Recovering Circles and Spheres from Point Data

Christoph Witzgall, Geraldine S. Cheok, and Anthony J. Kearsley

National Institute of Standards and Technology
Gaithersburg, MD 20899
`witzgall@nist.gov, cheok@nist.gov, ajk@nist.gov`

Summary. Methods for fitting circles and spheres to point sets are discussed. LADAR (LAser Detection And Ranging) scanners are capable of generating "point clouds" containing the (x, y, z) coordinates of up to several millions of points reflecting the laser signals. In particular, coordinates collected off objects such as spheres may then be used to model these objects by fitting procedures. Fitting amounts to minimizing what is called here a "gauge function," which quantifies the quality of a particular fit. This work analyzes and experimentally examines the impact of the choice of three such gauge functions. One of the resulting methods, termed here as "algebraic" fitting, formulates the minimization problem as a regression. The second, referred to as "geometric" fitting, minimizes the sum of squares of the Euclidean distances of the data points from the tentative sphere. This method, based on orthogonal distance minimization, is most highly regarded and widely used. The third method represents a novel way of fitting. It is based on the directions in which the individual data points have been acquired.

Key words: algebraic fitting; circles; coordinate search; directional fitting; geometric fitting; LADAR; optimization; quasi-Newton; registration; spheres.

1 Introduction

In 1997, an article by Rorres and Romano [19] addressing an archaeological problem caught the attention of Saul Gass. In the Greek city of Corinth, the circular starting line for an ancient (circa 550 B.C.) track for foot races had been found. The other end of such a racetrack had to be marked by a turning pole [18]. The circular starting line was chosen, presumably, to equalize the distances between starting positions and the turning pole at the far end. The location of the turning pole could thus be inferred as the center of the circle passing through the starting line. This is precisely what Rorres and Romano did: they surveyed points on the starting line and fit a circle to these points.

Saul was intrigued by this problem of recovering a circle from a set of points as he had encountered circle and sphere fitting problems earlier in con-

nection with Coordinate Measuring Machines (CMMs), which are crucial to today's precision manufacturing. Their metrology is a major concern at the National Institute of Standards and Technology (NIST) [17], with which he has been associated for nearly three decades as an academic advisor. The computer-guided probe of a CMM touches an object and records to a high level of accuracy the (x, y, z) coordinates of the point on the object where contact was made. A collection of such measured points then permits an accurate assessment of the shape of the object, say, the roundness of a sphere, or its dimensions, say, the radius of a sphere. Captivated by the obvious parallels between such geometric problems and those encountered in classical Operations Research [6], he explored the use of linear programming in this context [9]. He also dug deeper into the racetrack problem, using several different methods for fitting the data and, in particular, compared the widths of "annuli," the areas between two concentric circles containing all data points. He was able to achieve tighter results than those reported by Rorres and Romano. A beautifully crafted unpublished manuscript [10] summarizes this work. It is also telling that his emphasis was not so much on "how" to compute, but rather on "what" to compute, in other words, the task of "modeling" so as to capture a particular aspect of reality.

Our work aspires to follow his example. It was prompted by the rapid growth of 3D imaging technology and its applications, and the corresponding need for metrological analysis. 3D imaging systems include laser scanners and optical range cameras. The former category covers LADARs (LAser Detection And Ranging) or laser radars. Similarly to a CMM, a LADAR also determines 3D coordinates of points on an object, but does so by sending out a laser signal and analyzing its reflection back to the instrument as indicated in Figure 1. Also, a LADAR is typically a scanning device that can obtain millions of measurements in a very short time, resulting in large "point clouds" of possibly millions of data points. Applications include the monitoring of construction sites, the development of "as built" models of existing structures, mapping, visualization of hidden objects, guiding of unmanned vehicles, just to mention a few.

1.1 LADAR Technology; Point Clouds

The metrology of LADARs is a current research issue at NIST. This work also supports the development of standard test protocols for the performance evaluation of LADARs. Figure 2 shows the indoor, artifact-based LADAR facility at NIST. Outdoor test facilities are planned for instruments with ranges of 100 m and above.

Figure 3 presents a LADAR visualization of a rubble pile. The casual reader may see this picture as just what a photographic camera might produce. But this would miss the point that, once the point cloud has been captured in 3D, it can be displayed as seen from different view points. For instance, the
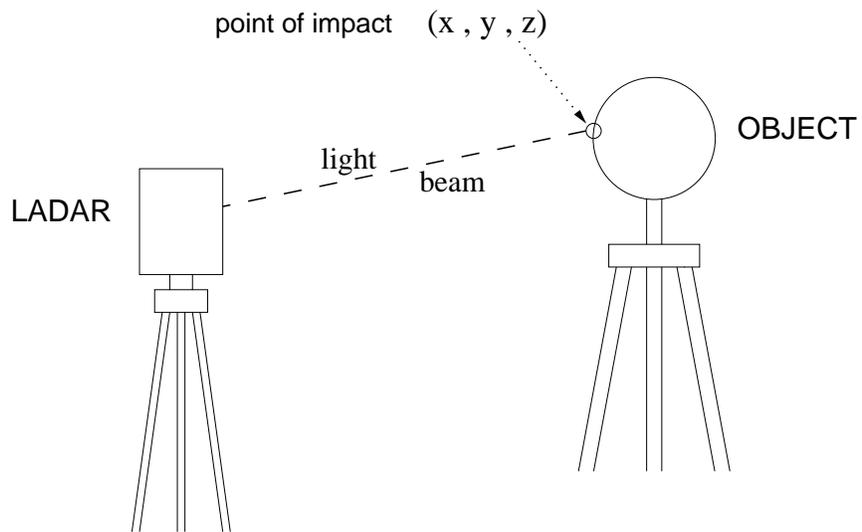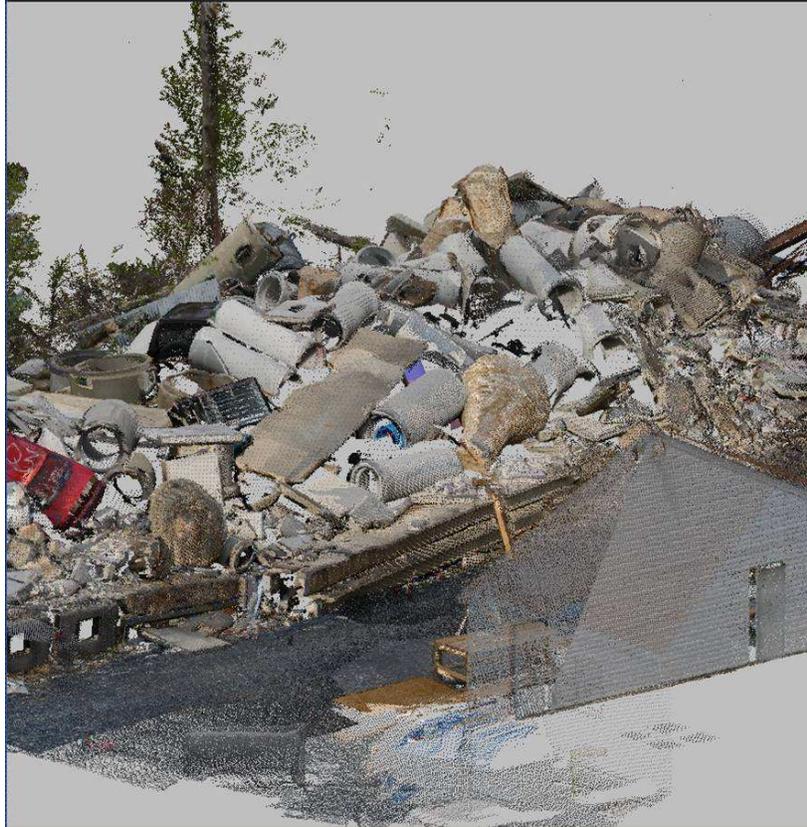
point of impact   (x , y , z)

