Numer. Math. 72: 173–196 (1995)

Numerische Mathematik

The index of general nonlinear DAEs

Stephen L. Campbell[1,⋆], **C. William Gear**[2]

[1] Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 USA
phone: 1-919-515-3300; email: slc@math.ncsu.edu; fax: 1-919-515-3798
[2] NEC Research Institute, 4 Independence Way, Princeton, NJ 08540, USA

Received November 15, 1993 / Revised version received December 23, 1994

Summary. In the last few years there has been considerable research on differential algebraic equations (DAEs) $F(t,y,y')=0$ where $F_{y'}$ is identically singular. Much of the mathematical effort has focused on computing a solution that is assumed to exist. More recently there has been some discussion of solvability of DAEs. There has historically been some imprecision in the use of the two key concepts of solvability and index for DAEs. The index is also important in control and systems theory but with different terminology. The consideration of increasingly complex nonlinear DAEs makes a clear and correct development necessary. This paper will try to clarify several points concerning the index. After establishing some new and more precise terminology that we need, some inaccuracies in the literature will be corrected. The two types of indices most frequently used, the differentiation index and the perturbation index, are defined with respect to solutions of unperturbed problems. Examples are given to show that these indices can be very different for the same problem. We define new "maximum indices," which are the maxima of earlier indices in a neighborhood of the solution over a set of perturbations and show that these indices are simply related to each other. These indices are also related to an index defined in terms of Jacobians.

Mathematics Subject Classification (1991): 65L20

1. Introduction

Many physical problems are most easily initially modeled as a nonlinear implicit system of differential and algebraic equations (DAEs),

⋆ Research supported in part by the U. S. Army Research Office under DAAL03-89-D-0003 and the National Science Foundation under DMS-9122745

$$F(t, y, y') = 0 \tag{1}$$

with $F_{y'} = \partial F / \partial y'$ identically singular [Brenan, Campbell and Petzold (1989)]. A variety of numerical methods have been developed for (1) ranging from backward differentiation (BDF) to implicit Runge-Kutta (IRK) methods [Brenan, Campbell and Petzold (1989), Griepentrog and März (1986), Hairer, Lubich and Roche (1989), Potra and Rheinboldt (1991)]. These methods are only directly suitable for lower index problems and often require that the problem have special structure. Although many important applications can be solved by these methods there is a need for more general approaches [Campbell (1989), Campbell and Moore (1994a,b)].

Most of the mathematical research on DAEs has focused on computing a solution that is assumed to exist. More recently there have been some existence results using differential geometry either in a recursive manner [Griepentrog (1992), Kunkel and Mehrmann (1994), Rabier and Rheinboldt (1991), Rabier and Rheinboldt (1994a), Rabier and Rheinboldt (1994d), Reich (1990), Reich (1991)], or under somewhat different assumptions [Campbell and Griepentrog (1995)]. Existence and uniqueness results for some index two and three systems using linearization techniques are also given in [März (1992)]. Two key concepts, variously defined, are solvability and the index. In this paper we shall assume that the DAE (1) is solvable and focus on the index. Solvability involves existence and uniqueness and is carefully defined in the next Section. Most of the DAE literature has concerned solvable DAEs. While many DAEs are solvable, the restriction to solvable systems does rule out some interesting DAEs including those with impasse points [Rabier and Rheinboldt (1994a,b)] and those with nonunique solutions [Bender and Laub (1987), Kunkel and Mehrmann (1993)]. Consideration of these other cases would not only greatly lengthen this paper but bury the points we wish to make in a even more technical detail.

The index is also important in systems and control theory where it is closely related to such nonlinear control theory concepts as the relative degree [Campbell (1995), Fliess, Lévine and Rouchon (1993a), Fliess, Lévine and Rouchon (1993b)]. We shall comment on this briefly in Sect. 4.

Historically most work on nonlinear DAEs has focused on systems with a special structure for which most definitions of the index were closely related. Recent consideration of more complex nonlinear DAEs has shown the need to carefully examine these concepts.

The theory for linear DAEs was developed by considering systems of the form

$$E(t)y' + B(t)y = \delta(t) \tag{2}$$

for all smooth forcing functions δ. The nonlinear theory was developed from (1). Even though (2) is in the form of (1), the nonlinear theory has traditionally corresponded to considering (2) with *fixed* δ. The presence of the (arbitrary) forcing function δ in (2) and its absence in (1) is one source of some of the confusion in the literature. In this paper we shall try to clarify some of this discussion and

put it into what we hope is a better perspective. Our current understanding owes much to the several papers listed in the necessarily incomplete bibliography.

In order to avoid technical difficulties that while sometimes important greatly complicate the discussion, we shall assume that (1) possesses all the derivatives with respect to the variables (t, y, y') that one needs. Similarly in this paper we shall not worry about the smoothness of solutions although that is also obviously important. Results guaranteeing smoothness can be found in [Rabier and Rheinboldt (1991), Rabier and Rheinboldt (1994c)]. We shall focus on the amount of differentiation of the DAE that is required to determine y'. Even with these simplifying assumptions we shall see that several of the existing conceptions about the index of DAEs need modification. Unfortunately, we shall have to introduce some new terminology, but we shall try to keep it to a minimum.

First we need to be careful about what we mean by solvability.

2. Solvability

Intuitively solvability means the existence of a well behaved family of solutions. However, the term is used in different ways.

2.1. Geometric solvability

For convenience we only consider square systems. Suppose that the DAE (1) is a system of n equations in the $(2n + 1)$-dimensional variable (t, y, y'). We will sometimes denote y' by v when we wish to emphasize it is an algebraic variable and not a derivative. Open sets are always taken to be connected. We define $\mathbb{R}^0$ to be $\{0\}$.

Definition 1. Let $\Omega \subset \mathbb{R}^{2n+1}$ be a connected open set. The DAE (1) is *geometrically solvable* on Ω if there are connected open sets $\Lambda \subset \mathbb{R}^\rho$ and $\mathcal{I} \subset \mathbb{R}$ and a function $\Phi(t, \lambda)$ such that the following properties hold

1. $\Theta(t, \lambda) = (t, \Phi(t, \lambda))$ is a diffeomorphism of $\mathcal{I} \times \Lambda$ into $\mathbb{R}^{n+1}$.
2. $\Phi(t, \lambda)$ is a solution of (1) for each value of $\lambda \in \Lambda$.
3. $(t, \Phi(t, \lambda), \Phi_t(t, \lambda)) \in \Omega$ for each $\lambda \in \Lambda$ and $t \in \mathcal{I}$.
4. If y is a solution of (1) such that $(t, y(t), y'(t)) \in \Omega$ for some $t \in \mathcal{I}$, then $y(t) = \Phi(t, \lambda)$ for some $\lambda \in \Lambda$.

A value (t, y) is called *(geometrically) consistent* if $y = \Phi(t, z)$ for some z. Hence consistent initial conditions for geometrically solvable DAEs uniquely determine solutions in Ω.

Geometric solvability is close to the usual definition of solvability. It says that there is a well behaved manifold of solutions and that a given solution is uniquely determined by an initial condition. It does not say anything about how this determination is carried out. Geometric solvability is the same as regularity [Reich (1991)].

Example 1. Note that

$$ty = 0 \tag{3}$$

is geometrically solvable on $\Omega = \mathbb{R}^3$ by taking $\rho = 0$ and $\Phi(t, 0) = 0$.

In this paper, solvable, with no modifiers, will mean geometrically solvable.

2.2. Uniform solvability

We need a somewhat stronger form of solvable than that given in Definition 1. The next definition is closer to that used in the linear case [Campbell (1987)]. The usual Euclidean norm on $\mathbb{R}^n$ is $\|\cdot\|$. For integers $m \geq 1$, let $\mathscr{C}^m$ be the space of m times continuously differentiable $\mathbb{R}^n$-valued functions on the finite interval $\mathscr{T}$. Let $\mathscr{C}^0$ be the space of continuous $\mathbb{R}^n$-valued functions on the finite interval $\mathscr{T}$. For $m \geq 0$ we give $\mathscr{C}^m$ the norm $\|g\|_m = \sum_{i=0}^m \|g^{(i)}\|_\infty$ where $\|h\|_\infty = \sup_{t\in\mathscr{T}} \|h(t)\|$. Let $\mathscr{B}_\epsilon = \{g \in \mathscr{C}^m : \|g\|_m < \epsilon\}$ and $\mathscr{B}_\epsilon(h) = \{g \in \mathscr{C}^m : \|h - g\|_m < \epsilon\}$. For notational convenience let $\|h\|_{-1} = \int_{\mathscr{T}} \|h(t)\|\, dt$.

