

# A Dynamic Systems Perspective on Qualitative Simulation

Elisha Sacks<sup>1</sup>

Department of Computer Science

Princeton University

Princeton, NJ 08544

## Abstract

This paper examines qualitative simulation (QS) from the phase space perspective of dynamic systems theory. QS consists of two steps: *transition analysis* determines the sequence of qualitative states that a system traverses and *global interpretation* derives its long-term behavior. I recast transition analysis as a search problem in phase space and replace the assorted transition rules with two algebraic conditions. The first condition determines transitions between arbitrarily shaped regions in phase space, as opposed to QS which only handles  $n$ -dimensional rectangles. It also provides more accurate results by considering only the boundaries between regions. The second condition determines whether nearby trajectories approach a fixed point asymptotically. It obtains better results than QS by exploiting local stability properties. I recast global interpretation as a search for attractors in phase space and present a global interpretation algorithm for systems whose local behavior determines global behavior uniquely.

---

<sup>1</sup>This research was performed while I was in the Clinical Decision Making Group of the M.I.T. Laboratory for Computer Science. It was supported (in part) by National Institutes of Health Grant No. R01 LM04493 from the National Library of Medicine.

# 1 Introduction

Qualitative reasoning seeks to automate the analysis of dynamic systems whose exact workings are unimportant to the problem task, complicated, or unknown. For example, a computer chef should be able to infer that turning up the flame under a pot increases the temperature of its contents without solving heat conduction equations. Current qualitative reasoning programs derive the abstract behavior of a system by simulating qualitative versions of its governing equations. This paper examines qualitative simulation (QS) from the phase space perspective of dynamic systems theory.

The next section translates the qualitative reasoning terminology of AI researchers into the phase space representation. QS consists of two steps: *transition analysis* determines the sequence of qualitative states that a system traverses and *global interpretation* derives its long-term behavior. Sections 3 and 4 recast transition analysis as a search problem in phase space and replace the assorted transition rules with two algebraic conditions. These conditions are more general and more accurate than the original transition rules. Section 5 recasts global interpretation as a search for attractors in phase space. It presents a global interpretation algorithm for systems whose local behavior determines global behavior uniquely. The final section contains conclusions and plans future work on global interpretation.

This paper addresses a generic form of QS, couched in terms of abstract differential equations that contain arbitrary constants and functions. The ideas apply to every formalism in the qualitative reasoning literature as well. The discussion pertains solely to input/output behavior and efficiency, not to cognitive or philosophical issues. It makes no claims about *human* reasoning methods or about the putative causal ordering that QS imposes on systems. See de Kleer and Brown [4, 5] and Iwasaki and Simon [9, 10] for a discussion of these issues.

## 2 Qualitative Reasoning in Phase Space

The phase space for a system of first-order differential equations

$$x'_i = f_i(x_1, \dots, x_n); \quad i = 1, \dots, n \quad (1)$$

is the Cartesian product of the  $x_i$ 's domains. One can convert higher-order equations to first-order ones by introducing new variables as synonyms for higher derivatives. Points in phase space represent states of the system. Curves on which equation (1) is satisfied, called

trajectories, represent solutions. A *phase diagram* for a system depicts its phase space and trajectories graphically. The topological and geometric properties of trajectories characterize the qualitative behavior of solutions. For example, a point trajectory, called a *fixed point*, indicates a constant solution, whereas a closed curve indicates a periodic solution. A fixed point is *stable* if every nearby trajectory approaches it asymptotically, *repellent* if every nearby trajectory diverges from it, and *unstable* otherwise.

For example, the motion of a block attached to a track from below and to an elastic band from above (Figure 1) is typically modeled by the abstract equations

$$\begin{cases} x' = v \\ v' = g(x) + h(v) \end{cases} \quad (2)$$

where  $x$  and  $v$  denote the location and velocity of the block. The functions  $g(x)$  and  $h(v)$  decrease monotonically and vanish at zero, but are otherwise arbitrary. They represent the elastic and frictional forces. (I ignore static friction for simplicity.) Figure 2 shows a typical phase diagram for equation (2). The system has a single stable fixed point,  $(0,0)$ , toward which all other trajectories spiral. The fixed point represents the steady state in which the block lies at rest directly beneath the ceiling. The spirals represent the decaying oscillations that arise when the block is displaced from the steady state.

