1. Introduction

A large class of dynamic problems in engineering lead to systems of differential and algebraic equations (DAEs). The differential equations arise from the conservation balances for mass, momentum, and energy. Mathematical models of chemical equilibrium, additional system constraints such as equations of state, and constitutive equations contribute algebraic equations to the dynamic problem. The solutions of DAE problems involve variable trajectories given by DAEs among variables and their derivatives. DAE systems arise naturally in optimal control or prescribed path problems.

DAEs occur in practice in the initial value (IVP) as well as in the boundary value (BVP) problems. For the numerical solution of BVPs, global methods such as orthogonal collocation,\textsuperscript{4} finite elements,\textsuperscript{2} or spectral methods\textsuperscript{3} convert differential into algebraic equations defined over the bounded problem domain. The fully discretized set of algebraic equations can be solved simultaneously by nonlinear solution algorithms. DAE discretization in BVP is discussed elsewhere.\textsuperscript{4}

Partial discretization of dynamic distributed systems with the method of lines (MOL) and transient lumped models leads to IVPs. Direct solution techniques for DAE IVPs by multistep integration were pioneered by Gear.\textsuperscript{5} However, convergence and accuracy of the numerical solution depend on the problem-specific mathematical relations among unknowns and their derivatives. A crucial property known as the index of a DAE determines the difficulty of convergence of numerical procedures. Existing solvers often fail in problems with a high index of 2 or more. Moreover, algebraic equations also prescribe quantitative relations between the initial values of the DAE problem. If the initial variable values or their derivatives violate these equations, existing numerical solution schemes break down.\textsuperscript{6} In high-index systems, there are also “hidden” constraints associated with derivatives of algebraic equations that consistent initial values have to satisfy.\textsuperscript{7} The number of hidden constraints correlates with the index. Although there is progress in directly solving high-index systems,\textsuperscript{8}--\textsuperscript{10} numerical convergence and consistency of the initial values require a priori knowledge of the DAE’s index before a numerical treatment should be attempted. In chemical engineering applications, multistaged unit operations such as catalytic reactors or separation networks, dynamics simulation of entire plants, or reacting flow problems discretized with MOL lead to very large DAEs. Hence, there is a need for developing algorithms for automatic a priori analysis of DAE properties applicable to industrial-size problems.

Outline. Earlier work for the DAE analysis adopted structural analysis for determining index and consistent initial values. Unfortunately, purely structural approaches may, in fact, over- or underestimate the actual index of a system. Our new proposal expands the resolution of automatic index analysis by considering both numerical and symbolic information. This hybrid approach renders a better estimate of the DAE’s actual index. We will demonstrate the new methodology’s correctness for the important class of linear DAEs and offer qualitative improvements for the general case. The paper is organized as follows: Section 2 discusses DAE properties and provides a concise index definition. Section 3 reviews the theory and background for analytical and structural index analysis. Section 4 introduces a new symbolic numerical algorithm with superior properties over existing approaches. Section 5 demonstrates chemical engineering applications of the methodology.

2. DAE Properties: The Differential Index of a DAE

Early direct numerical solution methods were based on kth-order backward differentiation methods.\textsuperscript{5} Inten-
sive research and further development has led to powerful numerical algorithms such as DASSL and DASPG.\textsuperscript{11,12} LIMEX,\textsuperscript{13,14} PSIDE,\textsuperscript{15} or DAEPACK.\textsuperscript{16} Some numerical solvers are limited to special problem formats; all perform reliably only when the index is low. The most important DAE categories suitable for state-of-the-art integrator codes are discussed next. More details can be found elsewhere.\textsuperscript{12} The most general form is the fully implicit DAE as given in eq 1. A direct solution for the fully implicit form is admissible with DASSL and DASPG.

**Implicit DAE form:**

$$F(y', y, t) = 0$$

(1)

An important special case is the explicit DAE of eq 2 where \(\mathbf{M}\) is a matrix of derivative expressions, which may also include time-varying terms. This is the problem format required by LIMEX,\textsuperscript{14} an extrapolation integrator, and RADAU5\textsuperscript{17} based on a three-stage Radau collocation method.

**Explicit DAE form:**

$$\mathbf{M}y' = f(y, t)$$

(2)

If all variable expressions and their derivatives occur linearly, the system is termed linear constant coefficient DAE as prescribed by eq 3. \(\mathbf{A}\) and \(\mathbf{B}\) are square matrices of real or complex numbers.

**Linear constant coefficient DAE:**

$$\mathbf{A}y' + \mathbf{B}y = f(t)$$

(3)

The existence of a unique solution to a DAE system can be guaranteed rigorously only for the constant coefficient type:

- Existence theorem of a unique solution—DAE solvability: The linear constant coefficient DAE system of eq 3 is solvable if and only if the matrix pencil \(\mathbf{A} + \mathbf{B}\) is regular; i.e., \(\det(i\mathbf{A} + \mathbf{B})\) is not identically vanishing as a function of \(i\).

