Automated Model Switching

Daniel S. Weld

Technical Report 89-08-01
Draft of August 6, 1989

Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195

ABSTRACT

Although computers are widely used to simulate complex physical systems, crafting the underlying models that enable computer analysis remains a difficult job with only a small number of computer tools for support. When a model is created for one task, it is often difficult to reuse the model for another purpose because each analytic task requires a different set of simplifying assumptions. Through the use of explicit representation of modeling assumptions and qualitative reasoning techniques, we are developing a theory of automated model selection and validation and have partially implemented the ideas in the SAM system.

Contents

1 Introduction 1
- 1.1 Three Phases of Reasoning . . . 1
- 1.2 Validation . . . 1
- 1.3 Model Switching . . . 2
- 1.4 Approximations . . . 2
- 1.5 Scenario . . . 3
- 1.6 Roadmap . . . 4

2 Preliminaries 5
- 2.1 Parameters . . . 5
- 2.2 Constraints . . . 6
- 2.3 Models . . . 6
- 2.4 Behaviors . . . 7

3 Multiple Models 7
- 3.1 Reformulations . . . 8
- 3.2 Behavior Difference . . . 9
- 3.3 Approximation Reformulations . . . 10
- 3.4 Summary . . . 12

4 Inter-Model Comparative Analysis 12
- 4.1 Qualitative Behavior . . . 13
- 4.2 Relative Change Values . . . 13
- 4.3 Relative Change of a Model Switch . . . 14
- 4.4 Reformulating Relative Changes . . . 15
- 4.5 Exploiting Approximation Reformulations . . . 15
- 4.6 Generalizations to the Reduction Theorem? . . . 19

5 SAM 19

6 Related Work 20

7 Future Work 21

1 Introduction

Although the bulk of work in model-based reasoning has focussed on problems of analysis in the framework of a single model, we believe that no single model can be adequate for a wide range of tasks. This observation applies as much to people as intelligent machines: large, monolithic models are as inconvenient for human experts as they are intractable for computers. Imagine trying to diagnose a misbehaving automobile with a single, flat, molecular-level description of the whole system. Quantum mechanics may be the right level to reason about bond angle in impure hydrocarbon fuels, but it does not provide a useful model of the pistons, spark plugs, or transmissions.

To achieve robust performance when reasoning about complex systems, analytic programs must do what human experts do: switch between models, dynamically choosing perspectives and simplifying assumptions that are appropriate to the task at hand. Just as a human engineer disregards details of the differential gears when diagnosing a leaking coolant system, model-based reasoning systems must be able to choose the right level of abstraction for a model before undertaking analysis. But different assumptions are warranted depending on the analytic question being answered. A reasoner must simplify but not oversimplify. A program that uses multiple models should validate its choice in the context of the problem at hand. Thus the critical step is enabling the program to reason explicitly about modeling assumptions.

1.1 Three Phases of Reasoning

We have developed an extended framework of computer analysis that is composed of three phases: choice of a model, problem solving in that model, and validation that the model is appropriate. If validation instead shows that the model was an inappropriate choice then a new model must be selected. The bulk of our work involves a novel technique for performing this model switch: generate and test using inter-model comparative analysis.

At present we assume that domain dependent heuristics are used for the initial choice of model. We further assume that once a model has been selected, the problem solving or analysis performed is quantitative in nature. In this paper, we consider behavioral prediction (i.e., numerical simulation) as the sole type of analysis. (Our model switching technique, however, is qualitative.)

1.2 Validation

We assume that model validation is achieved through a combination of internal consistency rules and direct observation.

- Internal consistency checks are the prefered method for validation. They consist simply of rules associated with a set of assumptions (equivalently with the corresponding set of models). After quantitative analysis, the rules are run to check consistency. If a rule detects a contradiction, the rule is responsible for specifying which assumption has been violated. Reasoning then switches to a model without that assumption. For example, suppose a Newtonian model is used to solve a question about a collison between several particles. Afterwards, a rule could check a simulation trace to see if any velocity approached the speed of light. If a high velocity were detected, reasoning would switch to a relativistic model.

- Direct observation, comparing the values predicted with those actually measured, is the only verification mechanism that approaches completeness. If a measurement differs significantly[1] from the value predicted, a DISCREPANCY results. Of course, no finite set of discrepancy-free measurements can guarantee validity, but the probability of error can be lowered asymptotically close to zero.

Although internal consistency checks have many advantages (speed, simplicity, and the ability to specify a corrective model), they have two problems as well. First, for the same reason that physical systems have no perfect model, it is impossible to generate a complete set of consistency checks. Second, they are domain dependent so the burden is on a human modeler to provide a comprehensive set of rules.

To guarantee robust performance, an automated reasoner can not rely exclusively on consistency checks. Direct observation must also be used, but this raises the problem of model switching. Direct observations can provide discrepancies, but it does not dictate which model will alleviate the problem. This paper addresses that task: given a discrepancy, choose a model that will produce predictions in closer accordance.

1.3 Model Switching

We present a simple, domain-independent algorithm for discrepancy-driven model switching: generate and test. Our program is given a GRAPH OF MODELS (GoM) [16, 13, 2, 2], i.e. a directed graph in which nodes represent models of the system at hand and edges connecting nodes are labeled with the set of simplifying assumptions that distinguish the two models. The G&T generator suggests all neigboring models in the GoM that eliminate at least one assumption held by the current model. The testor determines whether a candidate model can account for the discrepancy. If it can, then reasoning switches to the new model.

One can imagine a variety of testor algorithms, but the weakness of the generate and test strategy requires that the testor be very efficient. This paper shows how a qualitative reasoning technique, inter-model comparative analysis, can be used to screen candidate models.

Inter-model comparative analysis provides a qualitative answer to the question "What will be the effect on behavior of switching from one model of a system to another?" In general, inter-model comparative analysis is a very difficult problem, but in many cases it can be done efficiently. In particular, this paper shows that if the relationship between the two models can be formalized as an APPROXIMATION then the problem of inter-model comparative analysis reduces to that of intra-model comparative analysis. This means that the well-studied techniques of DQ analysis [24, 23] and exaggeration [27, 25] can be applied.

1.4 Approximations

Thus, the central idea of this paper is that of an approximation. The basic idea is that a simple model approximates a more complex model if the complex

[1] An appropriate definition of 'significantly' is an interesting problem. Acceptable error is a function both of the reliability of the measuring technology and of the task under consideration (i.e. what are the consequences of an error).

model has a parameter, called a FITTING PARAMETER, such that the quantitative behaviors predicted by the two models get arbitrarily close as the fitting parameter goes to a limit.

For example, consider the domain of simple mechanical devices. Suppose that the simple model is frictionless and the complex model represents friction as the product of the normal force times a coefficient of friction U. In this case, the simple model approximates the complex model with fitting parameter U — as the coefficient of friction tends towards zero, the behaviors predicted by the two models converge.

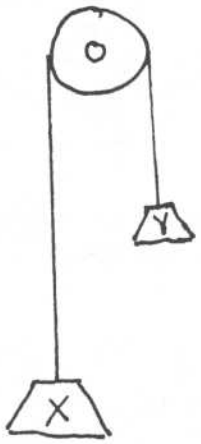

Figure 1: Two unequal weights hung over a pulley.

A more complex example results from figure 1. Many simplifying assumptions are possible, three of which are: massless pulley, massless rope, stretchless rope. Adopting the first assumption results in a model that approximate a model in which pulley-mass *is* represented; the fitting parameter is the mass of the pulley. The last assumption can also be seen as an approximation. Consider the model of an elastic rope which represents the rope as a spring obeying Hooke's law: $F = -KX$. Increasing the value of K results in a stiffer (less elastic) rope. And as K goes to infinity, the behaviors predicted by the two models converge. The next section shows how our implementation, SAM, will reason about this system.

1.5 Scenario

Suppose that SAM were directed to simulate the simple system shown in figure 1. At time zero the blocks are released, and the objective is to predict the velocities of the two blocks when block X hits the table. Figure 2 shows the GoM generated by the three independent assumptions mentioned above: I means inelastic rope, R means massless rope, and P means massless pulley.

Without domain-specific rules to the contrary, SAM chooses the simplest model (i.e., $\{IRP\}$) for reasoning. Using quantitative data and an ODE description of the system, suppose that SAM's numerical simulation routines predict the values S_x and S_y for the speed of the two blocks respectively.[2] SAM

[2]For expositional simplicity, we consider *speed*, not velocity.

