Error analysis for circle fitting algorithms

A. Al-Sharadqah[1] and N. Chernov[1]

Abstract

We study the problem of fitting circles (or circular arcs) to data points observed with errors in both variables. A detailed error analysis for all popular circle fitting methods – geometric fit, Kåsa fit, Pratt fit, and Taubin fit – is presented. Our error analysis goes deeper than the traditional expansion to the leading order. We obtain higher order terms, which show exactly why and by how much circle fits differ from each other. Our analysis allows us to construct a new algebraic (non-iterative) circle fitting algorithm that outperforms all the existing methods, including the (previously regarded as unbeatable) geometric fit.

Keywords: least squares fit, curve fitting, circle fitting, algebraic fit, error analysis, variance, bias, functional model.

1 Introduction

Fitting circles and circular arcs to observed points is one of the basic tasks in pattern recognition and computer vision, nuclear physics, and other areas [5, 9, 11, 23, 24, 27, 30, 32]. Many algorithms have been developed that fit circles to data. Some minimize the geometric distances from the circle to the data points (we call them geometric fits). Others minimize various approximate (or 'algebraic') distances, they are called algebraic fits. We overview most popular algorithms in Sections 3–4.

Geometric fit is commonly regarded as the most accurate, but it can only be implemented by iterative schemes that are computationally intensive and

[1]Department of Mathematics, University of Alabama at Birmingham, Birmingham, AL 35294; chernov@math.uab.edu; alsha1aa@uab.edu

subject to occasional divergence. Algebraic fits are faster but presumably less precise. At the same time the assessments on their accuracy are solely based on practical experience, no one has performed a detailed theoretical comparison of the accuracy of various circle fits. It was shown in [8] that all the circle fits have the same covariance matrix, to the leading order, in the small-noise limit. Thus the differences between various fits can only be revealed by a higher-order error analysis.

The purpose of this paper is to do just that. We employ higher-order error analysis (a similar analysis was used by Kanatani [22] in the context of more general quadratic models) and show exactly why and by how much the geometric circle fit outperforms the algebraic circle fits in accuracy; we also compare the precision of different algebraic fits. Section 5 presents our error analysis in a general form, which can be readily applied to other curve fitting problems.

Finally, our analysis allows us to develop a new algebraic fit whose accuracy exceeds that of the geometric fit. Its superiority is demonstrated by numerical experiments.

2 Statistical model

We adopt a standard *functional model* in which data points $(x_1, y_1), \dots, (x_n, y_n)$ are noisy observations of some *true points* $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_n, \tilde{y}_n)$, i.e.

$$x_i = \tilde{x}_i + \delta_i, \qquad y_i = \tilde{y}_i + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1}$$

where $(\delta_i, \varepsilon_i)$ represent isotropic Gaussian noise. Precisely, δ_i's and ε_i's are i.i.d. normal random variables with mean zero and variance σ^2.

The true points $(\tilde{x}_i, \tilde{y}_i)$ are supposed to lie on a 'true circle', i.e. satisfy

$$(\tilde{x}_i - \tilde{a})^2 + (\tilde{y}_i - \tilde{b})^2 = \tilde{R}^2, \qquad i = 1, \dots, n, \tag{2}$$

where $(\tilde{a}, \tilde{b}, \tilde{R})$ denote the 'true' (unknown) parameters. Therefore

$$\tilde{x}_i = \tilde{a} + \tilde{R}\cos\varphi_i, \qquad \tilde{y}_i = \tilde{b} + \tilde{R}\sin\varphi_i, \tag{3}$$

where $\varphi_1, \dots, \varphi_n$ specify the locations of the true points on the true circle. The angles $\varphi_1, \dots, \varphi_n$ are regarded as fixed unknowns and treated as additional parameters of the model (called incidental or latent parameters). For brevity we denote

$$\tilde{u}_i = \cos\varphi_i = (\tilde{x}_i - \tilde{a})/\tilde{R}, \qquad \tilde{v}_i = \sin\varphi_i = (\tilde{y}_i - \tilde{b})/\tilde{R}. \tag{4}$$

Note that $\tilde{u}_i^2 + \tilde{v}_i^2 = 1$ for every i.

Remark. In our paper δ_i and ε_i have common variance σ^2, i.e. our noise is homoscedastic. In many studies the noise is heteroscedastic [25, 35], i.e. the normal vector $(\delta_i, \varepsilon_i)$ has point-dependent covariance matrix $\sigma^2 C_i$, where C_i is known and depends on i, and σ^2 is an unknown factor. Our analysis can be extended to this case, too, but the resulting formulas will be somewhat more complex, so we leave it out.

3 Geometric circle fits

A standard approach to fitting circles to 2D data is based on orthogonal least squares, it is also called *geometric fit*, or *orthogonal distance regression* (ODR). It minimizes the function

$$\mathcal{F}(a, b, R) = \sum d_i^2, \tag{5}$$

where d_i stands for the distance from (x_i, y_i) to the circle, i.e.

$$d_i = r_i - R, \qquad r_i = \sqrt{(x_i - a)^2 + (y_i - b)^2}, \tag{6}$$

where (a, b) denotes the center, and R the radius of the circle.

In the context of the functional model, the geometric fit returns the maximum likelihood estimates (MLE) of the circle parameters [6], i.e.

$$(\hat{a}_{\mathrm{MLE}}, \hat{b}_{\mathrm{MLE}}, \hat{R}_{\mathrm{MLE}}) = \operatorname{argmin} \mathcal{F}(a, b, R). \tag{7}$$

A major concern with the geometric fit is that the above minimization problem has no closed form solution. All practical algorithms of minimizing $\mathcal{F}$ are iterative; some implement a general Gauss-Newton [6, 15] or Levenberg-Marquardt [9] schemes, others use circle-specific methods proposed by Landau [24] and Späth [30]. The performance of iterative algorithms heavily depends on the choice of the initial guess. They often take dozens or hundreds of iterations to converge, and there is always a chance that they would be trapped in a local minimum of $\mathcal{F}$ or diverge entirely. These issues are explored in [9].

A peculiar feature of the maximum likelihood estimates $(\hat{a}, \hat{b}, \hat{R})$ of the circle parameters is that they have infinite moments [7], i.e.

$$\mathbb{E}(|\hat{a}|) = \mathbb{E}(|\hat{b}|) = \mathbb{E}(\hat{R}) = \infty \tag{8}$$

for any set of true values $(\tilde{a}, \tilde{b}, \tilde{R})$; here $\mathbb{E}$ denotes the mean value. This happens because the distributions of these estimates have somewhat heavy tails, even though those tails barely affect the practical performance of the MLE (the same happens when one fits straight lines to data with errors in both variables [2, 3]).

To ensure the existence of moments one can adopt a different parameter scheme. An elegant scheme was proposed by Pratt [27] and others [13], which describes circles by an algebraic equation

$$A(x^2 + y^2) + Bx + Cy + D = 0 \tag{9}$$

with an obvious constraint $A \neq 0$ (otherwise this equation describes a line) and a less obvious constraint $B^2 + C^2 - 4AD > 0$. The necessity of the latter can be seen if one rewrites equation (9) as

$$\left(x - \frac{B}{2A}\right)^2 + \left(y - \frac{C}{2A}\right)^2 - \frac{B^2 + C^2 - 4AD}{4A^2} = 0. \tag{10}$$

It is clear now that (9) defines a circle if and only if $B^2 + C^2 - 4AD > 0$.

As the parameters (A, B, C, D) only need to be determined up to a scalar multiple, it is natural to impose a constraint

$$B^2 + C^2 - 4AD = 1, \tag{11}$$

because it automatically ensures $B^2+C^2-4AD > 0$. The constraint (11) was first proposed by Pratt [27]. Under this constraint, the parameters A, B, C, D are essentially bounded, see [9], and their maximum likelihood estimates can be shown to have finite moments.

The equation (9), under the constraint (11), conveniently describes all circles and lines (the latter are obtained when $A = 0$); the inclusion of lines is necessary to ensure the existence of the least squares solution [9, 26, 37].

After one estimates the algebraic circle parameters A, B, C, D, they can be converted to the natural parameters via

$$a = -\frac{B}{2A}, \qquad b = -\frac{C}{2A}, \qquad R^2 = \frac{B^2 + C^2 - 4AD}{4A^2}. \tag{12}$$

4 Algebraic circle fits

An alternative to the complicated geometric fit is made by fast non-iterative procedures called *algebraic fits*. We describe three most popular algebraic circle fits below.

Kåsa fit. One can find a circle by minimizing the function

$$\begin{aligned}\mathcal{F}_{\mathrm{K}} &= \sum (r_i^2 - R^2)^2 \\ &= \sum (x_i^2 + y_i^2 - 2ax_i - 2by_i + a^2 + b^2 - R^2)^2. \end{aligned} \tag{13}$$

In other words, one minimizes $\mathcal{F}_{\mathrm{K}} = \sum f_i^2$, where $f_i = r_i^2 - R^2$ is the so called *algebraic distance* from the point (x_i, y_i) to the circle. A change of parameters $B = -2a$, $C = -2b$, $D = a^2 + b^2 - R^2$ transforms (13) to a linear least squares problem minimizing

$$\mathcal{F}_{\mathrm{K}} = \sum (z_i + Bx_i + Cy_i + D)^2, \tag{14}$$

where we denote $z_i = x_i^2 + y_i^2$ for brevity (we intentionally omit symbol A here to make our formulas consistent with the subsequent ones). Now the problem reduces to a system of linear equations (normal equations) with respect to B, C, D that can be easily solved, and then one recovers the natural circle parameters a, b, R via (12).

This method was introduced in the 1970s by Delogne [11] and Kåsa [23], and then rediscovered and published independently by many authors, see references in [9]. It remains popular in practice. We call it *Kåsa fit*.

The Kåsa method is perhaps the fastest circle fit, but its accuracy suffers when one observes incomplete circular arcs (partially occluded circles); then the Kåsa fit is known to be heavily biased toward small circles [9]. The reason for the bias is that the algebraic distances f_i provide a poor approximation to the geometric distances d_i; in fact,

$$f_i = (r_i - R)(r_i + R) = d_i(2R + d_i) \approx 2Rd_i, \tag{15}$$

hence the Kåsa fit minimizes $\mathcal{F}_{\mathrm{K}} \approx 2R^2 \sum d_i^2$, and it often favors smaller circles minimizing R^2 rather than the distances d_i.

Pratt fit. To improve the performance of the Kåsa method one can minimize another function, $\mathcal{F} = \frac{1}{4R^2}\mathcal{F}_{\mathrm{K}}$, which provides a better approximation to $\sum d_i^2$. This new function, expressed in terms of A, B, C, D reads

$$\mathcal{F}_{\mathrm{P}} = \sum \frac{[Az_i + Bx_i + Cy_i + D]^2}{B^2 + C^2 - 4AD}, \tag{16}$$

due to (12). Equivalently, one can minimize

$$\mathcal{F}(A, B, C, D) = \sum [Az_i + Bx_i + Cy_i + D]^2 \tag{17}$$

subject to the constraint (11). This method was proposed by Pratt [27].

