Statistical inference in circular structural model and fitting circles to noisy data *

A. Donner† A. Goldenshluger ‡

Abstract

It is well known that commonly used algorithms for circle fitting perform poorly when sampling distribution of the points is not symmetric with respect to the circle center, e.g., when the points are sampled from a circle arc. To overcome this difficulty we introduce and study a parametric circular structural model. In this model the points on the circumference are assumed to be sampled according to the von Mises distribution with unknown concentration and mean direction parameters. We develop maximum likelihood and method of moments estimators of the circle center and radius, and study their statistical properties. In particular, we show that the proposed maximum likelihood estimator is asymptotically normal and efficient. We also develop a test of uniformity for the sampling distribution along the circle. Based on the derived theoretical results we propose a numerically stable circle fitting algorithm, investigate its accuracy in a simulation study, and illustrate its behavior in a real data example.

Keywords: circular structural model, circle fitting, latent variables, maximum likelihood estimators, test of uniformity, von Mises distribution.

2000 AMS Subject Classification: 62F20, 65D10.

1 Introduction

The problem of circle fitting to noisy data scattered on the plane arises in such diverse areas as archaeology (Thom 1955), (Freeman 1977), microwave engineering (Delogne 1972), GPS localization (Beck & Pan 2012), robotics (Núñes et al. 2008), imaging and computer vision (Witzgall et al. 2006), (Kanatani et al. 2016), to name but a few. In this problem we are given data points $\mathcal{S}_n = \{(X_i, Y_i),\, i = 1, \ldots, n\}$ on the plane which are thought to be noisy versions of points on a circle, and the goal is to find the "best" circular fit to $\mathcal{S}_n$.

A natural and straightforward idea is to define an estimator $\hat{\theta} = (\hat{a}, \hat{b}, \hat{\rho})$ of the circle's parameters $\theta = (a, b, \rho)$ as solution to the optimization problem

$$\hat{\theta} = \arg\min_{a,b,\rho} \sum_{i=1}^{n} \left[(X_i - a)^2 + (Y_i - b)^2 - \rho^2\right]^2, \tag{1.1}$$

where (a, b) is the circle center and ρ is its radius. This method is called the *algebraic fit*, and it is routinely used in practice. The optimization problem in (1.1) is reduced to the standard linear least

*The research is supported by the ISF grant 220/21.

†Department of Statistics, University of Haifa, Haifa 31905, Israel. e-mail: *adonner@campus.haifa.ac.il.*

‡Department of Statistics, University of Haifa, Haifa 31905, Israel. e-mail: *goldensh@stat.haifa.ac.il.*

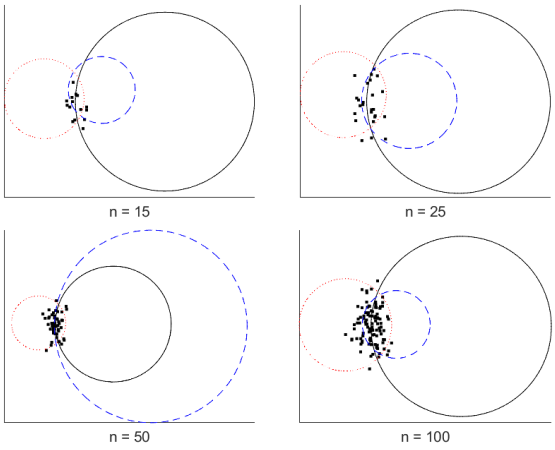


Figure 1: *The algebraic (red dotted line) and geometric (blue dashed line) fits fail to estimate the true circle (black line) when noisy observations (black dots) are sampled from a circular arc*

squares problem and solved efficiently. Another popular circle fitting algorithm is the *geometric fit* which is defined as solution to the following optimization problem:

$$\tilde{\theta} = \arg\min_{a,b,\rho} \sum_{i=1}^{n} \left[\sqrt{(X_i - a)^2 + (Y_i - b)^2} - \rho\right]^2. \tag{1.2}$$

The optimization problem in (1.2) is a non–linear least squares problem; it is non–convex, and it is typically solved using the Gauss–Newton, Levenberg–Marquardt or specially developed algorithms [see, e.g., Beck & Pan (2012)]. Both *algebraic* and *geometric* fits work well in practice when the distribution of the noisy data points along the circle is close to uniform, or if it is symmetric with respect to the circle center. However, in many practical situations, e.g., in computer vision applications, it is required to fit a circle to the data points which are sampled from an arc of the circle [see, e.g., Pratt (1987) and Landau (1987)]. Under these circumstances, performance of the *algebraic* and *geometric* fits is poor, and their accuracy does not improve as sample size n increases. A typical situation is displayed in Figure 1.

The main difficulty in the circle fitting problem is the presence of latent variables that determine sampling design along the circumference; the number of these variables grows with sample size. In mathematical terms the problem of circle fitting can be formulated as follows. We are given observations $\mathcal{S}_n = \{(X_i, Y_i),\ i = 1, \ldots, n\}$ generated by the equations

$$\begin{aligned} X_i &= a + \rho\cos\varphi_i + \xi_i \\ Y_i &= b + \rho\sin\varphi_i + \eta_i, \end{aligned} \tag{1.3}$$

where $\varphi_i \in [0, 2\pi)$, $i = 1, \ldots, n$ are unobservable angles, and (ξ_i, η_i), $i = 1, \ldots, n$ are independent bivariate normal random variables with zero mean and covariance matrix $\sigma^2 I$ which represent measurement errors. The problem is to estimate the circle center (a, b) and radius ρ from observations $\mathcal{S}_n$.

There are two common approaches for modeling latent variables $\{\varphi_i, i = 1, \ldots, n\}$. First, one can

regard $\{\varphi_i\}$ as unknown deterministic nuisance parameters; this assumption leads to the *circular functional model*. In this model the number of nuisance parameters grows with the sample size, and the classical work of Neyman & Scott (1948) shows that the maximum likelihood estimators can fail to be consistent under these circumstances [see also Kiefer & Wofowitz (1956)]. Secondly, $\{\varphi_i\}$ can be assumed to be independent identically distributed random variables with some distribution G on $[0, 2\pi)$. The latter assumption leads to the *circular structural model* which is a mixture model with mixing nuisance distribution G [see, e.g., Lindsay (1980)]. In this paper we focus on statistical inference in the circular structural model when the nuisance distribution G belongs to the parametric family of the von Mises distributions with unknown concentration and mean direction parameters.

Related literature. The problem of circle fitting is a subject of considerable literature; for a book–length survey of the area we refer to Chernov (2010). Below we provide a brief review of existing results that are most relevant to our work.

The majority of existing literature deals with numerical aspects of circle fitting and associated optimization problems. The most common approach for solving the geometric fit optimization problem (1.2) is the Levenberg-Marquardt algorithm. In addition, some special methods for solution of (1.2) have been developed. In particular, Landau (1987) suggested an iterative two step minimization algorithm, where the objective function is first minimized with respect to ρ for fixed a, b, and then with respect to a and b with the plugged–in current value of ρ. A similar approach was suggested by Späth (1996), but with a different parameterization. Kåsa (2011) suggested a simplistic approach for solving the algebraic fit problem (1.1) by reducing it to a standard linear least squares problem. These methods perform well when the data are almost noise–free, i.e., noise variance σ^2 is close to zero. However, for larger values of σ^2 these methods perform poorly when the observations are spread across less than half a circle. Motivated by this difficulty, Chernov & Ososkov (1984) suggested an approximation to the objective function of the geometric fit optimization problem. Specifically, the approximation is given by

$$\sum_{i=1}^{n} \left(\sqrt{(X_i - a)^2 + (Y_i - b)^2} - \rho\right)^2 \approx \frac{1}{4\rho^2} \sum_{i=1}^{n} \left((X_i - a)^2 + (Y_i - b)^2 - \rho^2\right)^2, \tag{1.4}$$

and it involves the objective function of the algebraic fit problem multiplied by factor $(2\rho)^{-2}$. The right hand side of (1.4) can be efficiently minimized numerically. The suggested method yields better results when the data are scattered across a small circular arc. Pratt (1987) and Taubin (1991) considered the same approximation (1.4) but with the algebraic parameterization of the circle equation, $A(x^2 + y^2) + Bx + Cy + D = 0$, which, with an additional constraint, provides an one–to–one mapping to the representation with parameters a, b, ρ. Both the Pratt and Taubin fits are represented as quadratic optimization problems and solved efficiently. Hence both approaches are very common and sometimes

their results serve as initial points for iterative solution of the geometric fit problem.

In contrast to numerical aspects, statistical properties of circle fitting algorithms have been studied to a much lesser extent. Chan (1965) investigated the circular functional model (1.3), where unobservable angles φ_i, $i = 1, \ldots, n$ are treated as unknown deterministic nuisance parameters. It is shown in the aforementioned paper that the maximum likelihood (ML) estimation of all parameters in this model (including nuisance parameters) leads to the geometric fit algorithm. Chan (1965) proved that necessary and sufficient conditions for consistency of the ML estimator of the circle center (a, b) are

$$\lim_{n\to\infty} \sum_{i=1}^{n} \cos\varphi_i = \lim_{n\to\infty} \sum_{i=1}^{n} \sin\varphi_i = 0. \tag{1.5}$$

The meaning of (1.5) is that unobservable angles are distributed symmetrically with respect to the circle center in asymptotics as $n \to \infty$. In what follows we refer to (1.5) as the *symmetry conditions*. Chernov (2011) showed that under the same assumptions as in Chan (1965), the ML estimators of the circle parameters have infinite expectation. Similar phenomenon holds in the functional linear regression model; see Anderson (1976). Berman & Culpin (1986) studied asymptotic bias of the algebraic fit estimators and showed that the bias converges to zero as $n \to \infty$ if and only if the symmetry conditions (1.5) hold. Kanatani & Rangarajan (2011) and Al-Sharadqah & Chernov (2009) considered both the algebraic and geometric fit estimators, and studied their mean squared errors in small noise asymptotics as $\sigma^2 \to 0$ and sample size n is fixed. Under these circumstances, both the algebraic and geometric fit estimators are consistent. Anderson (1981) introduced the circular structural model (1.3), where unobservable angles $\{\varphi_i\}$ are assumed to follow the uniform distribution on $[0, 2\pi)$ [see also Mardia & Holmes (1980, Section 6)]. Anderson (1981) derived estimating equations for the ML estimators in the considered setting and discussed some of their properties. It is worth mentioning that in the setting of Anderson (1981) the symmetry conditions (1.5) hold (the convergence in (1.5) should be understood as convergence in probability). Therefore both algebraic and geometric fit estimators are consistent.

The above brief review of existing literature shows that very little is known about statistical properties of circle fitting algorithms in the situation when the sampling distribution of angles is non–symmetric with respect to the circle center. In this paper we propose a parametric statistical model that allows non–symmetric sampling, solve related statistical inference problems for this model, and develop a numerically stable circle fitting algorithm with provable theoretical accuracy guarantees.

The paper contributions. In this paper our goal is two–fold. First, we propose a parametric circular structural model, when the unobservable angle variables $\{\varphi_i\}$ in (1.3) follow the von Mises distribution on the circle. The von Mises distribution is a symmetric unimodal distribution that is commonly used in modeling circular data [see, e.g., Mardia & Jupp (2000)]. It is a natural model for sampling design

concentrated on an arc of the circle. We study asymptotic properties of the ML estimators in the proposed model and show that they are asymptotically normal and efficient. Motivated by numerical difficulties in implementation of the ML estimators, we also develop the method of moments (MM) estimators of the circle parameters. The MM estimators are trivial in the uniform case; this case corresponds to the boundary of the parameter set in our model, where the concentration parameter of the von Mises distribution is equal to zero. Therefore, in order to distinguish between the uniform and non–uniform sampling distribution on the circle, and in order to improve stability of the circle fitting algorithm we develop a test of uniformity for the circular structural model. This test is a part of the proposed circle fitting algorithm; it is also of independent interest. Secondly, based on the derived theoretical results, a stable numerical algorithm for estimating the circle parameters is developed. The circle fitting algorithm works under broad conditions; it uses the MM estimates as initial points, and utilizes the uniformity test for efficient solution of the ML optimization problem. We illustrate performance of the proposed algorithm using extensive simulation experiments and compare it with the existing state-of-the-art circle fitting algorithms. In addition, application to analysis of a megalithic stone data is briefly discussed.

Organization. The rest of the paper is structured as follows. Section 2 presents the suggested parametric model along with discussion of its identifiability. Section 3 deals with the ML estimation in the circular structural model. Section 4 contains detailed development of the MM estimators. In Section 5 we develop a test of uniformity for the angle distribution in the circular structural model. Section 6 presents our final circle fitting algorithm, discusses its implementation, and reports on results of an extensive simulation study and on a real data example. Finally, Section 7 brings some concluding remarks. Appendix A collects some known facts and results about the von Mises distribution and the modified Bessel functions of the first kind; these facts are extensively used in the paper. The proofs of all results are given in Appendix B.

Notation. The following notation is used throughout the paper. The symbol $\mathbb{E}_\theta$ stands for the expectation with respect to the probability measure $\mathbb{P}_\theta$ of observations when the unknown parameter of the circular structural model is θ. The modified Bessel function of the first kind of order k is denoted $I_k(z)$, $z \in \mathbb{C}$. The ratios of the Bessel functions appear frequently in our calculations; we denote $R_k(z) := I_k(z)/I_0(z)$, $k = 0, \pm 1, \ldots$. For random variable φ having the von Mises distribution with concentration $\varkappa \geq 0$ and mean direction $\mu \in [0, 2\pi)$ we write $\varphi \sim \mathrm{vM}(\mu, \varkappa)$. The density of the von Mises distribution is denoted

$$g(\varphi) = \frac{e^{\varkappa \cos(\varphi - \mu)}}{2\pi I_0(\varkappa)}, \qquad \varphi \in [0, 2\pi). \tag{1.6}$$

2 Circular structural model

The suggested model we discuss in this paper is the structural version of (1.3). We assume that the available observations (X_i, Y_i), $i = 1, \dots, n$ are generated by equations

$$X_i = a + \rho \cos\varphi_i + \xi_i \tag{2.1}$$

$$Y_i = b + \rho \sin\varphi_i + \eta_i, \tag{2.2}$$

where (a, b) and $\rho > 0$ are the circle center and radius respectively, $(\xi_i, \eta_i) \sim \mathcal{N}_2(0, \sigma^2 I)$, $i = 1, \dots, n$ are independent bivariate normal measurement errors, and φ_i, $i = 1, \dots, n$ are independent identically distributed random variables, $\varphi_i \sim \mathrm{vM}(\mu, \varkappa)$. We assume that $\{\varphi_i, i = 1, \dots, n\}$ and $\{(\xi_i, \eta_i), i = 1, \dots, n\}$ are independent.

The introduced model is parametrized by $\theta := (a, b, \rho, \sigma^2, \mu, \varkappa)$, where a, b, and ρ are the parameters of interest, while σ^2, μ and $\varkappa$ are nuisance parameters. The joint probability density $f_\theta(x, y)$ of a generic observation (X, Y) is a mixture distribution

$$\begin{aligned} f_\theta(x,y) &= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \exp\Big\{ -\frac{1}{2\sigma^2}\Big[(x - a - \rho\cos\varphi)^2 + (y - b - \rho\sin\varphi)^2\Big]\Big\} g(\varphi)\mathrm{d}\varphi \\ &= \frac{1}{2\pi\sigma^2} \exp\Big\{ -\frac{1}{2\sigma^2}\Big[(x-a)^2 + (y-b)^2 + \rho^2\Big]\Big\} \frac{I_0\big(D_\theta(x,y)\big)}{I_0(\varkappa)}, \end{aligned} \tag{2.3}$$

where we have denoted

$$D_\theta(x,y) := \sqrt{A_\theta^2(x) + B_\theta^2(y)}, \quad A_\theta(x) := \varkappa\cos\mu + \frac{\rho}{\sigma^2}(x-a), \quad B_\theta(y) := \varkappa\sin\mu + \frac{\rho}{\sigma^2}(y-b).$$

The second line in (2.3) is obtained by substitution of (1.6), and using (A.1) from Appendix A.

