

STATISTICAL ANALYSIS OF CURVE FITTING METHODS IN ERRORS-IN-VARIABLES MODELS

UDC 519.21

A. AL-SHARADQAH AND N. CHERNOV

ABSTRACT. Regression models in which all variables are subject to errors are known as errors-in-variables (EIV) models. The respective parameter estimates have many unusual properties: their exact distributions are very hard to determine, and their absolute moments are often infinite (so that their mean and variance do not exist). In our paper, *Error analysis for circle fitting algorithms*, *Electr. J. Stat.* **3** (2009), 886–911, we developed an unconventional statistical analysis that allowed us to effectively assess EIV parameter estimates and design new methods with superior characteristics. In this paper we validate our approach in a series of numerical tests. We also prove that in the case of fitting circles, the estimates of the parameters are absolutely continuous (have densities).

1. INTRODUCTION

Regression models in which all variables are subject to errors are known as errors-in-variables (EIV) models [8, 12, 19]. The EIV regression problem is quite different (and far more difficult) than the classical regression where the independent variable is assumed to be error-free. The EIV regression, even in the linear case, presents challenging questions and leads to some counterintuitive results (see below). In EIV regression, many issues remain largely unresolved.

This work is a part of a bigger project whose purpose is to revisit some difficulties in the EIV regression studies and develop an unconventional approach to their resolution. Our approach is tailored for image processing applications, where the number of observed points (pixels) is limited but the noise is small. Our general goal is to develop a simplified error analysis that nonetheless allows us to effectively assess the performance of existing fitting algorithms. In addition, based on our approach one can design algorithms with increased accuracy. This work is devoted to an experimental validation of our error analysis scheme. We probe it on several test cases of linear and nonlinear EIV regression and demonstrate that it remains adequate in most cases and identify situations where further investigation may be necessary.

For one nonlinear model, circular regression, we prove theoretically that the estimates of the center and radius have probability densities, provided the observed points have probability distributions with densities. This is a novel fact.

Our paper is organized as follows. In Section 2 we describe a standard EIV model and highlight its principal difficulties. In Section 3 we review basic steps of our error

2010 *Mathematics Subject Classification.* Primary 62H10, 62J02; Secondary 62H35.

Key words and phrases. Errors-in-variables, regression, curve fitting, circle fitting, line fitting, functional model.

The second author was partially supported by National Science Foundation, grant DMS-0652896.

analysis. In Section 4 we test it on several model cases. In Appendix we derive some technical formulas.

2. EIV REGRESSION MODEL

Functional model. Suppose we are fitting a curve described by an implicit equation $P(x, y; \Theta) = 0$, where $\Theta = (\theta_1, \dots, \theta_k)^T$ represents a vector of unknown parameters. It is standard to assume that the observed points $(x_1, y_1), \dots, (x_n, y_n)$ are noisy images of some true points $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$ that lie on the true curve, i.e. satisfy

$$(1) \quad P(\tilde{x}_i, \tilde{y}_i, \tilde{\Theta}) = 0, \quad i = 1, \dots, n,$$

where $\tilde{\Theta}$ denotes the true value of the vector parameter. The coordinates of the true points $(\tilde{x}_i, \tilde{y}_i)$ are not random; they can be treated as extra parameters (in addition to $\theta_1, \dots, \theta_k$) that one may want to estimate. This set of assumptions is known as a *functional model* in the EIV regression analysis.

We assume that the noise is Gaussian and isotropic, i.e.

$$(2) \quad x_i = \tilde{x}_i + \delta_i, \quad y_i = \tilde{y}_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where δ_i, ε_i represent independent normal variables having distribution $N(0, \sigma^2)$ where σ^2 is an unknown parameter. In image processing it is natural to assume that the errors δ_i and ε_i are uncorrelated and have a common variance, though in other studies the pair $(\delta_i, \varepsilon_i)$ is allowed to have a more general covariance matrix (but we will not consider correlated noise here).

Under the above assumptions, the Maximum Likelihood Estimate (MLE) of Θ is obtained by minimizing $\mathcal{F}(\Theta) = \sum d_i^2$, where the d_i denote the geometric (orthogonal) distances from the observed points to the fitting curve (see a proof in [9]). This procedure is called *geometric fit* or *orthogonal distance regression* (ODR).

Fitting lines. Suppose we fit a straight line $y = \alpha + \beta x$ to observed points. Then $\Theta = (\alpha, \beta)$ and the objective function takes the form

$$(3) \quad \mathcal{F}(\alpha, \beta) = \sum d_i^2 = \frac{1}{1 + \beta^2} \sum (y_i - \alpha - \beta x_i)^2.$$

Its minimum is attained at

$$(4) \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad \text{and} \quad \hat{\beta} = \frac{s_{yy} - s_{xx} + \sqrt{(s_{yy} - s_{xx})^2 + 4s_{xy}^2}}{2s_{xy}},$$

where we use standard notation for sample means $\bar{x} = n^{-1} \sum x_i$ and $\bar{y} = n^{-1} \sum y_i$ and for the components of the so-called ‘‘scatter matrix’’:

$$(5) \quad s_{xx} = \sum (x_i - \bar{x})^2, \quad s_{yy} = \sum (y_i - \bar{y})^2, \quad s_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}).$$

The formula (4) holds whenever $s_{xy} \neq 0$, which is true almost surely. The geometric fit (3)–(4) was proposed in the late 1800s [1] and since then has been widely used in practice.

Only in 1976 explicit formulas were derived for the density functions of the estimates $\hat{\alpha}$ and $\hat{\beta}$; see [4, 5]. It turns out that those densities are not normal and do not belong to any standard family of probability densities. Those formulas are overly complicated, involve double-infinite series, and it was promptly noted [4] that they were not very useful for practical purposes.

It was also pointed out [4] that the estimates $\hat{\alpha}$ and $\hat{\beta}$ do not have finite moments, i.e. $E(|\hat{\alpha}|) = \infty$ and $E(|\hat{\beta}|) = \infty$. As a result, they have infinite mean squared errors! These facts pose immediate methodological questions.

- (Q1) How can we characterize, in practical terms, the accuracy of estimates whose theoretical MSE is infinite (and whose bias is undefined)?
- (Q2) Is there any precise meaning to the widely accepted notion that the MLE $\hat{\alpha}$ and $\hat{\beta}$ are best?

The main goal of our unconventional error analysis is to answer these and related questions; see below.

Fitting circles. Suppose now we fit a circle $(x - a)^2 + (y - b)^2 - R^2 = 0$ to observed points. Then $\Theta = (a, b, R)$ and the objective function takes the form

$$(6) \quad \mathcal{F}(a, b, R) = \sum d_i^2 = \sum [\sqrt{(x_i - a)^2 + (y_i - b)^2} - R]^2.$$

It is known [16, 28, 31] that, under certain general conditions, the minimum of (6) exists and is unique. But the minimization of (6) is a nonlinear problem that has no closed form solution. Thus there are no explicit formulas for the MLE \hat{a} , \hat{b} , \hat{R} . There are no known formulas for their densities either. (Even the existence of densities is a novel fact that we prove here in this paper.) Hence the situation here is even more difficult than it is for the linear model.

Furthermore, it was recently discovered [13] that the MLE \hat{a} , \hat{b} , \hat{R} have infinite moments, too, i.e. $\mathbf{E}(|\hat{a}|) = \infty$, $\mathbf{E}(|\hat{b}|) = \infty$, and $\mathbf{E}(|\hat{R}|) = \infty$. Thus one faces the same methodological questions as in the linear case.

