Extending the Constraint Propagation of Intervals

Allen C. Ward* — Lecturer, Mechanical Engineering

Tomas Lozano-Perez — Associate Professor, Artificial Intelligence Laboratory

Warren P. Seering — Associate Professor, Mechanical Engineering

Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

We show that the usual notion of constraint propagation is but one of a number of similar inferences useful in quantitative reasoning about physical objects. These inferences are expressed formally as rules for the propagation of "labeled intervals" through equations. We prove the rules' correctness and illustrate their utility for reasoning about objects (such as motors or transmissions) which assume a continuum of different states. The inferences are the basis of a "mechanical design compiler", which has correctly produced detailed designs from "high level" descriptions for a variety of power transmission and temperature sensing systems.

1 Introduction

"Constraint propagation" is often thought to be a key element in design [1, 2, 3, 4, 5, 6, 7, 8, 10], hardware debugging [11] and spatial reasoning [12]. Intervals are among the most general constraints propagated; for example, given y — 2x and $1 < x < 2$, one concludes $2 < y < 4$. The meaning and validity of this inference seem intuitively clear, and research attention has generally focused on its computational characteristics.

In fact, we show here that the meaning of these statements and the validity of this inference, as applied to physical objects, requires more attention. More precisely, the statement $1 < x < 2$ can be considered a relationship between a variable name, an interval of values, and the permissible states of the physical object being described. Reasoning about physical objects can involve at least four different kinds of such relationships. Further, the inference shown exemplifies only one of three useful computations on equations and intervals; each of the three performs correct inferences only for appropriate interval-variable relationships.

We begin with an example demonstrating the utility of three kinds of interval propagation, then introduce four "labels" for interval-variable relationships. The bulk of the paper defines and proves the correctness of a variety of propagation inferences over "labeled intervals". Finally, we briefly discuss the application of these ideas in a "mechanical design compiler"— a program which takes as input a schematic, specifications, and a utility function for a mechanical design, and returns a description detailed enough to allow construction of an optimal implementation.

*This work was done at the MIT AI Lab, with funding from the Industrial Technology Institute of Ann Arbor, Michigan, and the Office of Naval Research

1.1 An Example

Figure 1 shows graphically the governing equation, $t_0 = rt_i$, for an ideal variable-speed mechanical transmission; here t_0 and t_i, are the output and input torques, and r is the continuously variable "transmission ratio". We use this equation to illustrate three different inferences.

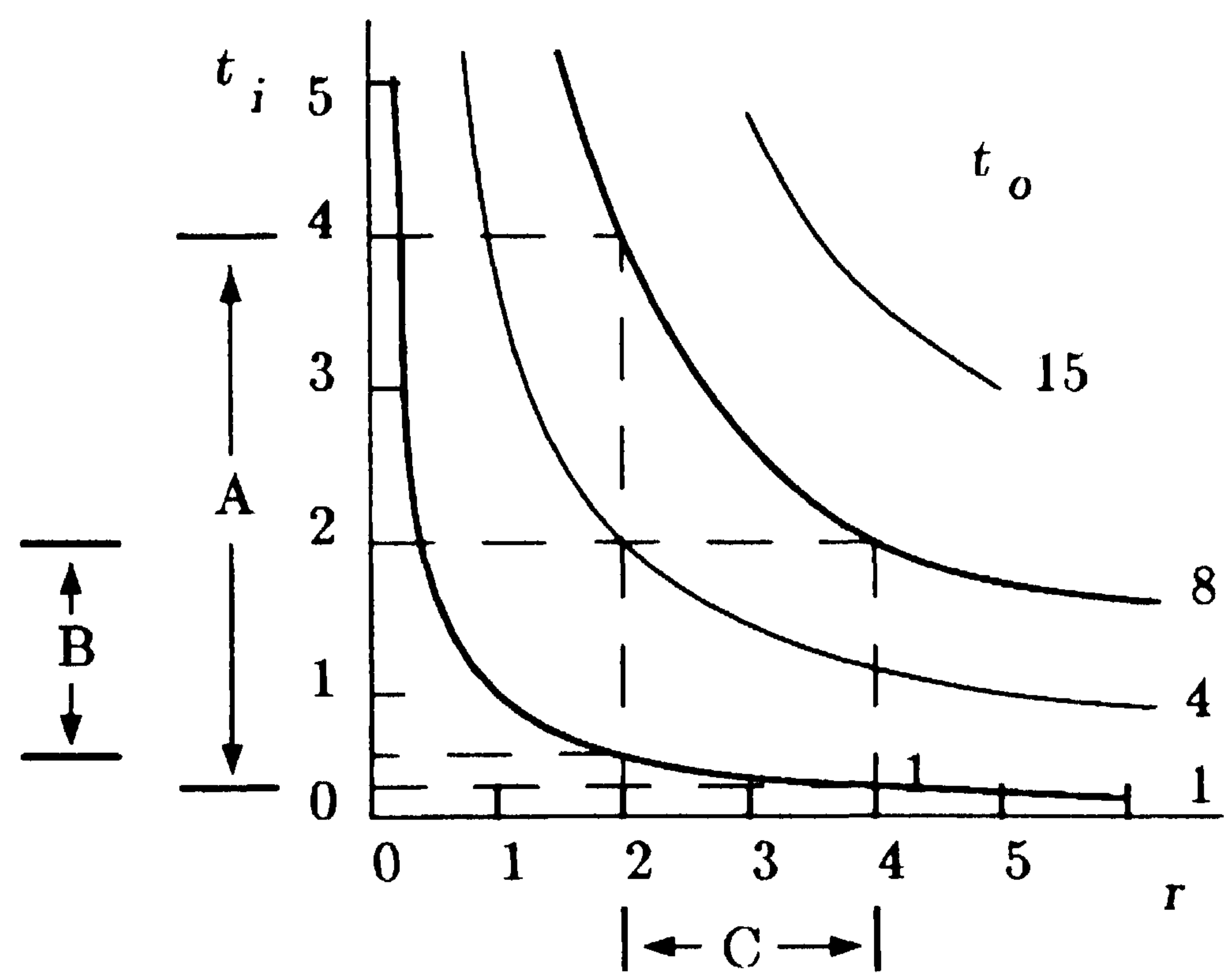

Figure 1: Inferences on a Mechanical Transmission

Case A: Suppose that the transmission ratio is limited to the interval from 2 to 4, and that if the output torque goes above 8 or falls to less than 1, it will damage the attached load. This seems clear enough: $2 < r < 4$, and $1 < t_0 < 8$. We want to pick motors which cannot damage the load, and conclude that the input or motor torque must fall in the interval A, from 0.25 to 4; $0.25 < t_i < 4$. This is the usual notion of interval constraint propagation.

