Dimensional Analysis Based Causal Ordering

Qiang Shen† Taoxin Peng† and Robert Milne‡

†: School of Artificial Intelligence, University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN

‡: Intelligent Applications Ltd. 1 Michaelson Square
Livingston, West Lothian EH54 7DP

email: †{qiangs, taoxinp}@dai.ed.ac.uk, ‡rmilne@bcs.org.uk

Abstract

This paper presents a novel approach for generating causal dependencies between system variables, from an acausal description of the system behaviour, *and* for identifying the end causal impact, in terms of whether a change in the value of an influencing variable will lead to an increase or a decrease in the value of the influenced variables. This work is based on the use of the conventional method for dimensional analysis developed in classical physics. The utility of the work is demonstrated with its application to providing causal explanations for selected physical systems. The results reflect well the common-sense understanding of the causality in these systems.

Introduction

Knowledge of causality is essential in handling many application problems. In model-based diagnosis, for example, when one or more variables are observed to have an abnormal value, it is necessary to find out what internal variables (which are perhaps unmeasurable) could have been causing the observed abnormalities, what other variables might be affected, and what other observations would be spuriously correlated to the observations of these variables. An explicit expression of the causal relations amongst system variables will enable the diagnostic system to generate convincing explanations for its findings.

Having recognised the great application potential, significant work has been developed for deriving a causal ordering between variables in a physical system (de Kleer & Brown 1984; Iwasaki & Simon 1986; 1994; Top & Akkermans 1991; Lee & Compton 1994; Travé-Massuyés & Pons 1997). However, most of these existing approaches only address the issue of what variables may affect what other variables, without identifying the directions of causal effects in terms of whether an increase in the value of an influencing variable will lead to an increase or a decrease in the value of the influenced variables. Also, some limitations remain when attempting to utilise the existing causal ordering techniques to draw inferences on the causal relations among variables in a dynamic system, e.g., requiring system models being self-contained (Iwasaki & Simon 1994; Travé-Massuyés & Pons 1997). Therefore, a technique which may reduce such restrictions is clearly very desirable.

To better understand causal relations between system variables, it is useful to investigate what represents the fundamental features of physical quantities and their relationships. When describing the behaviour of a physical variable, the most distinct characteristic is its physical dimensions. For this reason, dimensional analysis (Buckingham 1914; Huntley 1952) has long been serving as a basis upon which to create quantitative system models and/or to perform model-based simulation in classical physics and control engineering. Recently, it also has been applied to developing a method for qualitative reasoning (Bhaskar & Nigam 1990), though this method does not establish an explicit causal ordering among system variables. Inspired by this observation, this paper presents a novel approach for analysing causality in physical systems by means of dimensional analysis. Given the structural and behavioural description of the system considered, and the knowledge of the dimensions of the system variables, the proposed technique can be applied both to produce a causal ordering amongst the variables and to provide an identification for the directions of causal effects on the changes of variable values.

The rest of this paper is structured as follows. Theoretical foundations are introduced in the next section, showing the basis upon which to develop the present work. The causal ordering algorithm based on the dimensional analysis is then described in section 3. Illustrative examples of this method are given in section 4. The paper is concluded and further work pointed out in section 5.

Underlying Theories

Dimensional Analysis

Informally, dimensional analysis (Buckingham 1914; Huntley 1952) is a method by which information about a physical variable can be deduced from that on cer-

tain other variables, given a dimensionally consistent relationship between them. Based on the laws of motion formulated by Newton, three units are regarded as fundamental, namely, *length* (L), *mass* (M) and *time* (T) (Huntley 1952). These are termed *base units* in this paper. Any other physical unit is regarded as a *derived unit*, since it can be represented by a combination of these base units. Each base unit represents a *dimension*. For instance, the units of velocity and acceleration are derived ones and have two dimensions because they are defined by reference to two of the base units - length and time. The units of force, momentum and power are a composite of all the three base units, and therefore have three dimensions. In what follows, a variable whose unit is a base unit is called a *base variable*, otherwise the variable is called a *derived variable*. In this paper, the mass (MLT) system is adopted for the expression of physical dimensions. However, it is not the unique option. For instance, the study of electricity and magnetism has shown the value of including other dimensions as base units.

The present work makes use of the following three most basic properties of dimensional analysis.

- *Principle of dimensional homogeneity.* Given that $y = \sum_i a_i f_i(x_i)$ represents a physical law governing the behaviour of a certain system, with y and x_i being the system variables or their temporal derivatives and a_i being the corresponding parameters, all the $a_i f_i(x_i)$ must have the same dimensions as y.
- *Product theorem.* If the value of a variable can be derived from measurements of given base variables $u, v, w, ...$, that value can then be written in the form of $Cu^{\alpha}v^{\beta}w^{\gamma}$ ..., where $C, \alpha, \beta, \gamma, ...$ are constants.
- *Buckingham's* Π*-theorem.* Given physical quantities $u, v, w, ...$ such that $f(u, v, w, ...) = 0$ is a complete equation (that reflects the underlying physical laws characterising the physical situation), then its solution can be written in the form $F(\Pi_1, \Pi_2, ..., \Pi_{n-r}) = 0$, where n is the number of arguments of f and r is the number of basic dimensions needed to express the variables $u, v, w, ...$; for all i, Π_i is a dimensionless number.

x_i having the same dimensions with y means the following: the exponents of the three base units (dimensions) that make up these two variables must be the same.