Taubin fit. A slightly different method was proposed by Taubin [32] who minimizes the function

$$\mathcal{F}_{\mathrm{T}} = \frac{\sum\big[(x_i-a)^2+(y_i-b)^2-R^2\big]^2}{4n^{-1}\sum\big[(x_i-a)^2+(y_i-b)^2\big]}. \tag{18}$$

Expressing it in terms of A, B, C, D gives

$$\mathcal{F}_{\mathrm{T}} = \sum \frac{[Az_i+Bx_i+Cy_i+D]^2}{n^{-1}\sum[4A^2z_i+4ABx_i+4ACy_i+B^2+C^2]}. \tag{19}$$

Equivalently, one can minimize (17) subject to a new constraint

$$4A^2\bar{z}+4AB\bar{x}+4AC\bar{y}+B^2+C^2=1. \tag{20}$$

Here we use standard 'sample means' notation: $\bar{x}=\frac{1}{n}\sum x_i$, etc.

General remarks. Note that the minimization of (17) must use some constraint, to avoid a trivial solution $A=B=C=D=0$. Pratt and Taubin fits utilize constraints (11) and (20), respectively. Kåsa fit also minimizes (17), but subject to constraint $A=1$.

While the Pratt and Taubin estimates of the parameters A, B, C, D have finite moments, the corresponding estimates of a, b, R have infinite moments, just like the MLE (8). On the other hand, Kåsa's estimates of a, b, R have finite moments whenever $n \geq 4$; see [37].

All the above circle fits have an important property – they are independent of the choice of the coordinate system, i.e. their results are invariant under translations and rotations; see a proof in [14].

Practical experience shows that the Pratt and Taubin fits are more stable and accurate than the Kåsa fit, and they perform nearly equally well, see [9]. Taubin [32] intended to compare his fit to Pratt's theoretically, but no such analysis was ever published. We make such a comparison below.

There are many other approaches to the circle fitting problem in the modern literature [4, 10, 35, 28, 29, 31, 33, 36, 38], but most of them are either quite slow or can be reduced to one of the algebraic fits [14, Chapter 8].

Matrix representation. We can represent the above three algebraic fits in matrix form. Let $\mathbf{A}=(A,B,C,D)$ denote the parameter vector,

$$\mathbf{Z} \stackrel{\text{def}}{=} \begin{bmatrix} z_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ z_n & x_n & y_n & 1 \end{bmatrix} \tag{21}$$

the 'data matrix' (recall that $z_i = x_i^2 + y_i^2$) and

$$\mathbf{M} \stackrel{\text{def}}{=} \frac{1}{n}\mathbf{Z}^T\mathbf{Z} = \begin{bmatrix} \overline{zz} & \overline{zx} & \overline{zy} & \bar{z} \\ \overline{zx} & \overline{xx} & \overline{xy} & \bar{x} \\ \overline{zy} & \overline{xy} & \overline{yy} & \bar{y} \\ \bar{z} & \bar{x} & \bar{y} & 1 \end{bmatrix} \tag{22}$$

the 'matrix of moments'. All the algebraic circle fits minimize the same objective function $\mathcal{F}(\mathbf{A}) = \mathbf{A}^T\mathbf{M}\mathbf{A}$, cf. (17), subject to a constraint $\mathbf{A}^T\mathbf{N}\mathbf{A} = 1$, where the matrix $\mathbf{N}$ corresponds to the fit. The Pratt fit uses

$$\mathbf{N} = \mathbf{P} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 \end{bmatrix}, \tag{23}$$

the Taubin fit uses

$$\mathbf{N} = \mathbf{T} \stackrel{\text{def}}{=} \begin{bmatrix} 4\bar{z} & 2\bar{x} & 2\bar{y} & 0 \\ 2\bar{x} & 1 & 0 & 0 \\ 2\bar{y} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{24}$$

and the Kåsa uses $\mathbf{N} = \mathbf{K} \stackrel{\text{def}}{=} \mathbf{e}_1\mathbf{e}_1^T$, where $\mathbf{e}_1 = (1, 0, 0, 0)^T$.

Reduction to a generalized eigenvalue problem. To solve the above constrained minimization problem one uses a Lagrange multiplier η and reduces it to an unconstrained minimization of the function

$$\mathcal{G}(\mathbf{A}, \eta) = \mathbf{A}^T\mathbf{M}\mathbf{A} - \eta(\mathbf{A}^T\mathbf{N}\mathbf{A} - 1). \tag{25}$$

Differentiating with respect to $\mathbf{A}$ and η gives

$$\mathbf{M}\mathbf{A} = \eta\mathbf{N}\mathbf{A} \tag{26}$$

and

$$\mathbf{A}^T\mathbf{N}\mathbf{A} = 1, \tag{27}$$

thus $\mathbf{A}$ must be a generalized eigenvector for the matrix pair $(\mathbf{M}, \mathbf{N})$, which also satisfies $\mathbf{A}^T\mathbf{N}\mathbf{A} = 1$. The problem (26)–(27) may have several solutions. To choose the right one we note that for each solution $(\eta, \mathbf{A})$

$$\mathbf{A}^T\mathbf{M}\mathbf{A} = \eta\mathbf{A}^T\mathbf{N}\mathbf{A} = \eta, \tag{28}$$

thus for the purpose of minimizing $\mathbf{A}^T\mathbf{M}\mathbf{A}$ we should choose the solution of (26)–(27) with the smallest η. Note that η is automatically non-negative, since $\mathbf{M} = \frac{1}{n}\mathbf{Z}^T\mathbf{Z}$ is a positive semi-definite matrix.

Since multiplying $\mathbf{A}$ by a scalar does not change the circle it represents, it is common in practical applications to require that $\|\mathbf{A}\| = 1$, and accordingly one needs to replace the constraint (27) with $\mathbf{A}^T\mathbf{N}\mathbf{A} > 0$; the latter can be further relaxed as follows.

For generic data sets, $\mathbf{M}$ is positive definite. Thus if $(\eta, \mathbf{A})$ is any solution of the generalized eigenvalue problem (26), then $\mathbf{A}^T\mathbf{M}\mathbf{A} > 0$. In that case, due to (28), we have $\mathbf{A}^T\mathbf{N}\mathbf{A} > 0$ if and only if $\eta > 0$. Thus it is enough to solve the problem (26) and choose the smallest positive η and the corresponding unit vector $\mathbf{A}$.

The singular case and summary. The matrix $\mathbf{M}$ is singular if and only if the observed points lie on a circle (or a line); in this case the eigenvector $\mathbf{A}_0$ corresponding to $\eta = 0$ satisfies $\mathbf{Z}\mathbf{A}_0 = \mathbf{0}$, i.e. it gives the interpolating circle (line), which is obviously the best possible fit. However it may happen that for some (poorly chosen) matrices $\mathbf{N}$ we have $\mathbf{A}_0^T\mathbf{N}\mathbf{A}_0 < 0$, so that the geometrically perfect solution has to be rejected. Such algebraic fits are not worth considering. For all constraint matrices $\mathbf{N}$ in this paper we have $\mathbf{A}^T\mathbf{N}\mathbf{A} \geq 0$ whenever $\mathbf{A}^T\mathbf{M}\mathbf{A} = 0$ (the reader can verify this directly).

Summarizing, we conclude that each algebraic circle fit can be computed in two steps: first we find all solutions $(\eta, \mathbf{A})$ of the generalized eigenvalue problem (26), and then we pick the one with the minimal non-negative η.

5 Error Analysis: a general scheme

We employ an error analysis scheme based on a 'small noise' assumption. That is, we assume that the errors δ_i and ε_i (Section 2) are small and treat their standard deviation σ as a small parameter. The sample size n is fixed, though it is not very small.

This approach goes back to Kadane [16] and was employed by Anderson [2] and other statisticians [3]. More recently it has been used by Kanatani [19, 22] in image processing applications, who argued that the 'small noise' model, where $\sigma \to 0$ while the sample size n is kept fixed, is more appropriate than the traditional statistical 'large sample' approach, where $n \to \infty$ while $\sigma > 0$ is kept fixed. We use a combination of these two models: our main

assumption is $\sigma \to 0$, but n is regarded as a slowly increasing parameter; more precisely we assume $n \ll \sigma^{-2}$.

Suppose one is fitting curves defined by an implicit equation

$$P(x, y; \mathbf{\Theta}) = 0, \tag{29}$$

where $\mathbf{\Theta} = (\theta_1, \dots, \theta_k)^T$ denotes a vector of unknown parameters to be estimated. Let $\tilde{\mathbf{\Theta}} = (\tilde{\theta}_1, \dots, \tilde{\theta}_k)^T$ be the 'true' parameter vector corresponding to the 'true' curve $P(x, y; \tilde{\mathbf{\Theta}}) = 0$. As in Section 2 let $(\tilde{x}_i, \tilde{y}_i)$, $i = 1, \dots, n$, denote true points, which lie on the true curve, and (x_i, y_i) observed points satisfying (1). Let $\hat{\mathbf{\Theta}}(x_1, y_1, \dots, x_n, y_n)$ be an estimator. We assume that $\hat{\mathbf{\Theta}}$ is a regular (at least four times differentiable) function of observations (x_i, y_i). The existence of the derivatives of $\hat{\mathbf{\Theta}}$ is only required at the true points $(x_i, y_i) = (\tilde{x}_i, \tilde{y}_i)$, and it follows from the implicit function theorem under general assumptions provided $P(x, y; \mathbf{\Theta})$ in (29) is differentiable (in most cases P is a polynomial in all its variables); we omit the proof.

For brevity we denote by $\mathbf{X} = (x_1, y_1, \dots, x_n, y_n)^T$ the vector of all observations, so that $\mathbf{X} = \tilde{\mathbf{X}} + \mathbf{E}$, where $\tilde{\mathbf{X}} = (\tilde{x}_1, \tilde{y}_1, \dots, \tilde{x}_n, \tilde{y}_n)^T$ is the vector of the true coordinates and $\mathbf{E} = (\delta_1, \varepsilon_1, \dots, \delta_n, \varepsilon_n)^T$ is the 'noise vector'; the components of $\mathbf{E}$ are i.i.d. normal random variables with mean zero and variance σ^2.

We use Taylor expansion to the second order terms. To keep our notation simple, we work with each scalar parameter θ_m of the vector $\mathbf{\Theta}$ separately:

$$\hat{\theta}_m(\mathbf{X}) = \hat{\theta}_m(\tilde{\mathbf{X}}) + \mathbf{G}_m^T \mathbf{E} + \tfrac{1}{2}\, \mathbf{E}^T \mathbf{H}_m \mathbf{E} + \mathcal{O}_P(\sigma^3). \tag{30}$$

Here $\mathbf{G}_m = \nabla \hat{\theta}_m$ and $\mathbf{H}_m = \nabla^2 \hat{\theta}_m$ denote the gradient (the vector of the first order partial derivatives) and the Hessian matrix of the second order partial derivatives of $\hat{\theta}_m$, respectively, taken at the true vector $\tilde{\mathbf{X}}$. The remainder term $\mathcal{O}_P(\sigma^3)$ in (30) is a random variable $\mathcal{R}$ such that $\sigma^{-3}\mathcal{R}$ is bounded in probability.

