Diagnosability Analysis Considering Causal Interpretations for Differential Constraints

Erik Frisk[1], **Anibal Bregon**[2], **Jan Åslund**[1],
Mattias Krysander[1], **Belarmino Pulido**[2], and **Gautam Biswas**[3]

[1] *Department of Electrical Engineering, Linköping University, 581 83 Linköping, Sweden*
{frisk,jaasl,matkr}@isy.liu.se
[2] *Department of Computer Science, University of Valladolid, 47011 Valladolid, Spain*
{anibal,belar}@infor.uva.es
[3] *Dept. of EECS/ISIS, Vanderbilt University, Nashville, TN 37235, USA*
gautam.biswas@vanderbilt.edu

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 ARRs. 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 case study. Finally, we summarize our work and provide a discussion of the advantages of mixed causality over just derivative or just integral causality.

1 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 FDI (Blanke *et al.*, 2006) and the DX (Reiter, 1987) communities in the last three decades. The two communities have developed different algorithms that are being proved to be complementary (Cordier *et al.*, 2004). The main advantage of the Model-based approach is inherent to using models: re-usability. In more detail, both the DX and FDI approaches use the redundancy in the system description and measurements to develop a set of residuals, which have been termed as conflicts (Reiter, 1987), ARRs (Cassar and Staroswiecki, 1997), Possible Conflicts (Pulido and Alonso, 2004) and Structurally Overdetermined sets (Krysander *et al.*, 2008) by different researchers.

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 (Dressler and Struss, 2003). Recently, the process has been carried out pre-computing the whole set of existing ARRs for a given set of sensors, and analyzing their discriminability properties (Travé-Massuyès *et al.*, 2006). Such approaches are infeasible for large systems. More recently, (Krysander and Frisk, 2008; Rosich *et al.*, 2009) have explored alternative and more efficient computational ways for diagnosability analysis.

This work extends the more recent work by studying the diagnosability properties given a system model and a fixed set of sensors when different causal interpretations are considered for differential constraints. 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 different 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. Finally, we compare the different diagnosability capabilities for integral, derivative and mixed causality approaches.[1]

We have tested these concepts and algorithms in a case study, a three-tank system, proving that the proposed algorithms can be used to analyze diagnosability of a 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 case study used to illustrate main concepts. Afterwards, the theoretical background for the basic concepts used in the proposal is provided. Later on, we analyze the diagnosability properties of a system model under different causal

[1] A Matlab implementation is available at `http://www.fs.isy.liu.se/Software/CausalIsolability/`.

interpretations for differential constraints. Finally, we present and discuss results obtained using the approach on the case study, and draw some conclusions.

2 PROBLEM FORMULATION

We use a simple three tank system model (Figure 1) to introduce the problem and formulate the different classes of residual generators that we discuss in the paper. The three tank system model is represented by the set of equations

$$
\begin{aligned}
&c_1: q_1 = \frac{1}{R_{V1}}(p_1 - p_2) && c_7: y_1 = p_1\\
&c_2: q_2 = \frac{1}{R_{V2}}(p_2 - p_3) && c_8: y_2 = q_2\\
&c_3: q_3 = \frac{1}{R_{V3}}(p_3) && c_9: y_3 = q_0\\
&c_4: \dot{p}_1 = \frac{1}{C_1}(q_0 - q_1) && c_{10}: \dot{p}_1 = \frac{dp_1}{dt}\\
&c_5: \dot{p}_2 = \frac{1}{C_2}(q_1 - q_2) && c_{11}: \dot{p}_2 = \frac{dp_2}{dt}\\
&c_6: \dot{p}_3 = \frac{1}{C_3}(q_2 - q_3) && c_{12}: \dot{p}_3 = \frac{dp_3}{dt}
\end{aligned}
$$

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_i the capacitance of tank i. Three sensors y_1, y_2, and y_3, measure p_1, q_2, and q_0, respectively.

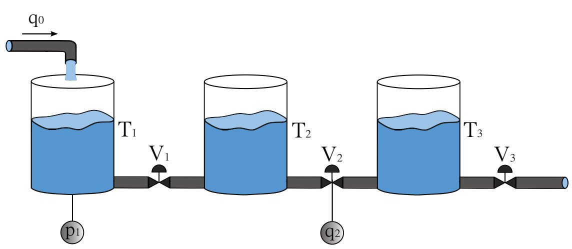

Figure 1: Diagram of the three-tank system.