From the product theory, given the use of the mass system, the dimensions of any physical variable can be represented in the general form of $M^{\alpha}L^{\beta}T^{\gamma}$. Such a representation of a variable is referred to as the dimensional representation of that variable. Following this representation, for example, the physical quantity force can be dimensionally represented by MLT^{-2}, pressure by $ML^{-1}T^{-2}$ and velocity by LT^{-1}.

In general, a dimensionless product Π can be expressed as follows:

$$\Pi_i = y_i \times (x_1^{\alpha_{i1}} ... x_r^{\alpha_{ir}})$$

where $x_1, ..., x_r$ are termed the repeating variables, $y_1, .., y_r$ are termed the performance variables and $\{\alpha_{ij} | 1 \leq i \leq n-r, 1 \leq j \leq r\}$ are the exponents.

Π-calculus

The Π-calculus (Bhaskar & Nigam 1990) is developed mainly for the purpose of reasoning about physical devices, processes and systems. It is especially useful for deriving the causal structure of the device's behaviour, given the input and output of a device, by means of partial derivatives. This technique is briefly summarised below. For simplicity, each dimensionless number, Π_i is called a *regime*, and a collection of regimes is called an *ensemble* (typically, in modelling physical systems, an ensemble can be seen as a distinct component). If x_k is a variable that occurs in both regimes Π_i and Π_j, then the x_k will be referred to as a *contact variable* between these two dimensional products. The set of variables x_j, $1 \leq j \leq r$ that repeat in each regime is called the *basis* of an ensemble. The size of the base of an ensemble, r can be determined directly by the number of base units that are involved in the dimensional representations of variables in that ensemble.

The analyses can be divided into three levels based on the relationships among the dimensionless products.

1. *Intra-regime analysis.* The analysis is within a regime and provides the following method for calculating partial derivatives:
$$\partial y_i / \partial x_j = -(\alpha_{ij} y_i)/x_j$$
2. *Inter-regime analysis.* The analysis is across regimes and gives the following method for calculating partial derivatives:
$$[\partial y_i / \partial y_j]^{x_p} = (\alpha_{ip}/\alpha_{jp})(y_i/y_j)$$
where x_p is a contact variable for regimes Π_i and Π_j.
3. *Inter-ensemble analysis.* The analysis is across ensembles and extends the inter-regime analysis to calculate inter-ensemble partial derivatives. However, no explicit formula for such calculation is generally available.

When the sign of a partial derivative, $\partial y_i / \partial x_j$ (or $[\partial y_i / \partial y_j]^{x_p}$), is obtained, the causal effect between ∂y_i and ∂x_j can then be inferred. For instance, if $\partial y_i / \partial x_j > 0$, then an increase in x_j will lead to an increase in y_i; if $\partial y_i / \partial x_j < 0$, then an increase in x_j will lead to a decrease in y_i. However, the calculation of the inter-ensemble partials is not straightforward. It will need some knowledge about connections between components, since an ensemble (of regimes) comes from a component in a device. This issue will be discussed further in the next section.

The Proposed Approach

Basic Notations

To start with, the dimensional representation of a variable x is hereafter denoted by $D(x)$. Therefore, for force f, pressure p and velocity v, the following holds: $D(f) = MLT^{-2}$, $D(p) = ML^{-1}T^{-2}$ and $D(v) = LT^{-1}$.

Definition 1. Two physical variables x_1 and x_2 are *equivalent* to each other if and only if they have the same dimensional representation, i.e. $D(x_1) = D(x_2)$.

This definition denotes the unique dimensional representation of physical variables within a given system.

Definition 2. A system variable is regarded to be *exogenous* if it is controlled only by factors external to the system; other variables are called non-exogenous. An equation that indicates a variable to be exogenous is called an *exogenous equation*; other equations are termed non-exogenous equations.

This definition includes the understanding that if a variable is exogenous, its derivative is treated as exogenous as well. Exogenous variables are determined during the modelling process. They jointly set up the scope of the physical system being modelled, separating it from the rest of the world. Such variables play an important role in causal ordering and, indeed, in any system modelling approaches.

In systems modelling, variables are normally represented as a function of time. Hence, the base unit *time* (T) is regarded more fundamental than the other two base units *length* (L) and *mass* (M). The following definition reflects this understanding.

Definition 3. For a given variable, x, the number of different base units appearing in its dimensional representation, excluding the dimension *time* if it has negative power, is called the *degree of commitment* of the variable, denoted as $cd(x)$; the algebraic sum of the exponents of all the base units involved is called the *degree of factorisation* of the variable, denoted as $fd(x)$; and the algebraic sum of the exponents of the base units, excluding that of dimension *time* if it has negative power is called the *degree of factorisation excluding negative time* of the variable, denoted as $fd\_t(x)$.

For example, given the dimensional representation of force $D(f) = MLT^{-2}$, the degree of commitment of force $cd(f) = 2$, the degree of factorisation $fd(f) = 1+1+(-2) = 0$ and the degree of factorisation excluding negative time $fd\_t(f) = 1 + 1 = 2$.

A functional relation between two variables which is represented by an equation is reversible (or symmetric as referred to in (Iwasaki & Simon 1994)), if x is expressed as a function of y and y can also be expressed as a function of x. However, causal relations are irreversible, i.e., that x causes y clearly does not imply that y causes x. The purpose of causal ordering is to find the causal relations among the variables in a given model, which will convert a set of reversible equations into a set of irreversible constraints amongst the system variables.