Expansion (30) shows that $\hat{\mathbf{\Theta}}(\mathbf{X}) \to \hat{\mathbf{\Theta}}(\tilde{\mathbf{X}})$ in probability, as $\sigma \to 0$. It is convenient to assume that

$$\hat{\mathbf{\Theta}}(\tilde{\mathbf{X}}) = \tilde{\mathbf{\Theta}}. \tag{31}$$

Precisely (31) means that whenever $\sigma = 0$, i.e. when the true points are observed without noise, then the estimator returns the true parameter vector, i.e. finds the true curve. Geometrically, it means that if there is a model curve that interpolates the data points, then the algorithm finds it.

With some degree of informality, one can assert that whenever (31) holds, the estimate $\hat{\mathbf{\Theta}}$ is *consistent* in the limit $\sigma \to 0$. This is regarded as a minimal requirement for any sensible fitting algorithm. For example, if the observed points lie on one circle, then every circle fitting algorithm finds that circle uniquely. Kanatani [20] remarks that algorithms which fail to follow this property "are not worth considering".

Under the assumption (31) we rewrite (30) as

$$\Delta\hat{\theta}_m(\mathbf{X}) = \mathbf{G}_m^T\mathbf{E} + \tfrac{1}{2}\,\mathbf{E}^T\mathbf{H}_m\mathbf{E} + \mathcal{O}_P(\sigma^3), \tag{32}$$

where $\Delta\hat{\theta}_m(\mathbf{X}) = \hat{\theta}_m(\mathbf{X}) - \tilde{\theta}_m$ is the statistical error of the parameter estimate.

The accuracy of an estimator $\hat{\theta}$ in statistics is characterized by its Mean Squared Error (MSE)

$$\mathbb{E}\big[(\hat{\theta} - \tilde{\theta})^2\big] = \mathsf{Var}(\hat{\theta}) + \big[\mathrm{bias}(\hat{\theta})\big]^2, \tag{33}$$

where $\mathrm{bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \tilde{\theta}$. But it often happens that exact (or even approximate) values of $\mathbb{E}(\hat{\theta})$ and $\mathsf{Var}(\hat{\theta})$ are unavailable because the probability distribution of $\hat{\theta}$ is overly complicated, which is common in curve fitting problems, even if one fits straight lines to data points; see [2, 3]. There are also cases where the estimates have theoretically infinite moments because of somewhat heavy tails, which on the other hand barely affect their practical performance. Thus their accuracy should not be characterized by the theoretical moments which happen to be affected by heavy tails; see also [2]. In all such cases one usually constructs a good approximate probability distribution for $\hat{\theta}$ and judges the quality of $\hat{\theta}$ by the moments of that distribution.

It is standard [1, 2, 3, 12, 34] to construct a normal approximation to $\hat{\theta}$ and treat its variance as an 'approximative' MSE of $\hat{\theta}$. The normal approximation is usually based on the leading term in the Taylor expansion, like $\mathbf{G}_m^T\mathbf{E}$ in (32). For circle fitting algorithms, the resulting variance (see below) will be the same for all known methods, so we will go one step further and use the second order term. This gives us a better approximative distribution and allows us to compare circle fitting methods. In our formulas, $\mathbb{E}(\hat{\theta}_m)$ and $\mathsf{Var}(\hat{\theta}_m)$ denote the mean and variance of the resulting approximative distribution.

The first term in (32) is a linear combination of i.i.d. normal random variables that have zero mean, hence it is itself a normal random variable with

zero mean. The second term is a quadratic form of i.i.d. normal variables. Since $\mathbf{H}_m$ is a symmetric matrix, we have $\mathbf{H}_m = \mathbf{Q}_m^T \mathbf{D}_m \mathbf{Q}_m$, where $\mathbf{Q}_m$ is an orthogonal matrix and $\mathbf{D}_m = \operatorname{diag}\{d_1, \dots, d_{2n}\}$ is a diagonal matrix. The vector $\mathbf{E}_m = \mathbf{Q}_m \mathbf{E}$ has the same distribution as $\mathbf{E}$ does, i.e. its components are i.i.d. normal random variables with mean zero and variance σ^2. Thus

$$\mathbf{E}^T \mathbf{H}_m \mathbf{E} = \mathbf{E}_m^T \mathbf{D}_m \mathbf{E}_m = \sigma^2 \sum d_i Z_i^2, \tag{34}$$

where the Z_i's are i.i.d. standard normal random variables, and the mean value of (34) is

$$\mathbb{E}\big(\mathbf{E}^T \mathbf{H}_m \mathbf{E}\big) = \sigma^2 \operatorname{tr} \mathbf{D}_m = \sigma^2 \operatorname{tr} \mathbf{H}_m. \tag{35}$$

Therefore, taking the mean value in (32) gives

$$\operatorname{bias}(\hat\theta_m) = \mathbb{E}(\Delta\hat\theta_m) = \tfrac{1}{2}\, \sigma^2 \operatorname{tr} \mathbf{H}_m + \mathcal{O}(\sigma^4). \tag{36}$$

Note that the expectations of all third order terms vanish, because the components of $\mathbf{E}$ are independent and their first and third moments are zero; thus the remainder term is of order σ^4.

Squaring (32) and again using (34) give the mean squared error (MSE)

$$\mathbb{E}\big([\Delta\hat\theta_m]^2\big) = \sigma^2 \mathbf{G}_m^T \mathbf{G}_m + \tfrac{1}{4}\, \sigma^4 \big([\operatorname{tr} \mathbf{H}_m]^2 + 2\|\mathbf{H}_m\|_F^2\big) + \mathcal{R}, \tag{37}$$

where $\|\mathbf{H}_m\|_F^2 = \operatorname{tr} \mathbf{H}_m^2$ is the Frobenius norm (note that $\|\mathbf{H}_m\|_F^2 = \|\mathbf{D}_m\|_F^2 = \operatorname{tr} \mathbf{D}_m^2$). The remainder $\mathcal{R}$ includes terms of order σ^6, as well as some terms of order σ^4 that contain third order partial derivatives, such as $\partial^3 \hat\theta_m / \partial x_i^3$ and $\partial^3 \hat\theta_m / \partial x_i^2 \partial x_j$. A similar expression can be derived for $\mathbb{E}\big(\Delta\hat\theta_m \Delta\hat\theta_{m'}\big)$ for $m \neq m'$, we omit it and only give the final formula below.

Classification of higher order terms. In the MSE expansion (37), the leading term $\sigma^2 \mathbf{G}_m^T \mathbf{G}_m$ is the most significant. The terms of order σ^4 are often given by long complicated formulas. Even the expression for the bias (36) may contain several terms of order σ^2, as we will see below. Fortunately, it is possible to sort them out keeping only the most significant ones, see next.

Kanatani [22] recently derived formulas for the bias of certain ellipse fitting algorithms. First he found all the terms of order σ^2, but in the end he noticed that some terms were of order σ^2 (independent of n), while the others of order σ^2/n. The magnitude of the former was clearly larger than that of the latter, and when Kanatani made his conclusions he ignored the

terms of order σ^2/n. Here we formalize Kanatani's classification of higher order terms as follows:
– In the expression for the bias (36) we keep terms of order σ^2 (independent of n) and ignore terms of order σ^2/n.
– In the expression for the mean squared error (37) we keep terms of order σ^4 (independent of n) and ignore terms of order σ^4/n.

These rules agree with our assumption that not only $\sigma \to 0$, but also $n \to \infty$, although n increases rather slowly ($n \ll 1/\sigma^2$). Such models were studied by Amemiya, Fuller and Wolter [1, 34] who made a more rigid assumption that $n \sim \sigma^{-a}$ for some $0 < a < 2$.

Now it turns out (we omit detailed proofs; see [14]) that the main term $\sigma^2 \mathbf{G}_m^T \mathbf{G}_m$ in our expression for the MSE (37) is of order σ^2/n; so it will never be ignored. Of the fourth order terms, $\frac{1}{2}\,\sigma^4 \|\mathbf{H}_m\|_F^2$ is of order σ^4/n, hence it will be discarded, and the same applies to all the terms involving third order partial derivatives mentioned above.

The bias $\sigma^2 \operatorname{tr} \mathbf{H}_m$ in (36) is, generally, of order σ^2 (independent of n), thus its contribution to the mean squared error (37) is significant. However the full expression for the bias may contain terms of order σ^2 and of order σ^2/n, of which the latter will be ignored; see below.

Now the terms in (37) have the following orders of magnitude:

$$\mathbb{E}\big([\Delta\hat{\theta}_m]^2\big) = \mathcal{O}(\sigma^2/n) + \mathcal{O}(\sigma^4) + \mathcal{O}(\sigma^4/n) + \mathcal{O}(\sigma^6), \tag{38}$$

where each big-O simply indicates the order of the corresponding term in (37). It is interesting to roughly compare their values numerically. In typical computer vision applications, σ does not exceed 0.05; see [5]. The number of data points normally varies between 10-20 (on the low end) and a few hundred (on the high end). For simplicity, we can set $n \sim 1/\sigma$ for smaller samples and $n \sim 1/\sigma^2$ for larger samples. Then Table 1 presents the corresponding typical magnitudes of each of the four terms in (37).

We see that for larger samples the fourth order term coming from the bias may be just as big as the leading second-order term, hence it would be unwise to ignore it. Earlier studies, see e.g. [5, 8, 17], usually focused on the leading, i.e. second-order, terms only, disregarding all the fourth-order terms, and this is where our analysis is different. We make one step further – we keep all the terms of order $\mathcal{O}(\sigma^2/n)$ and $\mathcal{O}(\sigma^4)$. The less significant terms of order $\mathcal{O}(\sigma^4/n)$ and $\mathcal{O}(\sigma^6)$ would be discarded.

Now combining all our results gives a matrix formula for the (total) mean

	σ^2/n	σ^4	σ^4/n	σ^6
small samples ($n \sim 1/\sigma$)	σ^3	σ^4	σ^5	σ^6
large samples ($n \sim 1/\sigma^2$)	σ^4	σ^4	σ^6	σ^6

Table 1: The order of magnitude of the four terms in (37).

squared error (MSE)

$$\mathbb{E}\Big[(\Delta\hat{\boldsymbol{\Theta}})(\Delta\hat{\boldsymbol{\Theta}})^T\Big] = \sigma^2\mathbf{G}\mathbf{G}^T + \sigma^4\mathbf{B}\mathbf{B}^T + \cdots, \tag{39}$$

where $\mathbf{G}$ is the $k \times 2n$ matrix of first order partial derivatives of $\hat{\boldsymbol{\Theta}}(\mathbf{X})$, its rows are $\mathbf{G}_m^T$, $1 \le m \le k$, and $\mathbf{B} = \frac{1}{2}[\operatorname{tr}\mathbf{H}_1, \dots \operatorname{tr}\mathbf{H}_k]^T$ is the k-vector that represents the leading term of the bias of $\hat{\boldsymbol{\Theta}}$, cf. (36). The trailing dots in (39) stand for all insignificant terms (those of order σ^4/n and σ^6).

