Toward a better understanding of differential algebraic equations (Introductory survey)

Eberhard Griepentrog,[1] *Michael Hanke,*[2] *and Roswitha Mär z*[2]

[1] Ingenieurhochschule Berlin, Fachbereich 2,
Marktstr. 9/13, O-1134 Berlin
[2] Humboldt-Universität zu Berlin, Fachbereich Mathematik,
PSF 1297, O-1086 Berlin

Differential algebraic equations (DAEs) are everywhere singular implicit ordinary differential equations (ODEs)

$$f(x'(t), x(t), t) = 0 \ , \tag{1}$$

where $f'_y(y, x, t)$ is singular for all values of its arguments. If $f'_y(y, x, t)$ were nonsingular, (1) could be solved for x', at least theoretically, and we would have an explicit ODE. Actually, **DAEs ARE ODEs** but those which cannot be solved with respect to the x'.

The notion "DAE" represents the fact that (1) consist of differential equations coupled with pure finite–dimensional ("algebraic") equations. Other catchwords for DAEs are e. g.: singular systems, descriptor systems, semistate equations, differential equations on manifolds.

DAEs arise in various fields of applications. The most popular ones are

- simulation of electrical circuits
- constrained dynamical systems, e.g. Euler–Lagrange equations of rigid bodies systems, and chemical reactions subject to balance invariants
- optimal control of lumped–parameter systems
- semi–discretization of partial differential equation systems, e. g. the Navier–Stokes system describing the flow of incompressible viscous fluids
- reduced equations in singularly perturbed systems.

For a fairly detailed survey of applications we refer to [**?**]. It should be mentioned that, in the last couple of years, DAEs have developed into a highly topical subject of applied mathematics. There is a rapidly increasing number of contributions devoted to DAEs in the mathemtical literature as well as in the field of mechanical engineering, chemical engineering, system theory, etc. Formally, models with singular implicit ODEs have already been known for a long time (cf. [**?**], [**?**]). It was C. W. Gear ([**?**]) who proposed to handle such ODEs numerically by backward differentiation formulas (BDF). In the sequel, powerful codes allowing to simulate large circuits sussessfully, provided one has a very special class of DAEs, have been developed. Obviously, for a long time DAEs were considered not to differ essentially from regular implicit ODEs in general. Only since about 1980

the mathematical community has been investigating DAEs more thoroughly, challenged by computation results that could not be brought into line with the above supposition, (cf. [**?**]). With their articles "DAE's revisited" ([**?**]) and "DAE's are not ODE's" ([**?**]), C. W. Gear, H. H. Hsu and L. R. Petzold have given rise to the discrussion on mathematical peculiarities of DAEs and on possibilities for their successful numerical treatment, a discussion that will surely be carried on still for a long time.

DAEs seem to be a typical example of the importance of keeping an eye upon the mathematical – inclluding the numerical – properties of the arising mathematical models already upon modelling them.

The fact that, in the meantime, also textbooks imparting general fundamentals of Numerical Analysis, as e. g. [**?**], treat DAE problems corresponds to the present importance of this field.

First attempts towards a better understanding of the peculiarities of ODEs referred to linear equations

$$A(t)x'(t) + B(t)x(t) = q(t) \tag{2}$$

where $A(t)$, $B(t)$ were continuous $m \times m$ matrix functions on an interval $I \subseteq R$. The notion of the *global index* by C. W. Gear and L. R. Petzold has provided an important essential possibility fo r classifying ODEs.

The DAE (2) is said to have the global index $\mu_K \in \mathbb{N}$ if there are regular matrix functions $E(\cdot) \in C$, $F(\cdot) \in C^1$ such that scaling (2) by $E(t)$ and transforming $x(t) = F(t)y(t)$ lead to the DAE

$$\begin{bmatrix} I & 0 \\ 0 & J \end{bmatrix} y'(t) + \begin{bmatrix} W(t) & 0 \\ 0 & I \end{bmatrix} y(t) = E(t)q(t) \ , \tag{3}$$

where J is a nilpotent Jordan block matrix $J^{\mu_K} = 0$, $J^{\mu_K - 1} \neq 0$. Eq. (3) is said to be the (Kronecker) canonical normal form of (2).
Clearly, this notion is closely related to canonical normal forms of regular matrix pencils (cf. [**?**]). Actually, (3) decouples into the system

$$\left.\begin{array}{ll} u'(t) + W(t)u(t) = r(t) \\ Jv'(t) + v(t) \quad\ = s(t) \end{array}\right\} \ . \tag{4}$$