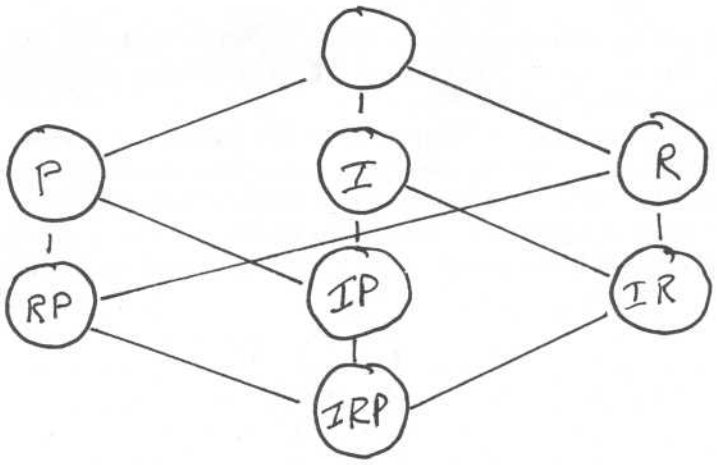

Figure 2: GoM for pulley system. For simplicity, only edges corresponding to changes in a single assumption are shown.

verifies the predictions by direct observation and finds a discrepancy: both X and Y are predicted to be moving too quickly.

The model-switching generator suggests (blindly) that perhaps the problem could be corrected by considering stretch in the rope. The testor prepares to perform inter-model comparative analysis to determine the chance in predicted velocities between models $\{IRP\}$ and $\{RP\}$. Since the former model approximates the latter, intra-model comparative analysis can be performed using spring constant K as a fitting parameter. DQ analysis is used to determine the effect of a more elastic spring $K\Downarrow_0$ and it is discovered that $S_x\Uparrow_1$ and $S_y\Downarrow_1$ because the extra stretch in the rope affects the two velocities differently.

Suppose the (stupid) generator next suggests retracting the assumption of massless rope. The testor performs inter-model comparative analysis on the switch from $\{IRP\}$ to $\{IP\}$. Again, this is an approximation reformulation, so intra-model comparative analysis can be performed using the rope mass as a fitting parameter. DQ analysis determines that the result is $S_x\Uparrow_1$ and $S_y\Uparrow_1$ because the extra length of (heavy) rope on the left happens to accentuate the difference in weight between X and Y.

Now the generator suggests retracting the assumption that the pulley is massless. Inter-model comparative analysis on the $\{IRP\}$ to $\{IR\}$ switch is done by DQ analysis of an increase in pulley mass in the $\{IR\}$ model. The results, $S_x\Downarrow_1$ and $S_y\Downarrow_1$ match the discrepancy so quantitative reasoning resumes in the $\{IR\}$ model. This time the predictions of numerical simulation are within tolerance, so SAM is confident of this model for similar problems in the future.

Note that SAM performs intra-model comparative analysis on a single fitting parameter solely for search control reasons. The theory allows for the retraction of multiple assumptions by simply composing the initial perturbations.

1.6 Roadmap

The next section defines necessary terms for use in the remainder of the paper: parameter, model and behavior. Section 3 makes precise the notion of a reformulation between two models and defines the important class of approxi-

mations. Section 4 presents the CA Reduction Theorem which enables efficient computation of inter-model comparative analysis between models related by an approximation. Section 5 describes the status of the SAM implementation, and section 6 discusses connections to related work.

2 Preliminaries

We consider a model to be a description of a physical system in terms of one or more parameters. Loosely speaking, some of the parameters act as inputs to the model which takes these initial conditions and produces values for all parameters over an interval of time. Since only these input parameters can be directly altered by external action, a notion of causality is established. Our models produce two kinds of behavioral descriptions: quantitative and qualitative representations of the change of parameteric values over time. We assume that the quantitative descriptions are of primary interest to the human user; the qualitative representation will be used to perform the model switching. This framework for reasoning is elaborated below.

2.1 Parameters

We describe systems with functions called PARAMETERS. All parameters are assumed to be continuous, continuously differentiable functions from an interval of the reals into an interval of the extended reals $[-\infty, \infty]$ and have only a finite number of points where the derivative crosses zero in any bounded interval [10]. The intent is that the domain of a parameter is time and the range is a value of interest. For example, the V parameter might denote the velocity of an object over time; $V(t_0)$ would thus denote the object's velocity at a particular time. The RANGE function maps from parameters to their ranges. Thus for a relativistic model, RANGE(V) might return $(-c, c)$, the open subinterval of $\Re^*$ where c denotes the speed of light.

Three kinds of parameters are distinguished: independent, boundary, and dependent. To specify the state of a system, independent parameters must be assigned values for all times in question. However, for simplicity we assume that all independent parameters are constant over time, so only a single value is necessary. Boundary parameters are a superset of the independents — once assigned values for some initial time (i.e., boundary conditions), they completely specify the state of the system. The remaining parameters are dependent — their values at a time point are completely determined by the values of the boundary parameters.

For example, consider the familiar horizontal, frictionless spring / block system described by $MA = -KX$. Mass M and spring constant K are typically independent parameters — their value must be specified for all times, an easy task if we assume that they are constant. Position X and velocity V are boundary parameters but not independent since knowledge of just the initial position and velocity is sufficient to specify a unique behavior. Acceleration A (and force, if defined) are dependent parameters. Let us emphasize that the selection of independent, boundary and dependent parameters is an underconstrained modeling issue and must be done by a human.

2.2 Constraints

To specify the interdependence between parameters in a physical system, models contain qualitative and quantitative constraints. By quantitative constraints we mean simply a system of ordinary differential equations (ODEs). A model's qualitative constraints are a finite set of instantiations of the six constraints used by QSIM [10]: ADD, MINUS, MULT, M^+, M^-, and $\frac{d}{dt}$. These constraints have the meaning one would expect: $P = \text{ADD}(Q, R)$ means $P(t) = Q(t) + R(t)$ for all times t, and $P = M^+(Q)$ means that P is a monotonicly increasing function of Q. For a formal definition see [10]. Naturally, it is important that the quantitative and qualitative descriptions are mutually coherent.

Definition 1 *A set of qualitative constraints* AGREES *with a set of ordinary differential equations (ODEs) iff every solution to the ODEs satisfies the constraints.*

For example, the qualitative constraint $P = M^+(Q)$ agrees with

$$P(t) = 15.3e^{2Q(t)}.$$

Note that a given set of qualitative constraints may agree with many ODEs. This is simply another way of stating that a qualitative description is an abstraction of a quantitative representation.[3]

2.3 Models

We define a model as a vector of parameters related by qualitative and quantitative constraints that agree.

Definition 2 *Let* $(P_1, \ldots, P_n)$ *be an ordered list of parameters. Let* C *be a set of qualitative constraints defined over* $\{P_i\}$. *Let* D *be a set of ordinary differential equations over* $\{P_i\}$. *Say that*

$$\mathcal{A} = ((P_1, \ldots, P_n), \text{C}, \text{D})$$

is a MODEL *if* C *agrees with* D *and* D *specifies a unique solution, when each parameter is restricted to its range.*

Let PARAM *be a function taking a model to the list of parameters for the model. Let* BOUND *be a function taking a model to the sublist* $(P_1, \ldots, P_l)$ *of boundary parameters. Let* INDEP *be a function taking a model to the sublist* $(P_1, \ldots, P_k)$ *of boundary parameters, where* $1 \leq k \leq l \leq n$.

We use calligraphic letters to denote models, lower case letters to denote real numbers, and capital letters to denote parameters. All parameters are numbered so we will frequently talk about the i-th parameter of a model as P_i, but when discussing a particular model we may use mnemonic names like V for velocity. When it is useful to emphasize that parameter P is part of model $\mathcal{A}$, we write it as ${}_{\mathcal{A}}\text{P}$. For simplicity, this definition of model ignores the issue of multiple operating regions [10, p294] which are useful for describing many complex systems.

[3] For the next few pages we are primarilly concerned with quantitative values and constraints, but qualitative constraints are necessary to perform intra-model comparative analysis (section 4.5).

2.4 Behaviors

Given a model, we wish to describe the behavior it predicts over time. Both quantitative and qualitative descriptions are necessary; we start with the former.

Definition 3 *Let $\mathcal{A}$ be a model with parameters $P_1, \ldots, P_n$ of which the first k are boundary parameters. An* INTERNAL STATE *of $\mathcal{A}$ is a k-tuple $\vec{p} = (p_1, \ldots, p_k)$ such that $p_i \in$ RANGE(P_i) forall $1 \leq i \leq k$. A* STATE *of $\mathcal{A}$ is an n-tuple such that the values $p_1, \ldots, p_k$ are an internal state and $p_{k+1}, \ldots, p_n$ are in their ranges and are consistent with the values for the boundary parameters under the model's quantitative constraints. A set of* INITIAL CONDITIONS *for $\mathcal{A}$ is an internal state of $\mathcal{A}$. The* BEHAVIOR *of $\mathcal{A}$ given initial conditions $\vec{p}$ is the unique function*

$$\mathcal{A}_{\vec{p}} : \Re \rightarrow \text{RANGE}(P_1) \times \ldots \times \text{RANGE}(P_n)$$

defined by

$$\mathcal{A}_{\vec{p}}(t) = (P_1(t), \ldots, P_n(t))$$

where the P_i are solutions to the model's ordinary differential equations given the boundary values $\vec{p}$.