LADAR

light beam

OBJECT

Fig. 1. Schematic of the operation of a LADAR scanner.

Fig. 2. Indoor artifact-based LADAR facility at NIST .

Disclaimer: Certain products are shown in this photograph. In no case does this imply recommendation or endorsement by NIST, nor does it imply that the products are necessarily the best available for the purpose.

Fig. 3. LADAR scan of a rubble pile

point cloud generated off a sphere is displayed in Figure 4 as seen sideways, that is, perpendicular to the direction of the scan.

The point cloud in Figure 4 consists of actual data points. The superimposed picture of the sphere does not depict the actual sphere but a "fitted" sphere, a sphere that in some sense best represents the point cloud, a sphere that was calculated using a fitting algorithm. The fitted sphere is expected to capture the radius and location of the actual sphere vis-à-vis the point cloud. Fitting the sphere thus provides a method for recovering the actual sphere from which the point cloud had been generated.

The term "recovery" suggests an emphasis on certain applications that are different from the ones pursued when using CMMs. LADARs are more likely to be used for locating, identifying, and recognizing an object. CMMs, on the other hand, emphasize quality control. For spheres, a typical question there would be how small an annulus between two concentric spheres would contain the data points. Also, as Figure 4 shows, the point clouds generated by
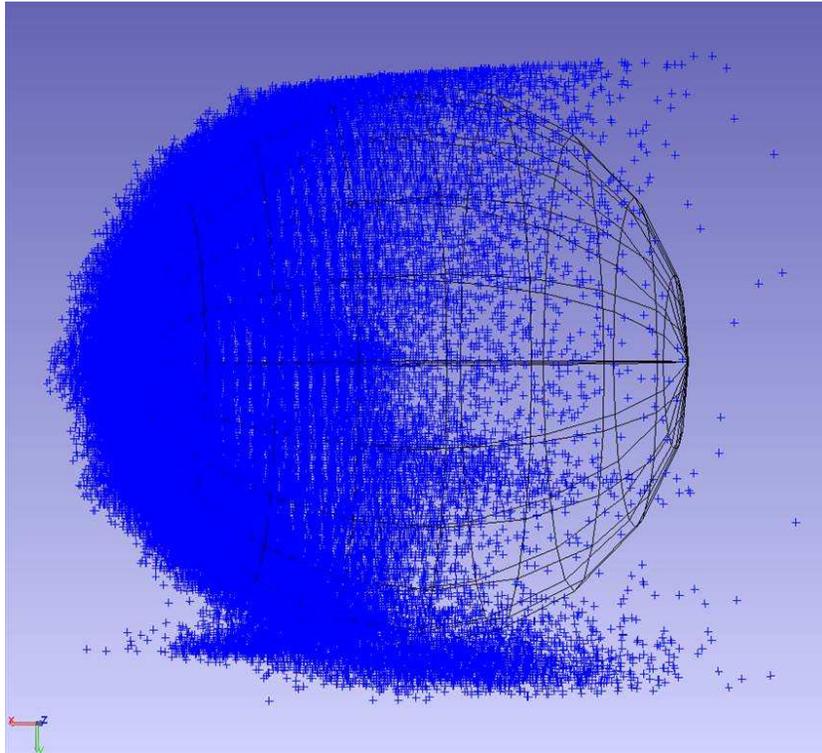
Fig. 4. Point cloud and fitted sphere.

LADARs tend to be big and "noisy," i.e., subject to significant random errors. The data generated by CMMs, on the other hand, tend to be less voluminous and less noisy.

While the emphasis in this work is on spheres, the results will typically hold for circles, too.

1.2 The Fitting Paradigm

The term "fitting" will be used here in a somewhat narrow and explicit sense. This sets it apart from other approaches to recovering scanned objects. Iterative Closest Point (ICP) [3] techniques have also been very effective. Tetrahedralization-based surface generation techniques may well be the way of the future [2].

Fitting requires the definition of a "gauge function," which expresses a measure of deviation of the point cloud as a whole from a putative geometric model such as a sphere. The parameters of the model governing its shape and size are the arguments of the gauge function. They are the variables of

an optimization problem: to minimize the deviation of data points from an object model.