Now it becomes obvious that only the first part of (4) is regular ODE. The second part leads immediately to

$$v(t) = \sum_{j=0}^{\mu_K - 1} (-1)^j (Js(t))^{(j)} \ . \tag{5}$$

Clearly, initial value problems for (4) may become solvable for *consistent initial values* u_0, $v_0 = \sum\limits_{j=0}^{\mu_K - 1} (-1)^j (Js(t_0))^{(j)}$ only, but for arbitrary u_0, v_0. This is an essential difference to regular ODEs and, in case of $\mu_K > 1$ this entails considerable numerical problems, which have not been solved by now. For $\mu_K = 1$ it holds that $J = 0$ and (5) simply provides $v(t) = s(t)$. In this case (4) resp. (2)

are similar in behaviour to a regular ODE, and it can be expected that many of the methods that proved their value for regular ODEs can be suitably modified for such index 1 DAEs. Much has already been done in this respect (cf. [**?**], [**?**] and the papers of Degenhardt and Lamour in this issue).

The treatment of all the so–called higher index DAEs with $\mu_K \geq 2$ is much more difficult. In contrast to index 1–equations, they cannot be solved for all continuous $q(\cdot)$. For the computation of solutions or only of consistent initial values, the differentiations given in (5) have to be filled. This causes numerical difficulties that become the greater, the greater μ_K is and the stronger the two parts of (4) are coupled with each other in the original DAE.

In [**?**], [**?**], [**?**] it is pointed out that higher index DAEs, in natural function space formulations, lead to essentially ill–posed problems in the sense of Tichonov. The resulting maps are not Fredholm. Their unbounded inverses cause numerical methods (discretizations) to become unstable. This is no surprise if we recall that the differentiation problem in the space of continuous functions is one of the simplest examples of

an ill–posed problem.

For constant coefficient linear higher index DAEs the mentioned instability is known to be a weak one only, that is, errors in the initial values or numerical round–off errors are amplified by factors $h^{\mu_K - 1}$ only, where h is the stepsize (e. g. [**?**], [**?**], [**?**]).

Are there still further classes DAEs which weak singularities arise for, and can they be tackled by a suitable application of numerical methods? Fortunately, there are such classes, among them even those which are relevant for important applications (e.g. [**?**], [**?**], [**?**],

[**?**], [**?**]). Surely, it will be possible to further extend these classes. As another possibility one can ensure that the direct discretizations do not refer to higher index DAEs, but to a regular ODE or an index 1 DAE. In order to do this reasonably, one has to know more about DAEs, of course.

The Kronecker canonical normal form (3) of (2) and the global index μ_K are highly informative – provided that they are known. Unfortunatley, as a rule $E(\cdot)$, $F(\cdot)$, and hence, (3) cannot be computed, not even numerically. So other, more easily accessible forms of linear DAEs have to be searched for. In the paper of B. Hansen the most important of these forms (standard canonical form, modified standard canonical form, Hessenberg form) are collected and compared.

A further possibility consists in determining the so–called *tractability index* $\mu_T \in \mathbb{N}$ by means of the following matrix chain, which can be derived directly from the coefficients $A(\cdot)$, $B(\cdot)$ of (2):

$$\begin{aligned} A_0 &:= A, \quad B_0 := B - AP_0' \\ A_{i+1} &:= A_i + B_i Q_i \\ B_{i+1} &:= B_i P_i - A_{i+1}(PP_1 \cdots P_{i+1})' PP_1 \cdots P_{i-1} \\ P_i &:= I - Q_i, \quad Q_i(t) \in L(\mathbb{R}^m) \text{ projects onto} \\ &\qquad \ker(A_i(t)), i = 0, \ldots, \mu_T - 1 \ . \end{aligned} \tag{6}$$

Here, the projections Q_j are chosen such that $Q_j Q_i = 0$ holds for $j > i$.