Thus a model $\mathcal{A}$ is an abstract description of a system. A state is a snapshot of the values of all the model's parameters at a given time. Combining a model $\mathcal{A}$ and a set of initial conditions $\vec{p}$ specifies a behavior $\mathcal{A}_{\vec{p}}$ that maps from times to states. Given a behavior or a state, one can use a projection function to isolate the parameter or parameter value of interest. For example, to extract the i-th parameter from the $\mathcal{A}_{\vec{p}}$ behavior, one would write $\pi_i(\mathcal{A}_{\vec{p}})$. If mnemonic names are used then the parameter name may be substituted in place of the index. For example, to determine the velocity (parameter V) specified by a state $\vec{p}$, one would write $\pi_V(\vec{p})$.

The definition above describes the relationship between a model, initial conditions and the resulting behavior but it does not say anything about how to compute the behavior. This is deliberate. Our objective is a general theory of model shifting that is independent of particular solution, simulation, symbolic algebra, or numeric approximation methods. The particular techniques used by the SAM implementation are described in section 5.

In addition to the quantitative behavior of a model, it is often useful to describe time-varying behavior qualitatively. To this end, we adopt Kuipers' QSIM representation as summarized in section 4.1.

3 Multiple Models

Suppose we have several models of the same system that embody different simplifying assumptions. We wish to characterize the difference between the behaviors they predict. Since two models may describe a physical system using different parameters, some work is necessary to enable behavioral comparison. In this section we show how REFORMULATION FUNCTIONS can be used to align distinct models. We discuss when the behavior of two comparable models can be considered equivalent. Finally, we consider a restricted class of reformulations, called APPROXIMATIONS, that have useful properties. Then in section 4 we discuss the problem of inter-model comparative analysis and describe an efficient technique for performing the analysis if one of the models approximates the other.

3.1 Reformulations

To provide a way to match descriptions in two different models we introduce a notion of a reformulation. The basic idea is that we can compare a complex model $\mathcal{B}$ to a simpler one $\mathcal{A}$ if an internal state of $\mathcal{B}$ allows us to construct a complete description of an internal state of $\mathcal{A}$. Although this notion is very general (almost any invertible, continuous function, meaningful or not, is a reformulation), it provides a useful foundation. Sections 3.2 and 3.3 refine the idea to a useful class of reformulations called approximations.

Definition 4 *Let $\mathcal{A}$ and $\mathcal{B}$ be models with n and m parameters such that* BOUND$(\mathcal{A}) = \{P_1 \ldots P_k\}$ *and* BOUND$(\mathcal{B}) = \{Q_1 \ldots Q_l\}$. *If there exists a continuous function Ψ' from* RANGE$(Q_1) \times \ldots \times$ RANGE(Q_l) *onto* RANGE$(P_1) \times \ldots \times$ RANGE(P_k) *then say that Ψ* COMPARES *$\mathcal{B}$ to $\mathcal{A}$ (written $\mathcal{A} \preceq_\Psi \mathcal{B}$) where Ψ is an extension of Ψ' that maps from states (rather than internal states) of $\mathcal{B}$ to states of $\mathcal{A}$ in the obvious way. Ψ is called a* REFORMULATION FUNCTION *from $\mathcal{B}$ to $\mathcal{A}$. For any state $\vec{q}$ of $\mathcal{B}$, if $\vec{p} = \Psi(\vec{q})$ then $\vec{p}$ is said to be the* CORRESPONDING STATE *of $\vec{q}$.*

For example, let $\mathcal{B}$ be a model of the two dimensional motion of a billiard ball using polar coordidates and let $\mathcal{A}$ be a model of the same system using rectangular coordinates. In this case $\mathcal{A} \preceq_\Psi \mathcal{B}$ because a reformulation function exists. Let Ψ be defined from $\{\theta\} \times \{R\}$ to $\{X\} \times \{Y\}$ as follows

$$\Psi(\theta, R) = (R\cos\theta,\ R\sin\theta).$$

For the rest of this paper, however, we assume that all reformulation functions are defined in terms of simple arithmetic operations (addition, subtraction, multiplication, and division). In fact, for many examples it suffices to specify trivial reformulations that equate parameters pairwise in the two models.

Note that this definition allows many possible reformulations between two nonempty models, most of which are uninteresting or irrelevant. Meaningful comparision between two models requires a good choice of Ψ, and much of this paper is concerned with characterizing useful classes of reformulations.

Proposition 1 *The compared-to relation $\preceq_\Psi$ is reflexive and transitive but not symmetric.*

Proof: Reflexivity is obvious from the existence of trivial (identity) reformulation functions. Transitivity follows from the fact that functional composition of reformulations preserves continuity and invertibility. To demonstrate the lack of symmetry, let $\mathcal{B}$ be a model with two parameters P and Q and $\mathcal{A}$ be a model with one parameter R. Assume that all of the parameters range over the whole real line. Define Ψ as a projection $\Psi = \pi_1$. Since Ψ is continuous and onto, Ψ is a reformulation and $\mathcal{A} \preceq_\Psi \mathcal{B}$. But an indirect proof shows $\mathcal{B} \not\preceq \mathcal{A}$. Suppose $\mathcal{B} \preceq_\Phi \mathcal{A}$. Then there exists some reformulation Φ from the reals onto the real plane. Consider Φ restricted to $\Re - \{0\}$. The image of this function is connected since removing one point from the real plane does not disconnect it, thus by continuity $\Re - \{0\}$ must be connected. But this is absurd. Thus Φ cannot be continuous and onto. Hence $\mathcal{B} \not\preceq \mathcal{A}$, and the compared-to relation is not symmetric. □

Intuitively this means that one can compare a "large" model to a "smaller" one but not vice versa. The lack of symmetry results from a reformulation mapping from a subset of one model's state-space to the other model's complete

state-space. For example, suppose that $\mathcal{B}$ is a model of two noninteracting oscillators (one blue and one red) and $\mathcal{A}$ is a model of the red oscillator, described using the same parameters as in $\mathcal{B}$. Clearly $\mathcal{B}$ can be compared to $\mathcal{A}$, since a reformulation could map all the $\mathcal{B}$ parameters describing the red oscillator into the the equivalent parameters in $\mathcal{A}$. An alternate (less meaningful) reformulation would map the $\mathcal{B}$ parameters describing the blue oscillator into the $\mathcal{A}$ parameters describing the red one. And, of course, it is possible to imagine a reformulation that mapped some several red and several blue parameters into a senseless $\mathcal{A}$ description of the red oscillator. However, despite the abundance of reformulations from $\mathcal{B}$ to $\mathcal{A}$, it is impossible to compare $\mathcal{A}$ to $\mathcal{B}$ because there isn't enough independent information in a state of $\mathcal{A}$ to flush out a complete $\mathcal{B}$ description that isn't redundant. Thus $\mathcal{B} \not\preceq \mathcal{A}$.

Another example begins to demonstrate the utility of this reformulation definition. Suppose $\mathcal{B}$ is a model of an oscillator assuming damping with frictional coefficient ${}_{\mathcal{B}}\mathrm{U}$ and $\mathcal{A}$ is simpler model of the same oscillator that does not account for damping. Although these two models predict different behaviors, $\mathcal{A} \preceq_\Psi \mathcal{B}$. In other words a description of the state of $\mathcal{B}$ lets one compute all the parameters of $\mathcal{A}$, but $\mathcal{B} \not\preceq \mathcal{A}$ since the value of ${}_{\mathcal{B}}\mathrm{U}$ can not be calculated from parameters in $\mathcal{A}$.

Proposition 2 *Let $\mathcal{A}$ and $\mathcal{B}$ be models with k and l boundary parameters respectively. $\mathcal{A} \preceq_\Psi \mathcal{B}$ iff $k \leq l$.*

Proof: This is an easy corollary of the Borsuk-Ulam theorem [12, p170]. □

Although this isn't a very strong constraint on reformulations (i.e. many uninteresting reformulations exist) it does reinforce the intuition that complex models have more independent parameters than simple models. The next two sections develop stronger (and more useful) constraints on reformulations based on their behavior.

3.2 Behavior Difference

Now that we have a way to connect descriptions in two separate models we are ready to define the difference in the models' predicted behaviors. We can define the behavior difference for any reformulation, but of course the difference will only be interesting if the reformulation is meaningful.

