SRC TR 89-43

Constraint Propagation with N-ary Semiquantitative Relations

By

M. L. Mavrovouniotis

Constraint Propagation with N-ary Semiquantitative Relations

Michael L. Mavrovouniotis

Systems Research Center and Chemical Engineering
University of Maryland
College Park, MD 20742, USA

Abstract

An approach for semiquantitative constraint propagation using both simple and complex nodes is presented. Each node has a label consisting of the union of a negative interval and a positive interval. Compared to simple interval labels, this representation provides significant increase in expressiveness, with only a moderate increase in complexity. In addition to simple nodes (variables), there are complex nodes, representing dimensionless products of variables. Previous efforts have focused on reasoning only with independent complex nodes; other nodes can be expressed as suitable algebraic expressions of independent nodes. The approach followed here involves the use of all irreducible complex nodes; these nodes are the simplest possible, in that they cannot be broken into smaller complex nodes. Since irreducible nodes are not necessarily independent, they are related by implicit constraints. The number of nodes and the number of implicit constraints are polynomial in the size of the problem. The coexisting layers of simple and complex nodes can be manipulated to limit the propagation: Labels on complex nodes are only propagated if they contain information that is not already provided by the simple nodes. This effectively reveals those complex nodes that bear interesting labels. These representation and reasoning choices are suited to engineering domains in which many dimensional kinds of variables are present and dimensionless ratios of variables are significant in defining the state of a system.

1 Introduction

There has been extensive investigation of constraint propagation with semiquantitative constraints, involving intervals and inequalities. Davis [1987] provides an excellent review and evaluation of the field.

It is advantageous to infer interval labels for variables, rather than perform constraint inferences, because in the latter it is hard to ensure that the derived constraints will be useful in answering queries. On the other hand, if the queries involve *relations among variables*, the interval labels may not contain that information. Davis [1987] points out that, by using complex nodes (new variables equal to expressions of simple variables), more information can be preserved by the labels, at the cost of increased complexity.

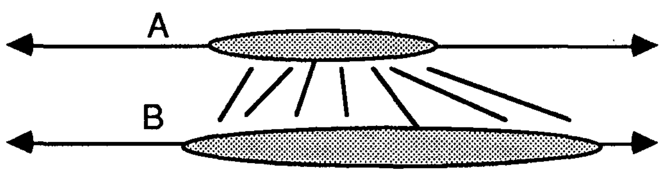

Figure 1:
While A and B vary over wide intervals, they do not vary independently. The two intervals are correlated.

Qualitative reasoning [Bobrow, 1985] with signs (+, –, 0) essentially uses only *simple* nodes, since it focuses on parameter values. Order-of-Magnitude Reasoning [Raiman, 1986, Dague *et al.*, 1987, Mavrovouniotis and Stephanopoulos, 1987 and 1988, Mavrovouniotis *et al.*, 1989], which was introduced to remove some of the inherent ambiguity of qualitative values, focuses completely on complex nodes. It is based on the representation of the relative orders of magnitude of the parameters of a physical system, in the form of binary relations among variables. Reasoning with such relations is based either on a prespecified set of rules [Raiman, 1986] or on rigorous quantitative semantics of the relations [Mavrovouniotis and Stephanopoulos, 1987 and 1988, Mavrovouniotis *et al.*, 1989]. Order-of-Magnitude reasoning enquires about the rough *correlation* between the intervals of variables that have the same physical dimensions (Figures 1 and 3), i.e., can be measured in the same units. Each Order-of-Magnitude relation can be viewed as an area in a plane whose coordinates are the correlated parameters (Figures 2 and 3).