Gauge functions are typically defined in two steps. First, a measure of the individual deviation

$$\Delta_i = \text{deviation}\,((x_i, y_i, z_i),\ \text{model})$$

of each data point from the putative model is defined. Then a norm $\|*\|$ is selected and applied to the vector of all deviations Δ_i in order to arrive at a single gauge value for an overall measure of deviation,

$$\boldsymbol{\Delta} \;=\; \|(\Delta_1\,,\ \Delta_2\,,\ ...\ \Delta_n)\| \ .$$

The quantity $\boldsymbol{\Delta}$ represents the desired gauge function, which depends on the model parameters as its arguments, since every one of the individual deviations Δ_i depends on those parameters.

The following are the most commonly selected norms:

- the maximum or Chebychev norm (L_∞): $\max_i |\Delta_i|$,
- the least-squares norm (L_2): $\sqrt{\sum_i \Delta_i^2}$,
- the sum of absolute values norm (L_1): $\sum_i |\Delta_i|$.

Directly minimizing the sum of squared deviation is, of course, equivalent to minimizing their L_2 norm. In practice, to avoid the very large numbers that typically arise for large data sets, both the L_2 and the L_1 norms are usually averaged by the number n of data points. The resulting measures are commonly referred to as

$$\text{RMS} \;=\; \sqrt{\frac{\sum_i \Delta_i^2}{n}}\,,\quad \text{ASD} \;=\; \frac{\sum_i |\Delta_i|}{n} \ .$$

The least-squares approach, as in using the RMS, is most commonly used. At NIST, a system for testing least-squares algorithms for a collection of types of objects including spheres and circles has been developed and implemented [20], as has been orthogonal distance regression for specified functional expressions [4]. Applications of the Chebychev norm are given in [1]. Chebychev fitting of circles has been examined by Saul Gass [9, 10] and others [22, 8]. As mentioned before, the interest here is in determining the minimum annulus containing all data points. The authors apply linear programming as well as farthest point Voronoi diagrams. In our work on spheres, the desired end result consists of the "center coordinates" (x_0^*, y_0^*, z_0^*) of a fitted sphere, and perhaps its "radius" r^*. Indeed, when fitting spheres, two different tasks may be encountered:

- fitting a sphere with its radius "free" to be determined;
- fitting a sphere with a specified "fixed" radius.

1.3 Layout of the Article

The remainder of the paper is organized as follows. In Section 2, we describe physical evidence on fitting spheres collected by the second author. In Section 3, two standard methods are discussed: "algebraic fitting" (Section 3.1) and "geometric fitting" (Section 3.2). Section 4 is devoted to the geometry (Section 4.1) and the process of scan-directed fitting (Section 4.2) based on an algorithm developed by the third author.

2 Results of Experiments

At NIST, considerable experience in LADAR scanning and fitting has been gathered. Key issues are applications to registration, volume determination, and object identification.

2.1 Locating I-Beams

The following demonstration experiment [12] was designed to demonstrate the feasibility of automated placing and pick-up of an I-beam by the computer-guided crane developed at NIST. The I-beam, residing on the laboratory floor, was scanned for location and orientation, with the data in the LADAR's coordinate system. A LADAR scanner was used to determine the pose (location and orientation) and the type of an I-beam residing on the floor of a laboratory at NIST. The idea was to scan the I-beam in order to determine shape and pose within the coordinate system of the LADAR instrument.

This instrument's coordinate system then had to be related to the coordinate system of the crane, a process generally called "registration". To this end, three "target" spheres, "A","B","C", were placed in the vicinity of the I-beam

The centers of the spheres were predetermined in the coordinate system of the crane. The LADAR scan covered these spheres along with the I-beam, and the fitting process yielded center coordinates in the instrument system. Thus, there were three target locations, each of which with coordinates known in both systems. The algorithm "Procrustes" [21] [14] was employed, which combines a translation with a rotation in order to transform one system into the other, matching the coordinates at each target point as well as possible. More precisely, this transformation is chosen so as to minimize the sum of the squares of the resulting coordinate differences at each target point. It is clear that the accuracy of the fitting algorithm as applied to the target spheres is the key to a correct registration.

This successful demonstration also provided an opportunity to experiment with sphere fitting. As described in [12], several fitting algorithms were implemented and tried in addition to a commercial software package. The radius of the target spheres was specified by the manufacture to be 76.2 mm (3 inches).
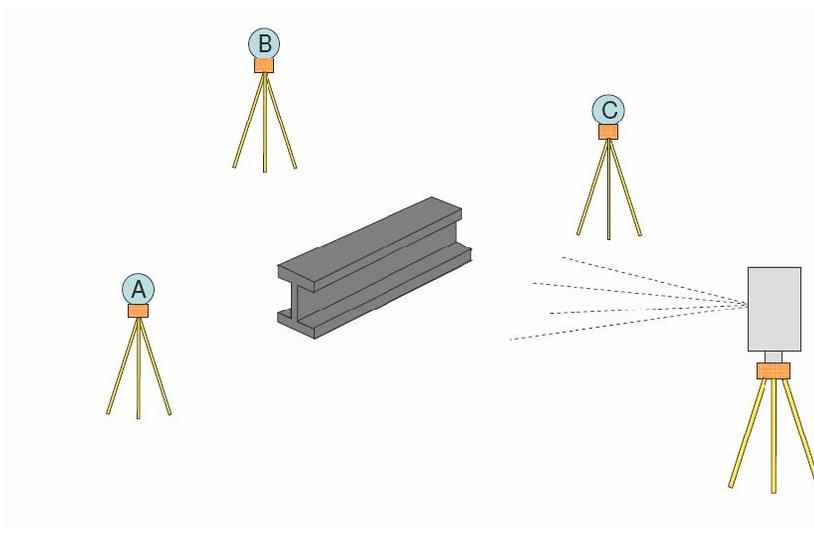
Fig. 5. Determining location and orientation of an I-beam by LADAR.

Thus, the fixed radius option for sphere fitting was used for the actual demonstration.

