

Research Article

Is the Best Fitting Curve Always Unique?

N. Chernov, Q. Huang, and H. Ma

Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Correspondence should be addressed to N. Chernov; chernov@math.uab.edu

Received 3 September 2012; Revised 22 October 2012; Accepted 23 October 2012

Academic Editor: Ada Che

Copyright © 2013 N. Chernov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fitting straight lines and simple curved objects (circles, ellipses, etc.) to observed data points is a basic task in computer vision and modern statistics (errors-in-variables regression). We have investigated the problem of existence of the best fit in our previous paper (see Chernov et al. (2012)). Here we deal with the issue of uniqueness of the best fit.

1. Introduction

This is a continuation of our paper [1] where we studied the problem of *existence* of the best fitting curve. Here we deal with its *uniqueness*.

Our interest in these problems comes from applications where one describes a set of points P_1, \dots, P_n (representing experimental data or observations) by simple geometric shapes, such as lines, circular arc, elliptic arc, and so forth. The best fit is achieved when the geometric distances from the given points to the fitting curve are minimized, in the least squares sense. Finding the best fit reduces to the minimization of the objective function

$$\mathcal{F}(S) = \sum_{i=1}^n [\text{dist}(P_i, S)]^2, \quad (1)$$

where S denotes the fitting curve (line, circle, ellipse, etc.). Here $\text{dist}(P, S) = \inf_{Q \in S} d(P, Q)$ denotes the *shortest* distance from P to S , and d stands for the Euclidean metric in \mathbb{R}^2 . We refer the reader to [1] for the background of the geometric fitting problem.

Most publications on the fitting problem are devoted to practical algorithms for finding the best fitting curve minimization (1) or statistical properties of the resulting estimates. Very rarely one addresses fundamental issues such as the existence and uniqueness of the best fit. If these issues do come up, one either assumes that the best fit exists and is unique or just points out examples to the contrary without deep investigation.

In our previous paper [1] we investigated the existence of the best fit. Here we address the issue of uniqueness. These issues turn out to be quite nontrivial and lead to unexpected conclusions. As a glimpse of our results, here and in [1], we provide a table summarizing the state of affairs in the problem of fitting most popular 2D objects (here **Yes** means the best fitting object exists or is unique in all respective cases; **No** means the existence/uniqueness fails in some of the respective cases).

We see that the existence and uniqueness of the best fitting object cannot be just taken for granted. Actually 2/3 of the answers in Table 1 are negative. In particular, the uniqueness can never be guaranteed. (For the exact meaning of *all cases* and *typical cases* we refer the reader to [1].)

The uniqueness of the best fit is not only of theoretical interest but also practically relevant. The nonuniqueness means that the best fitting object may not be stable under slight perturbations of the data points. An example is described by Nievergelt [2]. He presented a set of $n = 4$ points that can be fitted by three different circles equally well. Then by arbitrarily small changes in the coordinates of the points, one can make any of these three circles fit the points a bit better than the other two circles, thus the best fitting circle will change abruptly.

A similar example was described by Chernov in [3, Section 2.2], where the best fitting line to a given data set of $n = 4$ points is horizontal, but after an arbitrarily small change in the coordinates of the data points, it turns 90° and becomes vertical.

Such examples show that the best fitting object may be extremely sensitive to small numerical errors in the data or round-off errors of the calculation.

2. Uniqueness of the Best Fitting Line

We begin our study of the uniqueness problem with the simplest case—fitting straight lines to data points. We first introduce relevant statistical symbols and notation.

Given data points $(x_1, y_1), \dots, (x_n, y_n)$, we denote by \bar{x} and \bar{y} the *sample means*

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (2)$$

The point (\bar{x}, \bar{y}) is called the center of mass or the *centroid* of the given data set. We also denote by

$$s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2, \quad s_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (3)$$

$$s_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

the components of the so-called “scatter matrix”

$$\mathbf{S} = \begin{bmatrix} s_{xx} & s_{xy} \\ s_{xy} & s_{yy} \end{bmatrix}, \quad (4)$$

which characterizes the “spread” of the data set about its centroid (\bar{x}, \bar{y}) .

