an output equation. Then a feedback control is sought so that the resulting closed loop system is index one. In this feedback approach the way the control enters the system is fixed, the state is fixed, and a feedback loop is attached. Physically this means that extra connections are added to the original system. Also, what are considered state variables are not changed. The dynamics is changed. In contrast, in the problem we consider no extra feedback loops are added to the system. We are asking about what choices of inputs and state lead to the original system being index one.

2 Linear Time Invariant Case

We shall first briefly discuss the linear time invariant case since it is then easier to see what is happening. The next section concerns the technically more complicated linear time varying case. Finally in Section 5 we consider nonlinear systems.

We assume that we have a process in the form of

$$Ex' + Fx = 0. \tag{2.1}$$

Note that considering $Ex' + Fx = g$ where g are some source terms does not change the analysis so we omit these source terms. We also assume that we have a cost to be minimized of the form

$$J(x) = \frac{1}{2}\int_0^{\omega} x^T W x \, dt \tag{2.2}$$

where W is positive semi-definite. By performing a constant coordinate change in (2.1) we may assume without loss of generality that we have

$$\begin{aligned} x_1' - F_{11}x_1 - F_{12}x_2 &= 0, && (2.3a)\\ F_{21}x_1 + F_{22}x_2 &= 0. && (2.3b) \end{aligned}$$

Note here that F_{22} is not square and has more columns than rows.

The Hamiltonian is

$$H = \frac{1}{2}x^T W x - \lambda_1(x_1' - F_{11}x_1 - F_{12}x_2) - \lambda_2(F_{21}x_1 + F_{22}x_2),$$

which results in the necessary conditions

$$\begin{aligned} x_1' &= F_{11}x_1 + F_{12}x_2, && (2.4a)\\ 0 &= F_{21}x_1 + F_{22}x_2, && (2.4b)\\ -\lambda_1' &= F_{11}^T\lambda_1 + F_{21}^T\lambda_2 + W_{11}x_1 + W_{12}x_2, && (2.4c)\\ 0 &= F_{12}^T\lambda_1 + F_{22}^T\lambda_2 + W_{12}x_1 + W_{22}x_2. && (2.4d) \end{aligned}$$

The system (2.3) has a control selection which results in an index one plant if and only if F_{22} is full row rank. In this case there exists invertible, in fact orthogonal matrices, H so that

$$F_{22}H = \begin{pmatrix} \widehat{F}_{22} & \widehat{F}_{23} \end{pmatrix}, \tag{2.5}$$

where $\widehat{F}_{22}$ is nonsingular. If we let $x_2 = H\widehat{x}$, then we get that (2.3) becomes

$$\begin{aligned} x_1' - F_{11}x_1 - \widehat{F}_{12}\widehat{x}_2 - \widehat{F}_{13}\widehat{x}_3 &= 0, && (2.6a)\\ F_{21}x_1 + \widehat{F}_{22}\widehat{x}_2 + \widehat{F}_{23}\widehat{x}_3 &= 0. && (2.6b) \end{aligned}$$

Since $\widehat{F}_{22}$ is invertible, if we choose $\widehat{x}_3$ as the control variable, we get (2.6) is an index one DAE in $\{x_1, \widehat{x}_2\}$. For this choice we have that the state will generally have the same level of smoothness as the control since they are linked algebraically. If we choose H so that we also have $\widehat{F}_{23} = 0$, then the state depends on an integral of the control.

The requirement that $\widehat{F}_{22}$ be invertible in (2.5) does not uniquely determine either H or its action on any subspace. When we add the requirement that $\widehat{F}_{23} = 0$, then H^T must send the subspace $ker F_{22}$ to $0 \oplus R^p$.

We turn now to asking when (2.4) is index one in $\{x_1, x_2, \lambda_1, \lambda_2\}$. This will happen when

$$P = \begin{pmatrix} F_{22} & 0 \\ W_{22} & F_{22}^T \end{pmatrix}$$

is nonsingular. This implies that F_{22} is full row rank. Then using the H which makes $\widehat{F}_{23} = 0$ we get

$$\begin{pmatrix} I & 0 \\ 0 & H^T \end{pmatrix} P \begin{pmatrix} I & 0 \\ 0 & H \end{pmatrix} = \left(\begin{array}{cc|c} \widehat{F}_{22} & 0 & 0 \\ \hline \widehat{W}_{11} & \widehat{W}_{12} & \widehat{F}_{22}^T \\ \widehat{W}_{21} & \widehat{W}_{22} & 0 \end{array}\right).$$

This matrix will be invertible if and only if $\widehat{W}_{22} > 0$. That is, the weighting matrix W is positive definite on the subspace chosen for the control variable.

Proposition 2.1 *There exists a control selection for (2.3) which gives an index one DAE if and only if the Euler Lagrange equations for problem (2.2) are index one for any positive definite weighting matrix W.*

The particular choice of W is not important.

3 Linear Time Varying Case

We now turn to the more difficult linear time varying case. Suppose that, instead of (2.1), we have

$$E(t)x'(t) + F(t)x(t) = 0, \tag{3.1}$$

where we again have ommitted any source terms since they do not affect the solution of the control selection problem. If we wish to consider solutions where only some components of x are smooth, it is convenient to consider

$$(E(t)x(t))' + \widetilde{F}(t)x(t) = 0 \tag{3.2}$$

instead of (3.1). The problems to be solved include: characterizing when there exist a proper control selection for (3.1) and (3.2), giving an algorithm for computing these subspaces, and characterizing when we can get the response is smoother than the control.