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 numerical solvers. This assumption is realistic since commercial packages for simulating differential-algebraic equations, e.g. Dymola, successfully use similar techniques to solve large dynamical models (Fritzson, 2004).

A main concern is how to handle the dynamics in the model. For this class of residual generators, in the literature there have typically been two options; either integral causality or differential causality (Chantler *et al.*, 1996; Blanke *et al.*, 2006; Pulido and Alonso, 2004; Travé-Massuyès *et al.*, 2006). 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 (Brenan *et al.*, 1996). 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 tank model and assume differential causality. Then 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 numerical differentiation using equations c_{10} and c_{11}.

$$
\begin{aligned}
&p_1 := y_1, \quad \dot{p}_1 := \frac{d}{dt}p_1, \quad q_2 := y_2\\
&q_0 := y_3, \quad q_1 := q_0 - C_1\dot{p}_1\\
&p_2 := p_1 - R_{V1}q_1, \quad \dot{p}_2 = \frac{d}{dt}p_2
\end{aligned}
$$

A residual can then be computed using equation c_5 as $r := C_2\dot{p}_2 - q_1 + q_2$. The variables are computed in a sequential way using numerical differentiation where needed and in this case no algebraic loops need to be solved.

Integral causality works in a similar way and as another example, compute q_2 and q_0 using c_8 and c_9 as

$$q_2 := y_2, \quad q_0 := y_3$$

and then solve for q_1, p_1, p_2, $\dot{p}_1$, and $\dot{p}_2$ using the set of equation c_1, c_4, c_5, c_{10}, c_{11}:

$$
\begin{aligned}
&q_1 := \frac{1}{R_{V1}}(p_1 - p_2), \quad p_1 := \int_0^t \dot{p}_1 d\tau\\
&\dot{p}_1 := \frac{1}{C_1}(q_0 - q_1), \quad p_2 := \int_0^t \dot{p}_2 d\tau\\
&\dot{p}_2 := \frac{1}{C_2}(q_1 - q_2)
\end{aligned}
$$

Note that this is a differential loop that has to be solved numerically, which can be done with any ODE-solver technique. Loops are broken by integrators, and sometimes this is referred to as a spiral (Dressler and Struss, 1996). Finally, a residual can be computed using equation c_7 as $r := y_1 - p_1$.

These two examples illustrate the main principle of sequential residual generation with the differential 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. However, a similar loop cannot be solved with a differential causality assumption. This is because the loop corresponds to a differential equation which can not be solved by only differentiating variables.

As will be demonstrated in Section 5, there are 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 diagnosability properties for a given model. The problem studied in this paper is thus to derive efficient algorithms, which are not based on the set of ARRs/MSOs derived using integral, derivative, or mixed causality interpretation, and then use this model to determine the diagnosability properties of the system.

3 THEORETICAL BACKGROUND

This section recapitulates some basic formalism, concepts, and notation needed to describe the theoretical developments in Section 4.

3.1 Graph Representation of the Model

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

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

where $z \in \mathbb{R}^{n_z}$ is the vector of known variables, $x_1 \in \mathbb{R}^{n_1}$ the vector of unknown dynamic variables, and $x_2 \in \mathbb{R}^{n_2}$ 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 variable $x_{1,i} \in x_1$, constraints capturing the dynamics

$$\dot{x}_{1,i} = \frac{d}{dt} x_{1,i}, \quad i = 1, \ldots, n_1 \tag{2}$$

The constraints in (1) are algebraic and systems 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 numerical problems. The disadvantage is that the structural results may not be as precise as the analytical results. However, analytic results are computationally expensive, and they are often not possible to derive for nonlinear systems 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* $C = \{c_1, \ldots, c_m\}$ *is the set of constraints,* $X = \{x_1, \dot{x}_1, x_2\}$ *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* $c_i \in C$*.*

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 $\dot{x}_i$ in the differential constraints (2), E_I the non-differentiated variables x_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 1, where X, D, and I indicates edges in E_X, E_D, and E_I, respectively.

	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

Table 1: Bi-adjacency matrix for the structural model of the three-tank system.

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, 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 $E_1 \subset E_2$.