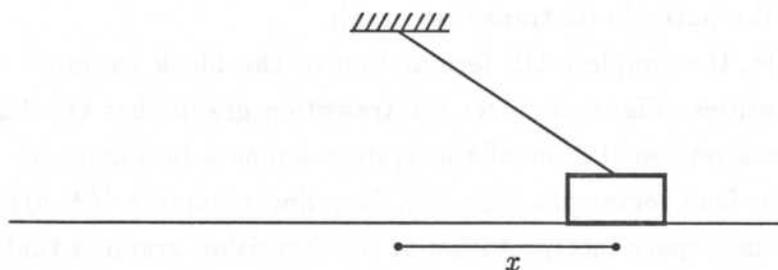


Figure 1: A block attached to an elastic band.

The QS representation translates directly into phase space terminology. QS manipulates abstract versions of numbers, called *qualitative values*, organized in *quantity spaces*: ordered sets of alternating intervals and points. One useful quantity space, called IQ, consists of the interval  $(-\infty, 0)$ , the point 0, and the interval  $(0, \infty)$ . These abbreviate to  $-$ , 0, and  $+$ . QS models systems with sets of *quantities*: functions from an independent variable (usually time) to quantity spaces. Define the *qualitative state* of a system to be the vector of the qualitative values of its quantities. QS characterizes the behavior of a system by the

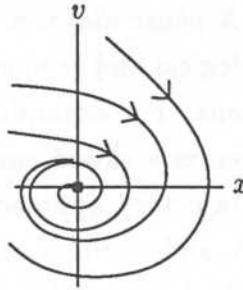


Figure 2: Phase diagram for the block equation (2).

sequences of qualitative states that it can go through. It represents this information with a *transition graph* whose nodes and links denote qualitative states and possible transitions.

Qualitative values correspond to intervals and points on the real line. Quantities correspond to phase variables. Let  $Q_i$  denote the quantity space for quantity  $q_i$ . Each qualitative state of a system with quantities  $q_1, \dots, q_n$  maps to an  $n$ -dimensional rectangle in the Cartesian product  $Q_1 \times \dots \times Q_n$ . QS cannot represent non-rectangular regions, such as circles, because the components of qualitative states are independent. Qualitative state transitions correspond to transitions between regions in phase space, so transition graphs specify the sequences of regions that trajectories traverse. The qualitative behaviors of a system are isomorphic to the paths in its transition graph.

For example, the simplest QS formulation of the block example represents  $x$  and  $v$  as IQ-valued quantities. Figure 3 shows the transition graph that QS derives for these values. The phase space representation of the system appears in Figure 4. The nine qualitative states map into four rectangles ( $\langle \pm, \pm \rangle$ ), four line segments ( $\langle \pm, 0 \rangle$ ), and one fixed point ( $\langle 0, 0 \rangle$ ). The phase space interpretation of the transition graph is that every solution either remains at  $\langle 0, 0 \rangle$  forever, cycles clockwise through the outer eight regions forever, or cycles for a while and then approaches  $\langle 0, 0 \rangle$  from within  $\langle -, + \rangle$  or  $\langle +, - \rangle$ . In fact, all trajectories are eventually trapped in every circle centered at the origin, but one cannot represent this information as a sequence of transitions between rectangular regions.

The next two sections examine transition analysis in a phase space setting. Transitions involving fixed points are different from those between other regions. From here on, let the term *region* denote regions other than fixed points. There are two cases to consider: transitions between regions and transitions from regions to fixed points. Transitions out of fixed points cannot occur by definition.

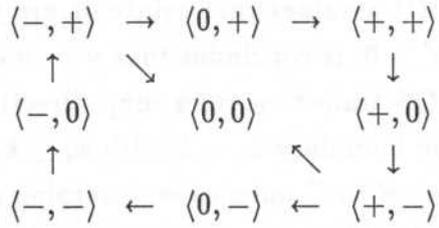


Figure 3: Transition graph for the block.