For fully implicit and even generally explicit DAEs, there are no general solvability criteria, although the general DAE in eq 1 can be considered solvable, at least locally, if the pencil \(x(t, \mathbf{A} + \mathbf{B})\) is regular. However, there are counterexamples to this statement in the literature.\textsuperscript{18} Theoretical existence of a unique solution is no guarantee for numerical solvability on a digital computer. For most existing integrators, the DAE index needs to be lower than or equal to 2. What exactly is the index of a DAE? Here we give a definition formulated by Ascher and Petzold\textsuperscript{19} known as the differential index of a DAE:

**Definition—differential index of a DAE:** For the generally implicit DAE of eq 1, the index along a solution trajectory \(y(t)\) is the minimum number of differentiations of the system that would be required to solve for \(y'(t)\) uniquely in terms for \(y\) and \(t\); in other words, the differentiation count needed to convert eq 1 into an ordinary differential equation (ODE).

The definition implies that the index depends not only on the DAE structure but also on the solution trajectory in the state space.\textsuperscript{19} Moreover, the index is equal to the number of hidden system constraints to which initial values must obey for successful initialization of the integration. Clearly, the problem of the DAE index and consistent initialization are intertwined. For small numbers of equations, index analysis can be done analytically as suggested in example A. For large systems occurring in chemical engineering applications, automatic computer procedures are needed.

**Example A.** The equations for the evolution of the liquid composition in an open evaporation operation are shown in eqs 4 and 5. The differential equations (4) describe the change of the liquid composition of the mixture; its trajectories \(x(t)\) sketch a residue curve in the composition space. The algebraic equation (5) relates the mixture’s bubble point, \(T(t)\), to the species’ properties such as the pure-component vapor pressure, \(p_0(T)\). The index of the system is 1 because a single derivation of eq 5 renders the derivative of the temperature, \(T(t)\). Clearly, this system cannot be directly integrated with an ODE solver. It is also obvious that the initial pot temperature, \(T(t_0)\), and compositions, \(x(t_0)\), must be chosen consistently for the problem to be physically meaningful. Accurate computation of the residue curves requires simultaneous solution of eqs 4 and 5. This example shows that DAE system formulations should not be avoided.

$$x'_i = (K_i(T) - 1)x_i \quad \forall i = 1, ..., n$$

(4)

$$\sum_{i=1}^{n} K_i(T) x_i = 1 \quad \text{with } K_i = p_0^i(T)/P$$

(5)

### 3. Principles of Index Analysis

Rigorous analytical index analysis was proposed by Gear.\textsuperscript{20,21} According to its definition, index analysis involves the search for derivatives of algebraic variables by repeated differentiations of algebraic system equations. Symbolic equation manipulations for chemical engineering models (e.g., physical property calculations, equations of states, models for effective reaction rates) are not reliable enough even when using specialized computer software (MAPLE, Symbolic toolbox in MATLAB). To overcome the obstacles with analytical treatment of general nonlinear expressions occurring in the chemical engineering practice, structural DAE analysis procedures were developed.

**Structural Index Analysis.** In structural analysis, the DAE system is replaced with an incidence matrix of variables and derivatives. If a variable \(v_i\) is contained in an equation \(e_i\), the corresponding entry in the structure matrix, \(e_{ij}\), is set to true; otherwise, it is set to false. A fully ranked derivative submatrix signifies expressions for all variable derivatives \(v'_j\) available, hence, the structural index of the DAE is zero. If its structural rank is smaller than the number of equations, i.e., some derivatives are missing, structural differentiations are performed. The rank of the augmented systems composed of the original and additional equations obtained by extra differentiations is examined again. The process of additional derivations is repeated until the derivative submatrix has full rank. This is equivalent to ascertaining that expressions for all variable derivatives are available in terms of known quantities. The structural index can be defined as follows:

**Definition—structural index:** The structural index of the DAE is the minimum number of structural differentiation steps necessary to obtain a derivative submatrix of structural rank equal to the number of equations.
The main assumption of the structural index analysis rests upon the equivalence between the structural index and the differential index. Unfortunately, this conjecture is wrong, as will be shown in this chapter. Before we can comprehend the shortcomings of structural approaches, it is necessary to introduce basic concepts by means of two classical methods: (i) graph theoretical method\textsuperscript{22} and (ii) structural analysis using Boolean algebra.\textsuperscript{18}