This matrix is symmetric and positive semidefinite. The scatter matrix \mathbf{S} defines the so called *scattering ellipse* whose center is (\bar{x}, \bar{y}) and whose axes are spanned by the eigenvectors of the scatter matrix \mathbf{S} (the major axis is spanned by the eigenvector corresponding to the larger eigenvalue).

Next we find the following best fitting line [3, Chapter 2]. We will describe lines in the xy plane by equation

$$Ax + By + C = 0, \quad (5)$$

where A, B , and C are the parameters of the line. Now the best fitting line is found by minimizing the objective function

$$\mathcal{F}(A, B, C) = \frac{1}{A^2 + B^2} \sum_{i=1}^n (Ax_i + By_i + C)^2. \quad (6)$$

The parameters (A, B, C) need only be specified up to a scalar multiple. Thus we can impose constraint $A^2 + B^2 = 1$. Since the parameter C is unconstrained, we can eliminate it, which gives

$$C = -A\bar{x} - B\bar{y}. \quad (7)$$

In particular, we see that the best fitting line always passes through the centroid (\bar{x}, \bar{y}) of the data set. Now the objective function is

$$\begin{aligned} \mathcal{F}(A, B) &= \sum_{i=1}^n [A(x_i - \bar{x}) + B(y_i - \bar{y})]^2 \\ &= s_{xx}A^2 + 2s_{xy}AB + s_{yy}B^2, \end{aligned} \quad (8)$$

or in matrix form

$$\mathcal{F}(\mathbf{A}) = \mathbf{A}^T \mathbf{S} \mathbf{A}, \quad (9)$$

where $\mathbf{A} = (A, B)^T$ denotes the parameter vector. Minimizing (9) subject to the constraint $\|\mathbf{A}\| = 1$ is a simple problem of the matrix algebra; its solution is the eigenvector of the scatter matrix \mathbf{S} corresponding to the smaller eigenvalue.

Observe that the parameter vector \mathbf{A} is orthogonal to the line (5), thus the line itself is parallel to the other eigenvector. In addition, it passes through the centroid, hence it is the major axis of the scattering ellipse.

The above observations are summarized as follows.

Theorem 1. *Every best fitting line $Ax + By + C = 0$ passes through the centroid and coincides with the major axis of the scattering ellipse.*

For typical data sets, the above procedure leads to a unique best fitting line. But there are certain exceptions.

If the two eigenvalues of \mathbf{S} coincide, then every vector $\mathbf{A} \neq \mathbf{0}$ is its eigenvector, and the function $\mathcal{F}(A, B)$ is actually constant on the unit circle $\|\mathbf{A}\| = 1$. In that case all the lines passing through the centroid of the data minimize \mathcal{F} ; hence, the problem has multiple (infinitely many) solutions. This happens if and only if \mathbf{S} is a scalar matrix, that is,

$$s_{xx} = s_{yy}, \quad s_{xy} = 0. \quad (10)$$

The above observations are summarized as follows.

Theorem 2. *A best fitting line is not unique if and only if the eigenvalues of the scatter matrix \mathbf{S} coincide. In this case the scattering ellipse becomes a circle. Moreover, in this case every line passing through the centroid (\bar{x}, \bar{y}) is one of the best fitting lines.*

Thus we have a dichotomy; either there is a *single* best fitting line or there are *infinitely many* best fitting lines. In the latter case, the whole bundle of lines passing through the centroid (\bar{x}, \bar{y}) are best fitting lines.

A simple example of a data set for which there are multiple best fitting lines is n points placed at the vertices of a regular polygon with n vertices (n -gon). Rotating the data set around its center by the angle $2\pi/n$ takes the data set back to itself. So if there is one best fitting line, then by rotating it through the angle $2\pi/n$ we get another line that fits equally well. Thus the best fitting line is not unique.

It is less obvious (but true, according to Theorem 2) that *every* line passing through the center of our regular polygon is a best fitting line; they all minimize the objective function.

Data points placed at vertices of a regular polygon seem like a very exceptional situation. However multiple best fitting lines are much more common. The following is true.