Determining the accuracy rather than the precision of such LADAR measurements would require the availability of an independent measurement system at a higher level of accuracy, which was not available at the time. Thus it was not possible to ascertain the actual centers of the target spheres with sufficient accuracy to check against the centers derived from fitting the LADAR data. The radii of the target spheres, however, were presumably known, and it could be determined how well the radii were reproduced if the free radius option of the sphere fitting algorithms were to be used. The results were disconcerting. The radii were generally underestimated. In particular, applying the commercial fitting package to target sphere "C" yielded for $n = 90$ the average radius,

$$r_{average} = 69 \text{ mm}, \ r_{actual} = 76 \text{ mm}, \ st.dev. = 3 \text{ mm},$$

a discrepancy of about 10 %. This result raised a "red flag". Could the center determination be trusted if established fitting methods produced systematic errors in the measurement of the radius? An effort was therefore started to find the reasons for such errors.

At first, suspicion centered on the quality of the target spheres which were made of styrofoam purchased at a local crafts store. It was thought that this material might permit some penetration by the scan beam, or perhaps the dimensions of the spheres were not quite accurate. These considerations led to the fabrication of a machined aluminum sphere. However, the same

discrepancies were encountered when this sphere was scanned and fitted, as will be seen in Section 2.2.

One other possible explanation concerned the LADAR instrument. Perhaps the distribution of the scan errors was not symmetric. In other words, there may have been more undershoots than overshoots, or vice versa. And finally, the instrument position with respect to the sphere may possibly matter. To check for this latter possibility, the experiment reported in Section 2.2 below was conducted.

2.2 An Additional Experiment

In this experiment, a data set was collected, and reduced to avoid boundary effects, off an aluminum sphere machined to a radius of 101.6 mm (4 in). This data set is displayed in Figure 6 together with two subsets, an upper and a lower subset into which the full set had been split as shown in Figure 7. The results of applying the commercial fitting package to these three data sets are displayed in Tables 1 and 2.
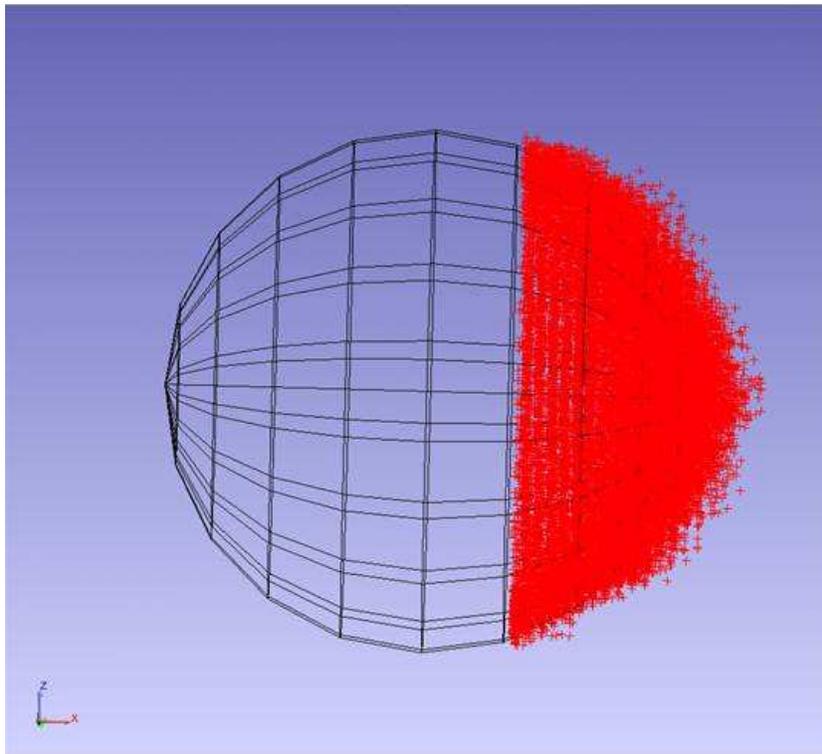
Fig. 6. Full "hemispherical" data set from aluminum sphere.
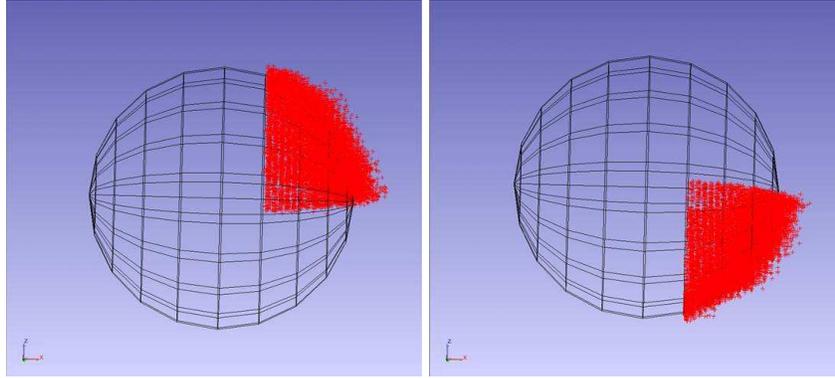
Fig. 7. Upper and lower portions of the hemispherical data set.

Table 1. Results of the experiment; variable radius.

	x	y	z	r
Full	-6254.99	-196.51	-78.85	98.41
Upper	-6258.27	-196.37	-83.02	102.36
Lower	-6258.61	-196.82	-72.61	103.66

Table 2. Results of the experiment; fixed radius.

	x	y	z	r
Full	-6259.19	-196.58	-78.87	101.6
Upper	-6257.52	-196.36	-82.55	101.6
Lower	-6256.59	-196.77	-73.98	101.6

The first observation concerns the result of fitting the full data set with the free radius option. As in the demonstration reported in Section 2.1, the radius was still underestimated:

$$r_{computed} = 98.41 \text{ mm}, \ r_{actual} = 101.6 \text{ mm},$$

but then it was overestimated for both the upper and the lower portion of the full data set. The next observation concerned the high level of sensitivity in the z-coordinate, which represents vertical elevation. Note that the same sensitivity in the z-coordinate showed up when the known radius of 101.6 mm had been kept fixed. Such variations are at odds with the fact that regions on the sphere are equivalent. Indeed, the upper and the lower data set occupy essentially symmetric positions on the sphere. Yet there is a substantial difference in fitting results.

