IMECE2016-67753

COMPUTATIONAL INVESTIGATIONS FOR A NEW, CONSTRAINED LEAST-SQUARES DATUM DEFINITION FOR CIRCLES, CYLINDERS, AND SPHERES

Craig M Shakarji Physical Measurement Laboratory National Institute of Standards and Technology Gaithersburg, Maryland 20899 craig.shakarji@nist.gov Vijay Srinivasan Engineering Laboratory National Institute of Standards and Technology Gaithersburg, Maryland 20899 vijay.srinivasan@nist.gov

ABSTRACT

For engineering drawings and CAD definitions, the problem of a suitable datum definition for datum features of circles, spheres, and cylinders has been sought by standards writers over decades. The maximum-inscribed and minimum-circumscribed definitions that have often been used have known problems relating to stability in many common, industrial cases. Examples of these problem cases include cylindrical datum features having an hourglass shape, barrel shape, or the shape of a tapered shaft and circular or spherical datum features that are dimpled. For these problematic cases, many resort to using a least-squares fit whose diameter is scaled to be just inside (or just outside) the datum feature. However, we show this shifted least-squares solution has its own drawbacks.

This paper investigates a new datum definition based on a constrained least-squares criterion. The use of this definition for datum planes has already elegantly solved the problem of providing a full contact solution when that solution is stable, while providing a balanced, stable solution in the case of rocker conditions. With that success as motivation, we now investigate using this definition for circles, spheres, and cylinders.

We demonstrate that the constrained least-squares is an excellent choice for several known problematic cases. This datum definition maintains stability in cases where the maximum-inscribed fits are not unique and thus not stable. Yet they also maintain close adherence to the maximum-inscribed solution when such solutions are stable. We also show that the constrained least-squares solution has clear benefits over the shifted least squares solution. This is the first computational investigation into the behavior of the constrained least-squares as a possible datum definition for these features. While not being fully comprehensive, these initial findings indicate that the constrained least-squares appears to be a safe and advantageous datum definition choice and provide substantial optimism that results in future investigated cases will be pleasing as well.

1. BACKGROUND AND INTRODUCTION

In the world of Geometric Dimensioning and Tolerancing (GD&T), datums are used extensively to locate and orient tolerance zones [1-7]. Given an (imperfect) datum feature on a real workpiece, a datum is a mathematically perfect geometry associated with that datum feature. An example of a datum plane associated with a datum feature is shown in Fig. 1. Some commonly used datums are planes, circles, cylinders, and spheres.



Fig. 1. Deriving a datum plane from a datum feature.

A decades-old problem in the field is creating rigorous definitions that determine which datums should be associated

with given datum features [8]. Apart from standardization, there are several different yet reasonable approaches by which a datum can be established from a datum feature. For example, given a planar feature, one could associate a least-squares planar datum (which in the case of Fig. 1 would pass through the material), or a least-squares plane shifted to be just outside the material (which in the case of Fig. 1 would contact the datum feature at one point), or a one-sided, full-contact fit, which corresponds to the datum shown in Fig. 1. For planes, the full-contact fit is described primarily in [9] (but one can also see [10-11]).

But a datum definition that is perhaps the most sought has also historically been most elusive to define mathematically. In fact, it has been difficult to define just in words, but is somewhat captured by the following properties:

- 1) The datum should lie outside the material
- The datum should contact the datum feature fully (unless that datum is not stable)
- 3) The datum should be stable in the sense that two datum features that are nearly the same should produce datums that are nearly the same.

It is the tension between (2) and (3) that produce the difficulty. In the case of datum planes, the ASME Y14.5 Standard [4-5] defines the datum plane as the full-contacting plane (as a planar surface resting on a surface plate) unless it is a "rocker" condition, which is handled separately. Another example of wording seeking to balance (2) and (3) is that some editions of ISO 5459 [6] indicate that a cylindrical datum given a datum feature is the maximum-inscribed (or minimum circumscribed) cylinder unless that is unstable, in which case the mobility is to be equalized to the surface. An example case is shown in Fig. 2.



Fig. 2. The maximum-inscribed cylinder (2D cross section shown) is not stable for the hourglass shape, since both inscribed cylinders shown have nearly the same diameter.

