Reasoning with Orders of Magnitude and Approximate Relations

Michael L. Mavrovouniotis and George Stephanopoulos

Laboratory for Intelligent Systems in Process Engineering
Department of Chemical Engineering, Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

The O[M] formalism for representing orders of magnitude and approximate relations is described, based on seven primitive relations among quantities. Along with 21 compound relations, they permit expression and solution of engineering problems without explicit disjunction or negation. In the semantics of the relations, strict interpretation allows exact inferences, while heuristic interpretation allows inferences more aggressive and human-like but not necessarily error-free. Inference strategies within O[M] are based on propagation of order of magnitude relations through properties of the relations, solved or unsolved algebraic constraints, and rules. Assumption-based truth-maintenance is used, and the physical dimensions of quantities efficiently constrain the inferences. Statement of goals allows more effective employment of the constraints and focuses the system's opportunistic forward reasoning. Examples on the analysis of biochemical pathways are presented.

1. Introduction

Numerous efforts have been made to apply Qualitative Reasoning to Physical Systems [Bobrow 84]. Major difficulties encountered in the reasoning effort, particularly in engineering applications, stem from the ambiguity inherent [de Kleer and Brown 84] in the qualitative values (-, 0, +) normally used. The incorporation of inequality relations through the quantity-space notion [Forbus 84] only partially resolves the ambiguities.

In engineering, apart from signs of quantities there is more partial knowledge available on rough relative magnitudes of quantities. It is thus desirable to examine ways of introducing more quantitativeness in qualitative reasoning, and employ this type of partial knowledge.

A quantitative approach for digital circuit diagnosis [Davis 84] uses hierarchic representation of time with several time granularities. The longest delay until quiescence at the finer level determines how many fine-grain units correspond to one coarse-grain unit, while events whose duration is shorter than the current granularity level are not represented. A similar concept in qualitative reasoning is *mythical time* [de Kleer 84a], a finer time granularity that can distinguish cause and effect among simultaneous events. Underlying time granularities and mythical time, is the notion of different orders of magnitude in time scales. It was recently pointed out that explicit Order-of-Magnitude reasoning, not just with time scales but with all variables, is the key to successful qualitative reasoning in engineering [Raiman 86], and the FOG formal system was introduced with three basic relations:

- A Ne B: A is negligible in relation to B.
- A Vo B: A is close to B (and has the same sign as B).
- A Co B: A has the same sign and Order-of-Magnitude as B.

The system has 30 rules of reasoning with its basic relations, classical qualitative values, addition, and multiplication. Although FOG is a good initial approach, it fails in several points particularly important in engineering applications:

1. It does not provide concrete semantics. If one does not intuitively understand what "A Co B" means, there is no further explanation available.
2. Its set of rules appears arbitrary, and it is not clear how it can be extended, e.g. to exponentials or integrals.
3. It does not allow incorporation of partial quantitative information, often available in engineering applications. For example, if FOG is told that "A Vo 0.1" and "B Vo 1000" it is unable to infer the obvious "A Ne B".
4. It lumps signs and magnitudes in single relations. The relation "A Co B" carries unnecessary sign connotations: Since the signs are kept track of separately anyway, why should this relation carry sign information? The engineer's intuitive Order-of-Magnitude notion does *not* carry such sign connotations.
5. It requires negation and disjunction to fulfill its reasoning even for very simple problems.
6. It uses knowledge only in the form of rules, and equations involving addition and multiplication.

The problem of applying qualitative reasoning in engineering, we address here with the O[M] formalism for reasoning about orders of magnitudes and approximations. We believe that O[M] lacks the basic faults of FOG described above. We will first describe Order-of-Magnitude relations and their semantics. After we mention the additional concepts of assignments, constraints, and rules, we will discuss how inferences in O[M] are guided and maintained. We will close with examples and a discussion of the O[M]'s potential.