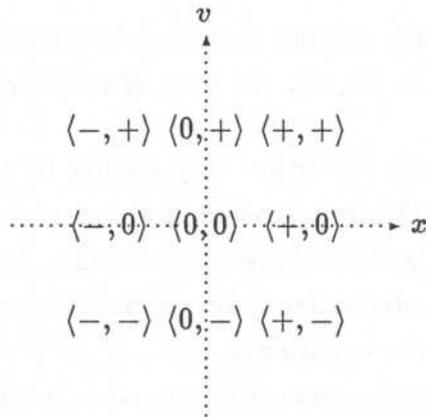


Figure 4: Phase space representation of qualitative states for the block.

### 3 Transitions Between Regions

QS calculates the successors of a qualitative state from the signs of its derivatives. It derives the signs from the governing equations and infers the transitions from the intermediate value and mean value theorems. (It assumes that solutions are piecewise smooth.) For example, given the state  $y < 0$  and  $y' > 0$ , it concludes that  $y$  may enter the state  $y = 0$ , but cannot enter  $y > 0$  directly. The QS transition rules map directly to phase space. Let regions  $R$  and  $S$  be adjacent along the boundary  $x_i = k$  with  $x_i \leq k$  for  $x \in R$  and  $x_i \geq k$  for  $x \in S$ . QS derives a transition from  $R$  to  $S$  unless the governing equations imply  $x'_i \leq 0$  for every  $x \in R$  and one from  $S$  to  $R$  unless the equations imply  $x'_i \geq 0$  for every  $x \in S$ . In the block example,  $x' > 0$  on  $\langle -, + \rangle$  and  $\langle 0, + \rangle$  by equation (2). Hence, there is a transition from  $\langle -, + \rangle$  to  $\langle 0, + \rangle$ , but not from  $\langle 0, + \rangle$  to  $\langle -, + \rangle$ . Figure 5 depicts this example as well as symmetric transitions between two regions.



Figure 5: Sample transitions. Arrows indicate  $x$ 's direction of change. (a) from  $\langle -, + \rangle$  to  $\langle 0, + \rangle$ , but not from  $\langle 0, + \rangle$  to  $\langle -, + \rangle$ . (b) from  $R$  to  $S$  and from  $S$  to  $R$ .

The QS transition analysis algorithm only applies to regions whose boundaries have the form  $x_i = k$ . This prevents QS from reasoning about regions such as  $u > v$  and  $u^2 + v^2 < k$  that have slanted and circular boundaries respectively. Also, the QS algorithm can produce spurious transitions when a derivative takes on positive and negative values within a single region. For example, given the equation  $y' = y - y^2$ ,  $y'$  is positive for  $y \in (0, 1)$  and negative for  $y \in (1, \infty)$ . QS derives a transition from  $(0, \infty)$  to 0 even though  $y$  moves away from 0 in  $(0, 1)$ , as shown in Figure 6. In this case, one can solve the problem by including the critical points of  $y$ , 0 and 1, in its value space. This solution becomes cumbersome for systems with many critical points and impossible when the critical points of one variable depend on another variable.

A phase space perspective generalizes transition analysis to regions bounded by smooth curves and eliminates the spurious transitions discussed above. For a trajectory to cross from



Figure 6: Phase diagram for  $y' = y - y^2$ . All solutions in  $(0,1)$  move away from 0.

region  $R$  to  $S$  via boundary  $u$ , its tangent  $t$  at the intersection point with  $u$  must form an acute angle with the normal  $n$ , as shown in Figure 7. In algebraic terms, the inner product  $t \cdot n$  must be positive. Hence, a transition exists from  $R$  to  $S$  unless  $t \cdot n \leq 0$  everywhere on  $u$ . This condition applies to any smooth boundary. It reduces to the QS transition condition,  $x'_i \leq 0$ , in the special case where  $u$  has the form  $x_i = k$ .