The corresponding model is the family of distributions $\{f_\theta : \theta \in \Theta\}$, where $\theta = (a, b, \rho, \sigma^2, \mu, \varkappa)$, and

$$\Theta := \mathbb{R} \times \mathbb{R} \times (0, \infty) \times (0, \infty) \times [0, 2\pi) \times (0, \infty). \tag{2.4}$$

We call $\{f_\theta : \theta \in \Theta\}$ *the full circular structural model*. Note that the special case $\varkappa = 0$ is not included in the parameter set Θ, even though the value $\varkappa = 0$ belongs to the parameter set of the von Mises distribution. For $\varkappa = 0$ the von Mises distribution reduces to the uniform distribution on $[0, 2\pi)$, and the mean direction parameter μ is not identified. For $\theta = (a, b, \rho, \sigma^2, \mu, 0)$, $f_\theta(x, y)$ does not depend on μ,

$$f_\theta(x,y) = \frac{1}{2\pi\sigma^2} \exp\Big\{ -\frac{1}{2\sigma^2}\Big[(x-a)^2 + (y-b)^2 + \rho^2\Big]\Big\} I_0\Big(\frac{\rho}{\sigma^2}\sqrt{(x-a)^2 + (y-b)^2}\Big).$$

Thus, if the value $\varkappa = 0$ is included in the parameter set Θ, the full circular structural model becomes non–identifiable. In the case $\varkappa = 0$ the model collapses to $\{f_{\theta_0} : \theta_0 \in \Theta_0\}$, where $\theta_0 := (a, b, \rho, \sigma^2)$, and the parameter set is $\Theta_0 := \mathbb{R} \times \mathbb{R} \times (0, \infty) \times (0, \infty)$. We call $\{f_{\theta_0} : \theta_0 \in \Theta_0\}$ *the reduced circular structural model*. This model was suggested and studied by Anderson (1981).

The next statement establishes identifiability of the full model.

Proposition 2.1 *The full circular structural model* $\{f_\theta : \theta \in \Theta\}$ *with* Θ *defined in* (2.4) *is identifiable, i.e., for any pair of parameters* $\theta_1, \theta_2 \in \Theta$ *such that* $\theta_1 \neq \theta_2$ *one has* $P_{\theta_1} \neq P_{\theta_2}$*, where* P_θ *stands for the probability measure on* $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ *corresponding to* f_θ*.*

3 Maximum likelihood estimator

It follows from (2.3) that in the full circular structural model the log–likelihood function based on the sample $S_n = \{(x_i, y_i), i = 1, \ldots, n\}$ is

$$\ell_n(\theta; S_n) := \sum_{i=1}^{n} \ln\left[I_0\left(D_\theta(x_i, y_i)\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i - a)^2 + (y_i - b)^2 + \rho^2\right] - n\ln\left[I_0(\varkappa)\right] - n\ln(2\pi\sigma^2), \tag{3.1}$$

while the log–likelihood function in the reduced model is

$$\ell_{0,n}(\theta; S_n) := \sum_{i=1}^{n} \ln\left[I_0\left(\frac{\rho}{\sigma^2}\sqrt{(x_i - a)^2 + (y_i - b)^2}\right)\right] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left[(x_i - a)^2 + (y_i - b)^2 + \rho^2\right] + n\ln(2\pi\sigma^2). \tag{3.2}$$

The ML estimator of $\theta = (a, b, \rho, \sigma^2, \mu, \varkappa)$ in the full circular structural model is defined as

$$\tilde{\theta}_n := \arg\max_{\theta \in \Theta} \ell_n(\theta; S_n), \tag{3.3}$$

where Θ is given in (2.4). The ML estimator of $\theta_0 = (a, b, \rho, \sigma^2)$ in the reduced model is

$$\tilde{\theta}_{0,n} = \arg\max_{\theta \in \Theta_0} \ell_{0,n}(\theta; S_n), \tag{3.4}$$

where $\Theta_0 = \mathbb{R} \times \mathbb{R} \times (0, \infty) \times (0, \infty)$. Some properties of $\tilde{\theta}_{0,n}$ are studied in Anderson (1981).

3.1 Asymptotic normality

First we show that the full circular structural model $\{f_\theta : \theta \in \Theta\}$ defined in (2.3)–(2.4) is *regular* in the sense of the following definition; see Ibragimov & Hasminskii (1981, Chapter 1, Section 7).

Definition 3.1 *A statistical model* $\{f_\theta : \theta \in \Theta\}$ *is called* regular *if*

(i) $f_\theta(x)$ *is a continuous function on* Θ *for almost all* x*;*

(ii) *there exists finite Fisher information* $J(\theta)$ *at each point* $\theta \in \Theta$*;*

(iii) *function* $\psi_\theta(\cdot) = \nabla_\theta[f_\theta^{1/2}(\cdot)]$ *is continuous in the* $\mathbb{L}_2$*–sense, where* ∇_θ *stands for the gradient (the vector of partial first order derivatives with respect to the components of the vector* θ*).*

First, condition (i) holds trivially for $f_\theta(x, y)$ in (2.3). Secondly, differentiating the log–likelihood function $\ell_1(\theta; x, y)$ based on one observation (x, y) with respect to parameters $\theta := (a, b, \rho, \sigma^2, \mu, \varkappa)$,

taking into account that $I_0'(z) = I_1(z)$ [see (A.2) in Appendix A] and using notation $R_k(z) = I_k(z)/I_0(z)$ we obtain

$$\frac{\partial \ell_1(\theta; x, y)}{\partial a} = \frac{1}{\sigma^2}(x-a)\Big(1 - \frac{R_1(D_\theta)}{D_\theta}\Big) - \frac{\rho}{\sigma^2}\Big(\frac{R_1(D_\theta)}{D_\theta}\Big)\varkappa\cos\mu, \tag{3.5}$$

$$\frac{\partial \ell_1(\theta; x, y)}{\partial b} = \frac{1}{\sigma^2}(y-b)\Big(1 - \frac{R_1(D_\theta)}{D_\theta}\Big) - \frac{\rho}{\sigma^2}\Big(\frac{R_1(D_\theta)}{D_\theta}\Big)\varkappa\sin\mu, \tag{3.6}$$

$$\frac{\partial \ell_1(\theta; x, y)}{\partial \rho} = \frac{R_1(D_\theta)}{\sigma^2 D_\theta}\Big\{\frac{\rho}{\sigma^2}\big[(x-a)^2 + (y-b)^2\big] + \varkappa\big[(x-a)\cos\mu + (y-b)\sin\mu\big]\Big\} - \frac{\rho}{\sigma^2}, \tag{3.7}$$

$$\begin{aligned}\frac{\partial \ell_1(\theta; x, y)}{\partial \sigma^2} &= \frac{1}{\sigma^4}\big[(x-a)^2 + (y-b)^2\big]\Big(\frac{1}{2} - \frac{\rho^2 R_1(D_\theta)}{\sigma^2 D_\theta}\Big) \\ &\quad + \frac{\rho}{\sigma^4}\Big(\frac{1}{2} - \frac{R_1(D_\theta)}{D_\theta}\varkappa\big[(x-a)\cos\mu + (y-b)\sin\mu\big]\Big),\end{aligned} \tag{3.8}$$

$$\frac{\partial \ell_1(\theta; x, y)}{\partial \mu} = \frac{R_1(D_\theta)}{D_\theta}\frac{\varkappa\rho}{\sigma^2}\big[-(x-a)\sin\mu + (y-b)\cos\mu\big], \tag{3.9}$$

$$\frac{\partial \ell_1(\theta; x, y)}{\partial \varkappa} = -R_1(\varkappa) + \frac{R_1(D_\theta)}{D_\theta}\Big\{\varkappa^2 + \frac{\rho}{\sigma^2}\big[(x-a)\cos\mu + (y-b)\sin\mu\big]\Big\}, \tag{3.10}$$

where for brevity we write D_θ for $D_\theta(x, y)$. The Fisher information matrix is

$$J(\theta) := \mathbb{E}_\theta\Big\{\nabla_\theta\, \ell_1(\theta; X_1, Y_1)\big[\nabla_\theta\, \ell_1(\theta; X_1, Y_1)\big]^T\Big\}.$$

The finiteness of the diagonal elements of the Fisher information matrix follows from the above relationships, finiteness of all moments of (X, Y), and the fact that (Amos 1974)

$$\frac{x}{1 + \sqrt{x^2+1}} \le R_1(x) \le \frac{x}{\sqrt{x^2+4}}, \quad \forall x.$$

Therefore condition (ii) of Definition 3.1 is fulfilled. Lastly, function $f_\theta^{1/2}(x, y)$ is continuously differentiable with respect to $\theta \in \Theta$, so that condition (iii) also holds. Thus the model is regular in the sense of Definition 3.1.

The established regularity of the circular structural model allows us to apply powerful general theory of Ibragimov & Hasminskii (1981) for proving asymptotic optimality of ML estimators in the circular structural model. The main result of this section is given in the next statement.

Theorem 3.1 *Let $\Theta \subset \mathbb{R}^6$ be given in (2.4), and $J(\theta)$ be a non–singular matrix for all $\theta \in \Theta$. Then for any compact set $K \subset \Theta$ there exists $\epsilon > 0$ such that with $\mathbb{P}_\theta$–probability one uniformly in $\theta \in K$ one has*

$$\sqrt{n}J(\theta)(\tilde{\theta}_n - \theta) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\nabla_\theta\, \ell(\theta; X_i, Y_i) + O(n^{-\epsilon}), \quad n \to \infty,$$

and therefore

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, J^{-1}(\theta)), \quad n \to \infty.$$

Proof of Theorem 3.1 is given in Section B.2; it is based on verification of general conditions for asymptotic normality of ML estimators derived in Ibragimov & Hasminskii (1981). In fact, using results in Ibragimov & Hasminskii (1981, Section 3, Chapter III) one can prove that the family of distributions $\{f_\theta : \theta \in \Theta\}$ satisfies the LAN (local asymptotic normality) condition so that the ML estimator $\tilde{\theta}_n$ is asymptotically efficient. Our proof of Theorem 3.1 also demonstrates that the same results hold for the ML estimator in the reduced circular structural model; see (3.4). We note that although the reduced model has been studied in Anderson (1981), the cited paper does not contain results on asymptotic normality and efficiency of the associated ML estimator.

3.2 Remarks on numerical implementation

The ML estimator $\tilde{\theta}_n$ has very attractive theoretical properties. However, its numerical implementation presents several difficulties. The objective function in (3.3) is non-convex, so there are no guarantees that any reasonable numerical optimization procedure will find the point of global maximum. It turns out that the choice of good initial points for iterative optimization procedures is crucial for solution of (3.3). In the context of circle fitting with other algorithms a similar remark appears, e.g., in Kanatani & Rangarajan (2011) and Beck & Pan (2012).

The next small experiment demonstrates the crucial importance of initial points selection for numerical solution of the optimization problem in (3.3). In the reduced circular structural model Anderson (1981) suggested the following intuitive initializations for parameters a, b, ρ, and σ^2:

$$\begin{aligned} &a_0 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad b_0 = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \\ &\sigma_0^2 = \frac{1}{2n}\sum_{i=1}^{n}[(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2 - \rho_0^2], \quad \rho_0 = \frac{1}{n}\sum_{i=1}^{n}\sqrt{(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2}. \end{aligned} \tag{3.11}$$

In the full circular structural model there is no natural initialization for parameters μ and $\varkappa$ of the von Mises distribution. Therefore we draw the initial values for μ and $\varkappa$ randomly from uniform distributions on $[0, 2\pi)$ and $[0, 100]$ respectively. For 1000 replications we generate a sample of size 5000 according to (2.1)-(2.2) with $a = b = 0$, $\rho = 35$, $\sigma^2 = 4$, $\mu = 0.35\pi$ and $\varkappa = 45$. Then we solve (3.3) numerically using the interior-point algorithm implemented in `fmincon` function of the Matlab Optimization Toolbox. The initial points for the optimization procedure are chosen as described above.

Figure 2 shows that approximately in 50% of the simulation runs the optimization process fails to estimate the values of a, b and ρ properly. In addition, Figure 3 displays the results of circle fits obtained on the same data set when the same initializations for a, b, ρ, σ^2 are used with the only difference in the random initial values of μ_0, $\varkappa_0$. We observe that panel (a) of Figure 3 shows a good fit, while in panel (b) the optimization process results in a poor circle estimate. Since the simulation runs differ only in the initial values for parameters $\varkappa$ and μ, this clearly demonstrates high sensitivity of the optimization

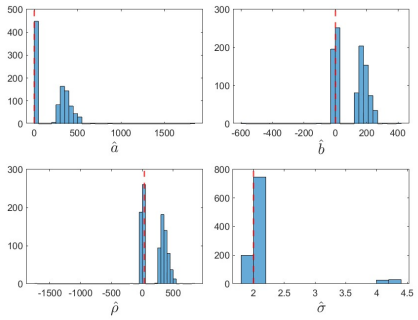
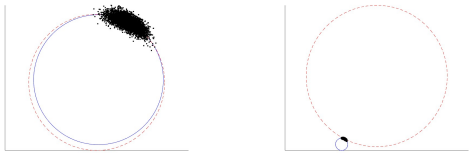


Figure 2: *Histograms of the ML estimates of a, b, ρ and σ from 1000 runs, where the true values $\theta = (0, 0, 35, 4, 0.35\pi, 45)$ are indicated by red dashed lines.*



(a) **(b)**

Figure 3: *Two circle fits for the same data set when the only change is in the initial values $\mu_0, \varkappa_0$. In panel (a) the estimated circle (red dashed line) is close to the true one (blue line). In panel (b) the numerical optimization algorithm diverges.*

process to initial values. It is worth mentioning that the same phenomenon holds for other choices of parameters.

In view of the presented results, the need for "good" initial points in order to obtain a stable circle fitting algorithm becomes clear. In the next section we develop method of moments estimators which can serve as initial points for the numerical optimization in (3.3). As we show below, the use of auxiliary method of moments estimators substantially improves stability of the proposed circle fitting algorithm.

4 Method of moments

In this section we develop the method of moments (MM) estimators for $\theta := (a, b, \rho, \sigma^2, \mu, \varkappa)$. It is remarkable that the mean direction parameter μ can be estimated separately from all other parameters. The proposed estimator of μ is intimately related to the principal component analysis (PCA) of the data from the circular structural model.

4.1 PCA of circular structural model

It will be convenient to rewrite the model (2.1)-(2.2) in the following form

$$Z = m + \rho e_\varphi + \varepsilon, \tag{4.1}$$

where $Z = [X, Y]^T$, $m = [a, b]^T$, and $e_\varphi = [\cos\varphi, \sin\varphi]^T$.

The next statement establishes expressions for the covariance matrix $\Sigma := \mathrm{cov}_\theta(Z)$ of the observations, and characterizes the mean direction parameter μ in terms of the spectral decomposition of Σ.

Proposition 4.1 (a). *The covariance matrix of Z, is given by*

$$\Sigma = \frac{1}{2}\begin{bmatrix} \rho^2(1 - R_1^2 + (R_2 - R_1^2)\cos 2\mu) + 2\sigma^2 & \rho^2(R_2 - R_1^2)\sin 2\mu \\ \rho^2(R_2 - R_1^2)\sin 2\mu & \rho^2(1 - R_1^2 - (R_2 - R_1^2)\cos 2\mu) + 2\sigma^2 \end{bmatrix}, \tag{4.2}$$

where R_i stands for $R_i(\varkappa)$, $i = 1, 2$.