2. O[M] Formalism

A variable in O[M] refers to a specific physical quantity, with known physical dimensions but unknown numerical value. Knowledge about the sign (-, 0, +) of the variable is kept as assertions, termed *sign specs*, stored within the variable. A landmark is similar to a variable, but it has known sign and value.

Variables and landmarks are collectively called *quantities*. Two quantities are *compatible* if they have the same physical dimensions. Within each quantity, there are *links*, each representing a compatible pair of quantities that can be interrelated. A link contains all the Order-of-Magnitude relations asserted between the two quantities, and information on where such relations can be obtained from and where they can be used (e.g. relevant constraints and rules, as we will describe later).

2.1. Primitive and Compound Relations

Order-of-Magnitude relations relate the non-negative magnitudes of quantities, regardless of their sign. Thus, there is no interference between signs and magnitudes, and reasoning with signs can be carried out with the normal qualitative reasoning principles. We introduce seven *primitive* irreducible binary relations among quantities, shown in Table 1.

We accept as a *compound* relation any implicit disjunction of two or more successive primitive relations. It should be emphasized that this restricted disjunction refers mainly to the semantics of the relations, and no *syntactic* disjunction is allowed. There are in total 21 compound relations. The notation for a compound relation produced from the primitives r_n through r_{n+m} is $r_n..r_{n+m}$ The compound relation standing for "A less than B" would be thus represented as "A <<..~< B".

Table 1: Primitive relations of the O[M] formalism

O[M]-RELATION		VERBAL EXPLANATION
r_1:	A << B	A is much smaller than B
r_2:	A -< B	A is moderately smaller than B
r_3:	A ~< B	A is slightly smaller than B
r_4:	A == B	A is exactly equal to B
r_5:	A >~ B	A is slightly larger than B
r_6:	A >- B	A is moderately larger than B
r_7:	A >> B	A is much larger than B

The 7 primitive relations and the 21 compound relations give a set R with a total of *28 legitimate relations* $r_1, ..., r_{28}$. This relation set allows full expressiveness without disjunction or negation. The inverse of every legitimate relation is also a legitimate relation. The negation of a legitimate relation is a legitimate relation if and only if that relation includes either of << or >>.

All of the 28 relations are physically meaningful and each can be given a short and intuitively appealing verbal description. They are powerful enough to express quantity-space partial ordering, all of FOG's relations, and other relations that engineers use in Order-of-Magnitude arguments. Negations of such commonsense relations are usually (but not always) expressible. For example the relation "less than or approximately equal to", frequently used in engineering, is expressed as <<..>~, and its negation as >-..>>. The relation ≈ "roughly equal to" is expressed as ~<..>~, but its negation cannot be expressed. Table 2 shows the correspondence of O[M] relations to commonsense and FOG relations.

2.2. Strict Interpretation Semantics

A relation A r_n B is equivalent to (A/B) r_n 1 and signifies an interval for the (A/B) ratio, as shown in Fig. 1.

To sanction the symmetry of the relations

A >~ B ≡ B ~< A (1)

A >> B ≡ B << A (2)

we impose the restrictions $e_3=1/e_2$ and $e_4=1/e_1$. To sanction the intuition that for A>B>0

A-B << B ≡ A >~ B (3)

we further impose e_1-$1=e_3$.

Under this strict semantics, the above constraints leave only one degree of freedom for the interpretation of our relations, as depicted in Fig. 2. We let the "accuracy" parameter *e* unspecified because it depends on the application domain. In the preliminary design of chemical processes for example, the designer tends to think of *e* between 0.05 and 0.20. On the other hand a physicist would only consider a parameter *e*<0.01. For many domains, this interval semantics (with some particular value for *e*) reflects the way human experts carry out their approximations and Order-of-Magnitude reasoning.

