Computer Techniques in Environmental Studies, (P. Zanetti ed.) Computational Mechanics Publications, Southampton, 1992.

The semiqualitative analysis of waste water treatment

Simon Parsons
Department of Electronic Engineering, Queen Mary and Westfield College, Mile End Road, London E1 4NS, United Kingdom.

Mirko Dohnal
Department of Chemical Engineering, Technical University of Brno, Technicka 2, 61669 Brno, Czechoslovakia.

ABSTRACT

The solution of environmental problems such as waste water treatment are usually based upon the development of sets of differential equations. A complete analytical solution of the model requires that every numerical constant in this set of equations is precisely known. This paper describes a computer program that implements a method from artificial intelligence which permits the solution of such sets of equations when constant values are unknown, whilst allowing those values that are known to be used. The use of the system is illustrated with the solution of a set of equations describing an anaerobic fermentor.

INTRODUCTION

The basis of modern engineering is knowledge of laws of nature such as the law of mass conservation, which form the foundation of any equation-oriented mathematical model. Given a particular environmental system it is usually not difficult to collect together a set of relevant differential equations, based on laws of nature, which will form the nucleus of the mathematical model of that system. These equations describe the dynamic behaviour of various key variables, whose values, when the equations are solved, predict the state of the system at any instant. However, to take this nucleus and flesh it out with all the necessary information that will make it an accurate and realistic working model is far from simple.

This is because the laws of nature on their own are not sufficient to build a good model. What is needed are the precise values of the constants that relate the variables in the differential equations. Unfortunately, real environmental systems are complex, integrated, ill-known, unique, and difficult to measure, in particular when their dynamic behaviour is considered. They may be subject to complex relations with their surroundings which may make it nearly impossible to isolate them without distorting any measurements made. Therefore knowledge of such systems may be inconsistent, and sparse, and it may be extremely time-consuming and difficult, and therefore expensive, to identify the value of every numerical constant. Without the values of all the constants the set of equations have no practical value. This is because the kind sets of complex equations that arise from the study of environmental systems are can only be solved by simulation, and it is not possible to run a classical simulation when constant values are unknown.

Thus incomplete and uncertain knowledge of the necessary numerical constants would seem to rule out projects such as modelling the scaling up of laboratory fermentors, risk evaluation, and cost estimation by conventional methods. It need not, however, prevent the use of artificial intelligence techniques such as qualitative reasoning[1].

QUALITATIVE REASONING

Qualitative reasoning is a method that attempts to capture the essence of human reasoning about complex systems. Rather than attempt to deal with a mass of numerical data, values are only distinguished as positive (+), zero (0), negative (-), or unknown (?). These values are sufficient to identify many of the interesting features of the behaviour of the important variables in a given system. Thus if we are interested in a key variable x, a substrate concentration say, whose value is known to be positive, and we can establish from the qualitative simulation of the model that the first time derivative of x is positive, while the second time derivative of x is negative:

$$x = + \qquad \frac{dx}{dt} = + \qquad \frac{d^2x}{dt^2} = - \tag{1}$$

then we know that the behaviour of the concentration over time will be to rise to some limiting value (Figure 1). We may not know what the limit is, but we do know that the concentration will eventually level off, and this less precise information may be sufficient.

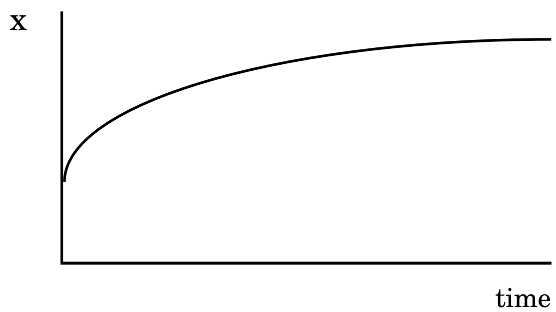


Figure 1. The qualitative behaviour described by (1)

Clearly if we are trying to establish that the substrate concentration is stable in the long term then the information we are able to deduce is quite adequate, and in many cases the fact that we can learn something from qualitative reasoning far outweighs the fact that what we can learn is not very detailed. That this is true is borne out by the success of the qualitative approach in practice[1–3].