Case B: In contrast, suppose that under the expected operating conditions the output torque must vary throughout the interval from 1 to 8 in order to drive the load. Note that we are not saying that the output torque is limited to the interval from 1 to 8; this interval means something else. With the same limits as in case A on the transmission ratio, we conclude that the motor torque must at least vary over the interval B, that is from 0.5 to 2, or the motor will fail to drive some load. This can't be "interval constraint propagation", since it gives different results with the same equation and interval inputs.

Case C: Now suppose that the transmission ratio is unknown, that the output torque must vary as in case B, and that the input torque is limited to the interval from 0.25 to 4. We conclude that the transmission must under some operating condition take on at least one value in the interval from 2 to 4, interval C; otherwise, at least one of the required output torques would be unattainable. "Interval constraint propagation" on $0.25 \leq t_i \leq 4$ and $1 \leq t_o \leq 8$ would give $0.25 \leq r \leq 32$. We will show later that this inference differs from that of Case B as well. Further, the ratio is not limited to the interval from 2 to 4, nor is it required to take on every value in this interval; this interval means something different still from those we have previously encountered.

The transmission equation relates the values for variables at a particular time. However, in each case, we used the equation to draw a conclusion about the set of values a variable could or should take on. Design is a natural area of application for such reasoning, because the designer must take into account the full variety of conditions under which his design must operate. Mechanical designers are in fact comfortable with the reasoning of the example, but if asked to justify it can provide only intuitive arguments. We will formalize these arguments, beginning by clarifying the possible relationships between variables, the states of an artifact, and intervals of values.

1.2 Assignment Intervals, and Equations

Let us suppose ourselves to be discussing an **object** of some sort. We describe this object using a set of **variable** names, and suppose that it can take on various **permissible states**; each state **assigns** each variable a value from the real number line.

We need some notation. We will use S to symbolize the set of permissible states, and s for an element of S. We will write $X = \langle x\ 0\ 2\rangle$ to mean the set of assignments of values in the interval $[0\ 2]$ to the variable x; $\mathbf{x} \in X$ to mean such an assignment; and $\mathbf{x}(s)$ to mean the function from states to assignments of the variable x.

Assignments inherit from the real numbers such relations as $<, =, \min$, in the obvious way. We can therefore refer to "intervals of assignments". If X is such an interval, then by $\mathbf{x}_l$ we will always mean $\min(X)$ and by $\mathbf{x}_h$, $\max(X)$. We will allow intervals to be infinite, e.g. $\langle x\ 2\ \infty\rangle$, but defer until the last section discussion of the computational implications of such intervals.

We can now introduce four kinds of statement about objects, their sets of permissible states, and intervals of assignments. We distinguish these by labeling the intervals, as in $\langle \mathit{label}\ x\ x_l\ x_h\rangle$, or just $\langle \mathit{label}\ X\rangle$, and refer to them as **labeled intervals**.

Definition 1 $\langle \overset{\text{only}}{[\ \ ]}\ X\rangle \overset{def}{\Longleftrightarrow} \forall s \in S, \exists \mathbf{x} \in X . \mathbf{x}(s) = \mathbf{x}$

That is, the permissible states assign values to x *only* from X; this is the interpretation given all three intervals in case A. This statement is actually a predicate on objects and sets of states, and could be written $\langle \overset{\text{only}}{[\ \ ]}\ X\rangle(\mathit{object}, S)$, but as we are considering only a single object and set of states we will leave them implicit.

Definition 2 $\langle \overset{\text{every}}{\leftrightarrow}\ X\rangle \overset{def}{\Longleftrightarrow} \forall \mathbf{x} \in X, \exists s \in S . \mathbf{x}(s) = \mathbf{x}$

That is, for *every* assignment in X, there exists a permissible state of the object making the assignment. This is the interpretation given both torque intervals in Case B.

Definition 3 $\langle \overset{\text{some}}{\cdots}\ X\rangle \overset{def}{\Longleftrightarrow} \exists s \in S . \exists \mathbf{x} \in X . \mathbf{x}(s) = \mathbf{x}$

That is, there exists *some* assignment $\mathbf{x}$ in X and state s in S such that s makes the assignment $\mathbf{x}$. Here we have the interpretation given the transmission ratio interval in case C.

Definition 4 $\langle \overset{\text{none}}{\bowtie}\ X\rangle \overset{def}{\Longleftrightarrow} \forall s \in S . \mathbf{x}(s) \notin X$

That is, there is *no* state $s \in S$ such that s makes any assignment in X. As an important exception to our normal custom, we here interpret X to be an open interval. If $\langle \overset{\text{none}}{\bowtie}\ X\rangle$ and X is semi-infinite, then we have $\langle \overset{\text{only}}{[\ \ ]}\ \overline{X}\rangle$, where $\overline{X}$ is the complement of X.

We will interpret equations describing the object as predicates on the permissible states of the object. More precisely, if the object is described using the equation $G(x, y, z) = 0$, then for every $s \in S$, $G(\mathbf{x}(s), \mathbf{y}(s), \mathbf{z}(s)) = 0$.

We impose tight restrictions on equations, discussing in [13] how these restrictions can be accommodated or loosened in practice. First, each equation must be implicit, and in three variables. (If we need equations of more than three variables to describe an object, we can use intermediate variables to convert them into systems of equations.) Second, over the domain of interest the equations must satisfy the **uniqueness** property; that is, if $G(x_0, y_0, z_1) = 0$ and $G(x_0, y_0, z_2) = 0$, then $z_1 = z_2$, and so on for permutations of variable names. Third, the domains of interest must be **compatible**; that is, for any permissible values of x, y there must be a permissible value of z satisfying the equation.

