Electronic Journal of Statistics
Vol. 8 (2014) 2741–2778
ISSN: 1935-7524
DOI: 10.1214/14-EJS971

Further statistical analysis of circle fitting

Ali Al-Sharadqah

Department of Mathematics, East Carolina University
e-mail: alsharadqaha@ecu.edu

Abstract: This study is devoted to comparing the most popular circle fits (the geometric fit, Pratt's, Taubin's, Kåsa's) and the most recently developed algebraic circle fits: hyperaccurate fit and HyperLS fit. Even though hyperaccurate fit has zero essential bias and HyperLS fit is unbiased up to order σ^4, the geometric fit still outperforms them in some circumstances. Since the first-order leading term of the MSE for all fits are equal, we go one step further and derive all terms of order σ^4, which come from essential bias, as well as all terms of order σ^4/n, which come from two sources: the variance and the outer product of the essential bias and the nonessential bias.

Our analysis shows that when data are distributed along a short circular arc, the covariance part is the dominant part of the second-order term in the MSE. Accordingly, the geometric fit outperforms all existing methods. However, for a long circular arc, the bias becomes the most dominant part of the second-order term, and as such, hyperaccurate fit and HyperLS fit outperform the geometric fit. We finally propose a 'bias correction' version of the geometric fit, which in turn, outperforms all existing methods. The new method has two features. Its variance is the smallest and has zero bias up to order σ^4. Our numerical tests confirm the superiority of the proposed fit over the existing fits.

MSC 2010 subject classifications: Primary 68T10, 68K45; secondary 68K40, 62P30.
Keywords and phrases: Circle fitting, computer vision, small-noise model, algebraic fit, geometric fit, mean squared error, bias, variance.

Received December 2013.

Contents

1. Introduction

Circle fitting is one of the most fundamental tasks in computer vision, image processing, and pattern recognition [15, 24, 26, 27, 31], as well as other sciences such as biology, nuclear physics, archaeology, and industry [13, 38, 30]. Suppose that n experimental observations (i.e., $\mathbf{m}_i = (x_i, y_i)^T$, $i = 1, \dots, n$) are recorded. Our goal is to estimate the coordinates of the circle center $\mathbf{c} = (a, b)^T$ and the radius R that describe the best fit for the experimental observations. Many algorithms were developed to fit circles to the data, but *geometric fit* is one of the most accurate ones. It is also called "*orthogonal distance regression (ODR)*", because it is based on the orthogonal least squares, and it minimizes the objective function

$$\mathcal{G}(a, b, R) = \sum_{i=1}^{n} d_i^2, \tag{1.1}$$

where d_i stands for the signed distance from $\mathbf{m}_i$ to the circle

$$d_i = r_i - R, \qquad r_i = \sqrt{(x_i - a)^2 + (y_i - b)^2}. \tag{1.2}$$

This is a nonlinear minimization problem, so it has no-closed form solution. All practical algorithms that minimize $\mathcal{G}$ are iterative; some implement general Gauss-Newton [8, 17] or Levenberg-Marquardt (LM) [12] schemes. Others use circle-specific methods proposed by Landau [26] and Späth [29]. All are computationally intensive and subject to occasional divergence, and their performances heavily depend on the choices of the initial guesses.

Consequently, alternative methods can be used to obtain accurate estimators. Usually, instead of computing and minimizing geometric distances, one minimizes various 'algebraic' distances, and as such, this type of fits is known as *algebraic fits*. Such fits are non-iterative. They are simpler and faster, but usually, all of them are less accurate than the geometric fit(this issue will be fully discussed later). In this paper, we will discuss the most popular fits: the Kåsa fit, the Taubin fit, the Pratt fit, hyper fit, and HyperLS fit. Indeed, algebraic fits use the standard form of the circle

$$A(x^2 + y^2) + Bx + Cy + D = 0. \tag{1.3}$$

If we denote $z = x^2 + y^2$, $\mathbf{z} = (z, x, y, 1)^T$, and denote the vector of algebraic parameters by $\mathbf{A} = (A, B, C, D)^T$, then Eq. (1.3) can be written as $\mathbf{z}^T\mathbf{A} = 0$. Thus, all algebraic fits minimize the objective function

$$\mathcal{F}(\mathbf{A}) = \frac{1}{n}\sum_{i=1}^{n}\big[Az_i + Bx_i + Cy_i + D\big]^2 := \mathbf{A}^T\mathbf{M}\mathbf{A} \tag{1.4}$$

but with different constraints imposed. The matrix of the moments $\mathbf{M}$ is a positive semidefinite matrix defined as

$$\mathbf{M} = \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{M}_i, \text{ where } \mathbf{M}_i = \mathbf{z}_i\mathbf{z}_i^T. \tag{1.5}$$

Since $\mathbf{M}$ is positive semidefinite, the minimum of $\mathcal{F}$ occurs at $\mathbf{A} = 0$. To remove the trivial solution, one must impose the constraint $\mathbf{A}^T\mathbf{N}\mathbf{A} = 1$ on the parametric space for the symmetric matrix $\mathbf{N}$.

Kåsa fit. The simplest and fastest method was introduced in the 1970s by Delogne [15] and Kåsa [24]. This method is one of the most beloved popular methods. Intuitively, it minimizes

$$\mathcal{F}_{\mathrm{K}}(\mathbf{A}) = \frac{1}{n}\sum_{i=1}^{n}(z_i + Bx_i + Cy_i + D)^2. \tag{1.6}$$

In other words, Kåsa's fit minimizes $\mathcal{F}$ in Eq. (1.4) but is subject to the constraint $A = 1$, which can be written as $\mathbf{A}^T\mathbf{N}_{\mathrm{K}}\mathbf{A} = 1$, where $\mathbf{N}_{\mathrm{K}} = \check{\mathbf{e}}_1\check{\mathbf{e}}_1^T$ with $\check{\mathbf{e}}_1 = (1, 0, 0, 0)^T$. The Kåsa method is perhaps the fastest circle fit, but its accuracy suffers when one observes incomplete circular arcs (partially occluded circles); then the Kåsa fit is known to be heavily biased toward small circles [12].

Pratt's fit. Pratt [27] proposed a clever fit based on a simple algebraic relation between circle's parameters. Pratt noticed that the radius of the circle R is related to the algebraic circle parameter $\mathbf{A}$. By completing the square, we can write Eq. (1.3) as

$$\left(x - \frac{B}{2A}\right)^2 + \left(y - \frac{C}{2A}\right)^2 = \frac{B^2 + C^2 - 4AD}{4A^2}, \tag{1.7}$$

which represents a circle with center (a, b) and radius $R = \sqrt{\frac{B^2+C^2-4AD}{4A^2}}$. Thus, it is natural to impose the constraint $B^2 + C^2 - 4AD > 0$. However, multiplying this constraint by a nonzero constant does not change the circle that it represents. As such, Pratt minimized $\mathcal{F}$ subject to the constraint

$$1 = B^2 + C^2 - 4AD = \mathbf{A}^T\mathbf{N}_{\mathrm{P}}\mathbf{A} \tag{1.8}$$

where

$$\mathbf{N}_{\mathrm{P}} \stackrel{\mathrm{def}}{=} \begin{bmatrix} 0 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 0 & 0 & 0 \end{bmatrix}. \tag{1.9}$$

Taubin's fit. The Taubin fit [31] is even more accurate than Pratt's fit, though both fall behind the geometric fit. Expressing $\mathcal{F}_{\mathrm{T}}$ in terms of A, B, C, D shows that

$$\mathcal{F}_{\mathrm{T}}(\mathbf{A}) = \frac{1}{n}\sum_{j=1}^{n}\frac{[Az_j + Bx_j + Cy_j + D]^2}{n^{-1}\sum_{i=1}^{n}[4A^2z_i + 4ABx_i + 4ACy_i + B^2 + C^2]}. \tag{1.10}$$

Equivalently, one can minimize $\mathcal{F}$ subject to a new constraint

$$4A^2\bar{z} + 4AB\bar{x} + 4AC\bar{y} + B^2 + C^2 = 1, \tag{1.11}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Again, the latter constraint can be written as $\mathbf{A}^T\mathbf{N}_{\mathrm{T}}\mathbf{A} = 1$, where

$$\mathbf{N}_{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{T}_i, \tag{1.12}$$

and

$$\mathbf{T}_i = \begin{bmatrix} 4z_i & 2x_i & 2y_i & 0 \\ 2x_i & 1 & 0 & 0 \\ 2y_i & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{1.13}$$

Hyperaccurate fit. In an attempt to compare between the above algebraic fits and the geometric fit through higher order error analysis, we found [1] an algebraic fit that outperforms other algebraic fits and *sometimes* outperforms the geometric fit! The new '*non-iterative*' algebraic fit is obtained by minimizing $\mathcal{F}$ subject to $\mathbf{A}^T\mathbf{N}_{\mathrm{H}}\mathbf{A} = 1$, where $\mathbf{N}_{\mathrm{H}} = 2\mathbf{N}_{\mathrm{T}} - \mathbf{N}_{\mathrm{P}}$. We called this novel algebraic fit 'hyperaccurate fit' or hyper fit for short.

Hyper fit and other previously mentioned fits have an important property. They are independent of the choice of the coordinate system (i.e., their results are invariant under translations and rotations; see a proof in [10]). Therefore, we will study them, even though there are many other approaches in the modern literature discussing circle fitting problem [7, 13, 35, 28, 30, 36, 38]. However, most of them are either quite slow or can be reduced to one of the algebraic fits [10, Chapter 8].

Hyper least squares fit (HyperLS). Motivated by our findings, Kanatani and Rangarajan [23] recently developed another fit that outperforms our fit when the sample size is small (but both fits are identical when n is large). Their fit minimizes $\mathcal{F}$ subject to $\mathbf{A}^T\mathbf{N}_{\mathrm{I}}\mathbf{A}$, where

$$\mathbf{N}_{\mathrm{I}} = \mathbf{N}_{\mathrm{H}} - \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{trace}(\mathbf{M}^{-}\mathbf{V}_i)\mathbf{M}_i + (\mathbf{z}_i^T\mathbf{M}^{-}\mathbf{z}_i)\mathbf{V}_i + 2\mathcal{S}\big[\mathbf{V}_i\mathbf{M}^{-}\mathbf{M}_i\big],$$

where we denote the symmetrization operator by $\mathcal{S}$, i.e. $\mathcal{S}(\cdot) = \frac{(\cdot)+(\cdot)^T}{2}$, while $(\,)^{-}$ denotes the Moore-Penrose pseudo-inverse matrix and $\mathbf{V}_i$ is the first-order covariance matrix of $\mathbf{z}_i$ (will be stated precisely later).

All previously mentioned algebraic fits can be obtained by solving the generalized eigenvector problem

$$\mathbf{MA} = \lambda\mathbf{NA}. \tag{1.14}$$

Eq. (1.14) determines $\mathbf{A}$ up to a scalar multiple (i.e., multiplying $\mathbf{A}$ by a nonzero scalar does not change the circle it represents), so we can set $\|\mathbf{A}\|_2 = 1$. Then we can convert the algebraic circle parameters A, B, C, D to the natural parameters via the relationships

$$a = -\frac{B}{2A}, \qquad b = -\frac{C}{2A}, \qquad R^2 = \frac{B^2 + C^2 - 4AD}{4A^2}. \tag{1.15}$$

In [1], we studied some of previously mentioned algebraic circle methods (Kåsa's fit, Taubin's fit, and Pratt's fit) and the geometric fit by applying error analysis in order to compare them. Our study allowed us to propose a new fit (hyper fit) that outperforms other algebraic fits and the geometric fit. Based on our results, Kanatani and Rangarajan [23] have proposed another algebraic fit that outperforms ours and the geometric fit (in some cases). Our results led to significant contributions to other curve fitting problems, such as conic fitting [3], and other applications in computer vision, such as 'Fundamental Matrix computation [4]. Moreover, their influences have appeared in research areas such as electrical and petroleum engineering [25], medical imaging [39], mechatronics, and automation (i.e., LED chip sorting process) [33], remote sensing [32], and others [14, 34].

On other hand, we noticed in some situations that the geometric fit still practically performs better than hyper and HyperLS. Of course, our conclusions in [1, 2] are solely based on first-order error terms of the variances and the first-order of the bias for each algebraic fits, while the second-order term of the variance dropped. Therefore, we believe that our findings in [1] do not reveal the whole picture about the performance of circle fits. This paper is devoted to studying circle fits further through computing the mean square errors (MSE) up to the third leading term, and as such, we can compare between them. To our knowledge, until now, no one has performed a detailed theoretical comparison of the accuracy of various circle fits (or even computed the second-order error terms of MSE). All studies in literature are based on the empirical MSE obtained by Monte Carol simulation. In this paper, we will show when the geometric fit outperforms the new novel algebraic fits, and we will propose a new fit that outperforms all existing fits.

Our paper is organized as follows. Section 2 is devoted to introducing notations, some statistical assumptions, and some related results. In Section 3, we conduct a detailed error analysis for geometric fit and algebraic fits. We will see when the geometric fit outperforms other fits. This rigorous analysis involves many lengthy derivations that are moved to Appendix. Section 4 presents a new fit and some numerical experiments that validate our findings.

2. Notations, assumptions, and previous results

For a more general consideration, suppose that a family of curves depend on p parameters, say $\theta_1, \ldots, \theta_p$. Our goal is to estimate the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ that corresponds to the best fitting curve. Let us assume that the *'true parameter vector'* $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_p)^T$ corresponds to the *'true curve'*. Also, we will assume that the true points $\tilde{\mathbf{m}}_i = (\tilde{x}_i, \tilde{y}_i)^T$ lie in the true curve. Mathematically, the true point satisfies the implicit equation

$$P(\tilde{\mathbf{m}}_i; \tilde{\boldsymbol{\theta}}) = 0, i = 1, \ldots, n. \tag{2.1}$$

In the circular regression, which is our model in this paper, the true points satisfy

$$(\tilde{x}_i - \tilde{a})^2 + (\tilde{y}_i - \tilde{b})^2 = \tilde{R}^2, \qquad i = 1, \ldots, n, \tag{2.2}$$

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

$$\tilde{x}_i = \tilde{a} + \tilde{R}\cos\varphi_i, \qquad \tilde{y}_i = \tilde{b} + \tilde{R}\sin\varphi_i, \tag{2.3}$$

where $\varphi_1, \ldots, \varphi_n$ specifies the locations of the true points on the true circle. We denote $\cos\varphi_i$ and $\sin\varphi_i$ by $\tilde{u}_i$ and $\tilde{v}_i$, respectively. Also, we consider $\mathbf{m}_i = (x_i, y_i)^T$ as an '*inaccurate measurement*' of the *true point* $\tilde{\mathbf{m}}_i$, i.e. $\mathbf{m}_i = \tilde{\mathbf{m}}_i + \mathbf{e}_i$, where $\mathbf{e}_i$ is the error vector. This means that this model assumes all variables are subject to errors and it is known in the literature as Errors-In-Variables (EIV) model, which is much more difficult than and different from the classical regression.

To understand the statistical properties of any estimator of $\boldsymbol{\theta}$, one should adopt realistic assumptions about the true points and the probability distribution of the observations. We assume here that the true points $(\tilde{x}_i, \tilde{y}_i)'s$ are fixed (but unknown), so they are treated as nuisance parameters. In EIV models, this model is known as a *functional model*, which is widely adopted in the applied community, especially in computer vision sciences.

On the other hand, we assume that the error vector $\mathbf{e}_i = (\delta_i, \varepsilon_i)^T \sim \mathbb{N}(\mathbf{0}, \sigma^2\mathbf{I}_2)$ and $\mathbf{I}_2$ is an identity matrix of size 2 for each $i = 1, \ldots, n$. The isotropic assumption about the noise is intuitively realistic for computer vision applications. For edge detection, in pattern recognition and computer vision, detecting (observing) any point $\mathbf{m}_i = (x_i, y_i)^T$ does not give any information about other points, and hence, the errors are independent. It is also reasonable to assume that the errors $\mathbf{e}_i$'s have the same covariance matrix for all points because we use the same algorithm for edge detection. With these statistical assumptions, the geometric fit returns the 'Maximum Likelihood Estimator' (MLE) for a circle's parameters [8], i.e. $\hat{\boldsymbol{\theta}} = (\hat{a}_{\text{MLE}}, \hat{b}_{\text{MLE}}, \hat{R}_{\text{MLE}}) = \operatorname{argmin} \mathcal{G}(a, b, R)$. To distinguish between estimates of the natural circle parameters (a, b, R) obtained by algebraic fits and geometric fit, we will generally denote by $\hat{\boldsymbol{\theta}}^a = (\hat{a}^a, \hat{b}^a, \hat{R}^a)$ the estimates obtained by algebraic fits.

Small-noise assumption. We assume that $\sigma \ll 1$ is known. The sample size n is fixed, though it is not very small. This approach refers back to Kadane [18] and was employed by Anderson [5] and other statisticians [6]. More recently, it was used by Kanatani [20, 22] in image processing applications, and Kanatani argued that the 'small noise' model, where $\sigma \to 0$ while the sample size n is kept fixed, is more appropriate than the traditional statistical 'large sample' approach, where $n \to \infty$ while $\sigma > 0$ is kept fixed. We use a combination of these two models. Our main assumption is that $\sigma \to 0$, but n is regarded as a slowly increasing parameter; more precisely, we assume that $n \ll \sigma^{-2}$. Therefore, it is convenient to assume that

$$\hat{\boldsymbol{\theta}}(\tilde{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}. \tag{2.4}$$

Precisely, Eq. (2.4) means that when $\sigma = 0$ (i.e., when the true points are observed without noise), then the estimator returns the true parameter vector (i.e., finds the true curve). Geometrically, this means that if there is a model curve that interpolates the data points, then the algorithm finds it. With some degree of informality, one can assert that when Eq. (2.4) holds, the estimator $\hat{\boldsymbol{\theta}}$ is *consistent* in the limit $\sigma \to 0$. Thus, we call this property *geometric consistency*. This is regarded as a minimal requirement for any sensible fitting algorithm. For example, if the observed points lie on one circle, then every circle fitting algorithm finds that circle uniquely. Kanatani [21] remarks that algorithms which fail to follow this property "are not worth considering".

