Proceedings of the ASME 2017 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC2017
August 6-9, 2017, Cleveland, Ohio, USA

DETC2017-67143

Optimality Conditions for Constrained Least-Squares Fitting of Circles, Cylinders, and Spheres to Establish Datums

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

This paper addresses the combinatorial characterizations of the optimality conditions for constrained least-squares fitting of circles, cylinders, and spheres to a set of input points. It is shown that the necessary condition for optimization requires contacting *at least* two input points. It is also shown that there exist cases where the optimal condition is achieved while contacting *only* two input points. These problems arise in digital manufacturing, where one is confronted with the task of processing a (potentially large) number of points with three-dimensional coordinates to establish datums on manufactured parts. The optimality conditions reported in this paper provide the necessary conditions to verify if a candidate solution is feasible, and to design new algorithms to compute globally optimal solutions.

1. Introduction

Datums are important for the specification and verification of product geometries before, during, and after manufacturing [1-7]. Traditionally, datums were established on a manufactured part using physical devices such as surface plates, expanding and closing vises, expanding mandrels, and contracting chucks [8]. In the current era of digital manufacturing, one also faces the task of establishing datums by *fitting* mathematical objects such as planes, circles, cylinders, and spheres to a cloud of points, which may number in the millions, sampled on a manufactured part using coordinate measuring systems (CMS). Standards development organizations such as ISO (International Organization for Standardization) and ASME are now responding to this trend by moving beyond supporting merely analog (i.e., physical) inspection devices to more general standards definitions that also support such digital (i.e., coordinate measurement) technologies. Experts in ISO/TC 213 and ASME Y14 standards committees are now engaged in defining the proper fitting criteria that simulate in the digital world what has been practiced in the physical world thus far.

Mathematically and computationally, fitting can be viewed as an optimization problem. Two recent papers [9, 10] addressed the problem of fitting primary and secondary datum planes using constrained total least-squares. Another recent paper [11] reported some promising results in the use of constrained least-squares fitting for circles, cylinders, and spheres to establish datums based on computational investigations.

This paper builds on the insights gained from [11], and addresses some of the theoretical aspects of such non-linear optimization problems. It will address only the combinatorial characterizations of the optimality conditions. Such optimality conditions provide two practical benefits [12]. First, they provide the necessary conditions that any solution to the non-linear optimization problem must satisfy; this comes in handy in (partially) verifying the correctness of a solution. Second, these optimality conditions suggest the design of new algorithms to find a global solution to the non-linear optimization problem; such algorithms may take advantage of modern advances in graphics processing units (GPUs) that are entering regular use in engineering and manufacturing.

The rest of the paper is organized as follows. Section 2 briefly reviews previously known optimality conditions for fitting datum planes to set the stage for non-linear datum features that are the main focus this paper. Section 3 describes the circle, cylinder, and sphere fitting problems as optimization problems in general, and summarizes the combinatorial characterizations of the optimality conditions under various optimization criteria. Then Section 4 summarizes a recent work on computational investigations of constrained least-squares fitting of circles, cylinders, and spheres. The main contribution of the paper appears in Section 5, where it is rigorously proved that the optimality conditions for constrained least-squares fitting involve a minimum of two contact points. Section 6 summarizes the results of the paper and offers some directions for future research.

2. Optimality Conditions for Fitting Datum Planes

Consider the problem of specification and verification involving datum planes. Figure 1(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 1(b) illustrates how such a system of primary and secondary datum planes may be established on a manufactured instance of the part.

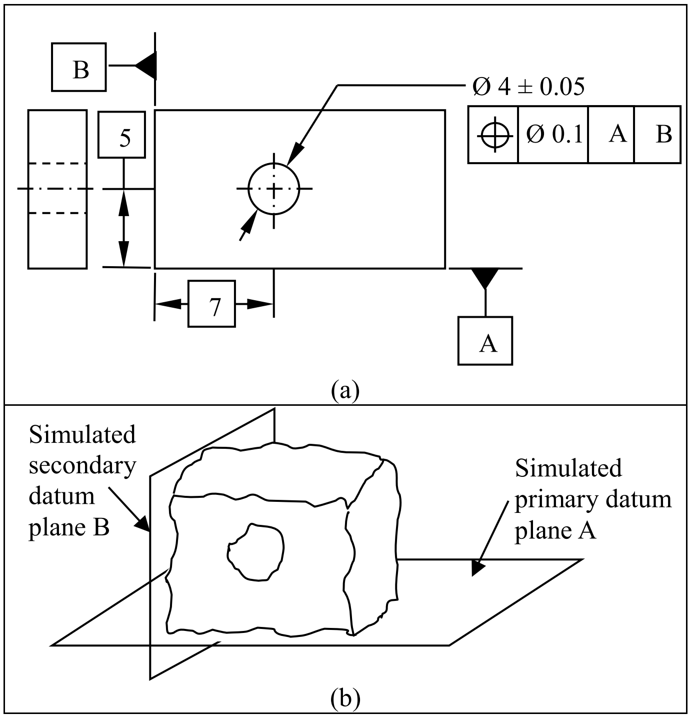

Figure 1. A simple example of (a) specification during design of a part, and (b) establishment of a system of primary and secondary datum planes on a manufactured instance of the part. The secondary datum plane B is required to be perpendicular to the primary datum plane A.

In the current era of digital manufacturing, a manufactured part, such as the one shown in Fig. 1(b), can be scanned by a CMS to generate a large number of discrete points each with three-dimensional Cartesian coordinates. Those points corresponding to the datum feature A can then be processed to fit a primary datum plane A, and similarly for a secondary datum plane B. The optimization problem for performing the plane fitting can be defined in several ways, depending on the choice of the objective function. In all cases, the datum planes are required to lie outside the material of the manufactured part while remaining as close to the part as possible.