These constraints are sufficient to guarantee that the equation can be solved for each of the three variables, and that the resulting functions are strictly monotonic[1]. Finally, we require that these functions be continuous.

Given $G(x, y, z)$, we will write $g(\mathbf{x}, \mathbf{y})$ to mean the associated function from assignments in x and y to assignments in z.

[1] By strictly monotonic, for these functions of two variables, we mean that if $x_1 < x_2$ and $g(x_1, y_1) < g(x_2, y_1)$, then for all $x > x_1$, $g(x, y_0) > g(x_1, y_0)$, and so on for permutations of variable names. It can be shown for these equations that if these inequalities hold, and $y_1 < y_2$ and $g(x_1, y_1) < g(x_1, y_2)$, then for all $x > x_1, y > y_1$, $g(x, y) > g(x_1, y_1)$. Note that the transmission equation is strictly monotonic only over the positive reals.

2 Interval Operations and Inferences

We can now formalize a number of operations on intervals and equations, asking for which permutations of labeled intervals they perform correct inferences.

2.1 Conventional Constraint Propagation

We introduce first the operation used in the introduction's Case A.

Definition 5 $\textsc{Range}(G, X, Y) = \{\mathbf{z} | \exists \mathbf{x} \in X, \exists \mathbf{y} \in Y.G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0\}$

That is, the RANGE of the equation G with respect to the intervals of assignment X, Y is the set of assignments to the variable z such that there exist assignments in X and Y satisfying $G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$. This is of course simply the usual image of X, Y under $g(\mathbf{x}, \mathbf{y})$. The continuity of $g(\mathbf{x}, \mathbf{y})$ ensures that Z is an interval. Trivially, RANGE is commutative in the intervals; that is, $\textsc{Range}(G, X, Y) = \textsc{Range}(G, Y, X)$.

Recall that by $\mathbf{x}_l$ and $\mathbf{x}_h$ we mean $\min(X)$ and $\max(X)$ respectively; then to compute RANGE we use:

Definition 6 $\textsc{Corners}(G, X, Y) = \{g(\mathbf{x}_l, \mathbf{y}_l), g(\mathbf{x}_h, \mathbf{y}_l), g(\mathbf{x}_l, \mathbf{y}_h), g(\mathbf{x}_h, \mathbf{y}_h)\}$.

This leads to

Lemma 1 $\textsc{Range}(G, X, Y) = Z = [\min(\textsc{Corners}(G, X, Y))\ \max(\textsc{Corners}(G, X, Y))]$. *Further, if* $\mathbf{z}_l = \min(Z) = g(\mathbf{x}_i, \mathbf{y}_1)$, *and* $\mathbf{z}_h = g(\mathbf{x}_j, \mathbf{y}_2)$, *with* $\mathbf{y}_1, \mathbf{y}_2 \in Y$, *then* $\{\mathbf{x}_i, \mathbf{x}_j\} = \{\mathbf{x}_l, \mathbf{x}_h\}$.

The idea is that the maximum and minimum of a monotonic function over a pair of intervals occur at the endpoints of the intervals; further, they occur at different endpoints. The lemma of course holds for permutations of the variable names. The proof follows directly from the monotonicity and continuity of $g(\mathbf{x}, \mathbf{y})$.

For which combinations of labeled intervals does the RANGE operation produce correct inferences? We begin with the most obvious.

Rule 1 $\langle \overset{\text{only}}{[\ ]}\ X\rangle \& \langle \overset{\text{only}}{[\ ]}\ Y\rangle \& G(x, y, z) = 0 \longrightarrow \langle \overset{\text{only}}{[\ ]}\ \textsc{Range}(G, X, Y)\rangle$

That is, if for every permissible state $G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0$ is satisfied, $\mathbf{x}(s)$ is in X and $\mathbf{y}(s)$ is in Y, then $\mathbf{z}(s)$ is in the image of X, Y under $g(\mathbf{x}, \mathbf{y})$. This follows directly from the definition of RANGE.

This rule expresses the inference of Case A. Recall that the output torque of the transmission should not go above 8 or below 1; $\langle \overset{\text{only}}{[\ ]}\ t_o\ 1\ 8\rangle$. The transmission ratio could not go below 2 or above 4; $\langle \overset{\text{only}}{[\ ]}\ r\ 2\ 4\rangle$. These, with the equation $t_o = rt_i$, match the antecedents of Rule 1. The CORNERS operation substitutes the endpoints of these intervals into $\frac{t_o}{r}$, returning assignments to t_i of $\{0.5, 0.25, 4, 2\}$; and the RANGE operation extracts the maximum and the minimum to form $\langle \overset{\text{only}}{[\ ]}\ t_i\ 0.25\ 4\rangle$, the limits on the input torque.

We also have

Rule 2 $\langle \overset{\text{only}}{[\ ]}\ X\rangle \& \langle \overset{\text{some}}{\cdots}\ Y\rangle \& G(x, y, z) = 0 \longrightarrow \langle \overset{\text{some}}{\cdots}\ \textsc{Range}(G, X, Y)\rangle$

Proof: By the definition of $\overset{\text{some}}{\cdots}$, there is some $s \in S$ such that $\mathbf{y}(s) \in Y$, and by the definition of $\overset{\text{only}}{[\ ]}$, $\mathbf{x}(s)$ is certainly in X. Then $\mathbf{z}(s) = g(\mathbf{x}(s), \mathbf{y}(s))$ is in $\textsc{Range}(G, X, Y)$, so $\langle \overset{\text{some}}{\cdots}\ \textsc{Range}(G, X, Y)\rangle$ is satisfied.