For the purposes of model switching we are mainly interested in a qualitative measure of the difference in predicted behaviors. However, a quantitative measure of this difference will also prove useful. In both cases we define the difference in terms of the parameters of the simpler of the two models (i.e. in terms of $\mathcal{A}$ if $\mathcal{A} \preceq_\Psi \mathcal{B}$) because of the inherent asymmetry of reformulations.

Definition 5 *Let $\mathcal{A}$ and $\mathcal{B}$ be models with* $\textsc{param}(\mathcal{A}) = (P_1, \ldots, P_n)$. *Let $\vec{p}$ be a state of $\mathcal{A}$ and $\vec{q}$ be a state of $\mathcal{B}$. Suppose that Ψ is a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Define the* DIFFERENCE IN P_i BETWEEN $\vec{p}$ AND $\vec{q}$ USING Ψ *as*

$$\textsc{pdiff}(P_i, \Psi, \vec{p}, \vec{q}) = \pi_i(\Psi(\vec{q})) - \pi_i(\vec{p})$$

In other words the difference in the value of a parameter in the two states is calculated by using the reformulation to convert the complex-system state into a corresponding simple-system state. Then the π_i projection functions extract the i-th parameter value from the two states and the difference is returned.

Using this definition we can now describe when two models make equivalent predictions. The intuition is that there should be no difference in predicted behavior given equivalent initial conditions. But it is important to recognize that while the difference in parameter values must be measured in the simpler model, the initial conditions must be specified in the more complex model to ensure that both models can be simulated.

Definition 6 *Let $\mathcal{A}$ and $\mathcal{B}$ be models with* PARAM$(\mathcal{A}) = (P_1, \ldots, P_n)$. *Let Ψ be a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Let $\vec{q}$ be an internal state of $\mathcal{B}$ representing a set of initial conditions. Let $\vec{p}$ be the internal state of $\mathcal{A}$ corresponding to the state $\Psi(\mathcal{B}_{\vec{q}}(0))$. Define the* BEHAVIOR DIFFERENCE BETWEEN $\mathcal{A} \preceq_\Psi \mathcal{B}$ OVER THE TIME INTERVAL $[t_s, t_f]$ GIVEN $\vec{q}$ *as*

$$\text{BDIFF}(\Psi, \vec{q}, t_s, t_f) = \max_{1 \le i \le n} \left(\sup_{t \in [t_s, t_f]} |\text{PDIFF}(P_i, \Psi, \mathcal{A}_{\vec{p}}(t), \mathcal{B}_{\vec{q}}(t))| \right)$$

If the BDIFF *is zero, then we say that the behaviors are* EQUIVALENT *over the interval $[t_s, t_f]$.*

In other words, for each parameter in the simple model, we compare corresponding complex values for all times and take the supremum (least upper bound) of the absolute differences. The behavior difference is the maximum value of the suprema.

A more flexible definition of behavioral difference over an interval of time would use perspectives [24, 23]. However, we defer these definitions since they are unnecessary for this paper.

3.3 Approximation Reformulations

In this section we present the most important idea in this paper, a restricted class of reformulations called approximations. Intuitively, one model approximates another when the behavior difference between them can be brought arbitrarily close to zero. This class of reformulations will prove to be very important. While inter-model comparative analysis is a difficult problem in general, it can be performed quite efficiently for models related by approximation reformulations.

Definition 7 *Let $\mathcal{A}$ and $\mathcal{B}$ be models, and suppose there exists a reformulation Ψ such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Say that $\mathcal{A}$* APPROXIMATES *$\mathcal{B}$* UNDER *Ψ if there exists a parameter $Q_f \in$* INDEP$(\mathcal{B})$ *and an endpoint l of the closure of* RANGE(Q_f) *such that for all internal states $\vec{q}$ of $\mathcal{B}$,*

$$\lim_{\pi_f(\vec{q}) \to l} \text{BDIFF}(\Psi, \vec{q}, 0, \infty) = 0$$

In this case, the parameter Q_f is called the FITTING PARAMETER *of Ψ and l is called its* APPROXIMATION LIMIT.

Since a fitting parameter is independent by definition, it is constant over time; this is why the definition only refers to its initial value $\pi_f(\vec{q})$. The idea behind the definition is that $\mathcal{A}$ approximates $\mathcal{B}$ if $\mathcal{B}$'s fitting parameter squeezes the behavior difference to zero as it goes to a limit. Some examples will clarify the definition.

Consider the simple system shown in figure 3. At time zero the block is released at the top of the θ degree inclined plane; under the force of gravity, it

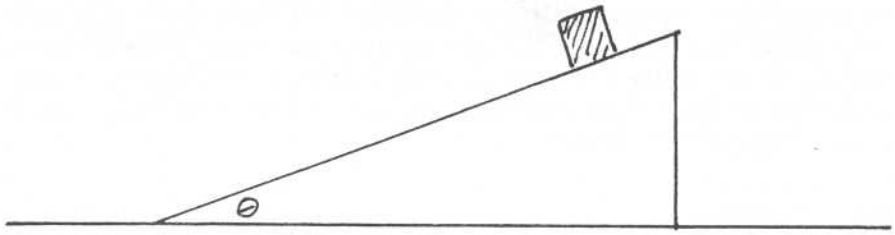

Figure 3: Block slides down an inclined plane.

moves downward (and to the side, but both models will ignore the horizontal component of movement). Let $\mathcal{A}$ be a model of this system with parameters T, G, Y, V, A denoting θ, gravity, height, and the vertical components of velocity and acceleration respectively. Let $\mathcal{B}$ be a model with all these parameters plus a coefficient of friction U. Let Ψ be the projection function:

$$\Psi(\theta, g, y, v, a, \mu) = (\theta, g, y, v, a)$$

It is clear that Ψ is a reformulation so $\mathcal{A} \preceq_{\Psi} \mathcal{B}$.

Before we can consider the behavior of $\mathcal{A}$ and $\mathcal{B}$, we must specify more details of the models. INDEP($\mathcal{A}$)=$\{G, T\}$ and BOUND($\mathcal{A}$) = $\{G, T, Y, V\}$. G and Y have range $[0, \infty)$ and T has range $[0°, 360°)$. The ODEs for $\mathcal{A}$ are:

$$\begin{aligned} A &= G\cos(T) \\ A &= \tfrac{d}{dt} V \\ V &= \tfrac{d}{dt} Y \end{aligned}$$

For the slightly more complex model, INDEP($\mathcal{B}$)=$\{G, T, U\}$ and BOUND($\mathcal{B}$) = $\{G, T, U, Y, V\}$. U has range $[0, \infty)$. The ODEs for $\mathcal{B}$ are:[4]

$$\begin{aligned} A &= G\cos(T) - UG\sin(T) \\ A &= \tfrac{d}{dt} V \\ V &= \tfrac{d}{dt} Y \end{aligned}$$

As the coefficient of friction U tends to zero, the frictional force diminishes and the first equation of $\mathcal{B}$ gets arbitrarily close to the first equation of $\mathcal{A}$. When there is no friction then the equations are identical so it is clear that the behavior difference is zero. Thus, we can say that $\mathcal{A}$ approximates $\mathcal{B}$ with fitting parameter U and approximation limit 0.[5]

While the system of figure 3 provides a clear example of approximating models, it is a bit misleading. The case where the fitting parameter can actually take on the limiting value (i.e. where it is legal for $\mathcal{B}$ to have zero friction) is really a degenerate case of the approximation definition. In general, this is not the case, and this is why the definition allows l to be in the closure of the

[4] These equations assume that the block is moving relative to the wedge, otherwise the frictional force will be smaller. This condition could be expressed as a range restriction on theta and the coefficient of friction, a range restriction on acceleration, or as an extension to the equation defining frictional force. We refrain from these details since they are irrelevant to the point at hand.

[5] In fact, either gravity or wedge angle θ could be used as a fitting parameter in this example.

parameter's range.[6] For example, consider two models of tension in a string. A simple model $\mathcal{A}$ might model the string as inelastic while the more complex $\mathcal{B}$ could use Hooke's law $F = -KX$ to model the elongation of the string under tension. As the spring coefficient K tends to infinity, the behaviors predicted by the two models become equivalent. Thus $\mathcal{A}$ approximates $\mathcal{B}$ even though ∞ is not a legal value for K in $\mathcal{B}$.

3.4 Summary

In this section, we showed how reformulations allowed a complex model to be compared to a simpler one. We characterized the difference in behaviors predicted by two models, and we isolated a particular class of reformulations, called approximations, where the behavior difference can be made arbitrarily small.

