Yan Wang

Assistant Professor University of Central Florida 4000 Central Florida Blvd. Orlando, FL 32816-2993 wangyan@mail.ucf.edu

Semantic Tolerance Modeling with Generalized Intervals

A significant amount of research has been performed to explore the mathematical basis for dimensional and geometric tolerance representation, analysis, and synthesis. However, tolerancing semantics such as logical dependency among variations and sequence of specifications is not maintained in these models. Consequently, their numerical results are not interpretable. In this paper, a semantic tolerance modeling scheme based on generalized intervals is proposed to improve the interpretability of tolerance modeling. Under certain optimality conditions, semantic tolerance models allow for true variation range estimations with simple computations. With the theoretical support of semantic tolerance modeling, a new dimension and tolerance specification scheme for semantic tolerancing is also proposed to better capture design intents and manufacturing implications, including flexible material selection, rigidity of specifications and constraints, component sorting in selective assembly, and assembly sequences.

1. Introduction

Tolerance modeling forms an important link between design and manufacturing. A significant amount of research has been carried out to explore the mathematical basis for dimensional and geometric tolerance representation, analysis, and synthesis. Relations among tolerances in components and assemblies are formulated in different ways and solved numerically. The typical analysis methods include variational estimation, kinematic formulation, statistical approximation, and Monte Carlo simulation. However, current tolerance modeling methods do not capture the semantics of tolerance specifications well.

First, traditional tolerance analysis methods assume all objects have rigid geometry. The variance is increasingly stacked up as components are assembled. The geometric variation of assembly is always assumed to be larger than those of its subassemblies and components. This rigid body tolerance analysis overlooks the role of flexible materials in assemblies, such as sheet metal and plastic components, which are common in aerospace, automobile, and electronics industries. For example, an airplane skin can be slightly warped, and yet riveted in place. Similarly, subassembly components of auto body with variations much larger than the specified ones can still meet the final assembly specification. The conventional addition theorem of tolerances is no longer valid in these applications. Given the specification of an assembly, unreasonably tight tolerance requirements may be assigned to subassemblies and components during tolerance synthesis if conventional methods are used. These methods treat tolerances for rigid and compliant assemblies with the same scheme of +/- range. The difference between rigid and flexible materials in assemblies is not captured.

Second, current modeling and analysis methods do not maintain the semantics of tolerance specifications during model formulation and numerical computing. These specifications and relationships among them imply manufacturing and assembly methods such as the sequence of fabrication. Tolerance analysis is usually simplified to the computation of numerical intervals. However, logical dependency and algebraic relations among variations are left out in existing approaches. This leads to the problem that numerical solutions are not interpretable, i.e., it is not viable to interpret and understand the relation between output range estimations and input variations. Instead of focusing only on mathematical and numerical convenience, a good model of tolerances should convey the full semantics of size and geometric tolerances and support analysis and synthesis with a simple yet comprehensive structure.

Third, both completeness and soundness of range estimations should be emphasized in tolerance analysis. A complete solution includes all possible occurrences, which is to check if the range estimation includes all possible stack-up results. Conversely, a sound solution does not include impossible occurrences, which consists in checking if the interval overestimates the actual variation range. True variation range estimations are both complete and sound. For example, completeness is the focus of the worst-case methods. It is usually assumed that tolerance variables are independent of each other. Thus the estimations are conservative and not sound when dependency exists among variables, i.e., variables are positively or negatively correlated. On the other hand, numerical estimations of statistical moments or kinematic variations are usually based on linearization or higherorder Taylor approximation, which makes it difficult to verify the completeness and soundness of solutions.

In this paper, we propose a new scheme, Semantic Tolerance Modeling, to represent and analyze tolerances based on generalized intervals. Unlike traditional set-based intervals, such as the interval [1,2] which represents a set of real values between 1 and 2, generalized (or modal) intervals also allow the existence of the interval [2,1]. With this extension, logic quantifiers (\forall and \exists) can be integrated to provide the interpretation of intervals. With tolerances represented by generalized intervals in semantic tolerance models, tolerancing semantics such as flexible material selection and logical relations among variations can be integrated into numerical results. In addition, modal interval analysis based on generalized intervals provides better variation estimation than the traditional worst-case interval analysis. If several optimality principles are followed, we can formulate tolerance models that estimate true variation ranges with simple algebraic calculation.

Based on interpretability principles of semantic tolerance modeling, a new dimension and tolerance specification scheme for semantic tolerancing is also proposed. The main difference between the semantic tolerancing scheme and the commonly used tolerancing practice is that *a priori* and *a posteriori* tolerances are differentiated in the new method. In fact, whenever defining a relationship among tolerances, we have implicitly differentiated these two types of tolerances. For instance, a working dimension is a dimension that is functionally critical and therefore explicitly specified in the design and blueprint. On the other hand, a balance dimension is not explicitly specified and its nominal and tolerance values are calculated from working dimensions. Compared to working dimensions, which are hard requirements imposed a priori, balance dimensions are soft and derived a posteriori. In general, a priori tolerances are tolerances with predetermined variations. They have the semantics of uncontrollable, unchangeable, critical, hard-constrained, specified, etc. A posteriori tolerances are tolerances with derived variations. They have the semantics of controllable, adjustable, flexible, softconstrained, feedback, etc. It should be noted that the semantic categories of a priori and a posteriori tolerances depend on the context of discourse.

The remainder of the paper is organized as follows: Section 2 gives an overview of related work on tolerance modeling, interval analysis, and an introduction to generalized intervals. Section 3 introduces the interpretability of semantic tolerance modeling. Section 4 presents the concept of semantic tolerancing based on the interpretability principles. Section 5 describes the true range estimation based on the optimality principles.

2. Background

2.1 3D Tolerance Modeling

There is a considerable amount of literature on tolerance modeling [1, 2]. Here, we only give a brief overview of 3D geometric tolerance zone representation related to the tolerance semantics. In variational approaches, tolerance zones are established either in 3D Euclidean space or in configuration space, such as offsetting tolerance zone [3], plane boundary representation [4, 5], and simplex based representation [6, 7]. In statistical approaches [8], geometric and size tolerances are not modeled separately. Statistical moments are estimated with linear or nonlinear tolerance stack-up. While the root-sum-square method yields optimistic estimations, alternatives were proposed to perform adjustment and correction for shifts and drifts [9]. Tolerance zone is also represented in mean-variance $(\mu - \sigma^2)$ space for analysis [10] and synthesis [11]. In kinematic approaches, geometrical variation and displacement are modeled by unified vectors and matrices [12, 13, 14], kinematic links in Euclidean space [15, 16] and configuration space [17]. The kinematic methods distinguish size tolerances and each type of geometric tolerances. However, relations among variations are not modeled, and estimation results are hard to interpret. In Monte Carlo simulation approaches [18, 19], large numbers of samples are

randomly generated and evaluated in statistical estimation. The drawback is that the computational cost for the sampling process is very high if an accurate estimation is required. The process also depends on the pre-assumption of certain statistical distributions for input random variates.