In contrast, the possible rule $\langle \overset{\text{some}}{\cdots}\ X\rangle \& \langle \overset{\text{some}}{\cdots}\ Y\rangle \& G(x, y, z) = 0 \longrightarrow \langle \overset{\text{some}}{\cdots}\ \textsc{Range}(G, X, Y)\rangle$ is invalid, because the assignment in X and the assignment in Y need not occur simultaneously. Consider, for example, the labeled intervals $\langle \overset{\text{some}}{\cdots}\ x\ 2\ 3\rangle$, $\langle \overset{\text{some}}{\cdots}\ y\ 1\ 4\rangle$, and equation $xy - z = 0$, and the following consistent and complete set of states:

State	x	y	z
s_1	2.5	0	0
s_2	0	2	0

The rule would incorrectly imply $\langle \overset{\text{some}}{\cdots}\ Z\ 2\ 12\rangle$. The same objection applies to the possible rule, $\langle \overset{\text{some}}{\cdots}\ X\rangle \& \langle \overset{\text{every}}{\leftrightarrow}\ Y\rangle \longrightarrow \langle \overset{\text{some}}{\cdots}\ \textsc{Range}(G, X, Y)\rangle$.

The possible rule $\langle \overset{\text{none}}{\bowtie}\ X\rangle \& \langle \overset{\text{none}}{\bowtie}\ Y\rangle \& G(x, y, z) = 0 \longrightarrow \langle \overset{\text{not}}{\bowtie}\ \textsc{Range}(G, X, Y)\rangle$ is also invalid. Consider an object described only by $\langle \overset{\text{none}}{\bowtie}\ x\ 2\ 3\rangle$, $\langle \overset{\text{none}}{\bowtie}\ y\ 1\ 4\rangle$, $xy - z = 0$. Then a state assigning $x = 1, y = 6, z = 6$ is permissible, and $\langle \overset{\text{none}}{\bowtie}\ Z\ 2\ 12\rangle$ is false. However, let us divide the complement of X into two intervals, $\overline{X}_l = \{\mathbf{x} | \mathbf{x} < \min(X)\}$, and $\overline{X}_h = \{\mathbf{x} | \mathbf{x} > \max(X)\}$. Then, using the symbol $\otimes$ for RANGE, we have

Rule 3 $\langle \overset{\text{none}}{\bowtie}\ X\rangle \& \langle \overset{\text{none}}{\bowtie}\ Y\rangle \& G(x, y, z) \longrightarrow \langle \overset{\text{none}}{\bowtie}\ \overline{\otimes(G, \overline{X}_l, \overline{Y}_l)} \cap \overline{\otimes(G, \overline{X}_h, \overline{Y}_l)} \cap \overline{\otimes(G, \overline{X}_l, \overline{Y}_h)} \cap \overline{\otimes(G, \overline{X}_h, \overline{Y}_h)}\rangle$

The intuition for this rather forbidding expression is that since $\mathbf{x}$ and $\mathbf{y}$ can't be in X and Y, they must be in *overlineX* and $\overline{Y}$; these complements can be divided into two intervals each; and $\mathbf{z}$ must be in the RANGE of one pair of such complement intervals. Hence, $\mathbf{z}$ cannot be in the intersection of the complements of those RANGEs. We might suppose that we could label the union of the RANGEs with a $\overset{\text{only}}{[\ ]}$ label, indicating that the assignment must fall in that union, but that union is not an interval. The intersection of the complements *is* an interval, because: $\overline{X}_l$, $\overline{X}_h$, $\overline{Y}$, and $\overline{Y}_h$ are semi-infinite intervals, the RANGEs of the pairs are also semi-infinite intervals, the complements of the RANGEs are semi-infinite intervals, and the intersection of intervals is an interval.

The formal proof of Rule 3 is simple. Let the consequent interval equal Z, and suppose there is some $s \in S$ such that $\mathbf{z}(s) \in Z$. By the antecedents, $\mathbf{x}(s) \in \overline{X}_j$ and $\mathbf{y}(s) \in \overline{Y}_k$, for j and k in $\{h, l\}$. But then $\mathbf{z}(s)$ is in $\textsc{Range}(G, \overline{X}_j, \overline{Y}_k)$, contrary to the definition of Z.

For the final rule of this section, we need:

Definition 7 $\textsc{Independent}(X, Y, S)$ *if and only if for any* $\mathbf{x} \in X$ *such that* $\mathbf{x} = \mathbf{x}(s_1)$ *and* $\mathbf{y} \in Y$ *such that* $\mathbf{y} = \mathbf{y}(s_2)$, *with* s_1 *and* s_2 *in* S, *then there is an* $s \in S$ *such that* $\mathbf{x}(s) = \mathbf{x}$ *and* $\mathbf{y}(s) = y$.

As usual, we will often leave S implicit.

If for every pair $\mathbf{x}, \mathbf{y}$ in $X \times Y$ there is a state making these assignments, and G is true in every state, then there is a state making every assignment in $\{\mathbf{z} | \exists \mathbf{x} \in X, \exists \mathbf{y} \in Y.G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0\}$. We therefore have:

Rule 4
$\text{INDEPENDENT}(X,Y)\&\langle \overset{\text{every}}{\leftrightarrow} X\rangle\&\langle \overset{\text{every}}{\leftrightarrow} Y\rangle\&G(x,y,z) = 0 \longrightarrow \langle \overset{\text{every}}{\leftrightarrow} Z\ \text{RANGE}(G,X,Y)\rangle$.

2.2 The DOMAIN Operation

We turn now to case B of the introduction, and define a partial inverse of RANGE. That is,

Definition 8
$\text{DOMAIN}(G,Z,X) = Y \overset{def}{\Longleftrightarrow} \text{RANGE}(G,Y,X) = Z$

DOMAIN is partial because for some G, Z, X there is no assignment interval Y which satisfying this definition; the computation process given below readily identifies such cases. The following rules apply only when such a Y exists.

Note that since RANGE is commutative with respect to its interval arguments, $\text{DOMAIN}(G,Z,X) = Y$ implies $\text{DOMAIN}(G,Z,Y) = X$.

DOMAIN has an equivalent direct definition.