Despite this undoubted success, there are some problems with qualitative reasoning that make it unsuitable for modelling certain systems. These problems stem from the limited number of values that any constant or variable can adopt. Raiman[4] illustrates this with a simple example from mechanics. Consider two masses which collide whilst travelling towards one another along the same line. One has a large mass M and velocity V, the other has a small mass and velocity m and v:

M, V $\qquad$ m, v

The net momentum from right to left above is given by the law of the conservation of momentum as:

$$MV - mv \tag{2}$$

Since M, V, m and v are all positive values, they all have qualitative value $+$, and the net leftwards momentum is established by the calculation:

$$+ \otimes + \ominus + \otimes + \tag{3}$$

where $\ominus$ is the operator representing the difference of two qualitative values, and $\otimes$ is the operator representing the multiplication of two such values. It is clear that the product of two positive values will itself be positive so that the calculation reduces to:

$$+ \ominus + \tag{4}$$

Again, brief consideration will show that the difference of two values which are only known to be positive can be either positive, negative or zero, depending on the relative sizes of the values. Thus qualitative reasoning can only deduce that the overall leftwards momentum will be ?, while intuitively we can see that it will be $+$. In the remainder of this paper we present a generalisation of qualitative reasoning, known as semiqualitative reasoning, which solves this problem, and has other advantages which make it especially suitable for solving complex sets of differential equations.

SEMIQUALITATIVE REASONING

In semiqualitative reasoning[5], the values of variables and constants are restricted to a set of $2k + 1$ intervals. This set of intervals covers all numbers from ∞ to $-\infty$, and the intervals are continuous and non-overlapping , so that any real number falls into one, and only one, interval. The intervals are symmetric about zero, which is itself an interval, and there are k positive and k negative intervals. The boundaries of the intervals may be set by an arithmetic or geometric progression, or may be chosen to reflect what are considered to be interesting values. Since the set of values used in qualitative reasoning corresponds to the set of semiqualitative intervals obtained for $k = 1$, it is clear that semiqualitative reasoning is a generalisation of qualitative reasoning.

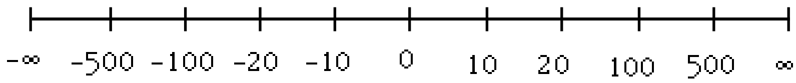
A basic understanding of how semiqualitative reasoning may be used to solve sets of differential equations may be obtained from a simple example. Consider the following set of equations as a model of a physical system:

$$x_1 + x_2 = x_3 \tag{5}$$

$$x_1 \cdot x_4 = x_3 \tag{6}$$

$$\frac{dx_4}{dt} = x_5 \tag{7}$$

The total number of variables is 5. Further consider that we have the set of intervals:



The only values that we consider as quantifiers for the values of a variable, and the values of its first and second derivatives, are the semiqualitative intervals. Since third and higher derivatives are usually unavailable, any qualitative variable is considered to be fully specified by the triplet of value, first derivative and second derivative. The following five triplets describes one set of assignments of values to the 5 variables, and thus one conceivable state of the system whose behaviour they describe:

		x	$\frac{dx}{dt}$	$\frac{d^2x}{dt^2}$	
x_1	<	(0–20)	(0–10)	0	>
x_2	<	(20–100)	(0–10)	(-20 –-100)	>
x_3	<	(10–20)	(20–100)	(20–100)	>
x_4	<	(0–10)	(0–10)	0	>
x_5	<	(10–20)	(0–-10)	0	>

This state is not however a physically possible state of the system since it is not a solution of equations (5)–(7). This is because x_3 is determined from $x_1 = [0, 20]$ and $x_2 = [20, 100]$ by (5), and $[0, 20] \oplus [20, 100] = [20, 500]$, where $\oplus$ represents the addition of two intervals which is carried out using interval arithmetic[6], whereas $x_3 = [10, 20]$. By similar means it is possible to identify all the 5-triplets which are solutions of the set of equations, and these correspond to all the qualitative solutions of the model.