The question is, given a set of system variables, how to determine which variable may have a higher degree of freedom to change its value? A variable is of a higher degree of freedom if it is more independent of, or less dependent upon the change of values of other system variables. Intuitively, variables with a lower cd and/or a lower fd_t value appear to be of a higher freedom degree and hence more independent, unless otherwise specified. For example, suppose that a given system model includes the following two variables, force f and velocity v, with $D(f) = MLT^{-2}$ and $D(v) = LT^{-1}$. This dictates that $cd(f) = 2$, $fd\_t(f) = 2$ and $cd(v) = 1$, $fd\_t(v) = 1$. If these two variables appear in the same equation and none of them is regarded as special (e.g. exogenous), then using the above heuristic, f is treated as more dependent than v. A natural deduction of this is that *force* can be regarded as depending on *velocity* at any given time instance, but not vice versa. This agrees with commonsense understanding of the physical relation between these two physical quantities. The conditions mentioned here are very important. Otherwise, the use of the heuristic may result in counter-intuitive conclusions. For instance, consider a simple, autonomous pendulum, without taking into account any special constraints, it may be concluded that the gravitational force would depend on the bob's velocity. This is not true because the gravitational force should have been initialised as an exogenous variable. Given these two variables appearing in the same equation, the correct explanation should be therefore that the bob's velocity depends on the gravitational force.

In addition, variables that have a dimension of *time* with negative power seem to be more independent, and the higher the negative exponent of the time dimension that a variable has, the less reliant its current value is upon the current values of other variables. In dynamic systems, this implies that, given two variables of different amounts of power on the time dimension and both being involved within one equation with the same cd and fd_t, the change of value of the variable with a more negative time exponent tends to occur before that of the other variable. For example, the change of physical quantity velocity $v(D(v) = LT^{-1})$ causes the change of quantity length $s(D(s) = L)$, given $s = vt$ (i.e. motion with a constant speed), where t stands for the absolute time. This again matches the intuitive understanding of the mechanics.

Generally speaking, a system is composed of some components. The behaviour of a system is determined by the behaviour of its components together with the specifications of their inter-connections, with the behaviour of each component being generally expressed as a set of equations. In this paper, the inter-connections between system components are specified using *structural constraints* which are imposed by the topological or geometrical linkages among these components. Such constraints imply the causal relations between the variables used to describe the boundary conditions of the components. This knowledge cannot be obtained from dimensional analysis, but from the design knowledge of the system (Bhaskar & Nigam 1990).

Definition 4. Given an ordinary equation or a structural constraint relating system variables $u, v, ...; x, y, ...$, which may be quantitative or qualitative and may include temporal derivatives, $[x, y, ...] = [u, v, ...]$ is named a symbolic causal equation, which signifies that the variables on the left hand side (LHS) causally depend on the variables on the right hand side (RHS) (if such a causal relationship between the system variables considered can indeed be established). The or-

der of the variables appearing on each side is arbitrary.

It is worth noting that, given an ordinary equation, it is not necessary to impose restriction over the number of the variables appearing on either side of its corresponding symbolic causal equation. This differs from the conventional representation of the irreversible (or asymmetric) causal equation (Iwasaki & Simon 1994), in which a constraint is imposed such that there is only one variable that is allowed to appear on the LHS.

For convenience, two binary predicates *Beless* and *Inequation* are introduced. Formula $Beless(x_1, x_2)$ states that x_1 is less independent than x_2. Formula $Inequation(x_1, x_2)$ states that variables x_1 and x_2 appear in the same equation within a given system model.

Conjecture. Let x_1 and x_2 be two variables in a given equation:

$$\begin{aligned} Beless(x_1, x_2) \iff & cd(x_1) > cd(x_2) \\ & \vee (cd(x_1) = cd(x_2) \wedge fd\_t(x_1) > fd\_t(x_2)) \\ & \vee (cd(x_1) = cd(x_2) \wedge fd\_t(x_1) = fd\_t(x_2) \\ & \quad \wedge fd(x_1) > fd(x_2)) \end{aligned}$$

Given two variables x_1 and x_2, if neither $Beless(x_1, x_2)$ nor $Beless(x_2, x_1)$ holds, the two variables are deemed to be *dimensionally equivalent*. A joint use of predicates *Beless* and *Inequation* allows the following notion to be defined.

Definition 5. A variable, x is called the *most dependent* variable in a given equation if it satisfies:

$$\neg \exists y (Inequation(x, y) \wedge Beless(y, x))$$

In order to deal with dynamic systems which always involve differential equations, two further notions have to be introduced. One is the well-known *Integral Causality* (Iwasaki & Simon 1994), which states that the value of a variable depends on the derivative of itself. The other is a weaker form of the differential causality rule also used in (Iwasaki & Simon 1994), which is herein referred to as the *Conditional Differential Causality Rule*. This rule requires that the most dependent temporal derivative or derivatives if there are more than one, among all the derivatives within a differential equation, be causally dependent on all the other variables and derivatives in that equation. For instance, given a differential equation:

$$f(x_1', x_2', x_3, x_4, x_5) = 0$$

if $beless(x_1', x_2')$, then x_1' is causally dependent on x_2' as well as on x_3, x_4 and x_5. If, however, x_1' and x_2' are dimensionally equivalent, then both are causally dependent on the variables x_3, x_4 and x_5.

