

Diagnosability Analysis Considering Causal Interpretations for Differential Constraints

Erik Frisk, Anibal Bregon, Jan Åslund, Mattias Krysander,
Belarmino Pulido and Gautam Biswas

Linköping University Post Print

N.B.: When citing this work, cite the original article.

©2012 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Erik Frisk, Anibal Bregon, Jan Åslund, Mattias Krysander, Belarmino Pulido and Gautam Biswas, *Diagnosability Analysis Considering Causal Interpretations for Differential Constraints*, 2012, IEEE transactions on systems, man and cybernetics. Part A. Systems and humans, (42), 5, 1216-1229.

<http://dx.doi.org/10.1109/TSMCA.2012.2189877>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-82059>

Diagnosability Analysis Considering Causal Interpretations for Differential Constraints

Erik Frisk, Anibal Bregon, Jan Åslund, Mattias Krysander, Belarmino Pulido, and Gautam Biswas

Abstract—This work is focused on structural approaches to studying diagnosability properties given a system model taking into account, both simultaneously or separately, integral and differential causal interpretations for differential constraints. We develop a model characterization and corresponding algorithms, for studying system diagnosability using a structural decomposition that avoids generating the full set of system Analytical Redundancy Relations. Simultaneous application of integral and differential causal interpretations for differential constraints results in a *mixed causality* interpretation for the system. The added power of mixed causality is demonstrated using a Reverse Osmosis Subsystem from the Advanced Water Recovery System developed at NASA Johnson Space Center. Finally, we summarize our work and provide a discussion of the advantages of mixed causality over just derivative or just integral causality.

Index Terms—Fault diagnosis, Diagnosability, Detectability, Isolability, Structural Analysis

I. INTRODUCTION

FAULT Detection and Diagnosis, FDD, are essential for Fault Tolerant Control and System Health Monitoring tasks. Model-based Reasoning has seen significant research activities from both the Systems Dynamics and Control Engineering (FDI) [1], [2], [3] and the Artificial Intelligence Diagnosis (DX) [4], [5] communities in the last three decades. The two communities have developed different algorithms that has been proved to be complementary [6]. An advantage of using Model-based techniques against other diagnosis approaches, like Expert Systems or Machine Learning, relies in the re-usability of models or diagnostic algorithms [7] (even though this re-usability process is quite difficult in practice).

In the last decade, a lot of work has been devoted to analyze diagnosability and sensor placement in the context of model-based diagnosis. Early works in the DX community on fault diagnosability were devoted to the definition and characterization of the diagnosability concept, based on fault detection and isolation results [8]. Recently, the process has been carried out by pre-computing the whole set of existing Analytical Redundancy Relations (ARRs) for a given set of sensors, and analyzing their discriminability properties [9], [10]. But such approaches are infeasible for large systems. More recently, [11], [12] have explored alternative and more efficient computational ways for diagnosability analysis.

Erik Frisk (frisk@isy.liu.se), Jan Åslund (jaasl@isy.liu.se), and Mattias Krysander (matkr@isy.liu.se) are with the Department of Electrical Engineering, Linköping University, Sweden.

Anibal Bregon (anibal@infor.uva.es) and Belarmino Pulido (belar@infor.uva.es) are with the Departamento de Informatica, Universidad de Valladolid, Spain.

Gautam Biswas (gautam.biswas@vanderbilt.edu) is with the Department of EECS/ISIS, Vanderbilt University, USA

Our work extends all these previous works by studying the diagnosability properties given a system model and a fixed set of sensors. The novelty of this work relies upon causality, i.e. the computational form, in constraints modeling dynamics. Our approach focuses on analyzing the structural model of the system to define and efficiently compute detectable and isolable parts of the system. The first contribution of this work is to consider either integral or derivative causality in differential constraints while performing system diagnosability analysis using the system model. The second contribution considers *mixed causality* –allowing the choice of derivative or integral causality on individual constraints – while analyzing system diagnosability. How to deal with loops for both approaches is discussed, and efficient algorithms for computing causal matchings in each case are developed. Then, we compare the diagnosability capabilities for the integral, derivative and mixed causality approaches. Finally, we discuss the applicability of sequential residual generators with the different causal interpretations.¹

To illustrate these concepts and algorithms we consider a complex case study, the Reverse Osmosis Subsystem from the Advanced Water Recovery System, developed at NASA Johnson Space Center [13], [14]. We have shown that the proposed algorithms can be used to analyze structural diagnosability of a complex system without exhaustive computation of the set of residuals. Moreover, we show that diagnosability is improved when mixed causality is considered.

This article is organized as follows. First, we introduce the problem formulation, together with a simple system to illustrate the main concepts of this work. The theoretical background for the basic concepts used in the proposal is provided next. This is followed by an analysis of the structural diagnosability properties of a system model under different causal interpretations for differential constraints. Next, we discuss the applicability of the proposed method to real systems, where both integral and derivative causalities are known to face different implementation issues. Finally, we present and discuss results obtained using the approach on the case study, and draw some conclusions.

II. PROBLEM FORMULATION

We use a simple three-tank system model (Fig. 1) to introduce the problem and formulate the different classes of residual generators that we discuss in the paper. The three-tank

¹A Matlab implementation is available at <http://www.fs.isy.liu.se/Software/CausalIsolability/>.

system model is represented by the set of equations

$$\begin{aligned}
 c_1 : q_1 &= \frac{1}{R_{V1}}(p_1 - p_2) & c_7 : y_1 &= p_1 \\
 c_2 : q_2 &= \frac{1}{R_{V2}}(p_2 - p_3) & c_8 : y_2 &= q_2 \\
 c_3 : q_3 &= \frac{1}{R_{V3}}(p_3) & c_9 : y_3 &= q_0 \\
 c_4 : \dot{p}_1 &= \frac{1}{C_{T1}}(q_0 - q_1) & c_{10} : \dot{p}_1 &= \frac{dp_1}{dt} \\
 c_5 : \dot{p}_2 &= \frac{1}{C_{T2}}(q_1 - q_2) & c_{11} : \dot{p}_2 &= \frac{dp_2}{dt} \\
 c_6 : \dot{p}_3 &= \frac{1}{C_{T3}}(q_2 - q_3) & c_{12} : \dot{p}_3 &= \frac{dp_3}{dt}
 \end{aligned}$$

where p_i is the pressure in tank i , q_i the flow through valve i , R_{Vi} the flow resistance of valve i , and C_{Ti} the capacitance of tank i . Three sensors y_1 , y_2 , and y_3 , measure p_1 , q_2 , and q_0 , respectively. For this study, six parametric faults have been considered in the plant: change in capacity of tanks C_{T1} , C_{T2} , and C_{T3} , and partial blocks in valves R_{V1} , R_{V2} , R_{V3} .

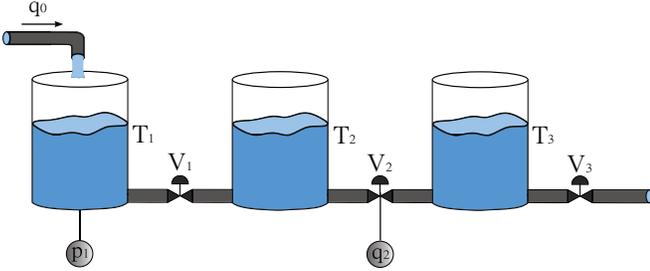


Fig. 1. Diagram of the three-tank system.

A sequential residual generator consists of a subset of equations that are used to compute the unknown variables included in these equations, and a redundant equation that checks the consistency between the observations and the considered subset of model equations. It is assumed here that all *algebraic* loops can be solved using symbolic or numeric solvers. This assumption is realistic since commercial packages for simulating differential-algebraic equations, e.g. Dymola, successfully use such techniques to solve large dynamical models [15].

A main concern is how to handle the dynamics in the model. For sequential residual generators, the literature report two options: (1) integral causality form, and (2) derivative causality form [1], [16], [9], [17]. This means that during computation, only differentiations or integrations are allowed. However, when solving *differential algebraic equations*, there is typically a need to include both differentiation and integration in the same solver [18]. For that reason, it is necessary to analyze the influence of combining both types of causal assignments, i.e., mixed causality, when using dynamic models.

As an example, consider the three-tank model and assume derivative causality. Fig. 2 graphically illustrates how variables p_1 , q_2 , q_0 , q_1 , p_2 can be sequentially computed from equations c_7 , c_8 , c_9 , c_4 , c_1 respectively, and \dot{p}_1 and \dot{p}_2 are computed by numeric differentiation using equations c_{10} and c_{11} . Finally, a residual is computed using equation c_5 as $r := C_{T2}\dot{p}_2 - q_1 + q_2$.

The variables are computed in a sequential way using numeric differentiation where needed, and no algebraic loops need to be solved.

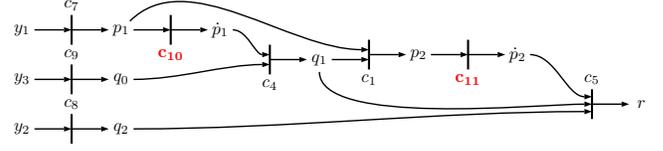


Fig. 2. A sequential residual generator in derivative causality.