In the linear time invariant case we first separated out the differential variables and then the algebraic variables. Now we suppose that, instead of (2.1), we have

$$D(t)(E(t)x(t))' + F(t)x(t) = 0, \;\; t \in \mathcal{I} := [0, \omega], \tag{3.3}$$

with continuous in t matrix coefficients. The weighting matrix in the cost now depends continuously on t. System (3.3) contains k equations, $x(t) \in I\!R^m$, and $D(t)$ has size $k \times n, E(t)$ is $n \times m, \; F(t)$ is $k \times m$. For continuously differentiable E, note that $Ex' = EE^+Ex' = EE^+(Ex)' - EE^+E'x$. Letting $D = EE^+$, where E^+ is the Moore-Penrose generalized inverse [10] of E, we see that equation (3.2) can be considered a special case of (3.3).

The weighting matrix $W(t)$ in (2.2) is supposed to be symmetric and positive semi-definite for all $t \in \mathcal{I}$. $W(t)$ is $m \times m, n \leq k < m$.

The leading term in the DAE (3.3) is assumed to be *stated properly*. That is, the coefficients D and E are well-matched so that there is no gap and no overlap between the subspaces $ker D(t)$, which is the nullspace of $D(t)$, and $im E(t)$ which is the range of $E(t)$ in $I\!R^n$. More precisely, we assume

$$ker D(t) \oplus im E(t) = I\!R^n, \;\; t \in \mathcal{I}, \tag{3.4}$$

($\oplus$ denotes the direct sum) and suppose these two subspaces are spanned by continuously differentiable basis functions. This will happen, for example, if D, E are continuously differentiable and have constant rank.

Let $R(t)$ denote the $n \times n$ projector matrix onto $im E(t)$ along $ker D(t)$. That is, $im R(t) = im E(t)$, $ker R(t) = ker D(t)$ for $t \in \mathcal{I}$. Then $R(t)$ is also continuously differentiable in t.

Under these assumptions $D(t), E(t)$ and the product $D(t)E(t) =: G_0(t)$ have common constant rank on $\mathcal{I}$, say r, where $r \leq n$. As pointed out in Theorem 4.5 of [2], in view of Hamiltonian properties in the inherent explicit ODE, it is preferable to model the problem with $r = n, \; ker D(t) = 0, \; R(t) = I_r$. In the following we drop the argument t almost everywhere. Then the relations are meant pointwise for each $t \in \mathcal{I}$. A solution of (3.3) is a continuous function $x : \mathcal{I} \to I\!R^m$, such that Ex is continuously differentiable, and (3.3) is satisfied for all $t \in \mathcal{I}$. Let $C^1_E(\mathcal{I}, I\!R^m)$ denote the corresponding function space, that is,

$$C^1_E(\mathcal{I}, I\!R^m) := \{x \in C(\mathcal{I}, I\!R^m) : Ex \in C^1(\mathcal{I}, I\!R^n)\}.$$

Define $G_0, P_0, Q_0, G_1, \mathcal{W}_0, \mathcal{G}_1$ by

$$\begin{aligned} G_0 &= DE, \; P_0 = G_0^+G_0 = E^+E, \; Q_0 = I_m - P_0, \\ \mathcal{W}_0 &= I_k - G_0G_0^+ = I_k - DD^+, \\ G_1 &= G_0 + FQ_0, \\ \mathcal{G}_1 &= G_0 + \mathcal{W}_0FQ_0. \end{aligned}$$

To understand the important role that these time varying matrices will play, note that $x = P_0x + Q_0x$. If we let $x_1 = P_0x$ and $x_2 = Q_0x$, then (3.3) becomes

$$D(Ex_1)' + FP_0x_1 + FQ_0x_2 = 0. \tag{3.5}$$

Also, if we multiply (3.5) by $\mathcal{W}_0$, then we get $\mathcal{W}_0F(Q_0x_2 + P_0x_1) = 0$. We will see that $\mathcal{W}_0FQ_0$ also plays an important role.

In terms of the linear time invariant matrix pencil $\{E, F\}$ from (2.1), we have that if E, F are square, then $G_1 = E + F(I - E^+E)$ being full rank is just the usual condition for the pencil to be is index one and regular where regular means that $det(sE + F)$ is not identically zero.

The orthoprojectors $P_0, Q_0, \mathcal{W}_0$ are continuous in t since G_0 is continuous and has constant rank. Observe that

$$G_1 = G_0 + \mathcal{W}_0FQ_0 + G_0G_0^+FQ_0 = \mathcal{G}_1(I_m + G_0^+FQ_0),$$

and

$$G_0 + \mathcal{W}_0F = (I_k + \mathcal{W}_0FG_0^+)\mathcal{G}_1,$$

with invertible factors $I_m + G_0^+FQ_0$ and $I_k + \mathcal{W}_0FG_0^+$. Therefore, all three matrices $G_1, \mathcal{G}_1$, and $G_0 + \mathcal{W}_0F$ have the same rank. Below, the conditions for G_1 to have full row rank k will play an important role. It means that there are no redundant equations in (3.3). In the context of (time invariant) controlled descriptor systems this is controllability at infinity [2, 3].