Three of the more popular strategies for datum plane fitting have been explored exhaustively [9, 10, 13-22], and are summarized in Table 1. (For the sake of simplicity, a tertiary datum is not shown in Fig.1.) In all these cases, the distances mentioned are the distances of the points measured perpendicular to the plane that is being fitted. In the first strategy, designated as CL_1, the 3-2-1 combinatorial combination of the (minimum number of) contact points on the primary, secondary, and tertiary datum planes is in harmony with the engineering folklore of planar datum system establishment [1, 22]. CL_1 is also known as fitting under constrained L_1-norm [20-22]. In the second strategy, designated as CL_2, the combinatorial nature of the contact condition can vary depending on the configuration of the input data points [9, 10]. CL_2 is the fitting under constrained L_2-norm; it is also known as the constrained total least-squares fitting. In the third strategy designated as SL_2, only one point can be guaranteed to contact each of the datum planes [16, 18]. SL_2 is just the shifted, unconstrained least-squares fitting.

Table 1. Summary of minimum number of contact points for three datum plane fitting strategies.

Designation	Optimization Problem	Primary Datum	Secondary Datum*	Tertiary Datum*
CL_1	Minimize the sum of distances, subject to the constraint that the datum plane lies outside the material.	3	2	1
CL_2	Minimize the sum of squared distances, subject to the constraint that the datum plane lies outside the material.	3, 2, or 1	2 or 1	1
SL_2	Minimize the sum of squared distances, without any constraint. Then shift the datum plane to the outer most point(s) of the material.	1	1	1

*secondary and tertiary datums also obey orientation constraints.

The numbers of contact points shown in Table 1 are the minimum numbers when the input points are in 'general position' – that is, there are no 'degenerate cases' such as four or more input points being co-planar. In those degenerate cases, it is possible that all the fitting strategies may result in contact points that exceed the minimum numbers shown in Table 1. However, even when the input points are in general position, the CL_2 fitting strategy may result in 1, 2, or 3 contact points for the primary datum plane, and 1 or 2 contact points for the secondary datum plane.

While all the results summarized in this section have been already reported exhaustively elsewhere, it is worth recalling a few facts that are relevant to the rest of the paper. The ISO and ASME standards committees, which examined these three (and a few more, including minimum-zone) plane fitting strategies, are in favor of the CL_2 fitting because it combines the mechanical stability of CL_1 and the numerical stability of SL_2. Also, the combinatorial characterizations of the optimality conditions that are presented in Table 1 as the minimum number of contact points have an important computational consequence. New algorithms for computing the fitted planes have been designed to exploit these conditions to generate a compact set of feasible solutions and to find a globally optimal solution [9, 10, 20-22].

3. Fitting Circles, Cylinders, and Spheres

Now consider the problem of fitting circles, cylinders, and/or spheres to sets of input points to establish datums. Figure 2 shows a simple example of how a designer may graphically present the specification of cylindrical datums (as primary and secondary datums) to position a pattern of holes in a part. Figs. 2(b) and 2(c) show how such primary and secondary datums may be established on manufactured instances of the part.

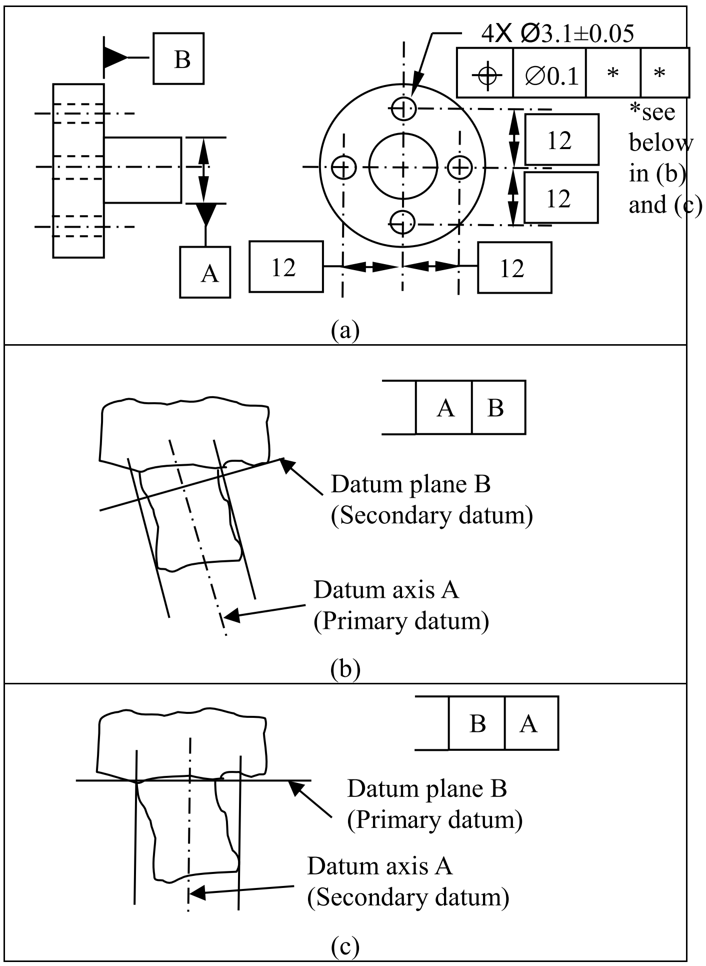

Figure 2. A simple example of (a) specification during design of a part, and (b, c) establishment of systems of primary and secondary datums involving cylindrical features on manufactured instances of the part.

In Fig. 2(b) the primary datum is the axis of a cylinder fitted to a set of points on the manufactured part that correspond to the datum feature A indicated in Fig. 2(a). In this case there is no constraint on the fitting, other than that the cylinder should enclose all the input points; in other words, it is a circumscribing cylinder. Then a secondary planar datum is fitted to the set of points on the manufactured part that correspond to the datum feature B indicated in Fig. 2(a). Here the constraint is added that the secondary datum plane B must be perpendicular to the primary datum axis A.

In Fig. 2(c) the specifications of the primary and secondary datums are reversed. The primary datum is a plane fitted to a set of points on the manufactured part that corresponds to datum feature B indicated in Fig. 2(a), following the procedure outlined in Section 2. Then a cylinder is fitted to the set of points on the manufactured part that corresponds to datum feature A indicated in Fig. 2(a). Here the additional constraint is that the fitted cylindrical axis, which is the secondary datum, must be perpendicular to the datum plane B.