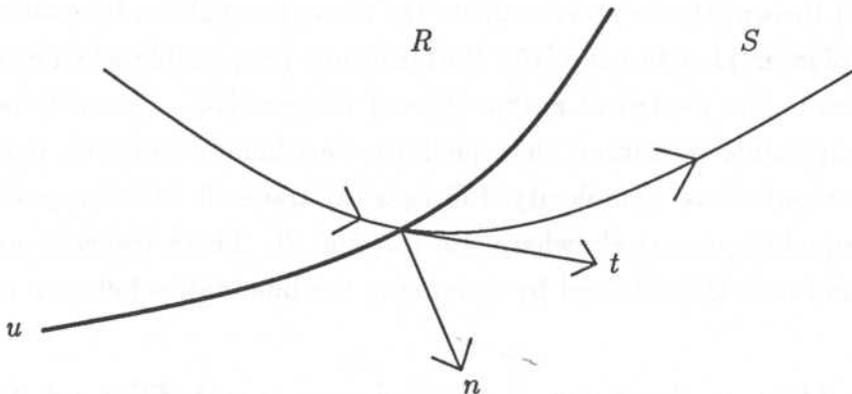


Figure 7: Trajectory crossing from  $R$  to  $S$  via  $u$ . The tangent  $t$  at the crossing point must form an acute angle with  $n$ , the normal to  $u$  that points into  $S$ .

The condition  $t \cdot n \leq 0$  is evaluated on the boundaries between regions, making it insensitive to changes in the signs of derivatives within a region. In our example,  $y' = 0$  at the boundary, 0, between  $(0, \infty)$  and 0, ruling out a transition to 0. In the block example,  $x' > 0$  on the boundary,  $\langle 0, + \rangle$ , between  $\langle -, + \rangle$  and  $\langle 0, + \rangle$ . This result agrees with the QS prediction because  $x' > 0$  throughout  $\langle -, + \rangle$ . The general case of regions  $R$  and  $S$  adjacent along  $x_i = k$  and governed by equation (1) is analogous. A transition from  $R$  to  $S$  occurs unless

$$\lim_{x_i \uparrow k} f_i(x_1, \dots, x_n) \leq 0 \quad (3)$$

for all  $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$  in  $R$ . The limit reduces to  $f_i(x_1, \dots, x_{i-1}, k, x_{i+1}, \dots, x_n)$  when  $f_i$  is continuous from below or when  $R$  has zero width in dimension  $i$ . Conversely, a

transition from  $S$  to  $R$  occurs unless

$$\lim_{x_i \downarrow k} f_i(x_1, \dots, x_n) \geq 0 \quad (4)$$

for all  $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$  in  $S$ . The limit reduces to  $f_i(x_1, \dots, x_{i-1}, k, x_{i+1}, \dots, x_n)$  when  $f_i$  is continuous from above or when  $S$  has zero width in dimension  $i$ . These reductions apply to all current QS systems because all researchers have assumed one-sided (but not two-sided) continuity.

The accuracy of these transition conditions depends on the specificity of the system equations and on the power of the inequality prover. If only sign equations are available, as in de Kleer and Brown [4], one must evaluate the above inequalities by *qualitative arithmetic*, a special case of interval arithmetic [15]. Better results are possible when algebraic equations are available, as in the models of Forbus [6] and Kuipers [14]. Powerful inequality provers can rule out impossible transitions that qualitative arithmetic endorses, perhaps at the cost of increased computational complexity. I discuss the tradeoffs between power and efficiency in existing inequality provers elsewhere [16] [17, Ch. 7]. These tradeoffs are orthogonal to the improvement over QS obtained by examining the boundaries between regions.

## 4 Transitions between Regions and Fixed Points

Some QS researchers, including Kuipers, derive transitions from regions to fixed points in the same manner as transitions between regions. Others, including de Kleer and Bobrow [3], rule them out altogether. An analytic function can be constant on an interval only if it is always constant; hence, a moving trajectory cannot come to rest. Neither approach accurately describes the underlying phenomena. Moving trajectories cannot, in fact, reach fixed points in *finite* time, but they can approach fixed points asymptotically. Moreover, the stability characteristics of fixed points govern the transitions from adjacent regions. Programs that apply the inter-regional transition rule to this case can overlook important inferences. This section describes an algorithm, called FPA (for fixed point analysis), that derives the transitions into fixed points from their stability characteristics.