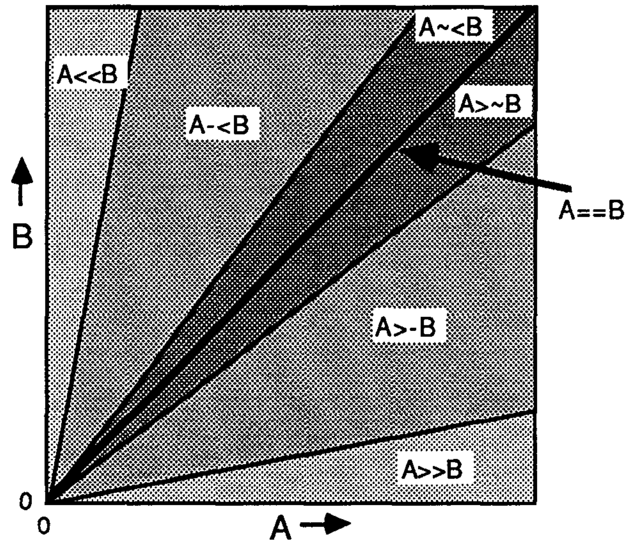

Figure 2:

Order-of-Magnitude relations [Mavrovouniotis and Stephanopoulos, 1987], can be expressed as a particular area on the graph indicating the correlation of A and B.

Drawing on all these ideas, this paper proposes the constraint propagation with N-ary (as opposed to binary) semiquantitative relations, encompassing several novel concepts:

- The obvious options for the representation of semiquantitative labels have so far been single intervals (the standard choice) or arbitrary sets of intervals (avoided by everyone, because of high computational complexity). In this work, **restricted sets of intervals** are used, providing moderate expressive power at a low cost in complexity.

- In the use of complex nodes (i.e., expressions involving the initial variables) the emphasis has been traditionally placed on selecting a set of independent nodes to replace the simple nodes. In this work, **simple and complex nodes** are allowed to coexist. Complex nodes are restricted not to a minimal independent set but to unbreakable or **irreducible nodes**, namely those that cannot be broken into smaller complex nodes.

- The **tightness** of a label on a complex node determines whether there is in that label any information not contained in the simple nodes; only *tight labels* are used in constraint propagation. In the absence of tight labels on complex nodes, only labels on simple nodes are propagated.

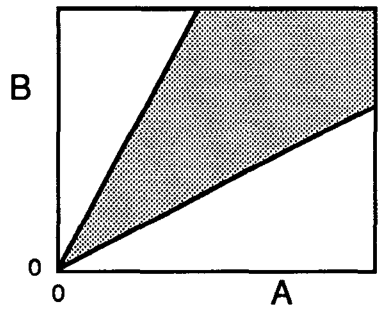

Figure 3:
The correlation of the intervals of A and B can be expressed as a rough dependency of B on A, i.e., as a region in a graph of B as a function of A.

2 Ranges

Systems for reasoning with interval labels have completely avoided the use of sets of intervals, because the number of intervals in the set may grow exponentially: In the worst case, an operation between a label consisting of n intervals and a label consisting of m intervals yields a set of $n \times m$ intervals. The advantage of interval sets is that the result is always correct, i.e., each value of the resulting interval set is indeed a potential outcome of the operation.

In the use of single intervals as labels (whose length is guaranteed constant), the result of every operation must be forced into the form of an interval – the narrowest interval that is a superset of the true result. In general, there may be parts of the resulting interval that could never be outcomes of the operation. For example, if a, b, c, and d are positive and a<b, the most accurate result for the division of (a, b) by (-c, d) is a set (union) of intervals, namely $(-\infty, -a/c) \cup (a/d, +\infty)$; if the result is required to be a single interval, however, the result becomes $(-\infty, +\infty)$.

In the pursuit of moderate accuracy and expressiveness, at a moderate complexity cost, this work employs a limited form of interval sets, called a **range** and defined as a union of three parts: A positive open interval, a negative open interval, and the point zero. One or more of these parts of a range can be empty. Some examples of acceptable ranges include: $((-2,-1) \cup (1,2))$; $((-1,0) \cup \{0\} \cup (0,2))$; $((-100,-10) \cup (10,100))$; and all simple open intervals. The remainder of this section discusses the advantages of the proposed representation for labels.