Figure 2 also illustrates an important fact. While fitting a cylinder is important in all cases that involve cylindrical datum features, one may want to fit a circle to a set points on a plane as in Fig. 2(c). In practice, fitting the cylinder in Fig. 2(c) is accomplished by first projecting all the relevant input points on to the primary datum plane B and then fitting a circle to these projected points.

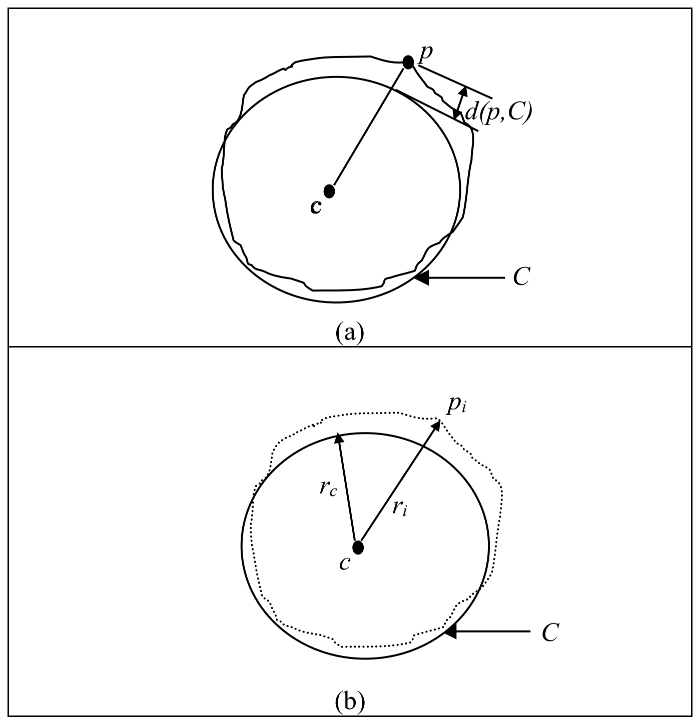

Figure 3. Fitting a circle C with center c to (a) an arbitrary, closed curve and (b) to a set of discrete points sampled on that curve.

With this in mind, this paper treats the circle fitting problem at par with those of fitting cylinders and spheres. In fact, it will be shown that theoretical arguments about circle fitting carry nicely to those about cylinder and sphere fitting as well. Figure 3 illustrates some of the notations and elements involved in fitting circles. In Fig 3(a) a circle C with center c is fitted to an arbitrary, closed curve. The distance of any point p on this curve to the circle C is the perpendicular distance between p and C, shown as $d(p,C)$.

Table 2. Summary of combinatorial characterizations of the optimality conditions for fitting circles, cylinders, and spheres.

	Designation	Optimization Problem	Minimum number of contact points
Circles/Cylinders	MIC/MIC_y	Maximum inscribing circle/cylinder: Maximize the radius so that the circle/cylinder keeps all the input points outside.	3
	MCC/MCC_y	Minimum circumscribing circle/cylinder: Minimize the radius so that the circle/cylinder keeps all the input points inside.	3 or 2
	SL_2C/SL_2C_y	Shifted (offset) least-squares circle/cylinder: Minimize the sum of the squares of the distances to the circle/cylinder. Then change the radius so that the circle/cylinder contacts the outermost point(s).	1
	CL_2IC/CL_2IC_y	Constrained least-squares inscribing circle/cylinder: Minimize the sum of the squares of the distances to the circle/cylinder, subject to the constraint that the circle/cylinder inscribes (keeps outside) all the input points.	2*
	CL_2CC/CL_2CC_y	Constrained least-squares circumscribing circle/cylinder: Minimize the sum of the squares of the distances to the circle/cylinder, subject to the constraint that the circle/cylinder circumscribes (keeps inside) all the input points.	2*
Spheres	MIS_p	Maximum inscribing sphere: Maximize the radius so that the sphere keeps all the input points outside.	4
	MCS_p	Minimum circumscribing sphere: Minimize the radius so that the sphere keeps all the input points inside.	4, 3, or 2
	SL_2S_p	Shifted (offset) least-squares sphere: Minimize the sum of the squares of the distances to the sphere. Then change the radius so that the sphere contacts the outermost point(s).	1
	$CL2IS_p$	Constrained least-squares inscribing sphere: Minimize the sum of the squares of the distances to the sphere, subject to the constraint that the sphere inscribes (keeps outside) all the input points.	2*
	$CL2CS_p$	Constrained least-squares circumscribing sphere: Minimize the sum of the squares of the distances to the sphere, subject to the constraint that the sphere circumscribes (keeps inside) all the input points.	2*

* indicates new results reported in this paper.

In practice, a discrete set of points (also called input points) will be sampled on the arbitrary, closed curve as shown in Fig. 3(b). Then the distance of any input point p_i to a circle C is just the difference between the radii r_i and r_c shown in Fig. 3(b). That is, $d(p_i, C) = | r_i - r_c |$. A circle C can be *fitted* to the set of input points by reducing the distances of the input points to the circle using some appropriate criteria.

The circle fitting problem can then be posed as an optimization problem involving the input points and the parameters of the fitted circle (e.g., the coordinates of its center c and radius r_c). In fact, the same definition and parameters hold for the sphere fitting problem and the cylinder fitting problem (where the 'center' is replaced by the 'axis of the cylinder'). With these preliminaries, the combinatorial characterizations of the optimality conditions for various fitting problems involving circles, cylinders, and spheres are summarized in Table 2. The input to each of these optimization problems is a discrete set of points. The starred items shown in the last column of Table 2 are new, and are proved in Section 5 of this paper. All the other items in the last column of Table 2 are quite well known in research literature and engineering practice [8, 15].

It is worth noting that shifting (offsetting) the radius for SL_2C, SL_2C_y, and SL_2S_p in Table 2 does not change the center or axis, and therefore has no effect on the established datum itself based on unconstrained least-squares fitting. Also worth noting is the absence of minimum-zone fittings of circles, cylinders, and spheres from Table 2. While the minimum-zone fittings are important for the specification and verification of form tolerances such as roundness and cylindricity [1, 3, 15], they are not relevant for datum establishment.