The problem of minimizing the cost (2.2) subject to the DAE (3.3) and the initial condition

$$D(0)E(0)x(0) = z_0 \tag{3.6}$$

where $z_0 \in im(D(0)E(0))$ is given, is closely related to the boundary value problem

$$\begin{aligned} D(Ex)' + Fx &= 0, && (3.7a)\\ -E^\top(D^\top\lambda)' + F^\top\lambda &= Wx, && (3.7b)\\ D(0)E(0)x(0) &= z_0, && (3.7c)\\ E(\omega)^\top D(\omega)^\top\lambda(\omega) &= 0. && (3.7d) \end{aligned}$$

This BVP states a sufficient optimality condition, which is also a necessary condition if it is also assumed that G_1 has full row rank (Proposition 3.2 below). We will refer to the DAE (3.7a), (3.7b) as the optimality DAE. Notice that it also has a properly stated leading term. From Theorem 3.3 of [2] we have

Proposition 3.1 *The optimality DAE (3.7a), (3.7b) is regular with (tractability) index one if and only if*

(i) G_1 *has full row rank* k *on* $\mathcal{I}$*, and*

(ii) $(\underbrace{G_0^\top + F^\top\mathcal{W}_0}_{k}\ \underbrace{WQ_0}_{m})$ *has full row rank* m*.*

Note that $G_0^\top + F^\top\mathcal{W}_0 = (G_0 + \mathcal{W}_0F)^\top$ and G_1 have the same rank. Further versions of condition (ii) are $ker \begin{pmatrix} G_0 \\ \mathcal{W}_0F \\ Q_0W \end{pmatrix} = 0$, or

$$\langle Wz, z\rangle > 0, \quad \text{for all } \ z \in kerG_0 \cap ker\mathcal{W}_0F, \ \ z \neq 0, \tag{3.8}$$

that is, W is strictly positive definite on this subspace.

Proposition 3.2 *If* $x_* \in C_E^1(\mathcal{I}, R^m)$, $\lambda_* \in C_{D^\top}^1(\mathcal{I}, R^k)$ *are a solution of the BVP (3.7), then* x_* *is a minimizer for (2.2),(3.3),(3.6). Conversely, if* G_1 *has full row rank* k*, and* $x_* \in C_E^1(\mathcal{I}, R^m)$ *minimizes (2.2),(3.3),(3.6), then there exists a* $\lambda_* \in C_{D^\top}^1(\mathcal{I}, R^k)$ *such that* x_*, λ_* *solve the BVP (3.7).*

Proof: We prove the first statement first. Let x_*, λ_* solve the BVP (3.7) and let $x \in C^1_E(\mathcal{I}, R^m)$ satisfy the DAE (3.3) as well as the initial condition (3.6). Let $\Delta x = x - x_*$. Then

$$\begin{aligned} J(x) - J(x_*) &= \frac{1}{2}\int_0^\omega \langle \Delta x(t), W(t)\Delta x(t)\rangle dt + \mathfrak{A}, \\ \text{where } \mathfrak{A} &= \int_0^\omega \langle W(t)\Delta x(t), x_*(t)\rangle dt = \int_0^\omega \langle \Delta x(t), W(t)\Delta x_*(t)\rangle dt \\ &= \int_0^\omega \langle \Delta x(t), -E(t)^\top (D(t)^\top \lambda_*(t))' + F(t)^\top \lambda_*(t)\rangle dt \\ &= \int_0^\omega \{-\langle E(t)\Delta x(t), (D(t)^\top \lambda_*(t))'\rangle + \langle F(t)\Delta x(t), \lambda_*(t)\rangle\} dt \\ &= \int_0^\omega \{\langle (E(t)\Delta x(t))', D(t)^\top \lambda_*(t)\rangle + \langle F(t)\Delta x(t), \lambda_*(t)\rangle\} dt \\ &= \int_0^\omega \langle D(t)(E(t)\Delta x(t))' + F(t)\Delta x(t), \lambda_*(t)\rangle dt = 0. \end{aligned}$$

Taking into account the positive semidefiniteness of W we find that $J(x) - J(x_*) \geq 0$ and x^* is a global minimum.

We now prove the second statement. Let $x_* \in C^1_E(\mathcal{I}, R^m)$ be a minimizer of the cost (2.2) subject to (3.3),(3.6). Since $\mathcal{G}_1 = G_0 + \mathcal{W}_0 F Q_0$ has full row rank k, there is a (continuous in t) orthogonal $m \times m$ matrix H such that

$$\mathcal{G}_1 H = (\underbrace{K}_{k} \ \ \underbrace{0}_{m-k}) \ , K \text{ invertible.}$$

This leads to $(G_0 + \mathcal{W}_0 F Q_0) H \begin{pmatrix} 0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = 0$. But $imG_0 = imD$ and $im\mathcal{W}_0 = (imD)^\perp$. Thus

$$G_0 H \begin{pmatrix} 0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = 0, \ \ \mathcal{W}_0 F Q_0 H \begin{pmatrix} 0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = 0. \tag{3.9}$$