The upper and the lower subset are, however, in a different position vis-à-vis the LADAR instrument. The angles of incidence certainly differ for these two subsets. This forces the conclusion that the instrument position has to

be taken into account when fitting. A method for this will be presented in Section 4.

3 Algebraic and Geometric Fitting

In a key paper [7], two generic concepts for defining deviations of data points were distinguished, and fitting methods based on those concepts were termed "algebraic" and "geometric," respectively. Algebraic fitting, in a broad sense, is based on an equation describing the object to be fitted, defining the deviation of a data point as the amount by which that equation is violated. In general geometric fitting, that deviation is defined as the orthogonal Euclidean distance to the object. For geometric fitting of functions, the term "orthogonal distance regression" is frequently used.

Specifically in [7], algebraic and geometric fitting methods based on least squares were examined for circles and ellipses. In this section, we will take the same approach to fitting circles and spheres.

3.1 Algebraic Fitting

Let (x_0, y_0, z_0) denote the center coordinates, and $r \geq 0$ the radius of a sphere. Then the following equation

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 - r^2 = 0$$

characterizes the points (x, y, z) of the sphere. Substituting

$$x_0 = -\frac{a}{2}, \; y_0 = -\frac{b}{2}, \; z_0 = -\frac{c}{2}, \; r^2 = \frac{a^2 + b^2 + c^2}{4} - d, \tag{1}$$

yields an alternate equation

$$x^2 + y^2 + z^2 + ax + by + cz + d = 0$$

of the above sphere in terms of linear parameters a, b, c, d. Note that the above equation has geometric meaning only if its parameters satisfy the condition

$$\frac{a^2 + b^2 + c^2}{4} - d \geq 0, \tag{2}$$

as otherwise the resulting radius would not be a real number.

The above equations for the sphere suggest the following definition of a deviation from the sphere by a data point (x_i, y_i, z_i):

$$\Delta_i = (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 - r^2, \tag{3}$$

and, equivalently,

$$\Delta_i = x_i^2 + y_i^2 + z_i^2 + ax_i + by_i + cz_i + d. \tag{4}$$

The algebraic method to be discussed here takes advantage of the fact that the deviation expression (4) is linear in its parameters. Choosing the least squares norm leads to the gauge function

$$F = \sum_i \left(x_i^2 + y_i^2 + z_i^2 + ax_i + by_i + cz_i + d \right)^2, \tag{5}$$

the minimization of which amounts to a straightforward linear regression,

$$-x_i^2 - y_i^2 - z_i^2 \sim ax_i + by_i + cz_i + d, \tag{6}$$

with an optimal solution,

$$a^*, b^*, c^*, d^*,$$

which always exists and is unique as long as atleast four data points do not lie in a single plane.

Also, the optimal parameters satisfy the condition (2). Indeed, it is well known that the data averages satisfy regression equations such as (6) exactly if the optimal parameters are inserted:

$$\frac{1}{n} \sum_i (-x_i^2 - y_i^2 - z_i^2) = a^* \frac{1}{n} \sum_i x_i + b^* \frac{1}{n} \sum_i y_i + c^* \frac{1}{n} \sum_i z_i + d^*.$$

Then

$$\frac{(a^*)^2 + (b^*)^2 + (c^*)^2}{4} - d^*$$

$$= \frac{1}{n} \sum_i \left(x_i^2 + y_i^2 + z_i^2 + a^* x_i + b^* y_i + c^* z_i + \frac{(a^*)^2 + (b^*)^2 + (c^*)^2}{4} \right)$$

$$= \frac{1}{n} \sum_i \left((x_i + \frac{a^*}{2})^2 + (y_i + \frac{b^*}{2})^2 + (z_i + \frac{c^*}{2})^2 \right) \geq 0.$$

Thus, for any data set containing at least four points not in a single plane, the regression (6) yields a unique result that, moreover, represents a real sphere.

By (1), the above derivation also yields an explicit expression of the optimal radius in terms of the optimal center coordinates:

$$(r^*)^2 = \frac{\sum_i \left((x_i - x_0^*)^2 + (y_i - y_0^*)^2 + (z_i - z_0^*)^2 \right)}{n}. \tag{7}$$

The expressions (3) and (4) represent the same deviation values and lead to the same gauge quantity, provided the same norm is used to combine the respective individual deviations, in this case, the least squares norm, However, the two resulting gauge functions differ in their parameters. As a positive

definite quadratic function, the former gauge function F is convex, whereas the latter gauge function

$$G = \sum_i \left((x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 - r^2\right)^2$$

is not. If the value of the radius has been prespecified, then this value needs to be accessible as one of the parameters, as in the above gauge function G, but not in gauge function F. As a result, regression will not work for the fixed radius option. Furthermore, there will always be minima if the radius is fixed. They may be, however, no longer unique as an example in Section 3.3 shows.

3.2 Geometric (Orthogonal) Fitting

Here, the actual Euclidean distance of a data point (x_i, y_i, z_i) from the sphere given by its center coordinates (x_0, y_0, z_0) and its radius r is specified as the deviation of this data point from this sphere:

$$\Delta_i = \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2} - r. \tag{8}$$

Following least squares, the gauge function

$$H = \sum_i \left(\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2} - r\right)^2 \tag{9}$$

characterizes the particular "geometric fitting" method examined in this paper.

By comparing results with a geometric fitting algorithm implemented by the authors, it was possible to ascertain that the commercial fitting package employed in the experiments described in Section 2 uses the geometric fitting method. In what follows, we will refer simply to "geometric fitting" regardless of which implementation has been used.