Theorem 3. *Given any data points $(x_1, y_1), \dots, (x_n, y_n)$, one can always move one of them so that the new data set will admit multiple best fitting lines. Precisely, there are always x'_n and y'_n such that the set $(x_1, y_1), \dots, (x_{n-1}, y_{n-1}), (x'_n, y'_n)$ admit multiple best fitting lines.*

TABLE 1

Objects	Existence in all cases	Existence in typical cases	Uniqueness
Lines	Yes	Yes	No
Circles	No	Yes	No
Ellipses	No	No	No
All conics	No	Yes	No

In other words, the $n - 1$ points can be placed arbitrarily, without any regular pattern whatever, and then we can add just one extra point so that the set of all n points will admit multiple best fitting lines, that is, will satisfy (10).

Still, the existence of multiple best fitting lines is a very unlikely event in probabilistic terms. If data points are sampled randomly from an absolutely continuous probability distribution, then this event occurs with probability zero. Indeed, (10) specifies a subsurface (submanifold) in the $2n$ -dimensional space with coordinates $x_1, y_1, \dots, x_n, y_n$. That submanifold has zero volume; hence, for any absolutely continuous probability distribution, its probability is zero.

However, if the data points are obtained from a digital image (say, they are pixels on a computer screen), then the chance of having (10) may no longer be negligible and may have to be reckoned with. For instance, a simple configuration of 4 pixels making a 2×2 square satisfies (10), and thus the orthogonal fitting line is not uniquely defined.

3. Uniqueness of the Best Fitting Circle

We have seen in Section 2 that the simplest fitting problem—that of fitting straight lines—can have multiple solutions, so it may not be too surprising to find out that more complicated problems also can have multiple solutions (we emphasize that the best fitting circle minimizes the sum of squares of *geometric* distances, as defined in the Introduction). Here we demonstrate the multiplicity of the best fit for circles.

However, we cannot describe all data sets for which the best fitting circle is not unique in the same comprehensive manner as we did that for lines in Section 2. We can only give some examples of such data sets.

All the known examples are based on the *rotational symmetry* of the data set. We already used this idea in Section 2. Suppose the data set can be rotated around some point O through the angle $2\pi/k$ for some integer $k \geq 2$, and after the rotation it comes back to itself. Then, if there is a best fitting circle, rotating it around O through the angle $2\pi/k$ would give us another circle that would fit the data set equally well. This is how we get more than one best fitting circle.

This is a nice idea but it breaks down instantly if the center of the best fitting circle happens to coincide with the center of rotation O . Then we would rotate the circle around its own center and obviously would get the same circle again. Thus one has to construct a rotationally symmetric data set more carefully to avoid best fitting circles centered on the natural center of symmetry of the set.

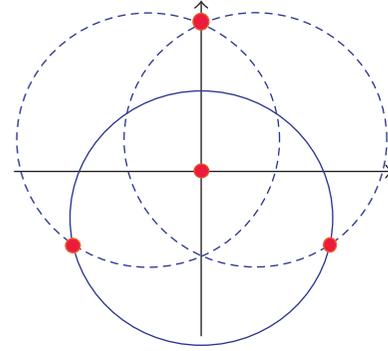


FIGURE 1: Four data points and three fitting circles.

The earliest and simplest example was given by Nievergelt [2]. He chose $n = 4$ data points as follows:

$$(0, 0), \quad (0, 2), \quad (\sqrt{3}, -1), \quad (-\sqrt{3}, -1). \quad (11)$$

The last three points are at the vertices of an equilateral triangle centered on $(0, 0)$. So the whole set can be rotated around the origin $(0, 0)$ through the angle $2\pi/3$, and it will come back to itself.

Nievergelt claimed that the best fitting circle has center $(0, -3/4)$ and radius $R = 7/4$. This circle passes through the last two data points and cuts right in the middle between the first two. So the first two points are at distance $d = 1$ from that circle, and the last two are right on it (their distance from the circle is zero). Thus the objective function is

$$\mathcal{F} = 1^2 + 1^2 + 0^2 + 0^2 = 2. \quad (12)$$

It is easy to believe that Nievergelt's circle is the best, indeed, as any attempt to perturb its center or radius would only make the fit worse (the objective function would grow). However a complete mathematical proof of this claim would be perhaps prohibitively difficult, so we leave it out.