Definition 2. Let $\mathscr{T}$ be an open subinterval of $\mathbb{R}$, Ω a connected open subset of $\mathbb{R}^{2n+1}$, and F a sufficiently differentiable function from Ω to $\mathbb{R}^n$. Then the DAE $F(t, y, y') = 0$ is *uniformly k-solvable* on $\mathscr{T}$ in Ω, $k \geq 0$, if there exists $\epsilon > 0$ such that

1. The DAE
$$F(t, y, y') = \delta(t) \tag{4}$$
is geometrically solvable for all $\|\delta\|_{k-1} < \epsilon$.
2. The solutions of (4) can be written as $\Phi(t, \lambda, \delta)$ where for each $\delta \in \mathscr{B}_\epsilon$, Φ satisfies *(1)-(4)* of Definition 1.
3. Φ, thought of as a nonlinear operator in δ, is continuous from $\mathscr{B}_\epsilon$ with the $\|\cdot\|_{k-1}$ norm to $\mathscr{C}_0$.

The dependence of Φ on δ in Definition 2 can include differentiations, integrations, and other operations.

Example 2. The linear differential equation $ty' + ty = 0$ is solvable but not k-uniformly solvable for any $k \geq 0$ on any interval $\mathscr{T}$ with $0 \in \mathscr{T}$.

In other applications one may might want to have restrictions on the types of perturbations. This is frequently the case in control theory when considering disturbances of certain inputs. Let Δ be a set of functions defined on $\mathscr{T}$. Usually there will be some structure to the set Δ.

Definition 3. Let $\mathscr{T}$ be an open subinterval of $\mathbb{R}$, Ω a connected open subset of $\mathbb{R}^{2n+1}$, and $F(t, y, y', \delta)$ a sufficiently differentiable function from Ω to $\mathbb{R}^n$ for each $\delta \in \Delta$. Then the DAE $F(t, y, y', 0) = 0$ is *uniformly k-solvable* on $\mathscr{T}$ in Ω, $k \geq 0$ *with respect to* Δ, if there exists $\epsilon > 0$ such that

1. The DAE

$$F(t, y, y', \delta) = 0 \tag{5}$$

is solvable for all $\|\delta\|_{k-1} < \epsilon$, $\delta \in \Delta$.
2. The solutions of (5) can be written $\Phi(t, \lambda, \delta)$ where for each $\delta \in \mathscr{B}_\epsilon$, Φ satisfies *(1)-(4)* of Definition 1.
3. Φ is continuous with respect to δ as a function on $\Delta \cap \mathscr{B}_\epsilon$ from the $\|\cdot\|_{k-1}$ norm to the $\|\cdot\|_0$ norm.

In some applications one might want Δ to be composed of different types of objects. For example, Δ could consist of parameters, perturbations, errors, and forcing functions. We will not address these important issues here. [Campbell (1988)] would be relevant to such a discussion. If a DAE is uniformly k-solvable [with respect to Δ] for some $k \geq 0$, we shall call it *uniformly solvable [with respect to Δ]*.

A more interesting example is the next one.

Example 3. Consider the DAE

$$x' = 1 \tag{6a}$$
$$y' = v \tag{6b}$$
$$0 = xv - y \tag{6c}$$

Let $\mathscr{M}$ be the zero set of $g(x, y, v) = xv - y$. This is a two dimensional manifold and

$$x' = 1 \tag{7a}$$
$$y' = v \tag{7b}$$
$$v' = 0 \tag{7c}$$

is a vector field on $\mathscr{M}$. The solutions of the DAE (6) are

$$x = t + c_1 \tag{8a}$$
$$y = (t + c_1)c_2 \tag{8b}$$
$$v = c_2 \tag{8c}$$

The DAE (6) is clearly geometrically solvable but not uniformly solvable if Ω is chosen so that $x = 0$ somewhere in Ω. To better understand what is happening let us reverse our point of view. We have the manifold $g(x, y, v) = xv - y = 0$ and a flow on the manifold given by (8). This flow is associated with a vector field given by (7). The DAE (6) can be viewed as an equation giving the manifold, (6c) and a projection of the vector field given by (6a) and (6b). The "singularity" at $x = 0$ is occurring because the vector field at that point is not transverse to the projection.

This type of singularity is not particularly pathological. Let $g(u) = 0$ be any two dimensional manifold in $\mathbb{R}^3$ which is not flat anywhere. Let

$$u' = \Psi(u) \tag{9}$$

define a nonzero vector field on this manifold. Let A be a 3×3 rank two matrix. Then on any open set Ω the set of matrices A for which the DAE

$$Aq' = A\Psi(q) \tag{10a}$$

$$0 = g(x, y, v) \tag{10b}$$

has a singularity like Example 3 has positive measure as a subset of the 3×3 rank two matrices. That is, this type of singularity is "generic" for certain specific classes of DAEs.

The type of singularity that appears in Example 3 poses several difficulties for numerical methods since near the singularity the accompanying linear algebra problems would experience extreme ill conditioning. For particular classes of problems it is possible to do a theoretical analysis of the vector field near the singularity. An analysis motivated by DAEs which arise in power systems is discussed in [Venkatasubramanian, Schättler and Zaborszky (1992)]. Computational approaches are developed in [Rabier and Rheinboldt (1994a,b)]. See also [Crouch, Ighneiwa and Lamnabhi-Lagarrigue (1991), Petzold, Ren and May (1993)].

3. "The" index

The approach of [März (1992)] for some index two and index three systems is based on linearizations. However, all of the existing approaches for general solvable nonlinear DAEs require some type of consideration of what we call the derivative array equations. The various indices are a measure of how much differentiation is required to determine y'. It is not our intention to survey all known definitions of the index. See for example [Griepentrog, Hanke and März (1992)]. In this section we shall show that there are essentially two types of indices, standard and uniform, which are philosophically distinct. Most of the indices studied to date have been standard indices. We shall give examples that show that for some systems the standard indices may vary greatly on the same problem with respect to the same solution. This runs counter to the usual perception in the literature that all the various types of indices that are currently considered (except for the local index) are essentially equivalent or differ by at most one when solutions and equations are sufficiently smooth. Section 3.1 discusses standard indices. In Sect. 3.2 we introduce the uniform indices and establish some of their properties.

3.1. Standard indices

In general, the solution y of (1) is known to depend on derivatives of F. If (1) is differentiated k times with respect to t, we get the $(k + 1)n$ *derivative array equations* [Campbell (1993)]

$$(11)\quad \begin{bmatrix} F(t,y,y') \\ F_t(t,y,y') + F_y(t,y,y')y' + F_{y'}(t,y,y')y'' \\ \vdots \\ \dfrac{d^k}{dt^k}F(t,y,y') \end{bmatrix} = F_k(t,y,y',w) = 0$$

where

$$(12)\qquad w = [y^{(2)}, \ldots, y^{(k+1)}]$$

Frequently in particular applications, different equations in $F = 0$ are differentiated a different number of times. This has no affect on the results presented here.

Consideration of (11) means that we must also consider open sets Γ in (t,y,v,w) space. We define the projection map π by

$$\pi(\Gamma) = \{(t,y,v) : \ (t,y,v,w) \in \Gamma \text{ for some } w\}$$

Suppose that we have an open set Ω that we are interested in. We shall frequently need to construct an open set Γ which, among its other properties, satisfies $\pi(\Gamma) = \Omega$. We will often denote this by writing $\Gamma = \Omega^{\mathrm{e}}$. Similarly, if we were to start with Γ we might denote $\pi(\Gamma)$ by Ω and Γ by Ω^{e}.

We define the *graph* of a solution y on $\mathscr{I}$ to be given by $\{(t,y(t),y'(t)) : t \in \mathscr{I}\}$. The *extended graph* is $\{(t,y(t),y'(t),\ldots,y^{(k+1)}(t)) : t \in \mathscr{I}\}$. Given a neighborhood Ω of part of the graph of a solution, we shall need to be careful in choosing sets Ω^{e} so that they include part of the extended graph.

A value (t,y) is said to be *consistent* for (11) *on* Ω *for* Ω^{e} if there exists (v,w) such that $F_k(t,y,v,w) = 0$ and $(t,y,v) \in \Omega$, $(t,y,v,w) \in \Omega^{\mathrm{e}}$. We shall often omit the "on omega" part of our terminology.