Knowledge about the absolute magnitude of a variable can be translated into the correct label. The constraint $|x| \in (2, 4)$ can be translated into the range label $x \in ((-4, -2) \cup (2, 4))$, whereas the use of single intervals would yield $x \in (-4, 4)$. Consider an inference on the value of $y=1/x$ from the above labels: Our range would yield $y \in ((-0.5, -0.25) \cup (0.25, 0.5))$, while the use of a single interval would yield $y \in (-\infty, +\infty)$, because the interval for x contains zero.

As was shown in the above example, ranges are more precise for certain arithmetic operations, notably:

◊ Division. For example, in interval arithmetic $(1, 1) / (-1, 1)$ yields $(-\infty, +\infty)$, but with ranges it yields $((-\infty, -1) \cup (1, +\infty))$.

◊ Square roots. The equation $y=x^2$, with $y \in (4, 9)$, yields $x \in (-3, 3)$ in interval arithmetic, but $x \in ((-3, -2) \cup (2, 3))$ with ranges.

Unlike operations on arbitrary sets of intervals, algebraic operations on ranges involve a *constant* number of interval operations. In general, arithmetic operations between two ranges can be performed as follows:

◊ Perform the operation among intervals for all possible combinations of a part for each variable (at most $3 \times 3=9$ combinations, since each range has 3 parts); all empty parts can be ignored.

◊ Separate each of the results into a positive, negative, and zero part.

◊ Take the interval union of all positive parts to form the positive part of the final answer; form similarly the negative and zero parts of the final answer. Since each part of the resulting range must be a simple interval, some information may be lost here. For example, in the addition of $(-2,-1) \cup (1,2)$ to $(-6,-5) \cup (5,6)$, the exact result with arbitrary sets of intervals should be $(-8,-6) \cup (-5,-3) \cup (3,5) \cup (6,8)$. With ranges the intervals must be collapsed to $(-8,-3) \cup (3,8)$. A representation with simple intervals would collapse the intervals further into $(-8,8)$. Thus, the use of ranges offers intermediate accuracy at an intermediate complexity cost.

For specific operations, taking advantage of the structure of the ranges, it is *a priori* known which combinations will form each part of the result. In multiplication, for example, the result will contain zero if, and only if, one or more of the multiplicands contain zero. For the other parts, ignoring the zeroes, the multiplication $((-a_1, -b_1) \cup (c_1, d_1)) \times ((-a_2, -b_2) \cup (c_2, d_2))$, where a_i, b_i, c_i, and d_i are positive, can be performed as follows:

◊ Negative part: $(-\max(-a_1 d_2, -a_2 d_1), -\min(-b_1 c_2, -b_2 c_1))$

◊ Positive part: $(\min(-b_1b_2, -c_1c_2), \max(a_1a_2, -d_1d_2))$

This is equivalent to only two multiplications of intervals. Note that, for all operations, whenever each range contains only one interval the operations can be performed as efficiently as on intervals.

Performing set operations (such as union and intersection) on ranges is only a matter of performing the operations on the positive parts, the negative parts, and the zero parts. In effect, it takes three set operations on intervals to carry out a set operation on ranges.

Sign-independent Order-of-Magnitude relations can be represented by ranges, but not by intervals. For example, under a quantitative semantics, the relation $1 \ll x$ (in any of the two existing Order-of-Magnitude formalisms) implies that either $x > 100$ or $x < -100$. Lumping this range into an interval can yield nothing more than $x \in (-\infty, \infty)$, i.e., no information whatsoever about x. Using a range, we get $x \in ((-\infty, -100) \cup (100, +\infty))$, retaining all the information of the Order-of-Magnitude relation.

One might observe that although zero was chosen as the point that splits the range into its parts, any numbers could have been used. For example, the parts of a range could be defined as an interval below -1, an interval between -1 and +1, and an interval above +1. The choice of zero makes sense because one normally does not deal with arbitrary mathematical formulas, but rather with simple formulas in which zero plays a role more important than any other number. Even if the critical number is not zero, using a set of two intervals may still capture critical information. Consider the equation $y=(x-5)^{-1}$, in which the critical point for x is 5, i.e., it would make sense to use intervals above and below 5. If the information available is that $x \in ((1, 2) \cup (8, 9))$, both ranges and intervals fail to capture the fact that $x \neq 5$. If, on the other hand, $x \in ((-2, -1) \cup (8, 9))$, then a range captures the set exactly.