The Conditional Differential Causality Rule is weaker because a) it does not require differential equations to be represented in first order, canonical form, which is required by some causal ordering methods (Iwasaki & Simon 1994; Lee & Compton 1994), and b) it allows more than one most dependent derivative to co-exist on the left hand side of a symbolic equation. A differential equation is said to be in canonical form if and only if there is only one derivative in the equation, and the derivative is the only term appearing on the left hand side of the equation (Iwasaki & Simon 1994). In addition, for the present approach, a derivative does not necessarily have to appear on the left hand side unless it is also a most dependent derivative among all the derivatives within the same equation.

Causal Ordering Algorithm

Having defined the basic notations for representing potential causal relationships amongst variables in a system model, an efficient algorithm is devised to rearrange the model into a set of symbolic irreversible causal equations. This results in the following *dimensional analysis based causal ordering algorithm*, where the identification of variables' dimensions are done by performing a simple pattern matching procedure.

1. For each component in the system, do
   (a) List the n variables that appear in the set of original equations, which define the behaviour of the component, and identify their dimensional representations.
   (b) For each non-exogenous variable, calculate its *cd*, fd_t and *fd*; for each exogenous variable, set its *cd*, fd_t and *fd* to zero.
   (c) For each non-exogenous equation, let $V = V_1 + V_2$ be the set of variables in it, such that V_1 contains all the derivatives and V_2 contains the others, do
      i. If the cardinality of V_1, $|V_1| = 0$, partition the set V_2 into two subsets *Left* and *Right*. Let *Left* contain all minimal elements of the partial order *beless*:
      $$Left = \{x | \neg \exists y \in V_2 \; Beless(x, y)\}$$
      and *Right* contain the remaining variables:
      $$Right = V_2 - Left$$
      ii. If $|V_1| > 0$, partition the set V_1 into *Right* and *Left* in the same way as partitioning V_2 as in (c.i), and then set
      $$Right = Right \cup V_2$$
      iii. For the two sets of variables, *Left* and *Right*, if both are not empty, rearrange their elements to form the corresponding symbolic causal equation, putting variables in set *Right* to the right hand side (RHS) of the symbolic equation and those in set *Left* to the left hand side (LHS).
      iv. If $|Right| > 1$, then set
      $$V_1 = \{v \mid v \in Right \wedge v\_is\_a\_derivative\}$$
      $$V_2 = Right - V_1$$
      and go to (i) (in order to identify, if any, further detailed causal relations amongst such variables).
2. For every derivative variable dx in the system model, create a symbolic equation such that $[x] = [dx]$.
3. For any two components which are connected, generate additional symbolic equations by directly putting the boundary variables of one component to the right hand side and those of the other to the left hand side, with respect to the causal implication indicated by the given structural constraints.

This algorithm produces a set of symbolic equations, which specifies the causal relations among all the system variables. It is mainly composed of two parts, the first (steps 1 and 2) dealing with equations within a single component and the second (step 3) coupling any two components. In this way, the algorithm is "context-free" within one component and becomes "context-dependent" between components. For different structural constraints (geometric or topological) between the components, the causal behaviour of the device may therefore be different accordingly. Thus, some variables may be treated as non-exogenous when one component is considered, while they may be treated as exogenous when another component is coupled to it. Structural constraints identify the causal relations among the variables involved.

The symbolic causal equations resulting from applying the algorithm can be interpreted as a causal graph, which is a directed graph with the nodes being system variables and the links being created by the following method:

> For any pair of variables, x and y, create a directed link from x to y if they jointly appear in one original equation or one structural constraint, and x is in the RHS set and y in the LHS set of the corresponding symbolic equation.

Formally, a directed graph $G = (V, E)$ consists of a finite non-empty set V of elements called *nodes* and a set E of ordered pairs between the elements of V called *links*. For simplicity, the set of all links pointing away from (pointing to) v, $v \in V$, is denoted by $E_{out}(v)$ ($E_{in}(v)$), and the set of all nodes associated with $E_{out}(v)$ ($E_{in}(v)$), excluding v, is denoted by $V_{out}(v)$ ($V_{in}(v)$). A node, v, is called an end node if $|E_{out}(v)| = 0$ or $|E_{in}(v)| = 0$.

In describing the behaviour of some system components or their relationships, the defining sets of equations and/or structural constraints may include more than one most dependent variable. This algorithm puts all such most dependent variables or derivatives into the LHS of the resulting symbolic causal equations, rather than artificially choosing just one of them to be put in the LHS. Although this may sound conservative, this treatment has so far appeared to produce reasonable and intuitive results. There are at least two reasons that support the present approach. Firstly, some of the variables in a model may represent physical quantities that stand in a fully equivalent position. It is too difficult, if not impossible, to tell the causal relations among them. Secondly, should there be additional information available on the description of the system, this may be utilised to discriminate further causal relations without being forced to assume one of the variables to be the most dependent and later to retract such assumptions. This approach is, therefore, more flexible compared to the approach where only one variable is allowed to be derived variable regarding any one equation, an approach that is typically employed in the existing causal ordering techniques.

It should be noted that step (1.c.iv) is a recursive procedure, allowing the algorithm to maximise the exploitation of dimensional information embedded in the system variables. This is justified on the ground that for any pair of variables that are both involved in one equation and, hence, are inter-related, if they are of a different independent level, then one is dependent on the other. However, in some cases this treatment may generate certain causal relations that are too detailed to be necessary for the explanation of the system considered.

Another point worth mentioning is that the above algorithm requires no explicit equations, quantitative or qualitative, to be actually given, but an implicit indication of there being a relation between the variables involved. This is due to the fact that the causal relationships among the variables appearing in a given relation are generated by analysing the dimensions of these variables alone.