The optimality conditions for MIC and MIS_p shown in Table 2 have inspired the design of some elegant algorithms to find global solutions. These involve nearest-neighbor Voronoi diagrams of the input points and their dual in the form of Delaunay triangles and Delaunay tetrahedra [23]. Since Delaunay triangles/tetrahedra can be derived from convex hulls (albeit in spaces with one dimension higher), clever algorithms can be designed to compute MIC and MIS_p within tractable computational complexity. Some of these algorithms can exploit the increasingly ubiquitous GPUs to speed up the computation.

In the same vein, the optimality conditions for MCC and MCS_p shown in Table 2 can be used to design algorithms to find globally optimal solutions. These algorithms involve furthest-neighbor Voronoi diagrams, which can also be derived from convex hulls in higher dimensions [23].

4. Computational Investigation of Constrained Least-squares Fitting

Buoyed by the initial success of the constrained least-squares fitting of planes to establish planar datums, a computational investigation of the constrained least-squares fitting for circles, cylinders, and spheres was conducted recently [11]. This empirical study used a general-purpose solver for the constrained non-linear optimization problems of CL_2IC, CL_2IC_y, and CL_2IS_p. The results were compared to the solutions of more commonly used fittings MIC, MIC_y, and MIS_p, and were found to exhibit several beneficial properties, such as the one described below.

Consider the case of a set of sampled points on a 'pinched' circle shown dotted in Fig. 4. The solid circle shows a constrained (inscribing) least-squares circle (CL_2IC) fitted to these input points. The MIC problem to the same set of input points will provide two solutions, one centered at c_l on the left half of the pinched circle and the other centered at c_r on the right half of the pinched circle. Since the datums critically depend on the location of these centers (and axes for cylinders), it is quite disconcerting that even a small perturbation of the input points can cause violent changes in the centers of the MIC solutions. The center of the CL_2IC solution, on the other hand, is very stable in the sense that small changes in the input points will lead only to small changes in the location of the datum.

Results from the computational investigations for constrained least-squares fitting of cylinders and spheres reproduced the same beneficial effects found for circles [11]. These computational investigations looked only into the inscribing circles, cylinders, and spheres. However, the results were sufficiently encouraging to prompt a theoretical investigation of the constrained least-squares fitting problem for the non-linear datum elements. A first step towards this theoretical investigation is taken in the next section.

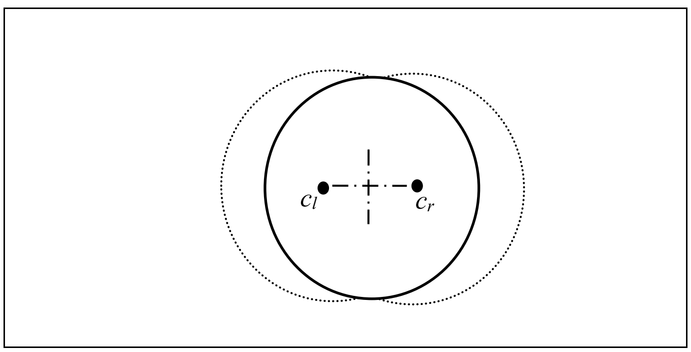

Figure 4. Fitting a constrained (inscribed) least-squares circle to a set of input points.

5. Optimality Conditions for the Constrained Least-squares Fitting

The combinatorial characterizations of the optimality conditions are proved in this section by considering the circles first. It is then shown that the equations, statements, and proofs can be easily extended to the cases of spheres and cylinders. Section 5.1 proves that there should be at least one contact point for optimality using a strategy of growing and shrinking *concentric* circles. The results are extended in Section 5.2, where proofs are provided that there should be at least two contact points for optimality using a strategy of growing and shrinking *co-tangential* circles. Then Section 5.3 presents cases that involve only two contact points at the optimal condition. The possibility of having more than two contact points for optimality is explored in Section 5.4 using a strategy of growing and shrinking *co-chordal* circles.

To set the stage for proving these results, consider a set of n input points and three concentric circles with center c in a plane as shown in Fig. 5. The radial distance of any point p_i from c is indicated as r_i. The radius of any arbitrary circle with an arbitrary center c is indicated as r_c. Also, let C_I be any (small) inscribing circle centered at c and C_C be any (large) circumscribing circle centered at c. Since their radii can be arbitrarily small or large, such circles C_I and C_C can always be found. (Note that an inscribing circle keeps all the input points outside of it, and a circumscribing circle keeps all the input points inside of it.)

With these notations, the optimization problems for constrained least-squares fitting of circles can be posed as

$$\min_{c,r_c} \sum_{i=1}^{n} (r_i - r_c)^2, \quad (1)$$

subject to

$$r_c \leq r_i, \forall i \text{ for } CL_2IC, \quad (2)$$

$$r_c \geq r_i, \forall i \text{ for } CL_2CC. \quad (3)$$

5.1 One point of contact

It can be easily shown that the optimum circle must contact at least one input point. For this, consider the small inscribing circle C_I in Fig. 5 that keeps all the input points outside of it. Such a small circle of arbitrarily small radius can always be found, and let r_c be its variable radius. Grow C_I *concentrically* (keeping its center at c) by increasing its radius r_c gradually while maintaining its inscribing constraints given by Eq. (2). Then every term in the sum of Eq. (1) is reduced, leading to a reduction in the objective function till a first input point is contacted. Stop at that instant because proceeding any further will violate the constraints in Eq. (2). Since this is true for any center c, including the one for the optimal solution, it leads to the following result.

Lemma 1a: CL_2IC must contact at least one input point.

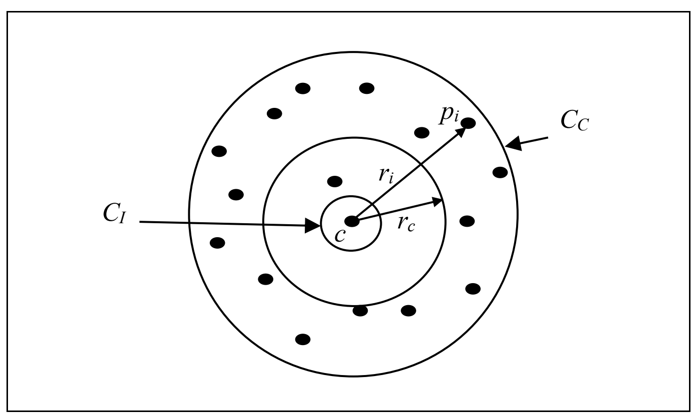

Figure 5. A set of points and concentric circles in a plane.