Our overall goal is to help automate the modeling process. Given the behavior predicted by a model, and a set of discrepancies between the values predicted and those observed, we wish to determine which other models would predict behaviors in closer agreement with the observations. As described in section 1, we advocate generate and test. This test must take two models and predict the effect on behavior of the switch from one model to the other. If the representation for observation discrepancies is qualitative, then the test can be implemented as a form of inter-model comparative analysis. Although inter-model comparative analysis is very difficult in general, the next section will show that it can be done efficiently if the reformulation is an approximation. In fact, it reduces to an intra-model comparative analysis problem and can be solved by DQ analysis [24, 23] or exaggeration [25, 27, 26]. Then in section 5 we present our implementation of these ideas. In section 6 we discuss related work, and in section 7 we describe issues for future research.

4 Inter-Model Comparative Analysis

Suppose we have two models of the same system that are related by a reformulation. We wish to characterize the difference between the behaviors they predict. By assuming one model as 'current' and considering a shift in models, we phrase this question as a comparative analysis problem: "What is the effect on predicted behavior of shifting from the current model to a different one?" But several kinds of comparative analysis are possible: intra-model comparative analysis predicts how the model's behavior will be affected by a perturbation in the value of some boundary parameters. Inter-model comparative analysis, on the other hand, describes how the behaviors predicted by two different models compare.

Clearly one could answer an inter-model comparative analysis question by comparing numerical approximations of solutions to the two models' sets of differential equations, but this is a laborious process. Instead, we consider qualitative techniques. The first step is to define qualitative behavior. Then we introduce the relative change language for describing differences in behaviors. Finally we show how inter-model comparative analysis reduces to the intra-model case if the reformulation is an approximation.

[6] The restriction that l be an endpoint of the closure is explained in section 4.5.

4.1 Qualitative Behavior

For simplicity, we adopt Kuipers' QSIM representation as summarized below. See [10, 24] for complete definitions.

A QUALITATIVE BEHAVIOR is a sequence of QUALITATIVE STATES alternating between states at time points and states that hold over open intervals of time. A qualitative state describes the qualitative value and qualitative derivative of each parameter. For example, the qualitative value of a parameter, P, at a time point t, is defined as the ordinal relationship between the parameter's real value at that point and its nearest LANDMARKS (special values of significance to the human modeler). This is written as follows:

$$\text{QVAL}(P,t) = \begin{cases} p_j & \text{if } P(t) = \text{landmark } p_j \\ (p_j, p_{j+1}) & \text{if } P(t) \in (p_j, p_{j+1}) \end{cases}$$

The sign of the parameter's derivative (known as qualitative direction) is represented symbolically:

$$\text{QDIR}(P,t) = \begin{cases} \text{inc} & \text{if } \frac{d}{dt}P(t) > 0 \\ \text{std} & \text{if } \frac{d}{dt}P(t) = 0 \\ \text{dec} & \text{if } \frac{d}{dt}P(t) < 0 \end{cases}$$

A collection of qualitative value and derivative pairs for every parameter forms the qualitative state of the model at a time point:

$$\text{QS}(\mathcal{A},t) = (\ldots, (\text{QVAL}(P_i,t), \text{QDIR}(P_i,t)), \ldots)$$

The qualitative state for a time interval is comparable. When any parameter's QVAL or QDIR changes, the parameter (and also the system as a whole) are said to TRANSITION. In other words, transitions are events when qualitatively significant changes happen in the value of a parameter. Thus each adjacent pair of states in a model's qualitative behavior represents a transition. Since it is often useful to be able to refer to transitions independent of the time at which they occur, the sequence of transitions for a qualitative behavior is denoted by the set $\{\gamma_i\}$. Every qualitative behavior also has a TIME FUNCTION, $\mathcal{T}$, which takes transitions to the points in time when they occur. Hence the qualitative behavior of a model can be written as:

$$\text{QS}(\mathcal{A}, \mathcal{T}(\gamma_0)), \text{QS}(\mathcal{A}, (\mathcal{T}(\gamma_0), \mathcal{T}(\gamma_1)), \text{QS}(\mathcal{A}, \mathcal{T}(\gamma_1)), \ldots$$

Just as our theory of model switching is independent of any particular quantitative solution method, it does not depend on a particular algorithm for generating qualitative descriptions. However, we note in passing that Kuipers' QSIM program [10] generates the set of possible qualitative behaviors for a model given a set of qualitative initial conditions. In the ideal case, QSIM produces a single behavior, but often ambiguity causes multiple behaviors to be generated. These are represented as a STATE TREE where every path through the tree represents a possible qualitative behavior. As described in section 5, our implementation uses QSIM.

4.2 Relative Change Values

To compare two behaviors (qualitative or quantitative), they must be distinguishable. In intra-model comparative analysis a hat accent is used to denote

a perturbed, second behavior. Thus $\widehat{V}$ might denote velocity in the second system, and $\widehat{V}(\widehat{\mathcal{T}}(\gamma_i))$ denotes the second system's value of velocity at the time (in the second system) of the i-th transition. To simplify the problem of comparative analysis, we assume that the behaviors are TOPOLOGICALLY EQUAL [24], i.e. they have identical sequences of transitions.

The relative change language is used to describe the difference between two behaviors of a single model. For values at transition points the definition is straightforward.[7]

Definition 8 *Let $\mathcal{A}$ be a model with* PARAM$(\mathcal{A}) = (P_1, \ldots, P_n)$ *as determined by initial conditions $\vec{p}$. For any transition γ_i in the qualitative behavior of $\mathcal{A}$, define the* RELATIVE CHANGE (RC) *of P given $\vec{p}$ at γ_i as:*

$$\text{RC}(P, \vec{p}, \gamma_i) = \begin{cases} \Uparrow & \text{if } \widehat{P}(\widehat{\mathcal{T}}(\gamma_i)) > P(\mathcal{T}(\gamma_i)) \\ \| & \text{if } \widehat{P}(\widehat{\mathcal{T}}(\gamma_i)) = P(\mathcal{T}(\gamma_i)) \\ \Downarrow & \text{if } \widehat{P}(\widehat{\mathcal{T}}(\gamma_i)) < P(\mathcal{T}(\gamma_i)). \end{cases}$$

These are written ${}^{\vec{p}}P\Uparrow_i$, ${}^{\vec{p}}P\|_i$, and ${}^{\vec{p}}P\Downarrow_i$.

This definition can be extended to handle comparisons over time intervals through the use of perspectives [24], but this complexity is unnecessary for this paper.

Relative change values are closely related to partial derivatives.

Proposition 3 *Let $\mathcal{A}$ be a model with* BOUND$(\mathcal{A}) = (P_1, \ldots, P_k)$ *and suppose* PARAM$(\mathcal{A}) = (P_1, \ldots, P_n)$ *as determined by the initial conditions $\vec{p}$. Suppose ${}^{\vec{p}}P_c\Uparrow_0$ for some c such that $1 \leq c \leq k$, and that ${}^{\vec{p}}P_j\|_0$ for all j such that $1 \leq j \leq k$ and $j \neq c$. Let P_e denote a dependent parameter $k+1 \leq e \leq n$. For all transitions γ_i if $\mathcal{T}(\gamma_i) = \widehat{\mathcal{T}}(\gamma_i)$ then*

$$\text{RC}(P_e, \vec{p}, \gamma_i) = sign\left(\frac{\partial P_e}{\partial P_c}(\mathcal{T}(\gamma_i))\right)$$

Proof: This is a direct corollary of proposition 15 in [24]. □

4.3 Relative Change of a Model Switch

Inter-model comparative analysis seeks to determine the effect on predicted parameter values of a switch from one model to another. These effects can be described using the following simple extension of the pointwise relative change notation.