Algorithm Extension

The above algorithm, as with other existing causal ordering methods, only returns a description of cause-effect relationships between system variables given a set of reversible equations os relations and a set of structural constraints. No identification of the causal impact in terms of whether an increase in the value of an influencing variable may lead to an increase or a decrease in the value of the influenced variables is provided. In general, this kind of identification can be derived from taking the partial derivative of the influenced variable with respect to the influencing variable. The question is how to calculate this kind of partials and what conditions are required of such calculation. Fortunately, a useful technique, the Π-calculus (Bhaskar & Nigam 1990), which is also based on the dimensional analysis, has been developed for deriving such identification for a given physical system. Although the method based on qualitative confluences (de Kleer & Brown 1984) may also be used for this purpose, the employment of confluences requires specialised restatements of physical laws, whilst the regimes used in the Π-calculus are directly derivable from dimensional analysis without explicit knowledge of physical laws.

To apply Π-calculus, a crucial step is the selection of r basis variables in a model component. In general, there are $\binom{n}{k}$ choices; however, many of these do not yield an ensemble. Although some heuristics that may help for such selection are given in (Bhaskar & Nigam 1990), a more explicit method is provided here, which covers those heuristics as specific cases:

1. For any component, write the set of all variables in the component as S, set $S_r = \phi$, where r is the size of the basis;
2. Find a variable x in S, satisfying

$$\neg\exists y \in S \; Beless(x, y)$$

   (a) If there is only one such variable, then this variable is selected and is put into set S_r;

(b) If there are more than one such variable, then choose any one of them and put it into the set S_r.

(c) Set $S = S - S_r$

3. If $|S_r| = r$ then stop, the set S_r is the basis for the corresponding component; otherwise go to step 2.

Although temporal derivatives are allowed in the causal graph generated by the above causal ordering algorithm, only the relations among variables are considered in the Π-calculus. In order to integrate the Π-calculus into this algorithm, the causal graph is needed to be modified to remove the derivatives. For this purpose, the *transitivity* that causality possesses as a basic characteristic is used. The transitivity states that that x causes y and y causes z implies that x causes z. Thus, two causal links: a link from a variable x to a derivative dy and a link from the derivative dy to a variable z can be replaced by a link from the variable x to z. With this transitional procedure, if z is itself also a derivative and there is a link from z to another variable u, then the path from x to u can be replaced by a single link from x to u, and so on. The graph resulting from the use of the transitivity shows the dependencies amongst system variables only, excluding all the temporal derivatives. Such graphs are hereafter named *derived causal graphs*.

Given $G = (V, E)$ being a causal graph including some derivatives as its nodes, a derived causal graph $G' = (V', E')$, which does not include any derivative as its node, can be obtained from G using the following procedure:

1. Set $V' = V$ and $E' = E$.
2. Select a derivative node dx in V', do

(a) If dx is not an end node, set

$$E' = E' - \{E'_{out}(dx) \cup E'_{in}(dx)\} \cup \{u \rightarrow v | u \in V'_{in}(dx), v \in V'_{out}(dx)\}$$

(b) If dx is an end node, set

$$E' = E' - \{E'_{out}(dx) \cup E'_{in}(dx)\}$$

(c) Set $V' = V' - \{dx\}$.

3. If there are no derivative nodes left, stop; otherwise go to step 2

A derived causal graph can then be analysed by means of the Π-calculus, in order to obtain a *causal graph with impact signs*, in which links are annotated with a causal influence effect sign, + or -. The + sign from variable x to variable y signifies that x causes y to change in a monotonically increasing way, and the - sign between them indicates that x causes y to change in a monotonically decreasing way. Such information is obviously very useful for performing many qualitative reasoning tasks (de Kleer & Brown 1984; Forbus 1984; Bhaskar & Nigam 1990; Lee & Compton 1994; Kuipers 1994).

The change from an ordinary causal graph to the corresponding causal graph with impact signs may appear to loose some of the representational power of the original, since derivatives are eliminated. However, the derivation of the new graph does not necessarily destroy the old one. Detailed causal links involving derivatives can be reserved by recording the original causal graph, although there is no information on the actual effect of how an influencing variable may cause the influenced variable to change there.

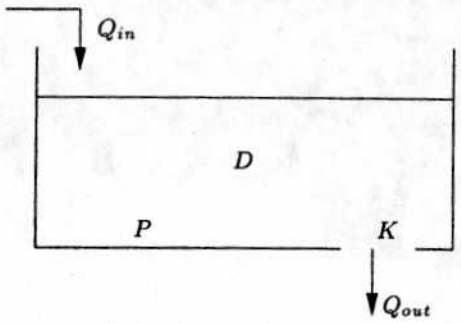

Figure 1: Bathtub System

Table 1: Meaning of the variables

Variable	Variable Meaning
Q_{oin}	the input flow rate
Q_{out}	the output flow rate
D	the mass of water in the tub
P	the pressure at the bottom
K	the size of the valve opening

Results

The above causal ordering algorithm has been implemented. To illustrate the basic ideas, this section presents some of the results from test-runs using a number of different system models.

Integral Causality

The proposed algorithm supports the Integral Causality rule (Iwasaki & Simon 1994). In fact, the *fd* value of dx/dt is always less than that of x, i.e. $Beless(x, dx/dt)$ holds in general. For instance, the temporal derivative of distance s is velocity v, whilst $D(s) = L$ and $D(v) = LT^{-1}$ which gives $fd(s) = 1$ and $fd(v) = 0$. The resulting explanation is of course very intuitive, showing that distance depends on velocity as mentioned before.