(b). *The eigenvalues of Σ are given by*

$$\lambda_{\min}(\Sigma) = \tfrac{1}{2}\rho^2\left(1 - 2R_1^2(\varkappa) + R_2(\varkappa)\right) + \sigma^2, \tag{4.3}$$

$$\lambda_{\max}(\Sigma) = \tfrac{1}{2}\rho^2\left(1 - R_2(\varkappa)\right) + \sigma^2. \tag{4.4}$$

(c). *The vector e_μ associated with the mean direction parameter $\mu \in [0, 2\pi)$ is the eigenvector of matrix $\Sigma = \mathrm{cov}_\theta(Z)$ corresponding to the minimal eigenvalue $\lambda_{\min}(\Sigma)$, i.e.*

$$e_\mu = \arg\min_{\zeta\in[0,2\pi)} e_\zeta^T \Sigma e_\zeta = \arg\min_{e_\zeta:\|e_\zeta\|=1} e_\zeta^T \Sigma e_\zeta.$$

Several remarks on the results of Proposition 4.1 are in order. First, Proposition 4.1 shows that the vector e_μ associated with the mean direction parameter μ is the eigenvector of Σ corresponding to the smallest eigenvalue. Thus, e_μ is the second principal direction, while $e_{\mu-\pi/2}$ is the first principal direction. Secondly, if $\varkappa = 0$ then $\lambda_{\min}(\Sigma) = \lambda_{\max}(\Sigma) = \frac{1}{2}\rho^2 + \sigma^2$ because $R_k(0) = 0$ for all $k = 1, 2, \ldots$. This is the case of the reduced circular structural model where the mean direction is not identifiable. The similar situation holds as $\varkappa \to \infty$: both eigenvalues of Σ tend to σ^2, and in this case the entire model tends to be unidentifiable because the von Mises distribution degenerates to a point mass.

4.2 Estimation of mean direction

Using the property proved in part (c) of Proposition 4.1 we can easily develop an estimator of the mean direction μ as follows. Let $\hat{\Sigma}$ be the empirical covariance matrix, $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T$, where $\bar{Z} = \frac{1}{n}\sum_{i=1}^n Z_i$. Then, the estimator $\hat{\mu}$ is defined by

$$\hat{\mu} = \arg\min_{\zeta\in[0,2\pi)} e_\zeta^T \hat{\Sigma} e_\zeta = \arg\min_{\zeta\in[0,2\pi)} e_\zeta^T \Big[\frac{1}{n}\sum_{i=1}^n (Z_i - \bar{Z})(Z_i - \bar{Z})^T\Big] e_\zeta.$$

The eigenvectors are determined up to a sign; therefore the solution is not unique, and we have two possible estimators $\hat{\mu}$, and $\hat{\mu} + \pi$ of mean direction μ. In particular, two possible solutions are given by

$$\hat{\mu}^{(1)} = \arctan\Big(\frac{e_{\hat{\mu},2}}{e_{\hat{\mu},1}}\Big) \mod 2\pi, \qquad \hat{\mu}^{(2)} = (\hat{\mu}^{(1)} + \pi) \mod 2\pi,$$

where $e_{\hat{\mu},i}$, $i = 1, 2$ stands for the ith element of the empirical eigenvector corresponding to the minimal eigenvalue. Note, however, that if we restrict the mean direction parameter to belong, e.g., to $[0, \pi)$ then the solution is unique. Under these circumstances, using standard methods one can establish consistency and asymptotic normality of the developed mean direction estimator.

4.3 Estimation of other model parameters

Having two possible estimates $\hat{\mu}$ of the mean direction, we proceed with development of estimators of the other model parameters a, b, ρ, $\varkappa$ and σ^2. For this purpose we derive estimating equations based on the moments of X and Y and other functionals of the joint distribution of (X, Y).

Since $\mathbb{E}_\theta \cos\varphi = R_1(\varkappa)\cos\mu$ and $\mathbb{E}_\theta \sin\varphi = R_1(\varkappa)\sin\mu$ [see (A.6) and (A.7) in Appendix A],

$$\mathbb{E}_\theta(X) = a + \rho R_1(\varkappa)\cos\mu, \tag{4.5}$$

$$\mathbb{E}_\theta(Y) = b + \rho R_1(\varkappa)\sin\mu. \tag{4.6}$$

Furthermore, (4.2) implies that

$$\mathrm{var}_\theta(X) + \mathrm{var}_\theta(Y) = \rho^2(1 - R_1^2(\varkappa)) + 2\sigma^2, \tag{4.7}$$

and it follows from (4.3)-(4.4) that

$$\lambda := \lambda_{\max}(\Sigma) - \lambda_{\min}(\Sigma) = \rho^2(R_1^2(\varkappa) - R_2(\varkappa)). \tag{4.8}$$

The additional estimating equation is obtained from the formula for the moment generating function of a generic observation (X, Y). In the proof of Propostion 2.1 in Appendix B we show that for all t_1, t_2

$$\begin{aligned}\Psi_\theta(t_1,t_2) &= \mathbb{E}_\theta e^{t_1X+t_2Y}\\ &= \exp\Big\{at_1 + bt_2 + \tfrac{1}{2}\sigma^2(t_1^2+t_2^2)\Big\}\frac{1}{I_0(\varkappa)} I_0\Big(\sqrt{(\rho t_1 + \varkappa\cos\mu)^2 + (\rho t_2 + \varkappa\sin\mu)^2}\Big).\end{aligned}$$

Therefore, if for some $r > 0$ we put $t_1 = r\cos\mu,\ t_2 = r\sin\mu$ then

$$\mathbb{E}_\theta\big[\exp\big\{r(X - \mathbb{E}_\theta(X))\cos\mu + r(Y - \mathbb{E}_\theta(Y))\sin\mu\big\}\big] = \exp\left\{\frac{\sigma^2 r^2}{2} - \rho r R_1(\varkappa)\right\}\frac{I_0(\rho r + \varkappa)}{I_0(\varkappa)}. \tag{4.9}$$

In all our simulations given in Section 6 we select $r = 1$. However, in some situations the data scaling by factor r may be useful and can facilitate computations.

Observe that the left hand sides of equations (4.5), (4.6), (4.7), (4.8) and (4.9) can be estimated from the data, so that we have five relationships that can serve as a basis for estimating equations. The estimating equations are obtained from the above relationships as follows. Denote S_{xx}, S_{xy} and S_{yy} the empirical estimators of $\mathrm{var}_\theta(X)$, $\mathrm{cov}_\theta(X,Y)$ and $\mathrm{var}_\theta(Y)$ respectively:

$$S_{xx} := \frac{1}{n}\sum_{i=1}^n (X_i - \bar X)^2,\ S_{yy} := \frac{1}{n}\sum_{i=1}^n (Y_i - \bar Y)^2,\ \ S_{xy} := \frac{1}{n}\sum_{i=1}^n (X_i - \bar X)(Y_i - \bar Y), \tag{4.10}$$

where $\bar X = \frac{1}{n}\sum_{i=1}^n X_i$ and $\bar Y = \frac{1}{n}\sum_{i=1}^n Y_i$. From equations (4.5)-(4.6) we get

$$\begin{aligned}\hat a &= \bar X - \hat\rho R_1(\hat\varkappa)\cos\hat\mu,\\ \hat b &= \bar Y - \hat\rho R_1(\hat\varkappa)\sin\hat\mu,\end{aligned} \tag{4.11}$$

and from (4.8) we have

$$\hat\rho = \sqrt{\frac{\hat\lambda}{R_1^2(\hat\varkappa) - R_2(\hat\varkappa)}}. \tag{4.12}$$

Expressing (4.7) in terms of σ^2 and plugging in $\hat\rho^2$, yields the equation

$$\hat\sigma^2 = \frac{1}{2}\left(S_{xx} + S_{yy} - \frac{1 - R_1^2(\hat\varkappa)}{R_1^2(\hat\varkappa) - R_2(\hat\varkappa)}\cdot\hat\lambda\right). \tag{4.13}$$

Substituting (4.11)-(4.13) and $\hat{\mu}$ in (4.9) we construct the last estimating equation:

$$\frac{1}{n}\sum_{i=1}^{n}\exp\{r(X_i-\bar{X})\cos\hat{\mu}+r(Y_i-\bar{Y})\sin\hat{\mu}\}=\exp\left\{\frac{r^2\hat{\sigma}^2}{2}-\hat{\rho}rR_1(\hat{\varkappa})\right\}\frac{I_0(\hat{\rho}r+\hat{\varkappa})}{I_0(\hat{\varkappa})}. \tag{4.14}$$

Solution of equations (4.11)–(4.14) with respect to $\hat{a},\hat{b},\hat{\rho},\hat{\sigma}^2$ and $\hat{\varkappa}$ proceeds as follows. First, by substituting equations (4.12) and (4.13) in (4.14), we obtain a single equation with respect to $\hat{\varkappa}$ only. This equation is solved numerically. Then $\hat{a},\hat{b},\hat{\rho}$ and $\hat{\sigma}^2$ are found by substitution in equations (4.11)–(4.13). The role of parameter r should be clear from (4.14): the choice of r can downscale large values of the expression under the exponent on the left hand side, making it computable. Details on numerical solution of (4.14) are provided below in Section 6.

As discussed in Section 4.2, the estimator of the mean direction $\hat{\mu}$ is not unique; therefore we solve equation (4.14) twice and obtain two sets of the method of moments estimators corrsponding to $\mu^{(1)}$ and $\mu^{(2)}$.

It is worth noting that estimating equations (4.11)–(4.14) are valid only under assumption that $\varkappa>0$, i.e., the angle distribution is non–uniform. The numerical algorithm solving (4.14) becomes unstable for small values of $\varkappa$. Therefore in the implementation of the circle fitting algorithm we need to distinguish between the cases of the uniform and non–uniform angle distribution. This fact motivates the development of a test for uniformity in the circular structural model. The test is of independent interest; it can be used independently of any circle fitting procedure.

5 Testing uniformity

We consider the following testing problem: test

$$H_0:\varkappa=0 \quad \text{versus} \quad H_1:\varkappa>0$$

on the basis of observations $\{(X_i,Y_i),i=1,\dots,n\}$ generated by the model (2.1)–(2.2).

Note that both the null and alternative hypotheses are composite: under the null hypothesis the vector of unknown parameters $\theta=(a,b,\rho,\sigma^2,\mu,\varkappa)$ reduces to $\theta_0=(a,b,\rho,\sigma^2)$. The formulated testing problem has an intereting feature. Under the null hypoheses the distribution of the observations do not depend on the mean direction μ which is a nuisance parameter in this setting. As it was pointed out in Davies (1977), in such situations the generalized likelihood ratio tests are not directly applicable, and the standard chi–squared approximation of the logarithm of the likelihood ratio does not hold.

We base the proposed testing procedure on the following simple observation. It follows from statement (b) of Proposition 4.1 that under the null hypothesis H_0 the eigenvalues of the covariance matrix Σ coincide, i.e., $\lambda_{\min}(\Sigma)=\lambda_{\max}(\Sigma)$. It is easily verified that the eigenvalues of the empirical covariance

matrix $\hat{\Sigma}$ are

$$\hat{\lambda}_{1,2} = \frac{1}{2}\left(S_{xx} + S_{yy} \pm \sqrt{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}\right),$$

where S_{xx}, S_{yy} and S_{xy} are given in (4.10). Define

$$\widehat{T}_n := (S_{xx} - S_{yy})^2 + 4S_{xy}^2.$$

Under the null hypothesis $\hat{T}_n$ should be close to zero, and H_0 should be rejected for large values of $\widehat{T}_n$. The next statement derives asymptotic distribution of $\widehat{T}_n$ under the null hypothesis.

Theorem 5.1 (a). *Under H_0 one has*

$$\sqrt{n}\begin{bmatrix} S_{xx} - S_{yy} \\ 2S_{xy} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, sI), \quad n \to \infty,$$

where $s := \frac{1}{2}\rho^4 + 4\sigma^2\rho^2 + 4\sigma^4$, and I is the identity matrix.

(b). *In addition, if*

$$\widehat{M}_n := \frac{1}{2n}\sum_{i=1}^{n}\left\{(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2\right\}^2$$

then under H_0

$$\frac{n\widehat{T}_n}{\widehat{M}_n} = n\left(\frac{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}{\widehat{M}_n}\right) \xrightarrow{d} \chi_2^2, \quad n \to \infty,$$

where χ_2^2 is the chi–square distribution with two degrees of freedom.

Using the results of Theorem 5.1 we propose the following test for uniformity:

reject the null hypothesis if

$$\frac{n\widehat{T}_n}{\widehat{M}_n} = n\left(\frac{(S_{xx} - S_{yy})^2 + 4S_{xy}^2}{\widehat{M}_n}\right) > \chi_{2,1-\alpha}^2,$$

where $\chi_{2,1-\alpha}^2$ is the $(1-\alpha)$–quantile of the chi-square distribution with two degrees of freedom.

According to Theorem 5.1 the significance level of the proposed test is equal to α asymptotically. The test we have developed for the hypotheses $H_0 : \varkappa = 0$ versus $H_1 : \varkappa > 0$ is based on testing equality of eigenvalues of the covariance matrix of the random vector (X, Y). In fact, the proposed procedure tests $H_0 : T = 0$ versus $H_1 : T > 0$, where

$$T := (\mathrm{var}_\theta(X) - \mathrm{var}_\theta(Y))^2 + 4\left(\mathrm{cov}_\theta(X, Y)\right)^2 = \rho^4(R_2(\varkappa) - R_1(\varkappa))^2.$$

It is clear that $T = 0$ whenever $\varkappa = 0$, but we also have $T \to 0$ as $\varkappa \to \infty$. Therefore, we expect that the test will have low power for very large values of $\varkappa$. However, in the case of large $\varkappa$ accuracy of any statistical procedure will be poor because the model becomes non–identifiable as $\varkappa$ approaches infinity.

6 Circle fitting algorithm, simulation study and example

Now we are in a position to describe the proposed circle fitting algorithm. The algorithm proceeds in two stages. First, we obtain auxiliary MM estimates of the circle parameters. At the second stage, these auxiliary estimates serve as starting points for numerical maximization of the log–likelihood function. In this section we also present results of an extensive simulation study and a real data example.

6.1 Implementation

The MM equations are given in (4.11)–(4.14). Some useful bounds on the parameter estimates can be derived from properties of the modified Bessel functions of the first kind. These bounds restrict search regions for numerical solution of moment equations and log–likelihood maximization problem, and improve stability of the proposed circle fitting algorithm.

Bounds on search region for $\hat{\varkappa}$, $\hat{\sigma}^2$ and $\hat{\mu}$. In view of (A.4) in Appendix A, $R_1^2(x) - R_2(x) \geq 0$ for all $x \geq 0$. Because $\hat{\lambda} \geq 0$, we have that

$$\frac{(1 - R_1^2(\hat{\varkappa}))\hat{\lambda}}{R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa})} \geq 0.$$

This leads to the inequality

$$\hat{\sigma}^2 = \frac{1}{2}\left(S_{xx} + S_{yy} - \frac{1 - R_1^2(\hat{\varkappa})}{R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa})} \cdot \hat{\lambda}\right) \leq \tfrac{1}{2}\left(S_{xx} + S_{yy}\right),$$

which provides an upper bound for $\hat{\sigma}^2$, in addition to the obvious lower bound of non-negativity. If $\hat{\sigma}^2 \geq 0$ then it follows from (4.13) that

$$\frac{1 - R_1^2(\hat{\varkappa})}{R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa})} \leq \frac{1}{\hat{\lambda}}\left(S_{xx} + S_{yy}\right).$$

The function $x \mapsto (1 - R_1^2(x))/(R_1^2(x) - R_2(x))$ is monotone decreasing on the positive real line. This property is easily proved by differentiating and observing that the derivative is strictly negative whenever $x > 0$. Therefore if $\varkappa_L$ denotes the value of $\varkappa$ that solves the equation

$$\frac{1 - R_1^2(\varkappa)}{R_1^2(\varkappa) - R_2(\varkappa)} = \frac{1}{\hat{\lambda}}(S_{xx} + S_{yy}), \tag{6.1}$$

then $\varkappa_L$ is a lower bound for $\hat{\varkappa}$. We remark that equation (6.1) is solved easily using the bisection method. As for the upper bound on $\hat{\varkappa}$, we note that numerical evaluation of Bessel functions fails for large values of the argument. Therefore in all our simulations we impose the upper bound of $\hat{\varkappa} \leq 500$.