Definition 9 *Let $\mathcal{A}$ and $\mathcal{B}$ be models and let Ψ be a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Let $P \in$* PARAM$(\mathcal{A})$. *Let $\vec{q}$ be an internal state of $\mathcal{B}$, and let $\vec{p}$ be the internal state corresponding to the state $\Psi(\mathcal{B}_{\vec{q}}(0))$. If the behaviors of $\mathcal{A}$ and $\mathcal{B}$ exhibit the same seqeunce of transitions given these initial conditions, Define the* RELATIVE CHANGE *of P* OVER *$\mathcal{A} \preceq_\Psi \mathcal{B}$ given $\vec{q}$ at transition γ_i as:*

[7] There are two differences between this definition and that used in [24, 23]. First, this definition uses signed semantics rather than the magnitude semantics; in other words we compare signed quantities rather than absolute values. Theoretically, the two approaches are equivalent, but we prefer signed semantics for this paper because of greater commonality with the quantitative definition of parameter difference. Secondly, in this paper we choose to make explicit the dependence on initial (boundary) conditions.

$$\text{RC}(P, \Psi, \vec{q}, \gamma_i) = \begin{cases} \Uparrow & \text{if } \text{PDIFF}(P, \Psi, \mathcal{A}_{\vec{p}}(_{\mathcal{A}}\mathcal{T}(\gamma_i)), \mathcal{B}_{\vec{q}}(_{\mathcal{B}}\mathcal{T}(\gamma_i))) > 0 \\ \| & \text{if } \text{PDIFF}(P, \Psi, \mathcal{A}_{\vec{p}}(_{\mathcal{A}}\mathcal{T}(\gamma_i)), \mathcal{B}_{\vec{q}}(_{\mathcal{B}}\mathcal{T}(\gamma_i))) = 0 \\ \Downarrow & \text{if } \text{PDIFF}(P, \Psi, \mathcal{A}_{\vec{p}}(_{\mathcal{A}}\mathcal{T}(\gamma_i)), \mathcal{B}_{\vec{q}}(_{\mathcal{B}}\mathcal{T}(\gamma_i))) < 0. \end{cases}$$

These are writte ${}^{\vec{q}}_{\Psi}P\Uparrow_i$, ${}^{\vec{q}}_{\Psi}P\|_i$, *and* ${}^{\vec{q}}_{\Psi}P\Downarrow_i$ *respectively.*

In other words, we define the multi-model version of relative change by comparing a parameter's value in a state of the simple model using Ψ to compute the corresponding value from the complex model. ${}_{\Psi}P\Uparrow_i$ means that switching from the simple model $\mathcal{A}$ to the more complex model $\mathcal{B}$ will cause P to get a larger predicted value given $\vec{\gamma}$ and its corresponding internal $\mathcal{A}$ state as initial conditions.

4.4 Reformulating Relative Changes

We are almost ready to tackle inter-model comparative analysis, but one observation must be made. Note that the relative change values are simply a form of the well-studied sign algebra $\{[+], [0], [-]\}$ [17, 28]. Because of this and because reformulation functions are defined solely with simple arithmetic operators, reformulation functions can be used to compute corresponding relative change values. In other words, given relative change values for all parameters in $\mathcal{B}$ at a transition, it is possible (with some potential loss of information due to ambiguity) to calculate a corresponding set of relative change values for $\mathcal{A}$. In fact, no information is lost in the most common case where the reformulation is a projection function.

Definition 10 *Let $\mathcal{A}$ be a model with* $\text{PARAM}(\mathcal{A}) = (P_1, \ldots, P_n)$ *as specified by initial conditions $\vec{p}$. For any transition γ_i, define*

$$\text{RCS}(\mathcal{A}_{\vec{p}}, \gamma_i) = (\text{RC}(P, \vec{p}, \gamma_i), \ldots, \text{RC}(P, \vec{p}, \gamma_i))$$

In other words, RCS takes a model's behavior and a transition and returns the vector of relative change values at that transition.

Suppose $\mathcal{A}$ and $\mathcal{B}$ are models with n and m parameters respectively. If $\mathcal{A} \preceq_{\Psi} \mathcal{B}$ we can extend the reformulation function Ψ to map from $\{\Uparrow, \|, \Downarrow\}^m$ to $\{\Uparrow, \|, \Downarrow\}^n$ using the standard qualitative interpretation of the arithmetic functions defining the reformulation. (For brevity, we will not include the full definition here).

Thus if $\text{RCS}(\mathcal{B}_{\vec{q}}, \gamma_i)$ is a vector of relative change values for model $\mathcal{B}$, an application of the reformulation $\Psi(\text{RCS}(\mathcal{B}_{\vec{q}}, \gamma_i))$ denotes the same changes in terms of the parameters of $\mathcal{A}$. To determine the corresponding relative change value for a specific $\mathcal{A}$ parameter P_j simply use the π_j projection function. Although we are overloading the Ψ function with two purposes (mapping the real values of parameters and mapping qualitative relative change values), the meaning is always clear because one can tell the usage from the type of the argument to Ψ.

4.5 Exploiting Approximation Reformulations

In general, inter-model comparative analysis appears quite difficult. Short of performing a complete numerical simulation of the two models and comparing the quantitative results (an expensive undertaking), there does not seem to be a solution.

However, if the reformulation linking the two models is an approximation, then inter-model comparative analysis reduces to an intra-model comparative analysis problem with an initial RC of the fitting parameter away from the approximation limit.

First note that if $\mathcal{A}$ approximates $\mathcal{B}$ then their time functions converge as the fitting parameter approaches the approximation limit.

Proposition 4 *Let $\mathcal{A}$ and $\mathcal{B}$ be models and let Ψ be a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$ and $\mathcal{A}$ approximates $\mathcal{B}$ under Ψ with fitting parameter Q_f and approximation limit l. Suppose that the closure of* RANGE$(Q_f) = [l, k]$. *Let $\vec{q}$ denote an internal state for $\mathcal{B}$ and let $\vec{q}$ the the corresponding internal state for $\mathcal{A}$. If there exists some open interval $U = (l, l + \rho) \subset [l, k]$ such that the behaviors of $\mathcal{A}_{\vec{p}}$ and $\mathcal{B}_{\vec{q}}$ exhibit the same sequence of transitions whenever $\pi_f(\vec{q}) \in U$, then forall γ_i*

$$\lim_{\pi_f(\vec{q}) \to l} {}_{\mathcal{B}}\mathcal{T}(\gamma_i) = {}_{\mathcal{A}}\mathcal{T}(\gamma_i).$$

Proof: This is a simple consequence of the approximation definition. Clearly some specification of topological equality is necessary since otherwise quantifying over transitions makes no sense. In fact, it is necessary that the behaviors be topologically equal in the vicinity of the approximation limit (not just at the limit), since otherwise discontinuities could result. □

Now to see how an inter-model comparative analysis problem reduces to the intra-model case consider the problem of predicting the velocity of the block in figure 3 (from initial conditions $\vec{q}$) when it hits the table (call this event transition γ_1). Inter-model comparative analysis might seek to determine whether switching from a frictionless model ($\mathcal{A}$) to a more complex model $\mathcal{B}$ that has coefficient of friction U will increase, decrease, or not affect the velocity. Because $\mathcal{A} \preceq_\Psi \mathcal{B}$ and Ψ is an approximation reformulation, this problem can be expressed as the following intra-model comparative analysis problem in model $\mathcal{B}$: "If ${}^{\vec{q}}U\Uparrow_0$ what will be the resulting relative change on V at transition one?" And as both DQ analysis [24] and exaggeration [27] will show, the answer is ${}^{\vec{q}}V\Downarrow_1$. This relationship between inter- and intra-model comparative analysis can be stated formally in terms of both partial derivatives and relative change values.

Proposition 5 *Let $\mathcal{A}$ and $\mathcal{B}$ be models with* $|$PARAM$(\mathcal{A})| = n$ *and* $|$BOUND$(\mathcal{B})| = k$. *Let Ψ be a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Suppose that $\mathcal{A}$ approximates $\mathcal{B}$ under Ψ with fitting parameter $Q_f \in$* INDEP$(\mathcal{B})$ *and approximation limit l where l is the greatest lower bound of* RANGE(Q_f). *Let $\vec{q}$ denote an internal state of $\mathcal{B}$, $\vec{p}$ the corresponding internal state of $\mathcal{A}$, and $q_f = \pi_f(\vec{q})$. For any $P_j \in$* PARAM$(\mathcal{A})$, *let $R = \pi_j(\Psi(\mathcal{B}_{\vec{q}}(t)))$. Then*

$$\text{if } \lim_{q_f \to l} \frac{\partial R}{\partial q_f}(t, \ldots, q_f, \ldots) = \rho > 0$$

$$\text{then } R(t, \ldots, q_f, \ldots) > P_j(t)$$

for arbitrary time t.

Proof:

First note that R is the function of time, defined in terms of parameters of $\mathcal{B}$, that corresponds to parameter P_j in model $\mathcal{A}$. Recall also that the solution

to $\mathcal{B}$'s quantitative constraints (ODEs) depends on the initial values for all of its k boundary parameters; this means that R is also a function of q_1 through q_k.

By definition of partial derivative,

$$\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots) = \lim_{\zeta\to 0}\frac{R(t,q_1,\ldots,q_f+\zeta,\ldots,)-R(t,q_1,\ldots,q_f,\ldots,)}{\zeta}.$$

Now, if $l \in \text{RANGE}(Q_f)$ then the partial derivative is defined there, and

$$\frac{\partial R}{\partial q_f}(t,\ldots,l,\ldots) = \lim_{q_f\to l}\frac{R(t,q_1,\ldots,q_f,\ldots,)-R(t,q_1,\ldots,l,\ldots,)}{q_f-l}. \qquad (1)$$

However, if $l \notin \text{RANGE}(Q_f)$ there is no problem because we can uniquely extend continuous functions (i.e. R and its partial derivative) to the closure of its domain by substituting the limiting value [?, p99]. As a convenience we assume that we are dealing with such an extension. Our conclusion does not depend on this, however, since we need only demonstrate a region near l where the implication holds. (If $l \in \text{RANGE}(Q_f)$ then the implication holds for $q_f = l$ as well.)