The DAE (2) is called *tractable with index* μ_T if the matrix $A_{\mu_T}(t)$ remains nonsingular, but the previous matrices $A_j(t)$, $j < \mu_T$, are all singular and have constant ranks.

It should be mentioned that constructing the matrix chain (6) we aim at a decoupling of (2) by utilizing the given projections and related subspaces. In 1987 it was only a hypothesis of R. März that this might apply to $\mu_T > 3$, too (cf. [**?**]). Now, B. Hansen has succeeded in proving this ([**?**]). The decoupling of (2) via (6) yields the solution decomposition

$$\begin{aligned} x &= P_0 x + Q_0 x \\ &= P_0 P_1 \cdots P_{\mu_T - 1} x + P_0 \cdots P_{\mu_T - 2} Q_{\mu_T - 1} x + \cdots + P_0 Q_1 x + Q_0 x \end{aligned}$$

as well as equations determining each of these components. Thereby, the component $P_0 P_1 \cdots P_{\mu_T - 1} x$ satisfies a regular ODE, $P_0 \cdots P_{\mu_T - 2} Q_{\mu_T - 1} x$ is the "algebraic" one. Further, to obtain $P_0 \cdots P_{\mu_T - 3} Q_{\mu_T - 2} x$ we have to perform one differentiation, and so on. Finally, the "worst" component is $Q_0 x$, which includes a $(\mu_T - 1)$th derivative.

Note that this decoupling itself is not to represent a numerical method, by no means. It has to be understood as a means for the corresponding investigatio n of the DAE itself as well as of its discretizations. Moreover, it may be useful

for determining consistent initial values (see the related paper of Hansen in this issue). In the other paper of Hansen (see also [**?**]), an important new result is reported about: All DAEs having a global index μ_K possess the tractability index

$$\mu_T = \mu_K \ . \tag{7}$$

The difference between the two classes of problems consists in slightly weaker requirements to smoothness for tractable DAEs.

Using the first projector P_0 from (6), we may rewrite (2) as

$$A(t)(P_0(t)x(t))' + (B(t) - A(t)P_0'(t))x(t) = q(t) \tag{8}$$

because of $A(t)P_0(t) \equiv A(t)$. This makes clear, that, naturally, solutions belong to the function space $\{x \in C : P_0 x \in C^1\}$, but they do not belong to C^1 in general (cf. [**?**]). The same fact becomes obvious from [**?**], [**?**]. The matrix chain concept has been applied, with reasonable success, to several nonlinear DAEs (1) via linearizations ([**?**], [**?**], [**?**] and the paper of März in this issue), however, some important problems in this respect have not been solved yet. In how far can a linearization be successful at all for such ill–posed problems?

The notion of an index most frequently used today is the notion of a *differentiation index* μ_D. Following [**?**], we consider the system of equations

$$\left.\begin{aligned} f(x', x, t) &= 0 \\ \tfrac{d}{dt} f(x', x, t) &= \tfrac{\partial}{\partial x'} f(x', x, t) x^{(2)} + \ldots = 0 \\ &\vdots \\ \tfrac{d^\mu}{dt^\mu} f(x', x, t) &= \tfrac{\partial}{\partial x'} f(x', x, t) x^{(\mu+1)} + \ldots = 0 \end{aligned}\right\} \tag{9}$$

as a system in the separate dependent variables $x', x^{(2)}, \ldots, x^{(\mu+1)}$ with x and t as independent variables.

The DAE (1) has the *differentiation index* μ_D, if μ_D is the smallest μ for which (9) can be solved for $x' = \mathcal{S}(x,t)$.

It should be mentioned that S. L. Campbell even earlier used the same idea (e. g. [**?**], [**?**]) for linear DAEs. E. Griepentrog has defined the same notion more precisely (see this issue): Define the compound functions (derivative arrays)

$$F_\mu(\bar{y}_\mu, x, t) := \begin{bmatrix} f(y_1, x, t) \\ \frac{\partial}{\partial x'} f(y_1, x, t) y_2 + \ldots \\ \cdots \\ \frac{\partial^\mu}{\partial x'^\mu} f(y_1, x, t) y_{\mu+1} + \ldots \end{bmatrix},$$