A comparative analysis of algebraic and geometric fitting has been provided by [16]. The difference in the performance of these methods is also illustrated by the example in Figure 8. Here, a LADAR scan of the sphere discussed in Section 2.1 has been fitted algebraically, and the algebraic deviations defined by (3) or (4) — the algebraic errors so to speak — have been plotted in the sequence in which the data points appear in the point cloud. It is seen that these algebraic errors are closely gathered around the zero horizontal, which is not surprising, since that is what was minimized. However, the geometric deviations defined by (8) are also displayed, exhibiting uncomfortably large swings. Both kinds of deviations show definite patterns such as the scalloped pattern of the algebraic deviations. The reasons for these patterns are not understood.

Moreover, the radius of that sphere determined by algebraic fitting falls even more short of the actual value than the one determined by geometric fitting:
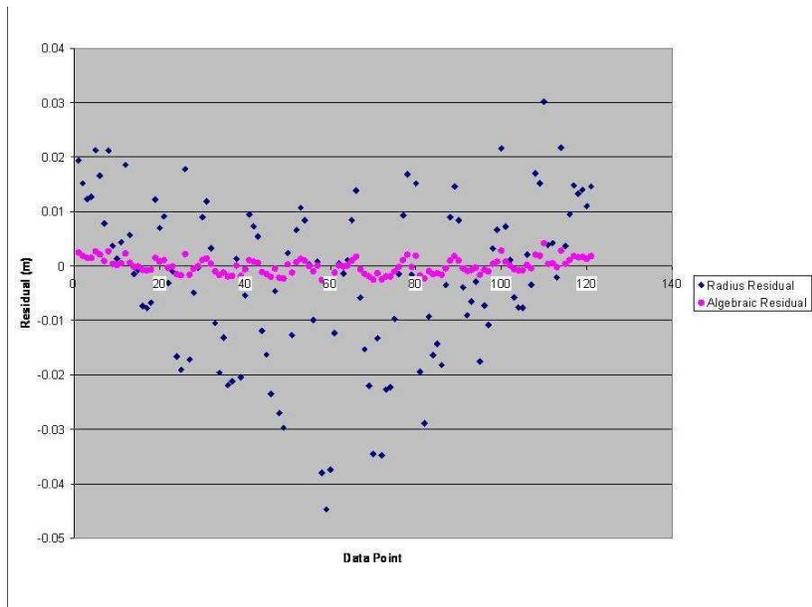
Fig. 8. Deviation display: geometric vs. algebraic fitting.

$$r_{algebraic} = 54 \text{ mm}, \ r_{geometric} = 69 \text{ mm}, \ r_{actual} = 76 \text{ mm}.$$

Experiences similar to the above, coupled with the strong intuitive appeal of orthogonal distance deviations and the general acceptance of the least squares approach, have made geometric fitting the main choice of the metrology community.

Contrary to algebraic fitting, the geometric fitting may not always have a solution. The reason is that an orthogonally best fitted plane, — in a sense a sphere with infinitely large radius —, competes with actual spheres. In such cases, the geometric fitting procedure may not converge as the radius grows ever larger. Furthermore, even if a solution exists, the solution may not be unique. Corresponding examples in the case of circles are given in Section 3.3. In addition, the gauge function H defined by Equation (9) is not differentiable in its parameters. The function displays an upward pointing cusp, wherever the center coordinates coincide with those of a data point:

$$(x_0, y_0, z_0) = (x_i, y_i, z_i).$$

3.3 Examples of Circles

Again, the following examples for circles are indicative of phenomena that pertain to spheres, as well.

The first extremely simple example shows that any fitting method with fixed radius may have different optimal solutions for geometric fitting.

Example A:

$$x_1 = -1, y_1 = 0,$$
$$x_2 = +1, y_2 = 0.$$

The second example admits four different optimal solutions in symmetric positions.

Example B:

$$x_1 = \quad 0, \ y_1 = \quad 0,$$
$$x_2 = +10, \ y_2 = +10,$$
$$x_3 = -10, \ y_3 = +10,$$
$$x_4 = -10, \ y_4 = -10,$$
$$x_5 = +10, \ y_5 = -10.$$

The center of one optimal circle is at $x_0^* = 3.892718$, $y_0^* = 0.0$, and its radius is $r^* = 12.312514$. The remaining three optimal circles are in symmetric position rotated by multiples of right angles around the origin. Algebraic fitting, on the other hand, has a unique solution, which for reasons of symmetry is centered at the origin. The optimal radius can then be calculated using (7): $r = \sqrt{160} = 12.64911$.

In order to establish optimality, the gradient and the Hessian for the parameters of the circle to the right were also computed. Up to factors of two, the gradient components compute to zero within seven digits after the decimal point, and the following Hessian results:

$$\begin{vmatrix} +0.5957504 & 0 & -0.31615945 \\ 0 & +0.06370123 & 0 \\ -0.31615945 & 0 & 1 \end{vmatrix}.$$

Since the eigenvalues of the Hessian, 0.0637, 0.4226, 1.1731, are all positive, local optimality follows. No other local optima have been found. The calculations were done with in-house geometric fitting software. The commercial fitting package has also been used, but produced a saddle point instead of a local minimum.

The third example is a data set without an optimal solution when geometric fitting is attempted. The $x-$axis is a limit solution.

Example C:

$$x_1 = +10, \ y_1 = +1,$$
$$x_2 = \quad +1, \ y_2 = \quad 0,$$
$$x_3 = -10, \ y_3 = +1,$$
$$x_4 = -10, \ y_4 = -1,$$
$$x_5 = \quad -1, \ y_5 = \quad 0,$$
$$x_6 = +10, \ y_6 = -1.$$

The claim that there are no finite local minima rests on use of the in-house software, which established that absence within a very large radius around the origin.

4 Fitting in Scan Direction

In what follows, we assume that the data points are acquired in scan direction, i.e., the data point (x_i, y_i, z_i) and its intended impact point on a sphere are found on a ray, referred to as "scan ray," emanating from the origin $(0, 0, 0)$, so that the intended target point is at the intersection of the scan ray with the sphere. Of course, the scan ray may not intersect the sphere, in which case we declare a "miss;" otherwise, a "hit." In reality, one might argue, every data point would arise from a hit, as otherwise it would not have gotten into the data set. But as long as a sphere is not in best-fit position, misses are to be expected.