Lemma 2 $\text{DOMAIN}(G,Z,X) = \{\mathbf{y} | \forall \mathbf{x} \in X, \exists \mathbf{z} \in Z.G(\mathbf{x},\mathbf{y},\mathbf{z}) = 0\}$

Proof: Let $\text{DOMAIN}(G,Z,X) = Y$, and $Y' = \{\mathbf{y}|\forall \mathbf{x} \in X, \exists \mathbf{z} \in Z.G(\mathbf{x},\mathbf{y},\mathbf{z}) = 0\}$; we must show that $Y = Y'$. Suppose $\mathbf{y}_0 \in Y$. By the compatibility property, for every $\mathbf{x} \in X$, there exists some $\mathbf{z}_0$ such that $G(\mathbf{x},\mathbf{y}_0,\mathbf{z}_0) = 0$. But by the definition of DOMAIN, RANGE(G,X,Y) = Z, and by the definition of RANGE, $Z = \{\mathbf{z}|\exists \mathbf{x} \in X, \exists \mathbf{y} \in Y.G(\mathbf{x},\mathbf{y},\mathbf{z}) = 0\}$, hence $\mathbf{z}_0$ is in Z; that is, for every $\mathbf{x} \in X$ there is a $\mathbf{z}_0 \in Z$ such that $G(\mathbf{x},\mathbf{y}_0,z_0) = 0$. $\mathbf{y}_0$ then satisfies $\forall \mathbf{x} \in X, \exists \mathbf{z} \in Z.G(\mathbf{x},\mathbf{y}_0,\mathbf{z}) = 0$, and $\mathbf{y}_0 \in Y'$.

For the converse, we show first that the endpoints of Y' are in Y. Let $\mathbf{y}'_l = \min(Y')$, and let $g(\mathbf{x}_l,\mathbf{y}'_l) = \mathbf{z}_1$, $g(\mathbf{x}_h,\mathbf{y}'_l) = \mathbf{z}_2$.

At least one of $\mathbf{z}_1, \mathbf{z}_2$ must be an endpoint of Z. To see this, assume the converse; $\mathbf{z}_l < \mathbf{z}_1 < \mathbf{z}_h$, and $\mathbf{z}_l < \mathbf{z}_2 < \mathbf{z}_h$. Since $g(\mathbf{x},\mathbf{y})$ is continuous and monotonic, we can choose some point $\mathbf{y}' < \mathbf{y}'_l$, but sufficiently close to $\mathbf{y}'_l$ that $\mathbf{z}_l < g(\mathbf{x}_l,\mathbf{y}') < z_h$ and $z_l < g(\mathbf{x}_h,\mathbf{y}') < \mathbf{z}_h$. But then, $\mathbf{z} = g(\mathbf{x},\mathbf{y}') \in Z$ for all $\mathbf{x} \in X$, and by the definition of Y', $\mathbf{y}' \in Y$. This contradicts the definition of $\mathbf{y}'_l$ as $\min(Y')$; hence, the assumption is false, and at least one of $\mathbf{z}_1,\mathbf{z}_2$ is an endpoint of Z.

Let $\mathbf{z}_k$ designate the element of $\{\mathbf{z}_1,\mathbf{z}_2\}$ which is an endpoint of Z, and let $\mathbf{x}_j$ designate the corresponding element of $\{\mathbf{x}_l,\mathbf{x}_h\}$; that is, $g(\mathbf{x}_j,\mathbf{y}'_l) = \mathbf{z}_k$. Lemma 1 and the uniqueness property of of G then imply that $\mathbf{y}'_l = \mathbf{y}_l$.

We can use symmetrical reasoning to conclude that, $\mathbf{y}'_h$ is also an endpoint of Y, then use Lemma 1 again to conclude that they are different endpoints (unless Y is a single point). Y is an interval by definition (RANGEis defined only on interval inputs), so all the assignments between $\mathbf{y}'_l$ and $\mathbf{y}'_h$ are also in Y, and $Y' = Y$.

To compute DOMAIN, we rely on:

Lemma 3 *If* $\text{DOMAIN}(G,Z,X) = Y$, *then* $\mathbf{y}_l = \min(Y)$ *and* $\mathbf{y}_h = \max(Y)$ *are in* $\text{CORNERS}(g,Z,X)$.

Proof: $\text{RANGE}(G,X,Y) = Z$, so $\mathbf{z}_l$ and $\mathbf{z}_h$ are in $\text{CORNERS}(G,X,Y)$. Thus, for some values i and j in $\{l,h\}$, $g(\mathbf{x}_i,\mathbf{y}_j) = \mathbf{z}_l$. But then by the uniqueness property of G, $g(\mathbf{x}_i,\mathbf{z}_l) = \mathbf{y}_i$, so $\mathbf{y}_i$ is in $\text{CORNERS}(G,Z,X)$. Using the same reasoning with $\mathbf{z}_h$ and using Lemma 1, we conclude that $\mathbf{y}_l$ and $\mathbf{y}_h$ must be in $\text{CORNERS}(G,Z,X)$.

We can therefore compute $\text{DOMAIN}(G,Z,X)$ by generating each possible $Y_t = \langle y\ y_1\ y_2\rangle$ where $y_1,y_2 \in \text{CORNERS}(G,Z,X)$ and $y_1 \leq y_2$, then testing whether $\text{RANGE}(G,X,Y_t) = Z$.

To formulate our next rule, we need one more definition.

Definition 9 *Let* $\mathbf{x}(s_1) < \mathbf{x} < \mathbf{x}(s_2)$ *for some* $s_1,s_2 \in S$; *if this implies that there exists some* $s \in S$ *such that* $\mathbf{x} = \mathbf{x}(s)$, *then* $\mathbf{x}$ *is* STATE-CONTINUOUS.

We then have

Rule 5 $\langle \overset{\text{every}}{\leftrightarrow} Z\rangle\&\langle \overset{\text{only}}{[\ ]} X\rangle\&G(x,y,z) = 0\&\text{STATE-CONTINUOUS}(y) \longrightarrow \langle \overset{\text{every}}{\leftrightarrow} \text{DOMAIN}(G,Z,X)\rangle$

Proof: $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ are either positive or negative throughout the domain. There are four permutations of these signs; we consider one in detail. Throughout, let $Y = \text{DOMAIN}(G,Z,X)$. The idea is to show that there must be states making assignments of y on either side of Y.

