Toward a New Mathematical Definition of Datums in Standards to Support Advanced Manufacturing

Conference Paper · June 2018

DOI: 10.1115/MSEC2018-6305

Vijay Srinivasan
National Institute of Standards and Technology

Proceedings of the ASME 2018 Manufacturing Science and Engineering Conference
MSEC2018
June 18-22, 2018, College Station, Texas, USA

MSEC2018-6305

Toward a New Mathematical Definition of Datums in Standards to Support Advanced Manufacturing

Craig M. Shakarji
Physical Measurement Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
craig.shakarji@nist.gov

Vijay Srinivasan
Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899, U.S.A.
vijay.srinivasan@nist.gov

Abstract

Recent advances in the digitization of manufacturing have prompted ASME and ISO standards committees to reexamine the definition of datums. Any new definition of datums considered by the standards committees should cover all datum feature types used in design, and support both traditional metrological methods and new digital measurement techniques. This is a challenging task that requires some careful compromise. This paper describes and analyzes various alternatives considered by the standards committees. Among them is a new mathematical definition of datums based on constrained least-squares fitting. It seems to provide the best compromise and has the potential to support advanced manufacturing that is increasingly dependent on digital technologies.

1. Introduction

Datums can be found everywhere in industry that is engaged in manufacturing. Standards define how datum features should be identified in design documents and used in manufacturing [1-7]. The design documents can be in the form of traditional two-dimensional (2D) projected views in drawings, or in the form of computer files that contain representations of three-dimensional (3D) models as presented in Fig. 1. During production, datum features are used to plan manufacturing processes; for example, which features should be machined first and within what tolerances, and how to position a workpiece in a machining fixture. During and after part production, datum features are used for inspection planning; for example, to make in-process measurements for quality control, and to check for conformance of a manufactured part to designer-specified tolerances [8].

Traditional inspection methods create datum feature simulators by employing physical devices such as surface plates, angle blocks, opening and closing vises, and expanding and contracting chucks. But this *physical* world is now being challenged by the emergence of the *digital* world. Recent advances in measurement science and technology have ushered in Coordinate Measurement Systems (CMS) and scanning devices that are Internet-enabled – as part of the current popular trend in Internet of Things (IoT) [9, 10]. These devices and systems can collect and communicate vast amounts of 3D coordinate data that can number in the millions and are sometimes referred to as 'point clouds,' and software has the potential to numerically simulate the datum features without resorting to the type of physical devices mentioned above.

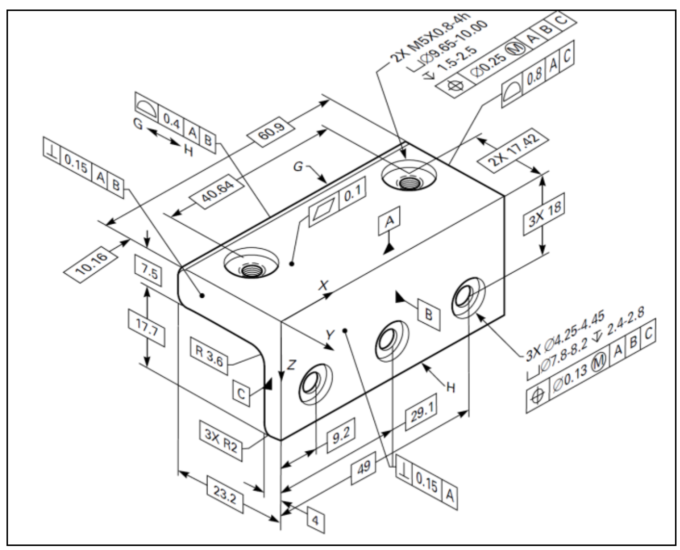


Figure 1. Datum features are indicated as A, B, and C within boxes in a 3D model [5].

Digitization of manufacturing is not confined to metrology. A recent industrial initiative called 'digital thread' envisions a seamless flow of 3D information in an enterprise all the way from design (as illustrated in Fig. 1) through production, and finally to retirement [11]. In this process, physical mock-ups are replaced by digital mock-ups – including all aspects related to datums – throughout a product's lifecycle in an engineering enterprise. Such an enterprise is also called a model-based enterprise [12] to emphasize the role of informational, computational, and mathematical models in digitization of manufacturing.

Digitization of manufacturing has also sparked an international re-thinking of manufacturing at large. This new way of thinking has given rise to terms and initiatives such as Smart Manufacturing, Cyber-manufacturing, Cyber-Physical Production Systems, and Industrie 4.0 [13, 14]. The rapid rise and adoption of these concepts have prompted standards development organizations such as ASME and ISO (International Organization for Standardization) to reexamine the definitions of geometrical product specifications and verifications. If these organizations don't step up to standardizing new definitions that can support rapid advances in digitization of manufacturing, some *ad hoc* practices and definitions will emerge, and chaos will ensue.

It is in this context that ASME and ISO are considering a new definition of datums and related concepts that will utilize digital technologies and support advanced manufacturing. But any new definition that will emerge from their standards committees should also support traditional datum establishment practices that use physical devices, such as surface plates mentioned earlier. This paper describes and analyzes various alternatives considered by the standards committees. Among them is a new mathematical definition of datums based on constrained least-squares. It seems to provide the best compromise and has the potential to support advanced manufacturing that is increasingly dependent on digital technologies.

The rest of the paper is organized as follows. Section 2 presents some fundamental mathematical concepts of datums that have been standardized by ASME and ISO. Establishment of physical and digital datums is discussed in Section 3. A new mathematical definition of datums based on constrained least-squares fitting is introduced and explained in Section 4. Finally, Section 5 summarizes the main contribution of the paper and offers some concluding remarks.

2. ASME and ISO Standards on Datums

Datums and datum features are defined in ASME and ISO standards using stylized symbols and textual prose. To illustrate this, first consider the problem of specification and verification involving planar datum features. Figure 2(a) shows how a designer may graphically present the specification of position tolerancing of a cylindrical hole in a part with respect to a system of primary and secondary planar datums. Figure 2(b) illustrates how such a system of primary and secondary datum planes may be established on a manufactured instance of the part.

As a second illustration, consider the problem of specification and verification involving a cylindrical datum feature. Figure 3 shows a simple example of how a designer may graphically present the specification of cylindrical datum feature (as primary and secondary datum feature) to position a pattern of holes in a part. Figures 3(b) and 3(c) show how such primary and secondary datums may be established on manufactured instances of the part.