Taking into account that $E = RE = RD^+DE = RD^+G_0$, we find that $EH \begin{pmatrix} 0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = RD^+G_0H \begin{pmatrix} 0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = 0$, and hence

$$EH = (\underbrace{\widetilde{E}}_{k} \ \ \underbrace{0}_{m-k}).$$

Let $\widetilde{P}_0 = \widetilde{E}^+\widetilde{E}$, $\widetilde{Q}_0 = I_k - \widetilde{P}_0$, and $FH = (\underbrace{\widetilde{F}}_{k} \ \ \underbrace{\widetilde{B}}_{m-k})$. Notice that $\begin{pmatrix} \widetilde{Q}_0 & 0 \\ 0 & I_{m-k} \end{pmatrix} = H^\top Q_0 H$ is the orthoprojector onto $kerEH$. Next we transform

$$x = H \begin{pmatrix} \widetilde{x} \\ \widetilde{u} \end{pmatrix} \begin{matrix} \}k \\ \}m-k \end{matrix} \tag{3.10}$$

so that (3.3) becomes

$$D(\widetilde{E}\widetilde{x})' + \widetilde{F}\widetilde{x} + \widetilde{B}\widetilde{u} = 0, \tag{3.11}$$

which looks like a controlled DAE. Because of $Ex = EH \begin{pmatrix} \widetilde{x} \\ \widetilde{u} \end{pmatrix} = (\widetilde{E} \ \ 0) \begin{pmatrix} \widetilde{x} \\ \widetilde{u} \end{pmatrix} = \widetilde{E}\widetilde{x}$, the smooth part of the variables in (3.11) is $\widetilde{E}\widetilde{x}$. The DAE (3.11) has a properly stated leading term, and, considered as a controlled DAE with $\widetilde{u}$ being the control, (3.11) is regular with (tractability) index one. Namely, we have

$$\begin{aligned} \widetilde{\mathcal{G}}_1 &:= D\widetilde{E} + \mathcal{W}_0\widetilde{F}\widetilde{Q}_0 = DEH \begin{pmatrix} I_k \\ 0 \end{pmatrix} + \mathcal{W}_0(\widetilde{F}\widetilde{B}) \begin{pmatrix} \widetilde{Q}_0 & 0 \\ 0 & I_{m-k} \end{pmatrix} \begin{pmatrix} I_k \\ 0 \end{pmatrix} \\ &= DEH \begin{pmatrix} I_k \\ 0 \end{pmatrix} + \mathcal{W}_0 FH \begin{pmatrix} \widetilde{Q}_0 & 0 \\ 0 & I_{m-k} \end{pmatrix} \begin{pmatrix} I_k \\ 0 \end{pmatrix} = DEH \begin{pmatrix} I_k \\ 0 \end{pmatrix} + \mathcal{W}_0 F Q_0 H \begin{pmatrix} I_k \\ 0 \end{pmatrix} \\ &= \mathcal{G}_1 H \begin{pmatrix} I_k \\ 0 \end{pmatrix} = (K \ \ 0) \begin{pmatrix} I_k \\ 0 \end{pmatrix} = K. \end{aligned}$$

Hence, $\widetilde{\mathcal{G}}_1$ is invertible, and so is $\widetilde{G}_1 := D\widetilde{E} + \widetilde{F}\widetilde{Q}_0$. This proves the index-one property. Moreover, since $P_0 H \begin{pmatrix} 0 \\ I_{m-k} \end{pmatrix} = G_0^+ G_0 H \begin{pmatrix} 0 \\ I_{m-k} \end{pmatrix} = 0$ and (3.9) holds, it follows that

$$\mathcal{W}_0 \widetilde{B} = \mathcal{W}_0 F H \begin{pmatrix} 0 \\ I_{m-k} \end{pmatrix} = \mathcal{W}_0 F (P_0 + Q_0) H \begin{pmatrix} 0 \\ I_{m-k} \end{pmatrix} = \mathcal{W}_0 F Q_0 H \begin{pmatrix} 0 \\ I_{m-k} \end{pmatrix} = 0.$$

This means that the control $\widetilde{u}$ in (3.11) is just directed to the explicit inherent ODE. Consequently, for each given control $\widetilde{u} \in C(\mathcal{I}, I\!R^{m-k})$, the initial value problem for (3.11) with the initial condition

$$D(0)\widetilde{E}(0)\widetilde{x}(0) = z_0 \tag{3.12}$$

has exactly one solution $\widetilde{x} \in C^1_{\widetilde{E}}(\mathcal{I}, I\!R^k)$.

Let $H^\top W H =: \begin{pmatrix} \widetilde{W} & \widetilde{S} \\ \widetilde{S}^\top & \widetilde{K} \end{pmatrix}$, where $\widetilde{W}, \widetilde{K}$ are symmetric and consider the transformed cost

$$\widetilde{J}(\widetilde{x}, \widetilde{u}) := J(H \begin{pmatrix} \widetilde{x} \\ \widetilde{u} \end{pmatrix}) = \frac{1}{2} \int_0^\omega \begin{pmatrix} \widetilde{x}(t) \\ \widetilde{u}(t) \end{pmatrix}^\top \begin{pmatrix} \widetilde{W}(t) & \widetilde{S}(t) \\ \widetilde{S}^\top(t) & \widetilde{K}(t) \end{pmatrix} \begin{pmatrix} \widetilde{x}(t) \\ \widetilde{u}(t) \end{pmatrix} dt. \tag{3.13}$$

where $\begin{pmatrix} \widetilde{x}_* \\ \widetilde{u}_* \end{pmatrix} = H^\top x_*$ is now an optimal pair for (3.13) subject to (3.11),(3.12). By Theorem 2.12 of [1], there is a function $\widetilde{\lambda}_* \in C^1_{D^\top}(\mathcal{I}, I\!R^k)$ such that $\widetilde{x}_*, \widetilde{\lambda}_*, \widetilde{u}_*$ form a solution of the BVP