Since Ψ is an approximation reformulation, we know that

$$\lim_{q_f\to l}\max_{1\le i\le n}\left(\sup_{t\in[0,\infty)}|\pi_i(\Psi(\mathcal{B}_{\vec{q}}(t)))-\pi_i(\mathcal{A}_{\vec{p}}(t))|\right)=0.$$

In other words, the maximum difference (overa all time and all parameters) between corresponding values is zero. So clearly there is no difference for the one pair of corresponding values we are interested in. I.e. for all time t

$$\lim_{q_f\to l} R(t,\ldots,q_f,\ldots) = P_j(t). \qquad (2)$$

Thus from equations 2 and 1

$\forall\epsilon>0,\exists\delta>0$ s.t. if $|q_f-l|<\delta$ then

$$\left|\frac{R(t,\ldots,q_f,\ldots)-P_j(t)}{q_f-l}-\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots)\right|<\epsilon.$$

In other words,

$\forall\epsilon>0,\exists\delta>0$ s.t. if $|q_f-l|<\delta$ then

$$\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots)-\epsilon<\frac{R(t,\ldots,q_f,\ldots)-P_j(t)}{q_f-l}<\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots)+\epsilon.$$

But $q_f - l$ is positive so,

$\forall\epsilon>0,\exists\delta>0$ s.t. if $|q_f-l|<\delta$ then

$$\left(\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots)-\epsilon\right)(q_f-l)<R(t,\ldots,q_f,\ldots)-P_j(t).$$

But since $\frac{\partial R}{\partial q_f}(t,\ldots,q_f,\ldots)$ converges to $\rho>0$ we simply need choose $\epsilon<\rho$ to show

$$0 < R(t, \ldots, q_f, \ldots) - P_j(t).$$

Or equivalently,

$$R(t, \ldots, q_f, \ldots) > P_j(t).$$

□

Expressing this result in terms of relative change values makes it easier to apply for model switching.

Corollary 6 (CA Reduction Theorem). *Let $\mathcal{A}$ and $\mathcal{B}$ be models such that* $|\text{PARAM}(\mathcal{A})| = n$ *and* $|\text{BOUND}(\mathcal{B})| = k$. *Let Ψ be a reformulation such that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Suppose that $\mathcal{A}$ approximates $\mathcal{B}$ under Ψ with fitting parameter $Q_f \in$* $\text{INDEP}(\mathcal{B})$ *and approximation limit l where l is the greatest lower bound of* $\text{RANGE}(Q_f)$. *Suppose the two time functions are equal: ${}_{\mathcal{A}}\mathcal{T} = {}_{\mathcal{B}}\mathcal{T}$. Then for any set of initial values $\{q_h\}$ to boundary parameters where $h \neq f$, there exists a positive real η such that if $q_f \in (l, l+\eta)$ and if*

$$\vec{q} = (q_1, \ldots, q_f, \ldots, q_k)$$

then for any parameter $P_j \in$ $\text{PARAM}(\mathcal{A})$ *and for any transition γ_i,*

$$\textit{if } ({}^{\vec{q}}Q_f\Uparrow_0 \rightarrow [\pi_j(\Psi(\text{RCS}(\mathcal{B}_{\vec{q}}, \gamma_i))) = \Uparrow]) \textit{ then } {}^{\vec{p}}_{\Psi}P_j\Uparrow_i$$

where $\vec{p}$ is the internal state of $\mathcal{A}$ corresponding to $\vec{q}$.

Proof:

As before, let $R = \pi_j(\Psi(\mathcal{B}_{\vec{q}}(t)))$. Using the definition of ${}^{\vec{p}}_{\Psi}P_j\Uparrow_i$ we can rewrite the consequent of our objective in terms of R as

$$\text{if } ({}^{\vec{q}}Q_f\Uparrow_0 \rightarrow [\pi_j(\Psi(\text{RCS}(\mathcal{B}_{\vec{q}}, \gamma_i))) = \Uparrow])$$

$$\text{then } R({}_{\mathcal{B}}\mathcal{T}(\gamma_i), \ldots, q_f, \ldots) > P_j({}_{\mathcal{A}}\mathcal{T}(\gamma_i)).$$

By applying proposition 3 we can write the antecedant of our objective in terms of R as

$$\text{if } \frac{\partial R}{\partial q_f}({}_{\mathcal{B}}\mathcal{T}(\gamma_i), \ldots, q_f, \ldots) > 0$$

$$\text{then } R({}_{\mathcal{B}}\mathcal{T}(\gamma_i), \ldots, q_f, \ldots) > P_j({}_{\mathcal{A}}\mathcal{T}(\gamma_i)).$$

And since the two time functions are necessarilly equal, it suffices to show that there exists a positive real η such that when $q_f \in (l, l+\eta)$ then

$$\frac{\partial R}{\partial q_f}(t, \ldots, q_f, \ldots) > 0 \;\rightarrow\; R(t, \ldots, q_f, \ldots) > P_j(t)$$

for arbitrary time t.

This follows from proposition 5. □.

This is the key result which enables model switching in SAM. Comparable results can be stated if the approximation limit is the ceiling of the range closure, and for other RC values. The net effect is that inter-model comparative analysis reduces to intra-model comparative analysis when performed in the complex model with a behavior generated from initial conditions that put the fitting parameter close to the approximation limit.

4.6 Generalizations to the Reduction Theorem?

The reduction theorem is important enough that we wish to be sure that it is as general as possible. In this section we consider two possible changes: eliminating the constraint on the value of the fitting parameter and relaxing the condition that the time functions be equal.

Unfortunately, it turns out that the the restriction that the fitting parameter be close to the approximation limit is crucial. To see this, consider the following example. Let $\mathcal{A}$ be a model with an independent parameter C and a dependent parameter X obeying the constraint that $X = C$. Let $\mathcal{B}$ be a model with two independent parameters C and D, one dependent parameter X, a range restriction that RANGE$(D) = (0, \infty)$ and obeying the constraint that $X = C + D + D^2$. Let Ψ be the obvious projection reformulation and it is clear that $\mathcal{A} \preceq_\Psi \mathcal{B}$. Furthermore, it is clear that $\mathcal{A}$ approximates $\mathcal{B}$ with fitting parameter D and approximation limit 0. What is the relative change on the predicted value of X of a switch from $\mathcal{A}$ to $\mathcal{B}$?

Well, it depends. If D is close to the approximation limit 0, then the value of D dominates the value of D^2 and X increases. But if D is large then a switch to $\mathcal{B}$ causes X to decrease. This illustrates the need to perform intra-model comparative analysis on a behavior of $\mathcal{B}$ in which the fitting parameter is close to the approximation limit.

On the other hand, it seems quite possible that the restriction that the time functions be equal could be relaxed. In fact proposition 4 shows that if $\mathcal{A}$ approximates $\mathcal{B}$, then the two time functions converge as the fitting parameter tends to the approximation limit. Thus it may be possible to generalize the reduction theorem by requiring only topological equality.

5 SAM

We are implementing a common lisp program, SAM, to test the ideas of validation, model shifting, and inter-model comparative analysis. SAM represents a physical system with a graph of models [16, 13, 2, 1] where each edge in the graph is an approximation reformulation labeled with the fitting parameter and approximation limit.

SAM solves a simulation task by reasoning in three phases. First a model is chosen; by default, the simplest model is used. Next, a Runge-Kutta algorithm approximates solutions for each of the parameters over an initial time interval. At this point SAM asks for measurements of parameter values and checks these against its predictions. If any discrepancies exceed a task dependent threshold, SAM seeks to switch models. It does this by generating a sequence of neighbor models to the current model in the graph. For each candidate model, SAM performs inter-model comparative analysis (using the DQ analysis technique [24] which calls upon the QSIM qualitative simulator [10]). The first model which appears to correct the discrepancy is tried next and the control loop repeats. Since each model switch moves upwards in the $\preceq_\Psi$ lattice, the routine is guaranteed to terminate either by producing an acceptable prediction or by failing at the most complex model.

As the implementation has just been started, no performance figures are available at this time.