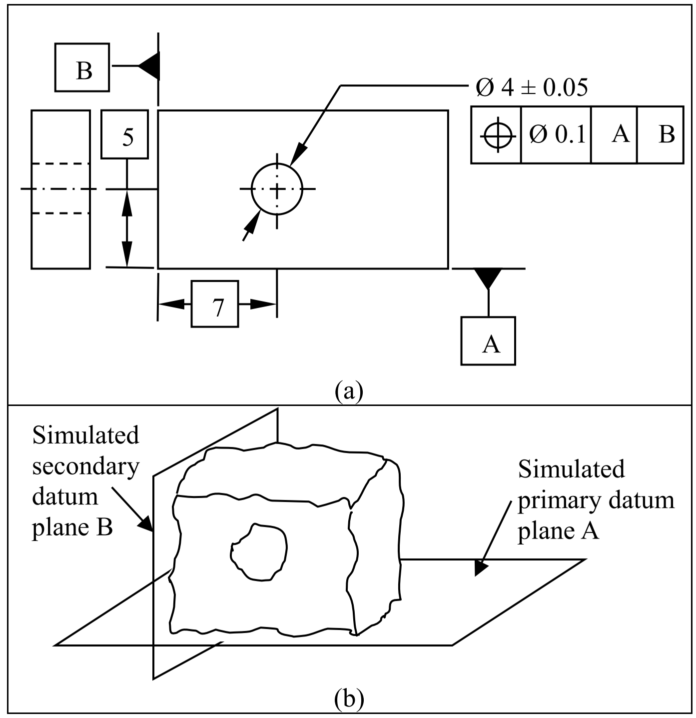


Figure 2. Illustration of planar datums. (a) Specification of planar datum features during design of a part, and (b) establishment of a system of primary and secondary datum planes on a manufactured instance of the part. The secondary datum plane B is required to be perpendicular to the primary datum plane A.

The standards cover more than planar and cylindrical datum features. The following Section 2.1 gives a brief description of a compact classification of *all* datum features based on group theory that has been adopted by both ASME and ISO standards. Features of size play an important role in the definition of datum features in standards, and these are classified and described in Section 2.2.

2.1 Continuous symmetry groups and datums

Datums are used to position one set of geometrical elements relative to another set of geometrical elements. For example, in Fig. 2 a datum system consisting of the primary datum plane A and a secondary datum plane B is used to position a cylindrical feature with respect to the set of planar datum features. Similarly, in Fig. 3 a datum system consisting of planar and cylindrical features is used for relative positioning of a pattern of four cylindrical holes with respect to the set of datum features. Needless to say, there are several different types of datum features in industrial practice than just planar and cylindrical datum features.

A major theoretical breakthrough occurred in the mid-1990s that provided a compact classification of *all* possible types of datum features encountered in industry [15-17]. This is based on a classification of continuous symmetry groups of geometric transformations in 3D, and is enumerated in Table 1. The second column in Table 1 gives the theoretical designations of seven classes of continuous symmetry in 3D. The third column provides some simple illustrative examples of the surface features that fall under these symmetry classifications.

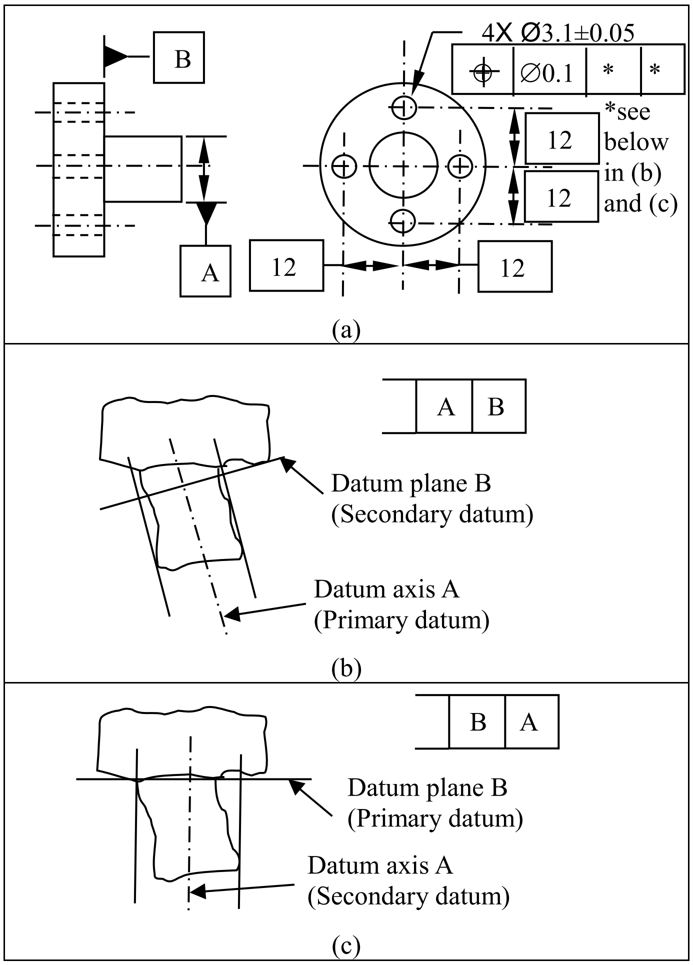
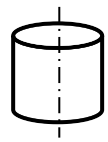
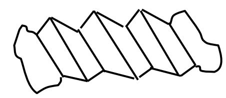
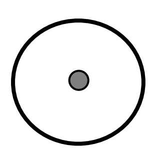
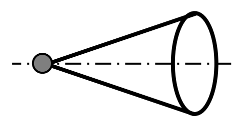
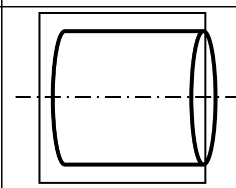
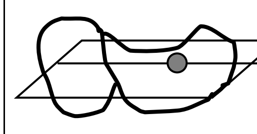


Figure 3. Illustration of cylindrical datum feature. (a) Specification during design of a part, and (b, c) establishment of systems of primary and secondary datums involving cylindrical features on manufactured instances of the part.

To explain what is meant by symmetry, first consider the planar class in Table 1. Any unbounded (that is, infinite) plane remains invariant (that is, remains the same from a point-set consideration) when it is subjected to independent translations along two directions in the plane and one rotation about an axis perpendicular to that plane. The plane thus has three degrees-of-freedom (two translations in the plane, and one rotation perpendicular to the plane) that keep the plane invariant. It should be noted that this observation holds true even for two or more parallel planes – thus parallel planes that bound 'slabs' and 'slots' also have planar symmetry. The three degrees-of-freedom mentioned above for a planar class are also the familiar 'kinematic' degrees-of-freedom. That is, a planar kinematic pair will exhibit these relative motions in a planar kinematic joint.

Table 1. Seven classes of continuous symmetry in 3D [15-17].