Bathtub Model

This simple system, as shown in figure 1, which is slightly revised from that given in (Iwasaki & Simon 1994) consists of five system variables as listed in table 1.

The only differences between these two models are that instead of using the depth of the water in the tub,

Table 2: Dimensional values for Bathtub

Variables	Dimensions	*cd*	*fd_t*	*fd*
Q_{out}	L^3T^{-1}	1	3	2
Q_{in}	L^3T^{-1}	0	0	0
D	M	1	1	1
P	$ML^{-1}T^{-2}$	2	0	-2
K	L^2	1	2	2

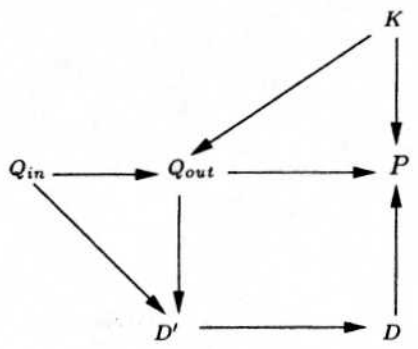

Figure 2: Causal graph for the bathtub

the mass of water is used to describe the volume of the water in the tub , and that no assumption of variable K being exogenous is made. The system is modelled by:

$$Q_{out} = c_1 K P \qquad D = c_2 P$$
$$D' = c_3(Q_{in} - Q_{out}) \qquad Q_{in} = c_4$$

As this system is composed of only one component, no structural constraints are involved. Given this model, it is straightforward to work out the dimensions and the *cd*, *fd_t* and *fd* values for each system variable. The results are listed in table 2.

With the input flow rate Q_{in} being an exogenous, running the dimensional-analysis based causal ordering algorithm leads to the following symbolic equations:

$$[P] = [K, Q_{out}] \qquad [Q_{out}] = [K]$$
$$[P] = [D] \qquad [D'] = [Q_{in}, Q_{out}]$$
$$[Q_{out}] = [Q_{in}] \qquad [D] = [D']$$

which can be depicted as shown in figure 2. This indicates:

> The input flow rate affects the rate of change of the mass and the output flow rate, the latter also affects the rate of change of the mass and the pressure at the bottom of the tub. The size of the valve opening determines the output flow rate and the pressure at the bottom of the tub, with the latter also depending on the mass of water in the tub.

It is important to notice that, without the assumption of variable K being exogenous, which is required by the existing approaches as presented in (Iwasaki & Simon 1994; Travé-Massuyés & Pons 1997) (in order to make the system model self-contained), the present algorithm can still produce a causal ordering for the system variables. It is, perhaps, more important to notice that although most cause-effect links produced herein are the same as those obtained using either of the existing approaches, there exists a significant difference. The present work gives an explanation that the change of the pressure at the tub bottom is caused by the change of the output flow rate. This matches physical and intuitive understanding of the modelled system (Gawthrop & Smith 1996) which involves a single component.

However, the conventional causal ordering theories, e.g. (Iwasaki & Simon 1994) would provide an explanation that the change of the pressure at the tub bottom determines the change of the outflow rate, resulting in a feedback loop within the system. Nevertheless, the derivation of this feedback loop relies upon an additional constraint that assumes that the mass (or depth) of the water in the tub remains constant. This assumption implies that another component, say, a regulator is needed to control the output flow rate. From this point of view, the outflow rate will become exogenous for the original single-component system model. Given this being the case, the application of the present causal ordering algorithm will then generate a causal graph involving the same feedback loop. In addition, the causal link between Q_{in} and Q_{out} will no longer become explicit. Also, the causal influence direction between the valve opening K and the outflow rate Q_{out} will be reversed, which correctly indicates that the required outflow rate will determine the size of the valve opening (which is exactly the job of the regulator component).

Another interesting observation from the causal graph of figure 2 is that a causal link between K and P is established, which is not achieved using either of the above mentioned existing approaches. From this graph, a derived causal graph can be generated using the algorithm extension, in order to obtain the influence sign for each link, by means of applying the Π-calculus.

In this example, there are five quantities and three dimensions. Q_{in}, D and K are selected as the basis by the algorithm extension.

The resultant Πs are:

$$\Pi_1 = Q_{out}/Q_{in}, \Pi_2 = PK^{7/2}/DQ_{in}^2$$

Then the relative intra-regime partials can be calculated such that

- From Π_1: $\partial Q_{out}/\partial Q_{in} > 0$
- From Π_2: $\partial P/\partial K < 0$, $\partial P/\partial D > 0$, $\partial P/\partial Q_{in} > 0$

Given Q_{in} as a contact variable, the corresponding inter-regime partials are:

$$[\partial P/\partial Q_{out}]^{Q_{in}} = (\partial P/\partial Q_{in})/(\partial Q_{out}/\partial Q_{in}) > 0$$

and

$$[\partial Q_{out}/\partial K]^{Q_{in}} = (\partial Q_{out}/\partial Q_{in})/(\partial K/\partial Q_{in}) > 0$$

This leads to the causal graph with signs as shown in figure 3.