Kanatani-Cramèr-Rao bound (KCR). As a standard measure in statistics, the efficiency of any unbiased estimator can be determined by Cramèr-Rao lower bound (CRB) obtained by taking the inverse of Fisher information matrix (FIM). Evaluating the constrained CRB for circles or other general geometric fittings has long history. For circle fitting problem, the CRB was derived in 1995 by Chan-Thomas [9], Kanatani [19] derived a general CRB for arbitrary curves for any unbiased estimators. Then, independently, Zelniker and Clarkson presented another proof in 2006 (see [37] for more details). That is, let $\hat{\boldsymbol{\theta}}$ be an unbiased estimator of $\boldsymbol{\theta}$ satisfying Eq. (2.1). Then, there is a symmetric positive semi-definite matrix $\tilde{\mathbf{V}}_{\min}$

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}) \geq \sigma^2 \mathbf{V}_{\min} = \sigma^2 \left(\sum_{i=1}^{n} \frac{P_{\boldsymbol{\theta} i}\, P_{\boldsymbol{\theta} i}^T}{\|P_{\mathbf{m} i}\|^2} \right)^{-}, \tag{2.5}$$

where the notation $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite matrix and

$$P_{\boldsymbol{\theta} i} = \Big(\partial P(\tilde{\mathbf{m}}_i; \tilde{\boldsymbol{\theta}})/\partial \theta_1, \ldots, \partial P(\tilde{\mathbf{m}}_i; \tilde{\boldsymbol{\theta}})/\partial \theta_p \Big)^T \tag{2.6}$$

$$P_{\mathbf{m} i} = \Big(\partial P(\tilde{\mathbf{m}}_i; \tilde{\boldsymbol{\theta}})/\partial x, \partial P(\tilde{\mathbf{m}}_i; \tilde{\boldsymbol{\theta}})/\partial y \Big)^T \tag{2.7}$$

stand for the gradient of P with respect to the model parameters $\theta_1, \ldots, \theta_p$ and for the gradient with respect to the planar variables x and y, respectively; both gradients are taken at the true point $\tilde{\mathbf{m}}_i = (\tilde{x}_i, \tilde{y}_i)^T$. In terms of the natural parameter $\boldsymbol{\theta} = (a, b, R)^T$, a circle is defined by $P = (x-a)^2 + (y-b)^2 - R^2$. Thus,

$$P_{\boldsymbol{\theta} i} = -2\big((\tilde{x}_i - \tilde{a}), (\tilde{y}_i - \tilde{b}), \tilde{R}\big)^T, \qquad P_{\mathbf{m} i} = 2\big((\tilde{x}_i - \tilde{a}), (\tilde{y}_i - \tilde{b})\big)^T. \tag{2.8}$$

Therefore,

$$\tilde{\mathbf{V}}_{\min} = \frac{1}{n} \begin{bmatrix} \overline{\tilde{u}\tilde{u}} & \overline{\tilde{u}\tilde{v}} & \tilde{u} \\ \overline{\tilde{u}\tilde{v}} & \overline{\tilde{v}\tilde{v}} & \bar{\tilde{v}} \\ \bar{\tilde{u}} & \bar{\tilde{v}} & 1 \end{bmatrix}^{-1} := (\tilde{\mathbf{W}}^T \tilde{\mathbf{W}})^{-1}, \tag{2.9}$$

where we use standard notation for sample means $\overline{\tilde{u}\tilde{u}} = \frac{1}{n}\sum \tilde{u}_i^2$, and $\overline{\tilde{u}\tilde{v}} = \frac{1}{n}\sum \tilde{u}_i \tilde{v}_i$, etc. And

$$\tilde{\mathbf{W}} = \begin{bmatrix} \tilde{u}_1 & \tilde{v}_1 & 1 \\ \vdots & \vdots & \vdots \\ \tilde{u}_n & \tilde{v}_n & 1 \end{bmatrix}. \tag{2.10}$$

Similarly, the lower bound of the algebraic parameters is simply

$$\tilde{\mathbf{V}}^a_{\min} = \frac{\tilde{\zeta}}{n}\tilde{\mathbf{M}}^-, \tag{2.11}$$

where $\tilde{\mathbf{M}}$ is the noiseless version of $\mathbf{M}$, and

$$\tilde{\zeta} = \tilde{B}^2 + \tilde{C}^2 - 4\tilde{A}\tilde{D}, \tag{2.12}$$

One obvious disadvantage of this lower bound is that it can be applied only to unbiased estimators! However, all estimators in EIV models are, roughly speaking, biased. In the early 2000's, Chernov and Lesort [11] realized that Kanatani's formula does not work for any (practical) geometrically consistent estimator in circle or other curve fitting problems because all existing estimators are biased. To overcome of this situation, they [11] employed first-order analysis for any geometrically consistent estimators and derived the minimal possible lower bound, to the first leading term. That is, there is a positive semi-definite matrix $\tilde{\mathbf{V}}_{\min}$, such that *the leading term of the variance matrix*, say $\mathbf{V}$, of any geometrically consistent estimator $\hat{\boldsymbol{\theta}}$ satisfying Eq.(2.4) has a natural bound $\sigma^2\tilde{\mathbf{V}}_{\min}$, i.e. $\mathbf{V} \geq \sigma^2\tilde{\mathbf{V}}_{\min}$. This bound of the leading term of variance coincides with Kanatani's formula (this is applied only to unbiased estimators). Therefore, they called this bound the *Kanatani-Cramér-Rao lower bound* (KCR).

2.1. Previous results

Our main goal in [1, 2] was to compare the most popular circle fits (geometric fit and other various algebraic fits such as Kåsa's fit, Pratt's fit, and Taubin's fit). We characterized the accuracy of estimators based on *(Total) Mean Squared Error (MSE)*:

$$\mathbb{E}\big[(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})^T\big] = \mathbf{b}(\hat{\boldsymbol{\theta}})\mathbf{b}(\hat{\boldsymbol{\theta}})^T + \operatorname{cov}(\hat{\boldsymbol{\theta}}),$$

where $\mathbf{b}(\hat{\boldsymbol{\theta}})$ is the bias of $\hat{\boldsymbol{\theta}}$, i.e. $\mathbf{b}(\hat{\boldsymbol{\theta}}) = \mathbb{E}(\hat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{\theta}}$. We used Taylor series expansion and geometric consistency (i.e. $\hat{\boldsymbol{\theta}}(\tilde{\mathcal{X}}) = \tilde{\boldsymbol{\theta}}$) to derive the quadratic approximation of $\hat{\boldsymbol{\theta}}$. i.e.,

$$\hat{\boldsymbol{\theta}}_{\mathrm{Q}} = \tilde{\boldsymbol{\theta}} + \Delta_1\hat{\boldsymbol{\theta}} + \Delta_2\hat{\boldsymbol{\theta}},$$

where the first-order error term is denoted by $\Delta_1\hat{\boldsymbol{\theta}} = (\Delta_1\hat{a}, \Delta_1\hat{b}, \Delta_1\hat{R})^T$. Its components form linear combinations of ε_i's and δ_i's, while the second-order error term is denoted by $\Delta_2\hat{\boldsymbol{\theta}} = (\Delta_2\hat{a}, \Delta_2\hat{b}, \Delta_2\hat{R})^T$, where$\Delta_2\hat{a}$, $\Delta_2\hat{b}$, and $\Delta_2\hat{R}$ are quadratic forms of ε_i's and δ_i's. Therefore,

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}_{\mathrm{Q}}) = \mathrm{MSE}(\Delta_1\hat{\boldsymbol{\theta}}) + \mathrm{MSE}(\Delta_2\hat{\boldsymbol{\theta}}). \tag{2.13}$$

Then, we classified each term in Eq. (2.13) based on its order of magnitude. The most significant term is the leading term of order $\mathcal{O}(\sigma^2/n)$ that comes from $\mathrm{MSE}(\Delta_1\hat{\boldsymbol{\theta}}) = \operatorname{cov}(\Delta_1\hat{\boldsymbol{\theta}})$, which is of order $\mathcal{O}(\sigma^2/n)$. A precise expression for the geometric fit is given in the following theorem (its proof is moved to appendix).

Theorem 2.1 ([1]). *Let $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{R})^T$ be the geometric fit and $\hat{\boldsymbol{\theta}}_{\mathrm{L}} = \tilde{\boldsymbol{\theta}} + \Delta_1\hat{\boldsymbol{\theta}}$ be its linear approximation. Then, $\Delta_1\hat{\boldsymbol{\theta}} = \tilde{\mathbf{K}}\mathbf{f}_1$ and the linear approximation of $\hat{\boldsymbol{\theta}}$ is $\hat{\boldsymbol{\theta}}_{\mathrm{L}} \sim \mathbb{N}(\tilde{\boldsymbol{\theta}}, \sigma^2(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1})$, where*

$$\tilde{\mathbf{K}} = (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^T, \quad \mathbf{f}_1 = \tilde{\mathbf{U}}\delta + \tilde{\mathbf{V}}\boldsymbol{\varepsilon}, \tag{2.14}$$

and $\delta = (\delta_1, \dots, \delta_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$, while $\tilde{\mathbf{U}} = \mathrm{diag}(\tilde{u}_1, \dots, \tilde{u}_n)$ and $\tilde{\mathbf{V}} = \mathrm{diag}(\tilde{v}_1, \dots, \tilde{v}_n)$.

This means that if we restricted our error analysis to the first leading term, then the geometric fit would be an unbiased and efficient estimator of $\boldsymbol{\theta}$ in the sense that the variance of its linear approximation achieves its minimal possible bound $\sigma^2\tilde{\mathbf{V}}_{\min}$ in the small-noise limit; this was proven by Fuller (Theorem 3.2.1 in [16]) for the geometric fit and independently by Chernov and Lesort [11]. In fact, the same property holds for all algebraic fits. That is, since $\mathbf{z}_i$ is a quadratic form of the random vector (x_i, y_i), it can be written as

$$\mathbf{z}_i = \tilde{\mathbf{z}}_i + \Delta_1\mathbf{z}_i + \Delta_2\mathbf{z}_i, \tag{2.15}$$

where the first-order error $\Delta_1\mathbf{z}_i$ is a linear combination of $(\delta_i, \varepsilon_i)$, for each $i = 1, \dots, n$. That is,

$$\Delta_1\mathbf{z}_i = \begin{bmatrix} 2\tilde{x}_i & 1 & 0 & 0 \\ 2\tilde{y}_i & 0 & 1 & 0 \end{bmatrix}^T \begin{bmatrix} \delta_i \\ \varepsilon_i \end{bmatrix} := \tilde{\mathbf{a}}_i\delta_i + \tilde{\mathbf{b}}_i\varepsilon_i, \tag{2.16}$$

where $\tilde{\mathbf{a}}_i = (2\tilde{x}_i, 1, 0, 0)^T$ and $\tilde{\mathbf{b}}_i = (2\tilde{y}_i, 0, 1, 0)^T$, while the second-order error term is $\Delta_2\mathbf{z}_i = (\delta_i^2 + \varepsilon_i^2, 0, 0, 0)^T$. Accordingly, for each $i = 1, \dots, n$, the covariance matrix of $\Delta_1\mathbf{z}_i$ (i.e., $\mathrm{cov}(\Delta_1\mathbf{z}_i)$) is equal to $\sigma^2\tilde{\mathbf{V}}_i$, where

$$\tilde{\mathbf{V}}_i = \tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + \tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T = \tilde{\mathbf{T}}_i. \tag{2.17}$$

Also, for each $i \neq j$, $\mathrm{cov}(\Delta_1\mathbf{z}_i, \Delta_1\mathbf{z}_j) = \mathbf{0}$ (a zero matrix of size 4) and $\mathbb{E}(\Delta_2\mathbf{z}_i) = 2\sigma^2\breve{\mathbf{e}}_1$. Let us write $\Delta\mathbf{A} = \mathbf{A} - \tilde{\mathbf{A}}$. We can decompose $\Delta\mathbf{A}$ by Taylor expansion as

$$\Delta\mathbf{A} = \Delta_1\mathbf{A} + \Delta_2\mathbf{A} + \mathcal{O}_{\mathrm{P}}(\sigma^3),$$

where $\Delta_1\mathbf{A}$ is the linear combination of δ_i's and ε_i's, $\Delta_2\mathbf{A}$ is a quadratic form of δ_i's and ε_i's, and all other higher order terms are represented by $\mathcal{O}_P(\sigma^3)$. Next, if we apply matrix perturbation to λ, $\mathbf{A}$, $\mathbf{M}$ and $\mathbf{N}$, then we get

$$\begin{aligned} \lambda &= \tilde{\lambda} + \Delta_1\lambda + \Delta_2\lambda + \mathcal{O}_{\mathrm{P}}(\sigma^3), \\ \mathbf{M} &= \tilde{\mathbf{M}} + \Delta_1\mathbf{M} + \Delta_2\mathbf{M} + \mathcal{O}_{\mathrm{P}}(\sigma^3), \\ \mathbf{N} &= \tilde{\mathbf{N}} + \Delta_1\mathbf{N} + \Delta_2\mathbf{N} + \mathcal{O}_{\mathrm{P}}(\sigma^3). \end{aligned} \tag{2.18}$$

Hence, $\mathbf{MA} = \lambda\mathbf{NA}$ becomes

$$\begin{aligned} (\tilde{\mathbf{M}} + \Delta_1\mathbf{M} + \cdots)(\tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \cdots) = (\tilde{\lambda} + \Delta_1\lambda + \cdots) \\ (\tilde{\mathbf{N}} + \Delta_1\mathbf{N} + \cdots)(\tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \cdots). \end{aligned} \tag{2.19}$$

Equating terms of the same order from both sides implies that $\tilde{\mathbf{M}}\tilde{\mathbf{A}} = \tilde{\lambda}\,\tilde{\mathbf{N}}\tilde{\mathbf{A}}$. However, $\tilde{\lambda} = 0$ because of the fact that

$$\tilde{\mathbf{A}}^T\tilde{\mathbf{M}}\tilde{\mathbf{A}} = \frac{1}{n}\sum(\tilde{\mathbf{z}}_i^T\tilde{\mathbf{A}})^2 = 0,$$

which follows from $\tilde{\mathbf{A}}^T\mathbf{z}_i = 0$ for all $i = 1, \dots, n$. In the same manner, we can show that $\Delta_1\lambda = 0$. That is, equating all terms of order $\mathcal{O}_{\mathrm{P}}(\sigma)$ in Eq. (2.19) and using $\tilde{\lambda} = 0$, we obtain

$$\tilde{\mathbf{M}}\Delta_1\mathbf{A} + \Delta_1\mathbf{M}\tilde{\mathbf{A}} = \Delta_1\lambda\tilde{\mathbf{N}}\tilde{\mathbf{A}}. \tag{2.20}$$

Premultiplying Eq. (2.20) by $\tilde{\mathbf{A}}$ makes the left-hand side of the resulting equation equal to zero. This follows immediately from the formal expression of $\mathbf{M}$ defined in Eq. (1.5). Thus,

$$\Delta_1\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\Delta_1\mathbf{M}_i, \quad \Delta_1\mathbf{M}_i = \tilde{\mathbf{z}}_i\,\Delta_1\mathbf{z}_i^T + \Delta_1\mathbf{z}\,\tilde{\mathbf{z}}_i^T. \tag{2.21}$$

Since $\tilde{\mathbf{A}}^T\tilde{\mathbf{z}}_i = 0$, one notices that $\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\tilde{\mathbf{A}} = 0$, and as such, $\Delta_1\lambda = 0$. Therefore,

$$\Delta_1\mathbf{A} = -\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}\tilde{\mathbf{A}}. \tag{2.22}$$

Our general error analysis in [1] shows that the most important part of MSE comes from the variance and has the order of magnitude $\mathcal{O}(\sigma^2/n)$. Namely, $\mathbb{E}(\Delta_1\mathbf{A}\Delta_1\mathbf{A}^T)$. This term has the form

$$\mathbb{E}(\Delta_1\mathbf{A}\Delta_1\mathbf{A}^T) = \tilde{\mathbf{M}}^-\mathbb{E}(\Delta_1\mathbf{M}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M})\tilde{\mathbf{M}}^-.$$

Using $\Delta_1\mathbf{M}\tilde{\mathbf{A}} = \frac{1}{n}\sum_{i=1}^n(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)\tilde{\mathbf{z}}_i$ and $\tilde{\zeta} = \tilde{B}^2 + \tilde{C}^2 - 4\tilde{A}\tilde{D}$ gives

$$\begin{aligned}\mathbb{E}(\Delta_1\mathbf{A}\Delta_1\mathbf{A}^T) &= \frac{1}{n^2}\sum_{i,j=1}^{n}\tilde{\mathbf{M}}^-\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_j^T\big)\tilde{\mathbf{M}}^- \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\tilde{\mathbf{M}}^-\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)^2\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T\big)\tilde{\mathbf{M}}^-.\end{aligned}$$

But

$$\mathbb{E}((\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)^2) = \sigma^2(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}) = \tilde{\zeta}\sigma^2.$$

As a standard fact in matrix theory, $\tilde{\mathbf{M}}^-\tilde{\mathbf{M}}\tilde{\mathbf{M}}^- = \tilde{\mathbf{M}}^-$, where $\tilde{\mathbf{M}} = \frac{1}{n}\sum_{i=1}^n\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T$. Thus

$$\mathbb{E}(\Delta_1\mathbf{A}\Delta_1\mathbf{A}^T) = \frac{\sigma^2\tilde{\zeta}}{n}\tilde{\mathbf{M}}^-\Big[\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_i^T\Big]\tilde{\mathbf{M}}^- = \frac{\sigma^2\tilde{\zeta}}{n}\tilde{\mathbf{M}}^- = \sigma^2\tilde{\mathbf{V}}^a_{\min}.$$

Furthermore, to express the covariance matrix of algebraic circle fits in terms of the natural parameters $\boldsymbol{\theta} = (a, b, R)^T$, we need to write $\Delta_1\mathbf{A}$ in terms of $\Delta_1\hat{\boldsymbol{\theta}}$.

Taking partial derivatives in Eq. (1.15) gives a 3×4 'Jacobian' matrix

$$\mathbf{J} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{B}{2A^2} & -\frac{1}{2A} & 0 & 0 \\ \frac{C}{2A^2} & 0 & -\frac{1}{2A} & 0 \\ -\frac{R}{A} - \frac{D}{2A^2R} & \frac{B}{4A^2R} & \frac{C}{4A^2R} & -\frac{1}{2AR} \end{bmatrix}. \tag{2.23}$$

Thus we have

$$\begin{aligned} \Delta_1 \hat{\boldsymbol{\theta}}^a &= \tilde{\mathbf{J}}\, \Delta_1 \mathbf{A} + \mathcal{O}_P(\sigma/\sqrt{n}), \\ \Delta_2 \hat{\boldsymbol{\theta}}^a &= \tilde{\mathbf{J}}\, \Delta_2 \mathbf{A} + \mathcal{O}_P(\sigma^2/n), \\ \Delta_3 \hat{\boldsymbol{\theta}}^a &= \tilde{\mathbf{J}}\, \Delta_3 \mathbf{A} + \mathcal{O}_P(\sigma^3/n) \end{aligned} \tag{2.24}$$

where the superscript here is used to distinguish the estimate of the algebraic fits from the geometric fit when the algebraic fits are expressed in terms of (a, b, R). The matrix $\tilde{\mathbf{J}}$ is $\mathbf{J}$ evaluated at the true parameters $(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})$.

It is remarkable to note that the leading term of the covariance matrix of any algebraic fit does not depend on the constraint matrix $\mathbf{N}$, and as such, all algebraic circle fits have the same variances (to the leading order). Moreover, the first-order approximation of their variances coincides with that of the geometric circle fit (see [1]). Thus, the geometric fitand all algebraic fits return estimators of (a, b, R) that attain KCR (i.e., $\sigma^2 \tilde{\mathbf{V}}_{\min}$). Thus, they are all optimal in this sense.

However, they behave differently in practice. In order to distinguish among them, one should adopt the following strategy. Since all fits attain KCR bound, one should employ the quadratic approximation $\hat{\boldsymbol{\theta}}_{\mathrm{Q}}$ (or even the third-order approximation, say $\hat{\theta}_{\mathrm{C}}$). Then one should compare between fits by considering their quadratic or cubic approximations.