By allowing variables to take on a wider range of values, semiqualitative reasoning permits the use of those numerical values that are known, and this means that it generates more precise solutions than are possible using qualitative reasoning. However, the fact that it is not necessary to have any more information than whether a quantity is positive, negative or zero means that semiqualitative reasoning is very robust, and may be used in situations where conventional methods cannot be used. In the next section, the use of semiqualitative reasoning is demonstrated by the solution of a problem from waste water treatment.

AN EXAMPLE

This section describes the semiqualitative simulation of an anaerobic digester. The digestion of organic waste is an important step in the treatment of waste water, and so the simulation of digesters that achieve this is important in order to establish the correct conditions under which the process should be carried out.

It is possible[7] to write down a complex set of equations which fully describe the action of an anaerobic fermentor and which, when solved, provide a suitable model of its behaviour. Unfortunately, the results of this analysis hinge upon the values of a number of key constants whose values not only vary from fermentor to fermentor, but are also extremely difficult to measure. As a result it is difficult and expensive to provide accurate solutions from a conventional analysis. A semiqualitative analysis is, however, possible using much less precise data.

The following set of equations provide a simplified model of the behaviour of an anaerobic fermentor. The full model may also be solved using semiqualitative methods, but the additional detail adds nothing to the understanding of the technique:

$$\frac{dx_1}{dt} + (k_{12} + k_{13}).x_1 + k_{11}x_1 . x_5 = +k_{21}.x_2 \qquad (8)$$

$$\frac{dx_2}{dt} + k_{21}x_2 = k_{12}x_1 \qquad (9)$$

$$\frac{dx_3}{dt} = k_{13}x_1 + k_{43}x_4 \qquad (10)$$

$$\frac{dx_4}{dt} + k_{43}x_4 = k_{11}x_1x_5 \qquad (11)$$

$$\frac{dx_5}{dt} + k_{53}x_5 + k_{11}x_1x_5 = 0 \qquad (12)$$

where x_1–x_5 are concentrations of various substrates, either those wastes being digested or the products of the digestion. In addition some of the values of the constants are known:

$$k_{11} = 100 \qquad k_{53} = 0.3$$
$$k_{13} = 5.0 \qquad \frac{k_{12}}{k_{21}} = 0.5$$
$$k_{43} = 1.0$$

and we also know that $k_{12} = 1.5$, and $k_{21} = 3.0$. This model is first solved with the boundaries of the semiqualitative intervals set as:

-∞ -1000 -500 -100 -20 -10 0 10 20 100 500 1000 ∞

We can specify additional constraints as:

		x	$\frac{dx}{dt}$	$\frac{d^2x}{dt^2}$	
x_1	<	(0–20)	(0–10)	0	>
x_2	<	(10–1000)	?	?	>
x_3	<	(0–20)	?	?	>
x_4	<	(10–1000)	?	?	>
x_5	<	(10–20)	(0–-10)	0	>

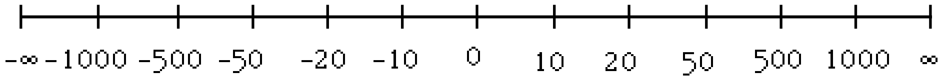
This set of constraints may be considered as a query, in this case asking the question:

"When x_1 is present in a concentration of less than 20, what are the ways in which it is possible to achieve a linear ($d^2x_1/dt^2 = 0$) increase of concentration of x_1 of less than 10 units per unit time while x_5 is present with a concentration of between 10 and 20, and changes linearly ($d^2x_5/dt^2 = 0$) at a rate of less than 10 units per unit time? Meanwhile x_3 is known to have a positive concentration of less than 20, while that of x_2 and x_4 is between 10 and 1000. The way that these last three variables change with time is not known."