We call the first (main) term $\sigma^2\mathbf{G}\mathbf{G}^T$ in (39) the *variance* term, as it characterizes the variance (more precisely, the covariance matrix) of the estimator $\hat{\boldsymbol{\Theta}}$, to the leading order. For brevity we denote $\mathbf{V} = \mathbf{G}\mathbf{G}^T$. The second term $\sigma^4\mathbf{B}\mathbf{B}^T$ is the 'tensor square' of the bias $\sigma^2\mathbf{B}$ of the estimator, again to the leading order. When we deal with particular estimators in the next sections, we will see that the actual expression for the bias is a sum of terms of two types: some of them are of order $\mathcal{O}(\sigma^2)$ and some others are of order $\mathcal{O}(\sigma^2/n)$, i.e.

$$\mathbb{E}(\Delta\hat{\boldsymbol{\Theta}}) = \sigma^2\mathbf{B} + \mathcal{O}(\sigma^4) = \sigma^2\mathbf{B}_1 + \sigma^2\mathbf{B}_2 + \mathcal{O}(\sigma^4), \tag{40}$$

where $\mathbf{B}_1 = \mathcal{O}(1)$ and $\mathbf{B}_2 = \mathcal{O}(1/n)$. We call $\sigma^2\mathbf{B}_1$ the *essential bias* of the estimator $\hat{\boldsymbol{\Theta}}$. This is its bias to the leading order, σ^2. The other terms, i.e. $\sigma^2\mathbf{B}_2$, and $\mathcal{O}(\sigma^4)$, constitute non-essential bias; they can be discarded. Now (39) can be written as

$$\mathbb{E}\Big[(\Delta\hat{\boldsymbol{\Theta}})(\Delta\hat{\boldsymbol{\Theta}})^T\Big] = \sigma^2\mathbf{V} + \sigma^4\mathbf{B}_1\mathbf{B}_1^T + \cdots, \tag{41}$$

where we only keep significant terms of order σ^2/n and σ^4 and drop the rest.

KCR lower bound. The matrix $\mathbf{V}$ representing the leading terms of the variance has a natural lower bound (an analogue of the Cramer-Rao bound):

for every curve family (29) there is a symmetric positive semi-definite matrix $\mathbf{V}_{\min}$ such that for every estimator satisfying (31)

$$\mathbf{V} \ge \mathbf{V}_{\min} = \left(\sum \frac{P_{\boldsymbol{\Theta} i}\, P_{\boldsymbol{\Theta} i}^T}{\|P_{\mathbf{x} i}\|^2} \right)^{-1}, \tag{42}$$

in the sense that $\mathbf{V} - \mathbf{V}_{\min}$ is a positive semi-definite matrix. Here

$$P_{\boldsymbol{\Theta} i} = \Big(\partial P(\tilde{\mathbf{x}}_i; \tilde{\boldsymbol{\Theta}})/\partial\theta_1, \ldots, \partial P(\tilde{\mathbf{x}}_i; \tilde{\boldsymbol{\Theta}})/\partial\theta_k \Big)^T \tag{43}$$

stands for the gradient of P with respect to the model parameters $\theta_1, \ldots, \theta_k$ and

$$P_{\mathbf{x} i} = \Big(\partial P(\tilde{\mathbf{x}}_i; \tilde{\boldsymbol{\Theta}})/\partial x, \partial P(\tilde{\mathbf{x}}_i; \tilde{\boldsymbol{\Theta}})/\partial y \Big)^T \tag{44}$$

for the gradient with respect to the planar variables x and y; both gradients are taken at the true point $\tilde{\mathbf{x}}_i = (\tilde{x}_i, \tilde{y}_i)$. For example in the case of fitting circles defined by $P = (x-a)^2 + (y-b)^2 - R^2$, we have

$$P_{\boldsymbol{\Theta} i} = -2\big((\tilde{x}_i - \tilde{a}), (\tilde{y}_i - \tilde{b}), \tilde{R} \big)^T, \qquad P_{\mathbf{x} i} = 2\big((\tilde{x}_i - \tilde{a}), (\tilde{y}_i - \tilde{b}) \big)^T. \tag{45}$$

Therefore,

$$\mathbf{V}_{\min} = (\mathbf{W}^T \mathbf{W})^{-1}, \tag{46}$$

where

$$\mathbf{W} \stackrel{\text{def}}{=} \begin{bmatrix} \tilde{u}_1 & \tilde{v}_1 & 1 \\ \vdots & \vdots & \vdots \\ \tilde{u}_n & \tilde{v}_n & 1 \end{bmatrix} \tag{47}$$

and $\tilde{u}_i, \tilde{v}_i$ are given by (4).

The general inequality (42) was proved by Kanatani [17, 18] for unbiased estimators $\hat{\boldsymbol{\Theta}}$ and then extended by Chernov and Lesort [8] to all estimators satisfying (31). The geometric fit (which minimizes orthogonal distances) always satisfies (31) and attains the lower bound $\mathbf{V}_{\min}$; this was proved by Fuller (Theorem 3.2.1 in [12]) and independently by Chernov and Lesort [8], who named the inequality (42) *Kanatani-Cramer-Rao* (KCR) lower bound. See also survey [25] for the more general case of heteroscedastic noise.

Assessing the quality of estimators. Our analysis dictates the following strategy of assessing the quality of an estimator $\hat{\mathbf{\Theta}}$: first of all, its accuracy is characterized by the matrix $\mathbf{V}$, which must be compared to the KCR lower bound $\mathbf{V}_{\min}$. We will see that for all the circle fitting algorithms the matrix $\mathbf{V}$ actually achieves its lower bound $\mathbf{V}_{\min}$, i.e. we have $\mathbf{V} = \mathbf{V}_{\min}$, hence these algorithms are optimal to the leading order.

Next, once the factor $\mathbf{V}$ is already at its natural minimum, the accuracy of an estimator should be characterized by the vector $\mathbf{B}_1$ representing the essential bias – better estimates should have smaller essential biases. It appears that there is no natural minimum for $\|\mathbf{B}_1\|$, in fact there exist estimators which have a minimum variance $\mathbf{V} = \mathbf{V}_{\min}$ and a zero essential bias, i.e. $\mathbf{B}_1 = \mathbf{0}$. We will construct such an estimator in Section 7.

6 Error analysis of geometric circle fit

Here we apply the general method of the previous section to the geometric circle fit, i.e. to the estimator $\hat{\mathbf{\Theta}} = (\hat{a}, \hat{b}, \hat{R})$ of the circle parameters minimizing the sum $\sum d_i^2$ of orthogonal (geometric) distances from the data points to the fitted circle.

Variance of the geometric circle fit. We start with the main part of our error analysis – the variance term represented by $\sigma^2\mathbf{V}$ in (41). The distances $d_i = r_i - R$ can be expanded as

$$\begin{aligned} d_i &= \sqrt{\big[(\tilde{x}_i + \delta_i) - (\tilde{a} + \Delta a)\big]^2 + \big[(\tilde{y}_i + \varepsilon_i) - (\tilde{b} + \Delta b)\big]^2} - \tilde{R} - \Delta R \\ &= \sqrt{\tilde{R}^2 + 2\tilde{R}\tilde{u}_i(\delta_i - \Delta a) + 2\tilde{R}\tilde{v}_i(\varepsilon_i - \Delta b) + \mathcal{O}_P(\sigma^2)} - \tilde{R} - \Delta R \\ &= \tilde{u}_i(\delta_i - \Delta a) + \tilde{v}_i(\varepsilon_i - \Delta b) - \Delta R + \mathcal{O}_P(\sigma^2), \end{aligned} \tag{48}$$

see (4). Minimizing $\sum d_i^2$ to the first order is equivalent to minimizing

$$\sum (\tilde{u}_i\, \Delta a + \tilde{v}_i\, \Delta b + \Delta R - \tilde{u}_i\delta_i - \tilde{v}_i\varepsilon_i)^2. \tag{49}$$

This is a classical least squares problem that can also be written as

$$\mathbf{W}\, \Delta\mathbf{\Theta} \approx \tilde{\mathbf{U}}\boldsymbol{\delta} + \tilde{\mathbf{V}}\boldsymbol{\varepsilon}, \tag{50}$$

where $\mathbf{W}$ is given by (47), $\mathbf{\Theta} = (a, b, R)^T$, as well as $\boldsymbol{\delta} = (\delta_1, \dots, \delta_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, while $\tilde{\mathbf{U}} = \operatorname{diag}(\tilde{u}_1, \dots, \tilde{u}_n)$ and $\tilde{\mathbf{V}} = \operatorname{diag}(\tilde{v}_1, \dots, \tilde{v}_n)$.

The solution of the least squares problem (50) is

$$\Delta\hat{\boldsymbol{\Theta}} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T(\tilde{\mathbf{U}}\boldsymbol{\delta} + \tilde{\mathbf{V}}\boldsymbol{\varepsilon}), \tag{51}$$

of course this does not include the $\mathcal{O}_P(\sigma^2)$ terms. Thus the variance of our estimator, to the leading order, is

$$\mathbb{E}\big[(\Delta\hat{\boldsymbol{\Theta}})(\Delta\hat{\boldsymbol{\Theta}})^T\big] = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbb{E}\big[(\tilde{\mathbf{U}}\boldsymbol{\delta}+\tilde{\mathbf{V}}\boldsymbol{\varepsilon})(\boldsymbol{\delta}^T\tilde{\mathbf{U}}+\boldsymbol{\varepsilon}^T\tilde{\mathbf{V}})\big]\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1}. \tag{52}$$

Now observe that $\mathbb{E}(\boldsymbol{\delta}\boldsymbol{\varepsilon}^T) = \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\delta}^T) = \mathbf{0}$, as well as $\mathbb{E}(\boldsymbol{\delta}\boldsymbol{\delta}^T) = \mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \sigma^2\mathbf{I}$, and we have $\tilde{\mathbf{U}}^2 + \tilde{\mathbf{V}}^2 = \mathbf{I}$. Thus to the leading order

$$\mathbb{E}\big[(\Delta\hat{\boldsymbol{\Theta}})(\Delta\hat{\boldsymbol{\Theta}})^T\big] = \sigma^2(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{W}(\mathbf{W}^T\mathbf{W})^{-1} = \sigma^2(\mathbf{W}^T\mathbf{W})^{-1}, \tag{53}$$

where the higher order (of σ^4) terms are not included. Comparing this to (46) confirms that the geometric fit attains the minimal possible covariance matrix $\mathbf{V}$.

Bias of the geometric circle fit. Now we do a second-order error analysis, which has not been previously done in the literature. According to a general formula (30), we put

$$\begin{aligned} a &= \tilde{a} + \Delta_1 a + \Delta_2 a + \mathcal{O}_P(\sigma^3),\\ b &= \tilde{b} + \Delta_1 b + \Delta_2 b + \mathcal{O}_P(\sigma^3),\\ R &= \tilde{R} + \Delta_1 R + \Delta_2 R + \mathcal{O}_P(\sigma^3). \end{aligned} \tag{54}$$

Here $\Delta_1 a$, $\Delta_1 b$, $\Delta_1 R$ are linear combinations of ε_i's and δ_i's, which were found above, in (51), and $\Delta_2 a$, $\Delta_2 b$, $\Delta_2 R$ are quadratic forms of ε_i's and δ_i's to be determined next.