$\bar{y}_\mu := (y_1^T, \ldots, y_{\mu+1}^T)^T$, according to (9). Let the partial Jacobians $H_\mu(\bar{y}_\mu, x, t)$ of $F_\mu(\bar{y}_\mu, x, t)$ with respect to $\bar{y}_\mu$ have constant rank. Define S_μ to be the manifold of all pairs (x,t) for which $F_\mu(\bar{y}_\mu, x, t) = 0$ is solvable. S_μ is called the *constraint manifold of order* μ. Further, introduce the manifold

$$M_\mu(x,t) := \{\bar{y}_\mu \in \mathbb{R}^{(\mu+1)m} : F_\mu(\bar{y}_\mu, x, t) = 0\}$$

for all $(x,t) \in S_\mu$.

The mapping $f(y,x,t)$ is said to be an *index–μ–mapping* if S_μ is non–empty and $M_\mu(x,t)$ is a singleton for all $(x,t) \in S_\mu$, and if μ is the smallest integer with these properties.
The DAE (1) is said to have the index μ_D if $F(y,x,t)$ is an index–μ_D–mapping. The main point (see the paper of Griepentrog) of this new version of the differentiation index is the geometrical background. The DAE (1) with index μ_D represents a unique vector field $v(x,t)$ defined on S_{μ_D} satisfying $f(v(x,t),x,t) = 0$. The solutions of the initial value problems

$$x' = v(x,t), \quad (x(t_0), t_0) \in S_{\mu_D} \tag{10}$$

proceed in S_{μ_D} and solve (1).

Thus, there has been established a close relation to the concept of W. Rheinboldt and S. Reich (e. g. [**?**], the paper of Reich in this issue), which understands a DAE as an (implicitly defined) vectorfield on an (implicitly given) manifold. Now, a regular DAE of geometrical index μ_G (degree μ_G) defined in this way is nothing but a DAE with the differentiation index

$$\mu_D = \mu_G \tag{11}$$

as shown by Griepentrog.

Comparing the global index μ_K (cf. (3), the tractability index μ_T (cf. (6)) and μ_D for linear systems (2) we have now (cf. the papers of Griepentrog and Hansen)

$$\mu_D = \mu_G = \mu_K = \mu_T \tag{12}$$

supposed the coefficients $A(t)$, $B(t)$ are smooth enough so that μ_D is defined.
It should be emphasized that the ranks of certain matrices are assumed to be

constant for μ_D as well as for μ_T (for μ_K this results automatically). If this does not apply, there arise new singularities which may lead, among other things, to bifurcations or impasse points (see the paper of März in this issue and [**?**]).

It should also be mentioned that μ_T is defined for nonlinear DAEs having the differentiation index $\mu_D \leq 3$ via linearization ([**?**], [**?**]), and it holds that

$$\mu_D = \mu_T \ . \tag{13}$$

On account of (11)–(13) we will only refer to index μ in the following. One of the most important intentions of the present monograph consists in regarding the indicated interrelation of the analytical, algebraic and geometrical background of the DAEs and to make it available for a constructive and numerical treatment.

First monographs on DAEs were written by S. L. Campbell ([**?**], [**?**]), and, almost unknown to the scientific world outside the U.S.S.R, by Ju. E. Boyarincev ([**?**]). Their books are devoted to linear problems almost exclusively.

A farily detailed discussion of linear and nonlinear problems of index 1, their solvability, Lyapunov stability, multistep and Runge–Kutta methods for initial value problems, finite difference methods and shooting methods for boundary value problems is contained in the monograph of E. Griepentrog and R. März ([**?**]).

K. E. Brenan, S. L. Campbell and L. R. Petzold ([**?**]), in their book, concentrated upon initial value problems for practically important nonlinear problems with index $\mu \leq 3$ and their numerical integration up to problems of implementation and software. The higher index DAEs with $\mu = 2$ and 3 that are investigated are of the so–called Hessenberg form as a rule. The methods discussed are multistep methods, in particular the BDF and Runge–Kutta methods. The latest and hitherto last book on DAEs is due to E. Hairer, Ch. Lubich and M. Roche ([**?**]) and deals with Runge–Kutta methods for ODEs with index $\mu \leq 3$ in Hessenberg form.