FPA calculates the fixed points of a system of differential equations

$$x'_i = f_i(x_1, \dots, x_n); \quad i = 1, \dots, n \quad (1)$$

by setting all derivatives to zero and solving the resulting equations. In the block example, it solves

$$\begin{cases} v = 0 \\ g(x) + h(v) = 0 \end{cases} \quad (5)$$

to obtain the unique fixed point:  $(0, 0)$ . The solution follows from the constraints on  $g$  and  $h$ . Sufficiently complicated systems can cause any equation solver to fail. This should rarely be a problem for FPA, given the simplicity of the equations in the qualitative reasoning literature.

FPA tests whether each fixed point that it calculates is stable or repellent. If a fixed point  $p$  is stable, every trajectory that passes near  $p$  converges to it. Transitions exist from each of  $p$ 's neighbors to  $p$  because each neighbor contains points arbitrarily near  $p$ . If  $p$  is repellent, every trajectory that passes near  $p$  diverges from it. Hence, there are no transitions into  $p$ .

A fixed point  $p$  of (1) is stable (repellent) exactly when all the eigenvalues of the  $n$ -by- $n$  matrix

$$A(x) = (a_{ij}) \text{ with } a_{ij} = \frac{\partial}{\partial x_j} f_i(x_1, \dots, x_n) \quad (6)$$

have negative (positive) real parts at  $x = p$  [8, Ch. 9]. To determine stability, FPA first ignores the magnitudes of the  $a_{ij}$  and tests whether their signs alone imply stability. This condition, called *sign stability*, is easier to ascertain than the weaker condition of stability, but rarer. The sign pattern of  $A$  expresses the direction of influence between variables in the neighborhood of  $p$ . The  $j$ th variable influences the  $i$ th variable positively if  $a_{ij} > 0$ , not at all if  $a_{ij} = 0$ , and negatively if  $a_{ij} < 0$ . A sign stable matrix is one in which this qualitative information implies stability.

Sign stability is the strongest condition derivable in formalisms, such as de Kleer's, where the magnitudes of the  $a_{ij}$  are unspecified. If  $-A = (-a_{ij})$  is sign stable, there are no transitions into  $p$ . If not, there exists an unstable instance of  $-A$ , hence an instance of  $A$  with a negative eigenvalue. For this instance, every neighborhood of  $p$  contains points whose trajectories approach  $p$  asymptotically [7, Ch. 1]. The transition graph must include links into  $p$  from each adjacent region to cover this case.

Sign stability does not distinguish between spiral and nodal approach to a fixed point (Figure 8) because the magnitudes of the  $a_{ij}$  determine these properties. In the two-dimensional case, for example, trajectories spiral for

$$\Delta = (a_{11} - a_{22})^2 + 4a_{12}a_{21} < 0 \quad (7)$$

and are nodal otherwise. Given only the signs of the  $a_{ij}$ , QS can do no better. I discuss constraints on the magnitudes of the  $a_{ij}$  at the end of this section.



Figure 8: Spiral versus nodal approach to a fixed point.

FPA calculates the  $a_{ij}$  from the system (1) and derives their signs with the BOUNDER inequality reasoner [16]. In the block example (2), the matrix

$$A(x, v) = \begin{bmatrix} \frac{\partial}{\partial x} v & \frac{\partial}{\partial v} v \\ \frac{\partial}{\partial x} (g(x) + h(v)) & \frac{\partial}{\partial v} (g(x) + h(v)) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ g'(x) & h'(v) \end{bmatrix} \quad (8)$$

has the sign pattern  $\begin{bmatrix} 0 & + \\ - & - \end{bmatrix}$  at  $(0, 0)$ . The bottom two entries are ‘-’ because  $g$  and  $h$  are monotone decreasing functions. The signs of the  $a_{ij}$  are also derivable in all the QS formalisms described in Bobrow [2].