Integral causality works in a similar way. Fig. 3 illustrate a sequential residual generator in integral causality where q_2 and q_0 are computed using c_8 and c_9 and then variables q_1 , p_1 , p_2 , \dot{p}_1 , and \dot{p}_2 are computed using the set of equations c_1 , c_4 , c_5 , c_{10} , c_{11} . Note that this is a differential loop [1] that has to be solved numerically, which can be done using any ODE-solver technique. Loops are broken by integrators, and sometimes this is referred to as a spiral [19]. Finally, the residual is computed using equation c_7 as $r := y_1 - p_1$.

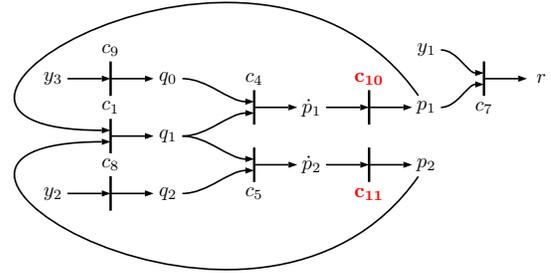


Fig. 3. A sequential residual generator in integral causality.

These two examples illustrate the main principle of sequential residual generation with the derivative and the integral causality assumptions. Here one can also note a fundamental difference between the two cases. A loop including dynamic constraints, i.e. any of c_{10} , c_{11} , c_{12} , in the integral causality case did not impose any difficulties since the loop could be directly solved by pure integration or using the structure to build a state observer [20]. However, a similar loop cannot be solved with a derivative causality assumption. This is because the loop corresponds to a differential equation which cannot be solved by only differentiating variables.

We demonstrate in Section V cases where neither derivative nor integral causality is enough and mixed causality has to be applied to compute all variables. Thus, different causality interpretations impose different constraints, which leads to the formulation of different residual generators. Therefore, different causality interpretations will likely result in different maximal structural diagnosability properties for a given model. Here, diagnosability refers to fault detectability and isolability properties of the model. See Section III-C for formal definitions. One possibility could be to compute the set of all Analytical Redundancy Relations (ARRs) [21], Possible Conflicts (PCs) [16], or Minimal Structurally Overdetermined (MSO) sets [22]. However, such an approach suffers from severe complexity properties since the number of MSO sets is exponential in the

model redundancy [22]. This paper addresses the problem of deriving efficient algorithms (which are not based on the set of ARR/PCs/MISO sets) using integral, derivative, or mixed causality interpretations, and then uses this model to determine the diagnosability properties of the system.

III. THEORETICAL BACKGROUND

This section recapitulates some basic formalism, concepts, and notation needed to describe the theoretical developments in Section IV.

A. Graph Representation of the Model

The class of models considered is a general class of first-order differential-algebraic equations in the form

$$g_i(x_1, \dot{x}_1, x_2, z) = 0, \quad i = 1, \dots, m \quad (1)$$

where $z \in \mathbb{R}^{n_z}$ is the vector of known variables, $x_1 \in \mathbb{R}^{n_1}$ is the vector of unknown dynamic variables, and $x_2 \in \mathbb{R}^{n_2}$ is the vector of unknown algebraic variables. Since the objective is to analyze the effect of causal assumptions it is convenient to add explicitly, for each component $x_{1,i}$ in x_1 , constraints capturing the dynamics

$$\dot{x}_{1,i} = \frac{d}{dt}x_{1,i}, \quad i = 1, \dots, n_1 \quad (2)$$

The constraints in (1) are algebraic and system dynamics are included in (2) which are referred to as dynamic or differential constraints. Note that the constraints expressed by equation (2) can be evaluated using two different causal interpretations:

- 1) *derivative causal interpretation (derivative causality*, for short), where $x_{1,i}$ is differentiated to obtain $\dot{x}_{1,i}$; and
- 2) *integral causal interpretation (integral causality*, for short), where $\dot{x}_{1,i}$ is integrated to obtain $x_{1,i}$.

Model analysis is based on the *model structure* rather than the analytical equations. This makes it possible to analyze large systems efficiently and with no numeric problems. The disadvantage is that the structural results may not be as precise as the corresponding analytical results. However, analytical results are computationally expensive if they are at all possible to derive. The reason is that the analytical results concern properties of solutions to the model equations, and often it is not possible to analytically solve nonlinear equations in the form (1).

The structure of a model is commonly represented by a bipartite graph as follows:

Definition 1. *The structural model graph for the model equations (1) and (2) is defined as a bipartite graph, $G(C, X, E)$, where C and X are node sets, such that $C = \{c_1, \dots, c_m\}$ is the set of constraints, X the set of unknown variables, and $E \subseteq C \times X$ edges such that $(c_i, x_j) \in E$ if $x_j \in X$ appears in constraint $c_i \in C$. The set of unknown variables are all components in x_1 , \dot{x}_1 , and x_2 .*

Since the objective is to analyze consequences of different causal interpretations of the differential constraints (2), the set of edges E is partitioned into $E = E_X \cup E_D \cup E_I$ where E_D is the set of edges corresponding to differentiated

variables $\dot{x}_{1,i}$ in the differential constraints (2), E_I the non-differentiated variables $x_{1,i}$ in the differential constraints, and E_X the remaining set of edges.

For example, the bi-adjacency matrix of the graph representing the three-tank model is shown in Table I, where X , D , and I indicate edges in E_X , E_D , and E_I , respectively.

TABLE I
BI-ADJACENCY MATRIX FOR THE STRUCTURAL MODEL OF THE THREE-TANK SYSTEM.

	q_0	q_1	q_2	q_3	p_1	p_2	p_3	\dot{p}_1	\dot{p}_2	\dot{p}_3
c_1		X			X	X				
c_2			X			X	X			
c_3				X			X			
c_4	X	X						X		
c_5		X	X						X	
c_6			X	X						X
c_7					X					
c_8			X							
c_9	X									
c_{10}					I			D		
c_{11}						I			D	
c_{12}							I			D

A number of simple graph operations and relations will be used in our algorithms. Let G be a structural model graph, E_1 a set of edges, X_1 a set of variables, and C_1 a set of constraints.

- $G(C, X, E) - E_1$ is the graph $G(C, X, E \setminus E_1)$.
- $G - X_1$ and $G - C_1$ are the graphs where a set of variables and constraints respectively are removed together with any corresponding connected edges.
- $C(G)$, $X(G)$, $E(G)$ are the constraints, variables, and edges respectively in a graph G .
- $C(G, E_1)$ and $C(G, X_1)$ are the set of constraints in graph G connected to edges E_1 or variables X_1 respectively.
- $G_1(C_1, X_1, E_1) \cup G_2(C_2, X_2, E_2)$ is the graph $G(C_1 \cup C_2, X_1 \cup X_2, E_1 \cup E_2)$.
- $G_1(C_1, X_1, E_1) \subset G_2(C_2, X_2, E_2)$ means that $C_1 \subset C_2$, $X_1 \subset X_2$, and $E_1 \subset E_2$.

B. The Dulmage-Mendelsohn Decomposition

A key tool when analyzing structural models is the Dulmage-Mendelsohn decomposition [23], used for diagnosis in [1], [11], [24]. The general Dulmage-Mendelsohn decomposition is illustrated in Fig. 4 where, by a suitable reordering of constraints and variables, the bi-adjacency matrix is converted to a triangular form. The sub-graph G^- with node sets C^- and X^- represents the underdetermined part of the model, G^0 with node sets C^0 and X^0 the exactly determined part, and G^+ with node sets C^+ and X^+ the overdetermined part. The overdetermined part contains redundancy, and can, therefore, be used for diagnosis. In the exactly determined part there is a finer structure of Hall-components, here denoted G_i . With some slight abuse of notation, $+$ and 0 will be used as operators on both graphs and set of constraints in the forthcoming sections.

A central concept used frequently in the following sections is *matching* [25]. A matching is a set Γ of edges such that no two edges in Γ have common nodes. A matching can, in the context of structural models, loosely be interpreted as which variable is solved in which equation, and also as a causal ordering or causal interpretation [26]. A matching is said to be

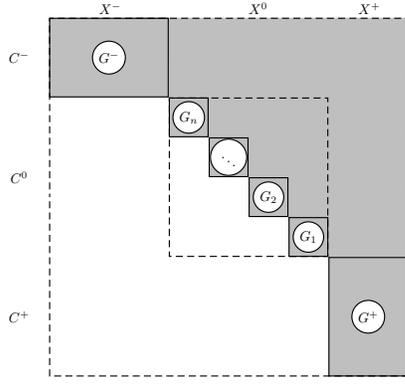


Fig. 4. Dulmage-Mendelsohn decomposition.

complete with respect to a node set if all nodes in the set are matched and perfect if the matching is complete with respect to both node sets in the bi-partite graph.

C. Structural Detectability and Isolability

Given a structural model and the Dulmage-Mendelsohn decomposition, we can now recapitulate standard definitions on structural detectability and isolability. These definitions will then be extended in Section IV to cover the cases where there are causal constraints. Without loss of generality, it is assumed that a fault f only influences one constraint, denoted c_f . In case a fault signal f appears in more than one constraint, extend the model with a new constraint

$$x_f = f$$

and substitute f for x_f in the rest of the model. In the three-tank system model, constraints $\{c_1, c_2, c_3, c_4, c_5, c_6\}$ are linked to faults in resistances and capacitances, constraints $\{c_7, c_8, c_9\}$ are linked to sensor faults, and constraints $\{c_{10}, c_{11}, c_{12}\}$ are not linked to faults.