It appears that the nonexistence of moments is a general phenomenon in the EIV regression analysis; some other instances of it were recently reported in [10] and [31]. We believe that if one fits an ellipse to observed points, then the MLE of its geometric parameters (center, focuses, axes) would have infinite moments, too, but this is yet to be proven mathematically.

In the case of lines, one can easily circumvent the lack of moments by choosing different parameters, e.g. by representing a line as $Ax + By + C = 0$ under a natural constraint $A^2 + B^2 = 1$; then the MLE \hat{A} and \hat{B} of the new parameters A and B would obviously have finite moments. But when one fits circles or ellipses, it is harder to find convenient parameters whose estimates would have finite moments; see some attempts in [16]. Besides in many applications the estimates of the center and radius are of natural interest and then one has to deal with their infinite moments.

Asymptotic models. Satisfactory answers to our questions (Q1) and (Q2) can be given in the context of various asymptotic models because then the MLE become optimal in some limit. In traditional statistics, it is common to take the limit $n \rightarrow \infty$ (“large sample model”). Then it can be shown that the MLE $\hat{\alpha}$ and $\hat{\beta}$ given by (4) are asymptotically efficient in the sense of Hajek bounds [18]. They are also optimal in other ways; see Gleser [17] and a survey [11]. We note that the so-called adjusted least squares estimators of α and β also have infinite moments [10], and they are efficient in the sense of Hajek bounds, too [26].

On the other hand, Anderson and Sawa [4, 5] investigated the asymptotic properties of the MLE $\hat{\alpha}$ and $\hat{\beta}$ assuming that n was fixed and $\sigma \rightarrow 0$. They called this limiting regime *small-sigma model*. It turns out that the small-sigma model is especially suitable for image processing and computer vision applications. On an image, the number of observed points (pixels on a computer screen) n is usually strictly limited, but the noise level σ is small (so that a naked eye can recognize the desired shape; otherwise a prior filtering should be applied and outliers should be removed). In image processing experiments, the number of observed points normally varies between 10–20 (on the low end) and a few hundred (on the high end). The noise σ is usually less than 5% of the size of the image [6], and in many cases it is below 1%. See also recent Kanatani’s papers [23]

presenting strong arguments in favor of using the “small-sigma” limit in computer vision applications.

We recall that for the linear regression (3), explicit formulas for the distributions of the MLE (4) are available. Using them, Anderson and Sawa [4, 5] treated σ as a small parameter and employed Taylor expansion (up to σ^4) to derive approximations for those distributions. Those approximations had finite moments, which could be regarded as “virtual” moments of the estimates (4) that characterize their accuracy, in practical terms. Anderson and Sawa [5] described their approximations as “virtually exact”, for small σ .

In nonlinear regression models, such as circles or ellipses, no explicit formulas for the distributions of the MLEs are known. We have developed an alternative approach to asymptotic approximations, which is described in the next section. Our approach works for arbitrary nonlinear models and for general estimates (not just MLE).

3. ERROR ANALYSIS

3.1. General scheme. Here we present our error analysis of curve fitting methods following [2]. For convenience we use vector notation

$$\mathbf{x} = (x_1, \dots, x_n)^T, \quad \mathbf{y} = (y_1, \dots, y_n)^T, \quad \boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T, \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T.$$

Then (2) can be written as $\mathbf{x} = \tilde{\mathbf{x}} + \boldsymbol{\delta}$ and $\mathbf{y} = \tilde{\mathbf{y}} + \boldsymbol{\varepsilon}$. We denote the “combined” noise vector by $\mathbf{h} = (\delta_1, \dots, \delta_n, \varepsilon_1, \dots, \varepsilon_n)^T$.

Let $\hat{\boldsymbol{\Theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$ be any estimate of the unknown parameters. For each scalar parameter $\theta = \theta_j$, $1 \leq j \leq k$ and its estimator $\hat{\theta} = \hat{\theta}_j$ we use the Taylor expansion

$$(7) \quad \hat{\theta}(\mathbf{x}, \mathbf{y}) = \hat{\theta}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) + \mathbf{G}^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T \mathbf{H} \mathbf{h} + \mathcal{O}_P(\sigma^3).$$

Here $\mathbf{G} = \nabla \hat{\theta}$ denotes the gradient vector (of the first order partial derivatives) and $\mathbf{H} = \nabla^2 \hat{\theta}$ the Hessian matrix (of the second order partial derivatives), all taken at the true point $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$. The remainder term $\mathcal{O}_P(\sigma^3)$ in (7) is a random variable \mathcal{R} such that $\sigma^{-3} \mathcal{R}$ is bounded in probability (for $\sigma < \sigma_0$). The second and third terms in (7) have typical values of order σ and σ^2 , respectively, and their moments are always finite (because \mathbf{h} is a Gaussian vector).

The expansion (7) is valid when the function $\hat{\theta}(\mathbf{x}, \mathbf{y})$ satisfies standard regularity conditions (more precisely, when it has a continuous third derivative). Such conditions can be verified by the implicit value theorem; see [2] for circle estimators.

If there is no noise, i.e. $\sigma = 0$, then the observed points lie right on the true curve, i.e. $(\mathbf{x}, \mathbf{y}) = (\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$. In that case the MLE (obtained by minimizing the geometric distances) will obviously return the true curve; hence $\hat{\theta}(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \hat{\theta}$. In fact, every decent fitting method should return the true curve in the absence of noise (as Kanatani [24] said, other methods are not worth considering). Thus the expansion (7) can be rewritten as

$$(8) \quad \Delta \hat{\theta}(\mathbf{x}, \mathbf{y}) = \mathbf{G}^T \mathbf{h} + \frac{1}{2} \mathbf{h}^T \mathbf{H} \mathbf{h} + \mathcal{O}_P(\sigma^3),$$

where $\Delta \hat{\theta}(\mathbf{x}, \mathbf{y}) = \hat{\theta}(\mathbf{x}, \mathbf{y}) - \hat{\theta}$ is the error of the parameter estimate.

The first term in (8), i.e. $\mathbf{G}^T \mathbf{h}$, is a linear combination of i.i.d. normal random variables (the components of \mathbf{h}) that have zero mean. Hence it is itself a normal random variable with zero mean, we denote it by $\Delta_1 \hat{\theta}$. Since $\Delta_1 \hat{\theta}$ is a linear transformation of \mathbf{h} , it is of order σ . In the same pattern, the second term is a quadratic form of i.i.d. normal variables, denoted by $\Delta_2 \hat{\theta}$; it is of order σ^2 . Accordingly, we define two approximations

to $\hat{\theta}$: the *linear approximation*

$$(9) \quad \hat{\theta}_L = \tilde{\theta} + \Delta_1 \hat{\theta},$$

and the *quadratic approximation*

$$(10) \quad \hat{\theta}_Q = \tilde{\theta} + \Delta_1 \hat{\theta} + \Delta_2 \hat{\theta}.$$

The distribution of $\hat{\theta}_L$ is normal with mean θ and variance $\sigma^2 \mathbf{G}^T \mathbf{G}$. Thus, to the leading order, the estimate $\hat{\theta}$ can be characterized as “unbiased” and having “variance” $\sigma^2 \mathbf{G}^T \mathbf{G}$ (at the same time, the theoretical bias and variance of θ may not exist; this leaves us wondering how good our approximations are – and it is the issue discussed below).