Additional bounds can be derived on the mean direction $\hat{\mu}$. From (4.2) we have that $\mathrm{cov}_\theta(X, Y) = \frac{1}{2}(R_2(\varkappa) - R_1^2(\varkappa))\sin(2\mu)$. Since $R_2(x) - R_1^2(x) \leq 0$ for all $x \geq 0$, the covariance sign is determined by the sign of $\sin(2\mu)$. For negative values of $\mathrm{cov}_\theta(X, Y)$, $\sin(2\mu)$ is positive; hence $\mu \in (0, \frac{1}{2}\pi) \cup (\pi, \frac{3}{2}\pi)$. For

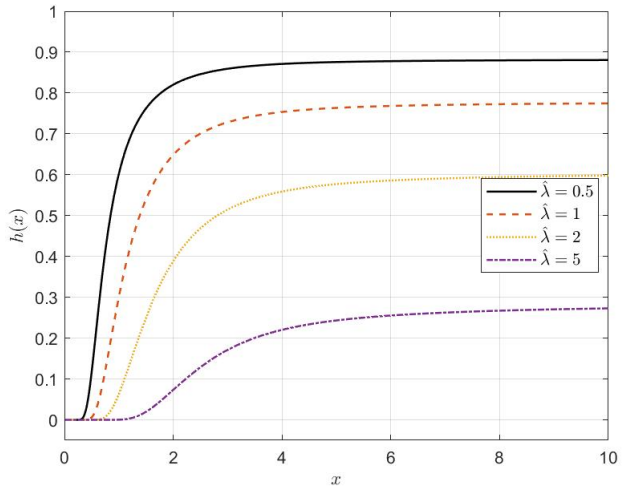


Figure 4: *Function $h(x)$ in (6.2) for $r = 1$ and different values of $\hat{\lambda}$.*

positive covariance values, $\mu \in (\frac{1}{2}\pi, \pi) \cup (\frac{3}{2}\pi, 2\pi)$. Using these facts, we can compute S_{xy} and according to its sign we can restrict the search for $\hat{\mu}$. Specifically, each value of $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$ has corresponding bounds, and we restrict the numerical algorithm to a specific quadrant.

Numerical solution of moment equations. Given lower and upper bounds on $\hat{\mu}, \hat{\varkappa}, \hat{\sigma}^2$ and estimate $\hat{\mu}$ of μ, we proceed to solve (4.11)–(4.14). First, substitution of (4.13) in (4.14) yields the equation

$$\begin{aligned} h(\hat{\varkappa}) &:= \exp\Big\{ -\frac{r^2\hat{\lambda}(1-R_1^2(\hat{\varkappa}))}{4[R_1^2(\hat{\varkappa}) - R_1(\hat{\varkappa})]} - \frac{r\sqrt{\hat{\lambda}}R_1(\hat{\varkappa})}{\sqrt{R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa})}} \Big\} \frac{I_0\Big(\hat{\varkappa} + r\sqrt{\hat{\lambda}/(R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa}))}\Big)}{I_0(\hat{\varkappa})} \\ &= \exp\Big\{ -\frac{r^2}{4}(S_{xx} + S_{yy}) \Big\} \cdot \frac{1}{n}\sum_{i=1}^{n} \exp\Big\{ r(X_i - \bar{X})\cos\hat{\mu} + r(Y_i - \bar{Y})\sin\hat{\mu} \Big\}. \end{aligned} \tag{6.2}$$

We recall that $\hat{\rho} = \sqrt{\hat{\lambda}/(R_1^2(\hat{\varkappa}) - R_2(\hat{\varkappa}))}$. The right hand side of (6.2) is a statistic, and the equation should be solved with respect to $\hat{\varkappa}$. In all our simulations we select $r = 1$.

Although function $h(\hat{\varkappa})$ on the left hand side of (6.2) depends on $\hat{\varkappa}$ in a rather complicated way, the equation (6.2) can be solved efficiently. This is equation of the type $h(x) = \text{const}$, and Figure 4 displays function $h(x)$ for $r = 1$ and different values of $\hat{\lambda}$. Note that function h depends on λ via $r\sqrt{\lambda}$ only, so curves in Figure 4 can be attributed to corresponding values of $r^2\lambda$. We solve (6.2) on the interval $[\varkappa_L, 500]$ using the trust region reflective algorithm, implemented in the function `lsqnonlin` of the Matlab Optimization Toolbox. The stopping criteria *MaxFunctionEvaluations, OptimalityTolerance* and *StepTolerance* were modified to 1e10, 1e−20 and 1e−20 respectively to account for flatness of function h for large values of the argument. Given estimates of $\varkappa$ and μ, we plug them in (4.11)-(4.13) and obtain the moment estimators of a, b, ρ, and σ^2.

Numerical solution of the likelihood maximization problem. Depending on the results of the uniformity test, the ML estimator is obtained by solution of the optimization problem for the full or reduced circular structural model. In both cases we use the interior-point algorithm implemented in `fmincon` function of the Matlab Optimization Toolbox. For the full circular model the initial points are the auxiliary MM estimators obtained as described above. For the reduced circular structural model we use the initial points (3.11), as proposed in Anderson (1981). Another remark concerns with the MM estimate $\hat{\sigma}^2$ as a starting point for the numerical likelihood maximization. Our experience shows that whenever $\hat{\sigma}^2$ is too close to zero (the boundary of parameter space), the numerical likelihood maximization with such initial points often diverges. This divergence is caused by numerical errors in computation of the log–likelihood function and its derivatives for small values of σ^2; see formulas (3.1) and (3.2). To cope with this problem, as an initial point for σ^2 we select the MM estimate $\hat{\sigma}^2$ perturbed by addition of an exponential random variable with unit expectation.

Circle fitting algorithm. The proposed circle fitting algorithm includes the following steps.

1. Perform the test of uniformity described in Section 5. In our implementation we use $\alpha = 0.05$.

2. If the test does not reject the null hypothesis H_0, then set $\hat{\varkappa} = 0$ and proceed with estimating $\theta_0 = (a, b, \rho, \sigma^2)$ in the reduced circular structural model as described in Anderson (1981). The corresponding optimization problem is given in (3.4). The initial points for the numerical procedure are chosen according to (3.11).

3. If the test rejects the null hypothesis then perform the following steps.

   (i). Compute the estimate of the mean direction $\hat{\mu}$ as described in Section 4.2 and obtain two potential values $\hat{\mu}^{(1)}$ and $\hat{\mu}^{(2)}$.

   (ii). For $\hat{\mu}^{(i)}$, $i = 1, 2$ solve the MM equations (4.11)–(4.14) and find two sets of moments estimates for θ.

   (iii). Solve the optimization problem in (3.3) twice, each time for different set of initial points given by the MM estimators of step (ii).

   (iv). Among the two solutions obtained in the step (iii) above, choose the one having higher log–likelihood value. This is the sought maximum likelihood estimate and the algorithm output.

The Matlab code implementing the above circle fitting algorithm along with documentation can be found on the Github repository. The average running time of the algorithm is ~ 0.1 seconds for $n = 50$, and ~ 2.7 seconds for $n = 10,000$ on a desktop with 32GB RAM and i7 10th generation processor.

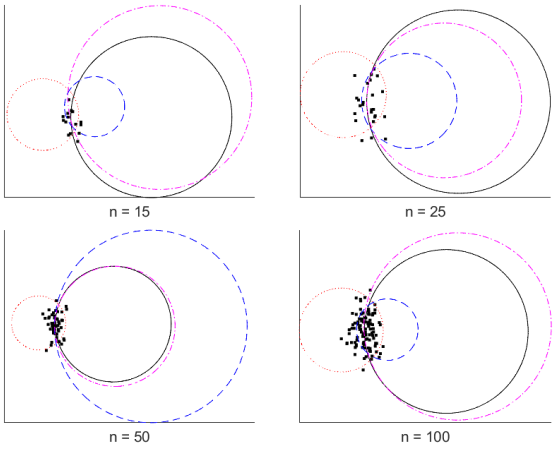


Figure 5: *Realizations of the algebraic (red dotted line), geometric (blue dashed line) and proposed fits (magenta dotted dashed line) for the true circle (black line) when noisy observations (black dots) are sampled from a circular arc.*

6.2 Simulation experiments

The goal of the following simulation study is to assess accuracy of the proposed algorithm under different combinations of model parameters, to study its sensitivity to distributional assumptions, and to compare it with commonly used circle fitting algorithms.

We begin with revisiting the example depicted in Figure 1, but now with the added circle fit proposed in this paper. Figure 5 displays a realization of the algebraic, geometric and proposed fits for different sample sizes when the points are sampled from an arc of the circle. The circular structural model studied in this paper is more adequate for such situations, and it is expected that performance of the proposed algorithm is preferable in such situations.

The following experiments have been performed in our simulation study.

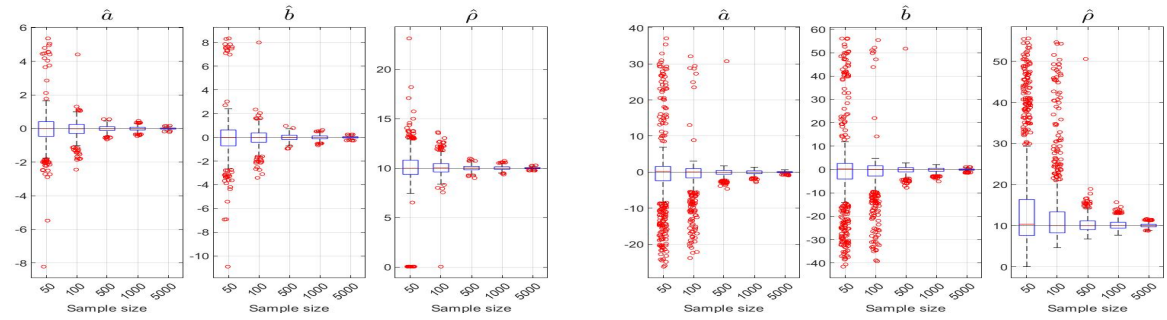
1. Experiment 1: we assess performance of the algorithm for different combinations of values of $\rho, \sigma^2, \varkappa$ and sample sizes n. The data are generated from the circular structural model with the von-Mises angle distribution.

2. Experiment 2: we compare our algorithm with the Levenberg-Marquardt algorithm (LM), the Pratt and Taubin fits for different combinations of parameter values.

3. Experiment 3: we study sensitivity of our algorithm to misspecification of angle distribution; the angles are generated from the uniform and beta distributions.

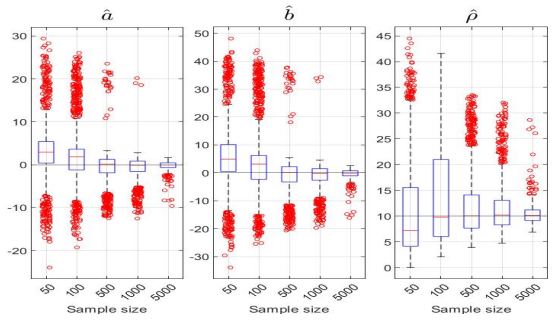
In our simulations we use the Matlab code implementing the Levenberg-Marquardt, Pratt and Taubin fits that was developed by N. Chernov.[1] All results reported below are based on $N = 1000$ simulation runs.

[1]The code is publicly available on the website https://people.cas.uab.edu/~mosya/cl/MATLABcircle.html

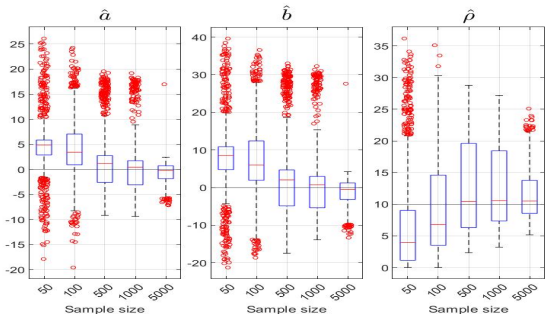


(a) $\varkappa = 5,\ \mu = \pi/3$

(b) $\varkappa = 20,\ \mu = \pi/3$



(c) $\varkappa = 50,\ \mu = \pi/3$



(d) $\varkappa = 80,\ \mu = \pi/3$

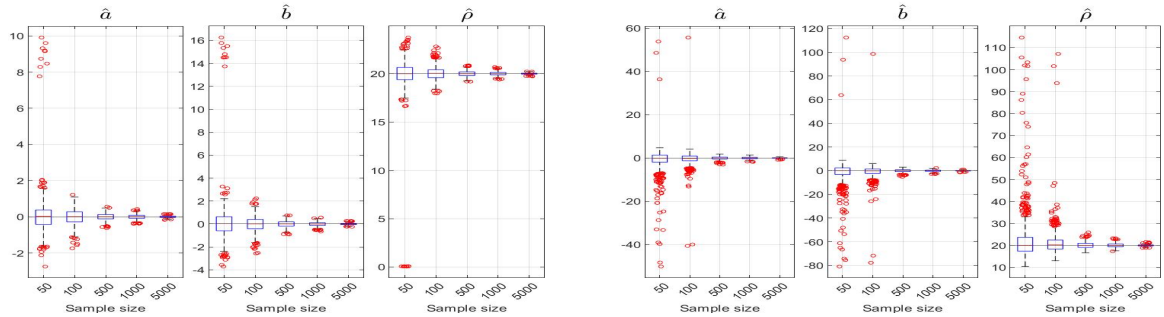
Figure 6: *Simulation results for* $a = b = 0,\ \rho = 10$ *and* $\sigma = 1$.

Experiment 1. First, we consider the simulation setup reported in Anderson (1981) and Chan (1965); here $a = b = 0, \rho = 10, \sigma = 1$. In addition, we fix $\mu = \pi/3$. Figure 6 presents boxplots of the obtained estimators for increasing sample sizes, where each panel is based on different value of $\varkappa = 5, 20, 50, 80$. The black solid lines indicate the true values of parameters. We can see that larger values of concentration parameter $\varkappa$ resulted in higher variability of the proposed fit. The graphs demonstrate that estimation accuracy improves with the sample size.

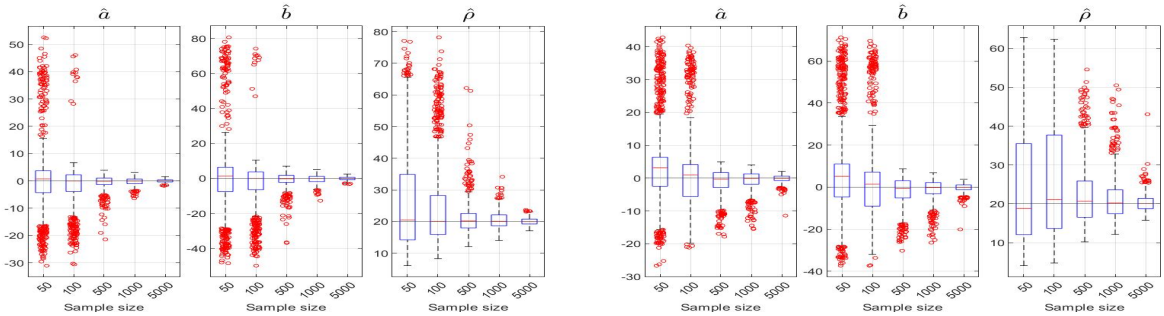
Figure 7 displays the results for the same values of $\varkappa$ and μ but with larger radius $\rho = 20$. In comparison with Figure 6 we observe some improvement in the estimator performance in terms of bias, but still the resulted variance is quite large. Note that the results in Figures 6 and 7 correspond to the high–noise scenario; here the noise standard deviation is $\sigma = 1$

The situation is completely different in the small–noise scenario. The setting with $\sigma = 0.05$ and $\rho = 1$ appears frequently in the literature on circle fitting [see, e.g., Berman (1989), Al-Sharadqah & Chernov (2009)]. Figure 8 presents the results, and we observe that the variance reduced dramatically compared to the case where $\rho = 20$, $\sigma = 1$, even though the ratio ρ/σ is the same as in the setting of Figure 7.