$$\begin{aligned} D(\widetilde{E}\widetilde{x})' + \widetilde{F}\widetilde{x} + \widetilde{B}\widetilde{u} &= 0, && (3.14a)\\ -\widetilde{E}^\top (D^\top \widetilde{\lambda})' - \widetilde{W}\widetilde{x} + \widetilde{F}^\top \widetilde{\lambda} - \widetilde{S}\widetilde{u} &= 0, && (3.14b)\\ -\widetilde{S}^\top \widetilde{x} + \widetilde{B^\top}\widetilde{\lambda} - \widetilde{K}\widetilde{u} &= 0, && (3.14c)\\ D(0)\widetilde{E}(0)\widetilde{x}(0) &= z_0, && (3.14d)\\ \widetilde{E}(\omega)^\top D(\omega)^\top \widetilde{\lambda}(\omega) &= 0. && (3.14e) \end{aligned}$$

Put equations (3.14b) and (3.14c) together to get

$$-\begin{pmatrix} \widetilde{E}^\top \\ 0 \end{pmatrix} (D^\top \widetilde{\lambda})' + \begin{pmatrix} \widetilde{F}^\top \\ \widetilde{B}^\top \end{pmatrix} \widetilde{\lambda} = \begin{pmatrix} \widetilde{W} & \widetilde{S} \\ \widetilde{S}^\top & \widetilde{K} \end{pmatrix} \begin{pmatrix} \widetilde{x} \\ \widetilde{u} \end{pmatrix},$$

and scale by H, which leads to

$$-E^\top (D^\top \widetilde{\lambda})' + F^\top \widetilde{\lambda} = W x,$$

that is, to (3.7b). Now we realize that $\lambda_* := \widetilde{\lambda}_*$ satisfies equation (3.7b) for $x = x_*$. Since

$$\begin{aligned} &E(\omega)^\top D(\omega)^\top \lambda_*(\omega) = (D(\omega)E(\omega))^\top \widetilde{\lambda}_*(\omega) \\ &= (D(\omega)E(\omega)H(\omega)H(\omega)^\top)^\top \widetilde{\lambda}_*(\omega) = (D(\omega)(\widetilde{E}(\omega)\ 0)H(\omega)^\top)^\top \widetilde{\lambda}_*(\omega) \\ &= H(\omega) \begin{pmatrix} \widetilde{E}(\omega)^\top \\ 0 \end{pmatrix} D(\omega)^\top \widetilde{\lambda}_*(\omega) = 0, \end{aligned}$$

this function λ_* satisfies the end conditions (3.7d), too. □

At this place it is worth mentioning that, if the full rank condition for G_1 does not hold, then a corresponding λ_* satisfying (3.7b), (3.7d) does not necessarily exist although there is a minimizer x_*. See Example 2.11 of [1].

The proof of the second assertion in Proposition 3.2 has its own value with respect to a good control selection in (3.3). Actually, we have proved

Proposition 3.3 *If $G_1 = DE + FQ_0$ in (3.3) has full row rank k, then there is a control selection $x = H\begin{pmatrix}\widetilde{x}\\ \widetilde{u}\end{pmatrix}$, H continuous and orthogonal, such that the resulting controlled DAE (3.11) is regular with index one, and the responses $\widetilde{x}$ are smoother than the controls ($\widetilde{x} \in C^1_{\widetilde{E}}$ for $\widetilde{u} \in C$).*

In Example 1.2 above, where no good control selection (leading to an index one DAE) exists, the corresponding matrix G_1 is

$$G_1 = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and it obviously does not have full row rank. The next assertion says that this is also the general situation: The full rank condition for G_1 is necessary for an index-one control selection.

Proposition 3.4 *If there is a control selection $x = H\begin{pmatrix}\widetilde{x}\\ \widetilde{u}\end{pmatrix}\begin{matrix}\}k\\ \}m-k\end{matrix}$ for (3.3), H invertible, such that $EH = (\widetilde{E}\ \ 0)$, $FH =: (\widetilde{F}\ \widetilde{B})$, and the resulting controlled DAE*

$$D(\widetilde{E}\widetilde{x})' + \widetilde{F}\widetilde{x} + \widetilde{B}\widetilde{u} = 0$$

is regular with (tractability) index one, then G_1 has full row rank k.

Proof: Due to the index-one property, the $k \times k$ matrix $D\widetilde{E} + \mathcal{W}_0\widetilde{F}$ is invertible. Compute $G_0 + \mathcal{W}_0 F = (G_0 H + \mathcal{W}_0 F H)H^{-1} = (D\widetilde{E} + \mathcal{W}_0\widetilde{F}\ \ \mathcal{W}_0\widetilde{B})H^{-1}$. Hence $rank(G_0 + \mathcal{W}_0 F) \geq rank(D\widetilde{E} + \mathcal{W}_0 F) = k$. Also $G_0 + \mathcal{W}_0 F$ has $rank\ k$, and so does G_1. □

In summary we have shown that if there is a good (index-one) control selection for the DAE (3.3), then the optimality DAE (3.7a),(3.7b) has index one for all weighting matrices W being positive definite on the subspace $kerG_0 \cap ker\mathcal{W}_0 F$. We have also shown that conversely, if the optimality DAE (3.7a),(3.7b) has index one, then there is a good control selection, and W is positive definite on the subspace $kerG_0 \cap ker\mathcal{W}_0 F$.