Then from [1], [11]:

Definition 2 (Structural Detectability). *A fault f is structurally detectable in a model if*

$$c_f \in C^+$$

Following the ideas in [11], a fault f_i is isolable from a fault f_j if f_i is detectable in the model $G - c_{f_j}$, i.e.

Definition 3 (Structural Isolability). *A fault f_i is structurally isolable from f_j in a model if*

$$c_{f_i} \in (C \setminus \{c_{f_j}\})^+$$

To illustrate the definition, consider the case of determining which faults that are structurally isolable from a fault in valve V_2 . Fig. 5 shows the just-determined and the over-determined part of $C \setminus \{c_{f_j}\}$ with $c_{f_j} = c_2$.

Since c_3 and c_6 , corresponding to faults in valve V_3 and tank T_3 , are not in the overdetermined part $\{c_1, c_4, c_5, c_7, c_8, c_9, c_{10}, c_{11}\}$, the definition states that a fault in valve V_3 and a fault in tank T_3 is not isolable from a fault in valve V_2 . All other faults are isolable from a fault in valve V_2 .

	p_3	q_3	\dot{p}_3	p_1	p_2	q_0	q_1	q_2	\dot{p}_1	\dot{p}_2
c_2	X				X			X		
c_{12}	X		X							
c_3	X	X								
c_6		X	X					X		
c_7				X						
c_{11}				X	X					X
c_4				X	X	X	X		X	
c_1				X	X		X			
c_8							X			
c_{10}				X				X	X	
c_5							X	X		X
c_9						X				

Fig. 5. Illustration of isolability from a fault in R_{V_2} , that is linked to constraint c_2 .

IV. DIAGNOSABILITY UNDER CAUSAL CONSTRAINTS

This section first introduces formal definitions on causal detectability and isolability and then proceeds to develop the algorithms and formal proofs of their correctness.

A. Basic Definitions

As discussed in Section II, a causal assumption imposes an ordering on how the unknown variables are computed in a system. If a proper order can be found, the system is solvable. This is better formulated as shown in Fig. 4, therefore, the first step en route to defining detectability and isolability under a causality assumption is to formally define solvability. Note that special attention has to be given to the Hall components, see Fig. 4, which correspond to a set of variables that has to be solved simultaneously in a set of constraints. For the case where no causality constraints are imposed, a solvability condition is then that there exists a complete matching with respect to the unknown variables.

As discussed in Section II, non-trivial loops involving integral constraints can be solved sequentially. Then, since it is assumed that algebraic loops can be solved, solvability for integral causality is defined as follows:

Definition 4. *A Hall component G is structurally solvable under integral causality if there exists a perfect matching Γ in G such that*

$$\Gamma \subseteq E_X \cup E_I$$

The definition is quite natural. A matching, with no derivative edges, that is complete in the unknown variables gives a computational sequence. The computational sequence may involve Hall components of size larger than 1, i.e., more than one variable has to be solved simultaneously. If the Hall component includes an integration, the computational loop is naturally broken [19], and if it is a pure algebraic loop, it is assumed that any such loop can be solved numerically.

For the derivative causality case, a similar definition can be stated. Here it is important to note that a non-trivial Hall component with a derivative edge can not be solved using differentiation only. This is because the equations in the Hall component correspond to a differential-equation to be solved which implies that integration is needed.

Definition 5. *A Hall component G is structurally solvable under derivative causality if there exists a perfect matching Γ in G such that*

- $\Gamma \subseteq E_X \cup E_D$, and

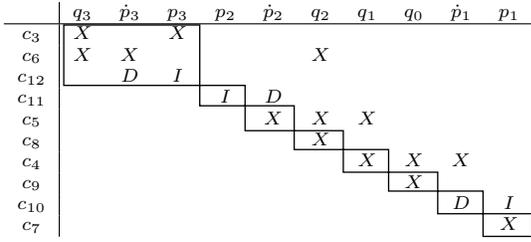


Fig. 6. A perfect matching Γ with causally solvable Hall components.

- $\Gamma \cap E_D \neq \emptyset$ implies that $|\Gamma| = 1$.

where $|\Gamma|$ is the cardinality of the set Γ .

The second condition ensures that there are no non-trivial loops with derivative edges.

For the mixed causality case, a matching Γ can structurally solve all variables if all variables that are not computed by integration can be solved using derivative causality. Thus, the solvability definition for the mixed causality case can then be stated based on Definitions 4 and 5.

Definition 6. A Hall component G is structurally solvable under mixed causality if there exists a perfect matching Γ in G such that all Hall components in $G'(\Gamma) = G - C(G, \Gamma \cap E_I) - X(G, \Gamma \cap E_I)$ are structurally solvable under derivative causality.

Fig. 6 illustrates a perfect matching, from the three-tank example, such that all Hall components, indicated by the boxes, are causally solvable.

Using Definition 2 in Section III that the monitorable part of a model is its overdetermined part, i.e., C^+ in Fig. 4, solvability can be defined for the three different causality interpretations. The set of constraints that is structurally monitorable can be directly defined using the following definition.

Definition 7. Given a causal assumption, a set of constraints C is structurally monitorable if

- $C = C^+$, and
- there exists a complete matching Γ with respect to all unknown variables in C such that all Hall components, induced by Γ , are structurally solvable under the causal assumption.

The Hall components induced by Γ means the Hall components of the sub-graph defined by the constraints and unknowns in Γ . This sub-graph is exactly determined and therefore are the induced Hall components given by a Dulmage-Mendelsohn decomposition as in Fig. 4. The union of two structurally monitorable sets is also monitorable and, therefore, there is a unique maximal monitorable set which is the union of all monitorable sets. This maximal set of constraints is of special importance since this set is a direct counterpart to the overdetermined part used in Definitions 2 and 3.

Definition 8. Given a causal assumption, $causal \in \{der, int, mix\}$, the set of structurally monitorable constraints under the causal assumption, C_{causal}^+ , is the maximal set of structurally monitorable constraints.

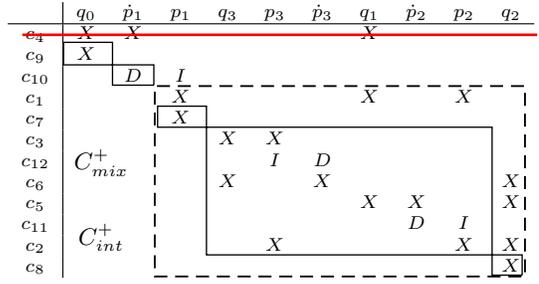


Fig. 7. Computation of $(C \setminus \{c_4\})_{causal}^+$.

With the definition of C_{causal}^+ , extensions of Definitions 2 and 3 are direct and summarized as:

Definition 9. A fault f is causally structurally detectable in a model if

$$c_f \in C_{causal}^+$$

A fault f_i is causally structurally isolable from f_j in a model if

$$c_{f_i} \in (C \setminus \{c_{f_j}\})_{causal}^+$$

Thus, algorithms that compute C_{causal}^+ for $causal \in \{der, int, mix\}$ are sufficient to evaluate structural diagnosability properties of a given model.

Again, consider the three-tank example. To determine causal isolability from a fault in tank 1 capacitance, remove equation c_4 from the model and compute C_{causal}^+ for the three cases. Fig. 7 shows the result. The boxes indicate a matching, with corresponding Hall components, and the dashed box show the overdetermined part. Then, since all hall components are solvable in both mixed and integral causality, $(C \setminus \{c_4\})_{mix}^+ = (C \setminus \{c_4\})_{int}^+ = \{c_1, c_2, c_3, c_5, c_6, c_7, c_8, c_{11}, c_{12}\}$, and therefore are all faults isolable from a fault in tank 1 in mixed and integral causality. Similarly, it can be shown that $(C \setminus \{c_4\})_{der}^+ = \emptyset$ and thus no faults are isolable from a fault in tank 1 in the derivative causality case.

B. Computing Monitorable Part Under a Causal Assumption

This section provides algorithms, and formal proofs, on how to compute C_{causal}^+ and, for a given set of equations, a causal matching. Computation of C_{causal}^+ makes it possible to determine isolability properties according to Definition 9 and with a causal matching it is possible to derive sequential residual generators as described in Section II. Integral, derivative, and mixed causality constraints will be treated separately.

1) *Integral causality:* The algorithm in [24] can be directly used to compute C_{int}^+ . An algorithm description is included here, which is equivalent to the one in [24], but uses the notation introduced in Section III.

Algorithm 1: Compute C_{int}^+

```

function computeInteg( $G(C, X, E_X \cup E_I \cup E_D)$ )
repeat
   $G := G^+$ ;
   $G_1 := G - E_D$ ;
   $G := G - C(G, X(G_1^-))$ ;
until  $G_1^- = \emptyset$ ;
return  $C(G)$ 

```

	\dot{p}_2	q_1	\dot{p}_1	q_0	\dot{p}_3	q_3	p_1	p_2	p_3	q_2	
c_5	X	X								X	
c_4		X	X	X							
c_9				X							
c_6					X	X					X
c_3						X			X		
c_7							X				
c_2							X	X	X	X	
c_{12}				D					I		
c_8											X
c_{10}							I				
c_{11}	D		D								

Fig. 8. Bi-adjacency matrix for the structural model used to illustrate Algorithm 1.