Expanding the distances d_i to the second order terms gives

$$\begin{aligned} d_i = {}& \tilde{u}_i(\delta_i - \Delta_1 a) + \tilde{v}_i(\varepsilon_i - \Delta_1 b) - \Delta_1 R\\ & - \tilde{u}_i\,\Delta_2 a - \tilde{v}_i\,\Delta_2 b - \Delta_2 R + \tfrac{\tilde{v}_i^2}{2\tilde{R}}(\delta_i - \Delta_1 a)^2 + \tfrac{\tilde{u}_i^2}{2\tilde{R}}(\varepsilon_i - \Delta_1 b)^2\\ & - \tfrac{\tilde{u}_i\tilde{v}_i}{\tilde{R}}(\delta_i - \Delta_1 a)(\varepsilon_i - \Delta_1 b). \end{aligned} \tag{55}$$

Since we already found $\Delta_1 a$, $\Delta_1 b$, $\Delta_1 R$, the only unknowns are $\Delta_2 a$, $\Delta_2 b$, $\Delta_2 R$. Minimizing $\sum d_i^2$ is now equivalent to minimizing

$$\sum(\tilde{u}_i\,\Delta_2 a + \tilde{v}_i\,\Delta_2 b + \Delta_2 R - f_i)^2, \tag{56}$$

where

$$
\begin{aligned}
f_i &= \tilde{u}_i(\delta_i - \Delta_1 a) + \tilde{v}_i(\varepsilon_i - \Delta_1 b) - \Delta_1 R \\
&+ \tfrac{\tilde{v}_i^2}{2R}(\delta_i - \Delta_1 a)^2 + \tfrac{\tilde{u}_i^2}{2R}(\varepsilon_i - \Delta_1 b)^2 - \tfrac{\tilde{u}_i\tilde{v}_i}{R}(\delta_i - \Delta_1 a)(\varepsilon_i - \Delta_1 b). \qquad (57)
\end{aligned}
$$

This is another least squares problem, and its solution is

$$\Delta_2\hat{\boldsymbol{\Theta}} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{F}, \qquad (58)$$

where $\mathbf{F} = (f_1, \ldots, f_n)^T$; of course this is a quadratic approximation which does not include $\mathcal{O}_P(\sigma^3)$ terms. In fact, the contribution from the first three (linear) terms in (57) vanishes, quite predictably; thus only the last two (quadratic) terms matter.

Taking the mean value gives, to the leading order,

$$\mathbb{E}(\Delta\hat{\boldsymbol{\Theta}}) = \mathbb{E}(\Delta_2\hat{\boldsymbol{\Theta}}) = \frac{\sigma^2}{2R}\big[(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{1} + (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{S}\big], \qquad (59)$$

where $\mathbf{1} = (1, 1, \ldots, 1)^T$ and $\mathbf{S} = (s_1, \ldots, s_n)^T$, here s_i is a scalar

$$s_i = [-\tilde{v}_i, \tilde{u}_i, 0](\mathbf{W}^T\mathbf{W})^{-1}[-\tilde{v}_i, \tilde{u}_i, 0]^T. \qquad (60)$$

The second term in (59) is of order $\mathcal{O}(\sigma^2/n)$, thus the essential bias is given by the first term only, and it can be simplified. Since the last column of the matrix $\mathbf{W}^T\mathbf{W}$ coincides with the vector $\mathbf{W}^T\mathbf{1}$, we have $(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{1} = [0, 0, 1]^T$, hence the essential bias of the geometric circle fit is

$$\mathbb{E}(\Delta\hat{\boldsymbol{\Theta}}) \stackrel{\text{ess}}{=} \frac{\sigma^2}{2\tilde{R}}\big[0, 0, 1\big]^T. \qquad (61)$$

Thus the estimates of the circle center, $\hat{a}$ and $\hat{b}$, have *no essential bias*, while the estimate of the radius has essential bias

$$\mathbb{E}(\Delta\hat{R}) \stackrel{\text{ess}}{=} \frac{\sigma^2}{2\tilde{R}}, \qquad (62)$$

which is independent of the number and location of the true points. These facts are consistent with the results obtained by Berman [5] under the assumptions that $\sigma > 0$ is fixed and $n \to \infty$.

7 Error analysis of algebraic circle fits

Here we analyze algebraic circle fits using their matrix representation. We recall that $\mathbf{A}$ is a solution of the generalized eigenvalue problem $\mathbf{MA} = \eta\mathbf{NA}$ corresponding to the smallest non-negative η. We also require $\|\mathbf{A}\| = 1$.

Matrix perturbation method. For every random variable, matrix or vector, $\mathbf{L}$, we write

$$\mathbf{L} = \tilde{\mathbf{L}} + \Delta_1\mathbf{L} + \Delta_2\mathbf{L} + \mathcal{O}_P(\sigma^3), \tag{63}$$

where $\tilde{\mathbf{L}}$ is its 'true', nonrandom, value (achieved when $\sigma = 0$), $\Delta_1\mathbf{L}$ is a linear combination of δ_i's and ε_i's, and $\Delta_2\mathbf{L}$ is a quadratic form of δ_i's and ε_i's; all the higher order terms (cubic etc.) are represented by $\mathcal{O}_P(\sigma^3)$. For brevity, we drop the $\mathcal{O}_P(\sigma^3)$ terms in our formulas. Therefore $\mathbf{A} = \tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \Delta_2\mathbf{A}$ and $\mathbf{M} = \tilde{\mathbf{M}} + \Delta_1\mathbf{M} + \Delta_2\mathbf{M}$, and (22) implies

$$\Delta_1\mathbf{M} = n^{-1}(\tilde{\mathbf{Z}}^T\Delta_1\mathbf{Z} + \Delta_1\mathbf{Z}^T\tilde{\mathbf{Z}}), \tag{64}$$

$$\Delta_2\mathbf{M} = n^{-1}(\Delta_1\mathbf{Z}^T\,\Delta_1\mathbf{Z} + \tilde{\mathbf{Z}}^T\,\Delta_2\mathbf{Z} + \Delta_2\mathbf{Z}^T\,\tilde{\mathbf{Z}}). \tag{65}$$

Since the true points lie on the true circle, $\tilde{\mathbf{Z}}\tilde{\mathbf{A}} = \mathbf{0}$, as well as $\tilde{\mathbf{M}}\tilde{\mathbf{A}} = \mathbf{0}$ (hence $\tilde{\mathbf{M}}$ is a singular matrix). Therefore

$$\tilde{\mathbf{A}}^T\,\Delta_1\mathbf{M}\,\tilde{\mathbf{A}} = n^{-1}\tilde{\mathbf{A}}^T\big(\tilde{\mathbf{Z}}^T\,\Delta_1\mathbf{Z} + \Delta_1\mathbf{Z}^T\,\tilde{\mathbf{Z}}\big)\tilde{\mathbf{A}} = 0, \tag{66}$$

hence $\mathbf{A}^T\mathbf{MA} = \mathcal{O}_P(\sigma^2)$, and premultiplying (26) by $\mathbf{A}^T$ yields $\eta = \mathcal{O}_P(\sigma^2)$. Next, substituting the expansions of $\mathbf{M}$, $\mathbf{A}$, and $\mathbf{N}$ into (26) gives

$$(\tilde{\mathbf{M}} + \Delta_1\mathbf{M} + \Delta_2\mathbf{M})(\tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \Delta_2\mathbf{A}) = \eta\tilde{\mathbf{N}}\tilde{\mathbf{A}} \tag{67}$$

(recall that $\mathbf{N}$ is data-dependent for the Taubin method, but only its 'true' value $\tilde{\mathbf{N}}$ matters, as $\eta = \mathcal{O}_P(\sigma^2)$, hence the use of the observed values only adds higher order terms). Now using $\tilde{\mathbf{M}}\tilde{\mathbf{A}} = \mathbf{0}$ yields

$$(\tilde{\mathbf{M}}\,\Delta_1\mathbf{A} + \Delta_1\mathbf{M}\,\tilde{\mathbf{A}}) + (\tilde{\mathbf{M}}\,\Delta_2\mathbf{A} + \Delta_1\mathbf{M}\,\Delta_1\mathbf{A} + \Delta_2\mathbf{M}\,\tilde{\mathbf{A}}) = \eta\tilde{\mathbf{N}}\tilde{\mathbf{A}} \tag{68}$$

The left hand side of (68) consists of a linear part $(\tilde{\mathbf{M}}\,\Delta_1\mathbf{A} + \Delta_1\mathbf{M}\,\tilde{\mathbf{A}})$ and a quadratic part $(\tilde{\mathbf{M}}\,\Delta_2\mathbf{A} + \Delta_1\mathbf{M}\,\Delta_1\mathbf{A} + \Delta_2\mathbf{M}\,\tilde{\mathbf{A}})$. Separating them gives

$$\tilde{\mathbf{M}}\,\Delta_1\mathbf{A} + n^{-1}\tilde{\mathbf{Z}}^T\,\Delta_1\mathbf{Z}\,\tilde{\mathbf{A}} = \mathbf{0} \tag{69}$$

(where we used (64) and $\tilde{\mathbf{Z}}\tilde{\mathbf{A}} = \mathbf{0}$) and

$$\tilde{\mathbf{M}}\,\Delta_2\mathbf{A} + \Delta_1\mathbf{M}\,\Delta_1\mathbf{A} + \Delta_2\mathbf{M}\,\tilde{\mathbf{A}} = \eta\tilde{\mathbf{N}}\tilde{\mathbf{A}}. \tag{70}$$

Note that $\tilde{\mathbf{M}}$ is a singular matrix (because $\tilde{\mathbf{M}}\tilde{\mathbf{A}} = \mathbf{0}$), but whenever there are at least three distinct true points, they determine a unique true circle, thus the kernel of $\tilde{\mathbf{M}}$ is one-dimensional, and it coincides with $\operatorname{span}(\tilde{\mathbf{A}})$. Also, we set $\|\mathbf{A}\| = 1$, hence $\Delta_1\mathbf{A}$ is orthogonal to $\tilde{\mathbf{A}}$, and we can write

$$\Delta_1\mathbf{A} = -n^{-1}\tilde{\mathbf{M}}^{-}\tilde{\mathbf{Z}}^T\,\Delta_1\mathbf{Z}\,\tilde{\mathbf{A}}, \tag{71}$$

where $\tilde{\mathbf{M}}^{-}$ denotes the Moore-Penrose pseudoinverse. Now one can easily check that $\mathbb{E}(\Delta_1\mathbf{M}\,\Delta_1\mathbf{A}) = \mathcal{O}(\sigma^2/n)$ and $\mathbb{E}(\Delta_1\mathbf{A}) = 0$; these facts will be useful in the upcoming analysis.