4.1 Geometry of Directed Deviations

We aim to determine the intersection, closest to the origin, of a sphere by a ray from the origin through a given data point (x_i, y_i, z_i). While this just amounts to solving a quadratic equation, there is also a straightforward geometric solution.
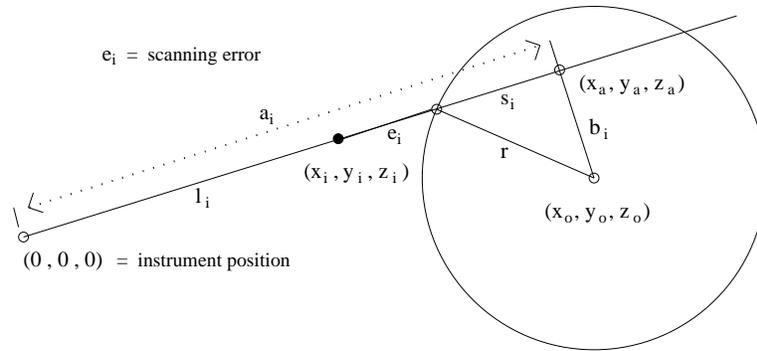
Fig. 9. Geometry of deviations in scan direction.

Consider the plane spanned by the ray and the center (x_0, y_0, z_0) of the sphere (see Figure 9). The direction of the ray is given by the direction cosines,

$$\xi_i = \frac{x_i}{\ell_i}, \ \eta_i = \frac{y_i}{\ell_i}, \ \zeta = \frac{z_i}{\ell_i},$$

where

$$\ell_i = \sqrt{x_i^2 + y_i^2 + z_i^2} > 0 \tag{10}$$

denotes the distance of the data point from the origin. Of interest is the orthogonal projection of the center of the sphere into the ray:

$$a_i(\xi_i, \eta_i, \zeta_i). \tag{11}$$

Here a_i denotes the distance of the projection point (11) from the origin. Using the orthogonality of the ray direction (ξ_i, η_i, ζ_i) to the difference vector

$$(a_i\xi_i, a_i\eta_i, a_i\zeta_i) - (x_0, y_0, z_0), \tag{12}$$

we find the expression

$$a_i = x_i\xi_i + y_i\eta_i + z_i\zeta_i$$

for the distance a_i.

Next we introduce the length of the difference vector (12), i.e., the distance between the sphere center and its projection into the scan ray. We have

$$b_i = x_0^2 + y_0^2 + z_0^2 - a_i^2 \ \geq \ 0, \tag{13}$$

by the Pythagorean theorem. Comparison against the radius

$$b_i \leq r, \tag{14}$$

yields the condition for the scan ray hitting the sphere. If this condition is met, then the sphere center, the projection point on the ray, and the intersection point on the sphere form a right triangle with sides b_i and s_i, and the hypotenuse r,

$$s_i = \sqrt{r^2 - b_i^2}. \tag{15}$$

It is now possible to express the distance along the scan ray from the origin to the impact point on the sphere. The difference between this distance and the distance ℓ_i of the data point from the origin (10),

$$a_i - s_i - \ell_i = a_i - \ell_i - \sqrt{r^2 - b^2}, \tag{16}$$

represents, in the case of a hit, the overshoot or undershoot of the measurement with respect to the sphere.

In general, the scan ray intersects the sphere at two points, one close to the origin of the instrument, and one on the other side of the sphere. Analogous to the above derivation, we find the expression

$$a_i + s_i - \ell_i = a_i - \ell_i + \sqrt{r^2 - b^2}$$

for the distance of the intersection point farther from the origin. It may be of interest to note that the product of those two distances equals the algebraic orthogonal deviation given by Equation (3)

$$(a_i - s_i - \ell_i)(a_i + s_i - \ell_i) = (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 - r^2. \tag{17}$$

Indeed, by (15),

$$(a_i - s_i - \ell_i)(a_i + s_i - \ell_i) = (a_i - \ell_i)^2 - s_i^2$$
$$= (a_i - \ell_i)^2 + b_i^2 - r^2.$$

By the Pythagorean theorem applied to the triangle formed by the data point, the projection point and the sphere center,

$$(a_i - \ell_i)^2 + b_i^2 = (x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2,$$

which establishes (17).

4.2 Geometric Scan-Directed Fitting

The expression (16) of the presumptive measurement error of a data point (x_i, y_i, z_i) was determined under the assumption of a hit. It, therefore, fails to provide a suitable basis for a gauge function. Indeed, any gauge function based solely on hits would yield a minimum of zero for any sphere position that avoids hits altogether. For this reason, we propose to define deviations also for data points whose scan rays are missing the sphere. The version of "directional" fitting reported here thus distinguishes between two kinds of deviations according to the value of b_i as follows:

If the data point causes a hit, that is, if $b_i \leq r$ (14), then following (15),

$$\Delta_i = a_i - s_i - \ell_i = a_i - \ell_i - \sqrt{r^2 - b^2}. \tag{18}$$

If the data point causes a miss, that is, if $b_i > r$, then

$$\Delta_i = \sqrt{(x_i - x_0)^2 + (y_0 - y_i)^2 + (z_i - z_0)^2} - r, \tag{19}$$

which in fact represents the orthogonal distance deviation (8).

The least squares approach leads to the gauge function

$$H = \sum_i \Delta_i^2. \tag{20}$$

Minimizing this gauge function may still not provide the desired result, because the so-fitted sphere may still permit some data points to miss, and the deviations assigned to these values may cause distortions. The procedure chosen here is to delete such data points temporarily, and re-minimize. The full point cloud is then again screened for data points that cause misses of the re-minimized sphere, and these points are deleted prior to the next re-minimization. That process of deleting followed by re-minimization is repeated until there are no more misses or the number of misses has stabilized. The goal is to arrive at a sphere fitted to a stable set of data points all providing hits.