The modeling and analysis methods mentioned above have been widely accepted and used in commercial software such as Vis VSA[®] and CE/Tol[®]. However, it is not easy to interpret the meanings of the estimated variations and relations among them in components and assemblies. Furthermore, tolerances of compliant assemblies tend to be overestimated with the rigid-body assumption.

2.2 Tolerance Analysis for Flexible Assembly

There is a relatively small amount of research on tolerance analysis for flexible materials. A combination of finite element structural analysis and Monte Carlo simulation was proposed to predict variations in sheet metal joining [20, 21]. Geometric and material covariance in compliant assemblies is modeled in finite element simulation [22, 23]. Process-oriented tolerancing for multi-station assembly has also been studied [24, 25].

The previously mentioned finite element approaches have been integrated into commercial software packages such as Vis VSA[®] and CATIA-TAA[®]. Nevertheless, the tradeoff between fidelity and performance is always associated with finite element methods. Accurate computations become expensive if the variance estimation involves complex assemblies. In most cases, the accurate calculation of structural deformation and stress distribution is not the main purpose of tolerance analysis. It is more important to analyze producibility and associated costs with a reasonable amount of computation.

2.3 Interval Analysis

Interval mathematics [26] is a generalization in which interval numbers replace real numbers, interval arithmetic replaces real arithmetic, and interval analysis replaces real analysis. Intervals inherently represent uncertainties and errors in technical constructions, measuring, computations, and ranges of fluctuation and variation. In engineering fields, interval analysis has been applied in computer graphics [27, 28, 29], robust geometry construction and evaluation [30, 31, 32, 33], set-based modeling [34], imprecise structural analysis [35], design optimization [36], finite-element formulation and analysis [37, 38], soft constraint solving [39, 40], and worst-case tolerance analysis and synthesis [41, 42].

Interval analysis captures intrinsic uncertainty and variance. However, it is based on a worst-case scenario as in traditional linear stack-up methods. The computational results usually are pessimistic in this variance addition scheme if dependencies exist among variables. In contrast, modal interval analysis based on generalized intervals is an extension of the traditional interval analysis, which differentiates semantics of interval specification in different application situations.

2.4 Modal Interval Analysis

Modal interval analysis (MIA) [43, 44, 45, 46, 47] is a logical and semantic extension of interval analysis. Unlike classical interval analysis which identifies an interval by a set of real numbers, MIA identifies the intervals by the set of predicates which is fulfilled by the real numbers. In MIA, a generalized interval is not restricted to ordered bounds. Operations are defined

in Kaucher arithmetic [48].

A modal interval or generalized interval $\mathbf{x} := [\underline{x}, \overline{x}] \in KR$ is called *proper* when $\underline{x} \le \overline{x}$ and *improper* when $\underline{x} \ge \overline{x}$. The set of proper intervals is denoted by $IR = \{[\underline{x}, \overline{x}] | \underline{x} \le \overline{x}\}$, and the set of improper interval is $\overline{IR} = \{[\underline{x}, \overline{x}] | \underline{x} \ge \overline{x}\}$.

Given a generalized interval $\mathbf{x} = [\underline{x}, \overline{x}] \in K\mathbb{R}$, two operators *pro* and *imp* return proper and improper values respectively, defined as

$$\operatorname{pro} \mathbf{x} := [\min(\underline{x}, \overline{x}), \max(\underline{x}, \overline{x})] \tag{1}$$

$$\operatorname{imp} \mathbf{x} := [\max(x, \overline{x}), \min(x, \overline{x})]$$
(2)

The relationship between proper and improper intervals is established with the operator *dual*:

$$\operatorname{dual} \mathbf{x} := [x, \underline{x}] \tag{3}$$

For example, $\mathbf{x} = [-1,1]$ and $\mathbf{y} = [1,-1]$ are both valid intervals. While **x** is a proper interval, **y** is an improper one. The relation between **x** and **y** can be established by $\mathbf{x} = \text{dual } \mathbf{y}$. The inclusion relation between generalized intervals is defined as $[\underline{x}, \overline{x}] \subseteq [\underline{y}, \overline{y}] \Leftrightarrow \underline{x} \ge \underline{y} \land \overline{x} \le \overline{y}$. The less than or equal to relation is defined as $[x, \overline{x}] \le [y, \overline{y}] \Leftrightarrow \underline{x} \le y \land \overline{x} \le \overline{y}$.

Given a set of closed intervals of real numbers in R, and the set of logical existential (\exists) and universal (\forall) quantifiers, each generalized interval has an associated quantifier. The semantics of $\mathbf{x} \in KR$ is denoted by $(Q_x x \in \text{pro } \mathbf{x})$ where $Q_x \in \{\exists, \forall\}$. An interval $\mathbf{x} \in KR$ is called *existential* if $Q_x = \exists$. Otherwise, it is called *universal* if $Q_x = \forall$. Similar to the way that real numbers are associated in pairs with the same absolute value but opposite + and - signs, generalized intervals are also associated in pairs. Each member of a pair corresponds to the same closed interval of real numbers, but has opposite existential or universal modalities.

Based on generalized intervals, we propose a semantic tolerance modeling scheme in which the implications of tolerance stacking can be embedded in tolerance models. Compared to traditional worst-case interval methods, MIA enables accurate range estimation with simple computations. The purpose of semantic tolerance modeling is to capture logical relationships and engineering implications with mathematical representation, which is to build a bridge between mathematical theory and engineering practice. Semantic tolerance modeling possesses important characteristics: (1) Interpretability: being able to interpret tolerance intervals during analysis and synthesis processes while providing the basic understanding of tolerancing semantics; and (2) Optimality: being able to analyze tolerance propagation and accumulation so that tolerances can be specified without invalidating the basic requirement of completeness and soundness. Interpretability allows tolerance semantics to be embedded in interval results. Optimality assures the tightness of variation estimation.