One may alternatively use labels consisting of a fixed number of intervals without restriction as to their location. This approach merits further consideration. It would differ from ranges in two main points. First, it is not clear how a set of several intervals would be simplified into two intervals. Second, certain operations (such as multiplication) would become less efficient.

3 Variables and Links

A **variable** refers to a specific physical parameter, with known physical dimensions (which define the set of units that can be used for the value of the variable). A range label, defaulting to the trivial infinite range, is attached to each variable. Propagation of these labels through algebraic constraints can be carried out in the usual way. Two variables are *compatible* if they have the same physical dimensions.

A link represents a compatible pair of variables that can be interrelated. They can also represent dimensionless products of *several* variables; specifically, a dimensionless expression of the form:

$$\prod_{i=1}^{k} X_i^{n_i}$$

can be represented by a link, where n_i are the *coefficients* of the link. This kind of combination of variables is assumed to be more important than other arbitrary or restricted expressions (such as linear combinations of compatible variables) in correlating the values of variables. The validity of this assumption depends on the kind of variables and the problem domain. For example, in reasoning about time, durations of events can be compared in pairs through the proposed expressions, but absolute time points cannot. For absolute time points, the use of linear combinations is more appropriate, as it allows indirect reasoning about durations. The expressions adopted here are particularly suitable when relations of an Order-of-Magnitude nature must be expressed.

A range for the dimensional value of each variable is available as the label of the variable. The purpose links serve is the representation of rough **correlations** among parameters, i.e., **relative magnitudes** (or orders of magnitude) of the parameters. This is why the expression of a link is required to be dimensionless. To reduce the number of allowable links, additional restrictions are imposed.

- To eliminate distinct links with coefficients n_i only differing by a constant factor, all n_i must be integers with $\gcd_i(n_i)=1$. This allows the existence of two links whose coefficients differ by a factor of -1, but this redundancy can be removed through additional conventions.

- To ensure a finite number of links, a link's expression is required to be *irreducible*. The link

$$\prod_{i=1}^{k} X_i^{n_i}$$

is irreducible if it cannot be broken down into other dimensionless expressions:

$$\left(\prod_{i=1}^{k} X_i^{n_i}\right)^N = \left(\prod_{i=1}^{k} X_i^{m_i}\right)^M \left(\prod_{i=1}^{k} X_i^{p_i}\right)^P$$

where $N>0$, $M\neq0$, $P\neq0$, and, for all i, $m_ip_iMP\geq0$, $n_ip_iNP\geq0$, $m_in_iMN\geq0$. If the initial expression is reducible, then it cannot represent a link.

The mathematical restrictions on the exponents ensure that none of the dimensionless quantities on the right-hand side of the equation is trivial (equal to one), and that, when they are multiplied together, no cancellations occur: Parameters that are in the numerator for the initial expression, remain in the numerator for each partial expression, and parameters in the denominator remain there.

The restrictions placed on the links behave well only because the labels are ranges, rather than simple intervals.[1]

4 Relations and Tightness

Apart from range labels on variables, which are propagated in the usual way, there are range labels on links denoting **semiquantitative relations**. It was mentioned before that only **tight relations** are propagated. The procedure that determines whether a relation is tight enough to be useful will be discussed below.

Let R_i be the label on each simple variable X_i. Let L be a link:

$$L = \prod_{i=1}^{k} X_i^{n_i}$$

with a current range label R_0. For a newly created link, R_0 has an implicit value derived from the labels of the variables:

[1] If simple intervals were used as labels, the first restriction placed on the links would be troublesome, because the square of a link can contain information not contained in the link itself. Similar problems arise with irreducibility when the exponents M, N, P are even.

$$R_0 = \prod_{i=1}^{k} R_i^{n_i}$$

Let R be a new candidate range for L.