The algorithm works by iteratively removing variables that can not be computed when no differential edges can be used in a matching. To obtain a causal matching, consider the graph G_1 after the final iteration. First, any matching in G_1 is causal according to Definition 4. Also, observe that when the iteration terminates it holds that $X(G) = X(G_1)$ and that $G = G^+$, which means that the causal matching in G_1 is also a causal matching for the variables in C_{int}^+ .

To illustrate how Algorithm 1 works consider the graph represented by Fig. 8 as input. This graph shows the structure of the model of three-tank system, when fault R_{V_1} is decoupled, i.e., constraint c_1 has been removed. In the first iteration G_1 is the graph in the figure when the differential edges marked D have been removed. The Dulmage-Mendelsohn decomposition of G_1 is indicated in the figure. The unknowns in the under-determined part $X(G_1^-) = \{\dot{p}_2, q_1, \dot{p}_1\}$ are not structurally solvable under integral causality and the constraints including these variables $C(G, X(G_1^-)) = \{c_4, c_5, c_{10}, c_{11}\}$ are not monitorable, hence removed from G . Since $G_1^- \neq \emptyset$, a second iteration starts with $G = G^2$ where the nodes sets of G^2 are indicated with lines on the top and on the right side of the figure. Since $(G^2)^+ = \emptyset$ the algorithm terminates after this iteration and returns $C(G) = C(\emptyset) = \emptyset$. Hence no equations are structurally monitorable under integral causality if R_{V_1} is decoupled.

2) *Derivative causality*: For the derivative causality case, note that the algorithm from [24] cannot be used since special attention has to be given to loops involving differential constraints, i.e., condition 2 in Definition 5. The algorithm works by first computing, again in an iterative manner, the set of all computable variables X_c under derivative causality and then the structurally monitorable part under derivative causality C_{der}^+ is computed.

Algorithm 2: Compute C_{der}^+

```

function computeDeriv( $G(C, X, E_X \cup E_I \cup E_D)$ )
 $X_c := \emptyset$ ;
repeat
   $G_{nc} := G - X_c$ ;
   $G_{ni} := G_{nc} - C(G_{nc}, E_I)$ ;
   $X_c := X_c \cup X(G_{ni}^+ \cup G_{ni}^0)$ ;
until  $X(G_{ni}^+ \cup G_{ni}^0) = \emptyset$ ;
 $G := G - C(G, X \setminus X_c)$ ;
return  $C(G^+)$ 

```

To illustrate how Algorithm 2 works, consider again the three-tank example. Fig. 9 shows the model when fault R_{V_1} is

	p_2	p_3	q_3	\dot{p}_3	\dot{p}_2	q_1	\dot{p}_1	p_1	q_2	q_0	
c_2	X	X							X		
c_3		X	X	X							
c_6			X	X					X		
c_{11}	I				D						
c_{12}		I		D							
c_5					X	X			X		
c_4					X	X				X	
c_{10}							D		I		
c_7								X			
c_8									X		
c_9										X	

$X_c^2 = X_c^3$

Fig. 9. Bi-adjacency matrix for the structural model used to illustrate Algorithm 2.

decoupled, i.e., constraint c_1 has been removed.

Let input G be the corresponding graph. No unknown variable has not yet been identified to be structurally solvable, thus $X_c = \emptyset$. In Algorithm 2, X_c represents the set of computable variables, G_{nc} represents the subgraph of G after removing the computable variables, and G_{ni} represents the subgraph of G_{nc} after removing the integral constraints. In the first iteration G_{ni} is set to G_{ni}^1 which is the graph with node sets indicated with lines in the figure. The difference between G_{ni}^1 and the input graph is that the differential constraints have been removed. The structurally solvable variables in G_{ni}^1 are $X_c^1 = \{p_1, q_2, q_0\}$ also indicated with a line in the figure. The outcome of the first iteration is that $X_c = X_c^1$. In the second iteration $G_{ni} = G_{ni}^2$ where the so far identified solvable unknowns X_c^1 have been removed. Note that c_{10} is included in G_{ni}^2 , which was not the case in G_{ni}^1 , since $p_1 \in X_c^1$. The interpretation is that c_{10} can be used to solve for \dot{p}_1 under derivative causality. The solvable unknowns in G_{ni}^2 are $\{\dot{p}_2, q_1, \dot{p}_1\}$ and the updated set of identified solvable unknowns becomes $X_c = X_c^2$ where X_c^2 is indicated in the figure. In the third iteration $G_{ni} = G_{ni}^3$ and $X_c = X_c^3 = X_c^2$. Since the third iteration did not identify any new solvable unknown, $X_c = X_c^3$ is the complete set of structurally solvable unknowns in the input graph and the iteration stops. In the second to last line of the algorithm the constraints $\{c_2, c_3, c_6, c_{11}, c_{12}\}$ containing unsolvable unknowns are removed. Finally, the remaining constraints contain no redundancy and the output is $C(G^+) = \emptyset$.

Theorem 1. *The output of Algorithm 2 satisfies the condition in Definition 8 for derivative causality.*

Proof: Consider C_{der}^+ and let the corresponding sub-graph be denoted G_d^+ . As a first step, we will show that C_{der}^+ is a subset of the output of the algorithm. According to the definition of C_{der}^+ , there exists a matching Γ , Hall components $H(\Gamma) = \{G_1, \dots, G_n\}$, and for each component G_k there exists a matching Γ_k such that $\Gamma_k \subset E_X \cup E_D$, and $\Gamma_k \cap E_D \neq \emptyset$ implies that $|\Gamma_k| = 1$. This is equivalent to the condition

$$E(G_k) \cap E_I = \emptyset \quad (3)$$

Assume that the Hall components are enumerated as in Fig. 4 and define $X_k = \cup_{j \leq k} X(G_j)$. In the first iteration $G_1 \subset G_{nc}$ since $X_c = \emptyset$. It follows from (3) that no part of G_1 is removed

when G_{ni} is created and $G_1 \subset G_{ni}$. Using that G_1 is a Hall component we get

$$G_1 = G_1^+ \cup G_1^0 \subset G_{ni}^+ \cup G_{ni}^0$$

Hence $X_1 \subset X_c$ after the first iteration.

In the second iteration, it follows from the definition of G_{nc} that $G_2 - X_c \subset G_{nc}$. Furthermore, $X_1 \subset X_c$ and condition (3) imply that none of the constraints in $C(G_2 - X_c)$ is removed when G_{ni} is computed, and hence $G_2 - X_c \subset G_{ni}$.

Using that $G_2 - X_c$ has no underdetermined part, $G_2 - X_c \subset G_{ni} - X_c$ and $X(G_{ni}) \cap X_c = \emptyset$ we get

$$\begin{aligned} G_2 - X_c &= (G_2 - X_c)^+ \cup (G_2 - X_c)^0 \\ &\subset (G_{ni} - X_c)^+ \cup (G_{ni} - X_c)^0 = G_{ni}^+ \cup G_{ni}^0 \end{aligned}$$

and it follows that

$$X_2 \subset X_c \cup X(G_2 - X_c) \subset X_c \cup X(G_{ni}^+ \cup G_{ni}^0)$$

where the set on the right-hand side is the set X_c after the second iteration.

We have shown that $X_2 \subset X_c$ after the second iteration and we can continue in the same way and show that $X_k \subset X_c$ after k iterations and $X(G_d^+) = X_n \subset X_c$ after at most n iterations. It follows that $C_{der}^+ \cap C(G, X \setminus X_c) = \emptyset$, $G_d^+ \subset G - C(G, X \setminus X_c)$, and we get $C_{der}^+ \subset C(G^+)$ in the final step of the algorithm.

The next step is to show that the output of the algorithm is a subset of C_{der}^+ . To do this it is sufficient to show that the output is structurally monitorable, since C_{der}^+ is the largest structurally monitorable subset of C . In iteration k there exists a complete matching Γ_k with respect to the variables in $G_{ni}^+ \cup G_{ni}^0$; see Fig. 10.

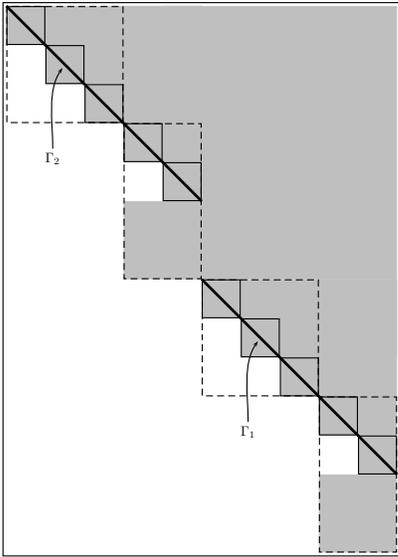


Fig. 10. Causal Matchings Γ_1 and Γ_2 after iteration 2.