Next consider the large circumscribing circle C_C in Fig. 5 that keeps all the input points inside of it. Such a large circle with arbitrarily large radius can always be found, and let r_c be its variable radius. Shrink C_C *concentrically* (keeping its center at c) by decreasing its radius r_c gradually while maintaining its circumscribing constraints given by Eq. (3). Then every term in the sum of Eq. (1) is reduced, leading to a reduction in the objective function till a first input point is contacted. Stop at that instant because proceeding any further will violate the constraints in Eq. (3). Since this is true for any center c, including the one for the optimal solution, it leads to the following result.

Lemma 1b: CL_2CC must contact at least one input point.

These results can be generalized to spheres. For this, just replace the word 'circle' with 'sphere' in every exposition so far in this section and consider the input points to be located in space instead of a plane. The following results then emerge.

Lemma 2a: CL_2IS_p must contact at least one input point.

Lemma 2b: CL_2CS_p must contact at least one input point.

It can be seen that these results apply to cylinders as well. Again, replace the word 'circle' with 'cylinder' and consider the input points located in space. Then 'concentric' circles become 'coaxial' cylinders. Also, set the viewpoint along the axis of the cylinder so that all the objects are projected onto a plane perpendicular to the axis. Then the axis itself is projected onto the center c as in Fig. 5, and all the arguments will hold for the cylinders as well leading to the following results.

Lemma 3a: CL_2IC_y must contact at least one input point.

Lemma 3b: CL_2CC_y must contact at least one input point.

5.2 Two points of contact

Can the optimal circles, cylinders, and spheres contact more than one input point? To answer this question, consider the circles again to establish the following result.

Theorem 1a: CL_2IC must contact at least two input points.

Proof: By Lemma 1a, there should be at least one contact point for the inscribing circle. Without any loss of generality, let p_1 be that first contact point from the input set. The proof relies on growing the inscribing circle *co-tangentially* at p_1 while maintaining the inscribing constraints of Eq. (2). Figure 6 shows the construction to aid the proof.

In Fig. 6, C_1 is the inscribing circle with center c_1 that makes the first contact with the input point p_1. C_2 is another inscribing circle with center c_2 that is co-tangential to C_1 at p_1 having T_1 as the co-tangent. C_2 is obtained by growing C_1 gradually and co-tangentially while maintaining the inscribing constraints of Eq. (2). By this assumption, all the other input points are outside of both C_1 and C_2. Let p_i be one such input point.

In Fig. 6, C_3 is a circle centered p_i and tangential to C_2 at the point q_2 with common tangent T_2. The radial line joining centers c_2 and p_i will obviously intersect circle C_2 at q_2. Let q_1 be the intersection point of the radial line between centers c_1 and p_i. It is clear from this construction that point q_1 lies strictly outside of the circle C_3 because circle C_1 is completely contained within circle C_2. Therefore, $p_i q_2 < p_i q_1$. (A special case arises when p_i lies directly below p_1 on the vertical line through c_2, c_1, and p_1 in Fig. 6. Then q_1, q_2, and p_1 coincide, resulting in $p_i q_2 = p_i q_1$.)

Note that $p_i q_2$ is the distance of the point p_i to the circle C_2 and $p_i q_1$ is the distance of the p_i to the circle C_1. Therefore, each term in the sum of the objective function in Eq. (1) is strictly reduced as C_1 is grown co-tangentially to C_2. (In the special case of a point p_i discussed above, where $p_i q_2 = p_i q_1$, that term remains unchanged.) Such reduction in the objective function continues till C_2 contacts a second input point. Stop at that instant, because proceeding any further will violate the inscribing constraints in Eq. (2). This proves the assertion that CL_2IC must contact at least two input points. ■

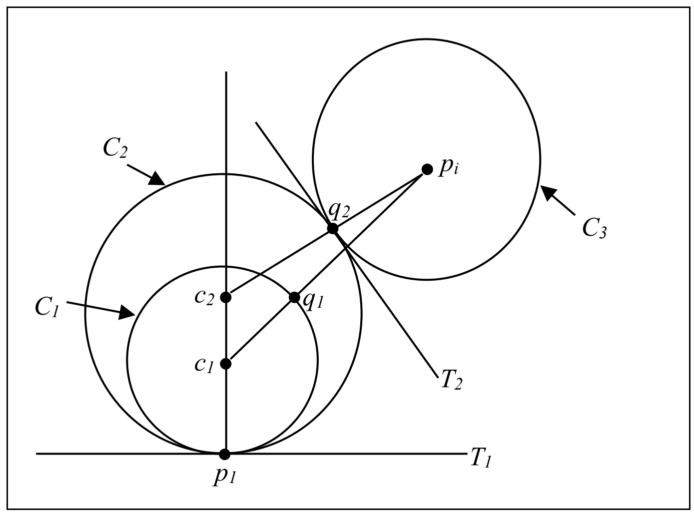

Figure 6. Construction for the proof of Theorem 1a.

Pursuing the same strategy, the following assertion can be proved.

Theorem 1b: CL_2CC must contact at least two input points.

Proof: By Lemma 1b, there should be at least one contact point for the circumscribing circle. Without any loss of generality, let p_1 be that first contact point from the input set. The proof relies on shrinking the circumscribing circle *co-tangentially* at p_1 while maintaining the circumscribing constraints of Eq. (3). Figure 7 shows the construction to aid the proof.

In Fig. 7, C_1 is the circumscribing circle with center c_1 that makes the first contact with the input point p_1. C_2 is another circumscribing circle with center c_2 that is co-tangential to C_1 at p_1 having T_1 as the co-tangent. C_2 is obtained by shrinking C_1 gradually and co-tangentially while maintaining the circumscribing constraints of Eq. (3). By this assumption, all the other input points are inside of both C_1 and C_2. Let p_i be one such input point.