3.2 The Dulmage-Mendelsohn Decomposition

A key tool when analyzing structural models is the Dulmage-Mendelsohn decomposition (Dulmage and Mendelsohn, 1958), used in diagnosis in e.g. (Blanke *et al.*, 2006; Krysander and Frisk, 2008; Flaugergues *et al.*, 2009). The general Dulmage-Mendelsohn decomposition is illustrated in Figure 2 where, by a suitable reordering of constraints and variables, the bi-adjacency matrix is converted to a triangular form. The sub-graph

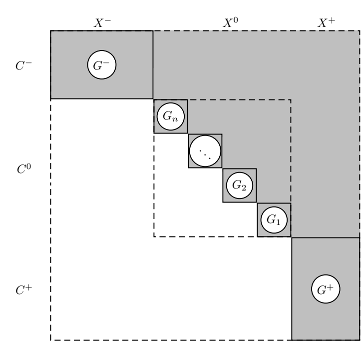

Figure 2: Dulmage-Mendelsohn decomposition.

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. It is the overdetermined part that contains redundancy and can therefore be used for diagnosis. In the exactly determined parts there is a finer structure of Hall-components, here denoted G_i, which will be explained later. 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* (Harary, 1969). 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. A Dulmage-Mendelsohn decomposition characterizes all matchings for a given model and, for example, a matching of variables X^0 is given by the diagonal of the exactly determined part in Figure 2.

3.3 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 in Section 4 be extended 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. Then from (Blanke *et al.*, 2006; Krysander and Frisk, 2008):

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

$$c_f \in C^+$$

Following the ideas in (Krysander and Frisk, 2008), 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*

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

4 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.

4.1 Basic Definitions

As discussed in Section 2, 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 Figure 2, 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 Figure 2, 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 2, 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 for simultaneously. If the Hall component includes an integration, the computational loop is naturally broken, 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 differentiation 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*
- $\Gamma \cap E_D \neq \emptyset$ *implies that* $|\Gamma| = 1$.

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.*

Using the definition in Section 3 that the monitorable part of a model is its overdetermined part, i.e., C^+ in Figure 2, 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 in $H(\Gamma)$ are structurally solvable under the causal assumption.*

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.*

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 diagnosability properties of a given model.

4.2 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 2. Integral, derivative, and mixed causality constraints will be treated separately.

Integral causality

To compute C^+_{causal} under integral causality, the algorithm in (Flaugergues *et al.*, 2009) can be used. An algorithm description is included here, which is equivalent to the one in (Flaugergues *et al.*, 2009), using the notation introduced in Section 3.

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 $C(G)$

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}.

Derivative causality

For the derivative causality case, note that the algorithm from (Flaugergues *et al.*, 2009) can *not* 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, 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
  $G_{nc} := G - X_c$;
  $G_{ni} := G_{nc} - C(G_{nc}, E_I)$;
  $X_c := X_c \cup X(G^+_{ni} \cup G^0_{ni})$;
until $X(G^+_{ni} \cup G^0_{ni}) = \emptyset$;
$G := G - C(G, X \setminus X_c)$;
return $C(G^+)$

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

Proof. Consider C^+_{der} and let the corresponding subgraph 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 $H(\Gamma) = \{G_1, \ldots, G_n\}$, 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

$$E(G_k) \cap E_I = \emptyset \tag{3}$$

Assume that the Hall components are enumerated as in Figure 2 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

$$G_1 = G^+_1 \cup G^0_1 \subset G^+_{ni} \cup G^0_{ni}$$

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

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

and it follows that

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

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 $G^+_{ni} \cup G^0_{ni}$; see Figure 3.

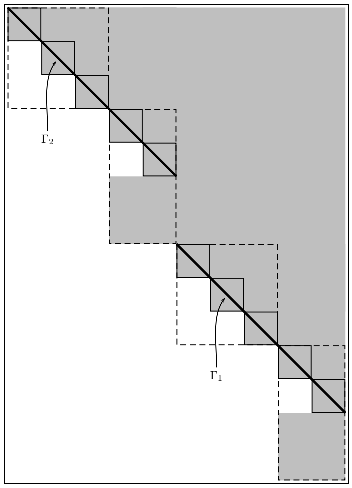

Figure 3: Causal Matching Γ_k in iteration k.

Let $H(\Gamma_k) = \{G_{k1}, \ldots, G_{kn_k}\}$ denote the induced Hall components. All Hall components G_{kj} fulfills 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}.