Let $H(\Gamma_k) = \{G_{k1}, \dots, G_{kn_k}\}$ denote the induced Hall components. All Hall components G_{kj} fulfill condition (3), since $E(G_2) \cap E_I = \emptyset$ by definition. Define $\Gamma = \cup_k \Gamma_k$, which is a complete matching with respect to the variables in X_c in $G - C(G, X \setminus X_c)$ and the induced Hall components are given

by $\cup_k H(\Gamma_k)$. After the operation $G := G - C(G, X \setminus X_c)$ the set $C(G)$ is a set of constraints where all unknown variables can be computed under the derivative causality assumption. By removing the exactly determined part of the model a structurally monitorable set of constraints is obtained. ■

The proof of Theorem 1 includes a constructive procedure to compute a causal matching Γ for the variables included in C_{der}^+ .

3) *Mixed causality*: The mixed causality case is treated by first considering an exactly determined model and proving, in a constructive manner, that there *always* exists a causal matching Γ . Then, this result is used to state an algorithm for computing the set C_{mix}^+ .

For an exactly determined model, i.e. the graph G satisfies that $G = G^0$, the set H of Hall components is uniquely defined and given by the Dulmage-Mendelsohn decomposition described in Section III-B. Let the set of admissible edges $A(G)$ be defined as

$$A(G) = \bigcup_{g \in H} E(g) \quad (4)$$

These edges are called admissible since these are the only edges in G included in some perfect matching of G . The following algorithm computes a causal matching, assuming mixed causality, for any exactly determined system.

Algorithm 3: Mixed causality matching

```

function mixedCausalityMatching( $G(C, X, E_X \cup E_I \cup E_D)$ )
 $\Gamma := \emptyset$ ;
while  $A(G) \cap E_I \neq \emptyset$  do
    Select any  $e \in A(G) \cap E_I$ ;
     $\Gamma := \Gamma \cup \{e\}$ ;
     $G := G - C(\{e\}) - X(\{e\})$ ;
end
Let  $\Gamma'$  be any perfect matching of  $G$ ;
 $\Gamma := \Gamma \cup \Gamma'$ ;
return  $\Gamma$ 

```

Correctness of the algorithm is proven in the following theorem:

Theorem 2. For a graph that satisfies $G = G^0$, Algorithm 3 returns a perfect matching Γ such that all Hall components in G are structurally solvable under mixed causality according to Definition 6.

Proof: Since e is included in a perfect matching, there exists a perfect matching in $G - C(\{e\}) - X(\{e\})$ as well, and $A(G)$ is well defined in each iteration. The set of admissible edges, $A(G)$, is decreasing in each iteration and after the final iteration a reduced graph G is obtained with the property $A(G) \cap E_I = \emptyset$. Let the Hall components in the original graph be denoted by $\{G_1, \dots, G_n\}$. After the reduction each Hall component G_k has a structure similar to the one illustrated in Fig. 11. Let the Hall components in the reduced Hall component G'_k be denoted by $H_k = \{G_{k1}, G_{k2}, \dots\}$. The Hall components in the reduced graph are then given by $H = \cup_k H_k$ and it follows from $A(G) \cap E_I = \emptyset$ that

$$E(G_{kj}) \cap E_I = \emptyset \quad (5)$$

The matching Γ obtained by the algorithm can be partitioned

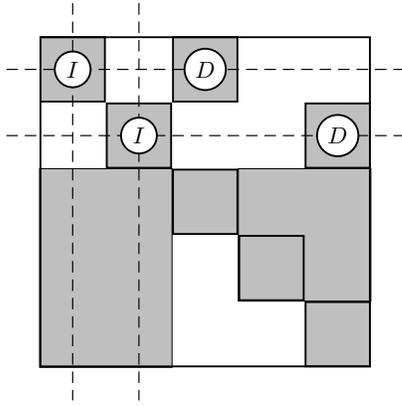


Fig. 11. Reduced Hall component G'_k

into sets Γ_k , $k = 1, \dots, n$, where each Γ_k is a perfect matching in G_k .

It follows from the construction that $G'_k = G_k - C(\Gamma_k \cap E_I) - X(\Gamma_k \cap E_I)$, and it follows from (5) that all Hall components in G'_k are solvable under derivative causality. ■

Based on Theorem 2, the following result on how to compute C_{mix}^+ is immediate.

Corollary 1. *Given a structural model graph G , the set of constraints in the overdetermined part G^+ equals C_{mixed}^+ as defined in Definition 8.*

Proof: The result in Theorem 2 states that, in an exactly determined system there *always* exists a mixed causal matching. From this follows that mixed causality is as general as the no causality case and thus $C_{mix}^+ = C(G^+)$. ■

The result can be summarized in the algorithm below.

Algorithm 4: Compute C_{mix}^+

function computeMixed(G)
 $C_1 := C(G^+)$;
return C_1

The three-tank system model (shown in Fig. 1) is used to illustrate the proposed approach. In this section, the fault detectability analysis is performed, and then, based on the detectability results, single fault isolability analysis is carried out.

C. Computational Complexity

The time-complexity of the presented algorithms is determined by the number of times the Dulmage-Mendelsohn decomposition is performed. Let n be the number of equations and m the number of unknown variables in an input graph G . The time-complexity of computing the Dulmage-Mendelsohn decomposition of G is of order $nm\sqrt{\min(m, n)}$ [27].

Algorithm 4 makes one Dulmage-Mendelsohn decomposition and thus have the same time-complexity. The number of Dulmage-Mendelsohn decompositions performed in Algorithms 1, 2, and 3 is linear in the number of unknowns m . Hence, the time-complexity of these algorithms is of order $nm^2\sqrt{\min(m, n)}$. This means that all algorithms have polynomial complexity.

D. Diagnosability Analysis of the Three-tank Model

Applying algorithms 1, 2, and 4 on the three-tank model automatically provides the monitorable part of the model for integral, derivative, and mixed causality respectively. The results show that all the constraints influenced by the faults considered belong to C_{int}^+ , C_{der}^+ , or C_{mix}^+ , hence, the system has full structural detectability when any of the three interpretations is considered.

To illustrate single fault isolability properties of the model, matrices representing isolability properties are computed for each one of the causal interpretations considered. Tables II, III, and IV show the isolability matrices when derivative, integral, and mixed causality, respectively, is considered. Columns and rows of the isolability matrix represent the faulty candidates considered. An X in position (i, j) indicates that fault j cannot be isolated from fault i . Isolability matrices were computed using the algorithms proposed and ideas of Definition 9.

When derivative causality is considered, the diagnosis system cannot provide full isolability, because faults in R_{V_1} and C_{T_1} cannot be isolated from the rest of the faults in the system, and faults in R_{V_2} , R_{V_3} , C_{T_3} cannot be isolated among themselves. Only faults in C_{T_2} can be uniquely isolated using derivative causality.

Integral causality provides better isolability than the derivative case: using integral causality, faults in C_{T_1} and C_{T_2} can be isolated from the rest of the faults in the system.

Finally, mixed causality provides the best results for isolability: faults in R_{V_1} , C_{T_1} , C_{T_2} can be isolated from the rest of the faults in the system.

TABLE II
ISOLABILITY MATRIX FOR THE THREE-TANK SYSTEM WHEN DERIVATIVE CAUSALITY IS CONSIDERED.

	R_{V_1}	R_{V_2}	R_{V_3}	C_{T_3}	C_{T_1}	C_{T_2}
R_{V_1}	X	X	X	X	X	X
R_{V_2}		X	X	X		
R_{V_3}		X	X	X		
C_{T_3}		X	X	X		
C_{T_1}	X	X	X	X	X	X
C_{T_2}						X

TABLE III
ISOLABILITY MATRIX FOR THE THREE-TANK SYSTEM WHEN INTEGRAL CAUSALITY IS CONSIDERED.

	R_{V_1}	R_{V_2}	R_{V_3}	C_{T_3}	C_{T_1}	C_{T_2}
R_{V_1}	X	X	X	X	X	X
R_{V_2}		X	X	X		
R_{V_3}		X	X	X		
C_{T_3}		X	X	X		
C_{T_1}					X	
C_{T_2}						X

V. APPLICABILITY OF SEQUENTIAL RESIDUAL GENERATORS WITH DIFFERENT CAUSAL INTERPRETATIONS

The objective of this section is, first, to discuss sequential residual generators, and then, discuss and compare implementation aspects of derivative, integral, and mixed causality.

Sections III and IV described theory and methodology for diagnosability analysis under the constraint of using *sequential*

TABLE IV
ISOLABILITY MATRIX FOR THE THREE-TANK SYSTEM WHEN MIXED CAUSALITY IS CONSIDERED.

	R_{V_1}	R_{V_2}	R_{V_3}	C_{T_3}	C_{T_1}	C_{T_2}
R_{V_1}	X					
R_{V_2}		X	X	X		
R_{V_3}		X	X	X		
C_{T_3}		X	X	X		
C_{T_1}					X	
C_{T_2}						X

residual generators. This particular method is attractive because of its general applicability to different types of models and the significant possibility for automatic syntheses of residual generators [28]. Models of industrial systems are typically large scale dynamic models including non-linearities such as look-up tables and saturations. Such models can without modifications be the input to sequential residual generation algorithms. This is an advantage compared to many other residual generation approaches where the model needs to be written in a more restrictive form, for example as a state-space, control affine, or polynomial form. Another attractive feature of sequential residual generation is the low computational complexity of the residual generation synthesis. A main reason for its low complexity is that no analytical computations are required, instead, a computation scheme is derived from efficient structural algorithms. This makes this approach feasible even for large scale models where analytical design methods have problems. For example, variable elimination based techniques [29] suffer from severe complexity problems [30, p. 108] and [31] has to deal with analytical difficulties wherever solutions to a non-linear partial-differential equation are needed in the design.