The difficulty of defining the desired datum is seen in the fact that when a threshold is given to define an unstable case, property (3) ends up being violated, because two similar datum features could be produced, each on other sides of the threshold, producing very different datums.

In the case of planes, a mathematical datum definition has recently been found that pleasingly solves the problem [12-13]. It is based on a constrained least-squares datum definition. This is different from a shifted least-squares solution and has the property that it makes full three-point contact with the surface when such full contact is not a rocker condition. When a rocker condition exists, the solution balances the rock, contacting only two points or one point as needed. In the case of Fig. 1, for example, the datum definition would provide the full contact datum shown. However, if the datum feature (in 3D) resembled and inverted pyramid, the datum feature would balance at only one point, the apex. If the datum feature were a concave, rectangular shape, meaning only the four corners are points of possible contact, the datum would make two diagonal points of contact and balance between the other two (like a four legged chair that rocks on two legs). In summary, a planar datum definition has been recently found that elegantly solves the elusive, decades-old datum problem.

The success of the planar problem has naturally led to questions about applying such a definition to other geometries. In this paper we investigate—largely through computational solutions—the behavior of applying a constrained least-squares definition to circles, spheres, and cylinders¹.

2. PROBLEMS WITH THE FULL-CONTACT SOLUTION

For the planar case, the full contact solution has been rigorously defined by using a constrained L_1 association [9]. The definition produces a pleasing result in cases like Fig. 1, but is unstable and undesirable for the cases of a "V", as shown in Fig. 3.



Fig. 3. The full-contact (L_1) planar datum will not balance with one point of contact as shown, but will coincide with the longer edge or the other of the datum feature, making it unstable.

For the cases of circles, spheres, and cylinders, the full contact solution corresponds to the maximum-inscribed or minimum-circumscribed definitions. For example, in the case of a circle, given a 2D set of data points, it is readily seen that the maximum-inscribed circle must contact at least three points. (If it did not, and only contacted two points, then another circle must

¹ The ISO 5459 revision Draft International Standard (DIS) has included the constrained least-squares definition as its default datum definition for all geometries.

This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States. Approved for public release; distribution is unlimited.

exist having greater diameter that is still inscribed, which can be seen by shifting the two-contact point circle's center perpendicular to the line connecting the two contact points.) In the continuous case, the full contact solution can contact two points—consider the maximum-inscribed circle of an ellipse. The same full contact applies to the cases of spheres and cylinders. However, just as in the planar case, these full contact solutions are unstable. Take, for example, the case of fitting a maximum-inscribed circle to the shape shown in Fig. 4.



Fig. 4. A shape having two maximum-inscribed circles, one centered at p and one at q.

The case of Fig. 4 is unstable in that the slightest change of shape can alter the center of the solution between the two centers shown. This has similarities to the planar datum case, where the full-contact solution varied in orientation between the two edges with slight changes of the datum feature (Fig. 3).

This problem with maximum-inscribed circles occurs more often in practice than a reader might realize. First, it can arise as a result of a simple combination of a two- and four-lobed form. For instance, the polar equation

 $r(\theta) = 10 + (0.02)\sin(4\theta) + (0.01)\sin(2\theta)$ has two maximum-inscribed circles.

For another case, imagine a workpiece where the primary datum feature is a plane and the secondary datum feature is a cylinder nominally perpendicular to the plane. Then the secondary datum problem is actually a circular datum problem arising from projecting the cylinder into the primary datum plane. If the datum feature were sampled poorly, the following case would arise (shown in Fig. 5) where two centers of maximum inscribed circles exist. Interestingly, neither center is the one desired.



Fig. 5. A practical, unstable case showing two center locations possible. The misalignment of the cylinder projected into the plane is exaggerated for clarity.

Another example where the case of Fig. 4 can be seen is with discrete sampling. Imagine sampling a perfect circle except that two points on opposite sides of the center are slightly closer to the center than the others. This dimpled effect also creates an instability in the solution similar to what is depicted in Fig. 4. This effect can also occur with spheres.