In Fig. 7, C_3 is a circle centered p_i and tangential to C_2 at the point q_2 with common tangent T_2. The radial line joining centers c_2 and p_i will obviously intersect circle C_2 at q_2. Let q_1 be the intersection point on C_1 of the radial line joining centers c_1 and p_i. It is clear from this construction that point q_1 lies outside of the circle C_3 because circle C_1 lies completely outside of circle C_2, which lies outside of circle C_3. Therefore, $p_i q_2 < p_i q_1$. (A special case arises when p_i coincides with c_2 in Fig. 7. Then q_1, q_2, and p_1 coincide, resulting in $p_i q_2 = p_i q_1$.)

Note that $p_i q_2$ is the distance of the point p_i to the circle C_2 and $p_i q_1$ is the distance of the p_i to the circle C_1. Therefore, each term in the sum of the objective function in Eq. (1) is strictly reduced as C_1 is shrunk co-tangentially to C_2. (In the special case discussed above, where $p_i q_2 = p_i q_1$, that term remains unchanged.) Such reduction in the objective function continues till C_2 contacts a second input point. Stop at that instant, because proceeding any further will violate the circumscribing constraints in Eq. (3). This proves the assertion that CL_2CC must contact at least two input points. ■

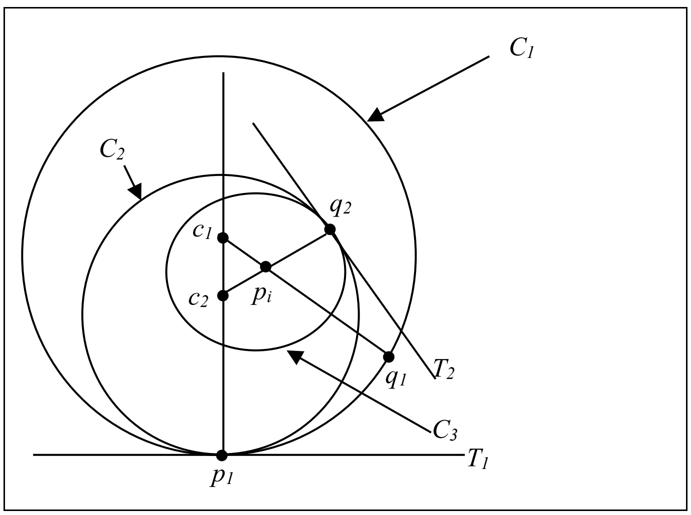

Figure 7. Construction for the proof of Theorem 1b.

By replacing the word ‘circle’ with ‘sphere’ in Eqs. (1), (2) and (3), and in the proofs and statements of Theorems 1a and 1b, the following results are obtained.

Theorem 2a: CL_2IS_p must contact at least two input points.

Theorem 2b: CL_2CS_p must contact at least two input points.

Obtaining similar results for cylinders requires only slightly more work. The words ‘circle’ and ‘center’ should be replaced by ‘cylinder’ and ‘axis,’ respectively, in Eqs. (1), (2), and (3). Also, the same replacement should be applied in the proofs and statements of Theorems 1a and 1b, while the constructions in the Figs. 6 and 7 should be viewed as projections on a plane perpendicular to the axes of the cylinders. (Note that the axes of these cylinders are parallel, and the term ‘coaxial’ replaces ‘concentric’; the term ‘co-tangential’ remains the same.) After making these replacements, it is possible to grow and shrink co-tangential cylinders in the proofs of Theorems 1a and 1b even if the first contact, shown as p_1 in Figs. 6 and 7, is made by cylinder C_1 at two or more points that lie on the same generator (also known as generatrix) line of C_1. Then, the following results are obtained.

Theorem 3a: CL_2IC_y must contact at least two input points that do not lie on the same generator line.

Theorem 3b: CL_2CC_y must contact at least two input points that do not lie on the same generator line.

5.3 Cases with only two points of contact

Can the minimum number of contact points be more than two for any of the geometries considered? No, as this subsection will show that there are cases that have only two contact points at the optimal solution. The first case involves circles in a plane, and this case is then expanded to space to produce a case for spheres. Finally, a case is constructed for cylinders in space.

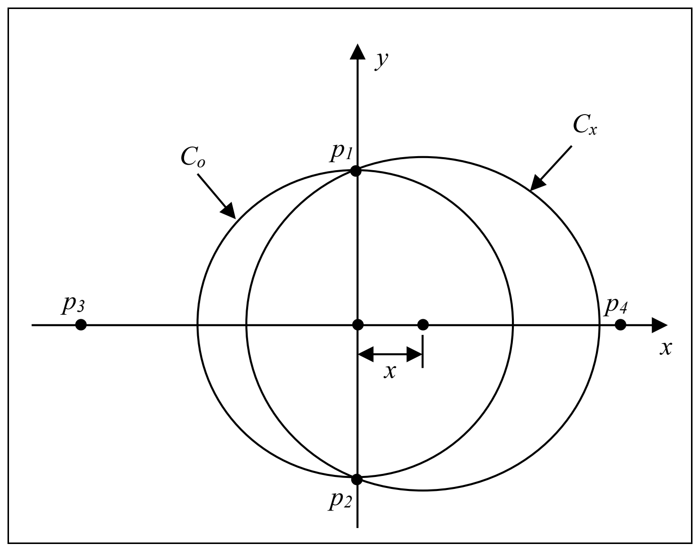

Figure 8. Construction of an example case with only two contact points for an optimal circle.

To start with circles, consider four input points in a plane as shown in Fig. 8, with the following coordinates:

$$p_1: (0, 1), p_2: (0, -1), p_3: (-1 - \Delta, 0), \text{and } p_4: (1 + \Delta, 0).$$

Here Δ can be any arbitrarily small positive or negative number. Initially, Δ will be assigned an arbitrarily small positive value for producing the case of an inscribing circle. In this case, one can view the four input points as representing a ‘pinched’ circle. Let C_o be the unit circle centered at the origin with unit radius, contacting points p_1 and p_2. With a small positive value for Δ, C_o is an inscribing circle for the four input points. It can be proved that C_o is also the inscribing circle that minimizes the sum of the squares of its distances from the input points; that is, C_o is the CL_2IC for these four input points. To prove this assertion, consider another inscribing circle C_x centered at $(x, 0)$ for a very small value of x, and contacting points p_1 and p_2 as shown in Fig. 8.