Our goal is actually more modest than finding the best fitting circle in Nievergelt's example. Our goal is to show that *there are* multiple best fitting circles (without finding them explicitly). And the multiplicity here can be proven as follows.

According to our general results [1], for every data set the best fit exists, which may be a circle or a line. If the best object is a circle, then its center is either at $(0, 0)$ or elsewhere. So we have three possible cases: (i) the best fitting object is a line, (ii) the best fitting object is a circle centered on $(0, 0)$, and (iii) the best fitting object is a circle with a center different from $(0, 0)$. In the last case our rotational symmetry will work, as explained above, and prove the multiplicity of the best fitting circle. So we need to rule out the first two cases.

Consider any circle of radius R centered on $(0, 0)$. It is easy to see that the respective objective function is

$$\mathcal{F} = R^2 + 3(2 - R)^2 = 4R^2 - 12R + 12. \quad (13)$$

Its minimum is attained at $R = 3/2$, and its minimum value is

$$\mathcal{F} = \left(\frac{3}{2}\right)^2 + 3\left(\frac{1}{2}\right)^2 = 3. \quad (14)$$

This is larger than $\mathcal{F} = 2$ in (12). Thus circles centered on the origin cannot compete with Nievergelt's circle and should be ruled out.

Next we consider all lines. As we have seen in Section 2, for rotationally symmetric data sets, all the best fitting lines pass through the center. All of those lines fit equally well. Taking the x axis, for example, it is easy to see that the corresponding objective function is

$$\mathcal{F} = 2^2 + 1^2 + 1^2 + 0^2 = 6. \quad (15)$$

This is greater than $\mathcal{F} = 2$ in (12) and even greater than $\mathcal{F} = 3$ in (14). Thus lines are even less competitive than circles centered on the origin, so they are ruled out as well. The proof is finished.

Therefore, the best fitting circle has a center different from $(0, 0)$. Thus by rotating this circle through the angles $2\pi/3$ and $4\pi/3$, we get two more circles that fit the data equally well. So the circle fitting problem has three distinct solutions. The alleged best fitting circles are shown in Figure 1.

After Nievergelt's example, two other papers presented, independently, similar examples of nonunique circle fits.

Chernov and Lesort [4] used a perfect square, instead of Nievergelt's regular triangle. They placed four points at the vertices of the square and another 4 points at its center, so the data set consisted of $n = 8$ points total. Then they used the above strategy to prove that at least four different circles achieve the best fit.

Zelniker and Clarkson [5] used a regular triangle again and placed three points at its vertices and three more points at its center (so that the data set consisted of $n = 6$ points). Then they showed that at least three different circles achieve the best fit.

These examples lead to an interesting fact that may seem rather counterintuitive. Let C be a circle of radius R with center O . Let us place a large number of data points on C and a single data point at the center O . Suppose the points on C are placed uniformly (say at the vertices of a regular polygon). Then it seems like C is an excellent candidate for the best fitting circle—it interpolates all the data points and misses only at O , so $\mathcal{F} = R^2$. It is hard to imagine that any other circle or line can do any better.

However, a striking fact proved by Nievergelt [6, Lemma 7] says that the center of the best fitting circle *cannot* coincide with any data point. Therefore in our example, C cannot be the best fitting circle. Hence some other circle with center $O' \neq O$ fits the data set better. And again, rotating the best circle about O gives other best fitting circles, so those are not unique.

Rotationally symmetric data sets described above are clearly exceptional; small perturbations of data points easily destroy the symmetry. But there are probably many other data sets, without any symmetries, that admit multiple circle fits, too. We believe that they are all unusual and can be easily destroyed by small perturbations. Below is our argument.