Our goal in [1] was to do so. We developed an unconventional approach that works for general curve fitting. Then, we applied it to circle fitting. We studied $\mathrm{MSE}(\Delta_2 \hat{\boldsymbol{\theta}})$, which has terms of order σ^4 that involve complicated formulas. Excluding n from the picture makes all terms of order σ^4 contribute equally. Our general analysis, however, distinguishes between them based on their dependence on n. Terms of order σ^4 are more significant than terms of order σ^4/n, and so on. Terms of order $\mathcal{O}(\sigma^4)$ in MSE come from different sources: $\mathbf{b}(\Delta_2 \hat{\boldsymbol{\theta}})\mathbf{b}(\Delta_2 \hat{\boldsymbol{\theta}})^T$ and $\mathrm{cov}(\Delta_2 \hat{\boldsymbol{\theta}})$. Our analysis showed that $\mathrm{var}(\Delta_2 \hat{\boldsymbol{\theta}}) \sim \mathcal{O}(\sigma^4/n)$. We also decomposed the bias into two parts: (1) *the essential bias*, say $\mathbf{b}_1(\hat{\boldsymbol{\theta}}_{\mathrm{Q}})$ of order of magnitude σ^2 and (2) *the nonessential bias* $\mathbf{b}_2(\hat{\boldsymbol{\theta}}_{\mathrm{Q}})$ with its order of magnitude σ^2/n.

Since all fits are optimal in the sense that their variances, to leading term, attain KCR, we went one step further and compared the second significant terms in the MSE. $\mathbf{b}_1 \mathbf{b}_1^T$ comes from the essential bias (our analysis in [1] discarded all terms of order σ^4/n and σ^4/n^2). The goal in this paper is to refine our analysis by including terms of order σ^4/n.

The next theorem summarizes one of our findings [1] in the context of the geometric fit. Its proof is important in our sequel analysis. However, it involves many useful details, which are summarized in the following two lemmas (see their proofs in the appendix).

Lemma 2.1. *For each $i, j = 1, \ldots, n$, define*

$$\tilde{\mathbf{t}}_i = (-\tilde{v}_i, \tilde{u}_i, 0)^T, \quad s_{i,j} = \tilde{\mathbf{t}}_i^T(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{t}}_j, \;\; \check{\mathbf{n}}_i = (\delta_i, \varepsilon_i, 0)^T.$$

Also, let $\tau_i = \tilde{\mathbf{t}}_i^T \Delta_1 \boldsymbol{\theta}$ and $\rho_i = \tilde{\mathbf{t}}_i^T \check{\mathbf{n}}_i$. Then, for all $i, j = 1, \ldots, n$, one has

$$\mathbb{E}(\tau_i\tau_j) = \sigma^2 s_{i,j}, \quad \mathbb{E}(\tau_i\rho_j) = 0, \quad \mathbb{E}(\rho_i\rho_j) = \sigma^2\delta_{i,j},$$

where δ_{ij} is the Kronecker Delta function.

Lemma 2.2. *Define $a_i = \rho_i^2$, $b_i = -2\tau_i\rho_i$, and $c_i = \tau_i^2$. Let $\tilde{\mathbf{s}} = (\tilde{s}_{1,1}, \ldots, \tilde{s}_{n,n})^T$, and similarly, define the $n-$dimensional vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$. Then,*

$$\mathbb{E}(\mathbf{a}) = \sigma^2\mathbf{1}, \quad \mathbb{E}(\mathbf{b}) = \mathbf{0}, \quad \mathbb{E}(\mathbf{c}) = \sigma^2\tilde{\mathbf{s}}.$$

Moreover,

$$\mathbb{E}(\mathbf{a}\mathbf{a}^T) = \sigma^4\big(\mathbf{1}_n\mathbf{1}_n^T + 2\,\mathbf{I}_n\big), \quad \mathbb{E}(\mathbf{b}\mathbf{c}^T) = \mathbb{E}(\mathbf{a}\mathbf{b}^T) = \mathbf{0}, \quad \mathbb{E}(\mathbf{a}\mathbf{c}^T) = \sigma^4\tilde{\mathbf{s}}\mathbf{1}^T,$$

and

$$\mathbb{E}(\mathbf{b}\mathbf{b}^T) = 4\sigma^4\mathrm{Diag}\big(\tilde{s}_{1,1}, \tilde{s}_{2,2}, \ldots, \tilde{s}_{n,n}\big), \qquad \mathbb{E}(\mathbf{c}\mathbf{c}^T) = \sigma^4\tilde{\mathbf{S}},$$

where $\mathbf{I}_n$ is an identity matrix of size n, $\mathbf{1}_n = (1, 1, \ldots, 1)^T$, and $\tilde{\mathbf{S}}$ is a square matrix of size n with an entry at the ij^{th} position equals to $\tilde{s}_{i,i}\tilde{s}_{j,j} + 2\tilde{s}_{i,j}^2$.

Next, we present the following theorem, and its proof is moved to the appendix.

Theorem 2.2. *Let $\hat{\boldsymbol{\theta}} = (\hat{a}, \hat{b}, \hat{R})^T$ be the geometric circle fit, and denote its quadratic approximation by $\hat{\boldsymbol{\theta}}_{\mathrm{Q}}$, then $\hat{\boldsymbol{\theta}}_{\mathrm{Q}} = \tilde{\boldsymbol{\theta}} + \tilde{\mathbf{K}}\mathbf{f}_2$, where the components of the $n-$dimensional vector $\mathbf{f}_2$ is given by*

$$f_{2,i} = \tfrac{1}{2\tilde{R}}\big(\tilde{\mathbf{t}}_i^T\check{\mathbf{n}}_i - \tilde{\mathbf{t}}_i^T\Delta_1\hat{\boldsymbol{\theta}}\big)^2 = \frac{1}{2\tilde{R}}(\rho_i - \tau_i)^2.$$

Moreover,

$$\mathrm{bias}(\hat{\boldsymbol{\theta}}_{\mathrm{Q}}) = \mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{2\tilde{R}}\big(\hat{\mathbf{e}}_3 + \tilde{\mathbf{K}}\tilde{\mathbf{s}}\big),$$

where $\hat{\mathbf{e}}_3 = (0, 0, 1)^T$.

In other words, the estimators of the circle's center, $\hat{a}$ and $\hat{b}$, have no essential bias, while the estimator of the radius $\hat{R}$ has essential bias, which is independent on n and the true points. In fact, we derived in [1] the bias of other algebraic fits as well. Then, we compared between fits based on their essential biases. Overall, a better estimator should have smaller essential bias. This strategy allowed us to compare between fits, and we showed why and by how much each estimator performs better than others. This explains a poor performance of the Kåsa fit, a moderate performance of the Pratt's fit, and a good performance of the Taubin fit and the geometric fit (in this order). Accordingly, we developed a new non-iterative algebraic fit, called hyperaccurate fit (HF) or hyper for short, which eliminates the essential bias. Therefore, HF was a doubly optimal fit in the sense: its variance attains KCR bound and has zero essential bias.

On the other hand, the 'nonessential bias', which is of order $\mathcal{O}(\sigma^2/n)$, was ignored in our study. Kanatani and Rangarajan [23] argued about the importance of the nonessential bias in such a case that n is relatively small (say n varies from 20 to 50). This means that terms of orders $\mathcal{O}(\sigma^2/n)$ is not negligible anymore. Therefore, they developed a new (efficient) algebraic fit (called hyper Least Square, or 'HyperLS' for short) that has zero bias up to $\mathcal{O}(\sigma^4)$.

Consequently, these two fits are supposed to perform much better than MLE (which has nonzero essential bias), especially that all of them are optimal in the sense that their covariance matrices attain KCR. Practically, however, the situation is still different. We observed that MLE is practically unbeatable in some cases, while hyper or HyperLs are unbeatable in other cases. Therefore, we go one step further by studying more terms in their MSE.

3. Third-order error analysis

It is quite interesting to find the MSE of geometric circle fit and other algebraic fits to the higher orders with the aid of their Taylor expansions up to the order $\mathcal{O}_{\mathrm{P}}(\sigma^4)$.

3.1. The error analysis of the geometric fit

Starting with the geometric fit, one can write $\hat{\boldsymbol{\theta}}$ as $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\mathrm{C}} + \mathcal{O}_{\mathrm{P}}(\sigma^4)$, where its cubic approximation is decomposed into

$$\hat{\boldsymbol{\theta}}_{\mathrm{C}} = \tilde{\boldsymbol{\theta}} + \Delta_1\hat{\boldsymbol{\theta}} + \Delta_2\hat{\boldsymbol{\theta}} + \Delta_3\hat{\boldsymbol{\theta}},$$

and as such,

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}_{\mathrm{C}}) = \mathbb{E}(\Delta_1\hat{\boldsymbol{\theta}}\Delta_1\hat{\boldsymbol{\theta}}^T) + \mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}\Delta_2\hat{\boldsymbol{\theta}}^T) + 2\mathcal{S}(\mathbb{E}(\Delta_3\hat{\boldsymbol{\theta}}\Delta_1\hat{\boldsymbol{\theta}}^T)) + \mathbb{E}(\Delta_3\hat{\boldsymbol{\theta}}\Delta_3\hat{\boldsymbol{\theta}}^T).$$

Including all terms of order σ^4, however, is difficult. Instead, we restrict our analysis to compute all terms of order σ^4 and σ^4/n, while other less important terms with order of magnitudes σ^4/n^2 and σ^6 will be discarded in our analysis. This means that

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}}) = \mathrm{MSE}(\Delta_1\hat{\boldsymbol{\theta}}) + \mathrm{MSE}(\Delta_2\hat{\boldsymbol{\theta}}) + 2\mathcal{S}\Big(\mathrm{cov}(\Delta_1\hat{\boldsymbol{\theta}}, \Delta_3\hat{\boldsymbol{\theta}})\Big) + \mathcal{O}(\sigma^4/n^2). \quad (3.1)$$

In the following theorem, we state the formal expression of the MSE of $\Delta_2\hat{\boldsymbol{\theta}}$ and $\mathrm{cov}(\Delta_3\hat{\boldsymbol{\theta}}, \Delta_1\hat{\boldsymbol{\theta}})$. The theorem also characterizes the terms of $\mathrm{MSE}(\Delta_2\hat{\boldsymbol{\theta}})$ according to their origins: the bias $\mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}})$ and the covariance matrix $\mathrm{var}(\Delta_2\hat{\boldsymbol{\theta}})$. Its derivation is quite lengthy, so it is moved to the appendix.

Theorem 3.1. *The MSE of* $\Delta_2\hat{\boldsymbol{\theta}}$ *is given by*

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}\Delta_2\hat{\boldsymbol{\theta}}^T) = \frac{\sigma^4}{4\tilde{R}^2}\Big(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2\tilde{\mathbf{V}}_{\min} + 2\mathcal{S}(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T)\Big) + \mathcal{O}(\sigma^4/n^2). \quad (3.2)$$

The terms in Eq. (3.2) come from different origins:

$$\mathrm{bias}(\Delta_2\hat{\boldsymbol{\theta}})\,\mathrm{bias}(\Delta_2\hat{\boldsymbol{\theta}})^T = \frac{\sigma^4}{4\tilde{R}^2}\Big(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2\mathcal{S}(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T)\Big) + \mathcal{O}(\sigma^4/n^2), \tag{3.3}$$

$$\mathrm{var}(\Delta_2\hat{\boldsymbol{\theta}}) = \frac{\sigma^4}{2\tilde{R}^2}\tilde{\mathbf{V}}_{\min} + \mathcal{O}(\sigma^4/n^2). \tag{3.4}$$

Moreover,

$$\mathrm{cov}(\Delta_1\hat{\boldsymbol{\theta}}, \Delta_3\hat{\boldsymbol{\theta}}) = \mathbb{E}\big(\Delta_3\hat{\boldsymbol{\theta}}\Delta_1\hat{\boldsymbol{\theta}}^T\big) = -\frac{\sigma^4}{2\tilde{R}^2}\tilde{\mathbf{V}}_{\min} + \mathcal{O}(\sigma^4/n^2). \tag{3.5}$$

Final formula for the mean squared error. Combining the resulting formulas obtained in Theorems 2.1, 2.2, and 3.1 yields

$$\mathbb{E}(\Delta\hat{\boldsymbol{\theta}}\Delta\hat{\boldsymbol{\theta}}^T) = \sigma^2\big(1 - \frac{\sigma^4}{2\tilde{R}^2}\big)\tilde{\mathbf{V}}_{\min} + \frac{\sigma^4}{4\tilde{R}^2}\Big(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2\mathcal{S}(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T)\Big) + \mathcal{O}(\sigma^4/n^2), \tag{3.6}$$

where $\tilde{\mathbf{V}}_{\min} = (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}$. The MSE can be decomposed into:

$$\mathrm{bias}(\Delta\hat{\boldsymbol{\theta}})\,\mathrm{bias}(\Delta\hat{\boldsymbol{\theta}})^T = \frac{\sigma^4}{4\tilde{R}^2}\big(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2\mathcal{S}\big(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T\big)\big) + \mathcal{O}(\sigma^4/n^2) \tag{3.7}$$

$$\mathrm{cov}(\Delta\hat{\boldsymbol{\theta}}) = \sigma^2\big(1 - \frac{\sigma^2}{2\tilde{R}^2}\big)\tilde{\mathbf{V}}_{\min} + \mathcal{O}(\sigma^4/n^2). \tag{3.8}$$

Remarkably, $\mathrm{cov}(\Delta_1\hat{\boldsymbol{\theta}}, \Delta_3\hat{\boldsymbol{\theta}}) = -\mathrm{Var}(\Delta_2\hat{\boldsymbol{\theta}})$, and as a result, *the second-order term* of the covariance of $\hat{\boldsymbol{\theta}}$ is negative. Besides, the bias of the center $\hat{\mathbf{c}} = (\hat{a}, \hat{b})$ does not contribute to MSE because $\mathbf{b}(\hat{\mathbf{c}})\mathbf{b}(\hat{\mathbf{c}})^T = \mathbf{0}_{2\times 2} + \mathcal{O}(\sigma^4/n^2)$. On other hand, the bias of $\hat{R}$ is significant, and it contributes negatively to the MSE in the sense that it increases MSE.

To understand these results, we will consider two special configurations. We positioned n equally spaced true points on (1) full circle and (2) semi-circle. Since the geometric fit is invariant under translations, rotations, and scaling, it is enough to set $\tilde{a} = \tilde{b} = 0$ and $\tilde{R} = 1$. Thus, $\tilde{u}_i = \tilde{x}_i$ and $\tilde{v}_i = \tilde{y}_i$. In sequel analysis we will use the standard statistical notations

$$S_{\tilde{x}\tilde{x}} = \sum_{i=1}^n \tilde{x}_i^2 - n\bar{\tilde{x}}^2, \quad S_{\tilde{y}\tilde{y}} = \sum_{i=1}^n \tilde{y}_i^2 - n\bar{\tilde{y}}^2, \quad S_{\tilde{x}\tilde{y}} = \sum_{i=1}^n \tilde{x}_i\tilde{y}_i - n\bar{\tilde{x}}\bar{\tilde{y}}.$$

Full circle. In this case, $\bar{\tilde{x}} = \bar{\tilde{y}} = \overline{\tilde{x}\tilde{y}} = 0$, so

$$S_{\tilde{x}\tilde{x}} = S_{\tilde{y}\tilde{y}} = \sum_{i=1}^n \tilde{x}_i^2 = \sum_{i=1}^n \cos^2\phi_i = \frac{n}{2},$$

and as such,

$$\tilde{\mathbf{V}}_{\min} = \mathrm{Diag}\big(\frac{1}{S_{\tilde{x}\tilde{x}}}, \frac{1}{S_{\tilde{y}\tilde{y}}}, \frac{1}{n}\big) = \frac{1}{n}\mathrm{Diag}(2,2,1),\ \tilde{\mathbf{s}} = \frac{2}{n}\mathbf{1}_n,\ \tilde{\mathbf{K}}\tilde{\mathbf{s}} = \frac{2}{n}\hat{\mathbf{e}}_3.$$

Then, up to the second leading term, we have

$$\mathrm{cov}(\Delta\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{n}(1-\frac{\sigma^2}{2})\mathrm{Diag}(2,2,1), \quad \mathrm{bias}(\Delta\hat{\boldsymbol{\theta}})\,\mathrm{bias}(\Delta\hat{\boldsymbol{\theta}})^T = \frac{\sigma^4}{4}\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T\big(1+\frac{4}{n}\big).$$

It is clear that both essential and nonessential biases of the center $\hat{\mathbf{c}} = (\hat{a}, \hat{b})^T$ have no contributions to the MSE. The second leading term of the MSE of $\hat{R}$ comes from the essential bias ($= \sigma^4/4$) and nonessential bias ($= \sigma^4/n$) and $\mathrm{var}(\hat{R}) = -\frac{\sigma^4}{2n}$. This shows that $\mathbf{b}_1(\hat{R})^2$ is much more important than other terms of order σ^4/n.