3. Interpretability

If a real relation $z = f(x_1, \dots, x_n)$ is extended to the interval relation $\mathbf{z} = \mathbf{f}(\mathbf{x}_1, \dots, \mathbf{x}_n)$, the interval relation \mathbf{z} is interpretable if

there is a semantic relation

$$\frac{\left(\mathbf{Q}_{1} x_{1} \in \operatorname{pro} \mathbf{x}_{1}\right) \dots \left(\mathbf{Q}_{n} x_{n} \in \operatorname{pro} \mathbf{x}_{n}\right) \left(\mathbf{Q}_{z} z \in \operatorname{pro} \mathbf{z}\right)}{\left(z = f\left(x_{1}, \dots, x_{n}\right)\right)}$$

$$(4)$$

As the basis of interpretation, two interval extensions of a real function $f(x): \mathbb{R}^n \to \mathbb{R}$, so-called semantic interval functions, are defined in a min-max form as

$$f^{*}(\mathbf{x}) \coloneqq [\min_{x_{p} \in \text{pro}\,\mathbf{x}_{p}} \max_{x_{i} \in \text{pro}\,\mathbf{x}_{i}} f(x_{p}, x_{i}), \max_{x_{p} \in \text{pro}\,\mathbf{x}_{p}} \min_{x_{i} \in \text{pro}\,\mathbf{x}_{i}} f(x_{p}, x_{i})]$$
(5)

$$f^{**}(\mathbf{x}) \coloneqq [\max_{x_i \in \text{prox}_i} \min_{x_p \in \text{prox}_p} f(x_p, x_i), \min_{x_i \in \text{prox}_i} \max_{x_p \in \text{prox}_p} f(x_p, x_i)]$$
(6)

where (x_p, x_i) is the component splitting corresponding to interval vector $\mathbf{x} = (\mathbf{x}_p, \mathbf{x}_i)$, with sub-vectors \mathbf{x}_p and \mathbf{x}_i containing proper and improper components respectively. Important properties of interpretability are available and proved based on these two semantic interval functions.

Theorem 3.1 [43] Given a continuous function $f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ and a generalized interval vector $\mathbf{x} \in \mathbb{KR}^n$, if there exists an interval $\mathbf{f}(\mathbf{x}) \in \mathbb{KR}$, then

$$f^{*}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{x}) \Leftrightarrow \left(\forall x_{p} \in \operatorname{pro} \mathbf{x}_{p}\right) \left(\operatorname{Q}_{\mathbf{f}} z \in \operatorname{pro} \mathbf{f}(\mathbf{x})\right) \left(\exists x_{i} \in \operatorname{pro} \mathbf{x}_{i}\right) \left(z = f(x_{p}, x_{i})\right)$$
(7)

Theorem 3.2 [43] Given a continuous function $f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ and a generalized interval vector $\mathbf{x} \in \mathbb{KR}^n$, if there exists an interval $\mathbf{f}(\mathbf{x}) \in \mathbb{KR}$, then

$$f^{**}(\mathbf{x}) \supseteq \mathbf{f}(\mathbf{x}) \Leftrightarrow (\forall x_i \in \operatorname{pro} \mathbf{x}_i) (\mathcal{Q}_{\operatorname{dual} \mathbf{f}} \ z \in \operatorname{pro} \mathbf{f}(\mathbf{x})) (\exists x_p \in \operatorname{pro} \mathbf{x}_p) (z = f(x_p, x_i))$$
(8)

Let $f(\mathbf{x}): \mathbb{R}^n \to \mathbb{R}$ be a rational continuous function. Its modal rational extension $\mathbf{f}: \mathbb{KR}^n \to \mathbb{KR}$ replaces the real variables of f with generalized interval variables and real operators with interval operators. The semantics of a modal interval relation or function is embodied in the relation's syntax. The syntax of a function $f(x_1,...,x_n):\mathbb{R}^n \to \mathbb{R}$ can be represented by a syntax tree. For example, the syntax tree of $f_1 = |x_1 + x_2|/(x_1 - x_2\sqrt{x_3})$ is shown in Figure 1. A component x_i is considered uni-incident in the function $f(x_1,...,x_n)$ if it occupies only one leaf of the syntax tree, such as x_3 in f_1 . Otherwise, it is multi-incident, such as x_1 and x_2 in f_1 . Leaves and branches of the syntax tree are connected with either onevariable operators such as $| \cdot |$ and $\sqrt{-}$, or two-variable operators such as $+,-,\times,/$.

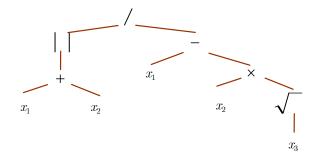


Figure 1. The syntax tree of $f_1 = |x_1 + x_2|/(x_1 - x_2\sqrt{x_3})$

3.1 Uni-incident interpretation

Theorem 3.3 [43] For a modal rational function $\mathbf{f}(\mathbf{x}) : \mathrm{KR}^n \to \mathrm{KR}$, if all arguments of $\mathbf{f}(\mathbf{x})$ are uni-incident, then

$$f^{*}(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{x}) \subseteq f^{**}(\mathbf{x}) \tag{9}$$

From Theorems 3.1, 3.2, and 3.3, we know modal rational functions of uni-incident variables are interpretable. For example, f(x, y) = x + y is considered for $x \in [1,3]$ and $y \in [2,5]$.

f([1,3],[2,5]) = [1,3] + [2,5] = [3,8], f([1,3],[5,2]) = [1,3] + [5,2] = [6,5], f([3,1],[2,5]) = [3,1] + [2,5] = [5,6],f([3,1],[5,2]) = [3,1] + [5,2] = [8,3],

have the meanings of

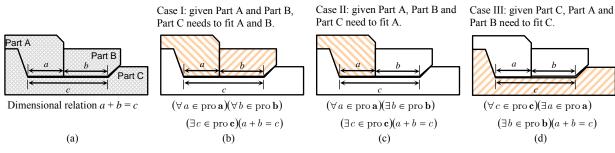
 $(\forall x \in [1,3])(\forall y \in [2,5])(\exists z \in [3,8])(z = x + y),$ $(\forall x \in [1,3])(\forall z \in [5,6])(\exists y \in [2,5])(z = x + y),$ $(\forall y \in [2,5])(\exists x \in [1,3])(\exists z \in [5,6])(z = x + y),$ $(\forall z \in [3,8])(\exists x \in [1,3])(\exists y \in [2,5])(z = x + y),$

respectively. Similarly, the modal natural extension of function $f(x) = (x_1 + x_2)(x_3 + x_4)$ with the generalized interval $\mathbf{x}_1 = [-2, 2]$, $\mathbf{x}_2 = [1, -1]$, $\mathbf{x}_3 = [-1, 1]$, and $\mathbf{x}_4 = [2, -2]$ is $\mathbf{f}(\mathbf{x}) = (\mathbf{x}_1 + \mathbf{x}_2)(\mathbf{x}_3 + \mathbf{x}_4) = [0, 0]$. It is interpreted as

$$(\forall x_1 \in [-2,2]) (\forall x_3 \in [-1,1]) (\exists x_2 \in [-1,1]) (\exists x_4 \in [-2,2]) ((x_1 + x_2)(x_3 + x_4) = 0)$$