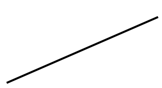
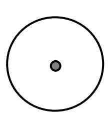
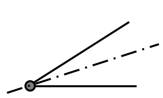
	Symmetry Class Type	Illustrative Surface Features	Reference Element(s)
1	Planar		plane
2	Cylindrical		line (axis)
3	Helical		helix
4	Spherical		point (center)
5	Revolute		line (axis), point-on-line
6	Prismatic		plane, line-on-plane
7	General		plane, line-on-plane, point-on-line

The observation made thus far about planar class can be extended to all the remaining six symmetry classes enumerated in Table 1 as follows. An unbounded right-circular cylinder (or any number of co-axial cylinders) has cylindrical symmetry because it remains invariant under rotation about its axis and translation along its axis; thus it has two degrees-of-freedom. A screw thread has helical symmetry with one-degree-of freedom;

it remains invariant under a helical motion, which is a combination of translation along the axis and rotation about the axis at a rate defined by the pitch of the helix. A sphere (or any number of concentric spheres) has spherical symmetry with three degrees-of-freedom because it remains invariant under the action of the three independent components of rotation in 3D about its center. A right-circular cone (or a torus, or an hour-glass, etc) belongs to a revolute class with only one degree-of-freedom because it remains invariant under rotation about its axis. An oval shaped elliptic cylinder, for example, belongs to the prismatic class with only one degree-of-freedom because it remains invariant under translation along a ruling on its surface. Finally, a saddle shaped object, such as the one shown in the last row of Table 1, has no degree-of-freedom left because it cannot remain invariant under any motion; so it belongs to a general class. A general class can also be defined using datum targets. More extensive theoretical discussions of these classes of symmetry can be found in [17].

An important link between the notion of symmetry and datums is provided by the reference elements listed in the last column of Table 1. These reference elements, which consist of just points, lines, planes, and helices, have the same invariant property as the parent surface features. This is significant because there is an important theorem that establishes the general property that *the relative positioning of any two geometrical objects is the same as the relative positioning of their respective reference elements* [17]. Thus, the reference elements in the last column of Table 1 serve as the datums for the datums features, and this holds true for *any* arbitrary geometric shape. This also brings out a clear distinction between a datum feature and a datum; for example, a cylindrical surface may be a datum feature but its axis is the datum.

Table 2. Three classes of continuous symmetry in 2D [16-17]

	Symmetry Class Type	Illustrative Curve Features	Reference Element(s)
1	Linear		line
2	Circular		point
3	General		line, point-on-line

To complete the datum classification, Table 2 enumerates the three classes of continuous symmetry in 2D. These cases come in handy while establishing secondary datums. Note that the reference elements for the general class in the last row establish a full-fledged two-dimensional reference frame.

The power of symmetry group classifications in defining *all* datum feature types in a compact manner was quickly seized upon by the ISO and ASME standards committees in their recent releases of standards that deal with datums [1, 4]. Table 3 summarizes how these two standards development organizations have chosen to define the datum feature types in their standards. While the ISO standards emphasize the set of rigid motions that keep the surface feature invariant, the ASME standards emphasize the degrees of freedom that are *constrained* by the established datums on a work-piece. For example, ISO takes the view that a planar datum remains invariant under two translational degrees-of-freedom in the plane and one rotational degree-of-freedom about an axis perpendicular to the plane. However, ASME focuses on the fact that a planar datum established on any work-piece *constrains* the following degrees-of-freedom of the work-piece: one translational degree-of-freedom perpendicular to the plane and two rotational degrees-of-freedom about two orthogonal axes on the plane. Even though ASME and ISO take complementary points of view on the degrees-of-freedom, the datum feature type classification still remains the same.

Table 3. ASME [1] and ISO [4] standardized datum feature types.

	Symmetry Class Type	ISO Invariant Classes (Kinematic Degrees-of-Freedom)	ASME Datum Features (Constrained Degrees-of-Freedom)
1	Planar	Planar	Planar
			Width
2	Cylindrical	Cylindrical	Cylindrical
3	Helical	Helical	(None)*
4	Spherical	Spherical	Spherical
5	Revolute	Revolute	Conical
6	Prismatic	Prismatic	Linear Extruded Shape
7	General	Complex	Complex

*not explicitly type classified in the standard

It is worth noting that the ASME standard separates single plane and two parallel planes (as width) even though they both fall under the same mathematical symmetry group (planar). Also, the ASME standard refers to revolute and prismatic classes as conical and linear extruded shape, respectively. These minor terminological variations from the ISO definitions may be attributed to simple engineering convenience and pose no major theoretical problems. Both ASME and ISO standards refer to the general symmetry group as ‘complex,’ not to be confused with mathematical analysis of geometry in the space involving complex numbers. Also, both ASME and ISO standards downplay the mathematically identified helical class because, in practice, helical threads are handled by other standardized specifications [18] and only the axis of the helix plays any significant practical role as a datum.

The compact classification of datum types, as shown in Table 3 and adopted by both ASME and ISO, will play an important role in a search for a unified mathematical datum fitting criterion described in Sections 3 and 4. But, before that, a brief examination of the special role played by 'features of size' in defining datums in ASME and ISO standards is in order.

2.2 Features of size and datums

Size plays an important role in industry and in standards [19]. In general, any feature that can be assigned a size dimension and tolerance can be called a *feature of size*. But standards reserve the designation of 'feature of size' to a limited number of features.

The current ASME standard [1] defines two major types of features of size. The first is a *regular feature of size*: it can be a cylindrical surface, a spherical surface, a circular element, a set of two opposed parallel elements, or a set of two opposed parallel surfaces. The second is an *irregular feature of size*, which is broken down into the following two subtypes: (1) a feature or collection of features that may contain or be contained by an envelope that is a sphere, a cylinder, or a pair of parallel planes, (2) a feature or collection of features that may contain or contained by an envelope other than a sphere, a cylinder, or a pair of parallel planes. In any case, a feature of size falls under one of the symmetry classes enumerated in Tables 1, 2 and 3. Any feature of size can be used as a datum feature in two ways: either (1) regardless of material boundary (RMB), or (2) at maximum material boundary (MMB) or least material boundary (LMB). This paper focuses on the mathematical datum definitions at RMB.

The ISO standards take a somewhat different view on features of size. In the recent ISO definition, a feature of size must satisfy two criteria: (1) it must belong to a one-parameter family of features, and (2) it must obey the monotonic containment property. The monotonic containment property refers to the fact that a feature with a larger parameter value (that is, size) should contain (or be contained in) the feature with a smaller parameter value (that is, size).

With this general stipulation, ISO standards define two types of sizes: linear sizes and angular sizes. For example, a cylinder is a feature of size with a linear size. On the other hand, a cone is a feature of size with an angular size (e.g., its apex angle). Both cylinder and cone satisfy the two criteria mentioned above: each belongs to a one-parameter family of surfaces and each satisfies the monotonic containment property. Any ISO feature of size also belongs to a symmetry group enumerated in Tables 1, 2 and 3. The notions of RMB, MMB, and LMB are also applicable in the ISO standards, albeit under different names. As stated earlier, this paper is concerned only with mathematical datum definitions at RMB.