Note that ? is equivalent to the interval [-∞, ∞]. Solving the model with this set of values gives the solution:

		x	$\frac{dx}{dt}$	$\frac{d^2x}{dt^2}$	
x_2	<	(20–100)	(20–100)	(0 –-10)	>
x_3	<	(10–20)	(500–1000)	(-20–-100)	>
x_4	<	(500–1000)	(-500–-1000)	(500-1000)	>

Which gives us a reasonably detailed idea of what values the substrate concentrations should have, and is a definite improvement on the information we had initially. The model may also be refined. for instance, if we want to further investigate the value and first derivative of x_2, say, we could choose a new set of intervals, choosing the upper limit of the third positive interval to be 50 instead of 100:



With the same set of constraints as before the following solution is generated:

		x	$\frac{dx}{dt}$	$\frac{d^2x}{dt^2}$	
x_2	<	(20–50)	(20–50)	(0 -–10)	>
x_3	<	(10–20)	(500–1000)	(-20–-50)	>
x_4	<	(500–1000)	(-500–-1000)	(500-1000)	>

which shows that by making certain intervals narrower, it is possible to make the solution more accurate.

A PROGRAM FOR SEMIQUALITATIVE ANALYSIS

Having seen the kind of results that semiqualitative analysis can generate, we consider a software system which can perform a semiqualitative analysis on a set of differential equations. In order to explain what the program does, and how it is operated, we discuss in detail the process of solving the example discussed above.

Preparing the input

The program for semiqualitative analysis works by considering the relations between the semiqualitative variables of the set of equations as a series of constraints upon their value. The analysis then consists of taking the known values, and propagating these through the network of constraints, seeing how they affect those values that are initially undefined. In order to do this the program needs some way of specifying the relationships between variable values, and this is done by means of a series of functional blocks (Figure 2).

Identification	Description
M1	Addition
M2	Multiplication
M3	Derivative
M4	X > Y
M5	DX > DY
M6	DDX > DDY

Figure 2. Functional blocks, note that DX is short for $\frac{dx}{dt}$, and DDX for $\frac{d^2x}{dt^2}$

Each block in the semiqualitative system has one or two inputs, and a single output, and specifies that a particular relationship holds between the inputs and the output. For instance the M1 block has two inputs and a single output, and specifies that the output must be equal to the semiqualitative sum of the inputs. In some ways to talk of input and output is a little misleading, since the propagation need not take place from input to output. Indeed what happens is that a graph is constructed whose arcs are semiqualitative variables and whose nodes are functional blocks, and the known constraints propagated around until the values are as refined as possible.

The first stage in the analysis is to write the equations (8)–(12) in form in which they may easily be specified using the functional blocks. Initially they are written as a series of variables related only by addition (there is no subtraction block since subtraction causes problems in interval arithmetic, and any equation written using subtraction may be rewritten using addition) and equality. This generates a set of equations:

$$x_6 + x_{11} + x_{19} = x_{12} \quad (13)$$
$$x_7 + x_{12} = x_{14} \quad (14)$$
$$x_{15} + x_{16} = x_8 \quad (15)$$
$$x_9 + x_{16} = x_{19} \quad (16)$$
$$x_{10} + x_{17} + x_{19} = x_{18} \quad (17)$$

which may be directly written down in terms of functional blocks. There are further equations which relate the variables in equations (13)–(17) to each other and the variables whose values are specified in the query:

$$x_6 = \frac{dx_1}{dt} \qquad x_{13} = 100 \times x_1$$
$$x_{14} = 1.5 \times x_1$$
$$x_7 = \frac{dx_2}{dt} \qquad x_{15} = 5.0 \times x_1$$
$$x_8 = \frac{dx_3}{dt} \qquad x_{16} = 1.0 \times x_4$$
$$x_{17} = 0.3 \times x_5$$
$$x_9 = \frac{dx_4}{dt} \qquad x_{18} = 0$$
$$x_{10} = \frac{dx_5}{dt} \qquad x_{19} = x_{13} \times x_5$$
$$x_{20} = x_{19} + x_{11}$$
$$x_{11} = 6.5 \times x_1 \qquad x_{21} = x_{19} + x_{17}$$
$$x_{12} = 3.0 \times x_2$$

The full set of equations, written in terms of functional blocks, form one part of the input to the program. The second part of the input is the semiqualitative query mentioned above, which sets the limit on the values of the variables in the original set of equations, in this case variables x_2-x_4. These limits may be any pair of the semiqualitative boundaries, which themselves form the third part of the input. The fourth and final part of the input is a list of variables whose value are required in the output. In the example since we are interested in the values of x_2-x_4 this part of the input will contain the names of these variables and their first and second derivatives, since we want to know the value of all three.