Given a consistent value (t,y), (11) viewed as an algebraic equation, will generally have a set of solutions for (y',w).

Definition 4. Suppose that $F(t,y,y') = 0$ is a solvable DAE on Ω. If v is uniquely determined by (t,y) and $F_k(t,y,v,w) = 0$ for all consistent values and ν_{d} is the least such integer k that this holds for, we call ν_{d} the *differentiation index* of the DAE.

Note that the definition of the differentiation index also assumes the specification of an open set Ω^{e}. The DAE is *higher index* if $\nu_{\mathrm{d}} \geq 2$. Higher index DAEs are sometimes also called algebraically incomplete. The differentiation index can also be defined with respect to a solution, or any other invariant manifold, but we will not do so.

If the DAE is geometrically solvable and $k \geq \nu_{\mathrm{d}}$, then v gives the vector field defined by the solutions on the manifold formed by the solutions.

There are several variants of the differentiation index. They are discussed in [Griepentrog, Hanke and März (1992)]. In some versions, the differentiations are accompanied by various coordinate changes in order to reveal the constraints

at that level. These approaches include the transversality ideas of März and colleagues [Griepentrog, Hanke and März (1992), Griepentrog and März (1986)] and the global index of Gear and Petzold [Gear and Petzold (1984)]. Alternatively, the terminology of differential geometry is used, see Griepentrog (1992), Reich (1991) or Rheinboldt and Rabier (1991). For linear time varying systems, theories exist for systems with impulsive and nonunique solutions [Kunkel and Mehrmann (1994), Rabier and Rheinboldt (1994d)]. The approaches differ somewhat in the amount of smoothness required and in whether some intermediate quantities must have constant rank Jacobians. Some only require the constant rank assumptions to hold on a manifold. In some approaches the w variables are eliminated as they occur. In others, the original $F = 0$ equation is augmented. Others replace part of the original equations with new equations using projections. However, when these indices are defined, they are equivalent when sufficient smoothness is present and constant rank assumptions give well defined constraints.

A different type of index is defined in [Hairer, Lubich and Roche (1989)]. Let $\|f\|_{\mathrm{p}}^t$ be $\|f\|_{\mathrm{p}}$ on the interval $[0,\ t]$ for $p \geq -1$.

Definition 5. The DAE $F(t, y, y') = 0$ has *perturbation index* ν_{p} along a solution y on the interval $\mathcal{I} = [0,\ T]$ if ν_{p} is the smallest integer such that if

$$F(t, \widehat{y}, \widehat{y}') = \delta(t) \tag{13}$$

for sufficiently smooth δ, then there is an estimate

$$\|\widehat{y}(t) - y(t)\| \leq C \left(\|\widehat{y}(0) - y(0)\| + \|\delta\|_{\nu_{\mathrm{p}}-1}^t \right) \tag{14}$$

for sufficiently small δ in the $\|\cdot\|_{\nu_{\mathrm{p}}-1}$ norm. C is a constant that depends on F and the length of the interval and the solution y.

We will always have $\nu_{\mathrm{p}} \geq 1$ since we assume that $F_{y'}$ is always singular.

For a pure Hessenberg system of index one, two, or three, the perturbation index is the same as the differentiation index [Hairer, Lubich and Roche (1989)]. However for index one systems of the form $B(y)y' = a(y)$, ν_{p} can be one higher than ν_{d}.

Example 4. Consider the following example from [Hairer, Lubich and Roche (1989)],

$$y_1' - y_3 y_2' + y_2 y_3' = 0,\ y_1(0) = 0 \tag{15}$$

$$y_2 = 0 \tag{16}$$

$$y_3 = 0 \tag{17}$$

Clearly, this DAE has $\nu_{\mathrm{d}} = 1$. Letting $\delta = [0, \epsilon \sin(\omega t), \epsilon \cos(\omega t)]$ we have $y_1' = \epsilon^2 \omega$ which involves ω. Thus the estimate depends on δ' so that $\nu_{\mathrm{p}} = 2$.

In [Gear (1990)] it is asserted that ν_d and ν_p differ by at most one. In Example 10, we shall show that this is not true.

The key thing to note about both of these definitions of the index, and others that we have not given, is that the index defined does not involve any continuity with respect to a class of perturbations. This is true even for ν_p since it is only talking about continuity as $\delta \to 0$ in some norm and not continuity in δ for $\delta \neq 0$. There is also little consideration given to actually carrying out the computations involved except in special cases [März (1992)].

We will now define a second class of indices which we call uniform indices. We shall see that unlike the previous indices these are more closely related to each other and are more easily computable for general unstructured DAEs.

3.2. Uniform indices

To simplify the following discussion we shall assume for the remainder of this paper that the DAE in question is uniformly k-solvable for some k. We shall also take functions $\delta(t)$ to be infinitely differentiable but with the norms $\|\cdot\|_i$. This choice is reasonable but not the only one that could be made. A different choice could alter some of our observations as noted after Example 6.

Definition 6. The *uniform index* ν_U of the DAE $F(t, y, y') = 0$ is the smallest integer k such that the DAE is uniformly k-solvable.

The perturbation index ν_p and the uniform index are different in several respects. First, ν_U can exist when ν_p does not exist as the next example shows.

Example 5. The algebraic equation

$$y^3 = 0 \tag{18}$$

has $\nu_U = 1$, but ν_p is not defined since $\delta^{1/3}$ is continuous but not Lipschitz continuous at $\delta = 0$.

Secondly, we shall see in Example 7 that ν_p can vary in the neighborhood of a solution whereas ν_U is locally the same whether it is defined with respect to a solution or on a neighborhood of solution.

We wish to focus more on the situation where some differentiation is possible as opposed to examples like (18). We return then to the derivative array equations (11) and define the Jacobians

$$\bar{J}_k = [\ G_{y'} \quad G_w\ ], \quad J_k = [\ G_{y'} \quad G_w \quad G_y\ ], \qquad \text{where } G = F_k$$

We shall say a system of algebraic equations

$$A \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b$$

is *1-full* with respect to x_1 if x_1 is uniquely determined for any consistent b [Campbell (1985)].

In practice, the equation $F_k = 0$ is a nonlinear system of equations with a singular Jacobian. It will sometimes be necessary to solve it numerically. Also, errors in y make it desirable to solve the derivative array equations in the least squares sense. The following assumptions on F_k permit a robust numerical least squares solution of the derivative array equations (11). The assumpitons (A1)–(A4) are to hold on a neighborhood Ω^e of the extended graph.

(A1) Sufficient smoothness of F_k.
(A2) Consistency of $G = F_k = 0$ as an algebraic equation.
(A3) $\overline{J}_k = [\ G_{y'} \quad G_w\ ]$ is 1-full with respect to v and has constant rank independent of (t, y, y', w).
(A4) $J_k = [\ G_{y'} \quad G_w \quad G_y\ ]$ has full row rank independent of (t, y, y', w).

The assumptions (A1)–(A4) provide the basis for a general numerical and analytical approach [Campbell (1989), Campbell (1993), Campbell and Griepentrog (1995), Campbell and Moore (1994a), Campbell, Moore and Zhong (1994)] for moderate sized nonlinear higher index DAEs. A constraint preserving version is introduced in [Campbell and Moore (1994b)]. A detailed discussion of the importance of each assumption can be found in the cited papers. The numerical solution of (11) is discussed there also. The conditions (A2), (A3), (A4) are numerically verifiable and can be used to establish solvability [Campbell and Griepentrog (1995)]. See also [Kunkel and Mehrmann (1993)].

Definition 7. The *uniform differentiation index* $\nu_{\rm UD}$ of the DAE (1) on Ω^e is the smallest integer k, if it exists, such that (A2) holds and (A1), (A3), (A4) hold on the open set Ω^e.

Although it is expressed in terms of (1) the uniform differentiation index actually implies solvability with respect to a parameter.