The prevalence of some of the common and popular features of size in ASME and ISO standards may be traced mathematically to the widespread use of quadric surfaces (that is, surfaces that are governed by second degree implicit equations) in industry. There is a remarkable classification theorem that states that all quadric surfaces can be classified under seventeen types, out of which twelve types correspond to surfaces in the real space (as opposed to the space involving complex numbers) [17]. Of these twelve, the popular ones used in industry that belong to one-parameter family of surfaces are listed in Table 4; these also satisfy the monotonic containment property. The non-degenerate quadrics in Table 4 are also known as 'natural quadrics' and are widely used, along with planes, in geometrical and solid modeling software systems.

Table 4. Popular one-parameter families of real quadrics [17].

	Surface Type	Canonical Equation	Parameter
Non-degenerate quadrics	Sphere	$x^2 + y^2 + z^2 = r^2$	r
	Right-circular cylinder	$x^2 + y^2 = r^2$	r
	Right-circular cone	$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = 0$	$\tan^{-1}(a/c)$
Degenerate quadrics	Parallel planes	$x^2 = a^2$	a
	Intersecting planes	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$	$\tan^{-1}(b/a)$

To complete the classification, Table 5 presents the popular one-parameter family of real conics (that is, planar curves of the second degree) in 2D. These appear as two-dimensional features of size (both linear and angular) in the standards.

Table 5. Popular one-parameter families of real conics [17].

	Curve Type	Canonical Equation	Parameter
Non-degenerate conic	Circle	$x^2 + y^2 = r^2$	r
Degenerate conics	Parallel lines	$x^2 = a^2$	a
	Intersecting lines	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$	$\tan^{-1}(b/a)$

In both ASME and ISO standards, features of size form only a subset of what can be specified as datum features. But the features of size play a special role in datum establishment because they can be specified with Ⓜ and Ⓛ modifiers to indicate when these datum features are used at MMB and LMB, respectively. As noted earlier, this paper does not deal with these modifiers, because datum feature simulators at MMB and LMB have fixed size values and can be handled by different means (e.g., using functional gages in the physical domain or soft gages in the digital domain).

3. Physical and Digital Establishments of Datums

In addition to standardizing the types of datums, the standards define how *theoretical* datums are established. The ASME standard [1] uses the notion of a 'theoretical datum feature simulator' to associate a surface of perfect form to a manufactured part feature that has an imperfect form. For example, as shown in Fig. 2(b), a mathematical plane serves as a theoretical datum feature simulator to establish a primary planar datum plane on a manufactured surface feature that may have irregularities and undulations. Similarly, in Fig. 3(b), a mathematical cylinder serves as a theoretical datum feature simulator on a manufactured surface feature with imperfect (that is, irregular) form; the axis of the mathematical cylinder then serves as the primary datum. Echoing the last column of Tables 1 and 2, the ASME standard asserts that "datums are theoretically exact points, axes, lines, and planes."

The ISO standard [4] uses a slightly different terminology to define the same process for theoretical datum establishment. In the ISO parlance, an 'associated integral feature' is a mathematical surface of perfect form that is fitted to an 'extracted integral feature' (which corresponds to a 'nominal integral feature') of a work-piece. The datum is then the 'derived feature' (e.g., an axis) of the associated integral feature. Without getting into the details of ISO terminology, suffice it to say that ASME's 'theoretical datum simulator' and ISO's 'associated integral feature' are equivalent, and they lead to the same theoretical datum definition.

Theoretical datums defined by ASME and ISO can be established on manufactured parts using either physical devices as outlined in the following Section 3.1 or using digital means as described subsequently in Section 3.2.

3.1 Physical datum establishment

Physical datums on manufactured parts can be established using what the ASME standard calls 'physical datum feature simulators,' which are the physical counterparts to the ASME's 'theoretical datum feature simulators' or the ISO's 'associated integral features.' For example, a surface plate can be used as a physical datum feature simulator for a planar datum feature depicted in Fig. 2(b). Similarly, a contracting chuck can be used as a physical datum feature simulator for a cylindrical datum feature shown in Fig. 3(b).

The world of physical metrology is full of wonderful devices such as surface plates, angle blocks, opening and closing vises, expanding mandrels, and contracting chucks to establish physical datums. Even though such physical devices do not have surfaces of mathematically perfect form, they are supposed to be produced with such high quality that they can be used to establish datums in industrial practice while their small form deviations contribute only very small measurement uncertainty.

Employing physical devices still requires some guidance on how they may be used to establish datums on manufactured parts. The ASME standard provides some tips for situations such as the one shown in Fig. 2. Let Q be the manufactured part in Fig. 2. The primary planar datum in Fig. 2(b) can be established by placing Q on a surface plate that serves as a physical datum feature simulator so that the surface plate contacts "a minimum of three points" on the manufactured part. An angle block that is perpendicular to the surface plate can serve as the secondary physical planar datum simulator in Fig. 2(b) by setting it up so that it contacts "a minimum of two points" on Q. Another angle block that is mutually perpendicular to the two physical devices can then serve the tertiary physical planar datum simulator (not shown in Fig. 2) by contacting "a minimum of one point" on Q. This process of establishing a mutually perpendicular planar datum system follows a '3-2-1 fixture principle,' as it is known in the engineering folklore. It is well entrenched in industrial practice for locking down all six degrees of freedom of a manufactured part.

The 3-2-1 fixture principle is also called the Kelvin principle [20]. There is another fixture principle known as the Maxwell principle, which is designated as '2-2-2' to indicate that all six degrees of freedom of a part can be constrained when each of the three mutually orthogonal datum planes contacts a minimum of two points. This brings up an interesting question. How can a stable primary datum plane be physically established if not enough contact points can be found on a manufactured part using either the Kelvin or the Maxwell principle? Figure 4 illustrates such a situation with a two-dimensional example.

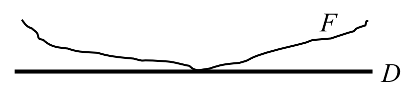


Figure 4. Associating a planar datum *D* to a manufactured feature *F*.

Even though it is drawn with some exaggeration for illustration, Fig. 4 presents the type of dilemma that a metrologist faces in establishing a physical datum. In the ASME standards [1, 2] this situation is described as a 'rocking' condition. Under the rocking condition, there may be a 'set of candidate datums' instead of just one datum *D* for a manufactured feature *F*. Recognizing the same problem, previous versions of ISO standards recommended 'shimming' to provide supports between *F* and *D* in Fig. 4 to ensure some stability for *F* relative to the datum *D*. Neither of the solutions is considered satisfactory because both leave the datum establishment to the skills of the metrologist who serves as an inspection technician.

The problem of rocking and non-uniqueness is not confined to planar datum establishment. Physical datum establishments for other datum features enumerated in Tables 1, 2 and 3 also encounter similar problems. One remedy to this problem is to come up with a good mathematical criterion to associate a *unique* perfect form surface to a datum feature of imperfect form. It turns out that there are several mathematical criteria, each of which has attempted to provide guidance to establishing physical datums as well as digital datums, as described next.