6 Related Work

A number of research projects have addressed issues similar to those described in this paper. Our work bears the closest resemblance to the work on PROMPT [16] a program which uses multiple models to perform innovative design [13]. This work introduced the notion of a graph of models (GoM). Each node in the graph denotes a model of the system at hand and the edge connecting two nodes is labeled with the set of simplifying assumptions (e.g., no friction) that distinguish the two models. One difference between their approach and ours is PROMPT's use of multiple GoMs corresponding to different domains of expertise [1]; this divide and conquer approach alleviates the exponential-space problem inherent in the GoM approach. Another difference is the manner in which a new model is chosen after a discrepancy is detected. PROMPT uses delta vectors to trigger domain-dependent parameter-change rules [2]. Their method is probably more efficient, but ours is domain independent. Future research may lead to techniques for compiling the conclusions of inter-model comparative analysis into parameter change rules, combining the best of the two approaches.

Falkenhainer and Forbus present an alternative paradigm for multiple perspective, multi-granular modeling [7]. Instead of a graph of models, Falkenhainer and Forbus define a generating set of model pieces that can be turned on and off by *consider assumptions*. By alternately assuming all consistent sets of consider assumptions, one could produce a graph of models, but by not doing so explicitly, considerable space savings are realized. In addition, Falkenhainer and Forbus introduce the important distinction between simplifying and operating assumptions. Simplifying assumptions abstract details from the device model while operating assumptions limit consideration to subcases of behavior such as equilibrium operation. They demonstrate the power of their approach by describing an implemented question-answering program that consideres only pertinent aspects of a Navy steam propulsion plant model for each question.

Davis' work on troubleshooting [4, 5] has been a major influence in our work. His decision to represent modeling assumptions explicitly allowed his program to perform diagnosis by sequentially relaxing the assumptions.

The program starts by assuming that the only possible faults are localized functional errors (e.g., a broken adder chip). If subsequent measurements guarantee that no single functional unit could be responsible for all observed symptoms, then the program retracts the assumption and considers a more detailed model that can represent more faults. Although Davis' implementation used just two models, he postulated an algorithm that would step through a sequence of increasingly complex models: e.g., representations for bridge faults, multiple faults, assembly errors, even design errors. Our work (and that of PROMPT) extends this switching paradigm by eliminating the need for a prespecified linear sequence of models.

There are many additional papers of relevance. Slices, a technique for reasoning about electronic circuits from multiple perspectives, is described in [19]. Time-scale abstraction [11] can be used to decompose complex systems into smaller parts that act with different rates, but current techniques require that both the models and the pattern of model switching be hand coded. Aggregation [22] can dynamically create abstract models of a system, but only for the limited class of repetitious or cyclic behavior. Patil's ABEL program [15, 14] constructed multiple models of a sick patient, each at a different level of detail,

with links to enforce correspondences. Collins and Forbus [3], building on the work of [8], describe a system that reasons about fluids using models constructed in two different ontologies.

Relevant work has also been done using nonphysical domains. [9] is a classic paper on reformulation — we are clearly inspired by his functional approach. Subramanian and Genesereth [18] discusses techniques for reformulating a first order predicate calculus theory into a simpler theory by "factoring out" facts that are irrelevant to a given class of queries. Van Baalen and Davis [21] describe a program that builds a specialized representation of a problem in order to capture the inherent constraints; solving the problem is easy given the new representation. In addition, the machine learning community has recognized the need for abstraction and reformulation of domain models; see [6, 20].

7 Future Work

Many things remain to be done:

- Can proposition 6 be generalized to eliminate the restriction that the time functions are identical? Proposition 4 shows that this constraint is very weak (i.e. easy to remove) given the existence of an approximation.
- The theory must be extended to deal with model whose behaviors do not satisfy topological equality. This should be simply a matter of defining a more flexible method for specifying CORRESPONDING EVENTS.
- Can the theory be extended to handle system with multiple operating regions? This could be difficult since proposition 6 depends crucially on continuity. For example, consider the assumption that string (in simple phsyical examples) never breaks. This can be seen as an approximation using the formalism above if one considers the BREAKING-POINT the fitting parameter. Wait, this isn't a parameter it's a value, so maybe everything works out fine. Anyway, more thought is necessary. My intuition tells me that uniform convergence (as opposed to pointwise convergence) is involved.
- We need to push on the idea of ONTOLOGIES as a way to implement space efficient graphs of models. (In a naive implementation, the size of a GoM is exponential in the number of possible assumptions.)
- Once SAM is complete, we need to test the power of our approach on real examples.

ACKNOWLEDGEMENTS

This work benefited from conversations with Dorothy Neville, Paul Beame, Alan Borning, and Richard Anderson. The suggestions of Paul Beame were crucial in proving proposition 6. Dorothy Neville read an early draft and made a number of helpful comments. Benjamin Kuipers kindly provided the common lisp code for QSIM, and Elisha Sacks supplied Runge-Kutta code; both of these are used in SAM.

This research was funded in part by National Science Foundation Grants IRI-8902010 and IRI-8957302.

References

[1] S. Addanki, R. Cremonini, and Penberthy J. Contexts: Dynamic Identification of Common Parameters in Distributed Analysis of Complex Devices. In *Proceedings of IJCAI-89*, August 1989.

[2] S. Addanki, R. Cremonini, and Penberthy J. Reasoning about Assumptions in Graphs of Models. In *Proceedings of IJCAI-89*, August 1989.

[3] J. Collins and K. Forbus. Reasoning About Fluids Via Molecular Collections. In *Proceedings of AAAI-87*, July 1987.

[4] R. Davis. Diagnosis via Causal Reasoning: Paths of Interaction and the Locality Principle. In *Proceedings of AAAI-83*, August 1983.

[5] R. Davis. Diagnostic Reasoning Based on Structure and Behavior. *Aritificial Intelligence*, 24, 1984.

[6] T. Ellman. Approximate Theory Formation: An Explanation-Based Approach. In *Proceedings of AAAI-88*, August 1988.

[7] B. Falkenhainer and K. Forbus. Setting up Large Scale Qualitative Models. In *Proceedings of AAAI-88*, August 1988.

[8] P. Hayes. Naive Physics I: Ontology for Liquids. In *Formal Theories of the Commonsense World.* Ablex, 1985.

[9] R. Korf. Towards a Model of Representation Changes. *Artificial Intelligence*, 14, September 1980.

[10] B. Kuipers. Qualitative Simulation. *Artificial Intelligence*, 29, September 1986.

[11] B. Kuipers. Abstraction by Time-Scale in Qualitative Simulation. In *Proceedings of AAAI-87*, July 1987.

[12] W. Massey. *Algebraic Topology: An Introduction.* Springer-Verlag, 1967.

[13] S. Murthy and S. Addanki. PROMPT: An Innovative Design Tool. In *Proceedings of AAAI-87*, pages 637–642, August 1987.

[14] R. Patil. Causal Representation of Patient Illness for Electrolyte and Acid-Base Diagnosis. Tr-267, MIT Laboratory for Computer Science, October 1981.

[15] R. Patil, P. Szolovits, and W. Schwartz. Causal Understanding of Patient Illness in Medical Diagnosis. In *Proceedings of IJCAI-81*, 1981.

[16] J.S. Penberthy. Incremental Analysis and the Graph of Models: A First Step towards Analysis in the Plumber's World. MS Thesis, MIT Laboratory for Computer Science, January 1987.

[17] M. Spivak. *Calculus.* W. A. Benjamin, Inc., Menlo Park, CA, 1967.

[18] P. Struss. Mathematical Aspects of Qualitative Reasoning. *International Journal for Artificial Intelligence in Engineering*, 3(3), July 1988.

[19] D. Subramanian and M. Genesereth. The Relevance of Irrelevance. In *Proceedings of IJCAI-87*, pages 416–422, August 1987.

[20] G.J. Sussman and G.L. Steele. CONSTRAINTS: A Language for Expressing Almost-Hierarchical Descriptions. *Artificial Intelligence*, 14:1–39, 1980.

[21] A. Unruh and P. Rosenbloom. Abstraction in Problem Solving and Learning. In *Proceedings IJCAI-89*, pages 682–687, August 1989.

[22] J. Van Baalen and R. Davis. Overview of an Approach to Representation Design. In *Proceedings of AAAI-88*, August 1988.

[23] D. Weld. The Use of Aggregation in Causal Simulation. *Artificial Intelligence*, 30(1), October 1986.

[24] D. Weld. Comparative Analysis. In *Proceedings of IJCAI-87*, August 1987.

[25] D. Weld. Comparative Analysis. *Artificial Intelligence*, 36(3), October 1988.

[26] D. Weld. Exaggeration. In *Proceedings of AAAI-88*, August 1988.

[27] D. Weld. Theories of Comparative Analysis. AI-TR-1035, MIT AI Lab, May 1988.

[28] D. Weld. Exaggeration. *Artificial Intelligence*, 43(2), To Appear in April 1990.

[29] B. Williams. MINIMA: A Symbolic Approach to Qualitative Algebraic Reasoning. In *Proceedings of AAAI-88*, August 1988.