From Theorem 1a, C_x can aspire to be an optimal solution because it is an inscribing circle maintaining contact with at least the two input points p_1 and p_2. The objective function (the sum of the squared distances of the four input points to a circle) for C_x in terms of x, and the first and second derivatives of the objective function can be easily seen to be the following:

$$f(x) = \left\{(1+\Delta) - x - \sqrt{x^2+1}\right\}^2 + \left\{(1+\Delta) + x - \sqrt{x^2+1}\right\}^2,$$

$$f'(x) = 8x - 4(1+\Delta)x(x^2+1)^{-1/2} \text{ , and}$$

$$f''(x) = 8 - 4(1+\Delta)\left\{(x^2+1)^{-1/2} - x^2(x^2+1)^{-3/2}\right\}.$$

Therefore at $x = 0$,

$$f(0) = 2\Delta^2,\ f'(0) = 0, \text{ and } f''(0) = 4(1-\Delta).$$

It means that $f''(0) > 0,\ \forall\ \Delta < 1$ at $x = 0$. This proves that C_0 achieves a local minimum for the objective function [12]. Since there are only four input points, C_0 also achieves the global minimum and is the CL_2IC that contacts only two input points. (Note that the two circles that contact three points without contacting both p_1 and p_2 (i.e., are not centered at $(x, 0)$) are not inscribing circles.)

To present a circumscribing case with only two contact points, consider Fig. 8 again and the related discussions with Theorem 2b in mind, but now with an arbitrarily small negative value for Δ. The expressions for the objective function and its first and second derivatives will still remain the same as above, leading to the result that $f'(0) = 0$ and $f''(0) > 0$ at $x = 0$. So C_0 achieves a local minimum. Since there are only four input points, C_0 also achieves the global minimum, and is the CL_2CC that contacts only two input points.

Proving the existence of optimal spheres with only two contact points requires a different construction. For this, consider six input points in space as shown in Fig. 9, with the following coordinates:

$$p_1: (0,0,1), p_2: (0,0,-1), p_3: (1+\Delta,0,0), p_4: (-1-\Delta,0,0),$$

$$p_5: (0,1+\Delta,0), \text{ and } p_6: (0,-1-\Delta,0).$$

For a small positive value of Δ, these six points may be viewed as representing a sphere ‘pinched’ at the poles. Let S_0 be a unit sphere centered at the origin with unit radius that contacts input points p_1 and p_2. The goal is to show that S_0 is the CL_2IS_p for a small positive value of Δ, and S_0 is the CL_2CS_p for a small negative value of Δ.

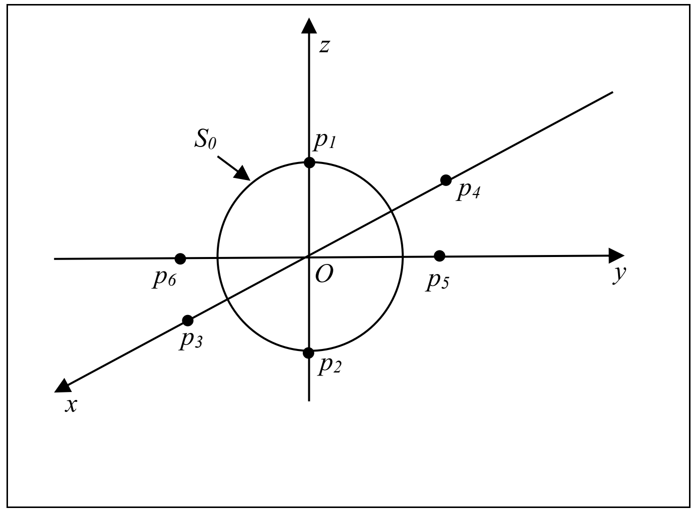

Figure 9. Construction of an example case with only two contact points for an optimal sphere.

For this, consider another sphere S_x centered at $(x, y, 0)$ for small values of x and y. Also assume that S_x contacts only p_1 and p_2 from the input set of points. An objective function $F(x, y)$ for the sum of the squared distances of the six input points to S_x, and the first and second partial derivatives of $F(x, y)$ can be derived in closed form. At $x = y = 0$ these closed form expressions yield the following values for the objective function, its gradient vector, and Hessian matrix:

$$F(0,0) = 4\Delta^2,$$

$$\nabla F(0,0) = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}, \text{ and}$$

$$H(0,0) = \begin{bmatrix} 8(1-\Delta) - \frac{4}{(1+\Delta)} & 0 \\ 0 & 8(1-\Delta) - \frac{4}{(1+\Delta)} \end{bmatrix}.$$

So the gradient vanishes at the origin. The eigenvalues of the Hessian matrix H are both close to 4 at the origin for small positive or negative values of Δ. Therefore, H is positive definite for any small positive or negative value of Δ. These conditions are sufficient to show that S_0 achieves a local minimum for the objective function [12]. Since there are only six input points, S_0 also achieves the global minimum contacting only two input points. Thus the following fact is established: when Δ is a small positive number, S_0 is the CL_2IS_p; similarly, when Δ is a small negative number, S_0 is the CL_2CS_p.