Proposition 1. *Suppose that* $F(t, y, y') = 0$ *is a solvable DAE and satisfies* (A1)–(A4) *on* Ω^e *with* $k = \nu_{\rm UD}$. *Then*

1. *For every* ϵ *sufficiently small there is an open set* $\Omega_\epsilon \subset \Omega$ *such that* $F(t, y, y') = \delta$ *is solvable on* Ω_ϵ *for* $\delta \in \mathscr{B}_\epsilon$.
2. *For any* $p \in \Omega$ *there is a function* $a(t)$ *and a neighborhood* $\widetilde{\Omega}$ *of* p *such that* $F(t, y, y') = a(t)$ *is solvable on* $\widetilde{\Omega}$, *satisfies* (A1)–(A4), *and* p *lies on the graph of a solution.*

Proof. We first prove statement one. Note that (A1), (A3), (A4) are independent of $a(t), \delta$ and will hold on subneighborhoods of Ω. Let $\widehat{\delta}_j$ be the column vector made up of $\delta, \ldots, \delta^{(j-1)}$. (A4) insures that consistency of $G = 0$ implies the consistency of $G = \widehat{\delta}_{k+1}$ as an algebraic equation. Also if the DAE $F = 0$ is solvable and (A1)–(A4) hold at k, then they hold for $k + 1$. If (A1)–(A4) hold for k and $k + 1$, then the DAE $F = \delta$ is solvable [Campbell and Griepentrog (1995)] and the first statement follows. Finally, given a point in Ω we just take a function with that point on its graph and use it to generate the $a(t)$. □

Since the uniform differentiation index can be reasonably computed [Campbell and Griepentrog (1995)] even for general unstructured problems, it is important to know how it relates to the others.

Proposition 2. *Suppose that ν_p and ν_{UD} are well defined. Then*

$$\nu_p \leq \nu_{UD} + 1 \tag{19}$$

Proof. Suppose that we consider the DAE $F(t, y, y') = \delta(t)$. This system has the same uniform differentiation index for all small δ. Using the results from [Campbell (1993), Campbell and Griepentrog (1995)] we get that the solutions of the DAE $F = \delta$ satisfy a smooth differential equation

$$y' = h(t, y, \delta, \ldots, \delta^{(k)}) \tag{20}$$

with $k = \nu_{UD}$. Thus the solutions will satisfy the estimate (14) if $\nu_p = \nu_{UD} + 1$. □

For Hessenberg systems [Brenan, Campbell and Petzold (1989)] of size r we can get that

$$y_1' = h_1(t, y, \delta) \tag{21a}$$

$$y_2 = h_2(t, y, \delta, \ldots, \delta^{(r-1)}) \tag{21b}$$

where $y = [y_1, y_2]$ so that $\nu_d = \nu_{UD} = \nu_p$.

However, in general ν_p can be quite different from ν_{UD}.

Example 6. Consider the DAE

$$\sin(y')y + x = 0 \tag{22a}$$

$$\sin(z')z + y = 0 \tag{22b}$$

$$z = 0 \tag{22c}$$

The DAE (22) has only one solution and $\nu_p = 1 = \nu_d$. However, $\nu_U = \nu_{UD} = 3$.

The system (22) has $\nu_p = 1$ because we assume that perturbations are smooth and only consider how they occur in the error estimate. If one changes the definition of ν_p to be the larger of the value from the error estimate and the smoothness of δ required, then ν_p would be 2 for Example 6. It is easy to modify Example 6 to give $\nu_p = \nu_d = 1$ and ν_{UD} any positive integer.

Example 7. If we replace (22c) in Example 6 with

$$z - w = 0 \tag{23a}$$

$$w' - w = 0 \tag{23b}$$

We have $\nu_p = 1$ along the solution $x = y = z = w = 0$ but $\nu_p = 3$ along any solution with $w \neq 0$. ν_{UD} is still 3.

Example 8. Note that $y^{1/3} = 0$ has $\nu_p = 1$ since $y = \delta^3$. However, ν_{UD} is undefined since $y^{-2/3}$ is not defined in a neighborhood of zero.

The next example is a variant of Example 4. It and the more general version in Example 10 will be used to illustrate several phenomena.

Example 9. Consider the DAE

$$y_2' y_2 + y_1 = 0 \tag{24a}$$
$$y_2 = 0 \tag{24b}$$

Differentiating (24a) and (24b) twice we get

$$y_2'' y_2 + (y_2')^2 + y_1' = 0 \tag{24c}$$
$$y_2' = 0 \tag{24d}$$
$$y_2''' y_2 + 3y_2' y_2'' + y_1'' = 0 \tag{24e}$$
$$y_2'' = 0 \tag{24f}$$

By considering the first four equations we have $\nu_d = 1$. To get ν_p note that if we consider

$$y_2' y_2 + y_1 = \delta_1 \tag{25}$$
$$y_2 = \delta_2 \tag{26}$$

then $y_2 = \delta_2,\ y_1 = \delta_1 - \delta_2 \delta_2'$. Thus $\nu_p = \nu_{UD} = 2$.

As mentioned earlier, several of the definitions of the standard indices and solvability in the literature are based on choosing a sequence of manifolds. It is instructive, to go through the calculation for (24) in Griepentrog type notation [Griepentrog (1992)]. Let S_k be the consistent y values and M_k the consistent y' values for some consistent y. Then

$$S_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ M_0 = \begin{bmatrix} * \\ * \end{bmatrix},\ \overline{J}_0 = \begin{bmatrix} 0 & y_2 \\ 0 & 0 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ M_1 = \begin{bmatrix} 0 \\ 0 \\ * \\ * \end{bmatrix},\ \overline{J}_1 = \left[\begin{array}{cc|cc} 0 & y_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 2y_2' & 0 & y_2 \\ 0 & 1 & 0 & 0 \end{array}\right]$$

$$S_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ M_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ * \\ * \end{bmatrix},\ \overline{J}_2 = \left[\begin{array}{cc|cc|cc} 0 & y_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 1 & 2y_2' & 0 & y_2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 3y_2'' & 1 & 3y_2' & 0 & y_2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

where $*$ is an arbitrary entry.

There are several points to be made about this example. First $\overline{J}_0$ does not have constant rank on S_0 so that the Griepentrog approach would not apply.

However, $\overline{J}_1$ does have constant rank and the v component of M_1 is unique so that the differentiation index can be taken to be 1. Note, however, that $\overline{J}_1$ is not 1-full if $y_2 \neq 0$. Thus (A3) does not hold for $k = 1$ on a neighborhood of $y = 0$. On the other hand, $\overline{J}_2$ is 1-full and constant rank in a neighborhood of $y = 0$ so that $\nu_{\mathrm{UD}} = 2$.

The key point of this example is not so much that the two types of indices differ by one but rather that the manifold approach of Griepentrog and others gives ν_{d} and not the uniform differentiation index.

We can expand on Example 9 as follows.

Example 10. Consider the DAE

$$F(y, y') = y_m N y' + y = 0 \tag{27}$$

where N is a $m \times m$ upper triangular nilpotent Jordan block, and $y = [y_1, \ldots, y_m]^T$. Then $\nu_{\mathrm{d}} = 1$ whereas $\nu_{\mathrm{p}} = \nu_{\mathrm{UD}} = m$.

This example utilized the fact that the index of nilpotency is not continuous, or even upper or lower continuous. Combining our previous examples we can construct the following example.

Example 11. Given any three integers $1 \leq k_1 \leq k_2 \leq k_3$ there exists a DAE with $\nu_{\mathrm{d}} = k_1$, $\nu_{\mathrm{p}} = k_2$, and $\nu_{\mathrm{UD}} = k_3$. This can be constructed by taking a DAE, $F^1(y, y') = 0$, like Example 6 with $\nu_{\mathrm{p}} = \nu_{\mathrm{d}} = 1,\ \nu_{\mathrm{UD}} = k_3$, a DAE, $F^2(z, z') = 0$, like Example 10 with $\nu_{\mathrm{d}} = 1,\ \nu_{\mathrm{p}} = \nu_{\mathrm{UD}} = k_2$, and a DAE, $F^3(w, w') = 0$, with $\nu_{\mathrm{d}} = \nu_{\mathrm{p}} = \nu_{\mathrm{UD}} = k_1$. Then the composite system $F^1 = 0,\ F^2 = 0,\ F^3 = 0$, has the desired properties.

As noted earlier the uniform differentiation index is the basis of a general numerical approach [Campbell and Moore (1994a), Campbell and Moore (1994b)]. While other indices can be computed for various classes of DAEs, the uniform differentiation index is the only index we are aware of which can be reasonably computed for general *unstructured* higher index DAEs. Since ν_{UD} is not equivalent to the standard indices ν_{p} and ν_{d}, it is important to determine how ν_{UD} relates to them. Unfortunately, we need to define two additional types of index in order to do this.