Table 2: O[M]-representation of relations from other systems

CLASSICAL COMMONSENSE RELATIONS	O[M]
less than (<)	<<..~<
less than or equal to (≤)	<<..==
greater than (>)	>~..>>
greater than or equal to (≥)	==..>>
equal to (=)	==
approximately equal to (≈)	~<..>~
less than or approximately equal to (≲)	<<..>~
greater than or approximately equal to (≳)	~<..>>
much less than	<<
much greater than	>>
FOG RELATIONS	**O[M]**
Negligible in relation to (Ne)	<<
Very close to (Vo)	~<..>~
Comparable to (Co)	-<..>-

With this clear semantics there is no need for prespecified rules since they can be derived from the intervals, which moreover allow incorporation of quantitative information. We named this interpretation strict because its solid intervals support only accurate correct inferences. For any primitive or compound relation the corresponding interval is continuous. The intervals produced from inferences are also continuous and the consequent relations can be expressed without disjunctions.

2.3. Heuristic Interpretation Semantics

The strict interpretation is accurate, but too strict compared to human reasoning. For example, from the relations A >~B and B >~ C the strict interpretation can only conclude A >~..>- C while human

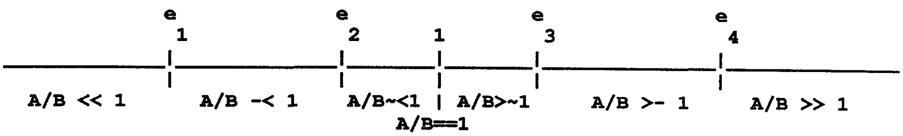

Figure 1: Strict interpretation of the relation A r_n B

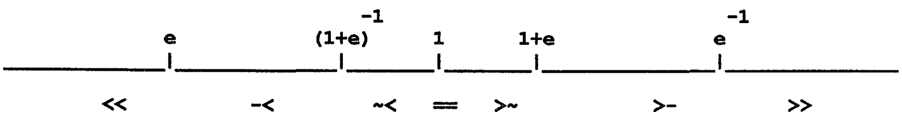

Figure 2: Constrained strict interpretation of the relation A r_n B

commonsense would aggressively conclude A >~ C. Clearly the latter result is heuristic. It is not guaranteed correct, but it is correct often enough for an engineer to be happy with. Any mechanism that can accommodate this result will *have* to accept the risk of wrong conclusions as the price for more aggressive inferences. Note that FOG sanctions even the more aggressive inference

(A Co B and B Ne C) → A Ne C (4)

a subcase of which in our notation would be

(A >- B and B << C) → A << C (5)

We feel this inference is too aggressive and error-prone, so we choose not to sanction it.

The heuristic interpretation we adopt replaces the boundary points of the intervals with regions (Fig. 3). We then construct two sets of primitive intervals: A set of non-exhaustive intervals and a set of overlapping ones, shown in Fig. 4. The following heuristic inference convention is adopted: For every inference step, assume the antecedent relations to denote non-exhaustive intervals, but allow the consequent relations to denote overlapping intervals. Thus, when the consequents are used as antecedents at a later step their intervals are "shrunk" and therein lies the power and the risk. Note that for compound relations this mechanism refers only to the end points of the compound intervals (i.e. the compound intervals do not have "holes").

The good properties that were mentioned for the strict interpretation are preserved by this transformation, with the exception of lost guaranteed accuracy of the inferences. The heuristic inference procedure resembles closely human reasoning. In the previous example it would infer

A >~ B and B >~ C → A >~ C (6)

Once an inference is made, people use the consequent without reconsidering its uncertainty and would infer further

A >~ C and C >~ D → A >~ D. (7)

Hence the "shrinking" of the expanded intervals when a consequent is used further.

To choose the new interval boundaries, we sanction the symmetry of the relations, as before, and the following inferences:

A >~ B → A-B << B (8)

A >~ B and B >~ C → A >~ C (9)

A >~ B and A >> C → B >> C . (10)

The interval boundary regions in the final form are shown in Fig. 5.

The exact choice of *e* depends on the domain of application. A very large number of inferences are valid regardless of the value of *e*. Apart from inferences based on addition and subtraction, this group also includes inferences with other functions:

x << 1 → exp(x) >~ 1+x (11)

x << 1 → sin(x) >~ x . (12)

2.4. Assignments, Constraints, and Rules

Assignments are "solved" algebraic relations that allow some quantities to produce relations among other quantities. The left hand side of an assignment can be either a ratio (link) of two quantities, or just a single variable.