3.2 Digital datum establishment

Digital datum establishment depends on mathematical definitions and related computational techniques, based on a *computable representation* of datum features on manufactured parts. Only in the past few decades has such a computable representation of datum features became widely available, thanks largely to the rise of CMS and scanning devices. And the availability of cheap computing power, along with powerful software, has introduced digital datum establishment as a potential competitor, and complement, to physical datum establishment in industry.

Any good mathematical definition of datums, whether it is for establishing physical datums or digital datums, starts with a fitting problem posed as an optimization problem. To set up the optimization problem with its objective function and constraints, consider a computable representation of a surface feature F on the boundary of a manufactured instance of a part and a corresponding mathematical surface S of perfect form, as shown in Fig. 5. Let q be any point on F and dA an elemental area around q. Also let $d(q, S)$ be the perpendicular distance between p and S; it is also the shortest distance between p and S. While F is a bounded set, S can be unbounded. For example, S can be an unbounded plane or an unbounded cylinder. Also, F comes from a manufactured part; so there is an unambiguous notion of when S lies outside the material side of F.

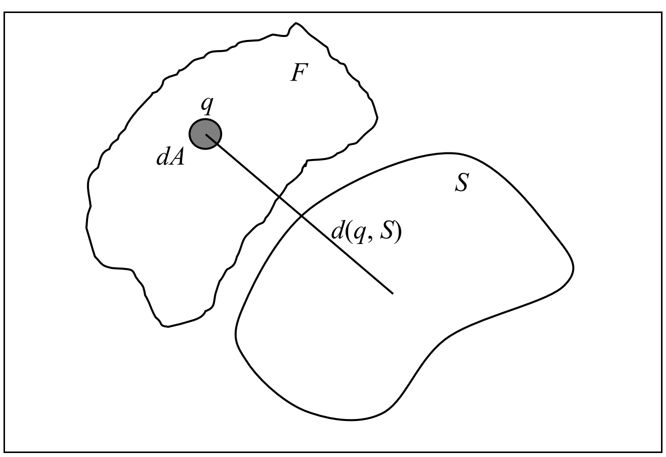


Figure 5. Notations for fitting a surface S to a feature F.

It was mentioned that F has a computable representation. It means that F may be a collection of n points, represented as a set $\{q_1, q_2, \dots, q_n\}$ with each point q_i having three coordinates that could be generated by a CMS. Then F is a finite set. In most applications, it is preferable to represent F as a continuum consisting of an infinite number of points. Then a computable representation of F may involve an interpolation (e.g., piece-wise linear tessellation) of measured points from a CMS.

The fitted surface S can be represented mathematically by the type of surface being fitted (e.g., plane, cylinder, sphere) along with its size and position parameters that may be represented as $\boldsymbol{u} = \{u_1, u_2, \dots, u_m\}$. In datum fitting problems, $m \ll n$. That is, the number of unknown parameters in $\boldsymbol{u}$ associated with S is much smaller that the number of known discrete points on F. Fitting a surface S to F is then an optimization problem of minimizing a distance measure $D(F, S)$ subject to the constraint that S lies on or outside the material side of F; the output is the parameter set $\boldsymbol{u}$.

Among many choices for the distance measure $D(F, S)$, the ones using the L_p-norm have been popular in industrial applications. Under this norm, the optimization problem is posed as

$$\min_{\boldsymbol{u}} \left[\frac{1}{A}\int_F |d(q, S)|^p dA\right]^{1/p} \tag{1}$$

subject to the constraint that S lies on or outside the material side of F. In Eq. (1), A stands for the area of the surface feature F. If the feature F is represented only discretely using uniform sampling, then the integration in Eq. (1) is replaced by a summation, resulting in the following optimization problem that uses the l_p-norm:

$$\min_{\boldsymbol{u}} \left[\frac{1}{n}\sum_{i=1}^{n} |d(q_i, S)|^p\right]^{1/p} \tag{2}$$

still subject to the constraint that S lies on or outside the material side of F. Three popular choices of p are 1,2, and ∞, and these three cases are discussed below in that order.

Case $p = 1$**.** Under this norm, the fitted surface S minimizes the integral (or sum) of the distances of the points on F to S (with the aforementioned material constraint). This problem, when a planar surface S is fitted to F, has been studied and solved recently [21-23]. It has been shown that minimization under the L_1-norm also minimizes the distance of the centroid of F to the planar datum S. This property is favored by engineers who seek mechanical stability in establishing datums

In particular, the advantage of the L_1-norm in establishing a planar datum system is that (in the discrete case) a primary planar datum will always contact a minimum of three points, a secondary planar datum will always contact a minimum of two points, and a tertiary planar datum will contact one point. Thus the '3-2-1 Kelvin fixture' principle will be observed, giving this digital datum system a seemingly strong advantage of mechanical stability.

There is, however, a problem of numerical stability with the L_1-norm that is illustrated in Fig. 6 using a simple 2D example. Consider a V-shaped manufactured feature F that is supposed to be flat, and a linear datum D fitted to it. Let s_1 and s_2 be the lengths of the two arms of F as indicated in Fig. 6. If $s_1 > s_2$ even by a very small amount, which can occur in numerical computations, the datum D will be aligned with the left arm of F as shown in Fig. 6(a). On the other hand, if the reverse were to occur and $s_1 < s_2$ by a very small amount, then the datum D will flip to the right arm of F as shown in Fig. 6(b). This type of flip-flop will occur even if the left and right arms of F are almost collinear (that is, their included angle differs from 180° by a very small amount).

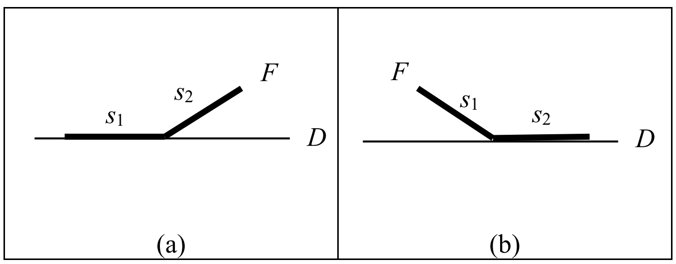


Figure 6. Illustration of numerical stability issues with the L_1-norm.

Figure 6 illustrates that seeking mechanical stability is not always the right objective in establishing datums. An engineer might be quite happy with a datum D that contacts F at just one point in Fig. 6 (at the junction of the two arms), as long as D has numerical stability. That is, if any small change in F results in only a small change in D with respect to F. In Section 4, it will be shown that a constrained least-squares fitting can accomplish this.

***Case* $p = 2$.** This is the well-known least-squares norm, also known as the Gaussian norm. In particular, it is called the 'orthogonal least-squares' or 'total least-squares' fit, since the distances of the points in F to S are measured perpendicular to S. Unconstrained least-squares fitting is very popular with numerical analysts and software developers in the field of computational coordinate metrology [24-27].