The above discussion motivates the usage of sequential residual generators. With that choice, there is still the choice of using derivative, integral, or mixed causality.

For the derivative case, approximations of derivatives are needed. For example, a simple way is to implement

$$\hat{y}(t) \approx \frac{p}{p/\omega_c + 1} y(t) \quad (6)$$

where p is the differentiation operator and ω_c some suitable cut-off frequency. In difficult situations with high noise levels, such a simple approach may not be sufficient and other more advanced methods, for example those based on spline interpolation [32], can be explored to better cope with noise. A main drawback is that imposing differential causality means a restriction in the possible residual generators. And, as seen in the three-tank example in Section IV, isolability performance may be affected. Another main drawback is the complication of differentiating noisy measurement signals and the higher the order of derivative, the more the noise is amplified. In addition, if the derivative approximation is done with a causal filter like (6), there will be a slight time delay for each differentiation which, for higher order redundancy relations, will make signals out of sync, and, therefore, introduce additional disturbances in the residual. However, the direct approach with derivative estimation may work very well in applications. For example, the work in [33] indicates positive results using derivative causality in a heavy duty vehicle engine application using real

measurement data from production sensors. In contrast with the integral causality case, stability of the residual generator is always ensured.

For the integral causality case, there is a similar restriction in possible residual generators as for the derivative causality case. There are no fundamental problems with noisy measurement signals, but this comes at the price of not being able to guarantee stable residual generators. Note that, just because the process model is stable, there is no guarantee that the square subsystem of an MSO set is stable. Another issue is the unknown initial condition and that biased models may introduce difficulties. Consider a model of a rotating machinery with a small bias in the friction model, which is not an unreasonable situation:

$$\dot{\varphi} = \omega \quad (7)$$

$$J\dot{\omega} = T_{control} - T_{fric} \quad (8)$$

$$y = \varphi \quad (9)$$

Pure integration will then not be a good idea and biased models are common in industrial cases, see for example [34], [35]. Anyway, these problems regarding biased models can also be avoided. In [20], a framework is proposed where state observers are used to estimate the initial conditions for simulation. Then, these initial conditions are used to reset the open loop integration at suitable simulation intervals.

The main advantage with the mixed causality case is that no restrictions are imposed on the possible residual generators. In the three-tank example, it was clear from Tables II and III that it is not possible to isolate fault in R_{V_1} from fault in R_{V_2} using neither derivative nor integral causality. Fig. 12 shows a computational graph for a residual obtained for mixed causality. Looking at the differential constraints in the residual, c_{10} uses derivative causality, while both c_{11} and c_{12} use integral causality. Only mixed causality allows for a residual to isolate faults in R_{V_1} (in this case this is the only residual not containing the constraint c_1). This residual was obtained from the causal matching automatically provided by Algorithm 3 when mixed causality is considered. Also, Corollary 1 states

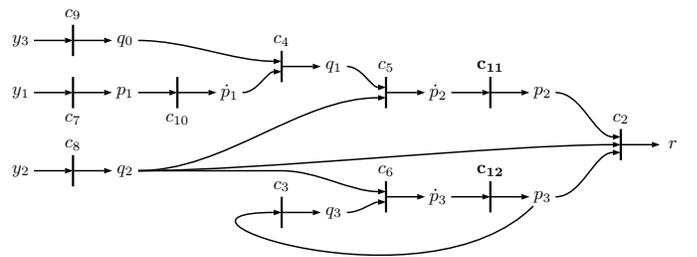


Fig. 12. Additional residual obtained when mixed causality is considered.

that performing analysis using mixed causality is equivalent to an analysis with no restrictions on causal interpretation of differential constraints. Thus, an upper limit of possible structural diagnosis performance is obtained with a free choice of residual generation method, including sequential generators. But of course, the mixed causality case inherits properties, the good ones as well as the bad ones, from both derivative and integral causality meaning that both noise sensitivity and stability of the residual generator has to be considered.

Typically, there is a possibility to lower the highest order derivative using a mixed approach compared to a purely derivative causality approach, which helps mitigating problems with sensor noise.

VI. CASE STUDY: THE REVERSE OSMOSIS SYSTEM

The use and storage of consumable resources during extended duration manned space missions entails high costs and requires to fulfill high safety constraints. The development of an *Life Support System* (LSS) will allow to reduce costs and increase safety by minimizing the current dependency on resupply missions [36]. To show the suitability and the improvement of our proposal, we apply the approach to a subsystem of the LSS, the Advanced Water Recovery System (AWRS), designed and built at NASA Johnson Space Center (JSC) [13], [14]. The AWRS (see Fig. 13) works in microgravity conditions converting waste water to potable water. As shown in Fig. 13, the AWRS is made up of four different subsystems: (1) the Biological Waste Processor, or BWP; (2) the Reverse Osmosis System, or ROS; (3) the Air Evaporation System, or AES; and (4) the Post Processing System, or PPS. Due to the high complexity of each one of these components, this case study only discusses the ROS results.

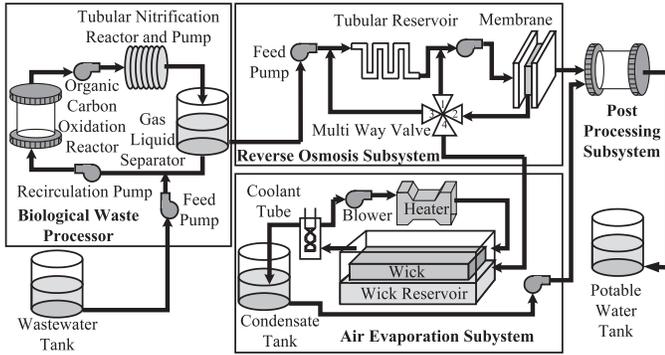


Fig. 13. Schema of the Advanced Water Recovery System with the Reverse Osmosis System [37].

The ROS uses a membrane system to remove inorganic matter from the effluent of the BWP. The inflow to the ROS, i.e. the effluent from the BWP, draws into a coiled section of pipe (Reservoir), and then cycles through the loop including the feed pump (Fp, which is always on), the tubular reservoir (TubRes), the recirculating pump (Rp), the membrane (Memb), and the recirculation path (Pipe) that includes a multi way valve. In this cycle the recirculating pump pulls the water at high pressure through the membrane to obtain clean water, called permeate. This water is then sent to the PPS, and the remaining water is recirculated in a feedback loop. The recirculation of water increases the concentration of impurities, called brine. Approximately 85% of the input effluent is cleaned by the ROS, and the remaining, the brine, is delivered to the AES.

Two measured variables have been selected for diagnosis experiments in this system: (1) the pressure immediately after the recirculation pump, P_{Pump} ; and (2) the pressure of the permeate at the membrane, P_{Memb} .

The ROS was initially modelled using the bond graph modeling approach [38], [39] (see Fig. 14). Using the bond graph model of the system, the set of equations describing the structural model of the ROS was automatically computed. Table V shows the bi-adjacency matrix for the structural model of the ROS. Rows and columns in the matrix represent equations and variables of the structural model, respectively. Variables named with an e in the matrix correspond to effort variables (i.e., fluid pressures) in the bond graph model, while variables named with an f correspond to flow variables (i.e., fluid flowrates) in the bond graph model. Numbers in the variables indicate the bond number in the bond graph model. Equations $c30$ to $c35$ are the differential equations of this system.

Table VI describes the ROS model parameters, that are also the fault candidates considered. Additional information regarding the ROS can be found in [39], [40].

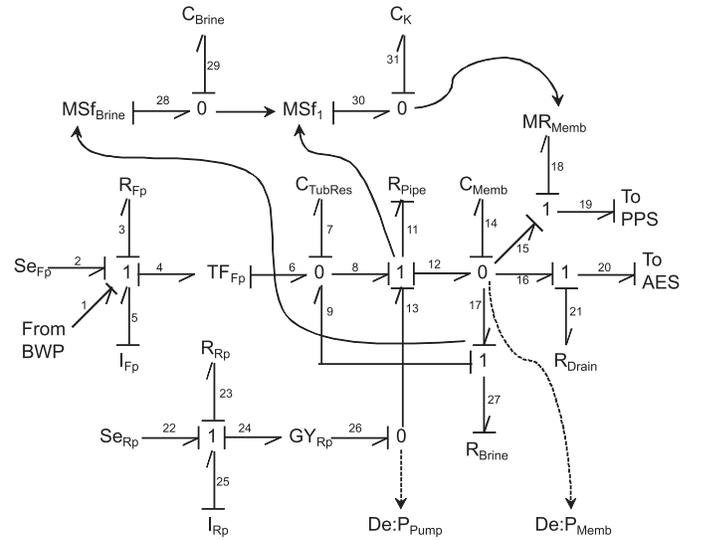


Fig. 14. Bond graph model of the Reverse Osmosis System [39], [40].