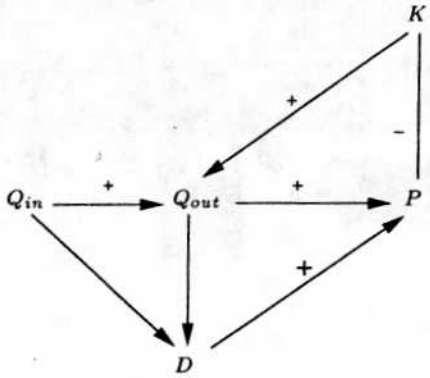

Figure 3: Causal graph with signs for the bathtub

It is worth noting that no signs are obtained for the derived causal links from Q_{in} and Q_{out} to D. This is because, for example, the link from Q_{in} to D comes from the path $Q_{in} \rightarrow D' \rightarrow D$ while both Q_{in} and D are in the basis of the same regime. These sign-free links are correct, however, since D depends on both Q_{in} and Q_{out}. An increase in Q_{in} does not necessarily cause the increase in D, unless it is known that the outflow rate is less than the inflow rate, a decrease in Q_{out} does not have to result in an increase in D, and so on.

Pressure Regulator

The pressure regulator model is adopted from (de Kleer & Brown 1984). The function of the pressure regulator is to maintain a constant pressure at its output. This device consists of two components: a pipe with an orifice and a spring valve as shown in figure 4. Each component is modelled individually first and the two components are then coupled to form the system model, subject to given structural constraints. The meaning of the system variables is listed in table 3.

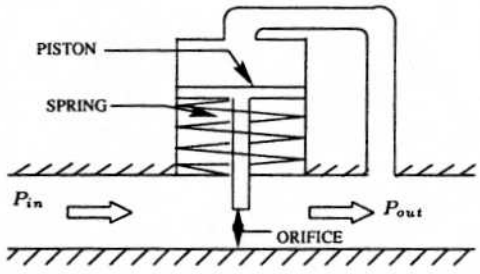

Figure 4: The pressure regulator

The component of the pipe with an orifice can be modelled by the following equations:

$$P_{out} = c_1 Q \qquad Q = (c_2 P_{in} + c_3 A_{open})/c_4 \rho$$
$$P_{in} = c_5 \qquad \rho = c_6$$

Where P_{in} and ρ are exogenous, and c_1 - c_6 are positive constants. The spring valve component can be modelled

Table 3: Meaning of the variables

Variable	Variable Meaning
P_{out}	the outlet pressure
P_{in}	the inlet pressure
ρ	the fluid density
Q	the orifice flow rate
A_{open}	the orifice opening
x	the spring displacement
P	the pressure on the piston
k	the spring constant

Table 4: Dimensional values

Variables	Dimensions	cd	fd_t	fd
P_{out}	$L^{-1}MT^{-2}$	2	0	-2
P_{in}	$L^{-1}MT^{-2}$	0	0	0
ρ	ML^{-3}	0	0	0
Q	L^3T^{-1}	1	3	2
A_{open}	L^2	1	2	2
x	L	1	1	1
P	$L^{-1}MT^{-2}$	0	0	0
k	MT^{-2}	0	0	0

by:

$$-kdx = c_7 dP \qquad P = c_8 \qquad k = c_9$$

Where c_7 - c_9 are positive constants and k and P are exogenous variables, and so is the derivative dP. The dimensional values for each variable are worked out as shown in table 4.

There are two structural constraints between the two components. One is the connection that transmits the outlet pressure in the pipe to the piston in the spring valve component. That is, the rate of change of the pressure on the piston, dP is determined by the outlet pressure in the pipe, P_{out}. This connection also yields the initialisation that the pressure on the piston is exogenous. The other constraint is that the motion of the piston affects the orifice opening; more specifically as the spring is compressed, the orifice reduces. This constraint indicates that the displacement x determines the opening A_{open}. Running the dimensional-analysis based causal ordering algorithm results in the following symbolic equations:

$$\begin{array}{llll} [P_{out}] = [Q] & & [Q] = [P_{in}, \rho, A_{open}] \\ [A_{open}] = [P_{in}, \rho] & & [dx] = [k, dP] \\ [dP] = [P_{out}] & & [A_{open}] = [x] \\ [P] = [dP] & & [x] = [dx] \end{array}$$

This leads to the causal graph as shown in figure 5, in which there is a feedback loop:

$$Q \rightarrow P_{out} \rightarrow dP \rightarrow dx \rightarrow x \rightarrow A_{open} \rightarrow Q$$

From this causal graph, a derived causal graph without derivatives involved can be generated by the

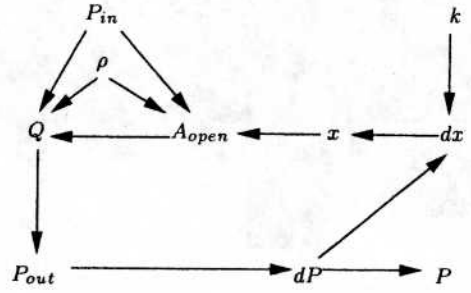

Figure 5: Causal graph for the pressure regulator

algorithm extension given in the previous section. For instance, the link from P_{out} to x comes from the path $P_{out} \rightarrow dP \rightarrow dx \rightarrow x$, while the link from P_{out} to P comes from the path $P_{out} \rightarrow dP \rightarrow P$.

In the pipe component, there are five variables and three dimensions are involved. It is easy to select Q_{in}, ρ and A_{open} as the basis by the algorithm extension. Similarly, P and k can be chosen as the basis for the spring component. Applying the Π-calculus to the derived causal graph, partial derivatives and hence the causal graph with impact signs for the pressure regulator can be obtained. The result is shown in figure 6. To show how the partials are calculated, for example, consider the pipe component, in which there are two regimes:

$$\Pi_1 = (Q\rho^{1/2})/(A_{open}P_{in}^{1/2}) \qquad \Pi_2 = P_{out}/P_{in}$$

The relative intra-regime partials can be calculated as

- From Π_1: $\partial Q/\partial P_{in} > 0$, $\partial Q/\partial A_{open} > 0$, $\partial Q/\partial \rho < 0$.
- From Π_2: $\partial P_{out}/\partial P_{in} > 0$.