At any rate, our approximation (9) gives us a quantitative measure of the accuracy of $\hat{\theta}$. If we denote $\mathbf{G}_i = \nabla \hat{\theta}_i$ for $1 \leq i \leq k$, then the entire vector estimate $\hat{\boldsymbol{\theta}}$ can be characterized by the covariance matrix $\sigma^2 \mathbf{V}$ where $\mathbf{V}_{ij} = \mathbf{G}_i^T \mathbf{G}_j$ for $1 \leq i, j \leq k$. This is our (partial) answer to question (Q1) in Section 2; see more of it below.

KCR lower bound and bias reduction. The matrix \mathbf{V} has a natural lower bound, i.e. $\mathbf{V} \geq \mathbf{V}_{\min}$ (in the sense that $\mathbf{V} - \mathbf{V}_{\min}$ is a positive semidefinite matrix), and there are explicit formulas for \mathbf{V}_{\min} ; see [15]. This fact was discovered and proved for unbiased estimates by Kanatani [21, 22] and for general estimates by Chernov and Lesort [15]; they called it the Kanatani–Cramer–Rao (KCR) lower bound.

The matrix \mathbf{V} for the MLE $\hat{\boldsymbol{\theta}}_{\text{MLE}}$ (which is obtained by minimizing geometric distances) always attains the KCR bound; see [15] and an earlier work by Amemiya and Fuller [3] where an explicit formula for \mathbf{V} was derived. In this precise sense the MLE is optimal, which gives a (partial) answer to question (Q2) in Section 2.

But is it a complete answer? In the nonlinear regression, there are several popular circle fitting methods (by Kåsa, Pratt, and Taubin; see references in [2, 16]). It turns out that they all (!) attain the KCR bound; see [15]. Many different ellipse fitting methods (such as FNS, HEIV, and the renormalization scheme; see references in [25]) attain the KCR bound, too (though some others, such as various algebraic fits, do not attain the KCR bound). Thus in order to distinguish between different fits attaining the KCR bound (and optimize their performances) one has to employ the quadratic approximation (10). Its last term $\Delta_2 \hat{\theta}$ is a quadratic form of normal random variables; it can be written as

$$(11) \quad \Delta_2 \hat{\theta} = \frac{1}{2} \sigma^2 \sum_{i=1}^{2n} d_i Z_i^2,$$

where the Z_i are i.i.d. standard normal random variables ($Z_i \sim N(0, 1)$) and the d_i are the eigenvalues of \mathbf{H} . Now we can approximate the bias of $\hat{\theta}$, i.e. $\mathbf{E}(\hat{\theta}) - \tilde{\theta}$, by

$$(12) \quad \text{bias}(\hat{\theta}) = \mathbf{E}(\hat{\theta}) - \tilde{\theta} \approx \frac{1}{2} \sigma^2 \sum d_i = \frac{1}{2} \sigma^2 \text{tr } \mathbf{H}.$$

This formula allows us to compare and optimize estimates which already minimize the covariance matrix \mathbf{V} . In particular, it turns out that the geometric circle fit has a very small bias, while other popular circle fits (by Taubin, Pratt, and Kåsa) have larger biases (in the order of increasing magnitude); see [2]. So the geometric circle fit is the most accurate among all popular circle fits (this also answers our question (Q2) in Section 2).

However, there is no natural minimum for the bias (12). In fact a novel circle fit was recently designed in [30] (based on our error analysis), for which the bias vanishes to the leading order, i.e. $\text{tr } \mathbf{H} = 0$. Thus our analysis allows one to design new and better methods which may outperform the geometric fit (previously regarded as unbeatable).

See the experimental evidence of the superior performance of the novel circle fits in [2, 30]. For the ellipse fitting problem, such work is currently in progress.

We note that linear approximations (9) and the respective minimization of the covariance matrix \mathbf{V} have been a dominant approach in the 1990s and early 2000s. Only very recent papers [2, 25, 30] are devoted to a consistent use of higher order expansions.

4. VALIDATION OF OUR APPROXIMATIVE FORMULAS

The main goal of this paper is to examine the accuracy of the approximations (9) and (10) numerically. We will see if they are good enough for typical image processing applications (where σ is small).

General criterion. We use the following criterion. Let $f(x)$ denote the density of an estimate $\hat{\theta}$ and $f_A(x)$ the density of its approximation $\hat{\theta}_L$ or $\hat{\theta}_Q$ (accordingly, f_A will also be denoted by f_L or f_Q). We say that the approximation $f_A(x)$ is good enough if it accounts for “almost all” of $f(x)$ in the sense that

$$(13) \quad f(x) = (1 - p)f_A(x) + pf_R(x), \quad -\infty < x < \infty,$$

where $f_R(x)$ is some other density function (the “remainder”) and $p > 0$ is sufficiently small. In this case we can think of the estimate $\hat{\theta}$ as a weighted combination of two random variables: a “good” one with density f_A and a “bad” one with density f_R . According to (13), the realizations of $\hat{\theta}$ are taken from the “good” distribution f_A with probability $1 - p$ and from the “bad” distribution f_R with probability p . If p is small (such as $p = 0.01$), then in practical terms the estimate $\hat{\theta}$ behaves almost like the ‘good’ random variable with density f_A . If, however, p is large (say, $p > 0.1$), then f_A does not constitute a good approximation. So the borderline between good and bad approximations may be set to 0.05. Thus there is a natural similarity between our p-value and the textbook P-value in modern statistics, for which 0.05 is a commonly used borderline between the null and alternative hypotheses.

The form of the remainder density f_R is irrelevant for our criterion; thus we only need to make sure that $f_R(x) \geq 0$ for all x . This leads us to the following formula for computing the “minimal” (i.e., “optimal”) p :

$$(14) \quad p_A = \min_{0 \leq p \leq 1} \{f(x) - (1 - p)f_A(x) \geq 0 \text{ for all } -\infty < x < \infty\}.$$

In our tests, we will use $f_A = f_L$ and $f_A = f_Q$, so we get two p ’s, p_L and p_Q . They will characterize the accuracy of the linear and quadratic approximations (9) and (10), respectively.

We have computed p_L and p_Q for two regression models: the linear regression (3)–(4) and nonlinear (circular) regression (6). To compute (14) we need the exact density $f(x)$ of the estimate $\hat{\theta}$ and its approximative density $f_A(x)$. The exact density is only known for the linear model $y = \alpha + \beta x$, but even in that case its formula is impractical (see above), so we have computed the empirical density $f(x)$ by using Monte-Carlo simulations.

The linear approximation density $f_L(x)$ is normal, so we just used an exact formula for it. The quadratic approximation density $f_Q(x)$ depends on the more complicated random variable (11). There exists an explicit formula for the density of (11), but that formula is too cumbersome for practical use—it involves doubly infinite series with coefficients being integrals of confluent hypergeometric functions; see [29]. Various approximations to that density have been developed, including Taylor expansions [27], numerical Fourier inversions [20], and saddle point approximations. But all such approximations are too computationally intense, subject to divergence, and heavily depend on initial guesses.

Therefore we again resorted to a simple Monte-Carlo simulation and constructed an empirical density; it turned out to be the cheapest way of achieving the desired accuracy.