The right hand side of an assignment, called *expression*, cannot be any arbitrary algebraic expression. It can involve only links, landmarks and numerical constants. The system attempts to automatically convert algebraic expressions to the acceptable form. The success of the automatic parsing depends on the form of the algebraic expression.

Constraints are "unsolved" algebraic relations among quantities. As with assignments, there are requirements on the form of the expressions, and the system attempts automatic conversion.

The first way to use constraints, is to simply "test" them, and accept or reject assumptions based on the outcome. The second is to form a set assignments by solving the constraint in all obvious ways. By "obvious" solutions we mean simply getting hold of one occurrence of a variable in the expression and solving with respect to that, regardless of its other occurrences. The O[M] system can apply automatically both approaches.

Knowledge of highly empirical nature often cannot be expressed in algebraic form. O[M] can accept knowledge in the form of simple if-then rules without free variables.

3. Control of Reasoning

We will briefly describe here how the system maintains consistency and how it expands and prunes the inference tree. The basic strategy of O[M] is depth-first data-driven reasoning. Any new fact is first checked for redundancy, created and used immediately, regardless of whether the use of its "parent" has been completed. It invokes all possible scenarios for further reasoning:

1. From the conjunction of relations new relations are inferred and redundant ones are retracted. From the symmetry and transitivity of relations new relations are inferred.
2. For relations between a variable and a landmark, numeric transitivity is applied. The idea is that if we find another variable related to another landmark compatible to the original one we can infer a relation between the two variables.
3. When a relation can serve as the antecedent of rules, the rules are invoked.
4. When a relation (actually its link) participates in the expression of assignments or constraints, these are invoked. Applying an assignment can yield knowledge about the magnitude as well as the sign of a variable.

In the domain of chemical engineering (our primary interest) there are many different kinds of variables present: temperatures, pressures, volumes, flowrates, masses, concentrations, etc. The requirement that only compatible quantities can be linked reduces

Figure 3: "Fuzzy" interval boundaries for the heuristic interpretation

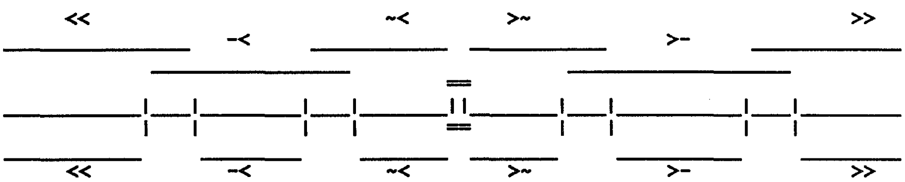

Figure 4: Overlapping intervals (top), and non-exhaustive intervals (bottom) for the heuristic interpretation

the search space in all inference scenarios.

3.1. Truth-Maintenance and Resolution of Contradictions

Assertions can be stated as assumptions rather than known facts. They can also be stated as dependent on assuming of other assertions. For each inference step then we form the assumption set under which the conclusion is valid and allow several relations between two quantities to coexist.

Assumption-based Truth-Maintenance is carried out using de Kleer's ATMS approach [de Kleer 84b], which avoids some serious problems of other truth-maintenance systems that use dependency-directed backtracking. In ATMS there is no backtracking involved, and important assumption sets can be parsed after the main problem-solving effort.

The resolution of contradictions requires more care in O[M] because with the heuristic interpretation, neighboring relations that apparently conflict may actually be both valid heuristically (since neighboring heuristic intervals are overlapping). We will delineate here the alternative ways of handling apparent contradictions.

The first way is to forbid any special treatment of neighboring conflicting relations. This would cause all kinds of assumption sets and eventually the whole problem (i.e. the empty assumption set) would be marked inconsistent, without being truly so.