Given P_{in} as a contact variable, the corresponding inter-regime partial is calculated by

$$[\partial P_{out}/\partial Q]^{P_{in}} = (\partial P_{out}/\partial P_{in})/(\partial Q/\partial P_{in}) > 0.$$

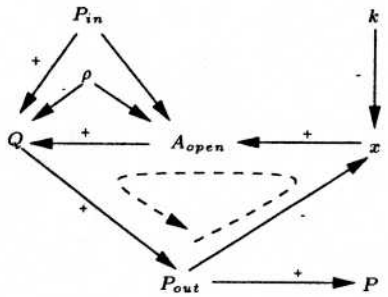

Figure 6: Causal graph with signs for the pressure regulator

From the reaching causal graph, the behaviour of the pressure regulator can be reasoned as follows:

> An increase in P_{in} leads to an increase in Q. This increase in Q leads to an increase in P_{out}. The increase in P_{out} leads to a decrease in the spring displacement x. This decrease in x leads to a decrease in A_{open} in the pipe orifice component. The decrease in A_{open} leads to the decrease in Q. Finally, this decrease in Q leads to a decrease in P_{out}.

This explanation confirms that there is a feedback loop: an increase in P_{out} eventually leads to a decrease in P_{out}. This reflects the correct desired functionality of the system - to prevent the outlet pressure from deviating from a preset constant value. The causal explanations given here are intuitive. However, the causal links from P_{in} to A_{open} and from ρ to A_{open} seem to be a bit too detailed, caused by the use of the recursive procedure of the algorithm. The value of A_{open} is indeed affected monotonically by the values of both P_{in} and ρ in general, although such influences are often ignored in many applications due to their considerably less significance in comparison with other causal links generated within the system.

It is interesting to compare the present technique with those proposed in (Iwasaki & Simon 1986; 1994; Lee & Compton 1994; Travé-Massuyés & Pons 1997) with respect to the treatment of multi-component devices. In this work, if a device is composed of more than one component, each component is considered separately first and the structural constraints among the components are then taken into account while coupling pairs of the components. The pressure regulator example shows that this treatment is quite successful. However, using the methods mentioned above, additional explicit equations should be given in order to represent the component connections or specific functionalities such as feedback loops. Finally, those approaches do not provide the impact signs of the actual causal effects. The present work provides a method to specify such impacts.

Conclusions

This paper has presented a novel approach to generate a description of the causal relationships between system variables and, also, to identify the actual impact of the causal effects by attaching a calculated positive or negative sign to each generated causal link. The sign indicates whether a change in the value of an influencing variable will lead to an increase or a decrease in the value of influenced variables. The work rests its theoretical foundations on the conventional dimensional analysis developed in classical physics and the Π-calculus reported in (Bhaskar & Nigam 1990). The causal ordering derived depends not only on the relationships among the system variables and their dimensional representation, but also on the system initialisation and the structural constraints.

Experimental results have shown that the present approach retains the most appealing characteristics of the existing causal ordering approaches, and enjoys being able to produce a causal explanation for considered systems, which reflects intuitive understanding of causal dependencies amongst the system variables. However,

the present approach is not able to produce any causal ordering between related varialbes which are dimensionally equivalent, though such variables might stand for quantities of the same physical position in the first place.

The work requires a number of further investigations, including: a) examining its application for more complex systems; b) investigating the possibility of integrating the present method with one or more other existing causal ordering algorithms, in order to maximise their benefits to generate better causal explanations; and c) studying its application for the purpose of fault diagnosis. The successful outcome of such future work would certainly enhance the performance of qualitative model-based reasoning systems.

Acknowledgements

This work is partially supported by the UK EPSRC grant GR/L88801, with kind contribution from Intelligent Applications Ltd., Edinburgh, Scotland.

References

Bhaskar, R., and Nigam, A. 1990. Qualitative physics using dimensional analysis. *Artificial Intelligence* 45:73–111.

Buckingham, E. 1914. On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev. IV (4)* 345–376.

de Kleer, J., and Brown, J. 1984. A qualitative physics based on confluences. *Artificial Intelligence* 24:7–83.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence* 24:85–168.

Gawthrop, P., and Smith, L. 1996. *Metamodelling: Bond graphs and dynamic systems.* Prentice Hall.

Huntley, H. 1952. *Dimensional Analysis.* MaCdonald & Co. (Publishers) Ltd.

Iwasaki, Y., and Simon, H. 1986. Causality in device behaviour. *Artificial Intelligence* 29:3–32.

Iwasaki, Y., and Simon, H. 1994. Causality and model abstraction. *Artificial Intelligence* 67:143–194.

Kuipers, B. 1994. *Qualitative Reasoning, Modeling and Simulation with incomplete knowledge.* MIT Press.

Lee, M., and Compton, P. 1994. Context-dependent causal explanations. In *Proceedings of QR-94*, 176–186.

Top, J., and Akkermans, H. 1991. Computational and physical causality. In *Proceedings of IJCAI-91*, 1171–1176.

Travé-Massuyés, L., and Pons, R. 1997. Causal ordering for multiple mode systems. In *Proceedings of QR-97*, 203–214.