Experiment 2. In this experiment we compare our algorithm with three state–of–the–art circle fitting algorithms. The Levenberg-Marquardt fit is a general purpose non–linear least squares optimization

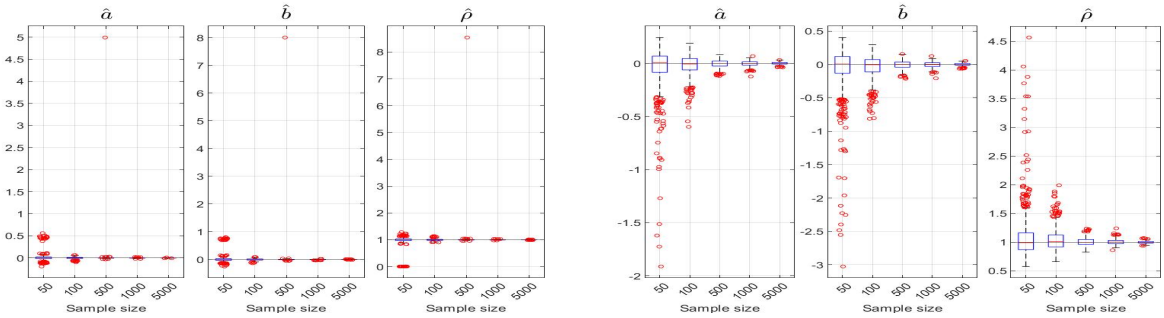


(a) $\varkappa = 5,\ \mu = \pi/3$ (b) $\varkappa = 20,\ \mu = \pi/3$

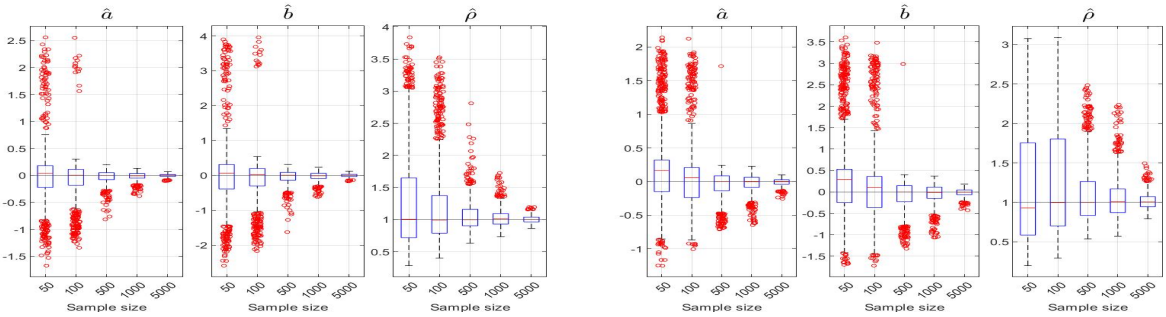


(c) $\varkappa = 50,\ \mu = \pi/3$ (d) $\varkappa = 80,\ \mu = \pi/3$

Figure 7: *Simulation results for* $a = b = 0,\ \rho = 20$ *and* $\sigma = 1$.

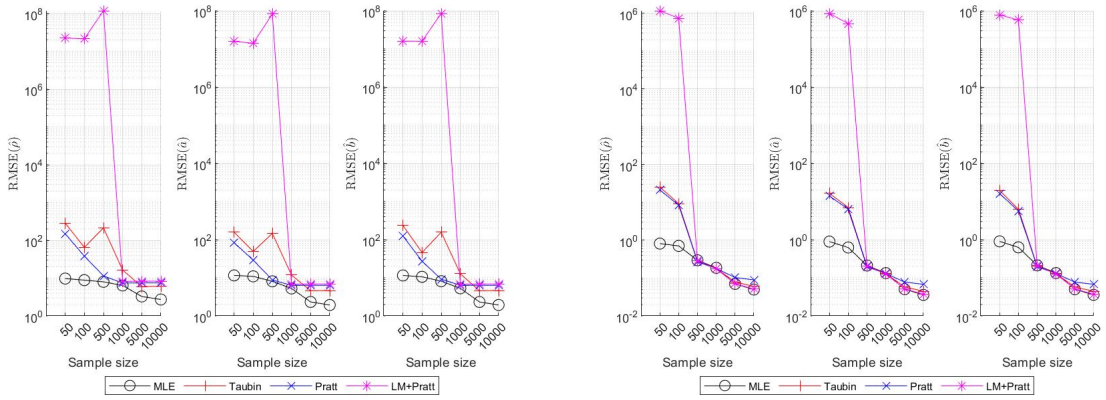


(a) $\varkappa = 5,\ \mu = \pi/3$ (b) $\varkappa = 20,\ \mu = \pi/3$



(c) $\varkappa = 50,\ \mu = \pi/3$ (d) $\varkappa = 80,\ \mu = \pi/3$

Figure 8: *Simulation results for* $a = b = 0,\ \rho = 20$ *and* $\sigma = 0.05$.



(a) $a = b = 0, \rho = 10, \sigma = 1$ and $\varkappa = 60$ **(b)** $a = b = 0, \rho = 1, \sigma = 0.05$ and $\varkappa = 60$

Figure 9: *The root MSE on logarithmic scale for different algorithms when* $\varphi_i \sim \mathrm{vM}(\frac{\pi}{4}, \varkappa)$.

algorithm applied to the geometric fit optimization problem (1.2). The Pratt and Taubin fits solve optimization problems of the algebraic fit type with algebraic parametrization of the circle equation. In particular, the Pratt fit solves the following optimization problem

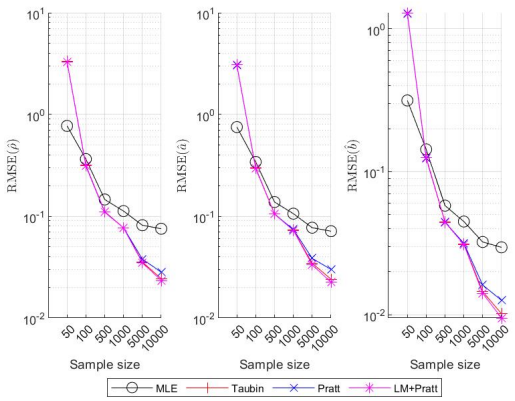
$$\begin{aligned} \min_{A,B,C,D} \quad & \sum_{i=1}^{n} \big(A(X_i^2 + Y_i^2) + BX_i + CY_i + D\big)^2 \\ \text{s.t.} \quad & B^2 + C^2 - 4AD = 1, \end{aligned}$$

while the Taubin fit is a solution to

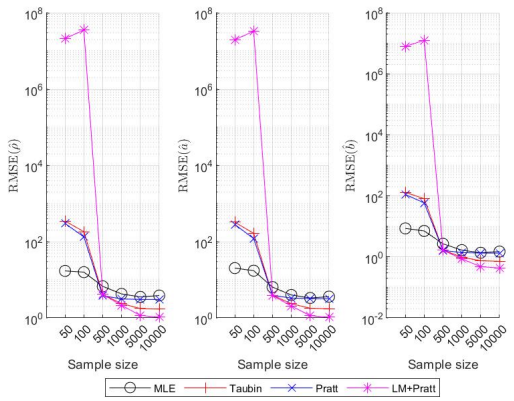
$$\begin{aligned} \min_{A,B,C,D} \quad & \sum_{i=1}^{n} \big(A(X_i^2 + Y_i^2) + BX_i + CY_i + D\big)^2, \\ \text{s.t.} \quad & 4A^2(\overline{X^2} + \overline{Y^2}) + 4AB\bar{X} + 4AC\bar{Y} + B^2 + C^2 = 1, \end{aligned}$$

where $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, and $\overline{Y^2} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$. Figure 9 displays mean squared errors of different methods in estimating the circle center and radius when the angles follow the von Mises distribution with indicated parameters. We observe that the ML estimator proposed in this paper is superior. In addition, for larger values of σ, the ML estimator shows even better performance in comparison with the other circle fitting algorithms.

Experiment 3. Now we consider the same settings as in Experiment 2 but with a different angle distribution so that the distributional assumptions of the ML estimator do not hold. First, we set $\varphi_i \sim U(0, \pi/4)$. Figure 10 presents the results for $\sigma = 0.05$ and $\sigma = 1$. We observe that the ML estimator still shows good performance for small samples sizes, but not as good as the other circle fitting algorithms if sample size grows. For larger measurement error variance σ^2, the ML estimator is still preferable for sample sizes $n \leq 500$, and the differences between the mean squared errors of the considered methods become smaller as n grows.

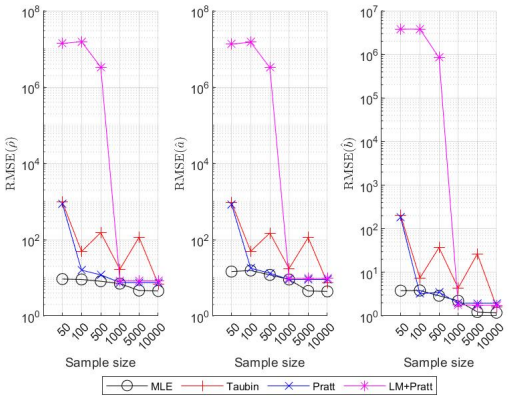


(a) $a = b = 0, \rho = 1, \sigma = 0.05$

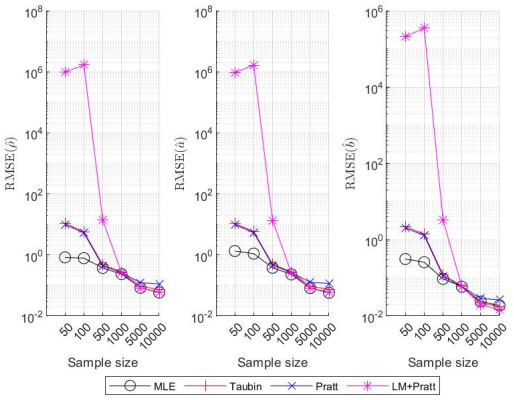


(b) $a = b = 0, \rho = 10, \sigma = 1$

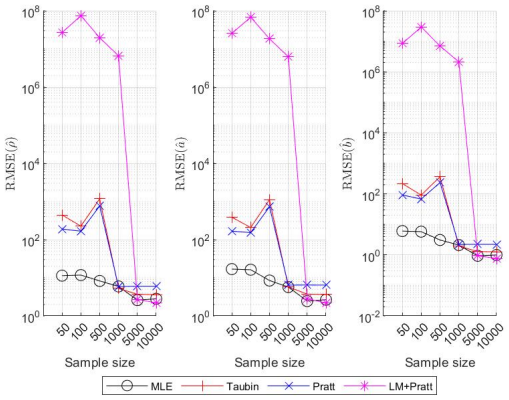
Figure 10: *The root MSE on logarithmic scale for different algorithms when* $\varphi_i \sim U(0, \pi/4)$.



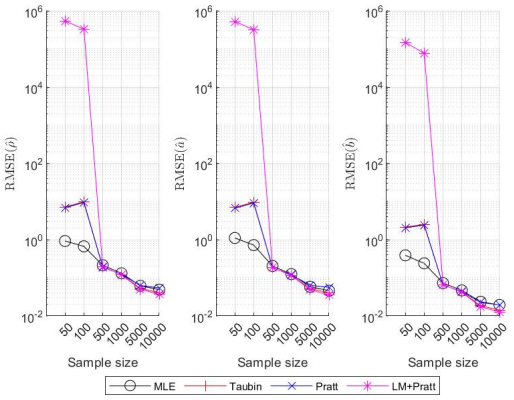
(a) The root MSE values on logarithmic scale for $a = b = 0, \rho = 10, \sigma = 1$



(b) The root MSE values on logarithmic scale for $a = b = 0, \rho = 1, \sigma = 0.05$



(c) The root MSE values on logarithmic scale for $a = b = 0, \rho = 10, \sigma = 1$



(d) The root MSE values on logarithmic scale for $a = b = 0, \rho = 1, \sigma = 0.05$

Figure 11: *In panels (a)-(b) the angles are generated from* Beta$(2, 5)$*, and in panels (c)-(d) from* Beta$(1, 6)$.

However, if we consider the setting with angles generated from Beta(2,5) and Beta(1,6) distributions then even though the distributional assumptions do not hold, the ML estimator outperforms the other circle fitting algorithms. Figure 11 displays the corresponding graphs.

In addition to Experiment 3, we also tested performance of the proposed algorithm in the case when the angle distribution is symmetric with respect to the circle center, but it is not uniform, i.e., the distributional assumptions of the proposed model do not hold. In particular, we considered a mixture of two von Mises distributions with the same concentration parameter $\varkappa$ and mean directions 0 and π. Under these circumstances, in general, the proposed estimator is inferior to the algebraic and geometric fit. This is particularly pronounced for large values of $\varkappa$ when the data are effectively sampled from two disjoint antipodally symmetric arcs of the circle.

Summary. The above simulation results demonstrate that cases with high concentration of the angles (large values of $\varkappa$) are more difficult, and they result in larger bias and variance of the ML estimator. In addition, accuracy is sensitive to changes in σ^2, and the proposed estimator shows better performance in comparison to the algebraic and geometric fits, when the noise variance is larger. In terms of 'signal-to-noise' ratio ρ/σ, it is clear that larger ratio yields better performance. In addition, for fixed value of ρ/σ the variance of the ML estimator is lower when ρ and σ are smaller. Finally, the ML estimator tends to obtain good results even when the distributional assumptions on the unobservable angle variables do not hold, but the data points are sampled from a single arc.

6.3 Real data example

In this section we illustrate usefulness of the proposed circular structural model by applying it to analysis of a megalithic stone data set. There is vast literature on megalithic stone structures in Great Britain that appear to lie on a circle or ellipse. This literature is concerned with reasons for erection of such structures, possible astronomical significance of stone alignments and so on. Thom (1955) surveyed more than two hundred different megalithic sites and, based on the collected data, hypothesized that the diameters of circular stone structures are integer multiples of *the megalithic yard* which is 0.83m or 2.72ft. This hypothesis stimulated extensive research on statistical analysis of Thom's megalithic stone data; we refer, e.g., to Kendall (1974), Freeman (1976) and Mardia & Holmes (1980). In order to test the validity of the megalithic yard hypothesis it is necessary to obtain accurate estimates of circle diameters and their variances. For this purposes circular structural models can be useful .

We consider the data set on stones' coordinates in the Avebury Main Ring documented in Table 1 in Thom et al. (1976). The cited paper posits that the stones are arranged in four circular arcs A, B, C and W "meeting at an angle instead of running smoothly into one another." The arcs contain 10, 7, 7, and 16

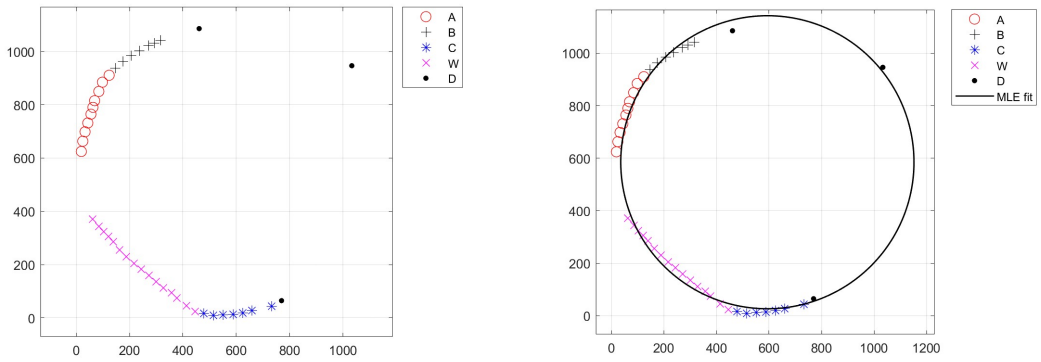


Figure 12: *The Avebury main ring: (a) stone positions and division in four circular arcs A, B, C, and W; (b) the circle fit obtained by the algorithm of Section 6.1.*

stones respectively; there are also three additional stones (denoted D here) that are considered belonging to other arcs. The stone positions and their division in arcs are displayed in panel (a) of Figure 12.