Suppose a set of data points P_1, \dots, P_n admits two best circle fits, and denote those circles by C_1 and C_2 . First consider a simple case; C_1 and C_2 are concentric, that is, have a common center, O . Let D_i denote the distance from the

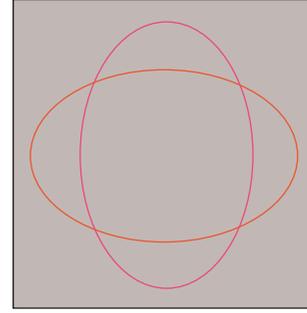


FIGURE 2: The best fit to a uniform distribution in a square.

point P_i to the center O . By direct inspection, for any circle of radius R centered on O the objective function is

$$\mathcal{F} = \sum_{i=1}^n (R - D_i)^2 = nR^2 - 2R \sum_{i=1}^n D_i + \sum_{i=1}^n D_i^2. \quad (16)$$

This is a quadratic polynomial in R , so it cannot have two distinct minima. So the two best fitting circles cannot be concentric.

Now suppose the circles C_1 and C_2 are not concentric, that is, they have distinct centers, O_1 and O_2 . Let L denote the line passing through O_1 and O_2 . Note that the data points cannot be all on the line L (because if the data points were collinear, the best fit would be achieved by the interpolating line and not by two circles). So there exists a point P_i that does not lie on the line L . Hence we can move it slightly toward the circle C_1 but away from the circle C_2 . Then the objective function \mathcal{F} changes slightly, and it will decrease at one minimum (on C_1) and increase at the other (on C_2). This will break the tie and ensure the uniqueness of the global minimum.

4. Uniqueness of the Best Fitting Ellipse

Based on the previous two sections, we should expect that data sets exist for which the best fitting ellipse is not unique. However, we could not find any explicit examples in the literature, so we supply our own.

Our previous paper [1] was the first to provide an example of that sort. We fitted conics to a uniform distribution in a perfect square, $[0, 1] \times [0, 1]$. We found, quite unexpectedly, that the best fit was achieved by *two* distinct ellipses; they were geometrically equal (i.e., they had the same major axis and the same minor axis), and they had a common center, but one was oriented vertically and the other horizontally. See Figure 2.

Strictly speaking, in this example we did not have a data set—we replaced it with a uniform distribution that is obtained as a limit of large samples, as $n \rightarrow \infty$. But we would get the same picture—two best fitting ellipses—if we place $N \times N$ data points in the square arranged as a perfect square lattice (e.g., the points have coordinates $(i/N, j/N)$, where $i = 1, \dots, N$ and $j = 1, \dots, N$).

A more elegant example can be constructed as follows. Recall (Section 3) that Nievergelt's example of multiple fitting circles consisted of $n = 4$ data points; three were placed at

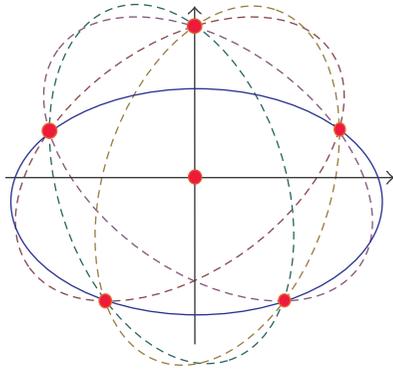


FIGURE 3: Six data points and five fitting ellipses.

vertices of an equilateral triangle and the fourth one at its center.

Note that a circle has three independent parameters, but an ellipse has five. So it is natural to generalize Nievergelt's example by placing *five* data points at vertices of a regular pentagon and the sixth one at its center. Thus we have $n = 6$ data points as follows:

$$(0, 0), \quad (0, 2), \quad \left(\pm 2 \cos \frac{\pi}{10}, 2 \sin \frac{\pi}{10}\right), \quad \left(\pm 2 \cos \frac{3\pi}{10}, -2 \sin \frac{3\pi}{10}\right). \quad (17)$$

We strongly believe that the best fitting ellipse passes through the last four data points and the point $(0, 1)$. These five points determine the ellipse uniquely. It is obviously symmetric about the y axis, so its major axis is horizontal. This ellipse cuts right in the middle between the first two data points. So those two points are at distance $d = 1$ from that ellipse and the last four are right on it (the distance is zero). Thus the objective function is

$$\mathcal{F} = 1^2 + 1^2 + 0^2 + 0^2 + 0^2 + 0^2 = 2. \quad (18)$$

Below we provide a partial proof of our claim that the above ellipse is the best. We also designed a full computer-assisted proof that involves extensive numerical computations.