• If $R_0 \subset R$, then the proposed label can be discarded, because it is not tighter than the existing one, and the analysis stops here.

• Otherwise, R is set to $R \cap R_0$, and is tentatively accepted. The new label is then propagated back to yield a potentially tighter label for each variable:

$$R_m = (R / \prod_{\substack{i=1 \\ i \neq m}}^{k} R_i^{n_i})^{1/n_m}$$

◊ If any variable changes label, and can yield a tighter (implicit) label R_0 then whole procedure is repeated, because if the check $R_0 \subset R$ is satisfied, the label R is useless.

◊ If no variable changes label, the new label R is accepted as tight and useful and propagated further.

The idea behind this procedure is that if the simple nodes contain enough information to allow a derivation of the label of the complex node, then this label is not tight enough to be useful. Under this tightness criterion, priority is always given to simple nodes, but when there is additional information not captured by them the complex nodes receive labels. The requirement that relations be tighter than the labels on individual variables limits the propagation of relations; because of the properties of interval operations, the chain of consequents of a relation is eventually terminated by relations that are not tight.

5 Propagation of Implicit Constraints among Links

The manipulation performed before accepting a label on a link involves the use of an implicit constraint relating the complex node (link) to simple nodes (variables). There are other implicit constraints that relate only complex nodes. They can be formed systematically by multiplying links together, then decomposing the product into different links, and equating the two parts.

To limit the number of implicit constraints, multiplication of only *two* links at a time is allowed, and the product is a *single irreducible link*. This implies that the two initial links must be *chained by a variable*: There must be a variable contained with opposite coefficients in the two links. The implicit constraints have the form:

$$\left(\prod_{i=1}^{k} X_i^{n_i}\right)^N = \left(\prod_{i=1}^{k} X_i^{m_i}\right)^M \left(\prod_{i=1}^{k} X_i^{p_i}\right)^P$$

with $m_i p_i MP<0$ for some i (meaning that X_i is the variable that chains the two links of the right-hand side).

In general, the decision that must be made is *how many links are allowed in each implicit constraint*. If this number is high, more information is propagated, at the cost of increased complexity.

6 Algebraic Constraints

When the labels assume values from a finite set, it is possible to perform propagation by checking the validity of a constraint for different combinations of values. Since ranges are continuously varying labels an algebraic constraint must be solved with respect to each variable, forming an *assignment* — an equation whose left-hand side is the variable in question. This is a difficult algebraic manipulation problem.

In a similar fashion, an equation can be manipulated to use and produce information on links: It can be structured in a dimensionless form, so that it relates links rather than simple variables. There are normally many different dimensional and dimensionless forms that each equation can assume. Although the forms are algebraically equivalent, they will behave differently in interval arithmetic (because operations are not distributive). All forms of an equation restructured in this way must be used in parallel, as it cannot be *a priori* predicted which exact form will yield the best results.

If relations for a particular link are desired, one strategy is to introduce the link in all equations that contain at least one variable from the link.

A particularly interesting class of equations is that of *balances*, such as mass balances, energy balances, charge balances, etc. A common form of a balance is:

$$\sum_e \prod_f a_{ef} = \prod_g b_g$$

where, for each f, all a_{ef} have the same physical dimensions. Thus, terms of the left-hand summation are similar in form, and represent inflows and outflows of a conserved quantity. The right-hand side product indicates either accumulation, generation, or consumption (there are sometimes many such terms). Links can be introduced in these equations quite efficiently: For each f, an a_{ef} is chosen as the scale; each term is then divided by the product of the selected scales. Alternatively, each term can be divided by the right-hand side product. An example of proper formation of links and their introduction into algebraic constraints is given in the next section.

After the propagation of all available labels through all constraints, information can be retrieved through queries. A query requests the label of a node (i.e., a variable or a link), which can simply be looked up. If a link does not possess an explicit label, its implicit range can be obtained from the ranges of its variables.