Semi-circle. In this case, we have $\bar{\tilde{x}} = \overline{\tilde{x}\tilde{y}} = 0$. Thus, $S_{\tilde{x}\tilde{x}} = \sum_{i=1}^n \tilde{x}_i^2 = \frac{n+1}{2}$ and $\sum_{i=1}^n \tilde{y}_i^2 = \frac{n-1}{2}$. Using Lagrange's trigonometric identities and small angle approximation gives

$$\sum_{i=1}^{n} \tilde{y}_i = \sum_{i=1}^{n-2} \sin i\omega = \frac{1}{2}\cot\frac{\omega}{2} - \frac{\cos(n-1/2)\omega}{2\sin\frac{\omega}{2}} = \cot\frac{\omega}{2} \sim \frac{2}{\omega} - \frac{1}{4}\omega,$$

where $\omega = \frac{\pi}{n-1}$. Thus $\bar{\tilde{y}} \sim \frac{2}{n\omega}$. Consequently, $S_{\tilde{y}\tilde{y}} = \frac{n-1}{2} - \frac{1}{n}\cot^2\frac{\omega}{2} \sim \frac{1}{2}(n-1)\kappa$, where $\kappa = 1 - \frac{8}{n\pi\omega} + \frac{2\omega}{n\pi}$. Thus, the contribution of the variance in the MSE is given by $\sigma^2(1-\frac{\sigma^2}{2})\tilde{\mathbf{V}}_{\min}$, where

$$\tilde{\mathbf{V}}_{\min} = \frac{1}{S_{\tilde{y}\tilde{y}}}\begin{bmatrix} \frac{S_{\tilde{y}\tilde{y}}}{S_{\tilde{x}\tilde{x}}} & 0 & 0 \\ 0 & 1 & -\bar{\tilde{y}} \\ 0 & -\bar{\tilde{y}} & \overline{\tilde{y}\tilde{y}} \end{bmatrix} = \frac{1}{S_{\tilde{y}\tilde{y}}}\begin{bmatrix} \frac{2S_{\tilde{y}\tilde{y}}}{n+1} & 0 & 0 \\ 0 & 1 & -\frac{1}{n}\cot\frac{\omega}{2} \\ 0 & -\frac{1}{n}\cot\frac{\omega}{2} & \frac{n-1}{2n} \end{bmatrix}$$
$$\sim \frac{1}{S_{\tilde{y}\tilde{y}}}\begin{bmatrix} \frac{2S_{\tilde{y}\tilde{y}}}{n+1} & 0 & 0 \\ 0 & 1 & -\frac{2}{\omega n} \\ 0 & -\frac{2}{\omega n} & \frac{n-1}{2n} \end{bmatrix}.$$

Next, we find how the bias contributes to MSE. First, note that

$$s_i = \frac{2\tilde{y}_i^2}{n+1} + \frac{\tilde{x}_i^2}{S_{\tilde{y}\tilde{y}}}, \quad \sum_{i=1}^{n}\tilde{y}_i^3 \sim \frac{4}{3\omega}, \quad \sum_{i=1}^{n}\tilde{x}_i^2\tilde{y}_i \sim \frac{2}{3\omega}.$$

Therefore,

$$\sum_{i=1}^{n} s_i \sim \frac{n-1}{n+1} + \frac{(n+1)n\omega^2}{n\pi\omega\kappa}, \quad \sum_{i=1}^{n} s_i y_i = \frac{8}{3\omega(n+1)} + \frac{4}{3\pi\kappa}.$$

Thus, $\tilde{\mathbf{W}}^T\mathbf{s} \sim (0, \sum_{i=1}^n s_i y_i, \sum_{i=1}^n s_i)^T$, and as such, the contribution of the bias in the MSE is $\mathbf{b}_1\mathbf{b}_1^T + 2\mathcal{S}(\mathbf{b}_1\mathbf{b}_2^T)$, which is equal to

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{1}{S_{\tilde{y}\tilde{y}}}(\sum_{i=1}^n s_i\tilde{y}_i - \frac{2}{n\omega}\sum_{i=1}^n s_i \\ 0 & \frac{1}{S_{\tilde{y}\tilde{y}}}(\sum_{i=1}^n s_i\tilde{y}_i - \frac{2}{n\omega}\sum_{i=1}^n s_i & 1 - \frac{4}{n\omega S_{\tilde{y}\tilde{y}}}\sum_{i=1}^n s_i\tilde{y}_i + \frac{n-1}{nS_{\tilde{y}\tilde{y}}}\sum_{i=1}^n s_i \end{bmatrix}.$$

Combining previous results gives us

$$\text{bias}(\hat{R})^2 = \frac{\sigma^4(1+\chi)}{4},$$
$$\text{var}(\hat{R}) = \frac{\sigma^2\left(1-\frac{\sigma^2}{2}\right)}{n - \frac{8}{\pi\omega} + \frac{2\omega}{\pi}} = \frac{\sigma^2\left(1-\frac{\sigma^2}{2}\right)(n-1)\pi^2}{(\pi^2-8)n^2 - (\pi^2-16)n + 2\pi^2 - 8},$$

where

$$\chi = \frac{2}{n\kappa}\left(\frac{-32}{3(n-1)(n+1)\omega^2} + \frac{n-1}{n+1} \right) + \frac{2}{n\pi\kappa^2}\left(\frac{-16}{3(n-1)\omega} + (n+1)\omega \right).$$

Thus, for large n

$$\text{bias}(\hat{R})^2 = \frac{\sigma^4}{4}\left(1 + \frac{(256 - 72\pi^2 + 9\pi^4)n^3}{3(-1+n^2)(8+n(-8+\pi^2))^2}\right), \quad \text{var}(\hat{R}) = \frac{\pi^2\sigma^2\left(1-\frac{\sigma^2}{2}\right)}{8+n(-8+\pi^2)}.$$

The ratio of $\mathbf{b}_1(\hat{R})\mathbf{b}_2(\hat{R})$ to $\text{var}(\hat{R})$ is approximately -2.1 to -2.5 for all values of n for not only long arcs but also for short arc as well. The ratio of $\mathbf{b}_1(\hat{R})^2$ to $\mathbf{b}_1(\hat{R})\mathbf{b}_2(\hat{R})$ is inversely proportional to n. This ratio varies from .5 to 1 when n is in the range 6–20. If $n \in (20, 50)$, the ratio will be between 1 to 2.2, and for $n \in (100, 200)$ the ratio lies between 2.2 and 8, etc. Therefore, it is clear that terms of order σ^4 will be more important than terms of order σ^4/n for moderately large values of n, say 150 or 200. The contribution of $\mathbf{b}_2(\hat{R})^2$ is quite small even for small values of n.

For shorter arcs, such as a quadrant of a circle, all of these three parts are important, while the discarded terms of order σ^4/n^2 are substantially small. This, in fact, supports our approach that keeps terms up to order σ^4/n and ignores the σ^4/n^2 terms in the MSE.

3.2. General error analysis of algebraic fits

Computing the MSE of algebraic fits is much more complicated than the error analysis of geometric fit. We will compute the MSE of algebraic fits, in a general context, up to order σ^4/n and discard all other terms of order σ^4/n^2, σ^6, etc. Namely,

$$\mathbb{E}(\Delta\mathbf{A}\Delta\mathbf{A}^T) = \text{MSE}(\Delta_1\mathbf{A}) + \mathbb{E}(\Delta_2\mathbf{A}\Delta_2\mathbf{A}^T) + 2\mathcal{S}\left[\mathbb{E}(\Delta_3\mathbf{A}\Delta_1\mathbf{A}^T)\right] + \mathcal{O}(\sigma^4/n^2).$$

Based on Eq. (2.19), we have

$$\tilde{\mathbf{M}}\,\Delta_2\mathbf{A} = -(\Delta_1\mathbf{M}\,\Delta_1\mathbf{A} + \Delta_2\mathbf{M}\,\tilde{\mathbf{A}}) + \Delta_2\lambda\tilde{\mathbf{N}}\tilde{\mathbf{A}}. \tag{3.9}$$

Next, we find $\Delta_2\lambda$. Premultiplying Eq. (3.9) and using Eq. (2.22) give us

$$\Delta_2\lambda = \frac{\tilde{\mathbf{A}}^T\mathbf{R}\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}} \text{ where } \mathbf{R} = \Delta_2\mathbf{M} - \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}. \tag{3.10}$$

The most important term of $\mathbf{R}$ is $\mathbf{R}_1 = \Delta_2\mathbf{M}$, which has a key role in our analysis, while $\mathbf{R}_2 = \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}$ has a less significant contribution to the

MSE. Accordingly, $\Delta_2\lambda$ can be decomposed into two components. Namely,

$$\Delta_2\lambda = \Delta_{2,1}\lambda - \Delta_{2,2}\lambda, \text{ where } \Delta_{2,k}\lambda = \frac{\tilde{\mathbf{A}}^T\mathbf{R}_k\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}, \quad k = 1, 2\,. \tag{3.11}$$

Based on Eqs. (3.9) and (3.10),

$$\Delta_2\mathbf{A} = \tilde{\mathbf{M}}^-\Big(\frac{\tilde{\mathbf{A}}^T\mathbf{R}\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} - \mathbf{R}\tilde{\mathbf{A}}\Big) = \tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\,\mathbf{R}\tilde{\mathbf{A}} \tag{3.12}$$

with

$$\mathbf{S}_{\tilde{\mathbf{N}}} = \frac{\tilde{\mathbf{N}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}} - \mathbf{I}_4. \tag{3.13}$$

Finally, our analysis needs $\Delta_3\mathbf{A}$. Combining all terms of order $\mathcal{O}_P(\sigma^3)$ in Eq. (2.19) gives

$$\begin{aligned}\Delta_3\mathbf{M}\tilde{\mathbf{A}} + \Delta_2\mathbf{M}\Delta_1\mathbf{A} + \Delta_1\mathbf{M}\Delta_2\mathbf{A} + \tilde{\mathbf{M}}\Delta_3\mathbf{A}&\\ = \Delta_3\lambda\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_2\lambda(\Delta_1\mathbf{N}\tilde{\mathbf{A}} + \tilde{\mathbf{N}}\Delta_1\mathbf{A}),&\end{aligned} \tag{3.14}$$

from which

$$\Delta_3\mathbf{A} = \tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\Big(\mathcal{J}_1 + \Delta_{2,1}\lambda\big(\mathcal{J}_2 - \Delta_1\mathbf{N}\tilde{\mathbf{A}}\big)\Big) + \mathcal{R}, \tag{3.15}$$

where $\Delta_3\mathbf{M} = \frac{2}{n}\sum_{i=1}^n \mathcal{S}(\Delta_2\mathbf{z}_i\Delta_1\mathbf{z}_i^T)$. Vectors $\mathcal{J}_1$ and $\mathcal{J}_2$ are defined by

$$\begin{aligned}\mathcal{J}_1 &= \Delta_3\mathbf{M}\tilde{\mathbf{A}} + \Delta_2\mathbf{M}\Delta_1\mathbf{A} - \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_2\mathbf{M}\tilde{\mathbf{A}}\\ &= \Big(\Delta_3\mathbf{M} - 2\mathcal{S}[\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_2\mathbf{M}]\Big)\tilde{\mathbf{A}},\\ \mathcal{J}_2 &= \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}} - \tilde{\mathbf{N}}\Delta_1\mathbf{A} = 2\mathcal{S}\big[\tilde{\mathbf{N}}\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}\big]\tilde{\mathbf{A}}.\end{aligned} \tag{3.16}$$

The complete analysis of the lengthy derivations of $\Delta_3\mathbf{A}$ is in the appendix. It is remarkable to note that we expressed $\Delta_3\mathbf{A}$ as $\Delta_3\mathbf{A} = \Delta_{3,1}\mathbf{A} + \mathcal{R}$, where the random vector $\mathcal{R}$ satisfies $\mathbb{E}(\mathcal{R}\Delta_1\mathbf{A}^T) \sim \mathcal{O}(\sigma^4/n^2)$, and as such, it will be dropped in our analysis. Next, two theorems summarize $\text{cov}(\Delta_3\mathbf{A}, \Delta_1\mathbf{A})$ and $\text{MSE}(\Delta_2\mathbf{A})$ up to order $\mathcal{O}(\sigma^4/n^2)$ (see their proofs in the appendix).

Theorem 3.2. *For each $i = 1, \ldots, n$, define*

$$\tilde{\alpha}_i = \tilde{\mathbf{a}}_i^T\tilde{\mathbf{A}} = 2\tilde{x}_i\tilde{A} + \tilde{B}, \quad \tilde{\beta}_i = \tilde{\mathbf{b}}_i^T\tilde{\mathbf{A}} = 2\tilde{y}_i\tilde{A} + \tilde{C}, \tag{3.17}$$

and

$$\tilde{\mathbf{Y}}_j = \begin{bmatrix} 8\tilde{x}_j & 2 & 0 & 0\\ 2 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix} \tag{3.18}$$

and

$$\tilde{\mathbf{P}}_j = \begin{bmatrix} 8\tilde{y}_j & 0 & 2 & 0\\ 0 & 0 & 0 & 0\\ 2 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}. \tag{3.19}$$

Also, define $\tilde{\mathbf{Q}}_i = \tilde{\alpha}_j \tilde{\mathbf{Y}}_i + \tilde{\beta}_i \tilde{\mathbf{P}}_i$ *and* $\tilde{\mathbf{\Pi}}_i = \tilde{\mathbf{z}}_i \check{\mathbf{e}}_1^T$*, where* $\check{\mathbf{e}}_1^T = (1,0,0,0)$*. Then,*

$$\operatorname{cov}(\Delta_1 \mathbf{A}, \Delta_3 \mathbf{A}) = \mathbb{E}(\Delta_3 \mathbf{A} \Delta_1 \mathbf{A}^T) = \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 + \mathbf{\Lambda}_3 + \mathbf{\Lambda}_4 + \mathcal{O}(\sigma^4/n^2),$$

where

$$\begin{aligned}
\mathbf{\Lambda}_1 &= -\frac{4\sigma^4}{n^2}\tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \sum_{i=1}^n \Big((\tilde{\mathbf{V}}_i \tilde{\mathbf{A}})(\tilde{\mathbf{\Pi}}_i \tilde{\mathbf{A}})^T + \tilde{\zeta} \tilde{\mathbf{\Pi}}_i^T \Big) \tilde{\mathbf{M}}^-,\\
\mathbf{\Lambda}_2 &= \frac{\sigma^4}{n^3} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \sum_{i \neq j} \Big(2\mathcal{S}\big[\mathbf{V}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T\big] \tilde{\mathbf{M}}^- \big(\mathbf{V}_i \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T + 2\tilde{A} \tilde{\mathbf{M}}_{ij} \big)\\
&\quad + \tilde{\zeta}\big((\tilde{\mathbf{V}}_i + 4\mathcal{S}[\tilde{\mathbf{\Pi}}_i]) \tilde{\mathbf{M}}^- \tilde{\mathbf{M}}_j \big) \Big) \tilde{\mathbf{M}}^-,\\
\mathbf{\Lambda}_3 &= -\frac{(n-1)\sigma^4 \tilde{\zeta}^2 \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \tilde{\mathbf{N}} \tilde{\mathbf{M}}^-}{n^2 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})}\\
&\quad - \frac{2\sigma^4 \tilde{\zeta} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}}}{n^3 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \sum_{i \neq j} \big[\mathcal{S}(\tilde{\mathbf{V}}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T) \tilde{\mathbf{M}}^- \tilde{\mathbf{N}} \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T \tilde{\mathbf{M}}^- \big].
\end{aligned}$$

Moreover, $\mathbf{\Lambda}_4 = \mathbf{0}$ *for the Kåsa and the Pratt fits. On the other hand,*

$$\begin{aligned}
\mathbf{\Lambda}_{4,\mathrm{T}} &= \frac{(n-1)\sigma^4}{n^3} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}_{\mathrm{T}}} \Big(\sum_{j=1}^n \tilde{\mathbf{Q}}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T \Big) \tilde{\mathbf{M}}^-,\\
\mathbf{\Lambda}_{4,\mathrm{H}} &= \frac{2(n-1)\sigma^4}{n^3} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}_{\mathrm{H}}} \Big(\sum_{j=1}^n \tilde{\mathbf{Q}}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T \Big) \tilde{\mathbf{M}}^-,\\
\mathbf{\Lambda}_{4,\mathrm{I}} &= \frac{(n-1)\sigma^4}{n^3\big(1 - \frac{3}{n^2}\big)} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}_{\mathrm{I}}} \sum_{j=1}^n \Big[2\tilde{\mathbf{Q}}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T - \frac{\tilde{\zeta} \operatorname{trace}(\tilde{\mathbf{M}}^- \tilde{\mathbf{V}}_j)}{n} \tilde{\mathbf{M}}_j \Big] \tilde{\mathbf{M}}^-,
\end{aligned} \tag{3.20}$$

for Taubin's, Hyper, and HyperLS fits, respectively.

Theorem 3.3. *For each* $i = 1, \ldots, n$*, define the scalars* $\tilde{\psi}_i = \tilde{\mathbf{z}}_i^T \tilde{\mathbf{M}}^- \tilde{\mathbf{z}}_i$ *and* $\tilde{\eta}_i = \tilde{\mathbf{A}}^T \tilde{\mathbf{V}}_i \tilde{\mathbf{M}}^- \tilde{\mathbf{z}}_i$ *and the vector* $\tilde{\Psi}_i = 2\tilde{\mathbf{\Pi}}_i + \tilde{\mathbf{V}}_i$*. Then*

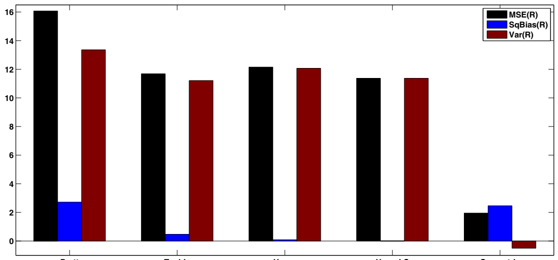
$$\mathbb{E}(\Delta_2 \mathbf{A} \Delta_2 \mathbf{A}^T) = \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \mathbf{\Phi} \mathbf{S}_{\tilde{\mathbf{N}}}^T \tilde{\mathbf{M}}^-,$$

where

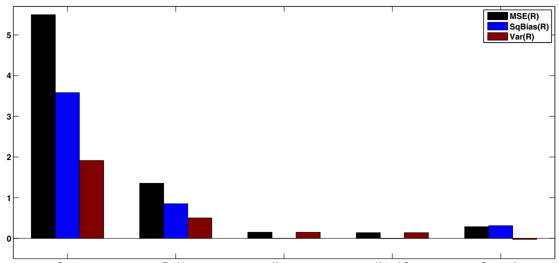
$$\begin{aligned}
\mathbf{\Phi} &= \frac{\sigma^4}{n^2} \sum_{i=1}^n \Big((\tilde{\Psi}_i \tilde{\mathbf{A}})(\tilde{\Psi}_i \tilde{\mathbf{A}})^T + \tilde{\zeta} \tilde{\mathbf{V}}_i + \sum_{j=1}^n (\tilde{\Psi}_i \tilde{\mathbf{A}})(\tilde{\Psi}_j \tilde{\mathbf{A}})^T \Big)\\
&\quad - \frac{\sigma^4}{n^3} \sum_{i \neq j} 2\mathcal{S}\big((\tilde{\Psi}_i \tilde{\mathbf{A}}) \big(\tilde{\eta}_j \tilde{\mathbf{z}}_j + \tilde{\psi}_j \tilde{\mathbf{V}}_j \tilde{\mathbf{A}} \big)^T \big).
\end{aligned} \tag{3.21}$$

Moreover, $\operatorname{cov}(\Delta_2 \mathbf{A}) = \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \mathbf{\Phi}_2 \mathbf{S}_{\tilde{\mathbf{N}}}^T \tilde{\mathbf{M}}^-$*, where*

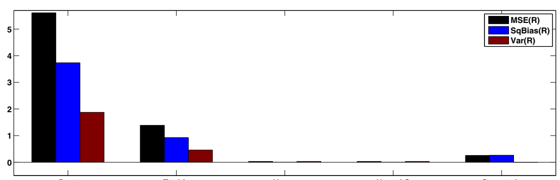
$$\mathbf{\Phi}_2 = \frac{\sigma^4}{n^2} \sum_{i=1}^n \big((\tilde{\Psi}_i \tilde{\mathbf{A}})(\tilde{\Psi}_i \tilde{\mathbf{A}})^T + \tilde{\zeta} \tilde{\mathbf{V}}_i \big) \tag{3.22}$$



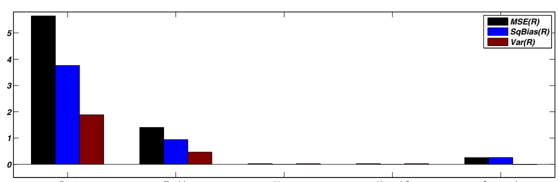
(a) Quadrant of a circle



(b) Semi-circle



(c) Three quadrants of a circle



(d) Full circle

FIG 1. *The theoretical fourth order term of MSE for five fits: Pratt's fit, Taubin's fit, hyper fit, HyperLS fit, and geometric fit respectively. The MSE decomposed into its origin: Variance (red) and the biased squared (blue).*

and

$$\mathrm{Bias}(\Delta_2\mathbf{A}) = \frac{\sigma^2}{n}\tilde{\mathbf{M}}^{-}\mathbf{S}_{\tilde{\mathbf{N}}}\sum_{i=1}^{n}\Big[\big(\tilde{\mathbf{\Psi}}_i - \frac{\tilde{\psi}_i}{n}\tilde{\mathbf{V}}_i\big)\tilde{\mathbf{A}} - \frac{\tilde{\eta}_i}{n}\tilde{\mathbf{z}}_i\Big].$$

Since hyper and HyperLS fits have zero bias up to order σ^2/n and σ^4, respectively, their MSEs are only controlled by $\mathrm{cov}(\Delta\hat{\boldsymbol{\theta}}^a) = \tilde{\mathbf{J}}\mathrm{cov}(\Delta\mathbf{A})\tilde{\mathbf{J}}^T$, which turn to be *roughly* equal as Figure 1 shows.