For each density function we simulated 10^8 – 10^9 values of the corresponding random variable and constructed a histogram with 10^3 bins covering the interval

$$(\mu' - 3\sigma', \mu' + 3\sigma'),$$

where μ' and σ' were estimated mean value and standard deviation of the corresponding variable. Then we applied a standard smoothing procedure using Gaussian kernels [7]. After that we computed p_A by (14) with the help of an adapted bisection algorithm.

Linear regression. The MLE $\hat{\beta}$ of the slope of the regression line is given by (4). Our error analysis gives the following approximations (see the proofs in Appendix). Let us denote by $\tilde{x}_j^* = \tilde{x}_j - \frac{1}{n} \sum \tilde{x}_i$ the ‘centered’ x -coordinates of the true points, $1 \leq j \leq n$, and use vector notation $\tilde{\mathbf{x}}^* = (\tilde{x}_1^*, \dots, \tilde{x}_n^*)$. Then the linear approximation is $\hat{\beta}_L = \tilde{\beta} + \Delta_1\beta$ with

$$(15) \quad \Delta_1\beta = \mathbf{G}^T \mathbf{h}, \quad \mathbf{G}_i = \begin{cases} -\tilde{\beta}\tilde{x}_i^*/\|\tilde{\mathbf{x}}^*\|^2, & \text{for } 1 \leq i \leq n, \\ \tilde{x}_{i-n}^*/\|\tilde{\mathbf{x}}^*\|^2, & \text{for } n+1 \leq i \leq 2n. \end{cases}$$

The quadratic approximation $\hat{\beta}_Q$ contains an extra term

$$(16) \quad \Delta_2\beta = \mathbf{h}^T \mathbf{N}^T \mathbf{Q} \mathbf{N} \mathbf{h}.$$

Here \mathbf{Q} is a $(2n) \times (2n)$ matrix given by

$$(17) \quad \mathbf{Q} = -\frac{1}{2\|\tilde{\mathbf{x}}^*\|^2} \begin{pmatrix} \frac{g-2\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2} \mathbf{Z}_n + (g\tilde{\beta}^2 + 2\tilde{\beta})\mathbf{I}_n & \frac{g\tilde{\beta}+2}{\|\tilde{\mathbf{x}}^*\|^2} \mathbf{Z}_n - (g\tilde{\beta} + 1)\mathbf{I}_n \\ \frac{g\tilde{\beta}+2}{\|\tilde{\mathbf{x}}^*\|^2} \mathbf{Z}_n - (g\tilde{\beta} + 1)\mathbf{I}_n & \frac{-g}{\|\tilde{\mathbf{x}}^*\|^2} \mathbf{Z}_n + g\mathbf{I}_n \end{pmatrix},$$

where \mathbf{I}_n denotes the $n \times n$ identity matrix, $\mathbf{Z}_n = \tilde{\mathbf{x}}^*(\tilde{\mathbf{x}}^*)^T$, $g = -2\tilde{\beta}/(1 + \tilde{\beta}^2)$, and \mathbf{N} is a $(2n) \times (2n)$ matrix defined by

$$(18) \quad \mathbf{N} = \begin{pmatrix} \mathbf{N}_n & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{N}_n \end{pmatrix}, \quad \mathbf{N}_n = \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n$$

and $\mathbf{0}_n$ and $\mathbf{1}_n$ denote $n \times n$ matrices consisting of zeroes and ones, respectively. All these formulas are derived in Appendix. Using these formulas we have simulated random values of $\hat{\beta}_L$ and $\hat{\beta}_Q$ and constructed their empirical densities.

Now we turn to our experimental results, i.e. present the computed values p_L and p_Q . They depend, generally, on several factors: (i) the true values of the model parameters (in this case α and β), (ii) the number and location of the true points, and (iii) the noise level σ . Since the geometric fit is invariant under translations, the value of α is irrelevant. For β , we tested four values: $\beta = 0$ (horizontal line, the easiest case), and $\beta = 1, 2, 10$ (the steepest line, $\beta = 10$, was the hardest to fit accurately). For each case we positioned $n = 10$ equally spaced true points on the line (spanning an interval of length $L = 1$; i.e., the distance between the first and last true point was one). The noise level σ was varied from 0 up to the point when the values of p became too large. We note that the geometric fit is invariant under scaling of coordinates x and y ; hence the p_L and p_Q only depend on the relative noise level σ/L (and in image processing applications, σ/L is usually below 0.05).

Our results are shown in Figure 1. It presents plots of p_L (the dashed line) and p_Q (the solid line) versus the noise parameter σ . The figure shows that for the simplest case $\beta = 0$ both approximations remain adequate up to a substantial noise $\sigma = 0.1$. But for more difficult cases $\beta = 1, 2, 10$ the linear approximation is only good for very small σ , and in the most challenging situation $\beta = 10$ it practically breaks down. The quadratic approximation remains good up to $\sigma = 0.05$ in all cases except the last one ($\beta = 10$).

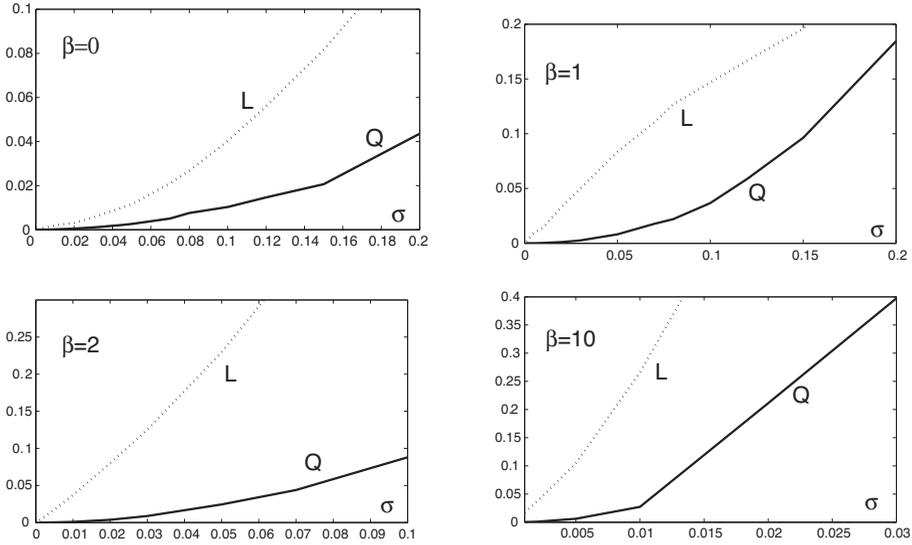


FIGURE 1. The p_L (dashed line) and p_Q (solid line) versus σ .

In that case it, too, becomes inaccurate when noise is not very small (roughly, when $\sigma > 0.01$).

Our experiments show that linear approximation is rather crude; it only works when the noise is very small, even by the image processing standards. The quadratic approximation is acceptable in most realistic cases, but not in all. In difficult situations, such as $\beta = 10$, it becomes less than adequate.