Definition 8. Let $\mu = \nu_{\mathrm{U}} - 1$. The *maximum perturbation index* of $F(t, y, y') = 0$ along a solution $\widetilde{y}$ is

$$\nu_{\mathrm{MP}} = \min_{\epsilon > 0} \max_{\|\delta\|_\mu < \epsilon} \{\nu_{\mathrm{p}} \text{ of } F(t, \widehat{y}, \widehat{y}') = \delta \text{ along solutions such that } \|\widehat{y} - \widetilde{y}\| < \epsilon\} \tag{28}$$

Note that while ν_{MP} is defined along a solution that, in fact, it is defined in a neighborhood of the solution since ν_{p} is a nonnegative integer valued variable and as noted before if $\|y-\widehat{y}\|_{r+n+1}$ is small enough, then $\|\delta\| = \|F(t,y,y') - F(t,\widehat{y},\widehat{y}')\|_r$ will also be small. If we take the maximum of ν_{MP} over solutions $\widetilde{y}$ with graphs in an open set Ω, we will talk of the value of ν_{MP} on Ω.

Definition 9. The *maximum differentiation index* of $F(t,y,y')=0$ on a set Ω is

$$\nu_{\mathrm{MD}} = \min_{\epsilon>0}\max_{\delta}\{\nu_{\mathrm{d}} \text{ of } F(t,y,y')=\delta, \text{ for } \|\delta\|_{\mu}<\epsilon \text{ and } F=\delta \text{ solvable on } \Omega\} \tag{29}$$

Clearly we have that on an open set

$$\nu_{\mathrm{MP}} \leq \nu_{\mathrm{UD}}+1 \tag{30}$$

$$\nu_{\mathrm{MD}} \leq \nu_{\mathrm{UD}} \tag{31}$$

$$\nu_{U} \leq \nu_{\mathrm{MP}} \tag{32}$$

The remainder of this section will establish the relationships between ν_{MD}, ν_{UD}, and ν_{MP}.

Theorem 1. *Suppose that the DAE (1) is solvable and that $\nu_{\mathrm{MD}}, \nu_{\mathrm{UD}}$ are both well defined on Ω and Ω^{e}. Then*

$$\nu_{\mathrm{MD}} = \nu_{\mathrm{UD}} \tag{33}$$

on a neighborhood $\overline{\Omega}$ which includes all the solutions whose (extended) graphs lie in (Ω^{e}) Ω.

Proof. From (31) it suffices to show that ν_{MD} cannot be less than ν_{UD}. Suppose that

$$\nu_{\mathrm{MD}} < \nu_{\mathrm{UD}} \tag{34}$$

and let $k_0 = \nu_{\mathrm{UD}} - 1$. By assumption, we have that F_{k_0} uniquely determines v as a function of y, δ for consistent y. This relationship is smooth in both y, δ since ν_{UD} is well defined.

Let $r = n(k_0+1) - \operatorname{rank}(\overline{J}_{k_0+1})$. Thus $\operatorname{rank}(\overline{J}_{k_0+1}) = nk_0 + n - r$. Then $n-r$ is the dimension of the solution manifold for a given δ at a given time t [Campbell (1993)]. By construction, $\operatorname{rank}(\overline{J}_{m+1}) \leq \operatorname{rank}(\overline{J}_m) + n$ since $\overline{J}_{m+1}$ has n more rows. Also $\operatorname{rank}(J_{m+1}) = \operatorname{rank}(J_m) + n$.

We shall show that, in fact, (A1)–(A4) hold for k_0 in four steps.

Step 1. The first n columns of $\overline{J}_{k_0}$ are full rank on a dense open subset $\widetilde{\Omega}_1$ of Ω^{e} whose closure includes all the solution graphs which lie in Ω^{e}. If this were not true, we could get a subset $\widehat{\Omega}$ of Ω^{e} on which $(F_{k_0})_v$ has constant rank and is not full column rank and which includes a solution. But then $F_{k_0}=0$ does not uniquely determine v on this set by the implicit function theorem.

Step 2. We can get an open dense subset $\widetilde{\Omega}_2$ of $\widetilde{\Omega}_1$ such that on each connected component of $\widetilde{\Omega}_2$ we have that $\overline{J}_{k_0}$ is 1-full and constant rank. To do this we first choose the open set so that $\overline{J}_{k_0}$ has constant rank on components. Now on each component we have $\overline{J}_{k_0}$ is constant rank. Thus if it were not 1-full at one point it would not be 1-full on a neighborhood of that point. But the first n columns are linearly independent by Step 1. These last two statements contradict the fact that v is determined independently of w.

Step 3. We now show $\overline{J}_{k_0}$ has constant rank on Ω^{e}. Let $\widetilde{\Omega}_{\mathrm{c}}$ be a component of $\widetilde{\Omega}_2$ containing a solution. Then (A1)–(A4) hold for $k = k_0$ on $\widetilde{\Omega}_{\mathrm{c}}$. By assumption they also hold for $k = k_0 + 1$ on $\widetilde{\Omega}_{\mathrm{c}}$. Thus the DAE is solvable on $\widetilde{\Omega}_{\mathrm{c}}$ and $r = nk_0 - \mathrm{rank}(\overline{J}_{k_0})$ on $\widetilde{\Omega}_{\mathrm{c}}$ [Campbell and Griepentrog (1995)]. This gives $\mathrm{rank}(\overline{J}_{k_0}) = nk_0 - r$ on all of $\widetilde{\Omega}_2$. But $\mathrm{rank}(\overline{J}_{k_0}) \geq \mathrm{rank}(\overline{J}_{k_0+1}) - n = nk_0 - r$. Since rank can only drop at a rank discontinuity, we have that $\mathrm{rank}(\overline{J}_{k_0}) = nk_0 - r$ on $\overline{\Omega}^{\mathrm{e}}$.

Step 4. Now that $\overline{J}_{k_0}$ has constant rank on $\overline{\Omega}^{\mathrm{e}}$ the argument of Step 2 shows that $\overline{J}_{k_0}$ must be 1-full on all of $\overline{\Omega}$. Thus $\nu_{\mathrm{UD}} \leq k_0$ which is a contradiction. □.

The need to reduce the neighborhood in the preceding theorem is illustrated by the next example.

Example 12. Let $\gamma(z)$ be an infinitely differentiable function which is zero for $|z| \leq \epsilon$, some $\epsilon > 0$, and positive otherwise. Now consider the DAE

$$\gamma(y_2)y_2' + y_1 = 0 \tag{35a}$$

$$y_2 = 0 \tag{35b}$$

Note that for "small" neighborhoods that we have $\nu_{\mathrm{UD}} = 1$ while $\nu_{\mathrm{UD}} = 2$ for neighborhoods which allow $|y_2| > \epsilon$.

Theorem 2. *Suppose that the DAE (1) is solvable and that $\nu_{\mathrm{MP}}, \nu_{\mathrm{UD}}$ are both well defined on Ω. Then*

$$\nu_{\mathrm{UD}} \leq \nu_{\mathrm{MP}} \leq \nu_{\mathrm{UD}} + 1 \tag{36}$$

Proof. From (30) and (33) it suffices to show that $\nu_{\mathrm{UD}} > \nu_{\mathrm{MP}}$ gives a contradiction. Let $k = \nu_{\mathrm{UD}}$. The proof of Theorem 2 is done in two parts. The first part will show that the vector field for

$$F(y', y, t) = \delta(t) \tag{37}$$

will depend on k derivatives of δ in a nonsingular way. The second part shows that this will violate the estimates given by the assumption $\nu_{\mathrm{MP}} < k$. To simplify our notation in this proof, let $b_i = b^{(i)}$ for $b = \delta, d, u$. Also, again let $\widehat{\delta}_j$ be the column vector made up of $\delta, \ldots, \delta^{(j-1)}$. Define $\widehat{G} = F_{k-1}$. We can write the derivative array equations $F_k = 0$ as

$$\widehat{G}(v_1, v_2, w_1, w_2, w_3, y_1, y_2, y_3) = \widehat{\delta}_k \tag{38a}$$

$$H(v_1, v_2, w_1, w_2, w_3, y_1, y_2, y_3) = \delta_k \tag{38b}$$

where the partition of the variables is taken so that the Jacobians

$$\begin{bmatrix} \widehat{G}_{v_1} & \widehat{G}_{w_1} & \widehat{G}_{y_1} \end{bmatrix}, \quad \begin{bmatrix} \widehat{G}_{v_1} & \widehat{G}_{v_2} & \widehat{G}_{w_1} & \widehat{G}_{w_2} & \widehat{G}_{y_1} & \widehat{G}_{y_2} \\ H_{v_1} & H_{v_2} & H_{w_1} & H_{w_2} & H_{y_1} & H_{y_2} \end{bmatrix} \tag{39}$$

are nonsingular. That is, $\{v_1, w_1, y_1\}$ are a subset of the original variables $\{v, w, y\}$ such that $[\widehat{G}_{v_1}\ \widehat{G}_{w_1}\ \widehat{G}_{y_1}]$ is a nonsingular submatrix of $[\widehat{G}_v\ \widehat{G}_w\ \widehat{G}_y]$. This can be done because of (A4). Similarly there is a strictly larger subset $\{v_1, v_2, w_1, w_2, y_1, y_2\}$ of the original variables so that the matrix on the right in (39) is a nonsingular submatrix of $[G_v\ G_w\ G_y]$ with $G = F_k$. The remaining variables in w, y are denoted w_3, y_3. We allow for some of the variables to be absent from (38). There might not, for example, be any y_1. From [Campbell (1993)] we know that there will always be a w_3 since there are always some unsolved for higher derivatives if a higher index DAE has all equations differentiated the same number of times. The t variable has been omitted for convenience. Given any point and time of interest, under our assumptions this choice of v_i, w_i, y_i can always be done in a sufficiently small neighborhood. The neighborhood is also local in time. We choose our neighborhood to be one where ν_{UD} is still k.