7 An Example

Consider the parameters of the Continuous Stirred-Tank Reactor (CSTR) of Figure 4, converting two reactants, A and B, to a product, P. There are five kinds of variables present:

- Six concentrations of the form C_{Xi}, in mol/m^3, where X is a chemical species (A, B, or P), and i is a stream (0 for the feed and 1 for the effluent of the reactor);
- two flowrates, F_i, in m^3/s, where i is a stream;
- the reaction volume, V, in m^3;
- the reaction rate, r, in mol/m^{3}s;
- the reaction rate constant, k, in m^3/mol s.

These kinds of variables can be combined to form the following irreducible dimensionless links:

(1) $r\,V\,F_i^{-1}C_{Xk}^{-1}$

(2) $r\,k^{-1}\,C_{Xi}^{-2}$

(3) $k\,C_{Xi}\,V\,F_k^{-1}$

(4) $F_i\,F_k^{-1}$

(5) $C_{Xi}\,C_{Yk}^{-1}$

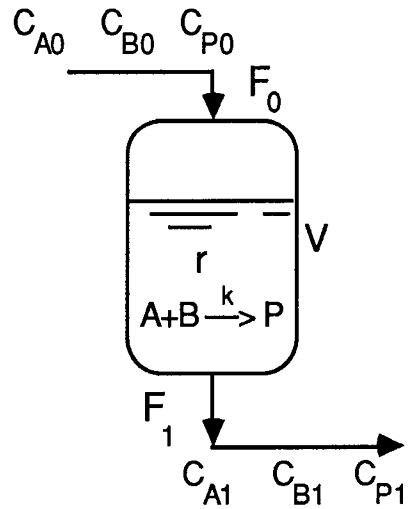

Figure 4:

A Continuous Stirred-Tank Reactor, in which the reaction A+B->P takes place. F_0 and F_1 are flowrates, in m^3/s. C_{Xi} is the concentration of species X (X=A,B,P) in stream i (i=0,1), in mol/m^3. V is the volume of the fluid in the tank, in m^3. r is the reaction rate, in mol/m^3s. k is the reaction rate constant, in $m^3/mol\ s$.

The last two kinds of links compare variables of the same kind; they are thus useful in capturing knowledge about the relative magnitudes of flowrates or concentrations. The first three kinds of links are related to the algebraic equations describing the steady-state behavior of the reactor. The reaction rate is related to concentrations:

$$r = k\, C_{A1}\, C_{B1}$$

This corresponds to a product of 2 links of the second kind. If more reactions were taking place in the reactor, then more of the links would be directly related to algebraic constraints. With only one reaction taking place, the links can still be introduced in the constraint, but they are unlikely to receive tight labels. Mass balances for each of the species A, B, and P yield:

$$F_0 C_{A0} - F_1 C_{A1} - rV = 0$$

$$F_0 C_{B0} - F_1 C_{B1} - rV = 0$$

$$F_0 C_{P0} - F_1 C_{P1} + rV = 0$$

Introduction of links into these mass balances (as discussed at the end of Section 6) leads to constraints that involve links of the first kind, along with links of the fourth and fifth kind. For example, if F_0 is chosen as the scale for flowrates and C_{P1} as the scale for concentrations, the last mass balance is transformed to:

$$C_{P0}C_{P1}^{-1} - F_0\,F_1^{-1} + rV\,F_1^{-1}C_{P1}^{-1} = 0$$

which is a constraint among three links; it is one of five possible transformations of the initial constraint. Given tight relations between F_0 and F_1 and between C_{P0} and C_{P1}, the constraint will yield a tight relation for the link rV/F_1C_{P1}. If F_0 is approximately equal to F_1:

$$F_0/F_1 \in (0.9,\ 1.1)$$

and C_{P0} is much smaller than C_{P1}:

$$C_{P0}/C_{P1} \in (0,\ 0.1)$$

then:

$$rV/F_1C_{P1} \in (0.8,\ 1.1)$$

This important inference can be made even without any knowledge about the ranges of individual parameters.

In this particular example the irreducible dimensionless links tie well with the algebraic equations. Any dimensionless ratio that occurs in an equation is either a link or a product of links. Thus, once an equation is nondimensionalized, it can always be cast in terms of irreducible links.