The second way is to simply allow neighboring relations to coexist, and mark them in a special way as non-conflicting. Since they will both propagate, this aggressive strategy amounts to implicitly asserting that indeed the overlapping part of the two neighboring intervals represents the "true" relation.

The third and most conservative way is to disclaim both relations (and mark them to avoid recurrence of the problem) and replace them by the compound relation representing their disjunction. If the initial relations are compound one need only consider the two primitive components (one from each initial relation) that are neighboring and take their disjunction.

We can try to take advantage of these pseudo-contradictions in the special case where one of the quantities involved is a variable and the other a landmark. After we apply the third strategy outlined above, we can assume that the true relation was indeed in the overlapping part of the intervals, select or create another compatible landmark, and relate it to the variable by a tighter relation.

3.2. Goal Direction

The search mode for O[M] is opportunistic forward chaining, but there are two ways to induce search for a particular relation. By stating that the goal is to relate two particular quantities, the user can induce additional ways to use constraints and assignments. Whenever one of the two goal quantities occur, the system uses the other one as well (for example, it divides both sides of the constraint by that variable).

Alternatively, the user may state that alternative relations between two quantities should be examined. Then, the system can create seven assumptions, one for each of the seven primitive relations and check them for consistency with available knowledge.

4. Implementation

The implementation of the O[M] system was done in Symbolics Common LISP, on Symbolics 3650 computers, running the Genera 7.0 environment. The Flavors Object-Oriented Programming system was heavily employed. All entities (quantities, relations, constraints, etc.) are implemented as objects.

Each of the simple problems on which O[M] was tested (such as reasoning about a single equipment piece or a three-reaction segment in a biochemical pathway) was handled in at most a few seconds. We have not yet tested the system on complex problems. Having many assumptions slows the system down, because expensive set-operations are required by ATMS. This problem can be remedied by using ordered data structures for assumption sets [de Kleer 84b].

5. Reasoning about Biochemical Pathways

The expressive power of O[M] is illustrated by the following relations involving sizes of molecules of biochemical interest.

- Enzymes have much larger Molecular Weight than small molecules: M_E >> M_S.
- In turn, H^+ has much smaller Molecular Weight than any other compound of biochemical interest: M_{H^+} << M_S.
- The molecular radius of an enzyme is only moderately larger than that of a small molecule (other than H^+): r_E >- r_S.
- For the molecular radius of H+: r_{H^+} << r_E and r_{H^+} -< r_S.

A higher concept in the analysis of biochemical pathways is that of the rate-limiting step of biochemical pathways, the "bottleneck" that limits the overall observable rate of the pathway.

For a linear pathway $P=\{r_1, r_2, \ldots, r_n\}$, where r_i is the i^{th} bioreaction of the pathway, K_i is the equilibrium constant of r_i, Q_i is the mass action ratio of r_i, a consistent observable rate-limiting step $H_L=r_L$ is a member of P such that the following relations are consistent with all knowledge available on P:

$\forall i \in [1,L-1]$: K_i >~ Q_i,

$\forall i \in [L+1, n]$: K_i >~..>> Q_i, and

K_L >> Q_L.

As a specific application, we will examine three consecutive reactions from the pathway of Glycolysis, to test the hypothesis that

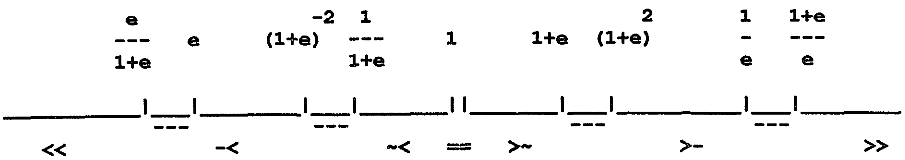

Figure 5: Final intervals for the heuristic interpretation

the first step is rate-limiting. We abbreviate Fructose Diphosphate as FDP, Dihydroxyacetone Phosphate as DHAP, Glyceraldehyde Phosphate as GAP, reduced and oxidized Nicotinamide Cofactors as NADH and NAD, Inorganic Phosphate as PI, Hydrogen Cations as H, and Diphosphoglycerate as DPG. The steps of interest are:

1. FDP $\rightarrow$ GAP + DHAP
2. DHAP $\rightarrow$ GAP
3. GAP + NAD + PI $\rightarrow$ DPG + NADH + H

The knowledge we have here is:

- Algebraic definition of the mass action ratios, [e.g. for the first reaction: G1 = (GAP DHAP) / FDP] and the catabolic reduction charge [CRC = NADH / (NADH + NAD)].
- Constant values for H, PI, CRC,and equilibrium constants (KE1, KE2, KE3).
- All concentrations are of the order of 100 μM.
- For the reactions to proceed in the specified direction, the mass-action ratios must be larger than the equilibrium constant [e.g. G3 >~..>> KE3].
- The goal to pursue relations among GAP, DPG, and landmark concentrations.
- The hypothesis that the first step is rate-limiting: G1 >> KE1.

In this example, O[M] would use the knowledge we provided to conclude that the hypothesis, that the first step is rate-limiting, is inconsistent. O[M] first narrows the range of GAP, using the first two reactions. The assumption yields that GAP -< 100μM. Propagating this through the last reaction step O[M] obtains DPG << 100μM, which conflicts with the given relation DPG -<..>- 100 μM.

6. Discussion

In the real world, there are always many positive and negative effects on any aggregate result. An intelligent approach in dealing with them, must concentrate on deciding which of the effects are important and which not. Only then should it attempt to determine the sign of the overall result. The O[M] formalism is aimed exactly at sorting out dominant effects.

Even in quantitative reasoning people use Order-of-Magnitude arguments to reduce algebraic complexity. This is often done systematically: As terms are dropped from equations, a term of the form O(x) does the bookkeeping, denoting that the largest dropped term is "of order x". Numerical constants are *not* introduced in the O(x) term. This type of reasoning resembles the O[M] formalism with the understanding that we keep track of orders O(e) and we additionally distinguish between O(e) and O(-e), but terms of order O(e^2) or higher are neglected.

The risks of the aggressive heuristic interpretation were pointed out earlier. Indeed the worst-case behavior of the strategy is miserable, but for real-world cases it performs much better. There is an additional safeguard in normal use of the strategy: We are normally interested in the relations ~<, >~, <<, and >> which are separated by the "buffer" regions -< and >-. It takes extremely bad cases for the error to propagate through the whole buffer region and convert e.g. a >~ to a >>.

We believe the O[M] formalism bridges the gap between traditional qualitative reasoning (with signs) and full quantitative reasoning (with numbers), as it can use mixed (quantitative and qualitative) knowledge. It will be suitable in many domains where extensive knowledge is naturally expressible in Order-of-Magnitude relations, especially since it is capable of handling numerical and algebraic knowledge as well.

REFERENCES

[Bobrow 84] Bobrow, D.G.
Qualitative Reasoning about Physical Systems: An Introduction.
Artificial Intelligence 24:1-7, 1984.

[Davis 84] Davis, R.
Diagnostic Reasoning Based on Structure and Behavior.
Artificial Intelligence 24:347-410, 1984.

[de Kleer 84a] de Kleer, J.
How Circuits Work.
Artificial Intelligence 24:205-280, 1984.

[de Kleer 84b] de Kleer, J.
Choices Without Backtracking.
In *Proceedings AAAI-84*, pages 79-85. American Association for Artificial Intellegence, 1984.

[de Kleer and Brown 84]
de Kleer, J., and Brown, J.S.
A Qualitative Physics Based on Confluences.
Artificial Intelligence 24:7-83, 1984.

[Forbus 84] Forbus, K.D.
Qualitative Process Theory.
Artificial Intelligence 24:85-168, 1984.

[Raiman 86] Raiman, O.
Order of Magnitude Reasoning.
In *Proceedings AAAI-86*, pages 100-104. American Association for Artificial Intellegence, 1986.