Different semantics of tolerance stack-up in assembly enclosure need to be differentiated in tolerance design. This includes the semantics associated with assembly sequence, accuracy of tolerance estimation, and controllability of variation. As tolerances are stacked up in a tolerance chain, a direct correlation exists between the time at which the part is assembled and the degree to which the corresponding variations are controllable in order to close the chain. The earlier a part is assembled in the sequential process, the less controllable the corresponding variations are in order to close the chain. In this sense, tolerances of earlier assembled parts are out of the current worker's control. They are *uncontrollable* tolerances. In contrast, the most recently assembled ones have *controllable* tolerances.

Based on manufacturing and assembly sequences, tolerances may be specified in different ways to designate desirable semantics. For example, in Figure 2, dimensions a, b, and c in three components have relation a+b=c. If Part A and B are provided by suppliers and Part C is to be built in house (Figure 2b, Case I), or if a and b are working dimensions and c is a balance dimension, the tolerance of c is determined by the tolerances of aand b and the tolerance chain should be closed. In this case, the semantics of "given A and B, C needs to fit A and B" is expressed $(\forall a \in \text{pro} \mathbf{a})(\forall b \in \text{pro} \mathbf{b})(\exists c \in \text{pro} \mathbf{c})(a+b=c)$, which is as different from the semantics of "given A, B and C need to fit A" when a is a working dimension while b and c are balance dimensions (Figure 2-c, Case II). These algebraic relations among tolerances should be compatible with the semantics of engineering specifications. In a semantic tolerance model, a priori and a posteriori tolerances are differentiated by the modalities of intervals. With the modal extension, semantics of specification sequence and rational can be embedded in algebraic relations of the model.





With the symbolic differentiation of a priori and a posteriori tolerances, different strategies of tolerance allocation could be applied in different scenarios. For example, in Figure 2-b, given two uncontrollable dimensions a and b, the controllable dimension c = a + b = [2,5] + [1,3] = [3,8]. In Figure 2-c, one extra controllable dimension b allows a tighter tolerance of c. c = a + b = [2,5] + [3,1] = [5,6]. The tolerance range of c is reduced from 5 to 1, which is smaller than the tolerance range of a. This indicates that the principle of selective assembly may be applied to achieve assembly. Selective assembly is a widely used process of sorting and selecting mating components in pairs so that high-precision assemblies can be achieved even with lowprecision components. This method is valuable when individual components cannot be produced with small enough tolerances to be fully interchangeable in assembly such as specialized roller bearings with micrometer level tolerances. However, selective assembly is a manual process, which signifies it may only be used in low-volume high-value products. In a cost-conscious mass production environment, choosing flexible materials is the alternative, as discussed in Section 3.3.

3.2 Multi-incident interpretation

Theorem 3.4 [43] For a modal rational function $\mathbf{f}(\mathbf{x}): \mathrm{KR}^n \to \mathrm{KR}$, if there are multi-incident improper arguments in $\mathbf{f}(\mathbf{x})$ and $\mathbf{t}^*(\mathbf{x})$ is obtained from \mathbf{X} , by transforming, for every multi-incident improper component, all incidences but one into its dual, then

$$f^*(\mathbf{x}) \subseteq \mathbf{f}(\mathbf{t}^*(\mathbf{x})) \tag{10}$$

Theorem 3.5 [43] For a modal rational function $\mathbf{f}(\mathbf{x}) : \mathrm{KR}^n \to \mathrm{KR}$, if there are multi-incident proper arguments in $\mathbf{f}(\mathbf{x})$ and $\mathbf{t}^{**}(\mathbf{x})$ is obtained from \mathbf{x} , by transforming, for every multi-incident proper component, all incidences but one into its dual, then

$$f^{**}(\mathbf{x}) \supseteq \mathbf{f}(\mathbf{t}^{**}(\mathbf{x})) \tag{11}$$

From Theorems 3.1, 3.2, 3.4, and 3.5, modal rational functions of multi-incident variables are interpretable with some modifications. For example, f(x, y) = xy/(x + y) is extended to $\mathbf{x} = [-1,3]$ and $\mathbf{y} = [15,7]$.

 $\mathbf{f}(\mathbf{x}, \mathbf{y}) = [-1,3] \times [15,7]/([-1,3] + [15,7]) = [-0.5,1.5]$ is not interpretable, whereas

$$\begin{split} \mathbf{f}(\mathbf{t}^*(\mathbf{x},\mathbf{y})) = [-1,3] \times [15,7]/([-1,3]+[7,15]) = [-1.1667,3.5], \\ \mathbf{f}(\mathbf{t}^*(\mathbf{x},\mathbf{y})) = [-1,3] \times [7,15]/([-1,3]+[15,7]) = [-1.0715,3.2143], \\ \mathbf{f}(\mathbf{t}^{**}(\mathbf{x},\mathbf{y})) = [-1,3] \times [15,7]/([3,-1]+[15,7]) = [-0.3889,1.1667], \end{split}$$

 $f(t^{**}(x, y)) = [3, -1] \times [15, 7]/([-1, 3] + [15, 7]) = [4.5, -1.5]$ are interpretable. They are interpreted as

 $(\forall x \in [-1,3]) (\exists y \in [7,15]) (\exists z \in [-1.1667,3.5]) (z = xy/(x+y)),$ $(\forall x \in [-1,3]) (\exists y \in [7,15]) (\exists z \in [-1.0715,3.2143]) (z = xy/(x+y))$

$$(\forall y \in [7,15]) (\forall z \in [-0.3889, 1.1667]) (\exists x \in [-1,3]) (z = xy/(x+y)), (\forall y \in [7,15]) (\exists x \in [-1,3]) (\exists z \in [-1.5,4.5]) (z = xy/(x+y))$$

respectively.