TABLE VI
ROS MODEL PARAMETERS AND DESCRIPTION OF POTENTIAL FAULTS.

Fault	Description
C_{Memb}	Buildup of impurities in the membrane
C_{TubRes}	Buildup of impurities in the tubular resistance
GY_{Rp}	Decrease in efficiency of the recirculation pump
I_{Fp}	Change in inertia of the feed pump
I_{Rp}	Change in inertia of the recirculation pump
R_{Brine}	Partial blockage in the pipe carrying brine
R_{Memb}	Partial blockage in flow through the membrane
R_{Pipe}	Partial blockage in pipe carrying water to the membrane
R_{Rp}	Increase in friction in the recirculation pump
R_{Fp}	Increase in friction in the feed pump

A. Fault Detectability and Isolability Results

Algorithms 1, 2, and 4, automatically provide the monitorable part of the ROS model for integral, derivative, and mixed causality, respectively. The results show that all the equations in the structural model influenced by the faults considered belong to the sets C_{int}^+ , C_{der}^+ , or C_{mix}^+ , hence, all faults are structurally detectable in any of the three causal interpretations.

TABLE V
BI-ADJACENCY MATRIX FOR THE STRUCTURAL MODEL OF THE REVERSE OSMOSIS SYSTEM. ROWS ARE EQUATIONS, COLUMNS ARE VARIABLES.

	e_3	e_5	e_1	e_{26}	e_{11}	e_{20}	f_{20}	f_7	f_{11}	f_{2127}	f_{30}	e_{16b}	e_{16a}	f_{16a}	e_{15b}	e_{15a}	f_{15a}	f_{14}	e_{23}	e_{24}	e_{25}	f_5	f_{25}	e_7	e_{28}	e_{31}	e_{14}	f_5	f_{25}	\dot{e}_7	\dot{e}_{28}	\dot{e}_{31}	\dot{e}_{14}			
c_1	X																					X														
c_2	X	X	X																					X												
c_3		X																											X							
c_4				X																				X												
c_5				X	X																			X			X									
c_6					X				X																X			X								
c_7						X																			X			X								
c_8						X	X																													
c_9							X	X	X														X													
c_{10}								X																									X			
c_{11}							X			X																									X	
c_{12}									X																									X		
c_{13}									X		X																X									X
c_{14}										X																										X
c_{15}											X	X																X								
c_{16}											X	X																	X							
c_{17}												X	X		X	X													X							
c_{18}													X		X	X												X								
c_{19}							X		X				X				X	X																		
c_{20}																		X	X																	X
c_{21}									X										X					X												
c_{22}																			X	X																
c_{23}																			X	X	X															
c_{24}																				X	X	X														
c_{25}		X																				X												X		
c_{26}			X																																	
c_{27}																																				
c_{28}												X																X								
c_{29}													X																							
c_{30}																							I													
c_{31}																								I												D
c_{32}																									I											D
c_{33}																										I										D
c_{34}																											I									D
c_{35}																												I								D

Regarding structural fault isolability, the single fault isolability matrices were computed using the proposed algorithms. Tables VII, VIII, and IX show the isolability matrices for derivative, integral, and mixed causality, respectively. Columns and rows of the isolability matrix represent the faulty candidates considered. An X in position (i, j) indicates that fault j cannot be isolated from fault i .

Looking at Table VII it is clear that the diagnosis system cannot provide full structural isolability when derivative causality is considered. Moreover, none of the faults can be uniquely isolated, and only two subsets of faults can be structurally isolated from the rest of the faults in the system, i.e., $\mathcal{I}_{der} = \{\{f_{R_{Fp}}, f_{I_{Fp}}, f_{C_{TubRes}}\}, \{f_{R_{Memb}}, f_{C_{Memb}}\}\}$. The situation is similar for integral causality, none of the faults are uniquely structurally isolable, and only two subsets of faults are structurally isolable, i.e. $\mathcal{I}_{int} = \{\{f_{GY_{Rp}}, f_{R_{Rp}}, f_{I_{Rp}}\}, \{f_{R_{Memb}}, f_{C_{Memb}}\}\}$. Finally, for mixed causality, the structural isolability results are improved: faults in R_{Pipe} and R_{Brine} can be uniquely isolated, and the rest of the fault candidates are included within subsets of fault candidates that can be isolated from other faults, i.e. $\mathcal{I}_{mix} = \{\{f_{R_{Pipe}}\}, \{f_{R_{Brine}}\}, \{f_{GY_{Rp}}, f_{R_{Rp}}, f_{I_{Rp}}\}, \{f_{R_{Fp}}, f_{I_{Fp}}, f_{C_{TubRes}}\}, \{f_{R_{Memb}}, f_{C_{Memb}}\}\}$

Thus, similar to the three-tank system, isolability analysis results for the ROS showed that $\mathcal{I}_{der} \prec \mathcal{I}_{mix}$ and $\mathcal{I}_{int} \prec \mathcal{I}_{mix}$, that can be guaranteed because always $C_{int}^+ \subseteq C_{mix}^+$ and $C_{der}^+ \subseteq C_{mix}^+$. Full structural isolability cannot be achieved for this system with its current sensors configuration, and maximum structural isolability can only be achieved when

mixed causality is considered.

VII. DISCUSSION AND CONCLUSIONS

In this paper we have presented a novel way to analyze structural diagnosability properties by analyzing the structural model of the system, under three different causality assumptions: derivative, integral, and mixed causality. We have proposed the theoretical framework and the algorithms to compute the monitorable part for a system model for the three causal interpretations considered, and used this to establish the diagnosability properties of the system. Moreover, using the computations for the monitorable part, we provide the mechanisms to efficiently compute causal matchings for each case.

Several approaches have been proposed in the literature to analyze diagnosability of systems, like the work done by [9], where diagnosability is analyzed after the computation of the complete set of ARRs. The approach presented in this paper provides diagnosability results with different causal interpretations by analyzing the model of the system, and not a previously designed diagnosis system. Other approaches (e.g., [24]) propose canonical decomposition methods that use invertibility information to analyze diagnosability of the system. The main difference with our approach is that in [24] differential constraints can be seen as non invertible constraints if information about the loops is not considered. Hence, this last approach is valid only when integral causality is considered, but not for the derivative and mixed causality cases.

TABLE VII
ISOLABILITY MATRIX FOR THE ROS SYSTEM WHEN DERIVATIVE CAUSALITY IS CONSIDERED.

	fR_{Fp}	fI_{Fp}	fC_{TubRes}	fGY_{Rp}	fR_{Rp}	fI_{Rp}	fR_{Pipe}	fR_{Brine}	fR_{Memb}	fC_{Memb}
fR_{Fp}	X	X	X							
fI_{Fp}	X	X	X							
fC_{TubRes}	X	X	X							
fGY_{Rp}	X	X	X	X	X	X	X	X	X	X
fR_{Rp}	X	X	X	X	X	X	X	X	X	X
fI_{Rp}	X	X	X	X	X	X	X	X	X	X
fR_{Pipe}	X	X	X	X	X	X	X	X	X	X
fR_{Brine}	X	X	X	X	X	X	X	X	X	X
fR_{Memb}									X	X
fC_{Memb}									X	X

TABLE VIII
ISOLABILITY MATRIX FOR THE ROS SYSTEM WHEN INTEGRAL CAUSALITY IS CONSIDERED.

	fR_{Fp}	fI_{Fp}	fC_{TubRes}	fGY_{Rp}	fR_{Rp}	fI_{Rp}	fR_{Pipe}	fR_{Brine}	fR_{Memb}	fC_{Memb}
fR_{Fp}	X	X	X	X	X	X	X	X	X	X
fI_{Fp}	X	X	X	X	X	X	X	X	X	X
fC_{TubRes}	X	X	X	X	X	X	X	X	X	X
fGY_{Rp}				X	X	X				
fR_{Rp}				X	X	X				
fI_{Rp}				X	X	X				
fR_{Pipe}	X	X	X	X	X	X	X	X	X	X
fR_{Brine}	X	X	X	X	X	X	X	X	X	X
fR_{Memb}									X	X
fC_{Memb}									X	X

TABLE IX
ISOLABILITY MATRIX FOR THE ROS SYSTEM WHEN MIXED CAUSALITY IS CONSIDERED.

	fR_{Fp}	fI_{Fp}	fC_{TubRes}	fGY_{Rp}	fR_{Rp}	fI_{Rp}	fR_{Pipe}	fR_{Brine}	fR_{Memb}	fC_{Memb}
fR_{Fp}	X	X	X							
fI_{Fp}	X	X	X							
fC_{TubRes}	X	X	X							
fGY_{Rp}				X	X	X				
fR_{Rp}				X	X	X				
fI_{Rp}				X	X	X				
fR_{Pipe}							X			
fR_{Brine}								X		
fR_{Memb}									X	X
fC_{Memb}									X	X

The primary conclusions from this work are that: (1) analysis of the diagnosability of a system considering different causal interpretations can be efficiently done using just the model of the system; and (2) when computing the monitorable part of the system using mixed causality, we can always guarantee that $C_{int}^+ \subseteq C_{mix}^+$ and $C_{der}^+ \subseteq C_{mix}^+$, which means that mixed causality will always provide equal or better isolability results than the integral or derivative causality approaches. In this paper we considered all algebraic constraints to be invertible. However, this work still needs to investigate two important issues: (1) all nonlinear constraints may not be invertible. Here the work of [12] may be applied, and we believe that this will produce superior diagnosability results; (2) implementation of the numerical derivative of variables. In reality, this will create a trade-off, especially when measurements are noisy. One will likely have to trade-off robustness of the approach to obtain better diagnosability results. This is an open question that we will investigate in future work.