This and similar procedures are still subject to experimentation and adjustments. Our first efforts, however, were clearly successful as attested by

Table 3. Results of fitting in the scan direction.

	x	y	z	r
Full	-6258.98	-198.07	-79.18	101.29
Upper	-6259.06	-198.15	-78.90	101.22
Lower	-6259.38	-198.01	-79.12	101.60

the results given in Table 3. They indicate that the abnormalities reported in Section 2 appear indeed to be caused primarily by "modeling error," namely the choice of an unsuitable gauge function for the fitting procedure.

The gauge function (20) is not differentiable, because a perturbation of the sphere parameters may cause the deviation of a data point switch from definition (18) to definition (19), or vice versa, and such transitions are not smooth. The minimization of such a gauge function thus requires an optimizer that does not require differentiability. The method that turned out to be successful was a method based on recent research in the optimization of noisy functions. Loosely speaking, this method is designed for minimizing functions with many non-differentiabilities as long as these are "shallow," i.e., they act as perturbations of an overall differentiable function. This method also permits constraining the minimization to avoid, for instance, values of variables that define spheres that would infringe upon the instrument location.

The algorithm proceeds in two stages. Initially, a quasi-Newton method ("BFGS") is employed to solve the nonlinear programming problem at hand, where gradients of the objective function are calculated using a centered finite-difference approximation with a large finite difference initial step-length [11]. As the algorithm progresses, the finite difference step-length is decreased until its size falls below the square root of machine precision. Subsequently, a simplex-based coordinate search method is employed [15]. This coordinate search method requires no gradient calculation or approximation, and has been applied successfully in the past on difficult non-differentiable constrained optimization problems (e.g. [13]).

5 Concluding Remarks

The analyses in this paper underscore the fact that the outcome of fitting methods strongly depends on the choice of the gauge function that is minimized in order to establish the fit. Three gauge functions were discussed. The associated "algebraic" and "geometric" fitting methods are most commonly used. In a novel third approach, the deviations are measured in the direction of the scan. In all three methods, the deviations were combined into single numbers by invoking the least squares L_2 norm.

The experimental evidence presented here, although limited so far, indicates that for certain applications it may not be enough to base fitting methods on the coordinates of points alone, but that it may be necessary to take into

account the directions in which those points had been acquired. This may require a general revision of currently used approaches to approximation.

Acknowledgments

References

1. G.T. Anthony, B. Bittner, B.P. Butler, M.G. Cox, R. Drieschner, R. Elligsen, A.B. Forbes, H. Gross, S.A. Hannaby, P.M. Harris, and J. Kok. Chebychev reference software for the evaluation of coordinate measuring machine data. *Report EUR 15304 EN*, National Physical Laboratory, Teddington, United Kindom, 1993.
2. J. Bernal. AGGRES: a program for computing power crusts of aggregates. *NISTIR 7306, National Institute Of Standards and Technology*, Gaithersburg, MD, 2006.
3. P.J. Besl and N.D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:239–256, 1992.
4. P.T. Boggs and J.E. Rogers. Orthogonal distance regression. *Contempory Math*, 112:183–194, 1990.
5. S-Y. Chou, T.C. Woo, and S.M. Pollock. On characterizing circularity. Dept. of Industrial and Operations Research, University of Michigan, Ann Arbor, 1994.
6. Z. Drezner, S. Steiner, and O. Weselowski. On the circle closest to a set of points. *Computers and Operations Research*, 29:637–650, 2002.
7. W. Gander, G.H. Golub, and R. Strebel. Least-squares fitting of circles and ellipses. *BIT*, 34:558–578, 1994.
8. J. Garcia-Lopez, P. A. Ramos, and J. Snoeyink. Fitting a set of points by a circle. *Discrete and Computational Geometry*, 20:389–402, 1998.
9. S.I. Gass, C. Witzgall, and H.H. Harary. Fitting circles and spheres to coordinate measuring machine data. *International Journal of Flexible Manufacturing Systems*, 10:5–25, 1998.
10. S.I. Gass. Comments on an ancient Greek racecourse: Finding minimum width annuluses. (unpublished manuscript), 1998.
11. P. Gilmore and C.T. Kelley. An implicit filtering algorithm for optimization of functions with many local minima. *SIAM Journal on Optimization*, 5:269–285, 1995.
12. D.E. Gilsinn, C. Witzgall, G.S. Cheok, and A. Lytle. Construction object identification from LADAR scans: an experimental study using I-beams. *NISTIR 7286, National Institute of Standards and Technology*, Gaithersburg, MD, 2005.
13. R. Glowinski and A.J. Kearsley. On the simulation and control of some friction constrained motions. *SIAM Journal on Optimization*, 5:681–694, 1995.
14. E. Golub and C.F. Van Loan. *Matrix Computation*. The Johns Hopkins University Press, Baltimore, MD, 1996.

15. R. M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9:1082–1099, 1999.
16. B.A. Jones and R.B. Schnabel. A comparison of two sphere fitting methods. *Proceedings of the Instrumentation and Measurement Technology Subgroup of the IEEE*, 1986.
17. S.D. Phillips, B. Borchardt, and G.C. Gaskey. Measurement uncertainty consideration for coordinate measuring machines. *NISTIR 5170, National Institute of Standards and Technology*, Gaithersburg, MD, 1993.
18. D.G. Romano. Athletes and mathematics in archaic Corinth: the origins of the Greek stadium. *Memoirs of the American Philosophical Society*, 206:xiv,117, 1993.
19. C. Rorres and D.G. Romano. Finding the center of a circular starting line in an ancient Greek stadium. *SIAM Review*, 39:745–754, 1997.
20. C.M. Shakarji. Least-squares fitting algorithms of the NIST algorithmic testing system. *Journal of Research of the National Institute of Standards and Technology*, 103:633–640, 1998.
21. P.H. Schoenemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrica*, 31:1–10, 1966.
22. C.K. Yap. Exact computational geometry and tolerancing metrology. *Snapshots on Computational and Discrete Geometry*, D. Avis and P. Bose, eds., McGill School of Computer Science, 3:34–48, 1994.