Under some prior assumptions on the positions of the circles' centres and radii, Thom et al. (1976) constructed estimates of these parameters for the arcs A, B, C and W, and suggested that the conducted measurements are in a full agreement with the megalithic yard hypothesis. Freeman (1977) reconsidered the Avebury data; he fitted circles to each arc separately using a least squares method. It is not clear from the exposition whether the algebraic or geometric fit was used. Since Freeman (1977) writes that the residual sum of squares is "the sum of the squares of the distances of each stone from its nearest point on the fitted arc", we suppose that his method is equivalent to the geometric fit. Freeman (1977) asserts that the results obtained from the least squares fit are significantly different from those reported in Thom et al. (1976); hence, it is claimed that there is no evidence for the megalithic yard hypothesis. The results of Thom et al. (1976) and Freeman (1977) are summarized in the first two columns of Table 1; the numbers in these columns are reproduced from Table 1 in Freeman (1977).

We used the algorithm developed in this paper in order to fit circles to the data points in arcs A, B, C and W. The estimated circle centers and radii are given in the third column of Table 1. It is clearly

		Thom et al. (1976)			Freeman (1977)			This paper		
Arc	n	$\hat{a}$	$\hat{b}$	$\hat{\rho}$	$\hat{a}$	$\hat{b}$	$\hat{\rho}$	$\hat{a}$	$\hat{b}$	$\hat{\rho}$
A	10	723.2	538.6	707.7	795.0	516.5	782.8	795.26	516.38	783.13
B	7	586.6	386.9	707.7	512.7	533.1	545.5	513.01	532.56	545.95
C	7	520.9	720.8	707.7	530.8	651.0	638.8	530.8	651.23	639.05
W	16	1612.8	1697.1	2041.5	1472.0	1553.4	1840.4	2091.1	2239.7	2762.3

Table 1: *The results of circular fits obtained by Thom et al. (1976), Freeman (1977), and by the method of this paper. The fitted circle centers and radii are given in the (a, b) and ρ–columns.*

seen that for arcs A, B and C the obtained estimates are very close to those derived in Freeman (1977), but the results for arc W are significantly different. Here the estimated radius is much bigger than the one obtained by the least squares method by Freeman (1977). The reason is that the data points in arc W are effectively lying on a straight line segment, and if this segment is modeled as a circular arc then the corresponding radius should be very big. This intuition is confirmed by the obtained estimate of the concentration parameter $\varkappa$ of the von Mises distribution: it is equal to 322.61 in this case. It is worth noting that in all these examples the sample sizes are so small that the reliability of obtained estimates seems to be rather low. In particular, in a separate numerical experiment Freeman (1977) showed that small perturbations in the data lead to completely different least squares estimates of the centers and radii. Finally, if the proposed algorithm is applied to all available 43 observations then the obtained circle estimate is $\hat{a} = 593.94$, $\hat{b} = 584.71$, and $\hat{\rho} = 558.01$. The corresponding fit is displayed in panel (b) of Figure 12.

7 Concluding remarks

We close this paper with several remarks about limitations and possible extensions of the proposed methodology.

The main theoretical assumption underlying statistical inference in the proposed model is that the unobservable angle variables follow the von Mises distribution. This assumption is natural in the context of fitting circles to noisy data. First, the von Mises distribution is one of the most widely used modeling assumptions for circular random variables. It is also rather flexible: as a particular case it includes the uniform distribution, the case that was extensively studied in the literature on circle fitting. Secondly, the von Mises distribution with non-zero concentration parameter is adequate when the data points are sampled from an arc of the circle. It is worth mentioning that, as any other parametric model, the circular structural model based on the von Mises distribution has limitations. The performance of the proposed estimator can be affected by deviations from the distributional assumptions. However, our simulations in Section 6.2 demonstrate that, when the data are sampled from a single arc, the developed algorithm is stable, even though the distributional assumptions do not hold.

The proposed methodology can be extended in different directions. In particular, the problem of fitting a sphere to a data cloud in $\mathbb{R}^d$ can be approached similarly. Such a problem has been discussed, e.g., in Witzgall et al. (2006). A reasonable assumption on the unobserved variables in this setting would be that they follow the Fisher–von Mises distribution on the sphere with unknown concentration $\varkappa \geq 0$ and mean direction $\mu \in \mathbb{S}^{d-1}$. The resulting parametric model can be analyzed using the ideas developed in this paper. In particular, the maximum likelihood estimator is defined similarly, and a preliminary

estimator of the mean direction parameter can be based on the principal component analysis. However, the full theoretical investigation of this extension is beyond the scope of the current paper.

The proposed circle fitting algorithm can be also used locally in order to recover smooth curves on the plane from noisy data. For this purpose, a circular arc can be fitted to a subset of data points in the way similar to the one in kernel smoothing methods. In general, we believe that structural distributional assumptions on unobservable variables can be very useful in many applications; they facilitate development of reasonable estimators with provable theoretical guarantees.

A Bessel functions and von Mises distribution

In this section for the ease of reference we collect well known facts about the modified Bessel functions of the first kind and the von Mises distribution. For comprehensive discussion we refer to classical books Watson (1944) and Mardia & Jupp (2000).

Modified Bessel function of first kind. The modified Bessel function $I_\nu(z)$ of the first kind of order ν is given by the infinite series

$$I_\nu(z) = (\tfrac{1}{2}z)^\nu \sum_{j=0}^{\infty} \frac{(\frac{1}{4}z^2)^j}{j!\Gamma(\nu+j+1)},$$

where $\Gamma(\cdot)$ is the Gamma function. For integer $\nu = k = 0, \pm 1, \pm 2, \ldots$ the following integral representation holds: $I_k(z) = \frac{1}{2\pi}\int_0^{2\pi} \cos(k\theta) e^{z\cos\theta} \mathrm{d}\theta$, $z \in \mathbb{C}$.

The following relationships involving the Bessel functions $I_k(z)$ are important for our purposes.

(i). For any $a, b \in \mathbb{R}$ one has

$$\int_0^{2\pi} e^{a\cos\theta + b\sin\theta} \mathrm{d}\theta = I_0\left(\sqrt{a^2+b^2}\right). \tag{A.1}$$

(ii). The modified Bessel function of first kind has the following property:

$$\frac{\mathrm{d}}{\mathrm{d}z} I_k(z) = I_{k+1}(z) + \frac{k}{z} I_k(z), \tag{A.2}$$

and specifically $\frac{\mathrm{d}}{\mathrm{d}z} I_0(z) = I_1(z)$.

(iii). The following recursive relationship holds:

$$I_k(z) = \tfrac{2(k+1)}{z} I_{k+1}(z) + I_{k+2}(z). \tag{A.3}$$

In addition, for $k \geq -1$ and real positive z on has

$$I_{k-1}(z) I_{k+1}(z) < I_k^2(z); \tag{A.4}$$

see Segura (2011). Dividing (A.4) by $I_{k-1}^2(z)$ we obtain that $I_{k+1}(z)/I_{k-1}(z) < I_k^2(z)/I_{k-1}^2(z)$.

(iv). For $\nu > -0.5$ and real $z > 0$ one has $I_\nu(z) > I_{\nu+1}(z)$ [see Soni (1965)]; therefore

$$0 \leq R_{k+1}(z) \leq R_k(z) \leq 1, \quad \forall z > 0, \ \ \forall k. \tag{A.5}$$

Moments of the von Mises distribution. We repeatedly use formulas for moments of the von Mises distribution. If $\varphi \sim \mathrm{vM}(\mu, \varkappa)$, then

$$\mathbb{E}\left(\cos(k\varphi)\right) = R_k(\varkappa)\cos(k\mu), \quad \forall k = 0, \pm 1, \pm 2, \dots. \tag{A.6}$$

$$\mathbb{E}\left(\sin(k\varphi)\right) = R_k(\varkappa)\sin(k\mu), \quad \forall k = 0, \pm 1, \pm 2, \dots. \tag{A.7}$$

Using trigonometric identities and (A.6)-(A.7) we also have

$$\mathbb{E}(\sin^2\varphi) = \tfrac{1}{2}\mathbb{E}(1-\cos 2\varphi) = \tfrac{1}{2}\big(1 - R_2(\varkappa)\cos(2\mu)\big) \tag{A.8}$$

$$\mathbb{E}(\cos^2\varphi) = \tfrac{1}{2}\mathbb{E}(1+\cos 2\varphi) = \tfrac{1}{2}\big(1 + R_2(\varkappa)\cos(2\mu)\big). \tag{A.9}$$

B Proofs

B.1 Proof of Proposition 2.1

First, we compute the moment generating function $\Psi_\theta(t_1, t_2) := \mathbb{E}_\theta(e^{t_1 X + t_2 Y})$ of (X, Y). By independence of $\{\varphi_i\}$, $\{\xi_i\}$ and $\{\eta_i\}$ we have

$$\begin{aligned}\Psi_\theta(t_1,t_2) &= e^{t_1 a + t_2 b} \cdot \mathbb{E}_\theta(\exp\left\{\rho(t_1\cos\varphi + t_2\sin\varphi)\right\}) \cdot \mathbb{E}_\theta(\exp\left\{t_1\xi + t_2\eta\right\}) \\ &= \exp\Big\{t_1 a + t_2 b + \tfrac{1}{2}\sigma^2(t_1^2+t_2^2)\Big\} \cdot \mathbb{E}_\theta(\exp\left\{\rho(t_1\cos\varphi + t_2\sin\varphi)\right\}).\end{aligned}$$

Furthermore,

$$\begin{aligned}\mathbb{E}_\theta(\exp\left\{\rho(t_1\cos\varphi + t_2\sin\varphi)\right\}) &= \frac{1}{2\pi I_0(\varkappa)}\int_0^{2\pi} \exp\{\rho(t_1\cos\varphi + t_2\sin\varphi)\} e^{\varkappa\cos(\varphi-\mu)}\mathrm{d}\varphi \\ &= \frac{1}{2\pi I_0(\varkappa)}\int_0^{2\pi} \exp\Big\{(\rho t_1 + \varkappa\cos\mu)\cos\varphi + (\rho t_2 + \varkappa\sin\mu)\sin\varphi\Big\}\mathrm{d}\varphi \\ &= \frac{1}{I_0(\varkappa)} I_0\big(\sqrt{(\rho t_1 + \varkappa\cos\mu)^2 + (\rho t_2 + \varkappa\sin\mu)^2}\big).\end{aligned}$$

Therefore

$$\Psi_\theta(t_1,t_2) = \exp\Big\{a t_1 + b t_2 + \tfrac{1}{2}\sigma^2(t_1^2+t_2^2)\Big\}\frac{1}{I_0(\varkappa)} I_0\Big(\sqrt{(\rho t_1 + \varkappa\cos\mu)^2 + (\rho t_2 + \varkappa\sin\mu)^2}\Big).$$

Let $\theta_1 = (a_1, b_1, \rho_1, \sigma_1^2, \mu_1, \varkappa_1)$, and $\theta_2 = (a_2, b_2, \rho_2, \sigma_2^2, \mu_2, \varkappa_2)$, $\theta_1, \theta_2 \in \Theta$. We will show that the equality

$$\Psi_{\theta_1}(t_1,t_2) = \Psi_{\theta_2}(t_1,t_2), \quad \forall t_1, t_2 \tag{B.1}$$

implies that $\theta_1 = \theta_2$. This fact will imply the identifiability of the model.

The moment generating function $\Psi_\theta(t_1, t_2)$ is factored to a product of two functions:

$$\begin{aligned}\Psi_\theta(t_1,t_2) &= \Psi^{(1)}_{a,b,\sigma^2}(t_1,t_2)\Psi^{(2)}_{\rho,\varkappa,\mu}(t_1,t_2) \\ &= \exp\Big\{t_1 a + t_2 b + \tfrac{1}{2}\sigma^2(t_1^2+t_2^2)\Big\} \times \frac{I_0\big(\sqrt{(\rho t_1 + \varkappa\cos\mu)^2 + (\rho t_2 + \varkappa\sin\mu)^2}\big)}{I_0(\varkappa)},\end{aligned} \tag{B.2}$$

where the first one is a normal factor depending on a, b and σ^2 only, while the second one depends on ρ, $\varkappa$ and μ only. In view of factorization (B.2), if (B.1) holds then necessarily

$$\Psi^{(1)}_{a_1,b_1,\sigma_1^2}(t_1,t_2) = \Psi^{(1)}_{a_2,b_2,\sigma_2^2}(t_1,t_2), \quad \forall t_1,t_2, \tag{B.3}$$

$$\Psi^{(2)}_{\rho_1,\varkappa_1,\mu_1}(t_1,t_2) = \Psi^{(2)}_{\rho_2,\varkappa_2,\mu_2}(t_1,t_2), \quad \forall t_1,t_2. \tag{B.4}$$

It follows from the equality of the normal factors (B.3) that $a_1t_1 + b_1t_2 + \frac{1}{2}\sigma_1^2(t_1^2+t_2^2) = a_2t_1 + b_2t_2 + \frac{1}{2}\sigma_2^2(t_1^2+t_2^2)$, $\forall t_1,t_2$, and the two polynomials are equal if and only if $a_1 = a_2$, $b_1 = b_2$ and $\sigma_1^2 = \sigma_2^2$.

Next consider equality (B.4). We have

$$\Psi^{(2)}_{\rho,\varkappa,\mu}(t_1,0) = \frac{1}{I_0(\varkappa)} I_0\Big(\sqrt{\varkappa^2 + \rho^2 t_1^2 + 2\varkappa\rho t_1\cos\mu}\Big), \quad \forall t_1$$
$$\Psi^{(2)}_{\rho,\varkappa,\mu}(0,t_2) = \frac{1}{I_0(\varkappa)} I_0\Big(\sqrt{\varkappa^2 + \rho^2 t_2^2 + 2\varkappa\rho t_2\sin\mu}\Big), \quad \forall t_2.$$

Therefore $\frac{\partial}{\partial t_1}\Psi^{(2)}_{\rho,\varkappa,\mu}(t_1,0)\Big|_{t_1=0} = R_1(\varkappa)\rho\cos\mu$, and $\frac{\partial}{\partial t_2}\Psi^{(2)}_{\rho,\varkappa,\mu}(0,t_2)\Big|_{t_2=0} = R_1(\varkappa)\rho\sin\mu$. Equating the above partial derivatives for parameters $(\rho_1,\varkappa_1,\mu_1)$ and $(\rho_2,\varkappa_2,\mu_2)$ we obtain

$$\mu_1 = \mu_2, \quad R_1(\varkappa_1)\rho_1 = R_1(\varkappa_2)\rho_2. \tag{B.5}$$

Furthermore, for given $(\rho_1,\varkappa_1,\mu)$ and $(\rho_2,\varkappa_2,\mu)$ assume for definiteness that $\cos\mu \neq 0$, and let $t_1^* = -2\varkappa_1\cos\mu/\rho_1$; then $\Psi^{(2)}_{\rho_1,\varkappa_1,\mu}(t_1^*,0) = 1$, and

$$\Psi^{(2)}_{\rho_2,\varkappa_2,\mu}(t_1^*,0) = \frac{1}{I_0(\varkappa_2)} I_0\Big(\sqrt{\varkappa_2^2 + 4\varkappa_1(\rho_2/\rho_1)^2\cos^2\mu - 4(\rho_2/\rho_1)\varkappa_1\varkappa_2\cos^2\mu}\Big).$$

Hence equation $\Psi^{(2)}_{\rho_1,\varkappa_1,\mu}(t_1^*,0) = \Psi^{(2)}_{\rho_2,\varkappa_2,\mu}(t_1^*,0)$ along with the monotonicity of $I_0(\cdot)$ imply that

$$\varkappa_2 = \sqrt{\varkappa_2^2 + 4\varkappa_1(\rho_2/\rho_1)^2\cos^2\mu - 4(\rho_2/\rho_1)\varkappa_1\varkappa_2\cos^2\mu},$$

which is equivalent to $(4\varkappa_1\rho_2/\rho_1)(\rho_2\varkappa_1/\rho_1 - \varkappa_2)\cos^2\mu = 0$, and hence $\varkappa_1\rho_2 = \varkappa_2\rho_1$. Combining this with (B.5) we obtain $\varkappa_1 R_1(\varkappa_1) = \varkappa_2 R_2(\varkappa_2)$. Below we show that function $x \mapsto xR_1(x)$ is monotone increasing on the positive real line. Therefore equality $\varkappa_1 R_1(\varkappa_1) = \varkappa_2 R_2(\varkappa_2)$ implies that $\varkappa_1 = \varkappa_2$ and hence $\rho_1 = \rho_2$. This proves the identifiability under assumption that $\cos\mu \neq 0$. If $\cos\mu = 0$ then we let $t_2^* = -2\varkappa_2\sin\mu/\rho_2$ and consider the equation $\Psi^{(2)}_{\rho_1,\varkappa_1,\mu}(0,t_2^*) = \Psi^{(2)}_{\rho_2,\varkappa_2,\mu}(0,t_2^*)$. Then the same reasoning as above yields the result.