ACKNOWLEDGMENTS

Anibal Bregon and Belarmino Pulido's work has been partially supported by the Spanish MCI DPI2008-01996 and MCI TIN2009-11326 grants. The work by Erik Frisk, Jan Åslund, and Mattias Krysander has been partially supported

by the Swedish Research Council within The Linnaeus Center CADICS. Gautam Biswas' work was supported by NASA NRA NNX07AD12A.

REFERENCES

- [1] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Springer, 2006.
- [2] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, Inc., 1998.
- [3] R. Patton, P. Frank, and R. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. Springer, 2000.
- [4] R. Reiter, "A Theory of Diagnosis from First Principles," *Artificial Intelligence*, vol. 32, pp. 57–95, 1987.
- [5] W. Hamscher, L. Console, and J. de Kleer, *Readings in Model-based Diagnosis*. Morgan-Kaufmann Pub., San Mateo, 1992.
- [6] M. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès, "Conflicts versus Analytical Redundancy Relations: a comparative analysis of the Model-based Diagnosis approach from the Artificial Intelligence and Automatic Control perspectives," *IEEE Trans. Syst. Man Cy. B.*, vol. 34, no. 5, pp. 2163–2177, 2004.
- [7] P. Struss, "Model-based diagnosis for industrial applications," in *Colloquium-Applications of Model-based Reasoning*. Savoy Place, London, United Kingdom: Institute of Electrical Engineers (IEE), 1997.
- [8] O. Dressler and P. Struss, "A toolbox integrating model-based diagnosability analysis and automated generation of diagnostics," in *Proc. of the 14th Intl. WS. on Principles of Diagnosis, DX03*, Washington D.C., USA, 2003.
- [9] L. Travé-Massuyès, T. Escobet, and X. Olive, "Diagnosability analysis based on component supported analytical redundancy relations," *IEEE Trans. Syst. Man Cy. A.*, vol. 36, no. 6, 2006.

- [10] X. Pucel, W. Mayer, and M. Stumptner, "Diagnosability analysis without fault models," in *Proc. of the 20th Intl. WS. on Principles of Diagnosis, DX09*, Stockholm, Sweden, 2009, pp. 67–74.
- [11] M. Krysander and E. Frisk, "Sensor placement for fault diagnosis," *IEEE Trans. Syst. Man Cy. A.*, vol. 38, no. 6, pp. 1398–1410, 2008.
- [12] A. Rosich, E. Frisk, J. Åslund, R. Sarrate, and F. Nejari, "Sensor Placement for Fault Diagnosis Based On Causal Computations," in *Proc. of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, SAFEPROCESS09*, Barcelona, Spain, 2009, pp. 402–407.
- [13] P. Bonasso, D. Kortenkamp, and C. Thronesbery, "Intelligent control of a water recovery system: Three years in the trenches," *AI Magazine*, vol. 24, no. 1, pp. 19–43, 2003.
- [14] K. D. Pickering, K. R. Wines, G. M. Pariani, L. A. Franks, J. Yeh, B. W. Finger, M. L. Campbell, C. E. Verostko, C. Carrier, J. C. Gandhi, and L. M. Vega, "Early results of an integrated water recovery system test," in *Proc. 29th Int. Conf. Environmental Systems*, Colorado Springs, Colorado, 2004.
- [15] P. Fritzon, *Principles of object oriented modeling and simulation with Modelica 2.1*. Wiley, 2004.
- [16] B. Pulido and C. Alonso, "Possible Conflicts: a compilation technique for consistency-based diagnosis," *IEEE Trans. Syst. Man Cy. B.*, vol. 34, no. 5, pp. 2192–2206, Oct. 2004.
- [17] M. Chantler, S. Daus, T. Vikatos, and G. Coghill, "The use of quantitative dynamic models and dependency recording engines," in *Proc. of the 7th Intl. WS. on Principles of Diagnosis, DX96*, Val Morin, Quebec, Canada, 1996, pp. 59–68.
- [18] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, ser. Classics in Applied Mathematics. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1996, vol. 14, revised and corrected reprint of the 1989 original.
- [19] O. Dressler, "On-Line Diagnosis and Monitoring of Dynamic Systems based on Qualitative Models and Dependency-recording Diagnosis Engines," in *European Conf. on Artificial Intelligence, ECAI96*, 1996, pp. 481–485.
- [20] B. Pulido, A. Bregon, and C. Alonso-Gonzalez, "Analyzing the influence of differential constraints in Possible Conflict and ARR computation," in *Current Topics in Artificial Intelligence, CAEPIA 2009 Selected Papers. P. Meseguer, L. Mandow, R. M. Gasca Eds.* Springer-Verlag Berlin, 2009.
- [21] J. Cassar and M. Staroswiecki, "A structural approach for the design of failure detection and identification systems," in *Proceedings of IFAC Control of Industrial Systems*, Belfort, France, 1997.
- [22] M. Krysander, J. Åslund, and M. Nyberg, "An Efficient Algorithm for Finding Minimal Over-constrained Sub-systems for Model-based Diagnosis," *IEEE Trans. Syst. Man Cy. A.*, vol. 38, no. 1, pp. 197–206, 2008.
- [23] A. L. Dulmage and N. S. Mendelsohn, "Coverings of bipartite graphs," *Canadian Journal of Mathematics*, vol. 10, pp. 517–534, 1958.
- [24] V. Flaugergues, V. Cocquempot, M. Bayart, and M. Pengov, "Structural Analysis for FDI: a modified, invertibility-based canonical decomposition," in *Proc. of the 20th Intl. WS. on Principles of Diagnosis, DX09*, Stockholm, Sweden, 2009, pp. 59–66.
- [25] F. Harary, *Graph theory*. Addison-Wesley, 1969.
- [26] N. Porté, S. Boucheron, J. Sallantin, and F. Arlabosse, "An algorithmic view at causal ordering," Centre de Recherche en Informatique, Technical report 45, 1988.
- [27] K. Murota, *Matrices and Matroids for System Analysis*. Springer-Verlag, 2000.
- [28] C. Svärd and M. Nyberg, "Residual generators for fault diagnosis using computation sequences with mixed causality applied to automotive systems," *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, vol. 40, no. 6, pp. 1310–1328, 2010.
- [29] M. Staroswiecki and G. Comtet-Varga, "Analytical Redundancy Relations for Fault Detection and Isolation in Algebraic Dynamic Systems," *Automatica*, vol. 37, no. 5, pp. 687–699, 2001.
- [30] D. Cox, J. Little, and D. O'Shea, *Ideals, varieties, and algorithms*, 2nd ed. Springer, 1996.
- [31] C. De Persis and A. Isidori, "A geometric approach to nonlinear fault detection and isolation," *IEEE Transactions on Automatic Control*, vol. 46, no. 6, pp. 853–865, 2001.
- [32] T. Wei and M. Li, "High order numerical derivatives for one-dimensional scattered noisy data," *Applied Mathematics and Computation*, vol. 175, pp. 1744–1759, 2006.
- [33] M. Johansson and J. Kingstedt, "Methods for residual generation using mixed causality in model based diagnosis," Master's thesis, Linköping University, SE-581 83 Linköping, 2008.
- [34] E. Höckerdal, L. Eriksson, and E. Frisk, "Air mass-flow measurement and estimation in diesel engines equipped with EGR and VGT," *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.*, vol. 1, no. 1, pp. 393–402, 2008.
- [35] E. Höckerdal, E. Frisk, and L. Eriksson, "Observer design and model augmentation for bias compensation with a truck engine application," *Control Engineering Practice*, vol. 17, no. 3, pp. 408–417, 2009.
- [36] B. Duffield and A. Hanford, "Advanced life support requirements document," NASA-Lyndon B. Johnson Space Center, Houston, TX, Tech. Rep. JSC-38571, Rev. B, Tech. Rep., September 2002.
- [37] I. Roychoudhury, G. Biswas, and X. Koutsoukos, "Designing Distributed Diagnosers for Complex Continuous Systems," *IEEE Transactions on Automated Science and Engineering*, vol. 6, no. 2, pp. 277–290, 2009.
- [38] J. Broenink, "Introduction to Physical Systems Modelling with Bond Graphs," *StE Whitebook on Simulation methodologies*, 1999.
- [39] G. Biswas, E. Manders, J. Ramirez, N. Mahadevan, and S. Abdelwahed, "Online model-based diagnosis to support autonomous operation of an advanced life support system," *Habitation - Intl. Journal for Human Support Research*, vol. 10, no. 1, pp. 21–38, 2004.
- [40] A. Bregon, "Integration of FDI and DX Techniques within Consistency-based Diagnosis with Posibles Conflicts," Ph.D. dissertation, University of Valladolid, Valladolid, Spain, 2010.