In assemblies, parametric relations with multi-incident variables are common. Compared to traditional tolerance modeling, semantic tolerance modeling allows us to explicitly interpret algebraic relations with the interpretability properties of modal intervals. Different numerical values and modalities can also be selected in order to derive specific semantics.

3.3 Rigidity interpretation

In the material property domain, the tolerance ranges for rigid materials correspond to proper intervals and those for flexible materials correspond to improper intervals.

In the one-way clutch example of Figure 3, the distance vector b, the length of the spring s, and the radius of the ball r satisfy the relation r + s = b. If ranges [5.2,5.7] and [7.8,8.0] are given to r and b respectively, the range for spring length s can be [2.1,2.8], as in relation

$$\mathbf{r} + \mathbf{s} = [5.2, 5.7] + [2.8, 2.1] = [8.0, 7.8] = \mathbf{b}$$

It is interpreted as

$$(\forall r \in [5.2, 5.7])(\forall b \in [7.8, 8.0])(\exists s \in [2.1, 2.8])(r + s = b)$$

The spring provides a "cushion" to absorb variance. If a larger range [7.8, 8.5] is allowed for b, no flexible material is absolutely required to absorb variance. Rigid material instead of spring for s can be chosen, as in relation

 $\mathbf{r} + \mathbf{s} = [5.2, 5.7] + [2.6, 2.8] = [7.8, 8.5] = \mathbf{b}$

It is interpreted as

$$(\forall r \in [5.2, 5.7])(\forall s \in [2.6, 2.8])(\exists b \in [7.8, 8.5])(r + s = b)$$

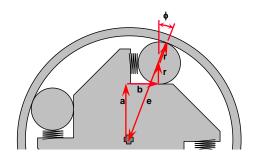
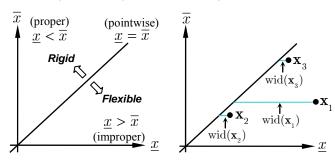


Figure 3. variations of size and geometry, shape deformation, and kinematics form a closed loop in assembly

As illustrated in Figure 4-a, the semantic difference between rigid and flexible materials is noted by interval modalities. If a function width of interval $\mathbf{x} = [\underline{x}, \overline{x}]$ is defined as wid $(\mathbf{x}) \coloneqq |\overline{x} - \underline{x}|$, the flexibility of materials is quantified by the width of improper intervals. The relative width of an improper interval indicates how flexible the material is. Compressibility may be indicated by the index $I^{-}(\mathbf{x}) = (\underline{x} - \overline{x})/\underline{x}$ and stretchability by $I^{+}(\mathbf{x}) = (\underline{x} - \overline{x})/\overline{x}$. For example, in Figure 4-b, material \mathbf{x}_{1} is more flexible than material \mathbf{x}_{2} , and \mathbf{x}_{2} is more flexible than \mathbf{x}_{3} . The rigidity diagram illustrates the relationship

between rigid and flexible materials. Selection of materials thus can be integrated into algebraic relations of generalized intervals.



(a) the improper interval domain corresponds to flexible materials

(b) the width of an improper interval exp indicates how flexible the material is

Figure 4. An inf-sup diagram is also a rigidity diagram

4. Semantic Tolerancing

Capturing semantics associated with design intents in engineering drawings is the main purpose of Geometric Dimensioning and Tolerancing (GD&T). Yet the current GD&T has some weakness such as not specifying the practice of measurement and inspection and lack of process semantics [49]. Semantic tolerancing with generalized intervals is a new dimension and tolerance specification scheme based on semantic tolerance modeling. With the differentiation of existential and universal modalities associated with ranges, design intents and manufacturing implications such as selection of flexible materials, rigidity of specifications and constraints, and sorting and sequencing of assembly can be captured. The major step of the proposed tolerancing practice is to differentiate a posteriori tolerances from a priori tolerances with symbols. Tolerances with universal modality are a priori tolerances, while those with existential modality are a posteriori tolerances. We use a minus*plus* notation $x \neq \Delta$ in combination with the traditional *plusminus* notation $x \pm \Delta$ to represent two modalities.

The modality of a tolerance is determined by the following rules. (1) If a closed tolerance chain $\sum_i d_i = z$ is formed, the dimensions on the left-hand side with the notation of $d_i \pm \Delta_i$ are a priori tolerances. Those with the notation of $d_i \mp \Delta_i$ on the left-hand side are a posteriori. However, on the right-hand side

s. Selection of materials thus ions of generalized intervals. \overline{x} \overline{x} \overline{x}

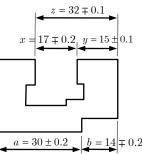


Figure 5. A priori and a posteriori tolerances in semantic tolerancing

Assembly sequence can be inferred from the semantic tolerance chain stack-up. As illustrated in Figure 6-a, four dimensions a, b, g, and z are specified within a closed chain a + g + b = [8.8, 9.2] + [2.8, 3.2] + [5.2, 4.8] = [16.8, 17.2] = z. $a = 9 \pm 0.2$ and $b = 5 \mp 0.2$ imply that subassembly B is assembled after subassembly A. If the functional requirement of working dimension g is not met, B needs to be adjusted. However, if the specifications are $a = 9 \mp 0.2$ and $b = 5 \pm 0.2$ as in Figure 6-b, A needs to be adjusted to meet the requirement of g. In Figure 6-c, $g = 3 \mp 0.2$ indicates that g is no longer functionally critical while a and b are.

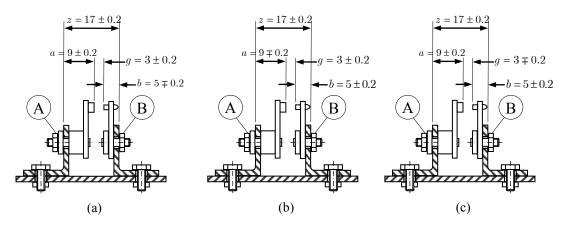


Figure 6. Semantic tolerancing implies assembly sequence

In semantic tolerancing, material selection and assembly methods can also be explicitly specified. Figure 7 illustrates flexible assembly and selective assembly examples of the Case III in Figure 2-d. The size tolerances of Part A and Part B are a posteriori. Both are larger than the size tolerance of Part C. Yet, three parts need to be assembled. The symbols of the tolerance specification in Figure 7-a indicate that flexible materials with the compressibility index in the range of $0.6/10.3 \approx 0.0583$ need to

be chosen for Parts A and B. If variation ranges of a and b are reduced to ± 0.03 and selective assembly process is intended to be used, the a posteriori tolerance symbol captures the intent that A and B need to be sorted and paired, as in Figure 7-b.