Mixed causality

The mixed causality case is treated by first considering an exactly determined model and proving, in a constructive manner, that there *always* exist 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 3.2. Let the set of admissible edges $A(G)$ be defined as

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

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 Γ' be any perfect matching of G;
$\Gamma := \Gamma \cup \Gamma'$;
return Γ

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 Γ fulfills the conditions in Definition 6 for all Hall components in G.*

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 $\{G_1, \ldots, G_n\}$. After the reduction each Hall component G_k has a structure similar to the one illustrated in Figure 4. Let the Hall components in the reduced Hall component G'_k be denoted by $H_k = \{G_{k1}, G_{k2}, \ldots, \}$. 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

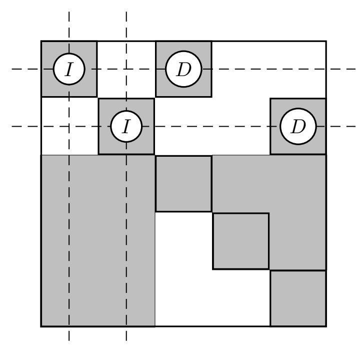

Figure 4: Reduced Hall component G'_k

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

The matching Γ obtained by the algorithm can be partitioned into sets Γ_k, $k = 1, \ldots, n$, 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^+ fulfills Definition 8 for a mixed causality assumption.*

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 $C^+_{mix} = C(G^+)$. $\square$

The result can be summarized in the algorithm below.

Algorithm 4: Compute C^+_{mix}

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

5 CASE STUDY: A THREE-TANK SYSTEM

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.

5.1 Structural Detectability Analysis

Algorithms 1, 2, and 4, automatically provide 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 detectability when any of the three interpretations is considered.

5.2 Structural Isolability Analysis

To illustrate single fault isolability properties of the model, we computed the isolability matrices for each one of the causal interpretations considered. Tables 2, 3, and 4 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, i.e., $\mathcal{I}_{der} = \{C_{T_2}\}, \{R_{V_2}, R_{V_3}, C_{T_3}\}$. Integral causality provide 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, i.e., $\mathcal{I}_{int} = \{C_{T_1}\}, \{C_{T_2}\}, \{R_{V_2}, R_{V_3}, C_{T_3}\}$. Finally, for mixed causality, provides the best results for isolability: faults in R_{V_1}, C_{T_1}, C_{T_2} can be isolated from the other faults in the system, i.e., $\mathcal{I}_{mix} = \{R_{V_1}\}, \{C_{T_1}\}, \{C_{T_2}\}, \{R_{V_2}, R_{V_3}, C_{T_3}\}$.

Thus, for this case study, isolability analysis results show that $\mathcal{I}_{der} \prec \mathcal{I}_{int} \prec \mathcal{I}_{mix}$ but in general, only $\mathcal{I}_{der} \prec \mathcal{I}_{mix}$ and $\mathcal{I}_{int} \prec \mathcal{I}_{mix}$ can be guaranteed because always $C^+_{int} \subseteq C^+_{mix}$ and $C^+_{der} \subseteq C^+_{mix}$. Maximum isolability cannot be achieved in this system using

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

Table 2: Isolability matrix for the three-tank system when derivative causality is considered.

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

Table 3: Isolability matrix for the three-tank system when integral causality is considered.

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

Table 4: Isolability matrix for the three-tank system when mixed causality is considered.

derivative or integral causality. When mixed causality is considered, additional residuals are created from the system models, and this improves the isolability. Fig. 5 shows a residual obtained for mixed causality. Looking at the differential constraints in the residual, e_{10} uses derivative causality, while both e_{11} and e_{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 e_1). This residual was obtained from the causal matching automatically provided by Algorithm 3 when mixed causality is considered.