Jeffries et al. [11] derive necessary and sufficient conditions for sign stability. Define  $R = \{i | a_{ii} \neq 0\}$ . Let  $G$  be an undirected graph with vertices  $V = \{1, \dots, n\}$  and edges

$$E = \{(i, j) | i \neq j \wedge a_{ij} \neq 0\}. \quad (9)$$

An  $R$ -coloring of  $G$  partitions  $V$  into black and white sets such that each vertex in  $R$  is black, no black vertex has exactly one white neighbor, and each white vertex has some white neighbor. A *complete matching* in  $G$  for  $Z \subset V$  is a set of edges satisfying: (1) no two edges touch the same vertex and (2) an edge touches each vertex in  $Z$ .  $A$  is sign stable iff:

1.  $a_{ii} \leq 0$  for all  $i$  and  $a_{jj} < 0$  for some  $j$ ,
2.  $a_{ij}a_{ji} \leq 0$  for all  $i \neq j$ ,
3.  $a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_{m-1} k_m} a_{k_m k_1} = 0$  for any  $m \geq 3$  distinct indices  $k_1, \dots, k_m$ ,
4. the only  $R$ -coloring of  $G$  is all black, and
5. a  $V - R$  complete matching exists for  $G$ .

Conditions 1 and 2 rule out positive feedback from a variable to itself and between pairs of variables. Klee and van den Driessche [13] present an algorithm that tests conditions 1–5 for an  $n$ -by- $n$  matrix  $A$  with  $e$  nonzero elements in  $O(\max\{n, e\})$  time.

The sign pattern of the block example is  $\begin{bmatrix} 0 & + \\ - & - \end{bmatrix}$  at the unique fixed point:  $(0, 0)$ . This pattern is sign stable, implying that the oscillations predicted by simulation die down for small initial displacements. Conditions 1–3 are immediate. The graph  $G$  has vertices  $V = \{1, 2\}$  and edges  $E = \{(1, 2)\}$ .  $R$  is  $\{2\}$ . Every  $R$ -covering of  $G$  paints 2 black by definition; hence 1 must be black to prevent 2 from having a single white neighbor, implying condition 4. The edge  $(1, 2)$  satisfies condition 5.

A *redundant system* contains a set of variables that can be expressed in terms of its constants and remaining variables. Redundancy prevents a stable system from being sign stable. For example, the following system describes the (inertia-free) flow of liquid through the U-shaped tube shown in Figure 9:

$$\left\{ \begin{array}{l} \text{level-}x = u(\text{volume-}x) \\ \text{pressure-}x = p(\text{level-}x) \\ \text{level-}y = v(\text{volume-}y) \\ \text{pressure-}y = p(\text{level-}y) \\ \text{flow-rate} = r(\text{pressure-}x - \text{pressure-}y) \\ \text{volume-}x' = -\text{flow-rate} \\ \text{volume-}y' = \text{flow-rate} \end{array} \right. \quad (10)$$

with  $p$ ,  $r$ ,  $u$ , and  $v$  monotone increasing functions that vanish at zero. The system is redundant because either of the variables  $\text{volume-}x$  and  $\text{volume-}y$  can be expressed as the difference between the constant *total-volume* and the other variable.

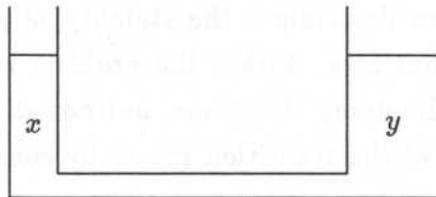


Figure 9: Fluid flow through a U-shaped tube.

Redundancy is primarily a modeling issue: flabby models produce weak conclusions. Nevertheless, FPA can ameliorate the problem by eliminating linear dependencies. It derives dependent sets of derivatives,  $\sum_i \alpha_i v_i' = 0$ , and integrates them into dependent sets of variables