Suppose $\frac{\partial z}{\partial x} < 0$, and $\frac{\partial z}{\partial y} < 0$. Then, $g(\mathbf{x}_h,\mathbf{y}_h) = \mathbf{z}_l$, and by the first antecedent of the rule we can choose $s_l \in S$ such that $\mathbf{z}(s_l) = \mathbf{z}_l$. By the second antecedent, $\mathbf{x}(s_l) \in X$, so $\mathbf{x}(s_l) \leq \mathbf{x}_h$. Recall that $G(\mathbf{x}(s),\mathbf{y}(s),\mathbf{z}(s)) = 0$ for all $s \in S$, hence by the uniqueness property of G, $g(\mathbf{x}(s_l),\mathbf{y}(s_l)) = \mathbf{z}(s_l)$. Now, assume that $\mathbf{y}(s_l) < \mathbf{y}_h$; by the assumed signs of the partial derivatives, $z(s_l) = g(\mathbf{x}(s_l),\mathbf{y}(s_l)) > g(\mathbf{x}(s_l),\mathbf{y}_h) > g(\mathbf{x}_h,\mathbf{y}_h) = \mathbf{z}(s_l)$. The assumption must be invalid, and $\mathbf{y}(s_l) \geq \mathbf{y}_h$.

Symmetrical reasoning leads to the conclusion that there is some $s_h \in S$ such that $\mathbf{y}(s_h) \leq \mathbf{y}_l$. Then, by the STATE-CONTINUOUS assumption, for each $\mathbf{y} \in [\mathbf{y}(s_h)\ \mathbf{y}(s_l)] \supseteq Y$, there is an $s \in S$ such that $\mathbf{y}(s) = \mathbf{y}$.

Rule 5 performs the inference of case B. Recall that we kept the limit specification on the ratio, $\langle \overset{\text{only}}{[\ ]}\ r\ 2\ 4\rangle$, but changed the output torque specification to require that the output torque take on every value in the interval: $\langle \overset{\text{every}}{\leftrightarrow} t_o\ 1\ 8\rangle$. The input torque is STATE-CONTINUOUS. The CORNERS operation again substitutes the endpoints of these intervals into $\frac{t_o}{r}$, returning $\{0.5, 0.25, 4, 2\}$. DOMAIN picks out $\langle \overset{\text{every}}{\leftrightarrow} t_i\ 0.5\ 2\rangle$ by substituting into $t_o = rt_i$ and checking the result against $\langle t_o\ 1\ 8\rangle$; $1 = (.5)(2)$ and $8 = (2)(4)$.

The STATE-CONTINUOUS assumption is physically significant. Suppose for our transmission we had $\langle \overset{\text{every}}{\leftrightarrow} t_o\ 6\ 12\rangle$, and $\langle t_i\ 2\ 3\rangle$. Rule 5 gives $\langle \overset{\text{every}}{\leftrightarrow} r\ 3\ 4\rangle$, which is correct for our variable speed transmission. However the requirements could be satisfied with a two-speed geared transmission with ratios 2.9 and 4.1.

Nothing in this section proves the existence of $\text{DOMAIN}(G,Z,X)$, and indeed this set may not exist. In Case C, for example, there is no set R of assignments to the ratio r such that $\text{RANGE}(t_o - rt_i =$

$0, \langle t_i\ 0.25\ 4\rangle, R) = \langle t_o\ 1\ 8\rangle$. In this case, however, we can apply the next operation.

2.3 The Sufficient-Points Operation

Let us begin by extending RANGE to operate on an interval and a point (of assignment), defining $\text{RANGE}(G, Y, \mathbf{x}_0) = \text{RANGE}(G, Y, [\mathbf{x}_0\ \mathbf{x}_0])$. Then the SUFFICIENT-POINTS operation is defined by

Definition 10
$\text{SUFPT}(G, Z, X) = \{\mathbf{y} | Z \subseteq \text{RANGE}(G, X, \mathbf{y})\}$

We need to show that if $Y = \{\mathbf{y} | Z \subseteq \text{RANGE}(G, X, \mathbf{y})\}$ exists it is an interval; that is, if $\mathbf{y}_1 < \mathbf{y}_2 < \mathbf{y}_3$, and $\mathbf{y}_1, \mathbf{y}_3 \in Y$, then $\mathbf{y}_2 \in Y$. Consider first the case where $\frac{\partial z}{\partial y} < 0$. Since $Z \subseteq \text{RANGE}(G, X, \mathbf{y}_1)$ we can find some $\mathbf{x}_i \in X$ such that $g(\mathbf{x}_i, \mathbf{y}_1) \leq \mathbf{z}_l$. Then, $g(\mathbf{x}_i, \mathbf{y}_2) < \mathbf{z}_l$. Alternatively, if $\frac{\partial z}{\partial y} > 0$, find $\mathbf{x}_i$ such that $g(\mathbf{x}_i, \mathbf{y}_3) \leq \mathbf{z}_l$, and again $g(\mathbf{x}_i, \mathbf{y}_2) < \mathbf{z}_l$. By symmetrical arguments there is also some $\mathbf{x}_j \in X$ such that $g(\mathbf{x}_j, \mathbf{y}_2) > \mathbf{z}_h$. Thus, $Z \subseteq \text{RANGE}(G, X, \mathbf{y}_2)$, and $\mathbf{y}_2 \in Y$.

There is an equivalent direct definition of SUFPT.

Lemma 4 $\text{SUFPT}(G, Z, X) = \{\mathbf{y} | \forall \mathbf{z} \in Z, \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0\}$

Proof: Let $\text{SUFPT}(G, Z, X) = Y$, and $Y' = \{\mathbf{y} | \forall \mathbf{z} \in Z, \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}, \mathbf{z}) = 0\}$; we need to show that Y and Y' are equal. If $\mathbf{y}_0 \in Y$, let $Z_0 = \text{RANGE}(G, X, \mathbf{y}_0) = \{\mathbf{z} | \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}_0, \mathbf{z}) = 0\}$. But by the definition of Y and SUFPT, Z_0 is a superset of Z, so certainly $\forall z \in Z, \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}_0, z) = 0$, and $\mathbf{y}_0 \in Y'$. Conversely, if $\mathbf{y}_0 \in Y'$, then $\forall \mathbf{z} \in Z \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}_0, \mathbf{z}) = 0$; but then Z is a subset of $\{\mathbf{z} | \exists \mathbf{x} \in X . G(\mathbf{x}, \mathbf{y}_0, \mathbf{z}) = 0\} = \text{RANGE}(G, X, \mathbf{y}_0)$, so $\mathbf{y}_0 \in Y$.