Circular regression. Let $\hat{\Theta} = (\hat{a}, \hat{b}, \hat{R})$ denote the MLE of the circle parameters (center and radius). We denote by $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ two vectors whose components are defined by

$$(19) \quad \tilde{u}_i = (\tilde{x}_i - \tilde{a})/\tilde{R}, \quad \tilde{v}_i = (\tilde{y}_i - \tilde{b})/\tilde{R}, \quad i = 1, \dots, n.$$

Let $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ denote $n \times n$ diagonal matrices whose main diagonals are $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$, respectively. Also let

$$(20) \quad \mathbf{W} = \begin{bmatrix} \tilde{u}_1 & \tilde{v}_1 & 1 \\ \vdots & \vdots & \vdots \\ \tilde{u}_n & \tilde{v}_n & 1 \end{bmatrix}.$$

Then our approximations are given by

$$(21) \quad \Delta_1 \hat{\Theta} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T (\tilde{\mathbf{U}} \boldsymbol{\delta} + \tilde{\mathbf{V}} \boldsymbol{\varepsilon}),$$

and

$$(22) \quad \Delta_2 \hat{\Theta} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{F},$$

where $\mathbf{F} = (f_1, \dots, f_n)^T$ is a vector with components

$$f_i = \tilde{u}_i(\delta_i - \Delta_1 a) + \tilde{v}_i(\varepsilon_i - \Delta_1 b) - \Delta_1 R + \frac{\tilde{v}_i^2}{2\tilde{R}}(\delta_i - \Delta_1 a)^2 + \frac{\tilde{u}_i^2}{2\tilde{R}}(\varepsilon_i - \Delta_1 b)^2 - \frac{\tilde{u}_i \tilde{v}_i}{\tilde{R}}(\delta_i - \Delta_1 a)(\varepsilon_i - \Delta_1 b).$$

All these formulas are derived in [2]. Using these formulas we have simulated random values of $\hat{\Theta}_L$ and $\hat{\Theta}_Q$ and constructed their empirical densities. We did that separately for the radius estimate \hat{R} and for the center estimates \hat{a} and \hat{b} .

Fact (existence of densities). The existence of the densities of the estimates \hat{a} , \hat{b} , and \hat{R} is a highly nontrivial fact that requires a proof. We provide it in Appendix.

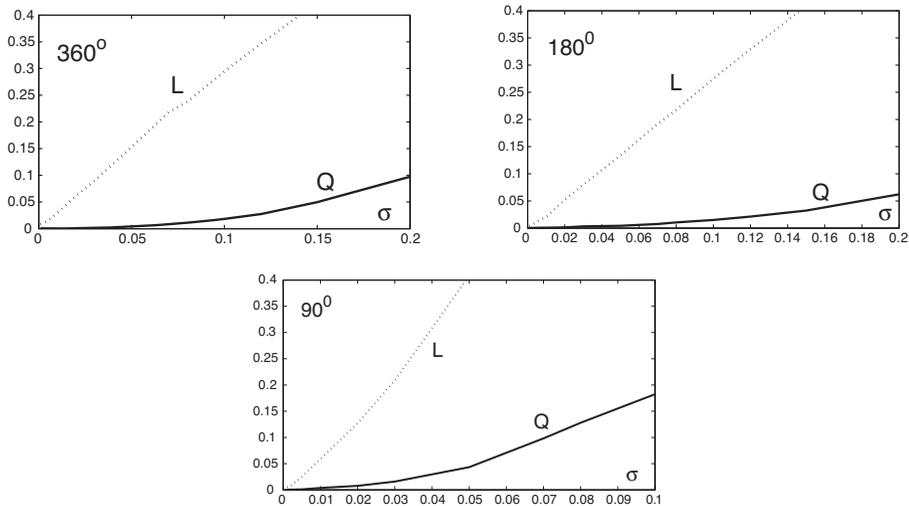


FIGURE 2. The p_L (dashed line) and p_Q (solid line) versus σ ; this figure corresponds to the radius estimate.

Next we turn to our experimental results, i.e. present the computed values p_L and p_Q . Again, they depend on (i) the true values of the circle parameters, (ii) the number and location of the true points, and (iii) the noise level σ . Since the geometric fit is invariant under translations, rotations, and scaling, it is enough to set $\tilde{a} = \tilde{b} = 0$ and $\tilde{R} = 1$. We positioned $n = 10$ equally spaced true points on an arc whose size was set to 360° (full circle), 180° (semi-circle), and 90° (quarter of a circle). The first case was the easiest for the fitting purposes, and the last one the hardest. The noise level σ was varied from 0 up to the point when the values of p_L and p_Q became too large. Due to the scaling invariance the p_L and p_Q only depend on the relative noise level σ/\tilde{R} (and in image processing applications, σ/\tilde{R} is usually below 0.05).

The results for the radius estimate are shown in Figure 2 and those for the center estimates in Figure 3. First we note that for the radius estimate the linear approximation \hat{R}_L is far worse than it is for the center estimate, i.e. for \hat{a}_L and \hat{b}_L . This happens because the radius estimate is known [2] to have a significant bias $\mathcal{O}(\sigma^2)$, which the linear approximation completely misses. The center estimate is known [2] to have a much smaller bias, this explains why the linear approximations \hat{a}_L and \hat{b}_L are quite close to the quadratic approximations \hat{a}_Q and \hat{b}_Q (at least for the full circle and semi-circle cases).

In other respects, the picture here is similar to the one we have seen for the linear regression. The linear approximation is only good for very small σ . The quadratic approximation remains good up to $\sigma = 0.05$ in all cases, but for the quarter circle it is barely adequate when $\sigma = 0.05$. Apparently, for smaller arcs its accuracy will further deteriorate, and quadratic approximation will not stay adequate anymore.

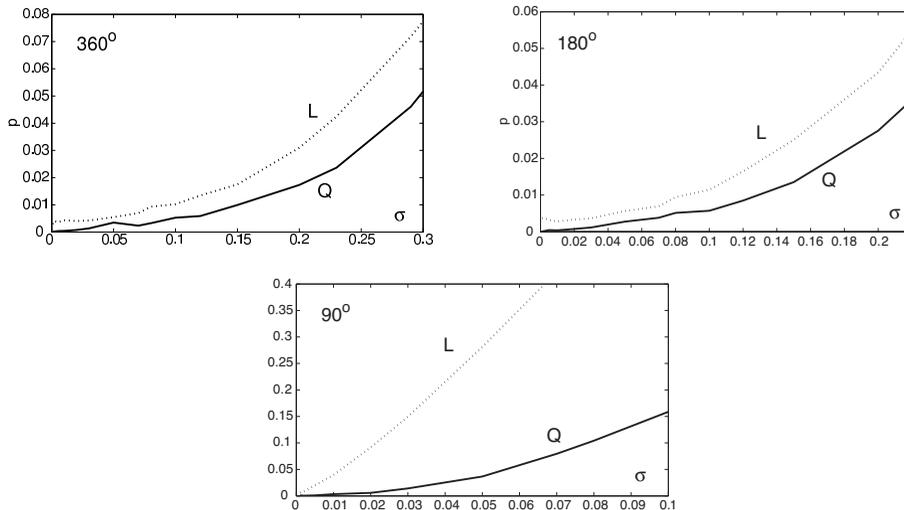


FIGURE 3. The p_L (dashed line) and p_Q (solid line) versus σ ; this figure corresponds to the center estimate.

5. CONCLUSIONS

Here we summarize our conclusions.