It remains to show that function $x \mapsto xR_1(x)$ is monotone increasing. We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}[xR_1(x)] &= \frac{I_1(x)}{I_0(x)} + x\frac{I_1'(x)}{I_0(x)} - x\left(\frac{I_1(x)}{I_0(x)}\right)^2 = \frac{I_1(x)}{I_0(x)} + x\frac{I_2(x)+\frac{1}{x}I_1(x)}{I_0(x)} - x\left(\frac{I_1(x)}{I_0(x)}\right)^2 \\ &= \frac{2I_1(x)}{I_0(x)} + \frac{x}{I_0(x)}\left(I_0(x) - \frac{2}{x}I_1(x)\right) - x\left(\frac{I_1(x)}{I_0(x)}\right)^2 = x(1-R_1^2(x)),\end{aligned}$$

where the second equality follows from $I_1'(x) = I_2(x) + \frac{1}{x}I_1(x)$ and the third one is a consequence of $I_2(x) + \frac{2}{x}I_1(x) = I_0(x)$; see (A.2) and (A.3). Since $0 \le R_1(x) \le 1$, the function $x \mapsto xR_1(x)$ is monotone increasing for positive x, as claimed. ∎

B.2 Proof of Theorem 3.1

The proof of Theorem 3.1 is based on verification of conditions of Theorem 8.1 in Chapter 1 of Ibragimov & Hasminskii (1981) for the circular structural model of Section 2. For ease of notation in the subsequent proof we write $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6) := (a, b, \rho, \sigma^2, \mu, \varkappa)$.

We need to verify the following conditions: there exists a number $0 < \delta \le 1$ such that for any compact set $K \subset \Theta$ one has

$$\sup_{\theta\in K} \mathbb{E}_\theta \Big| \frac{\partial}{\partial\theta_i}\ell_1(\theta; X_1, Y_1)\Big|^{2+\delta} < \infty, \quad \forall i = 1, \dots, 6, \tag{B.6}$$

$$\sup_{\theta\in K} \mathbb{E}_\theta \Big| \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_1(\theta; X_1, Y_1)\Big|^{1+\delta} < \infty, \quad \forall i, j = 1, \dots, 6, \tag{B.7}$$

$$\mathbb{E}_\theta \Big[\sup_{\theta,\theta'\in K} \Big| \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_1(\theta; X_1, Y_1) - \frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_1(\theta'; X_1, Y_1)\Big| \cdot \|\theta - \theta'\|^{-\delta}\Big] < \infty, \quad \forall i, j = 1, \dots, 6. \tag{B.8}$$

In fact, we will show that (B.6)–(B.8) are fulfilled with $\delta = 1$.

Verification of (B.6). Since $R_1(x)/x \le 1$ for all x we have from (3.5) that for any positive number $m \ge 1$

$$\mathbb{E}_\theta \Big| \frac{\partial}{\partial a}\ell_1(\theta; X_1, Y_1)\Big|^m \le \frac{2^{m-1}}{\sigma^{2m}}\big(\mathbb{E}_\theta|X_1 - a|^m + \rho^m\varkappa^m\big) < \infty.$$

Thus (B.6) is fulfilled for the random variable $\frac{\partial}{\partial a}\ell_1(\theta; X_1, Y_1)$: all moments are bounded when θ varies in a compact set. The same is true for the random variable $\frac{\partial}{\partial b}\ell_1(\theta; X_1, Y_1)$. We also have

$$\begin{aligned}\mathbb{E}_\theta \Big| \frac{\partial}{\partial\rho}\ell_1(\theta; X_1, Y_1)\Big|^m \le \frac{2^{m-1}}{\sigma^{2m}}\Big\{ \frac{2^{m-1}\rho^m}{\sigma^{2m}}\Big(&\mathbb{E}_\theta[(X_1 - a)^2 + (Y_1 - b)^2]^m \\ &+ \varkappa^m\mathbb{E}_\theta[(X_1 - a)^2 + (Y_1 - b)^2]^{m/2}\Big) + \frac{\rho^m}{\sigma^{2m}}\Big\} < \infty\end{aligned}$$

for any $\theta \in K$. The similar reasoning applies in order to establish boundedness of all moments of the remaining partial derivatives of $\ell_1(\theta; X_1, Y_1)$ in (3.6)–(3.10).

Verification of (B.7). First, we note that in view of (A.2) and (A.3) we have

$$R_k'(x) = R_{k+1}(x) + \frac{k}{x}R_k(x) - R_k(x)R_1(x),$$

and therefore

$$\frac{\mathrm{d}}{\mathrm{d}x}(R_k(x)/x) = \frac{1}{x}\Big[R_{k+1}(x) - \tfrac{k-1}{x}R_k(x) - R_k(x)R_1(x)\Big].$$

In particular, $\frac{\mathrm{d}}{\mathrm{d}x}\big(R_1(x)/x\big) = [R_2(x) - R_1^2(x)]/x$. Observing that $R_k(x) = \prod_{j=1}^{k}[I_j(x)/I_{j-1}(x)]$ and using bounds on $I_j(x)/I_{j-1}(x)$ derived in Amos (1974) we have

$$\prod_{j=1}^{k}\left(\frac{x}{j+\sqrt{x^2+j^2}}\right) \le R_k(x) \le \prod_{j=1}^{k}\left(\frac{x}{j-1+\sqrt{x^2+(j+1)^2}}\right).$$

This inequality implies that

$$R_k(x)/x^m \le 1, \quad \forall 0 \le m \le k. \tag{B.9}$$

By (3.5)–(3.10), the first order partial derivatives of $\ell(\theta; x, y)$ with respect to θ_i, $i = 1, \ldots, 6$ have the following representation

$$\frac{\partial \ell_1(\theta; x, y)}{\partial \theta_i} = \frac{R_1(D_\theta)}{D_\theta} u_i(\theta; x, y) + v_i(\theta; x, y), \quad i = 1, \ldots, 6. \tag{B.10}$$

For instance, for the partial derivative with respect to $\theta_1 = a$ we have

$$u_1(\theta; x, y) = -\frac{1}{\sigma^2}(x-a) - \frac{\rho}{\sigma^2}\varkappa\cos\mu, \quad v_1(\theta; x, y) = \frac{1}{\sigma^2}(x-a),$$

while for $\theta_3 = \rho$ we have $v_3(\theta; x, y) = -\rho/\sigma^2$ and

$$u_3(\theta; x, y) = \frac{\rho}{\sigma^4}\big[(x-a)^2 + (y-b)^2\big] + \frac{\varkappa}{\sigma^2}\big[(x-a)\cos\mu + (y-b)\sin\mu\big].$$

We note that all functions $u_i(\theta; x, y)$ and $v_i(\theta; x, y)$ are infinitely differentiable with respect to θ, so that

$$\frac{\partial^2 \ell_1(\theta; x, y)}{\partial \theta_i \partial \theta_j} = \frac{\partial}{\partial \theta_j}\Big(\frac{R_1(D_\theta)}{D_\theta}\Big) u_i(\theta; x, y) + \frac{R_1(D_\theta)}{D_\theta}\frac{\partial u_i(\theta; x, y)}{\partial \theta_j} + \frac{\partial v_i(\theta; x, y)}{\partial \theta_j}. \tag{B.11}$$

Because random variables $X - a$ and $Y - b$ have finite moments of all orders we have that for any compact set $K \subset \Theta$ and $m \ge 1$

$$\sup_{\theta\in K}\mathbb{E}_\theta\left|\frac{\partial u_i(\theta; X_1, Y_1)}{\partial \theta_j}\right|^m < \infty, \quad \sup_{\theta\in K}\mathbb{E}_\theta\left|\frac{\partial v_i(\theta; X_1, Y_1)}{\partial \theta_j}\right|^m < \infty, \quad \forall i, j = 1, \ldots, 6. \tag{B.12}$$

Furthermore, all second order partial derivatives of $\ell(\theta; x, y)$ with respect to θ_i, $i = 1, \ldots, 6$ contain partial derivatives of the form

$$\begin{aligned}\frac{\partial}{\partial \theta_i}\Big(\frac{R_1(D_\theta)}{D_\theta}\Big) &= \frac{\partial}{\partial x}\left(\frac{R_1(x)}{x}\right)_{x=D_\theta} \cdot \Big(\frac{\partial D_\theta}{\partial \theta_i}\Big) \\ &= \frac{1}{D_\theta^2}\Big[R_2(D_\theta) - R_1^2(D_\theta)\Big]\left(A_\theta(x)\frac{\partial A_\theta(x)}{\partial \theta_i} + B_\theta(y)\frac{\partial B_\theta(y)}{\partial \theta_i}\right).\end{aligned} \tag{B.13}$$

In view of (B.9), $|R_2(D_\theta) - R_1^2(D_\theta)/D_\theta^2| \le 1$ and $0 \le R_1(D_\theta)/D_\theta \le 1$ uniformly in $\theta \in K$. In addition, all moments of $X_1 - a$ and $Y_1 - b$ are bounded uniformly in $\theta \in K$, and

$$\sup_{\theta\in K}\mathbb{E}_\theta\left|A_\theta(X_1)\frac{\partial A_\theta(X_1)}{\partial \theta_i} + B_\theta(Y_1)\frac{\partial B_\theta(Y_1)}{\partial \theta_i}\right|^m < \infty, \quad \forall m > 1. \tag{B.14}$$

Combining (B.14), (B.13), (B.12) and (B.11) we obtain that

$$\sup_{\theta\in K}\mathbb{E}_\theta\left|\frac{\partial^2 \ell_1(\theta; X_1, Y_1)}{\partial \theta_i \theta_j}\right|^m < \infty, \quad \forall i, j = 1, \ldots, 6,$$

i.e., condition (B.7) holds.

Verification of (B.8). It is sufficient to demonstrate that

$$\mathbb{E}_\theta \sup_{\theta\in K}\left|\frac{\partial^3 \ell_1(\theta;X_1,Y_1)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| < \infty, \quad \forall i,j,k=1,\dots,6. \tag{B.15}$$

First we note that

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Big(\frac{R_1(x)}{x}\Big) = \frac{\mathrm{d}}{\mathrm{d}x}\Big(\frac{R_2(x)-R_1^2(x)}{x}\Big) = \frac{1}{x^2}[R_3(x)-3R_2(x)R_1(x)+2R_1^3(x)].$$

Then it follows from (B.9) that $|\frac{\mathrm{d}^2}{\mathrm{d}x^2}[R_1(x)/x]| \le 3$ for all $x \ge 0$. Differentiating (B.11), taking into account that

$$\sup_{\theta\in K}\mathbb{E}_\theta\left|\frac{\partial^2 u_i(\theta;X_1,Y_1)}{\partial\theta_j\partial\theta_k}\right|^m < \infty, \quad \sup_{\theta\in K}\mathbb{E}_\theta\left|\frac{\partial^2 v_i(\theta;X_1,Y_1)}{\partial\theta_j\partial\theta_k}\right|^m < \infty, \quad \forall i,j,k=1,\dots,6,$$

and applying the same reasoning as above we conclude that that (B.15) is fulfilled. ∎

B.3 Proof of Proposition 4.1

(a). Using representation (4.1), we have that

$$\mathrm{cov}_\theta(Z) = \rho^2\mathrm{cov}_\theta(e_\varphi) + \mathrm{cov}_\theta(\varepsilon). \tag{B.16}$$

In addition, it follows from (A.6)-(A.7) that $\mathbb{E}_\theta(e_\varphi) = R_1(\varkappa)e_\mu$, and from (A.8) - (A.9) that

$$\mathbb{E}_\theta(e_\varphi e_\varphi^T) = \begin{bmatrix}\mathbb{E}_\theta(\cos^2\varphi) & \mathbb{E}_\theta(\cos\varphi\sin\varphi)\\ \mathbb{E}_\theta(\cos\varphi\sin\varphi) & \mathbb{E}_\theta(\sin^2\varphi)\end{bmatrix} = \begin{bmatrix}\frac12+\frac12 R_2(\varkappa)\cos(2\mu) & \frac12 R_2(\varkappa)\sin(2\mu)\\ \frac12 R_2(\varkappa)\sin(2\mu) & \frac12-\frac12 R_2(\varkappa)\cos(2\mu)\end{bmatrix}$$

$$= \frac12\begin{bmatrix}1+R_2(\varkappa)\cos(2\mu) & R_2(\varkappa)\sin(2\mu)\\ R_2(\varkappa)\sin(2\mu) & 1-R_2(\varkappa)\cos(2\mu)\end{bmatrix}.$$

Then, after straightforward algebra,

$$\mathrm{cov}_\theta(e_\varphi) = \mathbb{E}_\theta(e_\varphi e_\varphi^T) - (\mathbb{E}_\theta e_\varphi)(\mathbb{E}_\theta e_\varphi)^T$$

$$= \frac12\begin{bmatrix}1+R_2(\varkappa)\cos 2\mu & R_2(\varkappa)\sin 2\mu\\ R_2(\varkappa)\sin 2\mu & 1-R_2(\varkappa)\cos 2\mu\end{bmatrix} - R_1^2(\varkappa)\begin{bmatrix}\cos^2\mu & \cos\mu\sin\mu\\ \cos\mu\sin\mu & \sin^2\mu\end{bmatrix}$$

$$= \frac12\begin{bmatrix}1-R_1^2(\varkappa)+(R_2(\varkappa)-R_1^2(\varkappa))\cos 2\mu & (R_2(\varkappa)-R_1^2(\varkappa))\sin 2\mu\\ (R_2(\varkappa)-R_1^2(\varkappa))\sin 2\mu & 1-R_1^2(\varkappa)-(R_2(\varkappa)-R_1^2(\varkappa))\cos 2\mu\end{bmatrix}.$$

Recalling that $\mathrm{cov}(\varepsilon) = \sigma^2 I$ and combining this with (B.16) we complete the proof of part (a). The proof of part (b) is immediate from (4.2).