4 Practical realization of the control selection

The preceding analysis provides the basis for a control selection algorithm. In general, if one has a continuous matrix function $G(t)$ and takes its singular value decompostion (SVD) the factors do not vary continuously in t due to discrete numerical decisions in the implementations. It is possible to get continuous factors by using a continuous SVD [7, 13, 15].

Apply a (continuous) SVD to $G_0 = DE$ to get

$$G_0 = \mathcal{U}^\top \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \mathcal{V}, \quad \Sigma \text{ is } r \times r, \quad \text{and invertible.} \tag{4.15}$$

This gives the projectors

$$Q_0 = \mathcal{V}^\top \begin{pmatrix} 0 & 0 \\ 0 & I_{m-r} \end{pmatrix} \mathcal{V}, \quad \mathcal{W}_0 = \mathcal{U}^\top \begin{pmatrix} 0 & 0 \\ 0 & I_{k-r} \end{pmatrix} \mathcal{U}.$$

Let

$$\mathcal{G}_1 = \mathcal{U}^\top \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \mathcal{V} + \mathcal{U}^\top \begin{pmatrix} I & 0 \\ 0 & I_{k-r} \end{pmatrix} \underbrace{\mathcal{U}F\mathcal{V}^\top}_{\begin{array}{c} \| \\ \begin{pmatrix} \bar{F}_{11} & \bar{F}_{12} \\ \bar{F}_{21} & \bar{F}_{22} \end{pmatrix} \begin{matrix} \}r \\ \}k-r \end{matrix} \\ \underbrace{\qquad}_{r}\ \underbrace{\qquad}_{m-r} \end{array}} \begin{pmatrix} 0 & 0 \\ 0 & I_{m-r} \end{pmatrix} \mathcal{V}.$$

Then

$$\mathcal{G}_1 = \mathcal{U}^\top \begin{pmatrix} \Sigma & 0 \\ \underbrace{0}_{r} & \underbrace{F_{22}}_{m-r} \end{pmatrix} \mathcal{V}, \quad F_{22} \text{ is } (k-r) \times (m-r).$$

Next we compute a continuous in t orthogonal $\check{H}$ such that

$$\check{H}^\top F_{22}^\top = \begin{pmatrix} \check{K} \\ 0 \end{pmatrix} \begin{matrix} \}k-r \\ \end{matrix} \tag{4.16}$$

and put

$$H = \mathcal{V}^\top \begin{pmatrix} I_r & 0 \\ 0 & \check{H} \end{pmatrix}. \tag{4.17}$$

We can see that this yields the correct H since

$$\begin{aligned} H^\top \mathcal{G}_1^\top &= \begin{pmatrix} I_r & 0 \\ 0 & \check{H}^\top \end{pmatrix} \mathcal{V}\mathcal{V}^\top \begin{pmatrix} \Sigma^\top & 0 \\ 0 & F_{22}^\top \end{pmatrix} \mathcal{U} = \begin{pmatrix} \Sigma^\top & 0 \\ 0 & \check{K} \\ 0 & 0 \end{pmatrix} \mathcal{U} \\ &=: \begin{pmatrix} K^\top \\ 0 \end{pmatrix} \begin{matrix} \}k \\ \}m-k \end{matrix}. \end{aligned}$$

Note that to actually carry out this procedure only requires first computing (4.15) and noting the size of Σ. Then $\mathcal{U}F\mathcal{V}^\top$ is computed and F_{22} extracted. Then we do (4.16) and finally (4.17).

The analysis and algorithms greatly simplify if we have a semi-explicit DAE such as

$$x_1' + F_{11}x_1 + F_{12}x_2 = 0, \quad \}k_1 \tag{4.18a}$$

$$F_{21}x_1 + F_{22}x_2 = 0, \quad \}k_2 \tag{4.18b}$$

as a special case of (3.3) with time varying F, $k = k_1 + k_2, r = n = k_1$ We merely have to choose an orthogonal (continuous) $(m-k_1) \times (m-k_1)$ matrix function $\check{H}$ such that

$$\check{H}^\top F_{22}^\top = \begin{pmatrix} \check{K} \\ 0 \end{pmatrix},$$

with $\check{K}$ nonsingular and let $H = \begin{pmatrix} I_r & 0 \\ 0 & \check{H} \end{pmatrix}$.

5 Nonlinear systems

Suppose now that we have a nonlinear system of the form

$$\mathcal{F}((E(t)x(t))', x(t), t) = 0. \tag{5.1}$$

If we have a trajectory x_* and linearize around x_*, we get a linear time varying DAE in $\delta = x - x_*$. If we have a continuous invertible function $H(t)$, it is straight forward to show that performing the change of coordinates $x = H\hat{x}$ either before or after the linearization results in the same linear time varying DAE. What happens if we apply the control selection algorithm to the linearization and then use this control choice for the nonlinear system?

It is easy to see that if the original DAE is semi-explicit, then the control selection computed by the algorithm of this paper will provide a local control selection for the original nonlinear DAE. The only thing that might be lost is the state being smoother than the control. In the following we show the same to be true for a further large class of nonlinear DAEs.