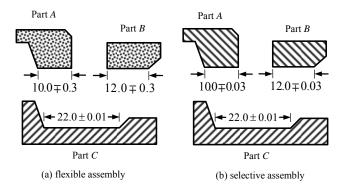


Figure 7. Semantic tolerancing captures intent of material selection and selective assembly

5. Completeness and Soundness

For $\mathbf{x} \in K\mathbb{R}^n$, if the modal rational extension $\mathbf{f}(\mathbf{x}): K\mathbb{R}^n \to K\mathbb{R}$ satisfies

$$f^*(\mathbf{x}) = \mathbf{f}(\mathbf{x}) = f^{**}(\mathbf{x}) ,$$

 $\mathbf{f}(\mathbf{x})$ is called *optimal*. In other words, if the evaluation of a modal rational function $\mathbf{f}(\mathbf{x})$ is both complete and sound, $\mathbf{f}(\mathbf{x})$ is

optimal. Optimal functions give tight bounds of complete estimation. The optimal modal interval extension unifies computable $f(\mathbf{x})$ and interpretable $f^*(\mathbf{x})$ and $f^{**}(\mathbf{x})$.

A real function $f(x, y_1, ..., y_m) : \mathbb{R}^{m+1} \to \mathbb{R}$ is called uniformly monotonic for x in an interval domain $(\mathbf{x}, \mathbf{y}_1, ..., \mathbf{y}_m)$, if f is monotonic for $\forall x \in \text{pro} \mathbf{x}$ and it keeps the same monotonicity for all the values of $y_i \in \text{pro} \mathbf{y}_i$. If each incidence of x in f is considered as an independent variable and each one of the incidences is uniformly monotonic, then f is called *totally* monotonic for x.

In the syntax tree of a modal rational function f(x), an operator is called *node-optimal* if its operands are all uniformly monotonic in their evaluated interval domains. f(x) is called *treeoptimal* in an interval domain \mathbf{X} , if the children of those operators that are not node-optimal in the syntax tree are either leaves or one-variable operators. For instance, the function $\mathbf{f}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = \mathbf{x}_1\mathbf{x}_2 + \mathbf{x}_3\mathbf{x}_4$ is tree optimal in any interval domain. Because $\mathbf{y}_1 + \mathbf{y}_2$ is node-optimal for any values of \mathbf{y}_1 and \mathbf{y}_{2} with the partial derivative of constant 1. Here, $\mathbf{y}_{1} = \mathbf{x}_{1}\mathbf{x}_{2}$ and $\mathbf{y}_2 = \mathbf{x}_3 \mathbf{x}_4$. Although $\mathbf{x}_1 \mathbf{x}_2$ or $\mathbf{x}_3 \mathbf{x}_4$ is not node-optimal if $0 \in \operatorname{pro} \mathbf{x}_1 \lor 0 \in \operatorname{pro} \mathbf{x}_2$ or $0 \in \operatorname{pro} \mathbf{x}_3 \lor 0 \in \operatorname{pro} \mathbf{x}_4$, the children of the multiplication operator are leaves. for $\mathbf{g}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = (\mathbf{x}_1 + \mathbf{x}_2)(\mathbf{x}_3 + \mathbf{x}_4)$ is tree optimal ([1,2],[0,2],[4,3],[3,2]), since $y_1 \cdot y_2$ is node-optimal for ([1,4],[7,5]); $\mathbf{x}_1 + \mathbf{x}_2$ and $\mathbf{x}_3 + \mathbf{x}_4$ are node-optimal. But $g(x_1, x_2, x_3, x_4)$ is not tree-optimal for ([-2,2],[1,-1],[-1,1],[2,-2]).

5.1 Uni-incident optimality

Theorem 4.1 [43] If $\mathbf{f}(\mathbf{x})$ is tree-optimal in a domain $\mathbf{x} \in K\mathbb{R}^n$ and all arguments of $\mathbf{f}(\mathbf{x})$ are uni-incident,

$$f^*(\mathbf{x}) = \mathbf{f}(\mathbf{x}) = f^{**}(\mathbf{x})$$

For example, $f(x, y) = (x + y)^2$ is optimal for $\mathbf{x} = [1,3]$ and $\mathbf{y} = [2,5]$. The true range of the function $R_f = [9,64]$. The natural extension is $\mathbf{f}([1,3],[2,5]) = ([1,3] + [2,5])^2 = [9,64]$. Similarly, $\mathbf{f}([3,1],[5,2]) = [64,9]$ is optimal.

5.2 Multi-incident optimality

Theorem 4.2 [43] If $\mathbf{f}(\mathbf{x})$ is tree-optimal in a domain $\mathbf{x} \in K\mathbb{R}^n$ and totally monotonous for all of its multi-incident arguments, and \mathbf{x}^p is obtained from \mathbf{x} , by transforming, for every multi-incident component, all incidences into its dual if the corresponding incidence has a monotonicity sense contrary to the global one, then

$$f^*(\mathbf{x}) = \mathbf{f}(\mathbf{x}^D) = f^{**}(\mathbf{x}) .$$

For example, f(x, y) = xy/(x + y) is extended to $\mathbf{x} = [1,3]$ and $\mathbf{y} = [15,7]$. The partial derivatives of f with respect to xand y are all positive within the domain. The partial derivatives of f with respect to the first incidences of x and y are

positive, and negative with respect to the second incidences of x and y . Therefore,

 $\mathbf{f}(\mathbf{x}^{D}) = [1,3] \times [15,7]/([3,1] + [7,15]) = [0.9375,2.1]$

is optimal.

When the syntax structure of a modal interval function is optimal within a given interval domain, true range estimation can be obtained. However, if the structure is not optimal, true range estimation is not guaranteed with the direct algebraic calculation. Theorems of optimality have been proved. Interested readers are referred to [43] for details. Following the above optimality principles, we can construct tolerance models that estimate true variation ranges with simple algebraic evaluation.

5.3 Example: true range estimation of one-way clutch

To illustrate the optimality of generalized intervals in range estimation, a comparison of the MIA method and the Direct Linearization Method (DLM) [50] (as implemented in CE/Tol[®] package) for the one-way clutch example is made, as shown in Figure 8. In this example, the true variation range can be derived analytically. The combination of the smallest roller (\underline{r}) and

largest gap ($\underline{a} \; \mbox{ and } \overline{e}$) gives the upper bound of displacement b .