Figure 1 exhibits the contribution of the second order terms of the bias and the variance of the estimate of R when $n = 100$ true points are distributed along circular arcs of sizes: 90° (Fig. 1(a)), 180° (Fig. 1(b)), 270° (Fig. 1(c)), and 360° (Fig. 1(d)). The fourth order term of MSE(black-colored bar) were compared for five fits: Pratt's fit, Taubin's fit, hyper fit, HyperLS fit, and the geometric fit, respectively. These terms are decomposed into their origins, the bias (blue-colored bars) and the second terms of the variance (red-colored).

When true points are distributed along a long circular arc, such as a semi-circle (Figures 1(d)), three quadrants of a circle, and a full circle, the covariance of $\hat{R}$ for all fits are quite smaller than their squared biases. Thus the dominant part in the MSE is the bias. Therefore, the mean square error of HyperLS fit is smaller than that of hyper fit, which is smaller than geometric fit, and so one. This validates our conclusions in [1, 2].

The short circular arcs, such as a quadrant of a circle (Fig. 1(a)), reveals interesting results. The covariance of $\hat{R}$ for each algebraic fit becomes quite bigger than the bias, and as such, it becomes the dominant part in the MSE for each algebraic fit. On other hand, the covariance of $\hat{R}$ obtained by the geometric fit is much smaller than that for algebraic fits (if fact it is negative and close to 0). Therefore, in total the geometric fit is the best in this case. Theen HyperLS has the second smallest MSE, and as such it is the second most accurate fit. Then hyper fit is the third most accurate fit, followed by Taubin's fit and Pratt's fit (in this order).

4. Adjusted maximum likelihood estimator

The MLE itself has bias up to the leading term. Thus, one can obtain a better estimate based on the MLE by using 'unbiasing' schemes, and as such, the bias (up to the second leading term) vanishes. The new estimator has the formal expression

$$\hat{\mathbf{\Gamma}} = \hat{\boldsymbol{\theta}} - \frac{\hat{\sigma}^2}{2\hat{R}}\big[\hat{\mathbf{e}}_3 + (\hat{\mathbf{W}}^T\hat{\mathbf{W}})^{-1}\hat{\mathbf{W}}^T\hat{\mathbf{s}}\big], \tag{4.1}$$

where $\hat{\mathbf{s}} = (\hat{s}_{1,1}, \hat{s}_{2,2} \dots, \hat{s}_{n,n})^T$. Here $\hat{s}_{i,i}$ is the random estimate of $\tilde{s}_{i,i}$, i.e.

$$\hat{s}_{i,i} = [-\hat{v}_i, \hat{u}_i, 0](\hat{\mathbf{W}}^T\hat{\mathbf{W}})^{-1}[-\hat{v}_i, \hat{u}_i, 0]^T,$$

$\hat{u}_i$, $\hat{v}_i$, and $\hat{\mathbf{W}}$ are regarded as estimates of $\tilde{u}_i, \tilde{v}_i$, and $\tilde{\mathbf{W}}$, which are calculated after computing the MLE. i.e.,

$$\hat{u}_i = \frac{x_i - \hat{a}}{\hat{R}}, \quad \hat{v}_i = \frac{y_i - \hat{a}}{\hat{R}},$$

and

$$\hat{\mathbf{W}} = \begin{bmatrix} \hat{u}_1 & \hat{v}_1 & 1 \\ \vdots & \vdots & \vdots \\ \hat{u}_n & \hat{v}_n & 1 \end{bmatrix}.$$

Also, we estimate σ^2 by

$$\hat{\sigma}^2 = \frac{1}{n-3}\sum_{i=1}^{n}\Big(\sqrt{(x_i-\hat{a})^2+(y_i-\hat{b})^2}-\hat{R}\Big)^2. \tag{4.2}$$

Next, we need to show that the quadratic approximation of $\hat{\mathbf{\Gamma}}$, say $\hat{\mathbf{\Gamma}}_{\mathrm{Q}}$, is an unbiased estimator of $\tilde{\boldsymbol{\theta}}$ (up to the leading term). This follows immediately if we verify that $\mathbb{E}(\hat{\sigma}^2)=\sigma^2$ (up to the leading term), and as such, $\mathbb{E}(\hat{\mathbf{\Gamma}}_Q)=\tilde{\boldsymbol{\theta}}+\mathcal{O}(\sigma^4)$. Indeed, let us define $\tilde{\mathbf{P}}=\tilde{\mathbf{W}}(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^T$, which acts as a projection matrix onto the column space of $\tilde{\mathbf{W}}$, and as such, $\mathbf{I}-\tilde{\mathbf{P}}$ is another projection matrix. Thus, both of them are idempotent and symmetric matrices. Since

$$\hat{\sigma}^2 = \frac{1}{n-3}\|\tilde{\mathbf{W}}\hat{\boldsymbol{\theta}}-(\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon})\|_2^2 = \frac{1}{n-3}\|(\mathbf{I}-\tilde{\mathbf{P}})(\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon})\|_2^2+\mathcal{O}_{\mathrm{P}}(\sigma^4)$$

and $\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon}\sim\mathbb{N}(\mathbf{0},\sigma^2\mathbf{I}_n)$, then

$$\frac{1}{\sigma^2}\|(\mathbf{I}-\tilde{\mathbf{P}})(\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon})\|_2^2$$

has a chi-square distribution with a degree of freedom $\mathrm{df}=\mathrm{tr}(\mathbf{I}_n-\tilde{\mathbf{P}})=n-3$. Thus,

$$\mathbb{E}\big(\hat{\sigma}^2\big)=\sigma^2+\mathcal{O}(\sigma^4).$$

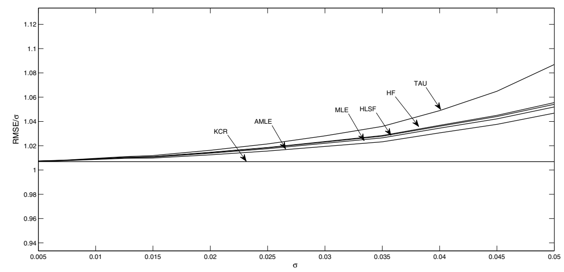
Therefore, the new estimator is an unbiased estimator of $\boldsymbol{\theta}$. Because this technique is an adjustment version of geometric fit, we call this estimator 'Adjusted Maximum Likelihood Estimator' (AMLE).

5. Numerical experiments

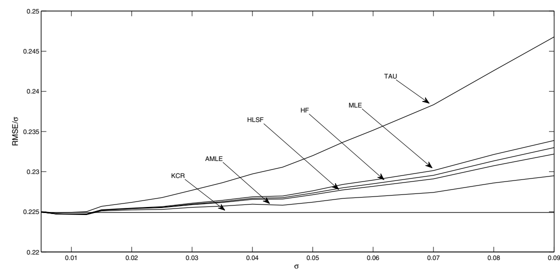
To demonstrate the superiority of AMLE, we turn our attention to some numerical experiments. We firstly positioned 100 equally spaced true points on an arc of size: (1) 90° (Fig. 2(a), (2) 180° (Fig. 2(b)), (3) 270° (Fig. 2(c)), and (4) 360° (Fig. 2(d)). For each case, we generated N random samples by adding Gaussian noise at level σ to each true point. For each sample, we applied various circle fits to estimate the parameters $(a; b; R)$. Then the Root Means Square Error (RMSE) for each fit is measured by computing

$$\mathrm{RMSE}(\hat{R})=\mathbb{E}(\hat{R}^2)\approx\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{R}-1)^2}, \quad \text{where } N=10^6. \tag{5.1}$$

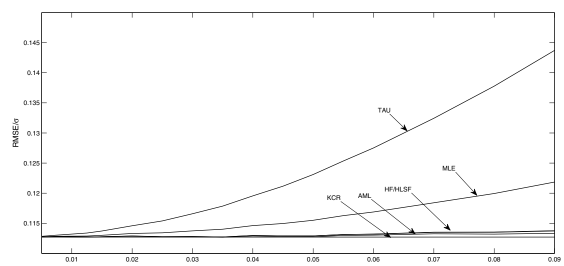
Since RMSE is asymptotically proportional to σ, we plotted the ratio RMSE $/\sigma$ against the noise level σ which varies from 0 up to the point at which the value of RMSE becomes large. We compared *five fits*: AMLE, MLE, hyper fit (HF), HyperLS fit (HLSF), and Taubin's fit (Tau). These fits depend on (i) the true values of the circle parameters, (ii) the number and location of the true points,



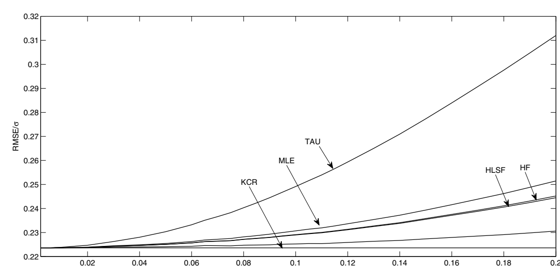
(a) Quadrant of a circle



(b) Semi-circle



(c) Three quadrants of a circle



(d) Full circle

Fig 2. *The simulated RMSE-to-noise ratio for five fits when* 100 *observations are distributed along arcs of lengths: 90° (Fig. 2(a)), 180° (Fig. 2(b)), 270° (Fig. 2(c)), and 360° (Fig.2(d)).*

and (iii) the noise level σ. Since Taubin's fit, hyper fit, and the geometric fit (and as such AMLE) are invariant under translations, rotations, and scaling, it is enough to set $\tilde{a} = \tilde{b} = 0$ and $\tilde{R} = 1$.

As can be seen in Figures 2(a)–2(d), as σ takes small values, all estimators approach KCR, which is the first leading term in MSE. As σ increases and reaches its typical values, AMLE exhibits a superior performance over other fits in all cases. The comparisons of other fits however depend on the circular arc. For long arcs, the second most accurate fits are HyperLS and hyper fits which perform better than MLE, followed by Taubin's fit at last (see Figs. 2(b)–2(d)). For short arcs, the geometric fit, however performs better than all other algebraic fits. This supports our theoretical results in this article.

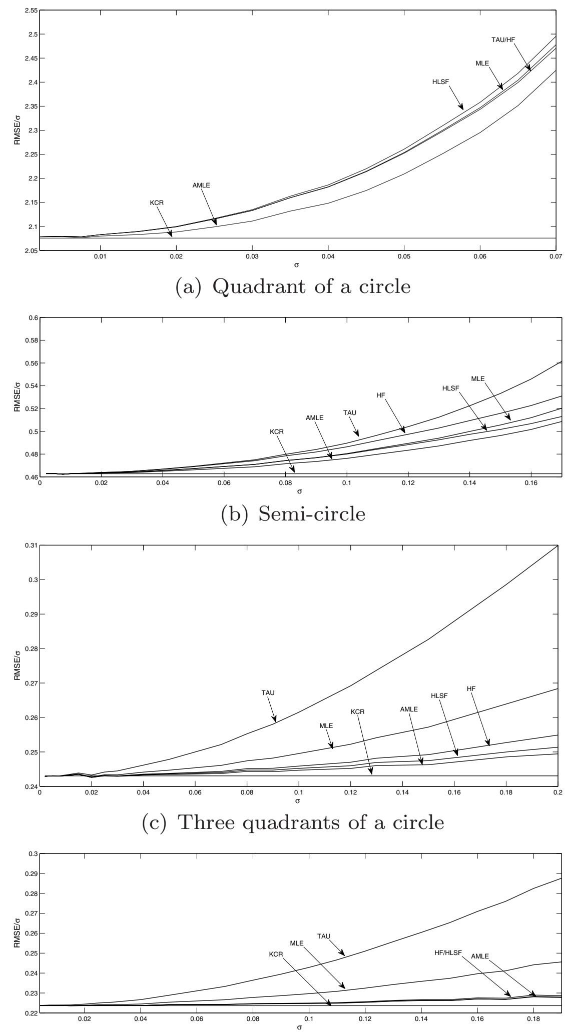
In the second experimental setting, 20 points are positioned on the four circular arcs, as set previously (see Figs. 3(a)–3(d)). The results confirm the superiority of AMLE over all other fits. Also, the same conclusions have been reached regarding the performance of other fits when data are sampled along long arcs (Figs. 3(b)–3(d)). That is, HyperLS is more accurate than hyper which is more accurate MLE and Taubins' fit (in this order). However, for shorter arcs (Figure 3(a)) the numerical results reveals an interesting observation. While the Taubin's fit and hyper fit have the same performance and both work better than MLE, HyperLS becomes unpredictability the least accurate fits. This turnaround in their behaviors for small values of n and short circular arc is left as future work.

6. Conclusion

In [1], we adopted an unconventional approach to study circle fittings by deriving the MSEs for the most important fits: algebraic fits (Kåsa's, Pratt's, Taubin's) and the geometric fit. We kept all terms of order σ^4, while other terms were discarded. We showed that the MSEs for all fits have a common leading term that coincides with KCR. Thus, we traced the second leading term of their MSE (partially) by tracking only terms of order σ^4 and discarding the others. This term comes from the essential bias. Based on that, we showed why and by how much the geometric fit performs betters than the Taubin fit, which is twice more accurate than Pratt's fit. We also developed the new algebraic fit (hyper fit) that has zero essential bias. Kanatani and Rangarajan immediately developed a new fit that removes both of the essential and the nonessential biases.

However the geometric fit still practically works better than the hyper fits in some cases. Apparently, other terms in the MSE have key roles on this aspect. We investigated this issue in this paper by deriving their MSE up to order σ^4/n and discarding σ^4/n^2. Terms of order σ^4/n come from the variance, and the product of the nonessential bias and the essential bias.

Our analysis shows that when data are sampled along short circular arcs, the second-order covariance parts in all algebraic fits become much larger than other parts in their second-order term in MSEs' expressions. The second-order covariance part for the geometric fit is negative though it is close to zero. This shows why the geometric fit is superior to all other algebraic fits when data are sampled along short arcs. On the other hand, if data are sampled along long circular arcs, the bias becomes the most dominant part among other second-order



(a) Quadrant of a circle

(b) Semi-circle

(c) Three quadrants of a circle

(d) Full circle

Fig 3. *The simulated RMSE-to-noise ratio for five fits when* 20 *observations are distributed along arcs of lengths: 90° (Fig. 3(a)), 180° (Fig. 3(b)), 270° (Fig. 3(c)), and 360° (Fig.3(d)).*

parts in the MSE, and as such hyperaccurate fit and HyperLS fit outperform the geometric fit which in turn outperforms Taubin's fit, followed by Pratt's fit (in this order). These conclusions are based on our assumptions that n is limited but fairly large, say $n > 70$. Special further investigations are needed in the case of small values of n and short circular arcs are chosen.

Our second contribution in this paper is proposing a bias-correction version of the geometric fit, which in turn, outperforms all existing methods in all cases whatever sample size n and the length of the circular arc are chosen. The new method has two features. Its variance is the smallest and has zero bias up to order σ^4. Our numerical tests confirm the superiority of the proposed fit over the existing fits.

Appendix

Proof of Theorem 2.1. The distance $d_i = r_i - R$ can be expanded as

$$\begin{aligned} d_i &= \sqrt{[(\tilde{x}_i+\delta_i)-(\tilde{a}+\Delta_1 a)]^2+[(\tilde{y}_i+\varepsilon_i)-(\tilde{b}+\Delta_1 b)]^2}-\tilde{R}-\Delta_1 R, \\ &= \sqrt{\tilde{R}^2+2\tilde{R}\tilde{u}_i(\delta_i-\Delta_1 a)+2\tilde{R}\tilde{v}_i(\varepsilon_i-\Delta_1 b)+\mathcal{O}_P(\sigma^2)}-\tilde{R}-\Delta_1 R \\ &= \tilde{u}_i(\delta_i-\Delta_1 a)+\tilde{v}_i(\varepsilon_i-\Delta_1 b)-\Delta_1 R+\mathcal{O}_P(\sigma^2). \end{aligned}$$

Minimizing $\sum d_i^2$ to the first order is equivalent to minimizing $\|\tilde{\mathbf{W}}\Delta_1\hat{\boldsymbol{\theta}}-\mathbf{f}_1\|_2^2$, where the vector $\mathbf{f}_1=(f_{1,1},\ldots,f_{1,n})^T$ is the vector whose components are $f_{1,i}=\tilde{u}_i\delta_i+\tilde{v}_i\varepsilon_i$. The previous problem is a classical least squares problem that can also be written as $\tilde{\mathbf{W}}\,\Delta_1\hat{\boldsymbol{\theta}}=\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon}$. If it is multiplied by $\tilde{\mathbf{K}}=(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^T$, then we get $\Delta_1\hat{\boldsymbol{\theta}}\approx\tilde{\mathbf{K}}\mathbf{f}_1$. Next, one can find the variance of $\hat{\boldsymbol{\theta}}$, to the leading order. i.e.

$$\mathbb{E}\big(\Delta_1\hat{\boldsymbol{\theta}}\Delta_1\hat{\boldsymbol{\theta}}^T\big)=\tilde{\mathbf{K}}\,\mathbb{E}\big[(\tilde{\mathbf{U}}\delta+\tilde{\mathbf{V}}\boldsymbol{\varepsilon})(\tilde{\mathbf{U}}\delta^T+\tilde{\mathbf{V}}\boldsymbol{\varepsilon}^T)\big]\tilde{\mathbf{K}}^T.$$

Now observe that $\mathbb{E}(\delta\boldsymbol{\varepsilon}^T)=\mathbb{E}(\boldsymbol{\varepsilon}\delta^T)=\mathbf{0}$, as well as $\mathbb{E}(\delta\delta^T)=\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)=\sigma^2\mathbf{I}_n$, and we have $\tilde{\mathbf{U}}^2+\tilde{\mathbf{V}}^2=\mathbf{I}_n$, where $\mathbf{I}_n$ is identity matrix of size n. Thus

$$\operatorname{cov}(\hat{\boldsymbol{\theta}}_L)=\mathbb{E}\big[(\Delta_1\hat{\boldsymbol{\theta}})(\Delta_1\hat{\boldsymbol{\theta}})^T\big]=\sigma^2\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T=\sigma^2(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}=\sigma^2\tilde{\mathbf{V}}_{\min},$$