Variance of algebraic circle fits. From (71) we conclude that

$$\begin{aligned}\mathbb{E}\big[(\Delta_1\mathbf{A})(\Delta_1\mathbf{A})^T\big] &= n^{-2}\tilde{\mathbf{M}}^{-}\mathbb{E}(\tilde{\mathbf{Z}}^T\,\Delta_1\mathbf{Z}\,\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\,\Delta_1\mathbf{Z}^T\,\tilde{\mathbf{Z}})\tilde{\mathbf{M}}^{-} \\ &= n^{-2}\tilde{\mathbf{M}}^{-}\mathbb{E}\Big[\big(\sum_i \tilde{\mathbf{Z}}_i\,\Delta_1\mathbf{Z}_i^T\big)\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\big(\sum_j \Delta_1\mathbf{Z}_j^T\,\tilde{\mathbf{Z}}_j^T\big)\Big]\tilde{\mathbf{M}}^{-}, \end{aligned} \tag{72}$$

where

$$\tilde{\mathbf{Z}}_i \stackrel{\text{def}}{=} \begin{bmatrix} \tilde{z}_i \\ \tilde{x}_i \\ \tilde{y}_i \\ 1 \end{bmatrix} \qquad \text{and} \qquad \Delta_1\mathbf{Z}_i = \begin{bmatrix} 2\tilde{x}_i\delta_i + 2\tilde{y}_i\varepsilon_i \\ \delta_i \\ \varepsilon_i \\ 0 \end{bmatrix} \tag{73}$$

denote the columns of the matrices $\tilde{\mathbf{Z}}^T$ and $\Delta_1\mathbf{Z}^T$, respectively. Next,

$$\mathbb{E}\big[(\Delta_1\mathbf{Z}_i)(\Delta_1\mathbf{Z}_j)^T\big] = \begin{cases} 0 & \text{whenever} \quad i \neq j \\ \sigma^2\tilde{\mathbf{T}}_i & \text{whenever} \quad i = j \end{cases} \tag{74}$$

where

$$\tilde{\mathbf{T}}_i \stackrel{\text{def}}{=} \begin{bmatrix} 4\tilde{z}_i & 2\tilde{x}_i & 2\tilde{y}_i & 0 \\ 2\tilde{x}_i & 1 & 0 & 0 \\ 2\tilde{y}_i & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{75}$$

Note $n^{-1}\sum\tilde{\mathbf{T}}_i = \tilde{\mathbf{T}}$ and $\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}} = \tilde{B}^2 + \tilde{C}^2 - 4\tilde{A}\tilde{D}$ for each i; recall (23) and (24). Hence

$$\sum\tilde{\mathbf{Z}}_i\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_i^T = \sum(\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}})\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^T = n(\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}})\tilde{\mathbf{M}}. \tag{76}$$

Combining the above formulas gives

$$\begin{aligned}\mathbb{E}\big[(\Delta_1\mathbf{A})(\Delta_1\mathbf{A})^T\big] &= n^{-2}\tilde{\mathbf{M}}^-\Big[\sum_{i,j}\tilde{\mathbf{Z}}_i\tilde{\mathbf{A}}^T\mathbb{E}\big(\Delta_1\mathbf{Z}_i^T\Delta_1\mathbf{Z}_j^T\big)\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_j^T\Big]\tilde{\mathbf{M}}^-\\ &= n^{-2}\sigma^2\tilde{\mathbf{M}}^-\Big[\sum\tilde{\mathbf{Z}}_i\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_i^T\Big]\tilde{\mathbf{M}}^-\\ &= n^{-1}\sigma^2\tilde{\mathbf{M}}^-(\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}}).\end{aligned}\tag{77}$$

Remarkably, the variance of algebraic fits does not depend on the constraint matrix $\mathbf{N}$, hence all algebraic fits have the same variance (to the leading order). In the next section we will derive the variance of algebraic fits in the natural circle parameters (a, b, R) and see that it coincides with the variance of the geometric fit (53).

Bias of algebraic circle fits. Since $\mathbb{E}(\Delta_1\mathbf{A}) = 0$, it will be enough to find $\mathbb{E}(\Delta_2\mathbf{A})$. Premultiplying (70) by $\tilde{\mathbf{A}}^T$ yields

$$\eta = \frac{\tilde{\mathbf{A}}^T\mathbf{M}\mathbf{A}}{\tilde{\mathbf{A}}^T\mathbf{N}\mathbf{A}} = \frac{\tilde{\mathbf{A}}^T\,\Delta_2\mathbf{M}\,\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^T\,\Delta_1\mathbf{M}\,\Delta_1\mathbf{A}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}} + \mathcal{O}_P(\sigma^2/n).\tag{78}$$

Recall that $\mathbb{E}(\Delta_1\mathbf{M}\,\Delta_1\mathbf{A}) = \mathcal{O}(\sigma^2/n)$, thus this term will not affect the essential bias and we drop it. Taking the mean value and using (71) gives

$$\mathbb{E}(\eta) = \frac{\tilde{\mathbf{A}}^T\mathbb{E}(\Delta_2\mathbf{M})\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}} + \mathcal{O}(\sigma^2/n),\tag{79}$$

We substitute (65) into (78), use $\tilde{\mathbf{Z}}\tilde{\mathbf{A}} = \mathbf{0}$, then observe that

$$\mathbb{E}(\Delta_1\mathbf{Z}^T\,\Delta_1\mathbf{Z}\tilde{\mathbf{A}}) = \sigma^2\sum\tilde{\mathbf{T}}_i\tilde{\mathbf{A}} = 2\tilde{A}\sigma^2\sum\tilde{\mathbf{Z}}_i + n\sigma^2\mathbf{P}\tilde{\mathbf{A}}\tag{80}$$

(here $\tilde{A}$ is the first component of the vector $\tilde{\mathbf{A}}$). Then, note that $\Delta_2\mathbf{Z}_i = (\delta_i^2 + \varepsilon_i^2, 0, 0, 0)^T$, and so

$$\mathbb{E}(\tilde{\mathbf{Z}}^T\,\Delta_2\mathbf{Z})\tilde{\mathbf{A}} = 2\tilde{A}\sigma^2\sum\tilde{\mathbf{Z}}_i.\tag{81}$$

Therefore the essential bias is given by

$$\mathbb{E}(\Delta_2\mathbf{A}) \overset{\text{ess}}{=} -\sigma^2\tilde{\mathbf{M}}^-\Big[4\tilde{A}n^{-1}\sum\tilde{\mathbf{Z}}_i + \mathbf{P}\tilde{\mathbf{A}} - \frac{\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\,\tilde{\mathbf{N}}\tilde{\mathbf{A}}\Big].\tag{82}$$

A more detailed analysis (which we omit) gives the following expression containing all the $\mathcal{O}(\sigma^2)$ and $\mathcal{O}(\sigma^2/n)$ terms:

$$\mathbb{E}(\Delta_2\mathbf{A}) = -\sigma^2\tilde{\mathbf{M}}^-\Big[4\tilde{A}n^{-1}\sum\tilde{\mathbf{Z}}_i + (1-\tfrac{4}{n})\mathbf{P}\tilde{\mathbf{A}} - \frac{(1-\frac{3}{n})(\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}})}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\,\tilde{\mathbf{N}}\tilde{\mathbf{A}}$$
$$- 4\tilde{A}n^{-2}\sum(\tilde{\mathbf{Z}}_i^T\tilde{\mathbf{M}}^-\tilde{\mathbf{Z}}_i)\tilde{\mathbf{Z}}_i\Big] + \mathcal{O}(\sigma^4). \tag{83}$$

This expression demonstrates that the terms of order σ^2/n (the *non-essential bias*) only add a small correction, which is negligible when n is large.

The expressions (82) and (83) can be simplified. Note that the vector $n^{-1}\sum\tilde{\mathbf{Z}}_i$ coincides with the last column of the matrix $\tilde{\mathbf{M}}$, hence

$$-\sigma^2\tilde{\mathbf{M}}^-\Big[4\tilde{A}n^{-1}\sum\tilde{\mathbf{Z}}_i\Big] = -4\sigma^2\tilde{A}\,[0,0,0,1]^T. \tag{84}$$

In fact, this term will play the key role in the subsequent analysis.

8 Comparison of various circle fits

Bias of the Pratt and Taubin fits. We have seen that all the algebraic fits have the same main characteristic – the variance (77), to the leading order. We will see below that their variance coincides with that of the geometric circle fit. Thus the difference between all our circle fits should be traced to the higher order terms, especially to their essential biases.

First we compare the Pratt and Taubin fits. For the Pratt fit, the constraint matrix is $\mathbf{N} = \tilde{\mathbf{N}} = \mathbf{P}$, hence its essential bias (82) becomes

$$\mathbb{E}(\Delta_2\mathbf{A}_{\text{Pratt}}) \overset{\text{ess}}{=} -4\sigma^2\tilde{A}\,[0,0,0,1]^T. \tag{85}$$

In other words, the Pratt constraint $\mathbf{N} = \mathbf{P}$ cancels the second (middle) term in (82); it leaves the first term intact.

For the Taubin fit, the constraint matrix is $\mathbf{N} = \mathbf{T}$ and its 'true' value is $\tilde{\mathbf{N}} = \tilde{\mathbf{T}} = \frac{1}{n}\sum\tilde{\mathbf{T}}_i$; also note that $\tilde{\mathbf{T}}_i\tilde{\mathbf{A}} = 2\tilde{A}\tilde{\mathbf{Z}}_i + \mathbf{P}\tilde{\mathbf{A}}$ for every i. Hence the Taubin's bias is

$$\mathbb{E}(\Delta_2\mathbf{A}_{\text{Taubin}}) \overset{\text{ess}}{=} -2\sigma^2\tilde{A}\,[0,0,0,1]^T. \tag{86}$$

Thus, the Taubin constraint $\mathbf{N} = \mathbf{T}$ cancels the second term in (82) *and* a half of the first term; it leaves only a half of the first term in place.

As a result, the Taubin fit's essential bias is twice as small as that of the Pratt fit. Given that their main terms (variances) are equal, we see that the Taubin fit is statistically more accurate than that of Pratt. We believe our analysis answers the question posed by Taubin [32] who intended to compare his fit to Pratt's.

'Hyperaccurate' algebraic fit. Our error analysis leads to another stunning discovery – an algebraic fit that has *no essential bias at all*. To our knowledge, this is the first such algorithm for curve fitting problems.

Let us set the constraint matrix to

$$\mathbf{N} = \mathbf{H} \stackrel{\text{def}}{=} 2\mathbf{T} - \mathbf{P} = \begin{bmatrix} 8\bar{z} & 4\bar{x} & 4\bar{y} & 2 \\ 4\bar{x} & 1 & 0 & 0 \\ 4\bar{y} & 0 & 1 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}. \tag{87}$$

Then one can easily see that $\mathbf{H}\tilde{\mathbf{A}} = 4\tilde{A}\,\frac{1}{n}\sum \tilde{\mathbf{Z}}_i + \mathbf{P}\tilde{\mathbf{A}}$, as well as $\tilde{\mathbf{A}}^T\mathbf{H}\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}}$, hence all the terms in (82) cancel out! The resulting essential bias vanishes:

$$\mathbb{E}(\Delta_2\mathbf{A}_{\text{Hyper}}) \stackrel{\text{ess}}{=} 0. \tag{88}$$

We call this fit *hyperaccurate*, or 'Hyper' for short. The term *hyperaccuracy* was introduced by Kanatani [21, 22] who was first to employ Taylor expansion up to the terms of order σ^4 for the purpose of comparing various algebraic fits and designing better fits.