We shall now show that the vector field must depend on δ_k.

From the implicit function theorem applied to (38a) we get that

$$v_1 = \overline{\phi}(v_2, w_2, w_3, y_2, y_3, \widehat{\delta}_k) \tag{40a}$$

$$w_1 = \overline{\psi}(v_2, w_2, w_3, y_2, y_3, \widehat{\delta}_k) \tag{40b}$$

$$y_1 = \overline{\theta}(v_2, w_2, w_3, y_2, y_3, \widehat{\delta}_k) \tag{40c}$$

while from (38) we get

$$v_1 = \phi_1(y_3, \widehat{\delta}_k) \tag{41a}$$

$$v_2 = \phi_2(y_3, \widehat{\delta}_k) \tag{41b}$$

$$w_1 = \psi_1(w_3, y_3, \widehat{\delta}_k, \delta_k) \tag{41c}$$

$$w_2 = \psi_2(w_3, y_3, \widehat{\delta}_k, \delta_k) \tag{41d}$$

$$y_1 = \theta_1(y_3, \widehat{\delta}_{k-1}) \tag{41e}$$

$$y_2 = \theta_2(y_3, \widehat{\delta}_{k-1}) \tag{41f}$$

The simpler form of (41a), (41b) occurs because v is uniquely determined by y, δ since $k = \nu_{\mathrm{MD}}$ by Theorem 1 and we are assuming for contradiction purposes that v does not depend on δ_k, that is $\phi_{\delta_k} = 0$. The simpler form of (41e), (41f) which does not involve δ_{k-1}, δ_k arises from [Campbell (1993)].

Now combine (40) with (38b) to get

(42) $$H(\overline{\phi}, v_2, \overline{\psi}, w_2, w_3, \overline{\theta}, y_2, y_3) = \delta_k$$

which is a function of $v_2, w_2, w_3, y_2, y_3, \widehat{\delta}_k$. Then (40) and (42) give a system locally equivalent to (38). Thus (42) has a nonsingular Jacobian with respect to v_2, w_2, y_2. Then by the implicit function theorem the solution of (40) and (42) is (40) and

(43a) $$v_2 = \widetilde{\phi}_2(w_3, y_3, \widehat{\delta}_k, \delta_k)$$

(43b) $$w_2 = \widetilde{\psi}_2(w_3, y_3, \widehat{\delta}_k, \delta_k)$$

(43c) $$y_2 = \widetilde{\theta}_2(w_3, y_3, \widehat{\delta}_k, \delta_k)$$

Since (42) is onto, the Jacobian of the right hand side of (43) is nonsingular with respect to δ_k. Since there are always more w components determined by (38) than by (38a), there is always a w_2 component. Hence

(44) $\widetilde{\psi}_{2\delta_k}$ is full row rank

However, the uniqueness part of the implicit function theorem says that $\widetilde{\phi}_2 = \phi_2$, $\widetilde{\psi}_2 = \psi_2$, $\widetilde{\theta}_2 = \theta_2$. But then (41b), (41d), (41f) have a nonsingular Jacobian with respect to δ_k since (43) does. Since the combined dimension of (41b), (41d), (41f) is the same as that of δ_k this is impossible unless there are no v_2, y_2 since ϕ_2 in (41b) and θ_2 in (41f) do not depend on δ_k. Then from (40) we would get

(45a) $$v = \overline{\phi}(w_2, w_3, y_3, \widehat{\delta}_k)$$

(45b) $$w_1 = \overline{\psi}(w_2, w_3, y_3, \widehat{\delta}_k)$$

(45c) $$y_1 = \overline{\theta}(w_2, w_3, y_3, \widehat{\delta}_k)$$

Again combining (45) with the $H = \delta_k$ equation which is now $H(\overline{\phi}, \overline{\psi}, w_2, w_3, \overline{\theta}, y_3) = \delta_k$ we get that

(46) $$w_2 = \widetilde{\psi}_2(w_3, y_3, \widehat{\delta}_k, \delta_k)$$

with

(47) $\widetilde{\psi}_{2\delta_k}$ nonsingular

But then (46), (45a) give

(48) $$v = \overline{\phi}(\widetilde{\psi}_2(w_3, y_3, \widehat{\delta}_k, \delta_k), w_3, y_3, \widehat{\delta}_k)$$

which by assumption does not depend on δ_k. Thus

$$\overline{\phi}_{w_2}\widetilde{\psi}_{2\delta_k} = 0$$

By (47) we get $\overline{\phi}_{w_2} = 0$ so that (48) implies that

(49) $$v = \overline{\phi}(w_3, y_3, \widehat{\delta}_k)$$

The dependence on w_3 is nontrivial since (38a) does not uniquely determine v. This follows since by Theorem 1 we can be on a neighborhood where $\nu_{\mathrm{MD}} = k$ also. However, there is no way to eliminate w_3 when we add (38b) since w_3 is still arbitrary and we have a contradiction. Thus, in fact

$$v_2 = \phi_2(y_3, \widehat{\delta}_k, \delta_k), \quad \phi_{2\delta_k} \neq 0 \tag{50}$$

This completes the first part of the proof of Theorem 2. Now assume that

$$\nu_{\mathrm{MP}} < k \tag{51}$$

We first prove the following Lemma.

Lemma 1. *Suppose that there is a $T^* > 0$ such that for all $0 < T \leq T^*$:*

1. *z is a solution of the differential equation $z'(t) = f(z(t), u(t), u_1(t), \ldots, u_r(t), t)$.*
2. *f is at least twice continuously differentiable in all components.*
3. *$f(0, 0, \ldots, t) = 0$ for $0 \leq t \leq T$.*
4. *There is a $\rho > 0$ such that*

$$\|z\| \leq K \left(\|z(0)\| + \sum_{i=0}^{r-2} \|u_i\| \right) \quad \text{for} \quad \left(\|z(0)\| + \sum_{i=0}^{r-2} \|u_i\| \right) < \rho$$

where $\|\cdot\|$ denotes the sup norm on the interval $[0, T]$ and K is independent of T.

Then $f_{u_r}(0, \ldots, 0) = 0$.

Proof of Lemma 1. Suppose that assumptions 1 through 4 hold. Let η be a constant vector. Define a perturbation d, to be used for δ in (4) by

$$d(t) = \epsilon^{(r+1)/2} Q\left(\frac{t}{\sqrt{\epsilon}}\right) \eta \tag{52}$$

where Q is a positive or negative sine or cosine chosen so that $Q^{(r)}(t) = \cos(t)$. Then

$$d_r(t) = \sqrt{\epsilon} \cos\left(\frac{t}{\sqrt{\epsilon}}\right) \eta \tag{53}$$

Also note that for (52) we have $\left(\sum_{i=0}^{r-2} \|d_i\|\right) = O(\epsilon^{3/2})$ and hence $\|z\| = O(\epsilon^{3/2})$ for small $\|z(0)\|$. Taking $u = d$ and expanding f in z and d_i we get

$$z' = f_z(0, t)z + \sum_{i=0}^{r} f_{u_i}(0, t)d_i + O_2(z, d, d_1, \ldots, d_r, t) \tag{54}$$