□

Proof of Lemma 2.1. Using Theorem 2.1, $\mathbb{E}(\tau_i\tau_j)=\tilde{\mathbf{t}}_i^T\mathbb{E}(\Delta_1\boldsymbol{\theta}\Delta_1\boldsymbol{\theta}^T)\tilde{\mathbf{t}}_j=\sigma^2\tilde{s}_{i,j}$. Next, we compute $\mathbb{E}(\tau_i\rho_j)$. Recall the definition of $\Delta_1\boldsymbol{\theta}=\tilde{\mathbf{K}}\mathbf{f}_1$, which can be written as $(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\mathbf{g}$, where

$$\mathbf{g}=\Big(\sum_{k=1}^n\tilde{u}_kf_{1,k},\sum_{k=1}^n\tilde{v}_kf_{1,k},\sum_{k=1}^nf_{1,k}\Big)^T,\quad f_{1,k}=\tilde{u}_k\delta_k+\tilde{v}_k\varepsilon_k$$

It also is remarkable notice that for each j, $\mathbb{E}(f_{1,k}(\check{\mathbf{n}}_j^T\tilde{\mathbf{t}}_j))=0$ for all $k=1,\ldots,n$. Thus $\mathbb{E}(\mathbf{g}(\check{\mathbf{n}}_j^T\tilde{\mathbf{t}}_j))=\mathbf{0}$, and as such $\mathbb{E}(\tau_i\rho_j)=0$. Finally, we simply observe $\mathbb{E}(\rho_i\rho_j)=\tilde{\mathbf{t}}_i^T\mathbb{E}(\check{\mathbf{n}}_i\check{\mathbf{n}}_j^T)\tilde{\mathbf{t}}_j=\delta_{i,j}\sigma^2$ since we have $\|\tilde{\mathbf{t}}_i\|_2^2=\tilde{u}_i^2+\tilde{v}_i^2=1$. This completes the proof of the lemma. □

Proof of Lemma 2.2. The first three assertions follow immediately from Lemma 2.1. Next we compute $\mathbb{E}(\mathbf{a}\mathbf{a}^T)$, $\mathbb{E}(\mathbf{a}\mathbf{b}^T)$, and $\mathbb{E}(\mathbf{a}\mathbf{c}^T)$ starting with the former. Using Isserlis' Theorem, then for each $i,j=1,\ldots,n$, we get

$$\mathbb{E}(a_ia_j)=\mathbb{E}(\rho_i^2\rho_j^2)=\mathbb{E}(\rho_i^2)\mathbb{E}(\rho_j^2)+2(\mathbb{E}(\rho_i\rho_j))^2=\sigma^4(1+2\delta_{i,j}),$$

where δ_{ij} is the Kronecker Delta function. Therefore $\mathbb{E}(\mathbf{a}\mathbf{a}^T) = \sigma^4(2\mathbf{I}_n + \mathbf{1}_n)$. To prove $\mathbb{E}(\mathbf{b}\mathbf{a}^T) = \mathbb{E}(\mathbf{b}\mathbf{c}^T) = \mathbf{0}$, we use

$$\begin{aligned}\mathbb{E}(a_i b_j) &= -2\mathbb{E}(\rho_i^2\rho_j\tau_j) = -2\big(\mathbb{E}(\rho_i^2)\mathbb{E}(\rho_j\tau_j) + 2\mathbb{E}(\rho_i\rho_j)\mathbb{E}(\rho_i\tau_j)\big) = 0\\ \mathbb{E}(b_i c_j) &= \mathbb{E}(\rho_i\tau_i\tau_j^2) = \mathbb{E}(\rho_i\tau_i)\mathbb{E}(\tau_j^2) + 2\mathbb{E}(\rho_i\tau_j)\mathbb{E}(\tau_i\tau_j) = 0\end{aligned}$$

since $\mathbb{E}(\rho_i\tau_j) = 0$ for all i and j. For the same reason

$$\mathbb{E}(a_i c_j) = \mathbb{E}(\tau_i^2\rho_j^2) = \mathbb{E}(\tau_i^2)\mathbb{E}(\rho_j^2) + 2\mathbb{E}(\tau_i\rho_j)^2 = \sigma^4\tilde{s}_{j,j},$$

which depends only on j regardless of i, thus $\mathbb{E}(\mathbf{a}\mathbf{c}^T) = \sigma^4\mathbf{1}_n\mathbf{s}^T$. Similarly, for each $i, j = 1, \ldots, n$, we have

$$\mathbb{E}(b_i b_j) = 4\mathbb{E}(\tau_i\rho_i\tau_j\rho_j) = 4\mathbb{E}(\tau_i\tau_j)\mathbb{E}(\rho_i\rho_j) = 4\sigma^4\delta_{i,j}\tilde{s}_{i,j}.$$

This means that $\mathbb{E}(\mathbf{b}\mathbf{b}^T)$ is an diagonal matrix whose i^{th} diagonal entity equals to $4\sigma^4\tilde{s}_{i,i}$. To end up, we prove

$$\mathbb{E}(c_i c_j) = \mathbb{E}(\tau_i^2\tau_j^2) = \big(\mathbb{E}(\tau_i^2)\mathbb{E}(\tau_j^2) + 2(\mathbb{E}(\tau_i\tau_j))^2\big) = \sigma^4\big(\tilde{s}_{i,i}\tilde{s}_{j,j} + 2\tilde{s}_{i,j}^2\big).$$

That is, $\mathbb{E}(\mathbf{c}\mathbf{c}^T) = \sigma^4\tilde{\mathbf{S}}$, where $\tilde{\mathbf{S}}$ is a square matrix of size n whose entries at the ij^{th} position is $\tilde{s}_{i,i}\tilde{s}_{j,j} + 2\tilde{s}_{i,j}^2$. □

Proof of Theorem 2.2. Expanding the distance d_i to the second order terms gives

$$d_i = q_i - \tilde{u}_i\,\Delta_2 a - \tilde{v}_i\,\Delta_2 b - \Delta_2 R + \mathcal{O}_{\mathrm{P}}(\sigma^3)$$

$$\begin{aligned}q_i = \tilde{u}_i(\delta_i - \Delta_1 a) + \tilde{v}_i(\varepsilon_i - \Delta_1 b) - \Delta_1 R + \tfrac{\tilde{v}_i^2}{2\tilde{R}}(\delta_i - \Delta_1 a)^2\\ + \tfrac{\tilde{u}_i^2}{2\tilde{R}}(\varepsilon_i - \Delta_1 b)^2 - \tfrac{\tilde{u}_i\tilde{v}_i}{\tilde{R}}(\delta_i - \Delta_1 a)(\varepsilon_i - \Delta_1 b).\end{aligned} \tag{A.1}$$

Next, we calculate $\Delta_2\hat{\boldsymbol{\theta}} = (\Delta_2 a, \Delta_2 b, \Delta_2 R)^T$. From the definition of $\check{\mathbf{n}}_i = (\delta_i, \varepsilon_i, 0)^T$ and $\tilde{\mathbf{t}}_i = (-\tilde{v}_i, \tilde{u}_i, 0)^T$ and revoke Theorem 2.2, then we have $q_i = l_i + f_{2,i}$ with $l_i = f_{1,i} - (\tilde{u}_i\Delta_1 a + \tilde{v}_i\Delta_1 b + \Delta_1 R)$. Now, let us write $\mathbf{q} = \mathbf{l} + \mathbf{f}_2$, $\mathbf{f}_2 = (f_{2,1}, \ldots, f_{2,n})^T$, $\mathbf{l} = (l_1, \ldots, l_n)^T$, and $\mathbf{q} = (q_1, \ldots, q_n)^T$. Thus minimizing $\sum d_i^2$ is now equivalent to minimizing

$$\sum_{i=1}^{n}\big(q_i - (\tilde{u}_i\,\Delta_2 a + \tilde{v}_i\,\Delta_2 b + \Delta_2 R)\big)^2,$$

from which we get another least squares problem, and its solution is $\Delta_2\hat{\boldsymbol{\theta}} \approx \tilde{\mathbf{K}}\mathbf{q}$. In fact, the contribution from the linear terms, l_i, vanishes, quite predictably; thus only the quadratic terms $f_{2,i}$ matter. That is, since $\tilde{\mathbf{K}}\tilde{\mathbf{W}} = \mathbf{I}_3$, $\tilde{\mathbf{K}}\mathbf{l} = \tilde{\mathbf{K}}\mathbf{f}_1 - \tilde{\mathbf{K}}\tilde{\mathbf{W}}\Delta_1\hat{\boldsymbol{\theta}} = \mathbf{0}$, and as such $\Delta_2\hat{\boldsymbol{\theta}} \approx \tilde{\mathbf{K}}\mathbf{f}_2$.

Next, we find the bias of $\hat{\boldsymbol{\theta}}_{\mathrm{Q}}$. Since $\mathbb{E}(\Delta_1\hat{\boldsymbol{\theta}}) = \mathbf{0}$, we need only to compute $\tilde{\mathbf{K}}\mathbb{E}(\mathbf{f}_2)$, where $2\tilde{R}\mathbf{f}_2 = \mathbf{a} + \mathbf{b} + \mathbf{c}$. Therefore, $\mathbb{E}(\mathbf{f}_2) = \frac{1}{2\tilde{R}}\sigma^2(\mathbf{1}_n + \tilde{\mathbf{s}})$, which is an immediately consequence of Lemma 2.2. Since the last column of the matrix $\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}$ coincides with the vector $\tilde{\mathbf{W}}^T\mathbf{1}_n$, we have $\tilde{\mathbf{K}}\mathbf{1}_n = (0, 0, 1)^T = \hat{\mathbf{e}}_3$. Thus pre-multiply $\mathbb{E}(\mathbf{f}_2)$ by $\tilde{\mathbf{K}}$ gives the desired result. □

Proof of Theorem 3.1. Let us write $\mathbf{f}_2 = \frac{1}{2\tilde{R}}(\mathbf{a}+\mathbf{b}+\mathbf{c})$ and use the results of Lemma 2.2 to show

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}\Delta_2\hat{\boldsymbol{\theta}}^T) = (\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1}\tilde{\mathbf{W}}^T\mathbb{E}(\mathbf{f}_2\mathbf{f}_2^T)\tilde{\mathbf{W}}(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1} \tag{A.2}$$

$$= \frac{1}{4\tilde{R}^2}\tilde{\mathbf{K}}\,\mathbb{E}\big((\mathbf{a}+\mathbf{b}+\mathbf{c})(\mathbf{a}+\mathbf{b}+\mathbf{c})^T\big)\tilde{\mathbf{K}}^T \tag{A.3}$$

The expected values of $\mathbf{aa}^T, \mathbf{ab}^T, \ldots, \mathbf{cc}^T$ are already found in Lemma 2.1, but not all terms will be kept in the MSE, if we take in our consideration their significant contribution to the MSE. That is, only $\mathbb{E}(\mathbf{aa}^T + \mathbf{ac}^T + \mathbf{ca}^T)$ is of our interest as being the important part of MSE. Now

$$\tilde{\mathbf{K}}\mathbb{E}\big(\mathbf{aa}^T + \mathbf{ac}^T + \mathbf{ca}^T\big)\tilde{\mathbf{K}}^T = \sigma^4\tilde{\mathbf{K}}\big(\mathbf{1}_n\mathbf{1}_n^T + 2\,\mathbf{I}_n + 2\mathcal{S}(\tilde{\mathbf{s}}\mathbf{1}_n^T)\big)\tilde{\mathbf{K}}^T \sim \sigma^4 + \sigma^4/n,$$

while the other terms, i.e. $\tilde{\mathbf{K}}\mathbb{E}[(\mathbf{b}+\mathbf{c})(\mathbf{b}+\mathbf{c})^T]\tilde{\mathbf{K}}^T$, will be dropped from the MSE because their orders of magnitude are $\mathcal{O}(\sigma^4/n^2)$. Accordingly,

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}\Delta_2\hat{\boldsymbol{\theta}}^T) = \frac{\sigma^4}{4\tilde{R}^2}\tilde{\mathbf{K}}\left(\mathbf{1}_n\mathbf{1}_n^T + 2\mathbf{I}_n + 2\mathcal{S}(\tilde{\mathbf{s}}\mathbf{1}_n^T)\right)\tilde{\mathbf{K}}^T + \mathcal{O}(\sigma^4/n^2).$$

Since $\tilde{\mathbf{W}}^T\mathbf{1}_n$ coincides with the third column of $\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}$, then $\tilde{\mathbf{K}}\mathbf{1}_n = \hat{\mathbf{e}}_3$, from which we have

$$\mathbb{E}(\Delta_2\hat{\boldsymbol{\theta}}\Delta_2\hat{\boldsymbol{\theta}}^T) = \frac{\sigma^4}{4\tilde{R}^2}\left(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}})^{-1} + 2\mathcal{S}(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T)\right) + \mathcal{O}(\sigma^4/n^2). \tag{A.4}$$

Next step is to classify terms according to their origins. Recall that $\text{bias}(\Delta_2\hat{\boldsymbol{\theta}}) = \frac{\sigma^2}{2\tilde{R}}(\hat{\mathbf{e}}_3 + \tilde{\mathbf{K}}\tilde{\mathbf{s}})$, thus

$$\text{bias}(\Delta_2\hat{\boldsymbol{\theta}})\,\text{bias}(\Delta_2\hat{\boldsymbol{\theta}})^T = \frac{\sigma^4}{4\tilde{R}^2}\left(\hat{\mathbf{e}}_3\hat{\mathbf{e}}_3^T + 2\mathcal{S}(\tilde{\mathbf{K}}\tilde{\mathbf{s}}\hat{\mathbf{e}}_3^T)\right) + \mathcal{O}(\sigma^4/n^2). \tag{A.5}$$

Finally, subtracting (A.5) from (A.4) gives (3.4).

Now we find $\text{cov}(\Delta_1\hat{\boldsymbol{\theta}}, \Delta_3\hat{\boldsymbol{\theta}})$, where the third-order error term $\Delta_3\hat{\boldsymbol{\theta}}$. Expanding the distances d_i to the third order error terms gives

$$d_i = t_i - (\tilde{u}_i\,\Delta_3 a + \tilde{v}_i\,\Delta_3 b + \Delta_3 R) + \mathcal{O}_{\mathrm{P}}(\sigma^4),$$

where

$$t_i = l_i + f_{2,i} - (\tilde{u}_i\Delta_2 a + \tilde{v}_i\Delta_2 b + \Delta_2 R) + f_{3,i}, \tag{A.6}$$

$f_{3,i}$ is the cubic function of $\mathbf{e}_i = (\delta_i, \varepsilon_i)$ for each $i = 1, \ldots, n$ and has a complicated expression, however, we are interested only in $\mathbb{E}(\mathbf{f}_3\mathbf{f}_1^T)$ up to order $\mathcal{O}(\sigma^4/n)$, thus

$$f_{3,i} = -\frac{1}{2\tilde{R}^2}\big(\tilde{u}_i\delta_i + \tilde{v}_i\varepsilon_i\big)\big(\tilde{u}_i\varepsilon_i - \tilde{v}_i\delta_i\big)^2 + \mathcal{R},$$

where $\mathcal{R}$ is a random vector such that $\mathbb{E}(\mathcal{R}\mathbf{f}_3^T) \sim \mathcal{O}(\sigma^4/n)$. Thus minimizing $\sum d_i^2$ is now equivalent to minimizing

$$\sum(t_i - (\tilde{u}_i\,\Delta_3 a + \tilde{v}_i\,\Delta_3 b + \Delta_3 R))^2,$$

Similarly, one can prove that $\Delta_3\hat{\boldsymbol{\theta}} \approx \tilde{\mathbf{K}}\mathbf{f}_3$, where $\mathbf{f}_3 = (f_{3,1}, \dots, f_{3,n})^T$. Thus up to the leading term $\Delta_3\hat{\boldsymbol{\theta}}\Delta_1\hat{\boldsymbol{\theta}}^T = \tilde{\mathbf{K}}\mathbf{f}_3\mathbf{f}_1^T\tilde{\mathbf{K}}^T + \mathcal{R}_1$ (Again $\mathcal{R}_1$ is a random matrix whose entities have typical values of order $\mathcal{O}(\sigma^4/n^2)$). Since $f_{1,i} = \tilde{u}_i\delta_i + \tilde{v}_i\varepsilon_i$, then after simple algebra we find $\mathbb{E}(f_{1,i}f_{3,j}) = -\frac{\sigma^4}{2\tilde{R}^2}$ if $i = j$ and 0 otherwise and hence $\mathbb{E}(\mathbf{f}_3\mathbf{f}_1^T) = -\frac{\sigma^4}{2\tilde{R}^2}\mathbf{I}_n + \mathcal{O}(\sigma^4/n^2)$. Therefore

$$\operatorname{cov}(\Delta_3\hat{\boldsymbol{\theta}}, \Delta_1\hat{\boldsymbol{\theta}}) = -\frac{\sigma^4}{2\tilde{R}^2}\tilde{\mathbf{K}}\mathbf{I}_n\tilde{\mathbf{K}} = -\frac{\sigma^4}{2\tilde{R}^2}\tilde{\mathbf{V}}_{\min}.$$

□

Derivation of Equation (3.15). All terms of (3.15) are already computed except $\Delta_3\lambda$. Pre-multiplying (3.14) by $\tilde{\mathbf{A}}$ gives

$$\begin{aligned}\Delta_3\lambda = \frac{1}{(\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}})}\Big[&\tilde{\mathbf{A}}^T\Delta_3\mathbf{M}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\Delta_1\mathbf{A} + \tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\Delta_2\mathbf{A}\\ &- \Delta_2\lambda\Big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{N}\tilde{\mathbf{A}}) + (\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\Delta_1\mathbf{A})\Big)\Big],\end{aligned} \tag{A.7}$$