- The linear approximation is only good for very small σ . For typical noise level in image processing applications the linear approximation is not quite accurate. Thus, the optimization of fitting methods to the ‘leading order’, i.e. the minimization of their variance matrix $\sigma^2 \mathbf{V}$ (as described in Section 3) can only ensure their optimality for very small noise. For a more realistic noise level, the minimization of \mathbf{V} is insufficient, and one needs to use the quadratic approximation for further improvements.
- In typical image processing applications, the optimization of fitting methods must consist of two steps: (i) minimizing the variance matrix \mathbf{V} and (ii) reducing the bias (i.e. reducing the terms coming from the quadratic approximation), according to our description in Section 3. Ideally, at step (ii) one should eliminate the $\mathcal{O}(\sigma^2)$ terms in the bias and leave only $\mathcal{O}(\sigma^4)$ terms.
- In the more difficult cases (such as fitting steep lines or small circular arcs), when estimates of the parameters are quite unstable, even the quadratic model may not be accurate enough. Then one may have to develop further expansion, up to σ^3 or σ^4 ; we regard such expansions as possible objectives for future studies.

APPENDIX

First we derive our formulas (15)–(18).

Linear approximation $\hat{\beta}_L$. It is clear that the factor $(1 + \beta^2)^{-1}$ in (3) can be replaced with its true value in our linear approximation. Eliminating α from the objective function (3) and keeping only terms of order σ^2 yields

$$\begin{aligned}
 \mathcal{F}(\beta) &= \frac{1}{1 + \tilde{\beta}^2} \sum (\tilde{y}_i^* + \varepsilon_i^* - (\tilde{\beta} + \Delta_1 \beta)(\tilde{x}_i^* + \delta_i^*))^2 + \mathcal{O}_P(\sigma^3) \\
 (23) \quad &= \frac{1}{1 + \tilde{\beta}^2} \sum (\varepsilon_i^* - \tilde{\beta} \delta_i^* - \tilde{x}_i^* \Delta_1 \beta)^2 + \mathcal{O}_P(\sigma^3),
 \end{aligned}$$

where $\delta_i^* = \delta_i - \bar{\delta}$ and $\varepsilon_i^* = \varepsilon_i - \bar{\varepsilon}$ denote the ‘centered’ errors. Now the objective function attains the minimum at

$$(24) \quad \Delta_1 \beta = (\tilde{\mathbf{x}}^*)^T (\boldsymbol{\varepsilon}^* - \tilde{\beta} \boldsymbol{\delta}^*) / \|\tilde{\mathbf{x}}^*\|^2,$$

where $\boldsymbol{\delta}^*$ and $\boldsymbol{\varepsilon}^*$ denote the vectors of δ_i^* ’s and ε_i^* ’s, respectively. Let \mathbf{h}^* denote the combined vector of δ_i^* ’s and ε_i^* ’s. The components of \mathbf{h}^* are not independent random variables. However, due to a simple relation $\mathbf{h}^* = \mathbf{N}\mathbf{h}$, we can express $\Delta_1 \beta$ as a linear function of a random vector \mathbf{h} whose components are independent:

$$(25) \quad \Delta_1 \beta = [-\tilde{\beta}(\tilde{\mathbf{x}}^*)^T \mathbf{N}_n : (\tilde{\mathbf{x}}^*)^T \mathbf{N}_n] \mathbf{h} / \|\tilde{\mathbf{x}}^*\|^2 = (\tilde{\mathbf{x}}^*)^T [-\tilde{\beta} \mathbf{I}_n : \mathbf{I}_n] \mathbf{h} / \|\tilde{\mathbf{x}}^*\|^2,$$

where we used the relation $(\tilde{\mathbf{x}}^*)^T \mathbf{N}_n = (\tilde{\mathbf{x}}^*)^T$. Note that $\Delta_1 \beta$ has normal distribution $N(0, \sigma^2(1 + \tilde{\beta}^2) / \|\tilde{\mathbf{x}}^*\|^2)$.

Quadratic approximation $\hat{\beta}_Q$. In this case we have to expand $(1 + \beta^2)^{-1}$:

$$\frac{1}{1 + \beta^2} = f_0(\tilde{\beta}) + f_1(\tilde{\beta}) \Delta_1 \beta + f_1(\tilde{\beta}) \Delta_2 \beta + f_2(\tilde{\beta}) \Delta_1 \beta^2 + \mathcal{O}_P(\sigma^3) \stackrel{\text{def}}{=} C + \mathcal{O}(\sigma^3),$$

where

$$f_0(\tilde{\beta}) = \frac{1}{1 + \tilde{\beta}^2}, \quad f_1(\tilde{\beta}) = -\frac{2\tilde{\beta}}{(1 + \tilde{\beta}^2)^2}, \quad f_2(\tilde{\beta}) = -\frac{1 - 3\tilde{\beta}^2}{(1 + \tilde{\beta}^2)^3}.$$

Now, keeping only terms of order σ^4 yields

$$(26) \quad \mathcal{F}(\beta) = \sum (\tilde{\beta} \delta_i^* + \tilde{x}_i^* \Delta_1 \beta + \tilde{x}_i^* \Delta_2 \beta + \delta_i^* \Delta_1 \beta - \varepsilon_i^*)^2 C + \mathcal{O}_P(\sigma^5).$$

We can find $\Delta_2 \beta$ by setting the derivative $\partial \mathcal{F} / \partial \Delta_2 \beta = 0$. To this end, we use the fact that $\sum (\tilde{\beta} \delta_i^* + \tilde{x}_i^* \Delta_1 \beta - \varepsilon_i^*) \tilde{x}_i^* = 0$:

$$\begin{aligned} \frac{d\mathcal{F}(\beta)}{d\Delta_2 \beta} &= \sum (f_1(\tilde{\beta}) d_i^2 + 2d_i x_i^* C) \\ &= f_1(\tilde{\beta}) \sum (\tilde{\beta} \delta_i^* + \tilde{x}_i^* \Delta_1 \beta - \varepsilon_i^*)^2 + 2 \sum d_i \tilde{x}_i^* C + 2 \sum d_i \delta_i^* C \\ &= f_1(\tilde{\beta}) \sum (\tilde{\beta} \delta_i^* + \tilde{x}_i^* \Delta_1 \beta - \varepsilon_i^*)^2 + 2f_0(\tilde{\beta}) \|\tilde{\mathbf{x}}^*\|^2 \Delta_2 \beta \\ &\quad + 4f_0(\tilde{\beta}) \sum \tilde{x}_i^* \delta_i^* \Delta_1 \beta + 2f_0(\tilde{\beta}) \sum \delta_i^* (\tilde{\beta} \delta_i^* - \varepsilon_i^*) \end{aligned}$$

where we omit the remainder term $\mathcal{O}_P(\sigma^3)$ for brevity. Let us denote

$$g = \frac{f_1(\tilde{\beta})}{f_0(\tilde{\beta})} = \frac{-2\tilde{\beta}}{1 + \tilde{\beta}^2}.$$

Then we arrive at

$$(27) \quad \begin{aligned} -2\|\tilde{\mathbf{x}}^*\|^2 \Delta_2 \beta &= g \sum (\tilde{\beta} \delta_i^* - \varepsilon_i^* + \tilde{x}_i^* \Delta_1 \beta)^2 + 4 \sum \tilde{x}_i^* \delta_i^* \Delta_1 \beta + 2 \sum \delta_i^* (\tilde{\beta} \delta_i^* - \varepsilon_i^*) \\ &= -2g \|\tilde{\mathbf{x}}^*\|^2 \Delta_1 \beta^2 + g \|\tilde{\mathbf{x}}^*\|^2 \Delta_1 \beta^2 + g \sum (\tilde{\beta} \delta_i^* - \varepsilon_i^*)^2 \\ &\quad + 2 \sum \delta_i^* (\tilde{\beta} \delta_i^* - \varepsilon_i^*) + 4 \sum \delta_i^* \tilde{x}_i^* \Delta_1 \beta \\ &\stackrel{\text{def}}{=} \mathbf{I} + \mathbf{II} + \mathbf{III} \end{aligned}$$