Lastly, by rotating this ellipse through the angles $2\pi k/5$ for $k = 1, 2, 3, 4$ we get four more ellipses that fit the data equally well. So the ellipse fitting problem has five distinct solutions; see Figure 3.

We will compare our ellipse to the best fitting circle centered on the origin and the best fitting lines. Consider any circle of radius R centered on $(0, 0)$. It is easy to see that the respective objective function is

$$\mathcal{F} = R^2 + 5(2 - R)^2 = 6R^2 - 20R + 20. \quad (19)$$

Its minimum is attained at $R = 5/3$, and its minimum value is

$$\mathcal{F} = \left(\frac{5}{3}\right)^2 + 5\left(\frac{1}{3}\right)^2 = \frac{10}{3}. \quad (20)$$

This is larger than $\mathcal{F} = 2$ in (18). Thus circles centered on the origin cannot compete with our ellipse.

Consider all lines. As we have seen in Section 2, for rotationally symmetric data sets all the best fitting lines pass through the center, and all of those lines fit equally well. Taking the x axis, for example, it is easy to see that the corresponding objective function is

$$\mathcal{F} = 2^2 + 2\left(2 \sin\left(\frac{\pi}{10}\right)\right)^2 + 2\left(2 \sin\left(\frac{3\pi}{10}\right)\right)^2 = 10. \quad (21)$$

This is greater than $\mathcal{F} = 2$ in (18) and even greater than $\mathcal{F} = 10/3$ in (20). Thus lines are even less competitive than circles centered on the origin.

Also, in the ellipse fitting problem, pairs of parallel lines are legitimate model objects; see [1]. We examined the fits achieved by pairs of parallel lines. The best fit we found was by two horizontal lines $y = y_1$ and $y = y_2$, where

$$y_1 = \frac{(2 + 4 \sin(\pi/10))}{4}, \quad y_2 = -2 \sin\left(\frac{3\pi}{10}\right). \quad (22)$$

Note that y_1 is the average y -coordinate of the first four points in our sample. Thus the first line is the best fitting line for the first four points, and the second line passes through the last two points. The objective function for this pair of lines is

$$\mathcal{F} = (2 - y_1)^2 + 2\left(2 \sin\left(\frac{\pi}{10}\right) - y_1\right)^2 + (0 - y_1)^2 \approx 2.146. \quad (23)$$

This is pretty good, better than the best fitting circle in (20). But still it is a little worse than the best fitting ellipse in (18).

Thus our ellipse fits better than any circle centered on the origin, any line, and any pair of parallel lines. In order to conclude that it is really the best fitting ellipse, we would have to compare it to all other ellipses and parabolas. This task seems prohibitively difficult if one uses only theoretical arguments as above. Instead, we developed a computer-assisted proof. It is a part of the Ph.D. thesis by Q. Huang, which we plan to post on the web [7].

References

- [1] N. Chernov, Q. Huang, and H. Ma, "Does the best fitting curve always exist?" *ISRN Probability and Statistics*, Article ID 895178, 2012.
- [2] Y. Nievergelt, "A finite algorithm to fit geometrically all midrange lines, circles, planes, spheres, hyperplanes, and hyperspheres," *Numerische Mathematik*, vol. 91, no. 2, pp. 257–303, 2002.
- [3] N. Chernov, *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*, vol. 117, Chapman & Hall/CRC Monographs on Statistics and Applied Probability, 2010.
- [4] N. Chernov and C. Lesort, "Least squares fitting of circles," *Journal of Mathematical Imaging and Vision*, vol. 23, no. 3, pp. 239–252, 2005.
- [5] E. E. Zelniker and I. V. L. Clarkson, "A statistical analysis of the Delogne-Kâsa method for fitting circles," *Digital Signal Processing*, vol. 16, no. 5, pp. 498–522, 2006.
- [6] Y. Nievergelt, "Perturbation analysis for circles, spheres, and generalized hyperspheres fitted to data by geometric total least-squares," *Mathematics of Computation*, vol. 73, no. 245, pp. 169–180, 2004.
- [7] <http://www.math.uab.edu/~chernov/cl>.