The Seminar Notes presented here provide a completion to the books mentioned above – with respect to analytical and geometrical investigations, the class of problems treated actually, as well as with respect to numerical methods. We have already mentioned those results (11) and (12) concerning the notion of an index that we consider to be important. Completing the two monographs above, the present one is mainly focused on boundary value problems, and, as a novelty in this context, on initial value problems for differential algebraic inclusions.

The contributions to this Seminar Notes are grouped around the 3 complexes:

1. analytical and geometric background of DAEs
2. approximating DAEs, numerical methods
3. differential–algebraic inclusions.

Above we have already discussed several aspects from the contributions of E. Griepentrog and B. Hansen to the complex 1. On the basis of the index concept mentioned above (cf. (11)) E. Griepentrog further described and compares various index reduction methods. S. Reich extends his geometrical understanding of regular DAEs which we developed in earlier papers for autonomous DAEs to the

case of nonautonomous DAEs (1). He, too, presents a reduction method which is taken up and described in a simple way by E. Griepentrog. In contrast to those reduction methods where differential equations are added, this method replaces "superflous" differential equations by equations. Equations reduced in this way have the advantage of correctly reflecting the asymptotic behaviour of the original DAE. However, their realization is probale to be much more difficult. We want to add that index reduction by generating algebraic equations is also reported about in [**?**], [**?**]. Earlier versions of this approach are to be found in [**?**].

In her contribution to the complex 1 R. März makes a first attempt to gain also local existence statements, continuations and, finally, criteria for the stability of equilibria by means of the projector– and subspace technique connected with the matrix chain (6). First, this attempt is successful for the index 2 DAE

$$Ax'(t) + g(x(t)) = 0$$

and for a disturbed form, respectively. Here it is essential that all criteria and conditions are formulated in terms of $A, g'(x_0), x_0$ fixed, hence they can be utilized in practice.

Let's now turn to the second complex. It is mainly concerned with some practical proposals for the direct numerical treatment of DAEs. Above we have tried to elucidate that the numerical treatment of higher index DAEs is problematical in general because of the occurring instabilities in the discretiyations caused by the ill–posedness of initial as well as boundary value problems. If possible, higher index DAEs should be avoided upon modelling.

For special higher index DAEs, e. g. for Euler–Lagrange equations describing constraint mechanical motions, special intergation methods overcoming these instabilities by an ingenuous utilization of the special structure of this DAE inclusively its geometry (e. g. [**?**] [**?**], [**?**]) have been developed. In general, however, one has only the possibility to trace back a higher index problem to an index problem or a regular ODE, or to approximate it by such a regular ODE. The most evident method in this respect are index reductions, which are closely related with the index notions (cf. see the paper of Griepentrog). However, one must heavily pay for the gain: Computational complexity, destroying the problem structure, drift–off, change of the stability of the system.

The second way is to *approximate* resp. *regularize* higher index problems by ones with lower index. In this field, there have only been a few approaches and results so far, unfortunately (e. g. [**?**], [**?**], [**?**]). Of course, all of them involve some difficulties and it has not been clarified yet whether they are sound. In our opinion, new possibilities of this kind should be sought for more intensively since the difficulties seem to be controllable.

M. Hanke devotes his contribution to such a regularization. In order to arrive at index 1 DAEs, appropriate small perturbations are introduced. Originally, the motivation for doing so was purely mathematical, but it turned out that this approach is closely related to certain pertubation methods in electric engineering (e. g. [**?**], [**?**]) and in mechanics (e. g. [**?**], [**?**]). Taking into account that higher index DAEs are ill–posed in naturally given topologies, one can successfully analyze

the approximate problems in the spirit of regularization methods. However, actually, we meet singularly perturbed regular ODEs, a fact calling for establishing asymptotic expansions. This is done for the linear DAE (2) and, what is more important, for some nonlinear semi–explicit systems. By a numerical example it is illustrated that competitive extrapolation methods should work well.

Note also that applying general regularization methods, e. g. to the ill–posed higher index DAEs seems to be a very heavy gun (cf. [**?**]).

As shown above (cf. (10)) one of the problems with DAEs is the computation of consistent initial values $(x_0, t_0) \in S_{\mu_D}$. This problem has not been solved yet since the manifold S_{μ_D} is implicitly given and accessible only with difficulties in practice. In [**?**] it is discussed how one can fall back on the definition of μ_D (cf. (9)) for this purpose.