(**I**, **II**, and **III** are defined and simplified below). In matrix notation, each term in the

previous expression is a quadratic form of \mathbf{h}^* :

$$(28) \quad \begin{aligned} \mathbf{I} &= -g\|\tilde{\mathbf{x}}^*\|^2\Delta_1\beta^2 = (\mathbf{h}^*)^T \begin{pmatrix} \frac{-g\tilde{\beta}^2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n & \frac{g\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n \\ \frac{g\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n & \frac{-g}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n \end{pmatrix} \mathbf{h}^*, \\ \mathbf{II} &= g\sum(\tilde{\beta}\delta_i^* - \varepsilon_i^*)^2 + 2\sum\delta_i^*(\tilde{\beta}\delta_i^* - \varepsilon_i^*) \\ &= (\mathbf{h}^*)^T \begin{pmatrix} (g\tilde{\beta}^2 + 2\tilde{\beta})\mathbf{I}_n & -(g\tilde{\beta} + 1)\mathbf{I}_n \\ -(g\tilde{\beta} + 1)\mathbf{I}_n & g\mathbf{I}_n \end{pmatrix} \mathbf{h}^* \end{aligned}$$

and

$$(29) \quad \mathbf{III} = 4\sum\delta_i^*\tilde{x}_i^*\Delta_1\beta = (\mathbf{h}^*)^T \begin{pmatrix} -\frac{4\tilde{\beta}}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n & \frac{2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n \\ \frac{2}{\|\tilde{\mathbf{x}}^*\|^2}\mathbf{Z}_n & \mathbf{0}_n \end{pmatrix} \mathbf{h}^*.$$

Combining equations (27)–(29) gives

$$\Delta_2\beta = \mathbf{h}^{*T}\mathbf{Q}\mathbf{h}^* = \mathbf{h}^T\mathbf{N}_n^T\mathbf{Q}\mathbf{N}_n\mathbf{h},$$

where \mathbf{Q} is already defined in (17).

Existence of densities for the circle fit. Here we prove that the circle parameter estimates \hat{a} , \hat{b} , and \hat{R} are absolutely continuous random variables, i.e. they have densities.

We only need to assume that the error vector $\mathbf{h} = (\delta_1, \dots, \delta_n, \varepsilon_1, \dots, \varepsilon_n)$ has an absolutely continuous distribution in \mathbb{R}^{2n} (the independence of its components and the exact type of their distributions are irrelevant). Now the estimates $(\hat{a}, \hat{b}, \hat{R})$ constitute a map from \mathbb{R}^{2n} to \mathbb{R}^3 , which is defined provided the best fitting circle exists and is unique. It has been pointed out by several authors [16, 28, 31] that the set of ‘data’ $(x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{R}^{2n}$ for which the best fitting circle fails to exist or is not unique has Lebesgue measure zero; thus the probability of such an event is zero. A precise proof can be found, for example, in [14, Section 3.9].

Now we have a map, call it G , from \mathbb{R}^{2n} to \mathbb{R}^3 that is defined at every point $X \in \mathbb{R}^{2n}$ where the estimator $(\hat{a}, \hat{b}, \hat{R})$ exists and is unique. It induces a probability distribution, call it μ , in \mathbb{R}^3 . If the latter is not absolutely continuous, then there is a set $A \subset \mathbb{R}^3$ such that $\text{Leb}(A) = 0$ but $\mu(A) > 0$. Our further arguments involve a bit of Lebesgue measure theory. By the Lebesgue decomposition theorem, $\mu = \mu_0 + \mu_1$, where μ_0 is singular and μ_1 is absolutely continuous. Let $A_0 \subset \mathbb{R}^3$ be a carrier of μ_0 , i.e., a set satisfying $\text{Leb}(A_0) = 0$, $\mu_0(A_0) > 0$, and $\mu_0(\mathbb{R}^3 \setminus A_0) = 0$. We note that $\mu(A_0) > 0$ implies $\text{Leb}(G^{-1}(A_0)) > 0$.

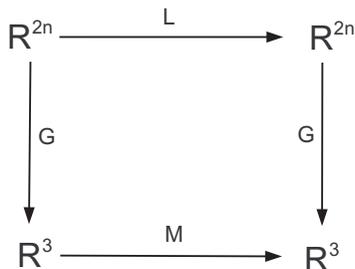


FIGURE 4. This diagram commutes, i.e., $G \circ L = M \circ G$.

Next, suppose we shift all the data points (x_i, y_i) by a vector $(\alpha, \beta) \in \mathbb{R}^2$. Then the center (\hat{a}, \hat{b}) is shifted by (α, β) as well (because the best fitting circle is invariant

under translations). If we expand (or contract) the set of data points (homothetically) by $\Lambda > 0$, i.e., transform $x_i \mapsto \Lambda x_i$ and $y_i \mapsto \Lambda y_i$ for all $i = 1, \dots, n$, then $\hat{a}, \hat{b}, \hat{R}$ change to $\Lambda \hat{a}, \Lambda \hat{b}, \Lambda \hat{R}$. Combining the above two types of transformations in \mathbb{R}^{2n} we get a family of transformations, $L_{\alpha, \beta, \Lambda}$, acting by

$$(x_i, y_i) \mapsto (\Lambda(x_i + \alpha), \Lambda(y_i + \beta)) \quad \forall i = 1, \dots, n$$

where $\Lambda > 0$. Let $M_{\alpha, \beta, \Lambda}$ be a transformation in \mathbb{R}^3 acting by

$$M_{\alpha, \beta, \Lambda}(a, b, R) \mapsto (\Lambda(a + \alpha), \Lambda(b + \beta), \Lambda R).$$

It follows that $G \circ L_{\alpha, \beta, \Lambda} = M_{\alpha, \beta, \Lambda} \circ G$; hence

$$G^{-1}(M_{\alpha, \beta, \Lambda}(A_0)) = L_{\alpha, \beta, \Lambda}(G^{-1}(A_0))$$

(see Figure 4). We will only use $L_{\alpha, \beta, \Lambda}$ for $\alpha, \beta \approx 0$ and $\Lambda \approx 1$, so we denote $\gamma = \Lambda - 1$ and treat (α, β, γ) as a small vector and denote $A_{\alpha, \beta, \gamma} = M_{\alpha, \beta, 1+\gamma}(A_0)$. Since $\text{Leb}(G^{-1}(A_0)) > 0$, we also have $\text{Leb}(L_{\alpha, \beta, 1+\gamma}(G^{-1}(A_0))) > 0$. Thus for all small α, β, γ we have $\mu(M_{\alpha, \beta, 1+\gamma}(A_0)) = \mu(A_{\alpha, \beta, \gamma}) > 0$. Since $\text{Leb}(A_{\alpha, \beta, \gamma}) = 0$, we have $\mu_0(A_{\alpha, \beta, \gamma}) = \mu(A_{\alpha, \beta, \gamma}) > 0$.