**Structural Analysis via Graph Theoretical Techniques.** Pantelides\textsuperscript{22} presented a graph theoretical method targeted for finding consistent initial values as well as its structural index. The variables are classified into two groups: dynamic and algebraic variables. A variable is classified as dynamic if its derivatives occur in the DAE equations and algebraic in their absence. The number of structural differentiations for obtaining derivatives for all algebraic variables stored in a bipartite incidence graph is equal to the structural index of the DAE system. A detailed discussion of the method and its implementation can be found in the literature.\textsuperscript{23}

**Structural Analysis with Boolean Algebra.** The efficiency of the Pantelides approach is offset by the abstract logic of the procedure. Unger and co-workers\textsuperscript{18} developed a more transparent structural analysis method, leading to exactly the same result. The DAE system with \( n \) equations containing a algebraic and \( d \) dynamic variables, with \( a + d = n \), is converted into its structural representation. The variable set includes all variables and their derivatives occurring in the DAE system cardinality of \( 2n \). For each equation \( e_i \), a structural representation is a \( 2n \)-dimensional vector of zero (0) and nonzero (+) elements. The structure of the DAE system is a matrix whose binary entries \( e_i \) record whether the variable \( v_j \) is a member of the equation \( e_i \) or not. This matrix is also known as the pattern matrix.

Missing derivative expressions of the algebraic variables are sought by means of structural differentiation and elimination steps. The equivalence transformations deploy an operator algebra defined over the Boolean set: zeros (0) and nonzeros (+). The procedure terminates when the rank of the derivative submatrix is equal to the number of equations. The dynamics of a condenser in eqs 6–9 used by several authors as a test case elucidates the mechanism.\textsuperscript{18,22}

**Example B.**

\[
N = F - L \tag{6}
\]

\[
NC_pT = FC_p(T_{IN} - T) + LAH + US(T_C - T) \tag{7}
\]

\[
0 = p - Ae^{-[B(T + C)]} \tag{8}
\]

\[
0 = pV - NRT \tag{9}
\]

with the variables \( N \) = molar liquid holdup, \( T \) = condenser temperature, \( p \) = condenser pressure, \( L \) = liquid flow out of the condenser, and \( F \) = feed rate and constant parameters \( A, B, C = \) in constants the Antoine vapor-pressure equation, \( \Delta H, U, S, C_p = \) condensation enthalpy, heat-transfer coefficient and area, and the specific heat capacity of the liquid, and \( T_C, T_{IN} = \) cooling water temperature and the temperature of the feed.

In the DAE’s pattern matrix (10), an asterisk denotes a particular variable occurrence in an equation. The last four columns of the matrix stand for the original variables, \( z = [N, p, T, L]^T \), while the first four columns belong to their derivatives, \( z' = [N, p', T', L']^T \). In this problem, \( p \) and \( L \) are algebraic, while \( N \) and \( T \) are dynamic variables. Nevertheless, it was observed\textsuperscript{18} that the two variables are coupled algebraically so that the dynamics of the entire system is determined fully by the trajectories of either \( N(t) \) or \( T(t) \), even though two dynamic variables occur in system equations (6)–(9). The pattern matrix (10) contains no expressions for \( p' \) and \( L' \); hence, its index is higher than zero. Index analysis involves a minimum number of structural differentiation steps necessary for producing the derivatives for \( p' \) and \( L' \). The new pattern after one structural differentiation step, entering a star in the derivative column for each nonzero variable, is given in eq 11.

\[
N' \ T' \ p' \ L' \ N \ T \ p \ L
\]

\[
e_1 = [ \star \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ] \tag{10}
\]

Using the information of already known derivatives (\( N', T' \)), we structurally eliminate \( T' \) in \( e_1 \) and \( e_4 \) with the help of \( e_3 \). The same process applied to \( N' \) and \( p' \) renders the pattern matrix (12). The algorithm is not complete,

\[
N' \ T' \ p' \ L' \ N \ T \ p \ L
\]

\[
e_1 = [ \star \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ] \tag{11}
\]

because no derivative for \( L' \) has been identified yet. Hence, the last algebraic equation \( e_4 \) is structurally differentiated a second time (eq 13). Elimination of the explicit derivatives results in the desired expressions for \( L' \) in eq 14. Because two structural differentiation

\[
N' \ T' \ p' \ L' \ N \ T \ p \ L
\]

\[
e_1 = [ \star \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ] \tag{12}
\]

\[
N' \ T' \ p' \ L' \ N \ T \ p \ L
\]

\[
e_1 = [ \star \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ] \tag{13}
\]

\[
N' \ T' \ p' \ L' \ N \ T \ p \ L
\]

\[
e_1 = [ \star \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 ] \tag{14}
\]

steps were required to convert the system into an ODE, the structural index of the original DAE is 2. In this example, the structural index is equal to the correct differential index as can be verified by analytical differentiations. Unfortunately, structural index analysis does not always give the correct result.

**Shortcomings of Structural Approaches.** Structural equivalence transformations as demonstrated above tend to deteriorate the pattern information by