As with DOMAIN, we can calculate SUFFICIENT-POINTS by testing various combinations of $\text{CORNERS}(G, Z, X)$. We need

Lemma 5 *If* $\text{SUFPT}(G, Z, X) = Y$, *then* $\mathbf{y}_l$ *and* $\mathbf{y}_h$ *are in* $\text{CORNERS}(G, Z, X)$.

Proof: Consider first $\mathbf{y}_h = \max(Y)$; let $Z' = [\mathbf{z}'_l\ \mathbf{z}'_h] = \text{RANGE}(G, X, \mathbf{y}_h)$
$= [\min(g(\mathbf{x}_l, \mathbf{y}_h), g(\mathbf{x}_h, \mathbf{y}_h))\ \ \max(g(\mathbf{x}_l, \mathbf{y}_h), g(\mathbf{x}_h, \mathbf{y}_h))]$. Then, either $\mathbf{z}_l$ or $\mathbf{z}_h$ must be an endpoint of Z'. Otherwise, $\mathbf{z}'_l < \mathbf{z}_l < \mathbf{z}_h$, and $\mathbf{z}_l < \mathbf{z}_h < \mathbf{z}'_h$, and since $g(\mathbf{x}, \mathbf{y})$ is continuous, we can choose some point $\mathbf{y}' > \mathbf{y}_h$ sufficiently close to $\mathbf{y}_h$ such that $\min(g(\mathbf{x}_l, \mathbf{y}'), g(\mathbf{x}_h, \mathbf{y}')) < z_l$ and $\max(g(\mathbf{x}_l, \mathbf{y}'), g(\mathbf{x}_h, \mathbf{y}')) > \mathbf{z}_h$. But then $Z \subseteq \text{RANGE}(G, X, \mathbf{y}')$, and y' should also be in Y.

We can therefore find a $\mathbf{z}_i$ which is an endpoint of both Z and Z'. By the definition of Z', $G(\mathbf{x}_j, \mathbf{y}_h, \mathbf{z}_i) = 0$ for some $\mathbf{x}_j \in \{\mathbf{x}_l, \mathbf{x}_h\}$. But then $\mathbf{y}_h = g(\mathbf{x}_j, \mathbf{z}_i)$, and by the definitions of CORNERS(,) $\mathbf{y}_h \in \text{CORNERS}(G, X, Z)$. A symmetric argument holds for $\mathbf{y}_l$.

We consider two inferences using SUFFICIENT-POINTS. First,

Rule 6 $\langle \overset{\text{every}}{\leftrightarrow} Z\rangle \& \langle \overset{\text{only}}{[\ ]} X\rangle \& G(x, y, z) = 0$
$\& \text{STATE-CONTINUOUS}(y) \longrightarrow \langle \overset{\text{some}}{\cdots} \text{SUFPT}(G, Z, X)\rangle$

Proof: Let $\text{SUFPT}(G, Z, X) = Y$, and $\overline{Y}_l = \{\mathbf{y} | \mathbf{y} < \min(Y)\}$. For $\mathbf{y} \in \overline{Y}_l$, at least one endpoint $\mathbf{z}_k$ of Z is such that $G(\mathbf{x}, \mathbf{y}, \mathbf{z}_k) \neq 0$ for any $\mathbf{y} \in \overline{Y}_l, \mathbf{x} \in X$. By the first antecedent, there is an $s_1 \in S$ such that $\mathbf{z}(s_1) = \mathbf{z}_k$, and by the second, $\mathbf{x}(s_1)$ is in X, so $\mathbf{y}(s_1)$ must be greater than $\max(\overline{Y}_l)$; thus, there is an $s_1 \in S$ such that $\mathbf{y}(s_1) \geq \min(Y)$. By a symmetrical argument, there is an $s_2 \in S$ such that $\mathbf{y}(s_2) \leq \max(Y)$. Either at least one of $\mathbf{y}(s_1), \mathbf{y}(s_2)$ is an element of Y, or Y is included in the interval between them, in which case the STATE-CONTINUOUS assumption requires a state for every $\mathbf{y} \in Y$.

This rule performs the inference of Case C. We required the output torque to take on all values in an interval, $\langle \overset{\text{every}}{\leftrightarrow} t_o\ 1\ 8\rangle$, but restricted the input torque, $\langle \overset{\text{only}}{[\ ]} t_i\ 0.25\ 4\rangle$. Rule 6 applies since the transmission ratio is continuously variable. CORNERS, using $r = \frac{t_o}{t_i}$ returns $\{4, 32, 0.25, 2\}$. Of these, $\text{RANGE}(t_o = rt_i, \langle t_i\ 0.25\ 4\rangle, r = 2)$ returns $\langle t_i\ .5\ 8\rangle$, which is a superset of $\langle t_o\ 1\ 8\rangle$. $r = 4$ also passes this test, but not $r = 0.25$ or $r = 32$. Hence the rule requires the transmission ratio to take on at least one value in [2 4]; $\langle \overset{\text{some}}{\cdots} r\ 2\ 4\rangle$.

For the second inference, we need another predicate on variables.

Definition 11 PARAMETER(x) *if and only if there is some single assignment* $\mathbf{x}_0$ *such that for all* $s \in S$, $\mathbf{x}(s) = \mathbf{x}_0$.

In the design context, PARAMETER(x) implies that the value of x is fixed at manufacture.

Rule 7 $\langle \overset{\text{every}}{\leftrightarrow} Z\rangle \& \langle \overset{\text{only}}{[\ ]} X\rangle \& G(x, y, z) = 0$
$\& \text{PARAMETER}(y) \longrightarrow \langle \overset{\text{only}}{[\ ]} \text{SUFPT}(G, Z, X)\rangle$

To prove this, one applies the same reasoning as for Rule 6, then notes that since y takes on only one value, that value must be between $\max(Y)$ and $\min(Y)$.