CMS vendors have implemented and used unconstrained least-squares fitting to discrete set of points, as defined in Eq. (2), for various geometric elements listed in Tables 1, 2 and 3, for a long time. The objective function is relatively simple, yielding a closed form expression for the gradient used in search algorithms. Such fits also have the advantage of numerical stability (that is, a small change in the input representation of F has only a small change in the output representation of S).

In spite of the popularity of unconstrained least-squares fitting, it has remained only in the background in datum establishment. This is because digital datum establishment typically requires the fitted surface to lie outside the material. The *constrained* least-squares fitting, which has been explored only recently, enforces the constraint that the fitted surface lies outside the material and it will be described in some detail in Section 4. It should be noted that constrained least-squares fitting is different from an unconstrained fit that has been shifted to lie just outside the material.

***Case* $p = \infty$.** This norm leads to minimax fitting. That is, minimizing the maximum deviation of points in F to S while maintaining the constraint that S lies to the outside of the material side of F. A class of problems that come under minimum-zone fitting can be posed as problems that have the L_∞-norm in the objective function, but without the constraint. The minimum-zone fitting is also known as Chebyshev fitting [28, 29].

Computational solutions to these unconstrained L_∞-norm problems are very useful for conformance assessment to form tolerancing problems, such as straightness, flatness, roundness, and cylindricity. Attempts have been made to use one of the surfaces (that lies outside the material) in the minimum-zone as the fitted surface associated with that feature for datum establishment. But these attempts have not met with much success because other fitting criteria described below have found favors in industry and standards.

Before looking at the constrained least-squares in detail in Section 4, it is useful to consider other fits, such as maximum inscribing and minimum circumscribing fitting of geometrical elements such as circles, parallel lines, parallel planes, spheres, and cylinders. (Interestingly, the maximum inscribing and minimum circumscribing fittings are also known as Chebyshev fittings [28].) The popularity of this type of fitting in standards stems from the fact that the minimum circumscribing and maximum inscribing fits can be achieved using the likes of contracting chucks and expanding mandrels when establishing physical datums. Hence maximum inscribing and minimum circumscribing fittings have been the default criteria in ASME and ISO standards for establishing datums involving circles, cylinders, spheres, and parallel planes.

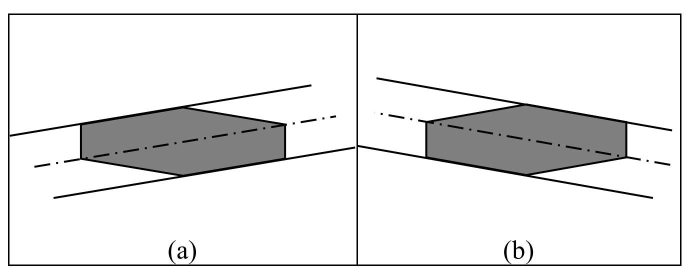


Figure 7. An illustration of the non-uniqueness and numerical stability issues with minimum circumscribing parallel lines fitting.

These fits (with the appropriate constraint) are considerably more complex than the fitting algorithms associated with the L_2-norm. In addition, there are some numerical stability issues with these fits for datums, as illustrated in Fig. 7 for a 2D problem. Consider the problem of establishing a planar datum using minimum circumscribing parallel planes (mimicking closing vises in the physical world) for a slab. The object in Fig. 7 is supposed to be a slab with nominally planar, parallel faces at the top and the bottom. But the manufactured part, as depicted in Fig. 7, has a V-shaped crown at the top and bottom faces of the slab. A minimal-separation fit can lead to non-unique solutions as seen in Figs. 7(a) and 7(b). With small numerical changes in the top and bottom faces of the slab, the median line indicated as the datum in Fig. 7 can flip-flop between Fig. 7(a) to Fig. 7(b), similar to the case seen in Fig. 6.

The maximum inscribing cylinders and spheres (mimicking expanding mandrels in the physical world) also run into a similar problem, as illustrated with maximum inscribing circles in Fig. 8. Consider a circular hole that has been manufactured as a combination two identical circles, shown dotted in Fig. 8. The maximum inscribing circle problem for this hole does not have a unique solution, because each of the dotted circles is a maximum inscribing circle having different centers as shown. This also illustrates a numerical stability problem. If one of the circles is slightly larger than the other, the fit will choose the center of that circle as the datum. This can lead to an undesirable flip-flop situation for digital datum establishment.

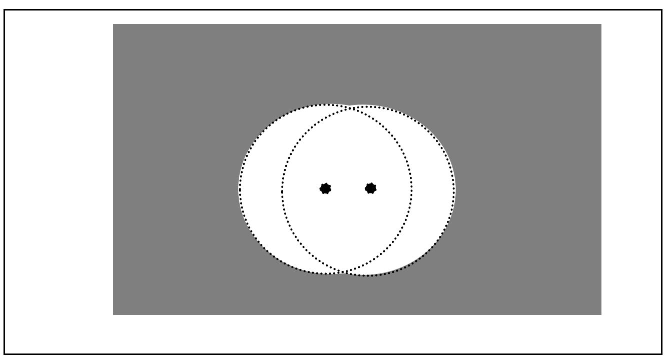


Figure 8. An illustration of the non-uniqueness and numerical stability issues with maximum inscribing circle fitting (Chebyshev fitting).

The non-uniqueness and stability problems with maximum inscribing and minimum circumscribing fits are not confined to digital datum establishment. As illustrated in Figs. 7 and 8, physical datum establishment employing closing vises and expanding mandrels will also face similar problems. In the physical world, the solution to the problem is left to the skills of the operator of these devices in the inspection shop.

In addition to the problems with traditional features of size such as circles, parallel planes, cylinders, and spheres, the maximum inscribing and minimum circumscribing fittings do not generalize to other features enumerated in Tables 1, 2 and 3. For example, solid wedges and angular slots (which are angular features of size that belong to the prismatic class) cannot be covered by these criteria. Also, solid and hollow cones (which are angular features of size that belong to the revolute class) are not covered by the maximum inscribing and minimum circumscribing optimization criteria.

4. Constrained Least-squares Fitting

Experience with various physical and digital datum establishments, as described in Section 3, has recently focused the attention of ASME and ISO standards committees to seek *one* fitting criterion that can be uniformly applied (1) to *all* the datum feature types enumerated in Tables 1, 2 and 3, and (2) to *both* physical and digital datum establishments with *mechanical* as well as *numerical* stability. It is clear from the discussions of Section 3 that this is a challenging task, and some compromise has to be made.

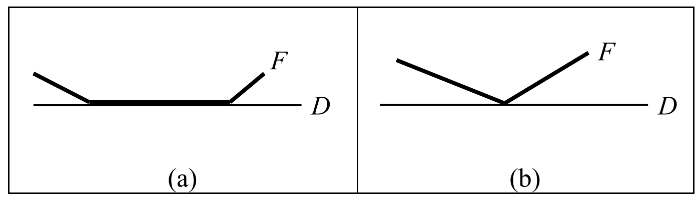


Figure 9. Illustration of linear datum fitting using constrained least-squares.