Conversely, we can derive the lower bound of displacement \underline{b} .

The MIA evaluation is based on the optimality analysis. It is not difficult to verify that the modal rational extension function $\mathbf{b}(\mathbf{a},\mathbf{e},\mathbf{r}) = \sqrt{(\mathbf{e}-\mathbf{r})^2 - (\mathbf{a}+\mathbf{r})^2}$ is tree-optimal within the given tolerance ranges of \mathbf{a} , \mathbf{e} , and \mathbf{r} listed in Table 1. The function is totally monotonous for r. $\partial b/\partial r < 0$ for r and for both of its incidences if regarded as independent individual variables. Thus the rational extension function is optimal. Compared to the methods of DLM with Root-Sum-Square (RSS) and Worst-Case (WC), the MIA evaluation result [4.0838, 5.4405] based on the modal rational extension function $\mathbf{b}(\mathbf{a}, \mathbf{e}, \mathbf{r})$ gives an accurate estimation of the true variation range.

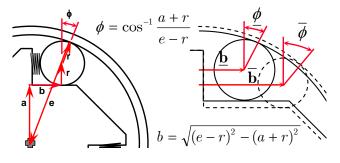


Figure 8. Modal intervals make complex algebraic relations with multi-incident variables interpretable. Interpretations are corresponding to different value sets

6. Concluding Remarks

A semantic tolerance modeling scheme based on generalized intervals is proposed to enrich tolerance modeling and analysis for interpretable and accurate variation estimations. Logical relationships among variations are embedded in the mathematical formulation. Semantic tolerance models capture more processoriented tolerancing semantics such as the difference between rigid and flexible materials in assemblies and component sorting in selective assembly and assembly sequence. The degeneracy of semantics during numerical computation is prevented. A new dimension and tolerance specification scheme for semantic tolerancing is proposed to symbolically differentiate a priori and a posteriori tolerances. Compared to traditional methods, semantic tolerance models with the optimal construction of relations estimate true variation ranges such that sound and complete solutions can be obtained.

The future work may include the study of linear and nonlinear tolerance stack-ups in flexible assemblies, where the optimal allocation of flexible materials helps to reduce the overall variations and costs. Interpretable linear and nonlinear tolerance analysis will also help to incorporate more process semantics in products' tolerance design.

Table 1. Result comparison between MIA and DLM method

Input			Output: position of roller (b) True Range is [4.0838,5.4405]		
Hub Height (a)	Ring Radius (e)	Roller Radius (r)	DLM with Root-Sum- Square (as in CE/Tol [®])	DLM with Worst- Case (as in CE/Tol [®])	MIA
[27.595, 27.695]	[50.7875, 50.8125]	[11.42, 11.44]	[4.3585, 5.2625]	[4.1368, 5.4842]	[4.0838, 5.4405]

7. Acknowledgements

The author appreciates the comments from anonymous reviewers for the improvement of the paper.

8. References

- Hong, Y.S. and Chang, T.-C., 2002, "A comprehensive review of tolerancing research," *Int. J. of Production Research*, 40(11), pp.2425-2459.
- [2] Zhang, H.C. eds., 1997, Advanced Tolerancing Techniques, John Wiley & Sons, New York.
- [3] Requicha, A.A.G., 1983, "Toward a theory of geometric tolerancing," *Int. J. of Robotics Research*, **2**(2), pp.45-60.
- [4] Roy, U. and Li, B., 1999, "Representation and interpretation of geometric tolerances for polyhedral objects – II Size, orientation and position tolerances," *Computer-Aided Design*, 31(4), pp.273-285.
- [5] Teck, T.B., Senthil Kumar, A., and Subramaian, V., 2001, "A CAD integrated analysis of flatness in a form tolerance zone," *Computer-Aided Design*, 33(11), pp.853-865.
- [6] Davidson, J.K., Mujezinovic, A., and Shah, J.J., 2002, "A new mathematical model for geometric tolerances as applied to round faces," *ASME J. of Mechanical Design*, 124(4), pp.609-622.
- [7] Mujezinovic, A., Davidson, J.K., and Shah, J.J., 2004, "A new mathematical model for geometric tolerances as applied to polygonal faces," *ASME J. of Mechanical Design*, **126**(3), pp.504-518.
- [8] Nigam, S.D. and Turner, J.U., 1995, "Review of statistical approaches to tolerance analysis," *Computer-Aided Design*, 27(1), pp.6-15.
- [9] Chase, K.W. and Greenwood, W.H., 1988, "Design issues in mechanical tolerance analysis," *Manufacturing Review*, 1(1), pp.50-59.
- [10] Srinivasan, V. and O'Connor, M.A., 1994, "On interpreting statistical tolerancing," *Manufacturing Review*, 7(4), pp.304-311.
- [11] Zhang, C., Luo, J., and Wang, B., 1999, "Statistical tolerance synthesis using distribution function zones," *Int. J.* of Production Research, **37**(17), pp.3995-4006.
- [12] Wirtz, A., Gachter, C., and Wipf, D., 1993, "From unambiguously defined geometry to the perfect quality control loop," *Annals of the CIRP*, **42**(1), pp.615-618.
- [13] Whitney, D.E., Gilbert, O.L., and Jastrzebski, M., 1994, "Representation of geometric variations using matrix transforms for statistical tolerance analysis in assemblies," *Research in Engineering Design*, 6, pp.191-210.
- [14] Desrochers, A. and Riviere, A., 1997, "A matrix approach to the representation of tolerance zones and clearances," *Int. J.* of Advanced Manufacturing Technology, **13**, pp.630-636.
- [15] Rivest, L., Fortin, C., and Morel, C., 1994, "Tolerancing a solid model with a kinematic formulation," *Computer-Aided Design*, 26(6), pp.465-476.
- [16] Chase, K.W., Magleby, S.P., Gao, J., and Sorensen, C.D., 1996, "Including geometric feature variations in tolerance

analysis of mechanical assemblies," *IIE Transactions*, **28**(10), pp.795-807.