In the process of finding $\Delta_3\lambda$, we keep in our minds that we are interested only in all terms in $\Delta_3\mathbf{A}$ whose contributions to the MSE (represented by $2\mathcal{S}(\mathbb{E}(\Delta_3\mathbf{A}\Delta_1\mathbf{A}^T))$) are of order $\mathcal{O}(\sigma^4/n)$, and we discard other less important terms. If we would multiply (3.14) by $\Delta_1\mathbf{A}^T = -\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-$, and then find the expected value of the equation, we find several terms with different magnitudes. For example, $\mathbb{E}(\Delta_3\mathbf{M}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T)$, $\mathbb{E}(\Delta_2\mathbf{M}\Delta_1\mathbf{A}\Delta_1\mathbf{A}^T)$ are of order σ^4/n, while

$$\mathbb{E}((\Delta_1\mathbf{M}\Delta_2\mathbf{A})\Delta_1\mathbf{A}^T) = \mathbb{E}\Big((\Delta_{2,1}\lambda - \Delta_{2,2}\lambda)\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T - \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-(\mathbf{R}_1 - \mathbf{R}_2)\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\Big)$$

involves terms with different magnitudes. The less important part is

$$\mathbb{E}\big(\Delta_{2,2}\lambda\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big) \text{ and } \mathbb{E}\big(\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\mathbf{R}_2\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big),$$

which are of order σ^4/n^2 (the expected value of the products of four averages, each of which is a linear combinations of $\mathbf{h}$, i.e. $\mathbb{E}(\prod_{i=1}^4 \mathbf{q}_i^T\mathbf{h})$). The significant part is

$$\mathbb{E}\big(\Delta_{2,1}\lambda\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big) \text{ and } \mathbb{E}\big(\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\mathbf{R}_1\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big),$$

each of which has order of magnitude σ^4/n. Based on that, instead of using the complete expressions of $\Delta_2\lambda$ and $\Delta_2\mathbf{A}$ we use $\Delta_{2,1}\lambda$ and

$$\Delta_{2,1}\mathbf{A} = \tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\mathbf{R}_1\tilde{\mathbf{A}} = \big(\Delta_{2,1}\lambda\tilde{\mathbf{M}}^-\tilde{\mathbf{N}} - \tilde{\mathbf{M}}^-\Delta_2\mathbf{M}\big)\tilde{\mathbf{A}},$$

and as such $\Delta_3\lambda = \Delta_{3,1}\lambda + \Delta_{3,2}\lambda$, where the contribution of $\Delta_{3,2}\lambda$ is of order σ^4/n^2, while the most important part is

$$\begin{aligned}\Delta_{3,1}\lambda = \frac{1}{(\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}})}\Big[&\tilde{\mathbf{A}}^T\Delta_3\mathbf{M}\tilde{\mathbf{A}} + \tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\Delta_1\mathbf{A} + \tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\Delta_{2,1}\mathbf{A}\\ &- \Delta_{2,1}\lambda\Big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{N}\tilde{\mathbf{A}}) + (\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\Delta_1\mathbf{A})\Big)\Big],\end{aligned} \tag{A.8}$$

and further

$$\Delta_{3,1}\lambda = \chi_1 + (\chi_2 - \chi_3)\Delta_{2,1}\lambda. \tag{A.9}$$

The scalars χ_1, χ_2, and χ_3 are defined as

$$\chi_1 = \frac{(\tilde{\mathbf{A}}^T\mathcal{J}1)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}, \quad \chi_2 = \frac{(\tilde{\mathbf{A}}^T\mathcal{J}2)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}, \text{ and } \chi_3 = \frac{(\tilde{\mathbf{A}}^T\Delta_1\mathbf{N}\tilde{\mathbf{A}})}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}, \tag{A.10}$$

where the vectors $\mathcal{J}_1$ and $\mathcal{J}_2$ are defined in (3.16). Replace $\Delta_3\lambda$ by $\Delta_{3,1}\lambda$ in (3.14) gives $\Delta_3\mathbf{A} = \Delta_{3,1}\mathbf{A} + \mathcal{R}$, where the random vector $\mathcal{R} \sim \mathcal{O}_{\mathrm{P}}(\sigma^3)$ such that $\mathbb{E}(\mathcal{R}\Delta_1\mathbf{A}^T) \sim \mathcal{O}(\sigma^4/n^2)$. Thus

$$\Delta_{3,1}\mathbf{A} = \tilde{\mathbf{M}}^-\Big[-\big(\Delta_3\mathbf{M}\tilde{\mathbf{A}} + \Delta_2\mathbf{M}\Delta_1\mathbf{A} + \Delta_1\mathbf{M}\Delta_{2,1}\mathbf{A}\big) + \Delta_{3,1}\lambda\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_{2,1}\lambda(\Delta_1\mathbf{N}\tilde{\mathbf{A}} + \tilde{\mathbf{N}}\Delta_1\mathbf{A})\Big]$$

and further it becomes

$$\Delta_{3,1}\mathbf{A} = \tilde{\mathbf{M}}^-\Big[-\mathcal{J}_1 - \Delta_{2,1}\lambda\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \big(\chi_1 + (\chi_2-\chi_3)\Delta_{2,1}\lambda\big)\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_{2,1}\lambda(\Delta_1\mathbf{N}\tilde{\mathbf{A}} + \tilde{\mathbf{N}}\Delta_1\mathbf{A})\Big]$$

Rearranging terms, after substituting the formal expressions of χ_1, χ_2, and χ_3, gives

$$\Delta_{3,1}\mathbf{A} = \tilde{\mathbf{M}}^-\Big[\Delta_{2,1}\lambda\big(-\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \frac{(\tilde{\mathbf{A}}^T\mathcal{J}2)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \tilde{\mathbf{N}}\Delta_1\mathbf{A}\big) - \mathcal{J}_1 + \frac{(\tilde{\mathbf{A}}^T\mathcal{J}1)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_{2,1}\lambda\Big(-\frac{(\tilde{\mathbf{A}}^T\Delta_1\mathbf{N}\tilde{\mathbf{A}})}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_1\mathbf{N}\tilde{\mathbf{A}}\Big)\Big]$$

which can be simply written as

$$\tilde{\mathbf{M}}^-\Big[-\mathcal{J}_1 + \frac{(\tilde{\mathbf{A}}^T\mathcal{J}1)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_{2,1}\lambda\big(\frac{(\tilde{\mathbf{A}}^T\mathcal{J}_2)}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} - \mathcal{J}_2\big) + \Delta_{2,1}\lambda\big(-\frac{(\tilde{\mathbf{A}}^T\Delta_1\mathbf{N}\tilde{\mathbf{A}})}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}\tilde{\mathbf{N}}\tilde{\mathbf{A}} + \Delta_1\mathbf{N}\tilde{\mathbf{A}}\big)\Big]$$

which reduces to $\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}[\mathcal{J}_1 + \Delta_{2,1}\lambda(\mathcal{J}_2 - \Delta_1\mathbf{N}\tilde{\mathbf{A}})]$ after using the defintion of the operator $\mathbf{S}_{\mathrm{N}} = \frac{\mathbf{N}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}$. □

Proof of Theorem 3.2. First let us state and prove the following lemma

Lemma A.1. *For each $i = 1, \dots, n$, define $\tilde{\mathbf{\Pi}}_i = \tilde{\mathbf{z}}_i\mathbf{e}_1^T$, and*

$$\mathbf{B}_i = (\Delta_2\mathbf{z}_i^T\tilde{\mathbf{A}})^2\tilde{\mathbf{M}}_i, \qquad \mathbf{C}_i = (\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_2\mathbf{z}_i)\Delta_1\mathbf{z}_i\tilde{\mathbf{z}}_i^T, \tag{A.11}$$

$$\mathbf{D}_i = (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})^2\Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T, \quad \mathbf{E}_i = (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})^2\Delta_2\mathbf{z}_i\tilde{\mathbf{z}}_i^T \tag{A.12}$$

then

$$\mathbb{E}(\mathbf{B}_i) = 2\sigma^4(2\tilde{\mathbf{\Pi}}_i\tilde{\mathbf{A}})(2\tilde{\mathbf{\Pi}}_i\tilde{\mathbf{A}})^T, \quad \mathbb{E}\big(\mathbf{C}_i\big) = 4\sigma^4\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}(\tilde{\mathbf{\Pi}}_i\tilde{\mathbf{A}})^T,$$

$$\mathbb{E}(\mathbf{D}_i) = \sigma^4\tilde{\zeta}\tilde{\mathbf{V}}_i + 2\sigma^4(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})^T, \quad \mathbb{E}(\mathbf{E}_i) = 4\sigma^4\tilde{\zeta}\tilde{\mathbf{\Pi}}_i^T$$

Proof of Lemma A.1. $\mathbb{E}(\mathbf{B}_i)$ follows immediately from

$$\mathbb{E}\big((\Delta_2 \mathbf{z}_i^T \tilde{\mathbf{A}})^2\big) = \mathbb{E}\big(\tilde{A}^2(\delta_i^2+\varepsilon_i^2)^2\big) = \tilde{A}^2\big(\mathrm{var}(\delta_i^2+\varepsilon_i^2) + \mathbb{E}(\delta_i^2+\varepsilon_i^2)^2\big) = 8\tilde{A}^2\sigma^4.$$

Note that $(\delta_i^2+\varepsilon_i^2)/\sigma^2$ has Chi-Squared distribution with $df = 2$. Using $\tilde{A}\tilde{\mathbf{z}}_i = \tilde{\mathbf{\Pi}}\tilde{\mathbf{A}}$ gives the desired result. Next, we find $\mathbb{E}(\mathbf{C}_i)$. Recall $\Delta_1 \mathbf{z}_i = \tilde{\mathbf{a}}_i\delta_i + \tilde{\mathbf{b}}_i\varepsilon_i$ and $\tilde{\alpha}_i = \tilde{\mathbf{A}}^T\tilde{\mathbf{a}}_i$ and $\tilde{\beta}_i = \tilde{\mathbf{A}}^T\tilde{\mathbf{b}}_i$, and hence $\tilde{\alpha}_i\tilde{\mathbf{a}}_i + \tilde{\beta}_i\tilde{\mathbf{b}}_i = \tilde{\mathbf{V}}_i\tilde{\mathbf{A}}$, then

$$\begin{aligned}
\mathbb{E}(\mathbf{C}_i) &= \tilde{A}\mathbb{E}\Big((\tilde{\alpha}_i\delta_i + \tilde{\beta}_i\varepsilon_i)(\delta_i^2+\varepsilon_i^2)(\tilde{\mathbf{a}}_i\delta_i + \tilde{\mathbf{b}}_i\varepsilon_i)\tilde{\mathbf{z}}_i^T\Big) \\
&= \tilde{A}\mathbb{E}\Big((\tilde{\alpha}_i\delta_i^4\tilde{\mathbf{a}}_i + \tilde{\beta}_i\varepsilon_i^4\tilde{\mathbf{b}}_i + (\tilde{\alpha}_i\tilde{\mathbf{a}}_i + \tilde{\beta}_i\tilde{\mathbf{b}}_i)\varepsilon_i^2\delta_i^2\Big)\tilde{\mathbf{z}}_i^T \\
&= 4\tilde{A}\sigma^4\Big(\tilde{\alpha}_i\tilde{\mathbf{a}}_i + \tilde{\beta}_i\tilde{\mathbf{b}}_i\Big)\tilde{\mathbf{z}}_i^T = 4\sigma^4\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}(\tilde{\mathbf{\Pi}}_i\tilde{\mathbf{A}})^T.
\end{aligned}$$

Now, after simple calculations, one can find

$$\mathbf{D}_i = (\tilde{\alpha}_i^2\delta_i^2 + 2\tilde{\alpha}_i\tilde{\beta}_i\delta_i\varepsilon_i + \tilde{\beta}_i^2\varepsilon_i^2)(\tilde{\mathbf{a}}_i\delta_i + \tilde{\mathbf{b}}_i\varepsilon_i)(\tilde{\mathbf{a}}_i^T\delta_i + \tilde{\mathbf{b}}_i^T\varepsilon_i).$$

Using

$$\tilde{\alpha}_i^2 + \tilde{\beta}_i^2 = \tilde{\mathbf{A}}^T(\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + \tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T)\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\mathrm{P}}\tilde{\mathbf{A}} = \tilde{\zeta}$$

and $\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} = \tilde{\alpha}_i\tilde{\mathbf{a}}_i + \tilde{\beta}_i\tilde{\mathbf{b}}_i$, one can find

$$\begin{aligned}
\mathbb{E}(\mathbf{D}_i) &= \mathbb{E}\Big(\tilde{\alpha}_i^2\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T\delta_i^4 + \tilde{\beta}_i^2\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T\varepsilon_i^4 + \big(\tilde{\alpha}_i^2\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T + \tilde{\beta}_i^2\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + 4\tilde{\alpha}_i\tilde{\beta}_i\mathcal{S}[\tilde{\mathbf{a}}_i\tilde{\mathbf{b}}_i^T]\big)\delta_i^2\varepsilon_i^2\Big) \\
&= \sigma^4\Big(3\tilde{\alpha}_i^2\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + 3\tilde{\beta}_i^2\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T + \big(\tilde{\alpha}_i^2\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T + \tilde{\beta}_i^2\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + 4\tilde{\alpha}_i\tilde{\beta}_i\mathcal{S}[\tilde{\mathbf{a}}_i\tilde{\mathbf{b}}_i^T]\big)\Big) \\
&= 2\sigma^4\big(\tilde{\alpha}_i^2\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + \tilde{\beta}_i^2\tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T + 2\tilde{\alpha}_i\tilde{\beta}_i\mathcal{S}[\tilde{\mathbf{a}}_i\tilde{\mathbf{b}}_i^T]\big) + \sigma^4(\tilde{\alpha}_i^2 + \tilde{\beta}_i^2)(\tilde{\mathbf{a}}_i\tilde{\mathbf{a}}_i^T + \tilde{\mathbf{b}}_i\tilde{\mathbf{b}}_i^T) \\
&= 2\sigma^4(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})^T + \sigma^4\tilde{\zeta}\tilde{\mathbf{V}}_i
\end{aligned}$$

Finally, since

$$\mathbf{E}_i = (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})^2\Delta_2\mathbf{z}_i\tilde{\mathbf{z}}_i^T = \big(\tilde{\alpha}_i^2\delta_i^2 + 2\tilde{\alpha}_i\tilde{\beta}_i\delta_i\varepsilon_i + \tilde{\beta}_i^2\varepsilon_i^2\big)(\delta_i^2+\varepsilon_i^2)\breve{\mathbf{e}}_1\tilde{\mathbf{z}}_i^T$$

one gets $\mathbb{E}(\mathbf{E}_i) = 4(\tilde{\alpha}_i^2 + \tilde{\beta}^2)\sigma^4\breve{\mathbf{e}}_1\tilde{\mathbf{z}}_i^T = 4\sigma^4\tilde{\zeta}(\tilde{\mathbf{\Pi}}\tilde{\mathbf{A}})^T$. This completes the proof of the lemma. □

Next we prove Theorem 3.2. First let us write

$$\mathrm{cov}(\Delta_1\mathbf{A}, \Delta_3\mathbf{A}) = \mathbb{E}(\Delta_3\mathbf{A}\Delta_1\mathbf{A}^T) = \mathbf{\Lambda}_1 + \mathbf{\Lambda}_2 + \mathbf{\Lambda}_3 + \mathbf{\Lambda}_4$$

where

$$\begin{aligned}
\mathbf{\Lambda}_1 &= -\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\mathbb{E}\Big(\Delta_3\mathbf{M}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Big) \\
\mathbf{\Lambda}_2 &= 2\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\mathbb{E}\Big(\mathcal{S}\big[\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_2\mathbf{M}\big]\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Big) \\
\mathbf{\Lambda}_3 &= -2\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\mathbb{E}\Big(\Delta_{2,1}\lambda\mathcal{S}[\tilde{\mathbf{N}}\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}]\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Big) \\
\mathbf{\Lambda}_4 &= -\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\mathbb{E}\Big(\Delta_{2,1}\lambda\Delta_1\mathbf{N}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\Big).
\end{aligned}$$

Now let us handel each $\mathbf{\Lambda}_i'$s starting with $\mathbf{\Lambda}_1$. Recall that $\mathbb{E}(\delta_i^3\varepsilon_j) = \mathbb{E}(\delta_i\varepsilon_j^3) = 0$, for all i, j and $\Delta_3\mathbf{M} = \frac{1}{n}\sum_{i=1}^n 2\mathcal{S}(\Delta_1\mathbf{z}_i\Delta_2\mathbf{z}_i^T)$ and $\Delta_1\mathbf{M}$. Using Lemma A.1, $\mathbb{E}(\Delta_3\mathbf{M}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M})$ is

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\big(2\mathcal{S}[\Delta_2\mathbf{z}_i\Delta_1\mathbf{z}_i^T]\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)\tilde{\mathbf{z}}_j^T\big) \\
&= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\Big((\Delta_2\mathbf{z}_i\Delta_1\mathbf{z}_i^T + \Delta_1\mathbf{z}_i\Delta_2\mathbf{z}_i^T)\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)\tilde{\mathbf{z}}_i^T\Big) \\
&= \frac{1}{n^2}\sum_{i=1}^n \Big(\mathbb{E}\big((\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})^2\Delta_2\mathbf{z}_i\tilde{\mathbf{z}}_i^T\big) + \mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_2\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)\Delta_1\mathbf{z}_i\tilde{\mathbf{z}}_i^T\big)\Big) \\
&= \frac{1}{n^2}\sum_{i=1}^n \big(\mathbb{E}(\mathbf{E}_i) + \mathbb{E}(\mathbf{C}_i)\big) \\
&= \frac{4\sigma^4}{n^2}\sum_{i=1}^n \Big((\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(\tilde{\mathbf{\Pi}}_i\tilde{\mathbf{A}})^T + \tilde{\zeta}\tilde{\mathbf{\Pi}}_i^T\Big)
\end{aligned}
$$

If the last expression is pre-multiplied by $-\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}$ and post-multiplied by $\tilde{\mathbf{M}}^-$, we get $\mathbf{\Lambda}_1$. Next we Compute $\mathbf{\Lambda}_2$ by firstly writing

$$
\mathbb{E}\big(\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_2\mathbf{M}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}\big) = \mathbf{\Upsilon}_1 + \mathbf{\Upsilon}_2 + \mathbf{\Upsilon}_3 + \mathbf{\Upsilon}_4
$$

where $\mathbf{\Upsilon}_i$'s are defined by

$$
\begin{aligned}
\mathbf{\Upsilon}_1 &= \frac{1}{n^3}\sum_{i,j,k=1}^n \mathbb{E}\big[\tilde{\mathbf{z}}_j\Delta_1\mathbf{z}_j^T\tilde{\mathbf{M}}^-(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_k)\Delta_1\mathbf{z}_i\tilde{\mathbf{z}}_k^T\big] \\
\mathbf{\Upsilon}_2 &= \frac{1}{n^3}\sum_{i,j,k=1}^n \mathbb{E}\big[\Delta_1\mathbf{z}_j\tilde{\mathbf{z}}_j^T\tilde{\mathbf{M}}^-(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_k)\Delta_1\mathbf{z}_i\tilde{\mathbf{z}}_k^T\big] \\
\mathbf{\Upsilon}_3 &= \frac{1}{n^3}\sum_{i,j,k=1}^n \mathbb{E}\big[(\tilde{\mathbf{z}}_j\Delta_1\mathbf{z}_j^T)\tilde{\mathbf{M}}^-(\tilde{\mathbf{A}}^T\Delta_2\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_k)\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_k^T\big] \\
\mathbf{\Upsilon}_4 &= \frac{1}{n^3}\sum_{i,j,k=1}^n \mathbb{E}\big[(\Delta_1\mathbf{z}_j\tilde{\mathbf{z}}_j^T)\tilde{\mathbf{M}}^-(\tilde{\mathbf{A}}^T\Delta_2\mathbf{z}_i)(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_k)\tilde{\mathbf{z}}_i\tilde{\mathbf{z}}_k^T\big]
\end{aligned}
$$

Remark A.1. Note that $\Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_2\mathbf{M}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}$ is a product of three averages; two of which are linear combinations of the combined error vector $\mathbf{h} = (\delta_1,\dots,\delta_n,\varepsilon_1,\dots,\varepsilon_n)^T$, and the third is a quadratic form of $\mathbf{h}$, thus it is enough to write the above expressions of $\mathbf{\Upsilon}_i$'s in terms of i and j because $\mathbb{E}(\delta_i^2\varepsilon_j\varepsilon_k) = \mathbb{E}(\varepsilon_i^2\delta_j\delta_k) = 0$ for all $j \neq k$.