(c). Let $\zeta \in [0, 2\pi)$ be a fixed number, and consider random variable $e_\zeta^T Z$. The variance of $e_\zeta^T Z$ is

$$\mathrm{var}_\theta(e_\zeta^T Z) = \rho^2\mathrm{var}_\theta(e_\zeta^T e_\varphi) + \mathrm{var}_\theta(e_\zeta^T\varepsilon) = \rho^2\mathrm{var}_\theta(e_\zeta^T e_\varphi) + \sigma^2.$$

Because $\mathbb{E}_\theta\big(e_\varphi^T e_\zeta\big) = \mathbb{E}_\theta\left(\cos(\varphi-\zeta)\right) = R_1(\varkappa)\cos(\mu-\zeta)$, and

$$\mathbb{E}_\theta\left((e_\varphi^T e_\zeta)^2\right) = \mathbb{E}_\theta\left(\cos^2(\varphi-\zeta)\right) = \tfrac12+\tfrac12\mathbb{E}_\theta\left(\cos(2(\varphi-\zeta))\right) = \tfrac12+\tfrac12 R_2(\varkappa)\cos(2(\mu-\zeta)).$$

we obtain that

$$\begin{aligned}\mathrm{var}_\theta(e_\zeta^T e_\varphi) &= \tfrac{1}{2} + \tfrac{1}{2}R_2(\varkappa)\cos(2(\mu-\zeta)) - R_1^2(\varkappa)\cos^2(\mu-\zeta)\\ &= \tfrac{1}{2}\big(1 - R_1^2(\varkappa)\big) + \tfrac{1}{2}\big(R_2(\varkappa) - R_1^2(\varkappa)\big)\cos(2(\mu-\zeta)),\end{aligned}$$

and therefore

$$\mathrm{var}_\theta(e_\zeta^T Z) = e_\zeta^T \Sigma e_\zeta = \tfrac{1}{2}\rho^2\Big[1 - R_1^2(\varkappa) + \big(R_2(\varkappa) - R_1^2(\varkappa)\big)\cos(2(\mu-\zeta))\Big] + \sigma^2. \tag{B.17}$$

Because $R_2(x) - R_1^2(x) \le 0$ for all $x \ge 0$ by (A.4), the minimum of the right hand side in (B.17) over ζ is achieved at $\zeta = \mu$, i.e.

$$\min_{\zeta\in[0,2\pi)} \mathrm{var}_\theta(e_\zeta^T Z) = \min_{e_\zeta:\|e_\zeta\|=1} e_\zeta^T\Sigma e_\zeta = e_\mu^T\Sigma e_\mu = \tfrac{1}{2}\rho^2\big[1 - 2R_1^2(\varkappa) + R_2(\varkappa)\big] + \sigma^2,$$

Thus e_μ is the eigenvector of the covariance matrix Σ corresponding to the minimal eigenvalue $\lambda_{\min}(\Sigma)$. Similarly, the maximum of the right hand side in (B.17) is achieved at $\zeta = \mu - \pi/2$, so that $e_{\mu-\pi/2}$ is the eigenvector corresponding to the maximal eigenvalue $\lambda_{\max}(\Sigma)$. The proof is complete. ∎

B.4 Proof of Theorem 5.1

(a). The proof is divided in two parts. First, we observe that the value of $\widehat{T}_n$ depends on S_{xx}, S_{yy} and S_{xy}, which are location invariant. Hence we start with derivation of the asymptotic distribution of the random vector with components

$$\sum_{i=1}^n (X_i - a),\ \sum_{i=1}^n (Y_i - b),\ \sum_{i=1}^n (X_i - a)^2,\ \sum_{i=1}^n (Y_i - b)^2,\ \sum_{i=1}^n (X_i - a)(Y_i - b).$$

Then we use the delta–method for derivation of the asymptotic distribution of $\widehat{T}_n$.

Consider vector $W := \big[X - a, (X-a)^2, Y - b, (Y-b)^2, (X-a)(Y-b)\big]^T$, and let W_i, $i = 1, \dots, n$ be the independent copies of W. By the central limit theorem,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n W_i - \mathbb{E}_{\theta_0}(W)\right) = \sqrt{n}(\bar{W} - \mathbb{E}_{\theta_0}(W)) \xrightarrow{d} \mathcal{N}(0, \Sigma_W), \quad n \to \infty,$$

where $\mathbb{E}_{\theta_0}(W) = \big[0, \frac{1}{2}\rho^2 + \sigma^2, 0, \frac{1}{2}\rho^2 + \sigma^2, 0\big]^T$, and

$$\Sigma_W = \begin{bmatrix} \frac{1}{2}\rho^2 + \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{8}\rho^4 + 2\sigma^2\rho^2 + 2\sigma^4 & 0 & -\frac{1}{8}\rho^4 & 0 \\ 0 & 0 & \frac{1}{2}\rho^2 + \sigma^2 & 0 & 0 \\ 0 & -\frac{1}{8}\rho^4 & 0 & \frac{1}{8}\rho^4 + 2\sigma^2\rho^2 + 2\sigma^4 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{8}\rho^4 + \sigma^2\rho^2 + \sigma^4 \end{bmatrix}.$$

Computation of $\mathbb{E}_{\theta_0}(W)$ is evident, and the entries of Σ_W are computed as follows:

$$[\Sigma_W]_{11} = \mathrm{var}_{\theta_0}(X - a) = \mathbb{E}_{\theta_0}((X-a)^2) = \tfrac{1}{2}\rho^2 + \sigma^2,$$

$$[\Sigma_W]_{22} = \mathrm{var}_{\theta_0}((X-a)^2) = \mathbb{E}_{\theta_0}((X-a)^4) - \left[\mathbb{E}_{\theta_0}((X-a)^2)\right]^2 = \tfrac{1}{8}\rho^4 + 2\sigma^2\rho^2 + 2\sigma^4,$$

$$[\Sigma_W]_{33} = \mathrm{var}_{\theta_0}(Y-b) = \mathbb{E}_{\theta_0}((Y-b)^2) = \tfrac{1}{2}\rho^2 + \sigma^2,$$

$$[\Sigma_W]_{44} = \mathrm{var}_{\theta_0}((Y-b)^2) = \mathbb{E}_{\theta_0}((Y-b)^4) - \left[\mathbb{E}_{\theta_0}((Y-b)^2)\right]^2 = \tfrac{1}{8}\rho^4 + 2\sigma^2\rho^2 + 2\sigma^4,$$

$$[\Sigma_W]_{55} = \mathbb{E}_{\theta_0}\left[(X-a)^2(Y-b)^2\right] = \tfrac{1}{8}\rho^4 + \sigma^2\rho^2 + \sigma^4,$$

$$[\Sigma_W]_{12} = \mathbb{E}_{\theta_0}(X-a)^3 - \mathbb{E}_{\theta_0}(X-a)\cdot\mathbb{E}_{\theta_0}(X-a)^2 = \mathbb{E}_{\theta_0}(X-a)^3 = \rho^3\mathbb{E}_{\theta_0}(\cos^3\varphi) = 0,$$

$$[\Sigma_W]_{13} = \mathbb{E}_{\theta_0}[(X-a)(Y-b)] = \rho^2\mathbb{E}_{\theta_0}(\cos\varphi\sin\varphi) = 0,$$

$$[\Sigma_W]_{14} = \mathbb{E}_{\theta_0}[(X-a)(Y-b)^2] - \mathbb{E}_{\theta_0}(X-a)\cdot\mathbb{E}_{\theta_0}(Y-b)^2 = \mathbb{E}_{\theta_0}[(X-a)(Y-b)^2]$$
$$= \mathbb{E}_{\theta_0}\left[\rho\cos\varphi(\rho^2\cos^2\varphi + \eta^2)\right] = 0,$$

$$[\Sigma_W]_{15} = \mathbb{E}_{\theta_0}[(X-a)^2(Y-b)] - \mathbb{E}_{\theta_0}(X-a)\cdot\mathbb{E}_{\theta_0}[(X-a)(Y-b)] = 0 \quad \text{[similarly to } [\Sigma_W]_{14}\text{]}$$

$$[\Sigma_W]_{23} = \mathbb{E}_{\theta_0}[(X-a)^2(Y-b)] - \mathbb{E}_{\theta_0}(X-a)^2\cdot\mathbb{E}_{\theta_0}(Y-b) = 0,$$

$$[\Sigma_W]_{24} = \mathbb{E}_{\theta_0}[(X-a)^2(Y-b)^2] - \mathbb{E}_{\theta_0}(X-a)^2\cdot\mathbb{E}_{\theta_0}(Y-b)^2$$
$$= \rho^4\mathbb{E}_{\theta_0}(\cos^2\varphi\sin^2\varphi) + \rho^2\sigma^2\mathbb{E}_{\theta_0}(\cos^2\varphi) + \rho^2\sigma^2\mathbb{E}_{\theta_0}(\sin^2\varphi) + \sigma^4 - \left(\tfrac{1}{2}\rho^2 + \sigma^2\right)^2$$
$$= \tfrac{1}{4}\rho^4\mathbb{E}_{\theta_0}(\sin^2 2\varphi) + \rho^2\sigma^2 + \sigma^4 - \left(\tfrac{1}{2}\rho^2 + \sigma^2\right)^2 = \tfrac{1}{8}\rho^4 + \rho^2\sigma^2 + \sigma^4 - \tfrac{1}{4}\rho^4 - \rho^2\sigma^2 - \sigma^4 = -\tfrac{1}{8}\rho^4,$$

$$[\Sigma_W]_{25} = \mathbb{E}_{\theta_0}[(X-a)^3(Y-b)] - \mathbb{E}_{\theta_0}(X-a)^2\cdot\mathbb{E}_{\theta_0}[(X-a)(Y-b)] = 0,$$

$$[\Sigma_W]_{34} = \mathbb{E}_{\theta_0}(Y-b)^3 - \mathbb{E}_{\theta_0}(Y-b)^2\cdot\mathbb{E}_{\theta_0}(Y-b) = 0,$$

$$[\Sigma_W]_{35} = \mathbb{E}_{\theta_0}[(Y-b)^2(X-a)] - \mathbb{E}_{\theta_0}(Y-b)\cdot\mathbb{E}_{\theta_0}[(Y-b)(X-a)] = 0,$$

$$[\Sigma_W]_{45} = \mathbb{E}_{\theta_0}[(X-a)(Y-b)^3] - \mathbb{E}_{\theta_0}(Y-b)^3\cdot\mathbb{E}_{\theta_0}[(Y-b)(X-a)] = 0.$$

Now we apply the delta–method in order to derive the asymptotic distribution of $\widehat{T}_n$. Define the function $g:\mathbb{R}^5\to\mathbb{R}^2$ by $g(x_1,x_2,x_3,x_4,x_5) = \left[-x_1^2 + x_2 + x_3^2 - x_4,\ 2x_5 - 2x_1x_3\right]^T$. With this notation $g(\bar{W}) = [S_{xx} - S_{yy},\ 2S_{xy}]^T$. Under the null hypothesis $g\left(\mathbb{E}_{\theta_0}(W)\right) = [0,0]^T$, and the matrix of partial derivatives of g is

$$D(x) := \begin{bmatrix} -2x_1 & 1 & 2x_3 & -1 & 0 \\ -2x_3 & 0 & -2x_1 & 0 & 2 \end{bmatrix} \ \Rightarrow\ D_0 := D\left(\mathbb{E}_{\theta_0}(W)\right) = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{bmatrix}.$$

Therefore by the delta–method

$$\sqrt{n}\begin{bmatrix} S_{xx} - S_{yy} \\ 2S_{xy} \end{bmatrix} \xrightarrow{d} \mathcal{N}(0, D_0\Sigma_W D_0^T), \quad n\to\infty,$$

and $D_0\Sigma_W D_0^T = \left(\tfrac{1}{2}\rho^4 + 4\sigma^2\rho^2 + 4\sigma^4\right)I = sI$. This completes the proof of (a).

(b). Under H_0 we have that $\mathbb{E}_{\theta_0}(X) = a$ and $\mathbb{E}_{\theta_0}(Y) = b$. In addition, observe that independently of the angle distribution the random variable $\sigma^{-2}[(X-a)^2 + (Y-b)^2]$ has non–central χ^2–distribution $\chi^2_2(\rho^2/\sigma^2)$ with two degrees of freedom and non–centrality ρ^2/σ^2. Thus,

$$\tfrac{1}{2}\mathbb{E}_{\theta_0}\left\{\left[\,(X-a)^2 + (Y-b)^2\,\right]^2\right\} = \tfrac{1}{2}\sigma^4\mathbb{E}_{\theta_0}\left\{\frac{1}{\sigma^2}\left[\,(X-a)^2 + (Y-b)^2\,\right]^2\right\} = \tfrac{1}{2}\rho^4 + 4\rho^2\sigma^2 + 4\sigma^4.$$

This formula shows that $\widehat{M}_n$ is an unbiased and consistent estimator of the parameter s.

By the results of part (a) we have that

$$n\big((S_{xx}-S_{yy})^2+4S_{xy}^2\big) \xrightarrow{d} \big(\tfrac{1}{2}\rho^4+4\rho^2\sigma^2+4\sigma^4\big)\cdot\chi_2^2, \quad n\to\infty.$$

The law of large numbers implies that $\widehat{M}_n \xrightarrow{p} \frac{1}{2}\rho^4+4\rho^2\sigma^2+4\sigma^4$ as $n\to\infty$; then applying the Slutzky theorem we complete the proof. ∎

References

Al-Sharadqah, A. and Chernov, N. (2009). Error analysis for circle fitting algorithms. *Electron. J. Stat.* **3**, 886–911.

Amos, D. E. (1974). Computation of modified Bessel functions and their ratios. *Math. Comp.* **28**, 239–251.

Anderson, D. A. (1981). The circular structural model. *J. Roy. Statist. Soc. B* **43**, 131–141.

Anderson, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. B* **38**, 1–36.

Beck, A. and Pan, D. (2012). On the solution of the GPS localization and circle fitting problems. *SIAM J. Optimization* **22**, 108–134.

Berman, M. (1989). Large sample bias in least squares estimators of a circular arc center and its radius. *Comput. Vision Graphics Image Process.* **45**, 126–128.

Berman, M. and Culpin, D. (1986). The statistical behaviour of some least squares estimators of the center and radius of a circle. *J. Roy. Statist. Soc. B* **48**, 183–96.

Chan, N. N. (1965). On circular functional relationships. *J. Roy. Statist. Soc. B* **27**, 45–56

Chernov, N. (2010). *Circular and Linear Regression: Fitting Circles and Lines by Least Squares.* CRC Press, Boca Raton.

Chernov, N. (2011). Fitting circles to scattered data: parameter estimates have no moments. *Metrika* **73**, 373–384.

Chernov, N. and Ososkov G. (1984). Effective algorithms for circle fitting. *Comput. Phys. Commun.* **33**, 329–333.

Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**, 247–254.

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

Freeman, P. R. (1976). Bayesian analysis of the megalithic yard. *J. Royal Stat. Soc. A* **139**, 20–55.

Freeman, P. R. (1977). Thom's survey of the Avebury ring. *J. Hist. Astronom.* **8**, 134–136.

Ibragimov, I. and Hasminskii, R. 1981. *Statistical Estimation. Asymptotic Theory.* Springer, New York.

Kanatani, K. and Rangarajan, P. (2011). Hyper least squares fitting of circles and ellipses. *Comput. Statist. Data Anal.* **55**, 2197–2208.

Kanatani, K., Sugaya, Y. and Kanazawa, Y. (2016). *Ellipse Fitting for Computer Vision. Implementation and Applications.* Synthesis Lectures on Computer Vision, 8. Morgan & Claypool Publishers, Williston, VT.

Kåsa, I. (1976). A circle fitting procedure and its error analysis. *IEEE Trans. Instr. Measurement* **1**, pp. 8–14.

Kendall, D.G. (1974). Hunting quanta. *Phil. Trans. Royal Soc. London. Series A, Math. Phys. Sci.* **276**, No. 1257, 231–266.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27**, 887–906.

Landau M. (1987). Estimation of a circular arc center and its radius. *Comput. Vision Graphics Image Process.* **38**, 317–326.

Lindsay, B. (1980). Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Phil. Trans. Royal Soc. London. Series A, Math. Phys. Sci.* **296**, No. 1427, 639–662.

Mardia, K.V. and Holmes, D. (1980). A statistical analysis of megalithic data under elliptic pattern. *J. Royal Stat. Soc. A* **143**, 293–302.

Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics.* John Wiley and Sons, Chichester.

Neyman, J. and Scott, E.L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.

Núñes, P., Vázquez-Martin, R., Bandera A., and Sandoval, F. (2008). An algorithm for fitting 2-D data on the circle: applications to mobile robotics. *IEEE Signal Process. Lett.* **15**, 127–130.

Pratt V. (1987). Direct least-squares fitting of algebraic surfaces. *Comput. Graphics* **21**, 145–152.

Segura, J. (2011). Bounds for ratios of modified Bessel functions and associated Turán-type inequalities. *J. Math. Anal. Applic.* **374**, 516-528.

Soni, R. P. (1965). On an inequality for modified Bessel hunctions. *J. Math. Phys.* **44**, 406-407.

Späth, H. (1996). Least-squares fitting by circles. *Computing* **57**, 179–185.

Taubin, G. (1991). Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1115-1138.

Thom, A. (1955). A statistical examination of the megalithic sites in Britain. *J. Roy. Statist. Soc. A* **118**, 275–295.

Thom, A., Thom A. S., and Foord, T. R. (1976). Avebury: a new assessment of the geometry and metrology of the ring. *J. Hist. Astron.* **7**, 183–192.

Watson G.N. (1944) *A Treatise on the Theory of Bessel Functions.* Second edition. Cambridge University Press.

Witzgall, C., Cheok, G. S. and Kearsley, A. J. (2006). Recovering circles and spheres from point data. *Perspectives in Operations Research.* Edited by F.B. Alt, M.C. Fu and B. Golden, 393–413. Springer.