We note that the Hyper fit is invariant under translations and rotations because its constraint matrix $\mathbf{H}$ is a linear combination of two others, $\mathbf{T}$ and $\mathbf{P}$, that satisfy the invariance requirements; see a proof in [14].

As any other algebraic circle fit, the Hyper fit minimizes the function $\mathcal{F}(\mathbf{A}) = \mathbf{A}^T\mathbf{M}\mathbf{A}$ subject to the constraint $\mathbf{A}^T\mathbf{N}\mathbf{A} = 1$ (with $\mathbf{N} = \mathbf{H}$), hence we need to solve the generalized eigenvalue problem $\mathbf{M}\mathbf{A} = \eta\mathbf{H}\mathbf{A}$ and choose the solution with the smallest non-negative eigenvalue η (see the end of Section 4).

We note that the matrix $\mathbf{H}$ is not singular, three of its eigenvalues are positive and one is negative (these facts can be easily derived from the following simple observations: $\det\mathbf{H} = -4$, $\operatorname{trace}\mathbf{H} = 8\bar{z} + 2 > 1$, and $\lambda = 1$ is one of its eigenvalues). If $\mathbf{M}$ is positive definite, then by Sylvester's law of inertia the matrix $\mathbf{H}^{-1}\mathbf{M}$ has the same signature as $\mathbf{H}$ does, i.e. the eigenvalues η

of $\mathbf{H}^{-1}\mathbf{M}$ are all real, exactly three of them are positive and one is negative. In this sense the Hyper fit is similar to the Pratt fit, as the constraint matrix $\mathbf{P}$ also has three positive and one negative eigenvalues.

The Hyper fit can be computed by a numerically stable procedure involving singular value decomposition (SVD). First, we compute the (short) SVD, $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, of the matrix $\mathbf{Z}$. If its smallest singular value, σ_4, is less than a predefined tolerance ε (we suggest $\varepsilon = 10^{-12}$), then $\mathbf{A}$ is the corresponding right singular vector, i.e. the fourth column of the $\mathbf{V}$ matrix. In the regular case ($\sigma_4 \geq \varepsilon$), one forms $\mathbf{Y} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^T$ and finds the eigenpairs of the symmetric matrix $\mathbf{Y}\mathbf{H}^{-1}\mathbf{Y}$. Selecting the eigenpair $(\eta, \mathbf{A}_*)$ with the smallest positive eigenvalue and computing $\mathbf{A} = \mathbf{Y}^{-1}\mathbf{A}_*$ completes the solution. The prior translation of the coordinate system to the centroid of the data set (which ensures that $\bar{x} = \bar{y} = 0$) makes the computation of $\mathbf{H}^{-1}$ particularly simple. The corresponding MATLAB code is available from our web page [14].

Transition between parameter schemes. Our next goal is to express the covariance and the essential bias of the algebraic circle fits in terms of the natural parameters $\mathbf{\Theta} = (a, b, R)^T$. Taking partial derivatives in (12) gives a 3×4 'Jacobian' matrix

$$\mathbf{J} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{B^2}{2A^2} & -\frac{1}{2A} & 0 & 0 \\ \frac{C^2}{2A^2} & 0 & -\frac{1}{2A} & 0 \\ -\frac{R}{A} - \frac{D}{2A^2R} & \frac{B}{4A^2R} & \frac{C}{4A^2R} & -\frac{1}{2AR} \end{bmatrix}. \tag{89}$$

Thus we have

$$\Delta_1\mathbf{\Theta} = \tilde{\mathbf{J}}\,\Delta_1\mathbf{A} \qquad \text{and} \qquad \Delta_2\mathbf{\Theta} = \tilde{\mathbf{J}}\,\Delta_2\mathbf{A} + \mathcal{O}_P(\sigma^2/n), \tag{90}$$

where $\tilde{\mathbf{J}}$ denotes the matrix $\mathbf{J}$ at the true parameters $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$. The remainder term $\mathcal{O}_P(\sigma^2/n)$ comes from the second order partial derivatives, for example

$$\Delta_2 a = (\nabla a)^T(\Delta_2\mathbf{A}) + \tfrac{1}{2}(\Delta_1\mathbf{A})^T(\nabla^2 a)(\Delta_1\mathbf{A}), \tag{91}$$

where $\nabla^2 a$ is the Hessian matrix of the second order partial derivatives of a with respect to (A, B, C, D). The last term in (91) can be actually discarded, as it is of order $\mathcal{O}_P(\sigma^2/n)$ because $\Delta_1\mathbf{A} = \mathcal{O}_P(\sigma/\sqrt{n})$. We collect all such terms in the remainder term $\mathcal{O}_P(\sigma^2/n)$ in (90).

Next we need a useful fact. Suppose a point (x_0, y_0) lies on the true circle $(\tilde{a}, \tilde{b}, \tilde{R})$, i.e.

$$(x_0 - \tilde{a})^2 + (y_0 - \tilde{b})^2 = \tilde{R}^2. \tag{92}$$

In accordance with our early notation we denote $z_0 = x_0^2 + y_0^2$ and $\mathbf{Z}_0 = (z_0, x_0, y_0, 1)^T$. We also put $u_0 = (x_0 - \tilde{a})/\tilde{R}$ and $v_0 = (y_0 - \tilde{b})/\tilde{R}$, and consider the vector $\mathbf{W}_0 = (u_0, v_0, 1)^T$. The following formula will be useful:

$$2\tilde{A}\tilde{R}\tilde{\mathbf{J}}\tilde{\mathbf{M}}^{-}\mathbf{Z}_0 = -n(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}_0, \tag{93}$$

where the matrix $(\mathbf{W}^T\mathbf{W})^{-1}$ appears in (59) and the matrix $\tilde{\mathbf{M}}^{-}$ appears in (77). The identity (93) is easy to verify directly for the unit circle $\tilde{a} = \tilde{b} = 0$ and $\tilde{R} = 1$, and then one can check that it remains valid under translations and similarities.

Equation (93) implies that for every true point $(\tilde{x}_i, \tilde{y}_i)$

$$4\tilde{A}^2\tilde{R}^2\tilde{\mathbf{J}}\tilde{\mathbf{M}}^{-}\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^T\tilde{\mathbf{M}}^{-}\tilde{\mathbf{J}}^T = n^2(\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}_i\mathbf{W}_i^T(\mathbf{W}^T\mathbf{W})^{-1}, \tag{94}$$

where $\mathbf{W}_i = (\tilde{u}_i, \tilde{v}_i, 1)^T$ denote the columns of the matrix $\mathbf{W}$, cf. (47). Summing up over i gives

$$4\tilde{A}^2\tilde{R}^2\tilde{\mathbf{J}}\tilde{\mathbf{M}}^{-}\tilde{\mathbf{J}}^T = n(\mathbf{W}^T\mathbf{W})^{-1}. \tag{95}$$

Variance and bias of algebraic circle fits in the natural parameters. Now we can compute the variance (to the leading order) of the algebraic fits in the natural geometric parameters. Notice that the third relation in (12) implies $\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}} = \tilde{B}^2 + \tilde{C}^2 - 4\tilde{A}\tilde{D} = 4\tilde{A}^2\tilde{R}^2$. Thus using (95) gives

$$\begin{aligned}
\mathbb{E}\big[(\Delta_1\mathbf{\Theta})(\Delta_1\mathbf{\Theta})^T\big] &= \mathbb{E}\big[\mathbf{J}(\Delta_1\mathbf{A})(\Delta_1\mathbf{A})^T\mathbf{J}^T\big] \\
&= n^{-1}\sigma^2(\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}})(\mathbf{J}\tilde{\mathbf{M}}^{-}\mathbf{J}^T) \\
&= \sigma^2(4\tilde{A}^2\tilde{R}^2 n^{-1}\mathbf{J}\tilde{\mathbf{M}}^{-}\mathbf{J}^T) \\
&= \sigma^2(\mathbf{W}^T\mathbf{W})^{-1}. \qquad (96)
\end{aligned}$$

Thus the variance of all the algebraic circle fits (to the leading order) coincides with that of the geometric circle fit, cf. (53). Therefore the difference between all the circle fits should be then characterized in terms of their biases, which we do next.

The essential bias of the Pratt fit is, due to (85),

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\Theta}}_{\text{Pratt}}) \overset{\text{ess}}{=} 2\sigma^2\tilde{R}^{-1}\big[0,0,1\big]^T. \tag{97}$$

Observe that the estimates of the circle center are essentially unbiased, and the essential bias of the radius estimate is $2\sigma^2/\tilde{R}$, which is independent of the number and location of the true points. We know that the essential bias of the Taubin fit is twice as small, hence

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\Theta}}_{\text{Taubin}}) \overset{\text{ess}}{=} \sigma^2\tilde{R}^{-1}\big[0,0,1\big]^T. \tag{98}$$

Comparing to (61) shows that the geometric fit has an essential bias that is twice as small as that of Taubin and four times smaller than that of Pratt. Therefore, the geometric fit has the smallest bias among all the popular circle fits, i.e. it is statistically most accurate.

The formulas for the bias of the Kåsa fit can be derived, too, but in general they are complicated. However recall that all our fits, including Kåsa, are independent of the choice of the coordinate system, hence we can choose it so that the true circle has center at $(0,0)$ and radius $\tilde{R}=1$. For this circle $\tilde{\mathbf{A}} = \frac{1}{\sqrt{2}}\,[1,0,0,-1]^T$, hence $\mathbf{P}\tilde{\mathbf{A}} = 2\tilde{\mathbf{A}}$ and so $\tilde{\mathbf{M}}^{-}\mathbf{P}\tilde{\mathbf{A}} = \mathbf{0}$, i.e. the middle term in (82) is gone. Also note that $\tilde{\mathbf{A}}^T\mathbf{P}\tilde{\mathbf{A}} = 2$, hence the last term in parentheses in (82) is $2\sqrt{2}\,[1,0,0,0]^T$.

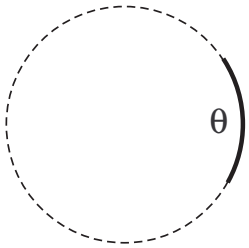

Figure 1: The arc containing the true points.