9 Complexity

The most basic parameters affecting the complexity of the propagation are defined below:

- P is the number of *primitive physical dimensions*. P is usually less than 6, considering that *length, mass, time, temperature, electric current,* and *number of moles* cover most cases. Often, P is even smaller.[1]

---

[1]For the discussion in this section, P should actually be defined as the number of *independent* (rather than primitive) physical dimensions. For example, if a system's parameters are only masses and velocities, the primitive dimensions are mass, length, and time, but length and time are not independent because they only occur in velocity. Thus, P should equal 2. If the system also possesses acceleration parameters, then length and time are independent and P equals 3.

- Q is the number of *dimensionally different kinds of quantities*, e.g:, volume, velocity, force, power, energy. Typically, Q is smaller than 10.
- N is the maximum number of *variables of each dimensional kind.*

The total number of variables is at most QN. Thus, P and Q are bounded, but N increases with the *size* or the *detail* of the physical system being analyzed.

There are two kinds of irreducible links:

- Links that are ratios of two compatible quantities, i.e., links of the form x^1y^{-1}. The number of these links is of order N^2.
- Links that include more than two quantities. It can be shown that these links can include at most P+1 variables, all of which must be of different dimensional kinds. Consequently, the number of these links is of order N^{P+1}.

Thus, for bounded P and Q, the total number of irreducible links is of order N^{P+1}, i.e., polynomial in N (the size of the problem). Consequently, the number of implicit constraints is also polynomial in N and the propagation through implicit constraints among links requires time polynomial in the size of the problem. However, the assimilation of labels is not necessarily complete.

Naturally, depending on the form of the algebraic equations involved, the propagation through explicit algebraic constraints could be much harder. It might require exponential time or it might not converge at all [Davis, 1987].

Various heuristic approaches can be used to limit the propagation through explicit constraints. For example, the number of invocations of an algebraic equation can be restricted, or the propagation can be stopped when new labels differ from old ones by only a small margin (using some appropriate metric).

10 Conclusions

This paper presents some attractive engineering options in the management of redundancy in nodes, labels, and constraints.

The use of ranges, rather than simple intervals, provides significant increase in expressive power, with little complexity burden.

By using physical dimensionality considerations, dimensionless irreducible complex nodes can be formed; they coexist with simple nodes. There is a polynomial number of complex nodes, and implicit constraints can be limited to involve only a certain number of links. This approach avoids the dilemma of independent complex nodes; it makes sense when dimensionally different quantities are present.

To decide which relations are worth propagating, and structure the coexisting layers of simple and complex nodes, the tightness of a relation is examined. Relations on complex nodes are only propagated if they contain more information than the labels on simple nodes. Priority in the propagation is always given to simple nodes, because there is a smaller number of them.

References

Bobrow, D., ed. *Qualitative Reasoning about Physical Systems.* MIT Press (1985). Note: This is the same as Volume 24 of *Artificial Intelligence.*

Dague, P., Raiman, O., and Devès, P. Troubleshooting: When Modeling is in Trouble. In *Proceedings AAAI-87,* p. 600-605, American Association for Artificial Intelligence (1987).

Davis, E. Constraint Propagation with Interval Labels. *Artificial Intelligence* **32**, 281-331 (1987).

Mavrovouniotis, M. L., and Stephanopoulos, G. Reasoning with Orders of Magnitude and Approximate Relations. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87),* p. 626-630, Morgan-Kaufmann, Palo Alto, 1987.

Mavrovouniotis, M. L., and Stephanopoulos G. Formal Order-of-Magnitude Reasoning in Process Engineering. *Computers and Chemical Engineering,* **12**(9/10):867-880, 1988.

Mavrovouniotis, M. L., Stephanopoulos, G., and Stephanopoulos, G. Formal Modelling of Approximate Relations in Biochemical Systems. *Biotechnology and Bioengineering,* in press, 1989.

Raiman, O. Order of Magnitude Reasoning. In *Proceedings AAAI-86,* p. 100-104, American Association for Artificial Intelligence (1986).