The examples for the case of circles also motivate similar examples for the case of cylinders. In addition to the four input points (for circles) in the $z = 0$ plane as shown in Fig. 8, consider additional eight points with coordinates given by

$$p_5: (0, 1+\Delta, -1),\ p_6: (0, -1-\Delta, -1),$$
$$p_7: (-1-\Delta, 0, -1),\ p_8: (1+\Delta, 0, -1),$$
$$p_9: (0, 1+\Delta, 1),\ p_{10}: (0, -1-\Delta, 1),$$
$$p_{11}: (-1-\Delta, 0, 1), \text{and } p_{12}: (1+\Delta, 0, 1).$$

One can view these twelve points representing a 'pinched' cylinder for a small positive value of Δ. Reasoning with these twelve input points in space in a manner very similar to the cases of the circles discussed above, it can be shown that there are the cases where CL_2IC_y and CL_2CC_y have only two contact points p_1 and p_2. For this, let a cylinder be represented by four parameters (x, y, a, b) using the point $(x, y, 0)$ where the axis of the cylinder pierces the xy-plane, and the direction cosines of the axis as $\left(a, b, \sqrt{1-a^2-b^2}\right)$. (The radius is taken to be the largest/smallest possible while still inscribing/circumscribing the points.) Then the values for the objective function $F(x, y, a, b)$, its gradient, and Hessian matrix entries can be obtained using a symbolic derivative calculations software as

$$F(0,0,0,0) = 10\Delta^2,$$

$$\nabla F(0,0,0,0) = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}, \text{and}$$

$$H(0,0,0,0) = \begin{bmatrix} 12 + O(\Delta) & 0 & 0 & 0 \\ 0 & \text{large positive} & 0 & 0 \\ 0 & 0 & 8 + O(\Delta) & 0 \\ 0 & 0 & 0 & 8 + O(\Delta) \end{bmatrix}.$$

The phrase 'large positive' in the Hessian matrix indicates that the function has a convex corner (much like $y = |x|$ has at $x = 0$). These results of a zero gradient and a positive definite Hessian (for a small positive or negative value of Δ) indicate that the solution is a local minimum, and the simplicity of the construction of the example is sufficient to show that this local minimum is also the global minimum.

5.4 Possibility of more than two points of contact

The existence of cases with only two input points of contact does not preclude the possibility that there can be more points of contact when the input points are still in general position. For example, the construction of Fig. 8 suggests a procedure for a *co-chordal* traversal to seek a better solution.

Consider the case shown in Fig. 10, where p_1 and p_2 are two contact points from the input set for an inscribing circle C_1. A one-parameter family of co-chordal circles, all contacting p_1 and p_2 (hence having $\overline{p_1 p_2}$ as the common chord) and yet inscribing all the input points, can be considered for possible reduction in the objective function. Let C_2 be one member in such a family of co-chordal circles.

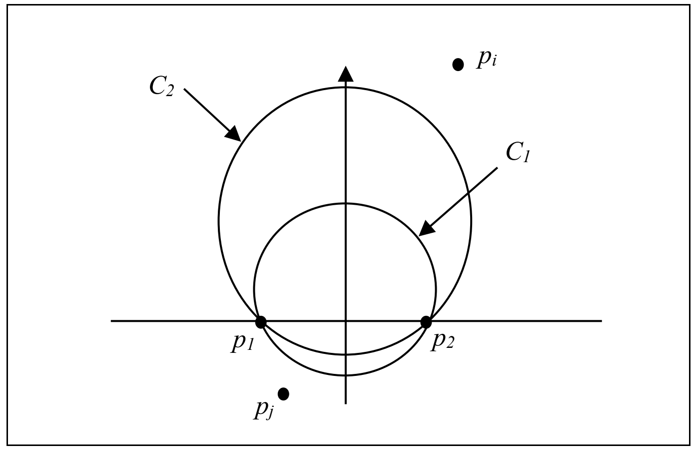

Figure 10. A one-parameter family of inscribing circles having a common chord.

It can be easily seen that for points such as p_i in Fig. 10, C_2 will be closer than C_1 thus contributing a reducing term to the objective function in Eq. (1). But for points such as p_j in Fig. 10, C_2 will be farther away than C_1 leading to an increase in the objective function. So it is possible that when all the terms are summed up in Eq. (1), there might be a net reduction in the objective function for C_2. This reduction might continue as C_2 is grown in this co-chordal traversal process till C_2 contacts a third input point.

Such optimal conditions involving three contact points for CL_2IC have been observed in computational investigations [11], leading to the pleasing result that MIC and CL_2IC may turn out to be the same in these cases. Similar results have been observed for cylinders and spheres. These empirical observations in numerical trials seem to suggest that in many cases, where the input points are in general position, the constrained least-squares solutions may coincide with the maximum inscribing and minimum circumscribing solutions. And where small perturbations in the input points can cause instabilities in the maximum inscribing and minimum circumscribing solutions, an example of which is shown in Fig. 4, the constrained least-squares approach can lead to a stable solution.

6. Summary and Concluding Remarks

This paper addressed the problem of establishing datums using constrained least-squares fitting to input points sampled on non-linear elements such as circles, cylinders, and spheres. Combinatorial characterizations of the optimality conditions for these constrained non-linear optimization problems were provided with rigorous proofs. These results have been verified by computational investigations using research software [11].

These necessary conditions, in the form of the minimum number of points of contact, not only enable a check of optimal solution candidates but also inspire design of new algorithms to find global solutions. For circles and spheres, the fact that the optimal solution must contact at least two input points implies that the centers of such circles and spheres must lie on the edges or surfaces of the Voronoi diagram of the input points. This leads to a dramatic reduction the search space for globally optimal solution. Algorithms that exploit this fact may also employ GPUs to speed up the search.

Further research is necessary to investigate such algorithmic issues. Another important open research problem is the extension of the constrained least-squares fitting to a continuous set of points, such as the one shown in Fig. 3(a). Robust implementations of constrained least-squares fitting algorithms in commercial software will be important for their acceptance by industry. Preliminary response from leading CMS vendors indicates that they are indeed beginning to implement and test their software using the results of this paper.

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 1101:2012, Geometrical product specifications (GPS) – Geometrical tolerancing – Tolerances of form, orientation, location and run-out, International Organization for Standardization, Geneva, 2012.

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

[5] Nielsen, H.S., The ISO Geometrical Product Specifications Handbook – Find your way in GPS, International Organization for Standardization, Geneva, 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. 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.

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

[11] 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.

[12] Gill, P.E., Murray, W., and Wright, M.H., Practical Optimization, Emerald Group Publishing Ltd., 1982.

[13] Hopp, T.H., The mathematics of datums, ASPE Newsletter, The American Society of Precision Engineers, Raleigh, NC, 1990.

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

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

[16] 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.

[17] Srinivasan, V., Reflections on the role of science in the evolution of dimensioning and tolerancing standards, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 227(1), pp. 3-11, 2013.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] O'Rourke, J., Computational Geometry in C, 2nd Edition, Cambridge University Press, 1998.