It remains to show that the above situation is impossible, i.e., for a finite singular measure μ_0 concentrated on a null set A_0 we cannot have $\mu_0(A_{\alpha, \beta, \gamma}) > 0$ for all small α, β, γ . Let us assume that $\mu_0(A_{\alpha, \beta, \gamma}) > 0$ for all small α, β, γ . Let χ_0 be the indicator function of A_0 , i.e., $\chi_0(x) = 1$ for $x \in A_0$ and $\chi_0(x) = 0$ for $x \in \mathbb{R}^3 \setminus A_0$. Then for all small α, β, γ ,

$$(30) \quad \mu_0(A_{\alpha, \beta, \gamma}) = \mu_0(A_0 \cap A_{\alpha, \beta, \gamma}) = \int_{\mathbb{R}^{2n}} \chi_0(x) \chi_0(M_{\alpha, \beta, 1+\gamma}^{-1}(x)) d\mu_0 > 0.$$

On the other hand, by the Fubini theorem,

$$(31) \quad \int \mu_0(A_{\alpha, \beta, \gamma}) d\alpha d\beta d\gamma = \int_{\mathbb{R}^{2n}} \chi_0(x) \left[\int \chi_0(M_{\alpha, \beta, 1+\gamma}^{-1}(x)) d\alpha d\beta d\gamma \right] d\mu_0 = 0$$

because the inner integral vanishes (recall that $\text{Leb}(A_0) = 0$). Equations (30) and (31) contradict each other; thus our claim is proved.

BIBLIOGRAPHY

1. R. J. Adcock, *Note on the method of least squares*, *Analyst* **4** (1877), 183–184.
2. A. Al-Sharadqah and N. Chernov, *Error analysis for circle fitting algorithms*, *Electr. J. Stat.* **3** (2009), 886–911. MR2540845 (2010j:62170)
3. Y. Amemiya and W. A. Fuller, *Estimation for the nonlinear functional relationship*, *Annals Statist.* **16** (1988), 147–160. MR924862 (90c:62065)
4. T. W. Anderson, *Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics*, *J. R. Statist. Soc. B* **38** (1976), 1–36. MR0411025 (53:14764)
5. T. W. Anderson and T. Sawa, *Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient*, *J. R. Statist. Soc. B* **44** (1982), 52–62. MR655374 (84j:62023)
6. M. Berman, *Large sample bias in least squares estimators of a circular arc center and its radius*, *CVGIP: Image Understanding* **45** (1989), 126–128.
7. A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Oxford Statistical Science Series, vol. 18, Oxford University Press, 1997.
8. R. J. Carroll, D. Ruppert, L. A. Stefansky, and C. M. Crainiceanu, *Measurement Error in Nonlinear Models: A Modern Perspective*, Chapman & Hall, London, 2006. MR2243417 (2007e:62004)
9. N. N. Chan, *On circular functional relationships*, *J. R. Statist. Soc. B* **27** (1965), 45–56. MR0189163 (32:6590)

10. C.-L. Cheng and A. Kukush, *Non-existence of the first moment of the adjusted least squares estimator in multivariate errors-in-variables model*, *Metrika* **64** (2006), 41–46. MR2242556 (2008b:62042)
11. C.-L. Cheng and J. W. Van Ness, *On estimating linear relationships when both variables are subject to errors*, *J. R. Statist. Soc. B* **56** (1994), 167–183. MR1257805
12. C.-L. Cheng and J. W. Van Ness, *Statistical Regression with Measurement Error*, Arnold, London, 1999. MR1719513 (2001k:62001)
13. N. Chernov, *Fitting circles to scattered data: parameter estimates have no moments*, *Metrika* **73** (2011), 373–384. MR2785031 (2012h:62252)
14. N. Chernov, *Circular and Linear Regression: Fitting Circles and Lines by Least Squares*, Monographs on Statistics & Applied Probability, vol. 117, CRC Press, Boca Raton–London–New York, 2010. MR2723019 (2012a:62005)
15. N. Chernov and C. Lesort, *Statistical efficiency of curve fitting algorithms*, *Comp. Stat. Data Anal.* **47** (2004), 713–728. MR2101548 (2005f:62038)
16. N. Chernov and C. Lesort, *Least squares fitting of circles*, *J. Math. Imag. Vision* **23** (2005), 239–251. MR2181705 (2007d:68165)
17. L. J. Gleser, *Functional, structural and ultrastructural errors-in-variables models*, *Proc. Bus. Econ. Statist. Sect. Am. Statist. Ass.*, 1983, pp. 57–66.
18. R. Z. Hasminskii and I. A. Ibragimov, *On Asymptotic Efficiency in the Presence of an Infinite-dimensional Nuisance Parameter*, *Lecture Notes in Math*, vol. 1021, Springer, Berlin, 1983, pp. 195–229. MR735986 (85h:62040)
19. S. van Huffel (ed.), *Total Least Squares and Errors-in-Variables Modeling*, Kluwer, Dordrecht, 2002. MR1951009 (2003g:00026)
20. J. P. Imhof, *Computing the distribution of quadratic forms in normal variables*, *Biometrika* **4** (1961), 419–426. MR0137199 (25:655)
21. K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, Amsterdam, 1996. MR1392697 (97k:62133)
22. K. Kanatani, *Cramer–Rao lower bounds for curve fitting*, *Graph. Mod. Image Process.* **60** (1998), 93–99.
23. K. Kanatani, *For geometric inference from images, what kind of statistical model is necessary?*, *Syst. Comp. Japan* **35** (2004), 1–9.
24. K. Kanatani, *Optimality of maximum likelihood estimation for geometric fitting and the KCR lower bound*, *Memoirs Fac. Engin. Okayama Univ.* **39** (2005), 63–70.
25. K. Kanatani, *Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis*, *Int. J. Computer Vision* **80** (2008), 167–188.
26. A. Kukush and E.-O. Maschke, *The efficiency of adjusted least squares in the linear functional relationship*, *J. Multivar. Anal.* **87** (2003), 261–274. MR2016938 (2004m:62155)
27. A. M. Mathai and S. B. Provost, *Quadratic Forms in Random Variables*, Marcel Dekker, New York, 1992. MR1192786 (94g:62110)
28. Y. Nievergelt, *A finite algorithm to fit geometrically all midrange lines, circles, planes, spheres, hyperplanes, and hyperspheres*, *J. Numerische Math.* **91** (2002), 257–303. MR1900920 (2003d:52011)
29. J. Robinson, *The distribution of a general quadratic form in normal variables*, *Austral. J. Statist.* **7** (1965), 110–114. MR0198584 (33:6739)
30. P. Rangarajan and K. Kanatani, *Improved algebraic methods for circle fitting*, *Electr. J. Statist.* **3** (2009), 1075–1082. MR2557129 (2011b:68220)
31. E. Zelniker and V. Clarkson, *A statistical analysis of the Delogne–Kåsa method for fitting circles*, *Digital Signal Proc.* **16** (2006), 498–522.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALABAMA AT BIRMINGHAM, BIRMINGHAM, ALABAMA 35294

E-mail address: `alsha1aa@gmail.com`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ALABAMA AT BIRMINGHAM, BIRMINGHAM, ALABAMA 35294

E-mail address: `chernov@math.uab.edu`

Received 4/MAR/2010
Originally published in English