What the program does

The first thing that the program does is to compile a set of combinator tables from the set of semiqualitative intervals, a measure designed to make it more efficient when it comes to applying the mathematical constraints. These tables are compiled by considering the pairwise combination of every possible set of intervals. Thus for the set of intervals in the example it would first consider adding [1000, ∞] to [1000, ∞], which gives [1000, ∞] since the addition of two numbers in the interval cannot lie outside the interval. Next it would try [1000, ∞] and [500, 1000] which would again give [1000, ∞], and continue until it

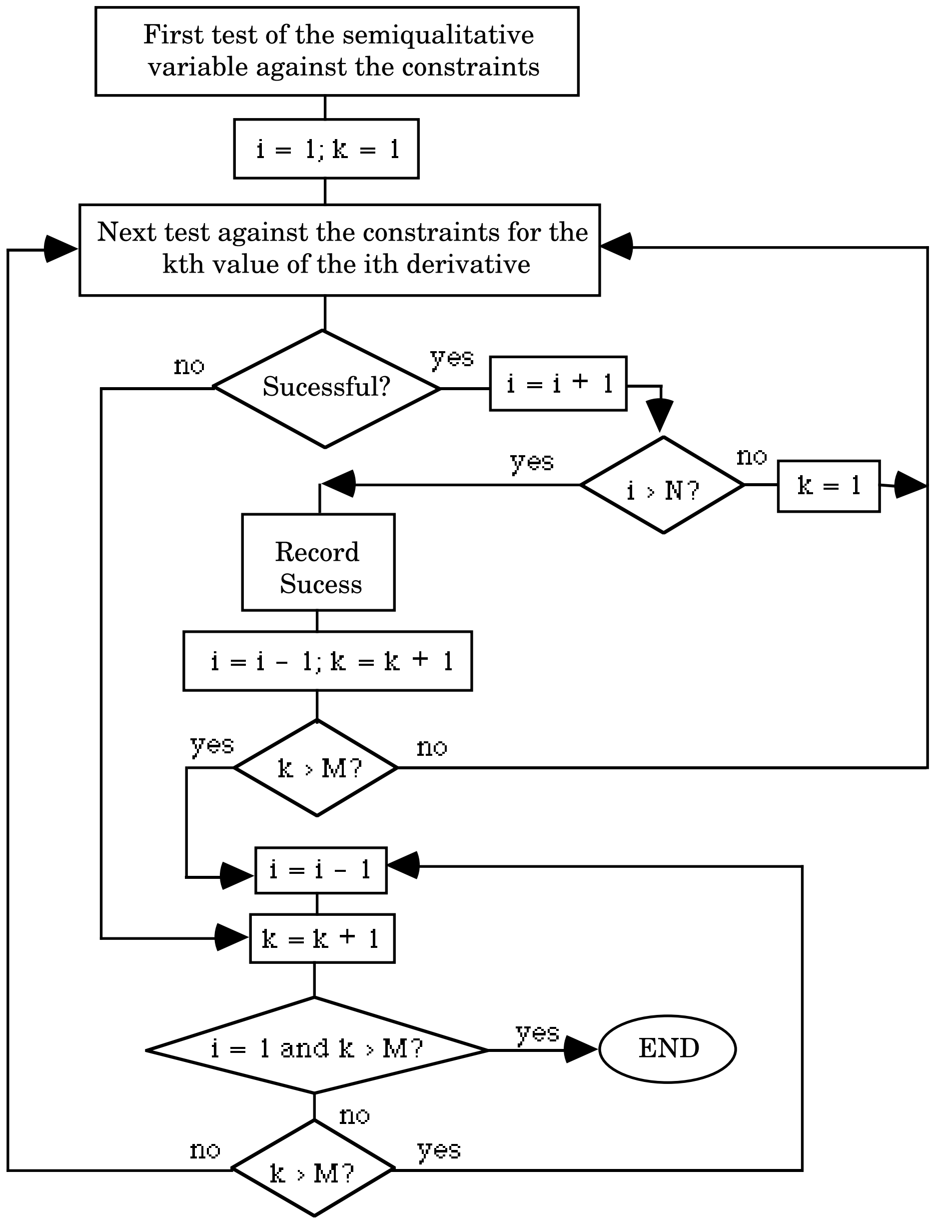


Figure 3. The basic algorithm used by the semiqualitative program

had added every interval to every other, including adding compound intervals such as [20, 500] and [-10, 10] to get [10, 1000]. Clearly this is a lengthy process, and provides a powerful argument for reducing the number of semiqualitative intervals, or at least keeping it as small as possible, but when it is complete the program can construct a look-up table which will provide very swift arithmetic operations later in execution. The program then assembles similar lookup tables for the other functional blocks.

Next the program decides on an order in which to test the values of the variables. This is done in such a way that the most constrained variable has its value propagated first, so that once its value is established the conceivable values of all the related variables may be evaluated as swiftly as possible. Again, establishing this order takes a little time, but it pays dividends in the long run.

After these two steps the program begins the process of propagating the constraints, essentially following the algorithm of Figure 3, for each and every variable, where N is the number of levels of derivative of the variable in question (so $N = 3$ in our example since we have value, first derivative and second derivative), i is the current level of derivative under consideration, M is the number of semiqualitative intervals that the value of the ith derivative can take on, and k is the index of the current value that is being considered for the ith derivative.