3 Some Application Problems

The rules derived above form part of a mechanical design compiler. This program accepts specifications, a utility function, and a schematic for a mechanical design, and returns catalog numbers for an optimal implementation[2]. Implementation of the compiler involves some difficulties we avoided in the preceding discussion.

3.1 Reasoning About Sets of Artifacts

The most important complication is that while throughout this paper we deal with representations of single objects, the compiler actually works with representations of sets of objects. [13] discusses these issues in detail; here we present only a sketch of some of the essential ideas.

Basic sets of objects are those corresponding to a particular catalog number; because of manufacturing tolerances, no two of these will be exactly the same. These

[2]The catalog numbers, together with the schematics, would usually be sufficient in the test domains to support construction by skilled mechanics. Extension to domains in which many components must be specially machined for the particular design remains a research issue.

can be described using labeled intervals; each labeled interval description is true of each object in the set.

From the basic sets, we automatically build an abstraction hierarchy, formulating labeled interval descriptions which are true of each object in the abstracted supersets. Thus, the "cylinder" symbol in a hydraulic system schematic represents all the hydraulic cylinders in a particular catalog. Since the describing statements are true for every cylinder, the rules we have described can be used to propagate labeled intervals describing the "load", thereby inferring statements about the pumps and motors. Conflicts between these statements, and those describing the basic sets, are used to eliminate inappropriate basic sets. A binary search is used to find the best of the surviving implementations.

While the rules derived here remain valid, the irreducibility of basic sets introduces additional rules, discussed in [13].

3.2 Relaxing the Monotonicity Assumption

Most of the equations describing mechanical artifacts are not monotonic over the real numbers. However, for a wide variety of designs it is possible to restrict values to the non-negative reals, producing strict monotonicity except perhaps at zero.

The CORNERS function may then involve divisions by zero. We extend division in the obvious ways: divisions of non-zero numbers by zero return ∞; divisions and multiplications of numbers by ∞ return zero and ∞ respectively. On dividing zero by zero, or multiplying zero by ∞, CORNERS returns a list including both zero and ∞.

The DOMAIN operation also needs modification. Consider again the transmission problem, where G is $t_o - rt_i = 0$. Suppose the output torque must assume every value in the operating region $\langle \overset{\text{every}}{\leftrightarrow}\ t_o\ 0\ 8\rangle$, while the input torque is limited by $\langle t_i\ 0\ 2\rangle$. Applying Rule 5, the CORNERS operation returns $\{0, \infty, \infty, 0, 4\}$. Now, $\text{RANGE}(G, \langle \overset{\text{only}}{[\ ]}\ t_i\ 0\ 2\rangle, \langle r\ 0\ 4\rangle) = \langle T_o\ 0\ 8\rangle$, but in fact there is no need for the transmission ratio to drop to 0; any transmission ratio greater than 4 will do. For this rule, then, we modify the DOMAIN operation so that it looks for the minimal interval in r such that $\text{RANGE}(G, T_i, R) \supseteq T_o$. In this case, there is no such interval, and this rule make no inference. Instead, Rule 6 returns $\langle \overset{\text{some}}{\cdots}\ r\ 4\ \infty\rangle$.

3.3 Performance

We discuss the expressive power of the labeled interval language and the performance of the compiler in detail in [14]. Here we remark only that the compiler has been tested on a wide variety of mechanical and hydraulic power train designs, as well a few temperature sensing systems. Some of these designs represent more than a million alternative solutions; the compiler has been able to select a solution, in each case, in less than twenty minutes. The solutions obtained seem consistently optimal; the time required to compile designs seems to grow as the logarithm of the number of alternatives represented, or linearly as the number of equations or variables used to describe them. The compiler has not been used on designs involving feedback loops, or where dynamic (as opposed to quasi-static) performance is important.

References

[1] Allen C. Ward and Warren Seering. An approach to computational aids for mechanical design. In *Proc. 1981 International Conference on Engineering Design.* ASME, 1987.

[2] David Serrano. *Constraint Management in Conceptual Design.* PhD thesis, Massachusetts Institute of Technology, 1989.

[3] R. J. Popplestone. The Edinburgh Designer system as a framework for robotics: the design of behavior. *Artifical Intelligence for Engineering Design, Analysis, and Manufacturing,* 1(1), 1987.

[4] Alan Fleming. Analysis of uncertainties in a structure of parts. In *International Joint Conference on Artificial Intelligence,* 1985.

[5] J. R. Dixon, E. C. Libardi Jr., S. C. Luby, M. Vaghul, and M. K. Simmons. Expert systems for mechanical design: Examples of symbolic representations of design geometries. In *Applications of Knowledge-based Systems to Engineering Analysis and Design.* ASME, New York, NY, 1985.

[6] Gerald Jay Sussman and Guy Lewis Steele. Constraints—a language for expressing almost heirarchical descriptions. *Artificial Intelligence,* 14, 1980.

[7] Jack Mostow, Lou Steinberg, Noshir Langrana, and Chris Tong. A domain-independent model of knowledge-based design: Progress report to the National Science Foundation. Technical Report Working Paper 90-1, Rutgers University, 1988.

[8] Mark Donald Gross. *Design as Exploring Constraints.* PhD thesis, Massachusetts Institute of Technology, 1985.

[9] S. Mittal, C. L. Dym, and M. Morjaria. PRIDE: An expert system for the design of paper paths. In *Applications of Knowledge-based Systems to Engineering Analysis and Design.* ASME, New York, NY, 1985.

[10] Walter Hamscher. Using structural and functional information in diagnostic design. Technical Report 707, MIT Artificial Intelligence Laboratory, June 1983.

[11] Ernest Davis. Constraint propagation with interval labels. *Artificial Intelligence,* 32, 1987.

[12] Allen C. Ward. *A Theory of Quantitative Inference for Artifact Sets, Applied to a Mechanical Design Compiler.* PhD thesis, Massachusetts Institute of Technology, 1989.

[13] Allen C. Ward and Warren Seering. The performance of a mechanical design compiler. In *Proceedings of the 1989 International Conference on Engineering Design,* August 1989.