6 DISCUSSION AND CONCLUSIONS

In this paper we have presented a new framework for analyzing the diagnosability of a system using differential, integral, and mixed causality. We have presented a novel way to analyze diagnosability properties by analyzing the structural model of the system. 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 by (Travé-Massuyès *et al.*, 2006), 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., (Flaugergues *et al.*, 2009)) propose canonical decomposition methods that use invertibil-

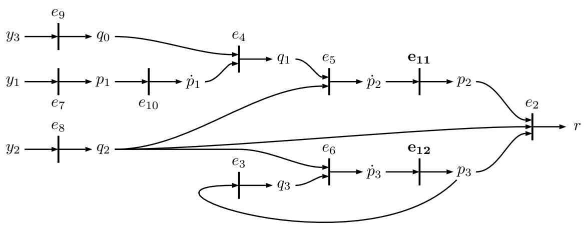

Figure 5: Additional residual obtained when mixed causality is considered.

ity information to analyze diagnosability of the system. The main difference with our approach is that in (Flaugergues *et al.*, 2009) differential constraints can be seen as non invertible constraints if information about the loops is not considered, hence, this approach is valid only when integral causality is considered, but not for the derivative and mixed causality cases.

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 (Rosich *et al.*, 2009) may be applied, and we believe that this will produce superior diagnosability results; (2) computation 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 JCyL VA005B09 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.

REFERENCES

(Blanke *et al.*, 2006) M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki. *Diagnosis and Fault-Tolerant Control*. Springer, 2006.

(Brenan *et al.*, 1996) K. E. Brenan, S. L. Campbell, and L. R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*, volume 14 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996. Revised and corrected reprint of the 1989 original.

(Cassar and Staroswiecki, 1997) J.P. Cassar and M. Staroswiecki. A structural approach for the design of failure detection and identification systems. In *In Prococeedings of IFAC Control of Industrial Systems*, Belfort, France, 1997.

(Chantler *et al.*, 1996) M. Chantler, S. Daus, T. Vikatos, and G. Coghill. The use of quantitative dynamic models and dependency recording engines. In *Proceedings of the Seventh International Workshop on Principles of Diagnosis, DX96*, pages 59–68, Val Morin, Quebec, Canada, 1996.

(Cordier *et al.*, 2004) M.O. 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.*, 34(5):2163–2177, 2004.

(Dressler and Struss, 1996) O. Dressler and P. Struss. The Consistency-based approach to automated diagnosis of devices. In Gerhard Brewka, editor, *Principles of Knowledge Representation*, pages 269–314. CSLI Publications, Standford, 1996.

(Dressler and Struss, 2003) O. Dressler and P. Struss. A toolbox integrating model-based diagnosability analysis and automated generation of diagnostics. In *Proceedings of the 14th International Workshop on Principles of Diagnosis, DX03*, Washington D.C., USA, 2003.

(Dulmage and Mendelsohn, 1958) A. L. Dulmage and N. S. Mendelsohn. Coverings of bipartite graphs. *Canadian Journal of Mathematics*, 10:517–534, 1958.

(Flaugergues *et al.*, 2009) V. Flaugergues, V. Cocquempot, M. Bayart, and M. Pengov. Structural Analysis for FDI: a modified, invertibility-based canonical decomposition. In *Proceedings of the 20th International Workshop on Principles of Diagnosis, DX09*, pages 59–66, Stockholm, Sweden, 2009.

(Fritzson, 2004) Peter Fritzson. *Principles of object oriented modeling and simulation with modelica 2.1*. Wiley, 2004.

(Harary, 1969) F. Harary. *Graph theory*. Addison-Wesley, 1969.

(Krysander and Frisk, 2008) M. Krysander and E. Frisk. Sensor placement for fault diagnosis. *IEEE Trans. Syst. Man Cy. A.*, 38(6):1398–1410, 2008.

(Krysander *et al.*, 2008) 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.*, 38(1):197–206, 2008.

(Pulido and Alonso, 2004) B. Pulido and C. Alonso. Possible Conflicts: a compilation technique for consistency-based diagnosis. *IEEE Trans. Syst. Man Cy. B.*, 34(5):2192–2206, Oct. 2004.

(Reiter, 1987) R. Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 32:57–95, 1987.

(Rosich *et al.*, 2009) A. Rosich, E. Frisk, J. Åslund, R. Sarrate, and F. Nejjari. Sensor Placement for Fault Diagnosis Based On Causal Computations. In *Proceedings of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, SAFEPROCESS09*, pages 402–407, Barcelona, Spain, 2009.

(Travé-Massuyès *et al.*, 2006) L. Travé-Massuyès, T. Escobet, and X. Olive. Diagnosability analysis based on component supported analytical redundancy relations. *IEEE Trans. Syst. Man Cy. A.*, 36(6), 2006.