The constrained least-squares fitting seems to offer this compromise. There is considerable experience in numerical analysis and software development in the theory and application of unconstrained least-squares fitting for all types of geometrical elements [24-27]. However, the problem of constrained least-squares fitting has been attacked only recently [30-34]. The early results are encouraging, as illustrated in Fig. 9 for simple examples in 2D. Figure 9(a) shows how a linear datum *D* will make a mechanically stable contact with the long arm of the feature *F*, even if the two arms shown not flush with *D* are of unequal lengths (to a reasonable extent). This full contact of the constrained least-squares is not obtained by using a shifted least-squares approach. In Fig. 9(b), the datum *D* contacts the junction of the two arms of *F*; this fitting is numerically stable because small changes in *F* will result in only small changes in *D*. It is worth comparing Fig. 9(b) with Fig. 6.

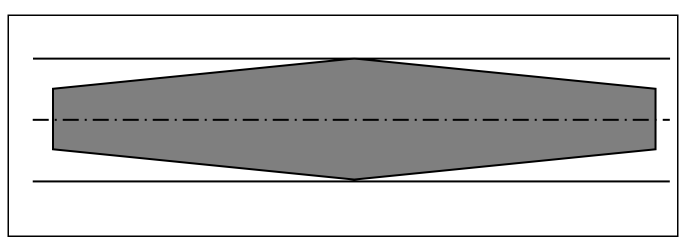


Figure 10. Illustration of datum fitting for a slab in 2D using constrained least-squares.

Figure 10 shows the result of constrained least-squares fitting to establish the digital datum for a slab using a 2D example. It leads to a unique and numerically stable solution, and avoids the type of flip-flop problem seen in Fig. 7 with Chebyshev fitting. More technical information on the constrained least-squares fitting for linear datum features can be found in [30, 31, 34].

The same constrained least-squares criterion can be applied to non-linear datum features as well [32, 33]. Figure 11 illustrates the constrained least-squares fitting of a circle to the same problem presented in Fig. 8. Such a fitting has uniqueness and numerical stability, and will not lead to the flip-flop situation encountered with Chebyshev fitting (in this case, maximum inscribing circle) of Fig. 8. As shown in [32], in many practical cases, the constrained least-squares fits similar to the one shown in Fig. 11 differ from a maximum inscribed circles (similar to the one shown in Fig. 8) by only a negligible amount. Further, [32] shows that the constrained least-squares fitting is preferable to the shifted least-squares fitting.

The initial reaction from the ASME and ISO standards committees has been very positive in favor of the constrained least-square fitting. It seems to offer the best compromise between mechanical stability and numerical stability for establishing physical and digital datums. It is also extendable to all the types of datum features enumerated in Tables 1, 2 and 3.

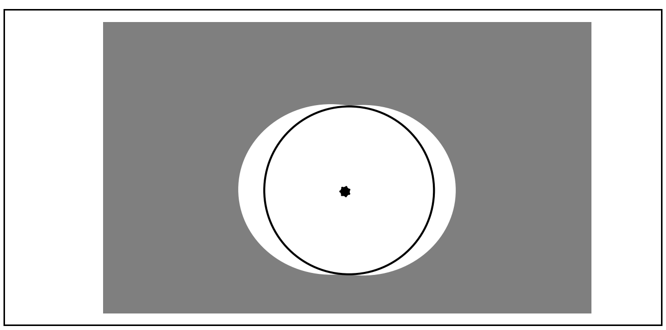


Figure 11. Illustration of constrained least-squares fitting to a non-linear datum feature.

To provide input to the standards committees and to enable efficient software implementation, a detailed research was initiated to study the structure and behavior of constrained least-squares fitting. It has been proved that the objective function for constrained least-squares fitting of planes and parallel planes is convex, and that the contact conditions for the optimal solution have a compact classification [34]. The optimality conditions for the constrained least-squares fitting of circles, cylinders, and spheres have also been proved [33]. It has been shown that the optimal circle, cylinder, or sphere must contact at least two points in the input set of feature F. Table 6 provides a snapshot of the coverage of current research on the convexity and optimality conditions (in the form of combinatorial characterization) for constrained least-squares fitting to establish datums. Using these results, major CMS vendors are beginning to implement and test their software for constrained least-squares fitting.

A careful study of Tables 1 through 6 reveals that there are still several research issues that remain to be explored. A combinatorial characterization of the optimality conditions (that is, what is the minimum number of points of F that will contact S) for the revolute, prismatic (beyond intersecting planes), and complex classes will be helpful to enable efficient implementations of the constrained least-squares fitting. Other than the cases that involve planes (single planes, parallel planes, and intersecting planes), characterizing the optimum solution using the constrained least-squares fitting to a continuous set (as opposed to a discrete set) of points remains an open problem.

Open standards development organizations such as ASME and ISO do not concern themselves with algorithmic issues. They focus only on the definition of the optimization problem with its objective function and constraints. Sometimes, national metrology research laboratories in the USA and Europe develop research software (that is correct but not necessarily fast) to test the industrial solutions to the optimization problems. But it is left to the private sector to come up with efficient implementations of algorithms to solve these optimization problems, and compete in the market place. The research community at large can and should undertake detailed algorithmic studies on constrained least-squares fitting to assist industry in this effort. The arrival of cheap Graphics Processing Units (GPUs) may dramatically improve the performance of industrial software to establish digital datums based on constrained least-squares from point clouds.

Table 6. Coverage of current research on constrained least-squares fitting.

Datum Feature	Symmetry Group	Feature of Size (RMB)	References
Line	Linear (2D)	No	[30, 31, 34]
Two parallel Lines	Linear (2D)	Linear size	
Two intersecting Lines	General (2D)	Angular size*	
Plane	Planar	No	
Two parallel planes	Planar	Linear size	
Two intersecting planes	Prismatic	Angular size*	
Circle	Circular (2D)	Linear size	[32, 33]
Cylinder	Cylindrical	Linear size	
Sphere	Spherical	Linear size	

*ISO terminology

5. Summary and Concluding Remarks

This paper described recent efforts in ASME and ISO standards committees to come up with a new mathematical definition of datums. A definition based on the constrained least-squares fitting criterion has been found to be the most promising mathematical definition to cover *all* the datum feature types defined by the standards and the best compromise to support both physical and digital datum establishments. The major contribution of this paper is to bring these developments to the attention of the research community, and to encourage further mathematical and algorithmic research to enable efficient software development. As manufacturing becomes more and more digitized, such software will play an increasingly dominant role in the advancement of manufacturing.

Acknowledgment and Disclaimer

The authors thank ISO and ASME standards experts whose advice and suggestions were invaluable in initiating and sustaining this research investigation. Any mention of commercial products or systems in this article is for information only; it does not imply recommendation or endorsement by NIST.