The issue for cylinders is even more problematic and more common. The reason is that the instability shown in Fig. 2 did not arise as a result of non-uniqueness. Mathematically, the unique maximum-inscribed fit is shown on the left. But the inscribed cylinder on the right has a significantly different orientation with almost no change in diameter, causing computational instability. This instability has plagued coordinate metrology software as documented in [14] and it continues to the present time.

Other cases of instability are seen in the minimumcircumscribed cylinder of a barrel shaped datum feature (Fig. 6) and either the minimum-circumscribed or maximum-inscribed cylinders of a taper-shaped datum feature (Fig. 7 shows the inscribed case).



Fig. 6. The minimum-circumscribed cylinder (2D cross section shown) is not stable for the barrel shape, since both

circumscribed cylinders shown have nearly the same diameter.



Fig. 7. The maximum-inscribed cylinder (2D cross section shown) is not stable for the taper shape, since rotating the axis as shown changes the diameter very little.

3. THE CONSTRAINED LEAST-SQUARES DEFINITION FOR THE PLANAR DATUM CASE

The problems that arise from the full contact solution were elegantly solved in [12, 13] by using the definition of a constrained least squares (called a constrained L_2 association in [12]). Figure 8 shows two typical cases where, on the left, one would seek to balance the rocking condition, and on the right, one would seek for the datum plane to be stably flush with the edge of the datum feature. This is what the constrained least-squares solution does automatically.



Fig. 8. Two typical cases of datum features with the associated constrained L_2 datums shown. The balanced rocking case is on the left and the stable, flush case is on the right.

For the rocker condition pictured on the left side of Fig. 8, if the line segment on the right were made longer, the constrained L_2 datum plane would roll to the right smoothly. For the stable case pictured on the right side of Fig. 8, if the line segment on the right were made somewhat longer, the L_2 constrained datum

plane would <u>not</u> move from its stable state. It would remain flush with the edge of the datum feature until the line segment on the right grew long enough to make a rocker condition, at which point the L_2 constrained datum would smoothly begin to roll to the right to balance the rocker.

In contrast, the shifted least-squares solution would achieve a flush mating with the datum feature (as pictured on the right of Fig. 8) for only an instant. That is, as the line segment on the right began to be extended, there would only be one length that resulted in a flush mating. This contrast shows the fascinating feature of the constrained L_2 , which stays flush with the datum feature—even while the line segment extends—until it reaches such a length that a rocking condition exists, like shown in Fig. 9.



Fig. 9. The line segment on the right is long enough for the constrained L_2 datum to treat it as a rocking condition and separate from the flush contact it had in the right hand picture of Fig. 8.

The constrained least-squares definition in this 2D case automatically chose between one and two points of contact in keeping the three properties desired as listed in Section 1. Similar behavior occurs in the 3D case where the datum plane makes one, two, or three points of contact with the datum feature, automatically and smoothly transitioning between these cases for varying datum features.

Since the instabilities that exist with the full contact solutions for the circle, sphere, and cylinder cases bear some similarity to the planar case, the hope is that the constrained least-squares criterion would also solve the problem of meeting all three desired properties for these cases as well.

4. THE CONSTRAINED LEAST-SQUARES DEFINITION FOR CIRCULAR DATUMS

Given an approximately circular curve, *C*, and a circle centered at (x_0, y_0) with radius, *r*, then the least-squares objective function is given by $\int_C d^2(\mathbf{p}) dc$, where $d(\mathbf{p})$ represents the distance from point \mathbf{p} (on *C*) to the circle (Fig. 10). With discrete sampling of points, we approximate the objective function

$$\int_{C} d^{2}(\boldsymbol{p}) dc \approx \sum_{i=1}^{N} d^{2}(\boldsymbol{p}_{i}) \Delta L_{i},$$

where p_i are the *N* sampled points, one in each subdivision of length ΔL_i . If *d* represents the signed distance to the circle (positive indicating that the point is outside the circle) then the constraint can be written—for the maximum-inscribed case—as $d(p) \ge 0$, for all points p.

The objective functions for the minimum-circumscribed case, as well as for cases of spheres and cylinders are similar.

Formulas for the distance calculations can be found in [14]. In this paper, our sampling was always uniform to allow the removal of the weighting factor ΔL_i from the objective function.



Fig. 10. Defining the objective function for fitting a circle to a curve, *C*.