Here O_2 represents the higher order z and d_i terms. Integrating (54) gives

$$z(T) = z(0) + \int_0^T f_z(0,t)z \; dt \tag{55a}$$

$$+ \sum_{i=0}^{r-2} \int_0^T f_{u_i}(0,t)d_i \; dt \tag{55b}$$

$$+ \int_0^T f_{u_{r-1}}(0,t)d_{r-1} \; dt \tag{55c}$$

$$+ \int_0^T f_{u_r}(0,t)d_r \; dt \tag{55d}$$

$$+ \int_0^T O_2(z, d, d_1, \ldots, d_r, t) \; dt \tag{55e}$$

By taking $\|z(0)\|$ small we have that $\left(\|z(0)\| + \sum_{i=0}^{r-2} \|d_i\| \right)$ is $O(\epsilon^{3/2})$. Then all the terms in (55a) and (55b) including $z(T)$ are $O(\epsilon^{3/2})$. The following integration by parts

$$\int_0^T f_{u_{r-1}}(0,t)d_{r-1} \; dt$$

$$= f_{u_{r-1}}(0,T)d_{r-2}(T) - f_{u_{r-1}}(0,0)d_{r-2}(0) - \int_0^T f_{tu_{r-1}}(0,t)d_{r-2}(t) \; dt \tag{56}$$

shows that (55c) is also $O(\epsilon^{3/2})$ since d_{r-2} is. Since $\|d_i\| = O(\sqrt{\epsilon})$ for $i = 0, \ldots, r$ and O_2 is quadratic in the d_i we have that (55e) is $TO(\epsilon)$. Every term has been estimated except (55d). Substituting all these estimates into (55) gives

$$O(\epsilon^{3/2}) = \int_0^T f_{u_r}(0,t)d_r(t) \; dt + TO(\epsilon) \tag{57}$$

$$= f_{u_r}(0,t)d_{r-1}\Big|_{t=0}^{t=T} - \int_0^T f_{tu_r}(0,t)u_{r-1}(t) \; dt + TO(\epsilon) \tag{58}$$

Another integration by parts shows that the integral in (58) is also $O(\epsilon^{3/2})$. Thus we have that (57) implies that

$$O(\epsilon^{3/2}) = f_{u_r}(0,t)d_{r-1}(t)\Big|_{t=0}^{t=T} + TO(\epsilon) \tag{59}$$

We are going to use the mean value theorem on the remaining term in (59). Since this is a vector function we technically need to treat each entry separately. However, since we then let the interval go to zero the following calculation is a correct outline of the actual calculation.

$$\begin{aligned} O(\epsilon^{3/2}) &= f_{u_r}(0,t)d_{r-1}(t)\Big|_{t=0}^{t=T} + TO(\epsilon) \\ &= [f_{u_r}(0,t)d_{r-1}(t)]'(\mu)T + TO(\epsilon) \quad \text{some } 0 \le \mu \le T \\ &= f_{tu_r}(0,\mu)d_{r-1}(\mu)T + f_{u_r}(0,\mu)d_r(\mu)T + TO(\epsilon) \end{aligned} \tag{60}$$

Now let $T = \epsilon^{2/3}$ so that (60) becomes

$$O(\epsilon^{3/2}) = f_{tu_r}(0,\mu)O(\epsilon)\epsilon^{2/3} + f_{u_r}(0,\mu)\sqrt{\epsilon}\cos\left(\frac{\mu}{\sqrt{\epsilon}}\right)\eta\epsilon^{2/3} + \epsilon^{2/3}O(\epsilon)$$

or

$$O(\epsilon^{3/2}) = f_{u_r}(0,\mu)\cos\left(\frac{\mu}{\sqrt{\epsilon}}\right)\eta\epsilon^{7/6} + O(\epsilon^{5/3})$$

Dividing both sides by $\epsilon^{7/6}$ yields

$$O(\epsilon^{1/3}) = f_{u_r}(0,\mu)\cos\left(\frac{\mu}{\sqrt{\epsilon}}\right)\eta + O(\epsilon^{1/2})$$

Note that

$$0 \le \frac{\mu}{\sqrt{\epsilon}} \le \frac{T}{\sqrt{\epsilon}} = \frac{\epsilon^{2/3}}{\sqrt{\epsilon}} = \epsilon^{1/6}$$

Letting $\epsilon \to 0^+$ and noting that η is arbitrary completes the proof of Lemma 1. □

To complete the proof of Theorem 2 observe that if $\widehat{y}$ is any solution of the original DAE we may introduce the change of variables $y = \widehat{y} + \widetilde{y}$ to get a new DAE in $\widetilde{y}$ where the solution of interest is $\widetilde{y} = 0$. Similarly given $F(y', y, t) = \widehat{\delta}$ we can rewrite it as $F(y', y, t) - \widehat{\delta} = \widetilde{\delta}$ and consider perturbations as $\widetilde{\delta} \to 0$. Similarly moving the t origin is trivial. Thus Lemma 1 contradicts (50) and Theorem 2 is proven. □

4. Numerical analysis and control

In this section we shall very briefly point out how the ideas of this paper relate to some issues in numerical analysis and systems theory. Additional references can be found in the cited papers, especially [Campbell (1995)].

4.1. Systems and control theory

The concept of index relates to several questions in systems and control theory and is of active interest in the engineering literature [Bachman, Brüll, Mrziglod, and Pallaske (1990), Blajer (1992), Chung and Westerberg (1990), Dai (1989), Fliess, Lévine and Rouchon (1993a), Lefkopoulos and Stadtherr (1993)]. The typical starting point in many nonlinear control problems is a system

$$F(x', x, u, t) = 0 \tag{61a}$$

$$y = h(x, u, t) \tag{61b}$$

Usually one refers to x as the state, u as the control, and y as an output or measurement. The system (61a) is sometimes referred to as the process or plant.

Often (61a) is in the form $x' = f(x,t) + g(x,t)u$. Constraints due to the physical process are included in (61a). Desired constraints are in (61b).

The path following or prescribed output problem [Fliess, Lévine and Rouchon (1993a), Fliess, Lévine and Rouchon (1993b), Hirschorn and Davis (1988)] considers y to be a given desired function $y(t)$. The output nulling problem is the special case where $y = 0$. In the inversion problem one allows y to vary over some class of outputs and the goal is to find u in terms of y and possibly x. If y is known, (61) is a DAE in $\{x,u\}$. Typically in the control development one performs a sequence of differentiations, sometimes using Lie bracket notation, until the control u can be recovered. In this situation the number of differentiations for (61) is one less than the differentiation index of (61). The number of differentiations is variably called the relative degree or order. Usually not all equations are differentiated the same number of times.

For the output nulling problem, in its simplest form, the appropriate concept is the differentiation index. There are no perturbation terms (δ's). However, suppose that one wants to consider the presence of small perturbations. In this setting y is allowed to be nonzero, but the minimal requirement might be continuity in terms of y as $y \to 0$. This leads to considering a perturbation type of index with the perturbation being y.

In the inversion problem, y varies over a class of functions and the most appropriate concepts are the uniform types of indices.

Another important issue in control theory is the effect of disturbances. Disturbances may vary from noise terms to unmodeled dynamics. We consider the later case. The disturbance is often taken to be smooth and the solution of some unknown dynamical system. This leads to a system of the form

$$F(x', x, u, t, \delta_1) = 0 \tag{62a}$$

$$y = h(x, u, \delta_2, t) \tag{62b}$$

Here either perturbation or uniform indices are most appropriate depending on whether we are interested in continuity in terms of δ_i or as $\delta_i \to 0$. Our analysis shows that these two types of continuity can lead to fundamentally different results.

Recently there has been an active discussion in the control literature about the concept of weak relative degree. That is, the relative degree depends on the value of y. This is closely related to the examples we gave earlier where the differentiation index was different than the perturbation or the uniform indices. Such problems arise in flight control and other problems where, for example, the control variable is multiplied by a trigonometric expression which can be zero for certain configurations of the plant. Extra differentiations in the control design sometimes smooth out discontinuities in the observer arising from singularities Campbell and Terrell (1991), Campbell, Nichols and Terrell (1991). That is, by using ν_{UD} differentiations instead of ν_{d} differentiations one got an expression without singularities. Sometimes the singularities correspond to actual singularities on the solution manifold. These singularities usually lead to places where

ν_{UD} is not defined. Note [Crouch, Ighneiwa and Lamnabhi-Lagarrigue (1991), Hirschorn and Davis (1988), Petzold, Ren and May (1993)].

4.2. Numerical methods for DAEs

Traditionally, in engineering, the only index that was considered was the differentiation index or the relative degree. The first different type of index was the perturbation index. This early work concentrated on systems with various special structure. There has been some interest in computing the index and also in modifying it because of the importance its value has for the behavior of various numerical methods [Duff and Gear (1986), Gear (1988), Gear (1990), Griepentrog (1992), Mattsson and Söderlind (1993)]. In recent years there has been increasing work on increasingly complex composite systems.