References

[1] ASME Y14.5-2009, Dimensioning and Tolerancing, The American Society of Mechanical Engineers, New York, 2009.

[2] ASME Y14.5.1-1994, Mathematical Definition of Dimensioning and Tolerancing Principles, The American Society of Mechanical Engineers, New York, 1994.

[3] ISO 1101:2017, Geometrical product specifications (GPS) – Geometrical tolerancing – Tolerances of form, orientation, location and run-out, International Organization for Standardization, Geneva, 2017.

[4] ISO 5459:2011, Geometrical product specifications (GPS) – Geometrical tolerancing – Datums and datum systems, International Organization for Standardization, Geneva, 2011.

[5] ASME Y14.41-2012, Digital Product Definition Data Practices, The American Society of Mechanical Engineers, New York, 2012.

[6] Krulikowski, A., Fundamentals of Geometric Dimensioning and Tolerancing, 3rd Edition, Effective Training Inc., Westland, MI, 2012.

[7] Srinivasan, V., Geometrical Product Specification, in CIRP Encyclopedia of Production Engineering, Laperrière, L. and Reinhart, G. (Eds), Springer, 2014.

[8] Wick, C. and Veilleux, R.F., Tool and Manufacturing Engineers Handbook, Vol. IV Quality control and assembly, 4th Edition, Society of Manufacturing Engineers, Dearborn, MI, 1987.

[9] Shakarji, C. M., 2011, "Coordinate Measuring System Algorithms and Filters," Coordinate Measuring Machines and Systems, R. J. Hocken, and P. H. Pereira, eds., CRC Press, pp. 153–182.

[10] The Wall Street Journal, The internet of things in here, and it isn't a thing, Aug. 21, 2016.

[11] Hedberg Jr, T. D., Lubell, J., Fischer, L., Maggiano, L., and Barnard Feeney, A., Testing the digital thread in support of model-based manufacturing and inspection, ASME Journal of Computing and Information Science in Engineering, *16*(2), 1-10, 2016.

[12] Hedberg Jr, T.D. and Carlisle M. (Eds), Proceedings of the 2017 Model-based Enterprise (MBE) Summit, NIST AMS 100-11, Gaithersburg, MD, 2017.

[13] The Wall Street Journal, Germany bets on smart factories to keep its manufacturing edge, Oct. 26, 2014.

[14] Jeschke, S., Brecher C., Song, H., and Rawat, D.B., (Eds) Industrial Internet of Things: Cybermanufacturing Systems, Springer, 2017.

[15] Clement, A., Riviere, A., and Temmerman, M., Cotation Tridimensionnelle des Systemes Mechaniques, Ivry-sur-Seine: PYC Edition, France, 1994.

[16] Srinivasan, V., A geometrical product specification language based on a classification of symmetry groups, Computer-Aided Design, Vol. 31, No. 11, pp. 659-668, 1999.

[17] Srinivasan, V., Theory of Dimensioning: An introduction to parameterizing geometric models, Marcel Dekker, New York, 2014.

[18] ASME B1.13M, Metric Screw Threads: M Profile, The American Society of Mechanical Engineers, NY, 2015.

[19] Morse, E.P., and Srinivasan, V., Size tolerancing revisited: A basic notion and its evolution in standards, Proc. Institution of Mech. Engg. Part B: Journal of Engineering Manufacture, Vol. 227, No. 5, pp. 662-671, 2013.

[20] Slocum, A., Kinematic couplings: A review of design principles and applications, Int. Journal of Machine Tools and Manufacture, Vol. 50, No. 4, pp. 310-327, 2010.

[21] Shakarji, C.M. and Srinivasan, V., Theory and algorithms for L_1 fitting used for planar datum establishment in support of tolerancing standards, ASME DETC2013-12372, Proceedings of the ASME 2013 IDETC/CIE Conferences, Portland, OR, 2013.

[22] Shakarji, C.M. and Srinivasan, V., An improved L_1 based algorithm for standardized planar datum establishment, ASME DETC2014-35461, Proceedings of the ASME 2014 IDETC/CIE Conferences, Buffalo, NY, 2014.

[23] Shakarji, C.M. and Srinivasan, V., Datum planes based on a constrained L_1 norm, ASME Journal of Computing and Information Science in Engineering, Vol. 15, December 2015.

[24] Butler, B.P., Forbes, A.B., and Harris, P.M., Algorithms for geometric tolerance assessment, NPL Report DITC 228/94, National Physical Laboratory, U.K., 1994.

[25] Forbes, A., Least-squares best-fit geometric elements, National Physical Laboratory Report, U.K., DITC 140/89, 1991.

[26] Srinivasan, V., Shakarji, C.M. and Morse, E.P., On the enduring appeal of least-squares fitting in computational coordinate metrology, ASME Journal of Computing and Information Science in Engineering, Vol. 12, March 2012.

[27] Shakarji, C.M. and Srinivasan, V., Theory and algorithms for weighted total least-squares fitting of lines, planes, and parallel planes to support tolerancing standards, ASME

Journal of Computing and Information Science in Engineering, 13(3), 2013.

[28] Anthony, G.T. et al., Chebyshev Best-Fit Geometric Elements, NPL Report DITC 221/93, National Physical Laboratory, U.K., 1993.

[29] Zhang, H. and Rou, U., Criteria for establishing datums in manufactured parts, Journal of Manufacturing Systems, 12(1), pp. 36-50, 1993.

[30] Shakarji, C.M. and Srinivasan, V., A constrained L_2 based algorithm for standardized planar datum establishment, ASME IMECE2015-50654, Proceedings of the ASME 2015 International Mechanical Engineering Congress and Exposition, Houston, TX, Nov. 13-19, 2015.

[31] Shakarji, C.M. and Srinivasan, V., Theory and algorithm for planar datum establishment using constrained total least-squares, 14^{th} CIRP Conference on Computer Aided Tolerancing, Gothenburg, Sweden, 2016.

[32] Shakarji, C.M. and Srinivasan, V., Computational investigations for a new, constrained least-squares datum definition for circles, cylinders, and spheres, ASME IMECE2016-66753, Proceedings of the ASME 2016 International Mechanical Engineering Congress and Exposition, Phoenix, AZ, Nov. 11-17, 2016.

[33] Shakarji, C.M. and Srinivasan, V., Optimality conditions for constrained least-squares fitting of circles, cylinders, and spheres, ASME DETC2017-67143, Proceedings of the ASME 2017 IDETC/CIE Conferences, Cleveland, OH, 2017.

[34] Shakarji, C.M. and Srinivasan, V., Convexity and optimality conditions for constrained least-squares fitting of planes and parallel planes to establish datums, ASME IMECE2017-70899, Proceedings of the ASME 2017 International Mechanical Engineering Congress and Exposition, Tampa, FL, 2017.