We developed maximum-inscribed and constrained leastsquares algorithms in order to investigate the behavior of the constrained least-squares in contrast to the maximum-inscribed fits. We begin with the "peanut shaped" case pictured in Fig. 11, where the maximum-inscribed fit skips from side to side depending on the slightest random perturbation. One such solution is shown on the left side of Fig. 11. In contrast, the constrained least-squares solution remained near the center as shown on the right side of Fig.11.



Fig. 11. The same set of data points fit with a maximum-inscribed fit (left) and a constrained least-squares fit (right).

To give a quantitative feel for the behavior of these two fits, we used the dimpled problem mentioned in Section 2, for which the fits behave similarly to that shown in Fig. 11. We generated 40 uniformly spaced data points on a 20 mm diameter circle centered at the origin and allowed those points to vary in the radial direction uniformly randomly up to $\pm 10 \ \mu$ m. For each data set we also added two dimple points at (0.0, 9.97) and (0.0, -9.97) (i.e., 30 μ m dimples) and fit the resulting data sets with both the

maximum-inscribed and constrained least-squares functions. This procedure was repeated 500 times and the histograms for the x-coordinate of both fits are shown in Fig. 12.



Fig. 12. Histograms showing the *x*-coordinate of the center location of fit circles (top) for the constrained least-squares fit and (bottom) for the maximum-inscribed fit. Units are mm, but the horizontal scale differs by a factor of 10 between the histograms.

As expected, the maximum-inscribed fit has a bimodal distribution for the *x*-coordinate while the constrained least-squares does not. But quantitatively, the standard deviation of the *x*-coordinate for the constrained least-squares is 1.3 μ m vs. 25 μ m for the maximum-inscribed fit, making them differ by a factor of nearly 20. The important lesson is not the precise numbers themselves, for different contrived cases would yield different numbers, except to note that the constrained least-squares is simply much more stable (by more than an order magnitude in the case shown here).

But one might then ask what is sacrificed with regard to the diameter of the fits. Interestingly, the mean diameter for the maximum-inscribed fit was 19.94006 mm vs. 19.94000 mm for the constrained least-squares. This means that the remarkable increase in stability came at the diameter "cost" of only 60 nanometers—an amount certainly negligible in light of the size of form deviations in this example case.

Before moving on to other geometries, we will make one observation about the constrained least-squares vs. the oftenused shifted least squares. For datum features having symmetric deviations (like shown in Fig. 11) the shifted least squares may work reasonably well. But for nonsymmetrical cases the shifted least squares can significantly underperform the constrained least squares.

Take, for example, eight uniformly spaced points on the 20 mm circle we have been considering. Assume all the points lie exactly on the circle except for the topmost point, which has coordinates (0.0, 10.01), meaning that for some reason this point

is off the circle by 10 μ m. Clearly this point will not affect the maximum-inscribed circle, which will be centered at the origin and have diameter 20 mm. However, the shifted least-squares fit has its center at (0, 0.0025) and a diameter of 19.9965—thus a 2.5 μ m shift in the center and a 3.5 μ m decrease in the diameter. By contrast, the constrained least-squares fit is centered at (0, 0.00037) with a diameter of 19.9995, making only a 0.37 μ m shift in center and a 0.5 μ m decrease in diameter. These differences from the maximum-inscribed fit are lower by a factor of about seven from the shifted least squares fit. While shifted least-squares is often used in practice due to its stability in some cases, it seems evident thus far that excellent stability can be achieved without such departure from the full-contact solution.

Lastly we note that given a set of data points, the shifted least-squares fit generally contacts only one point. This alone implies an inefficiency. For this inscribed circle case, the full contact (i.e., maximum-inscribed) solution always contacts three points, even at the cost of sacrificing stability. The constrained least-squares contacts three points or sometimes two, depending on the nature of the form deviations, allowing it to maintain stability.

5. THE CONSTRAINED LEAST-SQUARES DEFINITION FOR SPHERICAL DATUMS

The case of spheres is very much like that of circles, extended by one dimension. We do not show numerical results, which are similar to those shown in Fig. 12. But we do take time to note some differences that the extra dimension brings. First, the dimpled problem considered in the case of circles also exists in the case of spheres, but with some extra possibilities. When two dimples are present, the maximum inscribed sphere's center shifts away from the nominal center (similar to Fig. 11) but instead of shifting only left or right as in Fig. 11, it can shift at any angle within the plane perpendicular to the line formed by the dimples. It is also possible to have three dimples (say, on a great circle) that force the full contact solution to one of two sides.