Finally, after applying all the constraints to all the variables and establishing their possible value, the system outputs a list of all the possible interval values of all the derivatives of all the variables listed in the final part of the input.

SUMMARY

This paper has described the method of semiqualitative analysis and the application of a program for the semiqualitative analysis of a set of differential equations to a problem from the treatment of waste water. Semiqualitative analysis is an artificial intelligence method that combines the strengths of the human ability to reason about the qualitative behaviour of systems, exemplified by statements such as *"if volume decreases then pressure must increase"*, with the use of whatever numerical data is available. As a result semiqualitative analysis, and the program we have discussed, can be used to model complex physical systems for which sets of differential equations may be written, but for which the exact values of numerical constants are not known. Clearly the method cannot provide exact answers from inexact data, that is impossible, but it does permit the most exact possible answers to be established from the available information and they, as the example illustrated, can often be fairly exact.

ACKNOWLEDGEMENTS

The work of the first author was supported by a grant from ESPRIT Basic Research Action 3085 DRUMS (Defeasible Reasoning and Uncertainty Management Systems). He would, as ever, like to thank all his colleagues on the project for their continual help and encouragement.

REFERENCES

1. Weld, D. and de Kleer, J. *Readings in qualitative reasoning about physical systems* Morgan Kaufmann, San Mateo, California, 1990.
2. Davis, E. *Representations of commonsense knowledge* Morgan Kaufmann, San Mateo, California, 1990.
3. Dohnal, M. 'A methodology for common-sense model development', in Forum on Fuzziness, *Computers in Industry*, pp 1–18, 1991.
4. Raiman, O. 'Order of magnitude reasoning', in AAAI 86 pp 100–104, *Proceedings of the 5th National Conference on Artificial Intelligence*, Philadelphia, PA, 1986.
5. Parsons, S. and Dohnal, M. 'Qualitative, semiqualitative and interval algebras', Technical Report, Department of Electronic Engineering, Queen Mary and Westfield College, 1991.
6. Moore, R. E., *Interval Analysis*. Prentice-Hall, Inc., Englewood Cliffs, N.J. 1966.
7. Bailey, J. E. and Ollis, D. F. *Biochemical Engineering Fundamentals*. McGraw Hill, New York, 1986.