$\sum_i \alpha_i v_i = k$ . Next, it derives a maximal independent set of variables and eliminates the remaining variables by substitution. In the U-tube example, it obtains  $volume-x' + volume-y' = 0$  from the last two constraints, integrates the result into  $volume-x + volume-y = k$ , and substitutes  $k - volume-y$  for  $volume-x$ . Abbreviating  $volume-y$  to  $y$ , the reduced equation

$$y' = r(F(y)) \text{ with } F(y) = p(u(k - y)) - p(v(y)) \quad (11)$$

has  $A = (a_{11})$  with

$$a_{11} = \frac{\partial}{\partial y} r(F(y)) = -r'(F(y)) [p'(u(k - y))u'(k - y) + p'(v(y))v'(y)]. \quad (12)$$

The derivative terms in  $a_{11}$  are positive because the corresponding functions increase monotonically. Hence,  $a_{11}$  is negative and  $A$  satisfies the sign stability conditions. Small height differentials between the arms of the U-tube always disappear over time.

If neither  $A$  nor  $-A$  is sign stable at the fixed point  $p$ , FPA tests whether constraints on the magnitudes of the  $a_{ij}$  imply stability. For example, given  $a > b > 0$  it determines that

$$M = \begin{bmatrix} -a & b \\ b & -a \end{bmatrix} \quad (13)$$

is stable even though  $M$  violates condition 2 for sign stability and  $-M$  violates condition 1. FPA also determines that the fixed point is nodal from equation (7), since  $\Delta = 4b^2 > 0$ .

Stability of an  $n$ -by- $n$  matrix is decidable in  $O(n^3)$  time with the Routh-Hurwitz algorithm [1, Ch. 3] when the  $a_{ij}$  are numbers. That algorithm depends solely on inequalities involving terms in the determinant expansion of  $A$ . Hence, it can also handle symbolic  $a_{ij}$ , providing the signs of the determinant terms remain derivable. FPA tries to resolve the inequalities with the BOUNDER inequality reasoner. If it succeeds, as in the case of  $M$  above, the Routh-Hurwitz algorithm determines the stability of  $A$ . If BOUNDER cannot resolve some inequality, the algorithm fails. Either the problem is too hard for BOUNDER or the inequality is indeterminate, implying that some, but not all, instances of  $A$  are stable. FPA maintains the completeness of the transition graph by constructing transitions from every adjacent region to  $p$ .

The Routh-Hurwitz algorithm must resolve  $O(n^3)$  inequalities. The full BOUNDER program takes exponential time in each inequality, making the overall algorithm exponential. FPA can also use an  $O(n)$  subset of BOUNDER that performs interval arithmetic, making the overall algorithm  $O(n^4)$ , but sacrificing some predictive power. The restricted algorithm suffices for many matrices, including  $M$  above.

## 5 Global Interpretation

The task of global interpretation is to determine the long-term behavior of a system, as opposed to transient behavior. For example, the block asymptotically nears its rest position and the height differential in the U-tube asymptotically vanishes. The phase space equivalent of asymptotic behavior is an *attractor*: a connected subset of phase space that nearby trajectories approach. The fixed points of the U-tube and the block are point attractors. Other possibilities include limit cycles, separatrices, and strange attractors.

The global interpretation capabilities of QS are quite limited. It can detect possible point attractors by identifying sink regions in transition graphs, such as  $\langle 0, 0 \rangle$  for the block (Figure 3). However, it cannot tell whether all, some, or no neighboring trajectories actually approach such sinks. Similarly, QS can recognize cycles in the transition graph as possible limit cycles, but cannot tell whether nearby trajectories spiral inward, spiral outward, or wobble around.

Dynamic systems theory provides tools for determining the attractors of many extremely general equations. The simplest case, called *monostability*, is where all trajectories approach a single fixed point. A sufficient condition for monostability, proved by Kalman [12], is that the matrix  $A(x)$  be stable for every  $x$ . FPA tests this condition with the Routh-Hurwitz algorithm. If  $A(x)$  has a fixed sign pattern, it first tries the sign stability algorithm, which takes less time and avoids symbolic algebra and inequality reasoning. For example, the matrix

$$A(x, v) = \begin{bmatrix} 0 & 1 \\ g'(x) & h'(v) \end{bmatrix} \quad (14)$$

for the block equation (2) has the sign pattern  $\begin{bmatrix} 0 & + \\ - & + \end{bmatrix}$  for every  $x$  and  $v$ . The system is monostable because this pattern is sign stable, as proved in the previous section. This confirms the intuition that the block's oscillations die out as friction dissipates its initial energy. The U-tube equation (11) is also monostable, implying that arbitrary height differentials between the arms of the U-tube always disappear over time.