Given a set of data points for this inscribed sphere case, the full contact (i.e., maximum-inscribed) solution always contacts four points, even at the cost of sacrificing stability. The constrained least-squares contacts four points or sometimes three or sometimes even two, depending on the nature of the form deviations, allowing it to maintain stability. And similar to the 2D case, stability is gained with minimal cost to the diameter of the constrained least-squares fit.

The drawbacks that existed with the shifted least-squares on circles are present with similar magnitude in the case of spheres. Again, the constrained least-squares solution significantly mitigates these effects. Similar to the eight data point case with circles, a set of 14 points was created on a 20 mm sphere: six points were located at ± 10 mm on the axes and eight were located at $(10 \text{ mm}) \left(\pm \frac{\sqrt{3}}{3}, \pm \frac{\sqrt{3}}{3}, \pm \frac{\sqrt{3}}{3} \right)$. The exception was that the "north pole" point was shifted up 10 µm, making that point (0, 0, 10.01).

The maximum inscribed fit was unaffected by the shifted point. Its center was (0, 0, 0) and its diameter was 20 mm. The shifted least squares fit was centered at (0, 0, 0.0021) with a diameter of 19.9975 mm—thus a 2.1 µm shift in the center and a 2.5 µm decrease in the diameter. By contrast, the constrained least-squares fit was centered at (0, 0, 0.00004) with a diameter of 19.9995, making only a 0.4 µm shift in center and a 0.5 µm decrease in diameter. These differences from the maximum-inscribed fit are lower by a factor of more than five from the shifted least squares fit.

6. THE CONSTRAINED LEAST-SQUARES DEFINITION FOR CYLINDRICAL DATUMS

As discussed earlier and as shown in Figs. 2, 6, and 7, the maximum inscribed and minimum circumscribed cylindrical functions suffer from instabilities for cases that are quite common. Besides the examples shown, a few individual discrete points that are located closer to the axis than the others can cause similar dimpling effects as discussed in the case of circles. This means that instabilities can arise from the actual form of the workpieces or from measurement errors affecting individual points.

In 2002 it was shown that some commercial software has had problems in the past with maximum-inscribed and minimum-circumscribed fits for cylinders [15]. Anecdotal evidence suggests some of the problems persist. A contributing factor to this problem may be the objective function.

To illustrate the problem and to show why the constrained least-squares fit is so much more stable in this case, we will graph the objective functions for the maximum-inscribed tapered case (Fig. 7), the other unstable cases being similar.

We consider a cylindrical bore 20 mm in diameter and 100 mm in length. We will impose a tapered form such that the top of the cylinder is exactly 20 mm in diameter and the bottom is 20.1 mm in diameter. For simplicity we sample only two rings of points—one at the top and one at the bottom (the results being similar if more rings were sampled). We will graph the objective functions for the maximum-inscribed and constrained least-squares cases. If the axis of the tapered cylinder were along the z-axis, we will view the objective functions for inscribed cylinders having axes that pass through (0, 0, 100) and (x, 0) where x runs from -0.1 to 0.1 (see Fig. 13).



Fig. 13. The setup for the investigation into the objective functions for maximum-inscribed and constrained least-squares. All units in mm.

The optimal solution for both objective functions occurs when x = 0. The radius of the cylinder is 10 mm at that point. To illustrate the difference in the objective functions, we graph them together. We have scaled the constrained least-squares objective function to have the same value (of 10.0 mm) for the optimal case of x = 0. We also plot the square root of the sum-of-squares instead of the sum-of-squares to properly make the results have the same units and thus be a fair comparison. The results are shown in Fig. 14.



Fig.14. The maximum-inscribed objective function (nearly horizontal) and the constrained least-squares objective function (appearing parabolic) plotted for values of *x* ranging from -0.1 mm to 0.1 mm. All units are mm.