Next, assume for simplicity that the true points are equally spaced on an arc of size θ (a typical arrangement in many studies). Choosing the coordinate system so that the east pole $(1,0)$ is at the center of that arc (see Figure 1) ensures $\bar{y} = \overline{xy} = 0$. It is not hard to see now that

$$\tilde{\mathbf{M}}^{-}[1,0,0,0]^T = \tfrac{1}{4}(\overline{xx} - \bar{x}^2)^{-1}[\overline{xx}, -2\bar{x}, 0, \overline{xx}]^T. \tag{99}$$

	total MSE = variance + (ess. bias)^2 + rest of MSE			
Pratt	1.5164	1.2647	0.2500	0.0017
Taubin	1.3451	1.2647	0.0625	0.0117
Geom.	1.2952	1.2647	0.0156	0.0149
Hyper.	1.2892	1.2647	0.0000	0.0244

Table 2: Mean square error (and its components) for four circle fits ($10^4\times$values are shown). In this test $n = 100$ points are placed (equally spaced) along a semicircle of radius $R = 1$ and the noise level is $\sigma = 0.05$.

Using the formula (90) we obtain (omitting details as they are not so relevant) the essential bias of the Kåsa fit in the natural parameters (a, b, R):

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\Theta}}_{\text{Kasa}}) \overset{\text{ess}}{=} 2\sigma^2\big[0,0,1\big]^T - \frac{\sigma^2}{\overline{xx}-\bar{x}^2}\big[-\bar{x},0,\overline{xx}\big]^T. \tag{100}$$

The first term here is the same as in (97) (recall that $\tilde{R} = 1$), but it is the second term above that causes serious trouble: it grows to infinity because $\overline{xx} - \bar{x}^2 \to 0$ as $\theta \to 0$. This explains why the Kåsa fit develops a heavy bias toward smaller circles when data points are sampled from a small arc.

9 Experimental tests and conclusions

To illustrate our analysis of various circle fits we have run a few computer experiments where we set n true points equally spaced along a semicircle of radius $R = 1$. Then we generated random samples by adding a Gaussian noise at level $\sigma = 0.05$ to each true point, and after that applied various circle fits to estimate the parameters (a, b, R).

Table 2 summarizes the results of the first test, with $n = 100$ points; it shows the mean square error (MSE) of the radius estimate $\hat{R}$ for each circle fit (obtained by averaging over 10^7 randomly generated samples). The table also gives the breakdown of the MSE into three components. The first two are the variance (to the leading order) and the square of the essential bias, both computed according to our theoretical formulas. These two components

	total MSE = variance + (ess. bias)^2 + rest of MSE			
Pratt	25.5520	1.3197	25.0000	-0.76784
Taubin	7.4385	1.3197	6.2500	-0.13126
Geom.	2.8635	1.3197	1.5625	-0.01876
Hyper.	1.3482	1.3197	0.0000	-0.02844

Table 3: Mean square error (and its components) for four circle fits ($10^6\times$values are shown). In this test $n = 10000$ points are placed (equally spaced) along a semicircle of radius $R = 1$ and the noise level is $\sigma = 0.05$.

do not account for the entire mean square error, due to higher order terms which our analysis discarded. The remaining part of the MSE is shown in the last column, which is relatively small. (We note that only the total MSE can be observed in practice; all the other columns of this table are the results of our theoretical analysis.)

We see that all the circle fits have the same (leading) variance, which accounts for the 'bulk' of the MSE. Their essential bias is different, it is highest for the Pratt fit and smallest (zero) for the Hyper fit. Algorithms with smaller essential biases perform overall better, i.e. have smaller mean square error. The Hyper fit is the best in our experiment; it outperforms the (usually unbeatable) geometric fit.

To highlight the superiority of the Hyper fit, we repeated our experiment increasing the sample up to $n = 10000$, see Table 3 and Figure 2. We see that when the number of points is high, the the Hyper fit becomes several times more accurate than the geometric fit. Thus, our analysis disproves the popular belief in the statistical community that there is nothing better than minimizing the orthogonal distances.

Needless to say, the geometric fit involves iterative approximations, which are computationally intensive and subject to occasional divergence, while our Hyper fit is a fast non-iterative procedure, which is 100% reliable.

Summary. All the known circle fits (geometric and algebraic) have the same variance, to the leading order. The relative difference between them can be traced to higher order terms in the expansion for the mean square error. The

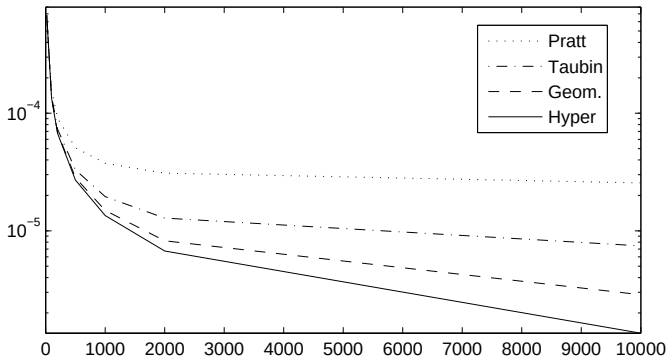

Figure 2: MSE for various circle fits (on the logarithmic scale) versus the sample size n (from 10 to 10^4).

second leading term in that expansion is the essential bias, for which we have derived explicit expressions. Circle fits with smaller essential bias perform better overall. This explains a poor performance of the Kåsa fit, a moderate performance of the Pratt fit, and a good performance of the Taubin and geometric fits (in this order). We showed that while there is a natural lower bound on the variance to the leading order (the KCR bound), there is no lower bound on the essential bias. In fact there exists an algebraic fit with zero essential bias (the Hyper fit), which outperforms the geometric fit in accuracy. We plan to perform a similar analysis for ellipse fitting algorithms in the near future.

The authors are grateful to the anonymous referees for many helpful suggestions. N.C. was partially supported by National Science Foundation, grant DMS-0652896.

References

[1] Y. Amemiya and W. A. Fuller. Estimation for the nonlinear functional relationship. *Annals Statist.*, 16:147–160, 1988.

[2] T. W. Anderson. Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics. *J. R. Statist. Soc. B*, 38:1–36, 1976.

[3] T. W. Anderson and T. Sawa. Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *J. R. Statist. Soc. B*, 44:52–62, 1982.

[4] A. Atieg and G. A. Watson. Fitting circular arcs by orthogonal distance regression. *Appl. Numer. Anal. Comput. Math.*, 1:66–76.

[5] M. Berman. Large sample bias in least squares estimators of a circular arc center and its radius. *CVGIP: Image Understanding*, 45:126–128, 1989.

[6] N. N. Chan. On circular functional relationships. *J. R. Statist. Soc. B*, 27:45–56, 1965.

[7] N. Chernov. Fitting circles to scattered data: parameter estimates have no moments. *Manuscript, see http://www.math.uab.edu/ chernov/cl.*

[8] N. Chernov and C. Lesort. Statistical efficiency of curve fitting algorithms. *Comp. Stat. Data Anal.*, 47:713–728, 2004.

[9] N. Chernov and C. Lesort. Least squares fitting of circles. *J. Math. Imag. Vision*, 23:239–251, 2005.

[10] N. Chernov and P. Sapirstein. Fitting circles to data with correlated noise. *Comput. Statist. Data Anal.*, 52:5328–5337, 2008.

[11] P. Delogne. Computer optimization of Deschamps' method and error cancellation in reflectometry. In *Proc. IMEKO-Symp. Microwave Measurement (Budapest)*, pages 117–123, 1972.

[12] W. A. Fuller. *Measurement Error Models.* L. Wiley & Son, New York, 1987.

[13] W. Gander, G. H. Golub, and R. Strebel. Least squares fitting of circles and ellipses. *BIT*, 34:558–578, 1994.

[14] http://www.math.uab.edu/ chernov/cl.

[15] S. H. Joseph. Unbiased least-squares fitting of circular arcs. *Graph. Mod. Image Process.*, 56:424–432, 1994.

[16] J. B. Kadane. Testing overidentifying restrictions when the disturbances are small. *J. Amer. Statist. Assoc.*, 65:182–185, 1970.

[17] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, Amsterdam, Netherlands, 1996.

[18] K. Kanatani. Cramer-Rao lower bounds for curve fitting. *Graph. Mod. Image Process.*, 60:93–99, 1998.

[19] K. Kanatani. For geometric inference from images, what kind of statistical model is necessary? *Syst. Comp. Japan*, 35:1–9, 2004.

[20] K. Kanatani. Optimality of maximum likelihood estimation for geometric fitting and the KCR lower bound. *Memoirs Fac. Engin. Okayama Univ.*, 39:63–70, 2005.

[21] K. Kanatani. Ellipse fitting with hyperaccuracy. *IEICE Trans. Inform. Syst.*, E89-D:2653–2660, 2006.

[22] K. Kanatani. Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis. *Int. J. Computer Vision*, 80:167–188, 2008.

[23] I. Kåsa. A curve fitting procedure and its error analysis. *IEEE Trans. Inst. Meas.*, 25:8–14, 1976.

[24] U. M. Landau. Estimation of a circular arc center and its radius. *CVGIP: Image Understanding*, 38:317–326, 1987.

[25] P. Meer. Robust techniques for computer vision. In G. Medioni and S. B. Kang, editors, *Emerging Topics in Computer Vision*, pages 107–190. Prentice Hall, 2004.

[26] Y. Nievergelt. A finite algorithm to fit geometrically all midrange lines, circles, planes, spheres, hyperplanes, and hyperspheres. *J. Numerische Math.*, 91:257–303, 2002.

[27] V. Pratt. Direct least-squares fitting of algebraic surfaces. *Computer Graphics*, 21:145–152, 1987.

[28] C. Rusu, M. Tico, P. Kuosmanen, and E. J. Delp. Classical geometrical approach to circle fitting – review and new developments. *J. Electron. Imaging*, 12:179–193, 2003.

[29] B. Sarkar, L. K. Singh, and D. Sarkar. Approximation of digital curves with line segments and circular arcs using genetic algorithms. *Pattern Recogn. Letters*, 24:2585–2595, 2003.

[30] H. Späth. Least-squares fitting by circles. *Computing*, 57:179–185, 1996.

[31] A. Strandlie, J. Wroldsen, R. Frühwirth, and B. Lillekjendlie. Particle tracks fitted on the Riemann sphere. *Computer Physics Commun.*, 131:95–108, 2000.

[32] G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to edge and range image segmentation. *IEEE Trans. Pattern Analysis Machine Intelligence*, 13:1115–1138, 1991.

[33] D. Umbach and K. N. Jones. A few methods for fitting circles to data. *IEEE Trans. Instrument. Measur.*, 52:1181–1885, 2003.

[34] K. M. Wolter and W. A. Fuller. Estimation of nonlinear errors-in-variables models. *Annals Statist.*, 10:539–548, 1982.

[35] S. J. Yin and S. G. Wang. Estimating the parameters of a circle by heteroscedastic regression models. *J. Statist. Planning Infer.*, 124:439–451, 2004.

[36] E. Zelniker and V. Clarkson. Maximum-likelihood estimation of circle parameters via convolution. *IEEE Trans. Image Proc.*, 15:865–876, 2006.

[37] E. Zelniker and V. Clarkson. A statistical analysis of the Delogne-Kåsa method for fitting circles. *Digital Signal Proc.*, 16:498–522, 2006.

[38] S. Zhang, L. Xie, and M. D. Adams. Feature extraction for outdoor mobile robot navigation based on a modified gaussnewton optimization approach. *Robotics Autonom. Syst.*, 54:277–287, 2006.