- [17] Sacks, E. and Joskowicz, L., 1998, "Parametric kinematic tolerance analysis of general planar systems," *Computer-Aided Design*, **30**(9), pp.707-714.
- [18] Turner, J.U. and Wozny, M.J., 1987, "Tolerances in computer-aided geometric design," *The Visual Computer*, 3, pp.214-226.
- [19] Ashiagbor, A., Liu, H. C., Nnaji, B.O., 1998, "Tolerance control and propagation for the product assembly modeler," *Int. J. of Production Research*, 36(1), pp.75-94.
- [20] Liu, S. and Hu, S., 1997, "Variation simulation for compliant sheet metal assemblies using finite element methods," ASME J. of Manufacturing Science & Engineering, 119(3), pp.368-374.
- [21] Liu, S.C., Hu, S.J., and Woo, T.C., 1996, "Tolerance analysis for sheet metal assemblies," ASME J. of Mechanical Design, 118(1), pp.62-67.
- [22] Merkley, K.G., 1998, *Tolerance Analysis of Compliant Assemblies*, Ph.D. thesis, Brigham Young University.
- [23] Camelio, J., Hu, S.J., and Marin, S.P., 2004, "Compliant assembly variation analysis using component geometric covariance," ASME J. of Manufacturing Science & Engineering, 126(2), pp.355-360.
- [24] Camelio, J., Hu, S.J., and Ceglarek, D., 2003, "Modeling variation propagation of multi-station assembly systems with compliant parts," ASME J. of Mechanical Design, 125(4), pp.673-681.
- [25] Ding, Y., Jin, J., Ceglarek, D., and Shi, J., 2005, "Processoriented tolerancing for multi-station assembly systems," *IIE Transactions*, 37(6), pp.493-508.
- [26] Moore, R.E., 1966, *Interval Analysis*, Prentice-Hall, Englewood Cliffs, N.J.
- [27] Mudur, S.P. and Koparkar, P.A., 1984, "Interval methods for processing geometric objects," *IEEE Computer Graphics & Applications*, 4(2), pp.7-17.
- [28] Duff, T., 1992, "Interval arithmetic and recursive subdivision for implicit functions and Constructive Solid Geometry," *Computer Graphics*, 26(2), pp.131-138.
- [29] Snyder, J., 1999, Generative Modeling for Computer Graphics and CAD: Symbolic Shape Design Using Interval Analysis, Academic Press, Cambridge.
- [30] Sederberg, T.W. and Farouki, R.T., 1992, "Approximation by interval Bezier curves," *IEEE Computer Graphics & Applications*, **12**(5), pp.87-95.
- [31] Abrams, S.L., Cho, W., Hu, C.Y., Maekawa, T., Patrikalakis, N.M., Sherbrooke, E.C., and Ye, X., 1998, "Efficient and reliable methods for rounded-interval arithmetic," *Computer-Aided Design*, **30**(8), pp.657-665.
- [32] Tuohy, S.T., Maekawa, T., Shen G., and Patrikalakis, N.M., 1997, "Approximation of measured data with interval B-Splines," *Computer-Aided Design*, 29(11), pp.791-799.
- [33] Wallner, J., Krasauskas, R., and Pottmann, H., 2000, "Error propagation in geometric constructions," *Computer-Aided Design*, **32**(11), pp.631-641.
- [34] Finch, W.W. and Ward, A.C., 1997, "A set-based system for eliminating infeasible designs in engineering problems dominated by uncertainty," in *Proc. 1997 ASME Design*

Engineering Technical Conference, No. DETC97/DTM-3886.

- [35] Rao, S.S. and Berke, L., 1997, "Analysis of uncertain structural systems using interval analysis," *AIAA Journal*, 35(4), pp.727-735.
- [36] Rao, S.S. and Cao, L., 2002, "Optimum design of mechanical systems involving interval parameters," ASME J. of Mechanical Design, 124(3), pp.465-472.
- [37] Muhanna, R.L. and Mullen, R.L., 1999, "Formulation of fuzzy finite-element methods for solid mechanics problems," *Computer-Aided Civil & Infrastructure Engineering*, 14(2), pp.107-117.
- [38] Muhanna, R.L., Zhang, H., and Mullen, R.L., 2007, "Interval finite element as a basis for generalized models of uncertainty in engineering mechanics," *Reliable Computing*, 13(2) pp.173-194.
- [39] Joan-Arinyo, R., Mata, N., and Soto-Riera, A., 2001, "A constraint solving-based approach to analyze 2D geometric problems with interval parameters," *ASME J. of Computing & Information Science in Engineering*, 1(4), pp.341-346.
- [40] Wang, Y. and Nnaji, B.O., 2007, "Solving interval constraints by linearization in computer-aided design," *Reliable Computing*, 13(2), pp.211-244.
- [41] Yang, C.C., Marefat, M.M., and Ciarallo, F.W., 2000, "Interval constraint networks for tolerance analysis and synthesis," *Artificial Intelligence for Engineering Design*, *Analysis & Manufacturing*, 14(4), pp.271-287.
- [42] Wu, W. and Rao, S.S., 2004, "Interval approach for the modeling of tolerances and clearances in mechanism analysis," ASME J. of Mechanical Design, 126(4), pp.581-592.
- [43] Gardenes, E., Sainz, M.A., Jorba, L., Calm, R., Estela, R., Mielgo, H., and Trepat, A., 2001, "Modal intervals," *Reliable Computing*, 7(2), pp.77-111.
- [44] Markov, S., 2001, "On the algebraic properties of intervals and some applications," *Reliable Computing*, 7(2), pp.113-127.
- [45] Shary, S.P., 2002, "A new technique in systems analysis under interval uncertainty and ambiguity," *Reliable Computing*, 8(2), pp.321-418.
- [46] Popova, E.D., 2001, "Multiplication distributivity of proper and improper intervals," *Reliable Computing*, 7(2), pp.129-140.
- [47] Armengol, J., Vehi, J., Trave-Massuyes, L., and Sainz, M.A., 2001, "Application of modal intervals to the generation of error-bounded envelopes," *Reliable Computing*, 7(2), pp.171-185.
- [48] Kaucher, E., 1980, "Interval analysis in the extended interval space IR," *Computing Supplement*, 2, pp.33-49.
- [49] Hopp, T.H., 1993, "The language of tolerances," in *Quality Through Engineering Design*, W. Kuo, eds., Elsevier, Amsterdam, pp.317-332.
- [50] Chase, K.W., Magleby, S.P., and Gao, J., 1997, "Tolerance analysis of two- and three- dimensional mechanical assemblies with small kinematic adjustments," in *Advanced Tolerancing Techniques*, H.C. Zhang, eds., John Wiley & Sons, New York, pp.103-137.