B. Hansen sketches two new methods providing consistent initial values for a class of nonlinear index 2 DAEs. The first one uses the matrix chain and projector technique (cf. (6)). On this background, B. Hansen proposes a special transformation to decompose the DAE into a regular explicit ODE and two nonlinear equations. This method requires symbolic calculations. It seems not to be effective. The second method uses special index reduction propsed in [**?**] and [**?**], as well as the fact that consistent initial values for index 1 DAEs may be computed more or less easily.

Although index 1 DAEs appear as simplest DAEs above, they also have their peculiarities, which have to be taken into account especially in the numerical treatment. Fundamental solution marices become singular, and in the consequence, the traditional shooting matrices, too. Standard symmetric finite difference schemes applied to index 1 DAE become unstable etc. (e. g. [**?**]). In any case, an uncritical generalization of methods aproved for regular ODEs should be avoided. The behaviour of DAEs should be regarded carefully. A. Degenhardt proposes a collocation method for solving boundary value problems in transferable DAEs (that is index 1 DAEs) (1), where the partial Jacobian $f'_y(y, x, t)$ is assumed to have a constant nullspace N.

Due to the solution concept proposed in [**?**] (cf. also (8)) allowing different smoothness for the solution components in N and out of N a collocation method is created which approximates the nullspace and the non–nullspace components by piecewise polynomials of different degrees. For regular ODEs, this method reduces to the well–known piecewise polynomial collocation (e. g. [**?**]). By this, some results of [**?**] are generalized. The resulting equivalent Runge–Kutta methods are symmetric. A. Degenhardt uses different basis representations leading to different implementation strategies. Feasibility, convergence and even superconvergence are proved.

The paper of R. Lamour also deals with boundary value problems in index 1 DAEs. Generalizing results of [**?**], the well–posedness of problems with general boundary conditions, including multipoint and integral conditions, is stated. To handle autonomous oscillations, it is carefully investigated how to turn to an equivalent well–posed problem in a similar way as one used to do for regular ODEs. Special interest is devoted to oscillations in constraint mechanical systems. Since this DAE has originally index 3, an index reduction method is used.

It is pointed out that index reduction has to be coupled with a careful selection of the additional initial and end conditions, respectively. Besides t he periodicity condition and phase normalization, properly selected conditions like energy conservation are appropriate to ensure equivalence of the index reduction process. Finally, let us turn to complex 3. A great variety of dynamical systems with discontinuities can be modelled by DAEs, where discontinuities with respect to the state variable do arise. Among them, we have control systems with discontinuous feedback (e. g. robotics, aircraft control), mechanical systems with dry friction, electric circuits with switch elements etc. One way to model such systems is to replace the discontinuous characteristics of a system component by a multivalued one, which leads to a differential–algebraic inclusion (DAI) like

$$A(t)x'(t) + B(t)x(t) \in g(x(t), t) \ . \tag{14}$$

D. Niepage's contribution starts with the DAE

$$A(t)x'(t) + B(t)x(t) = h(x(t), t) \ , \tag{15}$$

where the left hand part is index 1 and h is discontinuous. By the use of Filippov's regularization ([**?**]), (15) is transformed into a DAI (14). Via the usual splitting technique (cf. [**?**]), solvability is proved. Moreover, on the basis of a discretization theoretical approach for DAIs ([**?**], [**?**]), multistep methods and Runge–Kutta–methods are shown to converge. The contribution of W. Wendt is devoted to a system of differential and algebraic inclusions of monotone type, which is obtained when using set–valued characteristics for diodes in electrical networks consisting of resistors, inductors, diodes (switches) and sources. The unique solvability of initial value problems is proved by the well–known Rothe method. Moreover, for a class of appropriate discrete approximations, stability and convergence orders are verified. By this, results of Elliott ([**?**]) for ordinary differential inclusions are generalized, and stronger results as in [**?**] are derived. Further, a nice method for computing the solution step by step is proposed and illustrated by an example.

In our opinion, especially the contribution of W. Wendt shows that even modelling with inclusions may be of practical importance and use, which is still doubted quite often. Of course, the problem of really efficient numerical methods still remains open, here too. We hope that we will be wiser in this respect as well as concerning the above problems in the future.