Let the function $\mathcal{F}$ in (5.1) be given as

$$\mathcal{F}(y, x, t) = D(x, t)y + f(x, t), \quad x \in \mathbb{R}^m, y \in \mathbb{R}^n, t \in \mathcal{I}, \tag{5.2}$$

where $D(x,t)$ is a $k \times n$ matrix, $f(x,t) \in I\!R^k$. Let $E(t)$ and $D(x,t)$ be well-matched, that is,

$$ker D(x,t) \oplus im E(t) = I\!R^n, \quad x \in I\!R^m, t \in \mathcal{I}, \tag{5.3}$$

and $ker D(x,t)$ is independent of x. $\mathcal{F}$ is assumed to be continuous with continuous partial derivatives $\mathcal{F}_y = D$ and $\mathcal{F}_x$. The two subspaces involved in (5.3) are assumed to be spanned by continuously differentiable basis functions. In summary, equation (5.1) is now a quasi-linear DAE of the form

$$D(x(t),t)(E(t)x(t))' + f(x(t),t) = 0$$

that has a properly stated leading term [14]. The homogeneous part of the linear DAE resulting from linearization along a given function (not necessarily a solution) $x_* \in C^1_E(\mathcal{I}, I\!R^m)$ is [14]

$$D_*(t)(E(t)\delta(t))' + F_*(t)\delta(t) = 0, \quad t \in \mathcal{I}, \tag{5.4}$$

where

$$\begin{aligned} D_*(t) &= D(x_*(t),t) = \mathcal{F}_y((E(t)x_*(t))', x_*(t), t), \\ F_*(t) &= \mathcal{F}_x((E(t)x_*(t))', x_*(t), t). \end{aligned}$$

Observe that the linearized DAE (5.4) also has a properly stated leading term. Hence, we may apply the results given in Sections 3 and 4.

We continue using the orthoprojections

$$P_0(t) = E(t)^+ E(t), \quad Q_0(t) = I - P_0(t),$$

as well as the continuous matrices

$$G_{*0}(t) = D_*(t)E(t), \quad G_{*1}(t) = G_{*0}(t) + F_{*0}(t)Q_0(t),$$

introduced in Section 3 for the linear DAE (3.3). Here by $*$ we indicate the application of these ideas to the linearized DAE (5.4). Additionally we introduce for the nonlinear DAE (5.4) the matrices $G_0(x,t),\ G_1(y,x,t)$, which are defined pointwise for $y \in I\!R^n,\ x \in I\!R^n,\ t \in \mathcal{I}$, by

$$\begin{aligned} G_0(x,t) &= D(x,t)E(t), \\ G_1(y,x,t) &= G_0(x,t) + \mathcal{F}_x(y,x,t)Q_0(t). \end{aligned}$$

By construction it follows that

$$G_0(x_*(t),t) = G_{*0}(t), \quad G_1((E(t)x_*(t))', x_*(t), t) = G_{*1}(t). \tag{5.5}$$

Next, assuming $G_{*1}(t)$ to be of full row rank k, we compute a good (continuous) control selection $H_* = (H_{*1}\ H_{*2})$ for the linearized DAE (5.4) as described in Sections 3 and 4. Recall that we then have $EH_* = (\tilde{E}_*\ 0),\ F_*H =: (\tilde{F}_*\ \tilde{B}_*)$, and $\tilde{Q}_{*0} := H_{*1}^T Q_0 H_{*1}$ is the orthoprojector of $I\!R^k$ onto $ker \tilde{E}_*$. Moreover, $\tilde{G}_{*1} := D_*\tilde{E}_* + \tilde{F}_*\tilde{Q}_{*0}$ remains nonsingular due to the index-one property of the resulting controlled linear DAE.

Applying the transformation

$$x = H_* \begin{pmatrix} \tilde{x} \\ \tilde{u} \end{pmatrix} = H_{*1}\tilde{x} + H_{*2}\tilde{u} \tag{5.6}$$

to the nonlinear DAE (5.1) and taking into account that $Ex = \tilde{E}_*\tilde{x}$ holds, we obtain

$$\mathcal{F}((\tilde{E}_*(t)\tilde{x}(t))',\ H_{*1}(t)\tilde{x}(t) + H_{*2}(t)\tilde{u}(t), t) = 0, \tag{5.7}$$

which we can consider to be a controlled DAE. Note that (5.7) is again a DAE with properly stated leading term. We will now show this DAE is regular with index one locally around the transformed functions $\tilde{x}_* = (I\ 0)H_*^T x_*,\ \ \tilde{u}_* = (I\ 0)H_*^T x_*$. The corresponding test matrices for the controlled nonlinear DAE (5.7) are [14]

$$\begin{aligned} \tilde{G}_0(\tilde{x},t,\tilde{u}) &= D(H_{*1}(t)\tilde{x} + H_{*2}(t)\tilde{u}, t)\tilde{E}_*(t), \\ \tilde{G}_1(y,\tilde{x},t,\tilde{u}) &= \tilde{G}_0(\tilde{x},t,\tilde{u}) + \mathcal{F}_x(y, H_{*1}(t)\tilde{x} + H_{*2}(t)\tilde{u}, t)H_{*1}(t)\tilde{Q}_{*0}(t), \end{aligned}$$

for $y \in I\!R^n,\ \tilde{x} \in I\!R^k,\ t \in \mathcal{I},\ \tilde{u} \in I\!R^{m-k}$.