As seen in the graph, the objective function for the maximum-inscribed cylinder is extremely shallow (and looks flat to the eye). An algorithm searching for a minimum would have to distinguish between objective function values that differ by extremely small amounts even for significant changes in orientation.

We note an important distinction between this type of instability and those discussed in the circle and sphere cases. In those cases, the instabilities described would have existed even if computations were made with arbitrarily high precision. In this case, the instability arises as a result of numerical issues. Nonetheless, even the effects of these instabilities are significantly diminished by the use of the constrained least-squares objective function.

The constrained least-squares datum definition easily computed the desired solution in all of the example cases depicted in Figs. 2, 6, and 7.

7. CONCLUSIONS

Given the pleasing results of the constrained least-squares datum definition in the planar case, we were compelled to look at the behavior of this datum definition for other cases, namely of circles, spheres, and cylinders. Since this is the first investigation (e.g., little was mentioned for instance about minimum circumscribed cases) we recognize that the investigation is not finished. However, we investigated many of the known, major problem areas of instabilities of some other datum definitions, and the constrained least-squares definition has been shown to perform excellently in all of them. It seems to agree very closely with the full contact solution when that fit is stable, and it also maintains desired stability with little cost (i.e., with little change of diameter compared to the full contact fit). These initial findings indicate that the constrained least-squares appears to be a safe datum definition choice and provide substantial optimism that results in future investigated cases will be pleasing as well.

REFERENCES

[1] 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, Vol. 227, No. 1, pp. 3-11, 2013. DOI: 10.1177/0954405412464012

[2] Tandler, W. "All Those Datum Things" *Inside Metrology*, Quality Digest, Quality Digest Magazine, 13 February 2008.

[3] Tandler, W. "Establishing Datum Reference Frames," *Inside Metrology*, Quality Digest, 12 March 2008.

[4] ANSI/ASME Y14.5.1M-2009 "Dimensioning and Tolerancing," The American Society of Mechanical Engineers, New York.

[5] ANSI/ASME Y14.5.1M-1994 "Dimensioning and Tolerancing," The American Society of Mechanical Engineers, New York.

[6] ISO 5459:2011. "Geometrical product specifications (GPS)—geometrical tolerancing—datums and datum systems." Geneva: International Organization for Standardization, 2011.

[7] Zhang, Xuzeng, and Roy, Utpal "Criteria for establishing datums in manufactured parts" Journal of Manufacturing Systems, 12(1), pp 36–50, 1993.

[8] Hopp, T. H., 1990, "The Mathematics of Datums," ASPE Newsletter, September 1990, American Society for Precision Engineering, Raleigh, NC. A reprint is available at: <u>http://www.mel.nist.gov/msidlibrary/doc/hopp90.pdf</u>

[9] Shakarji, C. M., and Srinivasan, V., "Datum Planes Based on a Constrained L1 Norm," ASME Journal of Computing and Information Science in Engineering, 15(4), 2015.

[10] Shakarji, C. M., and Srinivasan V., "An improved L1 based algorithm for standardized planar datum establishment," DETC2014-35461, Proceedings, ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Buffalo, NY, 2014.

[11] Shakarji, C. M., and Srinivasan V., "Theory and Algorithms for L1 Fitting Used for Planar Datum Establishment in Support of Tolerancing Standards," DETC2013-12372, Proceedings, ASME 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Portland, OR, 2013.

[12] Shakarji, C. M., and Srinivasan V., "A Constrained L2 Based Algorithm for Standardized Planar Datum Establishment," IMECE2015-50654, Proceedings, ASME 2015 International Mechanical Engineering Congress and Exposition Conference, Houston, TX, 2015.

[13] Shakarji, C. M., and Srinivasan V., "Theory and Algorithm for planar datum establishment using constrained total least squares," Proceedings, CIRP Conference on Computer Aided Tolerancing, Gothenburg, Sweden, 2016.

[14] Shakarji, C. M., "Least-squares fitting algorithms of the NIST algorithm testing system," J. Res. Natl. Inst. Stand. Technol., 103, pp. 633-641, 1998.

[15] Shakarji, C. M., "Evaluation of One- and Two-sided Geometric Fitting Algorithms in Industrial Software," Proceedings, ASPE Annual Conference, St. Louis, MO, 2002.