The uniform differentiation index provides information on the amount of differentiation needed to insure continuity with respect to perturbations, and also constant ranks which are important numerically. Secondly it is, for moderate sized problems, a computable quantity and its computation is closely linked to establishing solvability [Campbell and Griepentrog (1995)].

IRK methods hold considerable promise for the numerical solution of a variety of DAE systems. There are major difficulties in extending them to more general classes of system but some very recent work suggests that some extension of IRK or even RK methods may be possible. It is to be expected that ν_{MP} will be important in discussing these developments.

5. Conclusion

Many definitions of indices have been given in the literature. We have shown that there are generally two types of indices. For indices defined without considering continuity with respect to a perturbation, we have shown by examples that contrary to popular belief the value of many of the different standard indices can vary widely. We have defined new maximal indices with respect to a class of perturbations. These maximal indices are more closely related.

References

1. Bachman, R., Brüll, L., Mrziglod, Th., Pallaske, U. (1990): On methods for reducing the index of differential algebraic equations, Computers Chem. Engng, **14** , 1271–1273
2. Bender, D.J., Laub, A.J., (1987): The linear quadratic optimal regulator for descriptor systems, IEEE Trans. Aut. Control **32**, 672–688
3. Blajer, W. (1992): Index of differential-algebraic equations governing the dynamics of constrained systems, Appl. Math. Modeling **16**, 70–77
4. Brenan, K.E., . Campbell, S.L., Petzold, L.R. (1989): Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations, Elsevier
5. Campbell, S.L. (1985): The numerical solution of higher index linear time varying singular systems of differential equations, SIAM J. Sci. Stat. Comp. **6**, 334–348

6. Campbell, S.L. (1987): A general form for solvable linear time varying singular systems of differential equations, SIAM J. Math. Anal. **18**, 1101–1115
7. Campbell, S.L. (1988): Bilinear nonlinear descriptor control systems. In Linear Algebra in Signals, Systems, and Control, edited by B. N. Datta et al, SIAM, 439–511
8. Campbell, S.L. (1989): A computational method for general higher index singular systems of differential equations, 1989 IMACS Transactions Scientific Computing **1.2** , 555–560
9. Campbell, S.L. (1993): Least squares completions for nonlinear differential algebraic equations, Numer. Math. **65**, 77–94
10. Campbell, S.L. (1995): High index differential algebraic equations, J. Mechanics of Structures and Machines **23**, 199–222
11. Campbell, S.L., Griepentrog, E. (1995): Solvability of general differential algebraic equations, SIAM J. Sci. Comp. **16**, 257–270
12. Campbell, S.L., Moore, E. (1994a): Progress on a general numerical method for nonlinear higher index DAEs II, Circuits Systems & Signal Processing **13**, 123–138
13. Campbell, S.L., Moore, E. (1994b): Constraint preserving integrators for general nonlinear higher index DAEs, Numerische Math. **69**, 383–399
14. Campbell, S.L., Moore, E., Zhong, Y. (1994): Utilization of automatic differentiation in control algorithms, IEEE Trans. Automatic Control **39**, 1047–1051
15. Campbell, S.L., Terrell, W.J. (1991): Observability of Linear Time Varying Descriptor Systems, SIAM J. Matrix Analysis **12**, 484–496
16. Campbell, S.L., Nichols, N., Terrell, W.J. (1991): Duality, observability, and controllability for linear time varying descriptor systems, Circuits Systems & Signal Processing **10**, 455–470
17. Chung, Y., Westerberg, W. (1990): A proposed numerical algorithm for solving nonlinear index problems, Ind. Eng. Chem. Res. **29**, 1234–1239
18. Crouch, P.E., Ighneiwa, I., Lamnabhi-Lagarrigue, F. (1991): On the singular tracking problem, Math. Control Signals Systems **4**, 341–362,
19. Dai, L. (1989): Singular Control Systems, Springer-Verlag, Berlin
20. Duff, I., Gear, C.W. (1986): Computing the structural index, SIAM. J. Alg. Disc. Methods **7**, 594–603
21. Fliess, F., Lévine, J., Rouchon, P. (1993a): Index of implicit time-varying linear differential equation: a noncommutative linear algebraic approach, Linear Algebra & Its Appl. **186**, 59–71
22. Fliess, F., Lévine, J., Rouchon, P. (1993b): Generalized state variable representation for a simplified crane description, Int. J. Control **58**, 227–283
23. Gear, C.W. (1988): Differential-algebraic equation index transformations, SIAM J. Sci. Stat. Comp., **9**, 39–47
24. Gear, C.W. (1990): Differential algebraic equations, indices, and integral algebraic equations, SIAM J. Numer. Anal. **27**. (1990), 1527–1534
25. Griepentrog, E. (1992): Index reduction methods for differential-algebraic equations, Seminarberichte Nr. 92-1, Humboldt-Universität zu Berlin, Fachbereich Mathematik, 14–29
26. Griepentrog, Hanke, E., März, R. (1992): Toward a better understanding of differential algebraic equations (Introductory survey), Seminarberichte Nr. 92-1, Humboldt-Universität zu Berlin, Fachbereich Mathematik, 1–13
27. Griepentrog, E. März, R. (1986): Differential-Algebraic Equations and Their Numerical Treatment, Teubner-Texte zur Mathematik, Band 88, Leipzig
28. Hairer, E., Lubich, C., Roche, M. (1989): The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods, Springer-Verlag, New York
29. Hirschorn, R.M. Davis, J.H. (1988): Global output tracking for nonlinear systems, SIAM J Control and Optimization **26**, 1321–1130
30. Kunkel, P., Mehrmann, V. (1993): A new class of discretization methods for the solution of linear differential algebraic equations with variable coefficients, preprint
31. Kunkel, P., Mehrmann, V. (1994): Canonical forms for linear differential algebraic equations with variable coefficients, J. Comp. Appl. Math. **69**, to appear
32. Lefkopoulos A., Stadtherr, M.A., (1993): Index analysis of unsteady-state chemical process systems-I. An algorithm for problem formulation, Computers Chem. Engng. **17**, 399–413
33. März, R. (1992): On quasilinear index 2 differential-algebraic equations, Seminarberichte 92-1, Humboldt-Universität zu Berlin, Fachbereich Math., 39-60

34. Mattsson, S., Söderlind, G. (1993): Index reduction in differential-algebraic equations using dummy derivatives, SIAM J. Sci. Stat. Comp., **14**, 677–692
35. Petzold, L.R., Ren, Y., May, T. (1993): Numerical solution of differential-algebraic equations with ill-conditioned constraints, preprint
36. Potra F.A., Rheinboldt, W.C. (1991): Differential-geometric techniques for solving differential algebraic equations. In Real-Time Integration Methods for Mechanical System Simulation, Eds. E. J. Haug and R. C. Deyo, Springer-Verlag Computer and Systems Sciences **69**, 155–191
37. Rabier, P.J., Rheinboldt, W.C. (1991): A general existence and uniqueness theorem for implicit differential algebraic equations, Diff. Int. Eqns. **4**, 563–582
38. Rabier, P.J., Rheinboldt, W.C. (1994a): On impasse points of quasilinear differential algebraic equations, J. Math. Anal. Appl. **181**, 429–454
39. Rabier, P.J., Rheinboldt, W.C. (1994b): On the computation of impasse points of quasilinear differential algebraic equations, Math. Comp. **62**, 133–154
40. Rabier, P.J., Rheinboldt, W.C. (1994c): A geometric treatment of implicit differential-algebraic equations, J. Diff. Eqns. **109**, 110–146
41. Rabier, P.J., Rheinboldt, W.C. (1994d): Classical and generalized solutions of time dependent linear differential algebraic equations, Linear Alg. Appl., to appear
42. Reich, S. (1990): On a geometric interpretation of differential-algebraic equations, Circuits Systems & Signal Processing **9**, 367–382
43. Reich, S. (1991): On an existence and uniqueness theory for nonlinear differential-algebraic equations, Circuits Systems & Signal Processing **10**, 343–359
44. Venkatasubramanian, V., Schättler, H., Zaborszky, J. (1992): A stability theory of large differential algebraic systems: a taxonomy, Systems Science and Mathematics Report SSM 9201, Washington University, St. Louis

This article was processed by the author using the LaTeX style file *pljour1m* from Springer-Verlag.