The $k \times k$ matrix $\tilde{G}_1(y, \tilde{x}, t, \tilde{u})$ depends continuously on its arguments, and we have

$$\begin{aligned}\tilde{G}_1((\tilde{E}_*(t)\tilde{x}_*(t))', \tilde{x}_*(t), t, \tilde{u}(t)) &= D_*(t)\tilde{E}_*(t) + F_*(t)H_{*1}(t)\tilde{Q}_{*0}(t) \\ &= D_*(t)\tilde{E}_*(t) + \tilde{F}_*(t)\tilde{Q}_{*0}(t) = \tilde{G}_{*1}(t).\end{aligned}$$

Since $\tilde{G}_{*1}(t)$ remains nonsingular on $\mathcal{I}$, there is a neighbourhood of the trajectory $T_* := \{((E(t)x_*(t))', \tilde{x}_*(t), t, \tilde{u}(t)) : t \in \mathcal{I}\}$ in $I\!R^n \times I\!R^k \times I\!R \times I\!R^{m-k}$, where $\tilde{G}_1(y, \tilde{x}, t, \tilde{u})$ is nonsingular and, hence, the DAE (5.7) has index one.

Proposition 5.1 *Suppose that the nonlinear DAE (5.1) has a full-row-rank matrix $G_1(y, x, t)$ for all x, y, t. Then for each arbitrary reference function $x_* \in C^1_E(\mathcal{I}, I\!R^m)$, there is a continuous linear control selection H_* which transforms the DAE (5.1) locally into a contolled regular index one DAE.*

Proof: Since $G_1(u, x, t)$ has uniformly rank k, relation (5.5) ensures the full-rank property for $G_{*1}(t)$, which is needed for computing the control selection H_*. □

If the nonlinear function $\mathcal{F}$ defining the DAE (5.1) is not given globally on $I\!R^n \times I\!R^m \times \mathcal{I}$ but on a subset $\mathcal{G} \times \mathcal{I},\ \mathcal{G} \subseteq I\!R^n \times I\!R^m$ open, one can proceed similarly, keeping all function values $x_*(t)$ etc. in the right domains.

6 Conclusion

We have presented a solution of the control selection problem which leads to computational algorihtms for determining a good control selection for both linear and many nonlinear system. The algorithm not only computes such a good control selection when it exists but also tells us when no such selection exits. Application of the algorithm does not require any prior manipulation of the system to put it into some special form. The algorithms for control selection presented here may be applied directly to the original system.

References

[1] A. Backes, *A necessary optimality condition for the linear-quadratic DAE control problem*, Humboldt-Universität zu Berlin, Preprint 16, 2003, www.mathematik.hu-berlin.de/publ/pre/2003/p-list-03.html

[2] K. Balla, G. Kurina and R. März, *Index criteria for differential algebraic equations arising form linear-quadratic optimal control problems*, Humboldt-Universität zu Berlin, Preprint 14, 2003, www.mathematik.hu-berlin.de/publ/pre/2003/p-list-03.html

[3] D.J. Bender and A.J. Laub, *The linear-quadratic optimal regulator for descriptor systems*, IEEE Transactions on Automatic Control, 32 (1987), 672– 688.

[4] J. T. Betts, *Sparse nonlinear programming test problems (Release 1)*, BCSTECK-93-047, Boeing Computer Services, 1993.

[5] J. T. Betts, *Practical Methods for Optimal Control using Nonlinear Programming*, SIAM, Philadelphia, 2001.

[6] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM Publications, 1996.

[7] A. Bunse-Gerstner, R. Byers, V. Mehrmann and N.K. Nichols, *Numerical computation of an analytic singular value decomposition of a matrix valued function*, Numerische Mathematik, 60 (1991), 1–40.

[8] R. Byers, P. Kunkel and V. Mehrmann, *Regularization of linear descriptor systems with variable coefficients*, SIAM J. Contr. Optimiz. 35 (1997), 117–133.

[9] R. Byers, A. Bunse-Gerstner, V. Mehrmann, and N.K. Nichols, *Feedback design for regularizing descriptor systems*, Linear Algebra and Applications, 299 (1999), 119–151.

[10] S. L. Campbell and C. D. Meyer, Jr. *Generalized Inverses of Linear Transformations*, Dover, 1991.

[11] D. Chu, V. Mehrmann, and N. K. Nichols *Minimum norm regularization of descriptor systems by mixed output feedback*, Linear Algebra and Applications, 296 (1999), 39–77.

[12] G. A. Kurina and R. März, *On linear-quadratic control problems for time-varying descriptor systems*, SIAM J. Control Opt., 42 (2004), 2062–2077.

[13] P. Kunkel and V. Mehrmann, *Smooth factorizations of matrix valued functions and their derivatives*, Numerische Mathematik, 60 (1991), 115–132.

[14] R. März, *Differential algebraic systems with properly stated leading term and MNA equations*, International Series of Numerical Mathematics, 146 (2003), 135–151.

[15] V. Mehrmann and W. Rath, *Numerical methods for the computation of analytic singular value decompositions*, Electronic Transactions in Numerical Analysis, 1 (1993), 72-88.

[16] J. W. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory: A Behavioral Approach*, Springer Texts in Applied Mathematics No. 26, 1997.

[17] R. Yu and D. H. Wang, *Further study on structural properties of LTI singular systems under output feedback*, Automatica, 39 (2003), 685–692.