## 6 Conclusions

This paper recasts the terminology and algorithms of qualitative reasoning into the phase space representation. It simplifies transition analysis from many rules to two algebraic

conditions. The first condition determines transitions between arbitrarily shaped regions in phase space, as opposed to QS which only handles  $n$ -dimensional rectangles. It also provides more accurate results by considering only the boundaries between regions. The second condition determines whether nearby trajectories approach a fixed point asymptotically. It obtains better results than QS by exploiting local stability properties. Both conditions apply to equations at all levels of abstraction: sign constraints, symbolic constraints, numeric values, and combinations thereof. The more information exists, the better they perform.

Qualitative reasoning includes more than transition analysis. Identifying attractors, which determine long-term behavior, is the next key problem. The monostability test of the previous section provides a method for deriving global fixed point attractors. Some additional methods for identifying attractors appear in my thesis [17], but many remain to be developed. Mathematicians, scientists, and engineers explore many properties of dynamic systems beyond transitions and attractors, including boundedness and periodicity of trajectories, the existence of unstable limit cycles, and the location of separatrices. Incorporating their techniques into qualitative reasoning is a topic for future research.

## References

- [1] Stephen Barnett. *Polynomials and Linear Control Systems*. Marcel Dekker, Inc., New York, 1983.
- [2] Daniel G. Bobrow, editor. *Qualitative Reasoning about Physical Systems*. MIT Press, 1985.
- [3] Johan de Kleer and Daniel G. Bobrow. Qualitative reasoning with higher-order derivatives. In *Proceedings of the National Conference on Artificial Intelligence*, pages 86–91. American Association for Artificial Intelligence, August 1984.
- [4] Johan de Kleer and John Seely Brown. A qualitative physics based on confluences. *Artificial Intelligence*, 24:7–83, 1984.
- [5] Johan de Kleer and John Seely Brown. Theories of causal ordering. *Artificial Intelligence*, 29:33–61, 1986.
- [6] Kenneth D. Forbus. Qualitative process theory. *Artificial Intelligence*, 24:85–168, 1984.
- [7] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1986.

- [8] Morris W. Hirsch and Stephen Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press College Division, Orlando, Florida, 1974.
- [9] Yumi Iwasaki and Herbert A. Simon. Causality in device behavior. *Artificial Intelligence*, 29:3–32, 1986.
- [10] Yumi Iwasaki and Herbert A. Simon. Theories of causal ordering: Reply to de Kleer and Brown. *Artificial Intelligence*, 29:63–72, 1986.
- [11] C. Jeffries, Victor Klee, and Pauline van den Driessche. When is a matrix sign stable? *Canadian Journal of Mathematics*, 29:315–326, 1977.
- [12] R. E. Kalman. Physical and mathematical mechanisms of instability in nonlinear automatic control systems. *Transactions of the ASME*, 79:553–566, April 1957.
- [13] Victor Klee and Pauline van den Driessche. Linear algorithms for testing the sign stability of a matrix and for finding z-maximum matchings in acyclic graphs. *Numerische Mathematik*, 28:273–285, 1977.
- [14] Benjamin Kuipers. Qualitative simulation. *Artificial Intelligence*, 29:289–338, 1986.
- [15] Ramon E. Moore. *Methods and Applications of Interval Analysis*. SIAM Studies in Applied Mathematics. SIAM, Philadelphia, 1979.
- [16] Elisha P. Sacks. Hierarchical reasoning about inequalities. In *Proceedings of the National Conference on Artificial Intelligence*, pages 649–654. American Association for Artificial Intelligence, 1987.
- [17] Elisha P. Sacks. Automatic qualitative analysis of ordinary differential equations using piecewise linear approximations. TR 416, Massachusetts Institute of Technology, Laboratory for Computer Science, 545 Technology Square, Cambridge, MA, 02139, March 1988. Also appears as AI-TR-1031.