Remark A.2. By Remark A.1, all $\mathbf{\Upsilon}_i$'s consist of two summations having different orders of magnitude; some of order $\mathcal{O}(n^2\sigma^4)$ (when $i \neq j$ and $j = k$) and others of order $\mathcal{O}(n\sigma^4)$ (when $i = j = k$), thus their contributions to MSE are of orders $\mathcal{O}(\sigma^4/n)$, $\mathcal{O}(\sigma^4/n^2)$, respectively. This means that we will ignore the latter case from our consideration.

Denote $\tilde{\mathbf{M}}_{ij} = \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_j^T$ (if $i = j$, set $\tilde{\mathbf{M}}_{ii} = \tilde{\mathbf{M}}_i$), and repeatedly apply the fact $\mathbb{E}(\delta_i^2 \delta_j^2) = \mathbb{E}(\delta_i^2)\,\mathbb{E}(\delta_j^2) = \sigma^4$ for all $i \neq j$, to each of $\boldsymbol{\Upsilon}_i$'s. We reach

$$\mathbb{E}\big(\Delta_1 \mathbf{M} \tilde{\mathbf{M}}^- \Delta_2 \mathbf{M} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M}\big) = \frac{2\sigma^4}{n^3} \sum_{i \neq j}^{n} \mathcal{S}\big(\mathbf{V}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T\big) \tilde{\mathbf{M}}^- \big(\mathbf{V}_i \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T + 2\tilde{A} \tilde{\mathbf{M}}_{ij}\big). \tag{A.13}$$

In the same analogue, we find $\mathbb{E}(\Delta_2 \mathbf{M} \tilde{\mathbf{M}}^- \Delta_1 \mathbf{M} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M})$. Recall that $\tilde{\boldsymbol{\Pi}}_i = \tilde{\mathbf{z}}_i \breve{\mathbf{e}}_1^T$, then

$$\frac{1}{n^3} \sum_{i \neq j}^{n} \Big(\mathbb{E}\big(\Delta_1 \mathbf{z}_i \Delta_1 \mathbf{z}_i^T + \Delta_2 \mathbf{z}_i \tilde{\mathbf{z}}_i^T + \tilde{\mathbf{z}}_i \Delta_2 \mathbf{z}_i^T\big) \cdot \tilde{\mathbf{M}}^- \cdot \mathbb{E}\big(\tilde{\mathbf{A}}^T \Delta_1 \mathbf{z}_j \Delta_1 \mathbf{z}_j^T \tilde{\mathbf{A}}\big) \tilde{\mathbf{M}}_j \Big)$$

and further

$$\mathbb{E}\big(\Delta_2 \mathbf{M} \tilde{\mathbf{M}}^- \Delta_1 \mathbf{M} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M}\big) = \frac{\sigma^4}{n^3} \tilde{\zeta} \sum_{i \neq j}^{n} \Big(\big(\tilde{\mathbf{V}}_i + 4\mathcal{S}[\tilde{\boldsymbol{\Pi}}_i]\big)\, \tilde{\mathbf{M}}^- \tilde{\mathbf{M}}_j \Big) \tag{A.14}$$

Now combine Equations (A.13) and (A.14) to get

$$\boldsymbol{\Lambda}_2 = \frac{\sigma^4}{n^3} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \sum_{i \neq j}^{n} \Big(2\mathcal{S}[\mathbf{V}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T] \tilde{\mathbf{M}}^- (\mathbf{V}_i \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T + 2\tilde{A} \tilde{\mathbf{M}}_{ij}) + \tilde{\zeta}\big((\tilde{\mathbf{V}}_i + 4\mathcal{S}[\tilde{\boldsymbol{\Pi}}_i]) \tilde{\mathbf{M}}^- \tilde{\mathbf{M}}_j\big) \Big) \tilde{\mathbf{M}}^- \tag{A.15}$$

To find $\boldsymbol{\Lambda}_3$, we use the definition of $\Delta_{2,1}\lambda$ given in equation (3.11). Firstly, observe that for all $i = 1, \dots, n$, we have $\mathbb{E}(\tilde{\mathbf{A}}^T \Delta_2 \mathbf{M}_i \tilde{\mathbf{A}}) = \sigma^2 \tilde{\zeta}$ and

$$\mathbb{E}\big((\Delta_1 \mathbf{M}_i \tilde{\mathbf{A}})(\Delta_1 \mathbf{M}_i \tilde{\mathbf{A}})^T\big) = \tilde{\mathbf{A}}^T \mathbb{E}(\Delta_1 \mathbf{z}_i \Delta_1 \mathbf{z}_i^T) \tilde{\mathbf{A}} \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i^T = \sigma^2 \tilde{\zeta} \tilde{\mathbf{M}}_i.$$

Consequently, $\mathbb{E}(\Delta_{2,1}\lambda \tilde{\mathbf{N}} \tilde{\mathbf{M}}^- \Delta_1 \mathbf{M} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M} \tilde{\mathbf{M}}^-) =$

$$\frac{\tilde{\mathbf{N}} \tilde{\mathbf{M}}^-}{n^3 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \sum_{i \neq j}^{n} \mathbb{E}(\tilde{\mathbf{A}}^T \Delta_2 \mathbf{M}_i \tilde{\mathbf{A}})\, \mathbb{E}\big((\Delta_1 \mathbf{M}_j \tilde{\mathbf{A}})(\Delta_1 \mathbf{M}_j \tilde{\mathbf{A}})^T\big) \tilde{\mathbf{M}}^-$$

and as such $\mathbb{E}(\Delta_{2,1}\lambda \tilde{\mathbf{N}} \tilde{\mathbf{M}}^- \Delta_1 \mathbf{M} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M} \tilde{\mathbf{M}}^-) =$

$$\sigma^4 \frac{\tilde{\zeta}^2 \tilde{\mathbf{N}} \tilde{\mathbf{M}}^-}{n^3 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \sum_{i \neq j}^{n} \tilde{\mathbf{M}}_j \tilde{\mathbf{M}}^- = \frac{(n-1)\tilde{\zeta}^2 \sigma^4}{n^2 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \tilde{\mathbf{N}} \tilde{\mathbf{M}}^- \tag{A.16}$$

In the same approach we find

$$\mathbb{E}\big(\Delta_2 \lambda_{2,1} \Delta_1 \mathbf{M} \tilde{\mathbf{M}}^- \tilde{\mathbf{N}} \tilde{\mathbf{A}} \tilde{\mathbf{A}}^T \Delta_1 \mathbf{M} \tilde{\mathbf{M}}^-\big) = \frac{\sigma^4 \tilde{\zeta} \sum_{i \neq j}^{n} \big(2\mathcal{S}[\mathbf{V}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T] \tilde{\mathbf{M}}^- \tilde{\mathbf{N}} \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T \tilde{\mathbf{M}}^-\big)}{n^3 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \tag{A.17}$$

Substitute (A.16) and (A.17) in $\boldsymbol{\Lambda}_3$ to get

$$\boldsymbol{\Lambda}_3 = -\frac{(n-1)\tilde{\zeta}^2 \sigma^4}{n^2 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}}\, \tilde{\mathbf{N}} \tilde{\mathbf{M}}^- - \frac{\tilde{\zeta} \sigma^4}{n^3 (\tilde{\mathbf{A}}^T \tilde{\mathbf{N}} \tilde{\mathbf{A}})} \tilde{\mathbf{M}}^- \mathbf{S}_{\tilde{\mathbf{N}}} \sum_{i \neq j}^{n} [2\mathcal{S}(\tilde{\mathbf{V}}_j \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T) \tilde{\mathbf{M}}^- \tilde{\mathbf{N}} \tilde{\mathbf{A}} \tilde{\mathbf{z}}_j^T \tilde{\mathbf{M}}^-] \tag{A.18}$$

Finally, the last term $\mathbf{\Lambda}_4$ depends on $\Delta_1\mathbf{N}$ which is exactly zero for Pratt's and Kåsa's fits, since their corresponding constraint matrices are independent of the data. On the other hand, the constraint matrices associated with Taubin's fit and Hyper fit are data-dependent, and as such we should compute $\mathbf{\Lambda}_4$ for both fits separately. Because $\mathbf{N}_{\rm H} = 2\mathbf{N}_{\rm T} - \mathbf{N}_{\rm P}$, $\Delta_1\mathbf{N}_{\rm H} = 2\Delta_1\mathbf{N}_{\rm T}$, and as such it is enough to compute $\mathbf{\Lambda}_4$ for Taubin's fit. Since $\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\rm T}\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\rm P}\tilde{\mathbf{A}} = \tilde{\zeta}$, then

$$\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\rm H}\tilde{\mathbf{A}} = 2\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\rm T}\tilde{\mathbf{A}} - \tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_{\rm P}\tilde{\mathbf{A}} = \tilde{\zeta} \tag{A.19}$$

Also, recall that $\Delta_{2,1}\lambda = \frac{(\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})}{\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}\tilde{\mathbf{A}}}$, then $\mathbf{\Lambda}_4$'s for both Hyper and the Taubin fits become

$$\begin{aligned}\mathbf{\Lambda}_{4,\rm T} &= -\frac{1}{\tilde{\zeta}}\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}_{\rm T}}\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})\Delta_1\mathbf{N}_{\rm T}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big)\\ \mathbf{\Lambda}_{4,\rm H} &= -\frac{2}{\tilde{\zeta}}\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}_{\rm H}}\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})\Delta_1\mathbf{N}_{\rm T}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\big)\end{aligned} \tag{A.20}$$

respectively. Also, since $\tilde{\mathbf{N}}_{\rm T} = \frac{1}{n}\sum_{i=1}^n\tilde{\mathbf{V}}_i$, then $\mathbb{E}((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})\Delta_1\mathbf{N}_{\rm T}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T)$ becomes

$$\begin{aligned}&= -\frac{1}{n^3}\sum_{i,j,k=1}^{n}\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}_i\tilde{\mathbf{A}})\Delta_1\mathbf{V}_j\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}_k\tilde{\mathbf{M}}^-\big)\\ &= -\frac{1}{n^3}\sum_{i,j=1}^{n}\mathbb{E}\big((\tilde{\mathbf{A}}^T[\Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T]\tilde{\mathbf{A}})\Delta_1\mathbf{V}_j\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)\tilde{\mathbf{z}}_j^T\big)\tilde{\mathbf{M}}^-\end{aligned}$$

$\Delta_1\mathbf{M}_i$ and $\Delta_2\mathbf{M}_i$ are the linear and the quadratic forms of the random vectors $(\delta_i, \varepsilon_i)$ for all $i = 1, \ldots, n$. Hence the expectation of all cross-product terms, which have different indices j and k ($j \neq k$), is exactly zero. Note that when $i = j$, the resulting expression will be dropped because it is of order $\mathcal{O}(\sigma^4/n^2)$. Thus we are interested only in the case of $i \neq j$. Thus $\mathbb{E}((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})\Delta_1\mathbf{N}_{\rm T}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T)$ becomes

$$\begin{aligned}&= -\frac{1}{n^3}\sum_{i\neq j}^{n}\big[\tilde{\mathbf{A}}^T\mathbb{E}(\Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T)\tilde{\mathbf{A}}\cdot\mathbb{E}(\Delta_1\mathbf{V}_j\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j))\tilde{\mathbf{z}}_j^T\big]\tilde{\mathbf{M}}^-\\ &= -\frac{\tilde{\zeta}\sigma^2}{n^3}\sum_{i\neq j}^{n}\mathbb{E}\big(\Delta_1\mathbf{V}_j\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)\big)\tilde{\mathbf{z}}_j^T\tilde{\mathbf{M}}^-\end{aligned} \tag{A.21}$$

Next, perturbating $\mathbf{V}_j$ gives $\Delta_1\mathbf{V}_j = \tilde{\mathbf{Y}}_j\delta_j + \tilde{\mathbf{P}}_j\varepsilon_j$, then we conclude

$$\mathbb{E}(\Delta_1\mathbf{V}_j\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)) = \sigma^2\tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}.$$

Combining all above equations together and using $\text{cov}(\Delta_1\mathbf{z}_i) = \sigma^2\tilde{\mathbf{N}}_{\rm T}$ simplifies $\mathbb{E}((\tilde{\mathbf{A}}^T\Delta_2\mathbf{M}\tilde{\mathbf{A}})\Delta_1\mathbf{N}_{\rm T}\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T)$ to

$$= -\frac{\tilde{\zeta}\sigma^4}{n^3}\Big(\sum_{i\neq j}^{n}\tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}\tilde{\mathbf{z}}_j^T\Big)\tilde{\mathbf{M}}^- = -\frac{(n-1)\tilde{\zeta}\sigma^4}{n^3}\Big(\sum_{j=1}^{n}\tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}\tilde{\mathbf{z}}_j^T\Big)\tilde{\mathbf{M}}^-$$

According to above discussion, $\mathbf{\Lambda}_4$ for Taubin's and Hyper fits can be formalized by the expressions given below

$$\mathbf{\Lambda}_{4,\mathrm{T}} = \frac{(n-1)\sigma^4}{n^3}\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}_\mathrm{T}}\Big(\sum_{j=1}^n \tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}\tilde{\mathbf{z}}_j^T\Big)\tilde{\mathbf{M}}^- \tag{A.22}$$

$$\mathbf{\Lambda}_{4,\mathrm{H}} = \frac{2(n-1)\sigma^4}{n^3}\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}_\mathrm{H}}\Big(\sum_{j=1}^n \tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}\tilde{\mathbf{z}}_j^T\Big)\tilde{\mathbf{M}}^- \tag{A.23}$$

respectively. In the same analogue we find $\mathbf{\Lambda}_{4,I}$ for HyperLS fit, say $\mathbf{\Lambda}_{4,I}$. First note that $\sum_{i=1}^n \tilde{\xi}_i = \sum_{i=1}^n \tilde{\mathbf{z}}_i^T\tilde{\mathbf{M}}^-\tilde{\mathbf{z}}_i = \mathrm{trace}(\tilde{\mathbf{M}}^-\tilde{\mathbf{M}}) = 3$, thus

$$\tilde{\mathbf{A}}^T\tilde{\mathbf{N}}_\mathrm{I}\tilde{\mathbf{A}} = \tilde{\zeta}\Big(1 - \frac{1}{n^2}\sum_{i=1}^n \tilde{\xi}_i\Big) = \tilde{\zeta}(1 - 3/n^2)$$

This fact together with Eq. (A.21) and $\mathrm{trace}(\tilde{\mathbf{M}}^-\tilde{\mathbf{V}}_i) \sim \mathcal{O}(n)$ give us

$$\mathbf{\Lambda}_{4,\mathrm{I}} = -\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}_\mathrm{I}}\Big[-\frac{2(n-1)\sigma^4(\sum_{j=1}^n\tilde{\mathbf{Q}}_j\tilde{\mathbf{A}}\tilde{\mathbf{z}}_j^T)\tilde{\mathbf{M}}^-}{n^3(1-\frac{3}{n^2})} - \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}(\Delta_2\lambda_{2,1}\mathrm{trace}(\tilde{\mathbf{M}}^-\tilde{\mathbf{V}}_i)\Delta_1\mathbf{M}_i\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T)\Big].$$

Finally, one can use

$$\frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\Big(\Delta_2\lambda_{2,1}\mathrm{trace}(\tilde{\mathbf{M}}^-\tilde{\mathbf{V}}_i)\Delta_1\mathbf{M}_i\tilde{\mathbf{A}}\Delta_1\mathbf{A}^T\Big) = -\frac{\tilde{\zeta}(n-1)\sigma^4\sum_{i=1}^n\mathrm{trace}(\tilde{\mathbf{M}}^-\tilde{\mathbf{V}}_i)\tilde{\mathbf{M}}_i\tilde{\mathbf{M}}^-}{n^4(1-\frac{3}{n^2})}.$$

to get the desired result. □

Proof of Theorem 3.3. Recall $\Delta_2\mathbf{A} = \tilde{\mathbf{M}}^-\tilde{\mathbf{S}}_{\tilde{\mathbf{N}}}(\mathbf{R}_1 - \mathbf{R}_2)\tilde{\mathbf{A}}$, where $\mathbf{R}_1 = \Delta_2\mathbf{M}$ and $\mathbf{R}_2 = \Delta_1\mathbf{M}\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}$. It is straightforward to show that

$$\mathbb{E}(\mathbf{R}_1\tilde{\mathbf{A}}) = \frac{\sigma^2}{n}\sum_{i=1}^n \tilde{\Psi}_i\tilde{\mathbf{A}},$$

where $\tilde{\Psi}_i = \tilde{\mathbf{V}}_i + 2\mathcal{S}[\tilde{\mathbf{\Pi}}_i]$; while

$$\begin{aligned}\mathbb{E}(\mathbf{R}_2\tilde{\mathbf{A}}) &= \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}\Big(2\mathcal{S}(\tilde{\mathbf{z}}_i\Delta_1\mathbf{z}_i^T)\tilde{\mathbf{M}}^-(\Delta_1\mathbf{z}_j^T\tilde{\mathbf{A}})\tilde{\mathbf{z}}_j\Big)\\ &= \frac{1}{n^2}\sum_{i=1}^n\Big((\tilde{\mathbf{z}}_i\tilde{\mathbf{A}}^T)\mathbb{E}\big(\Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T\big)\tilde{\mathbf{M}}^-\tilde{\mathbf{z}}_i + \tilde{\psi}_i\mathbb{E}\big(\Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T\big)\tilde{\mathbf{A}}\Big)\\ &= \frac{\sigma^2}{n^2}\sum_{i=1}^n\big(\tilde{\eta}_i\tilde{\mathbf{z}}_i + \tilde{\psi}_i\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}\big).\end{aligned}$$

Thus premultiplying $\mathbb{E}(\mathbf{R}_1 - \mathbf{R}_2)$ by $\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}$ gives the bias of $\Delta_2\mathbf{A}$. Next we derive the MSE of $\Delta_2\mathbf{A}$. Since $\mathbb{E}(\mathbf{R}_2\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbf{R}_2)$ has $\mathcal{O}(\sigma^4/n^2)$, it will be ignored in our analysis, and as such we are interested in

$$\mathbb{E}(\Delta_2\mathbf{A}\Delta_2\mathbf{A}^T) = \tilde{\mathbf{M}}^-\,\tilde{\mathbf{S}}_{\tilde{\mathbf{N}}}\,\mathbb{E}\Big(\mathbf{R}_1\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbf{R}_1 - 2\mathcal{S}[\mathbf{R}_1\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbf{R}_2]\Big)\,\tilde{\mathbf{S}}_{\tilde{\mathbf{N}}}^T\,\tilde{\mathbf{M}}^- \tag{A.24}$$

up to order $\mathcal{O}(\sigma^4/n^2)$. We commence with writing $\mathbb{E}(\mathbf{R}_1\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbf{R}_1) = \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2$, where

$$\boldsymbol{\Omega}_1 = \frac{1}{n^2}\sum_{i\neq j}^{n}\big[\mathbb{E}(\Delta_2\mathbf{M}_i)\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbb{E}(\Delta_2\mathbf{M}_j)\big] \tag{A.25}$$

$$\boldsymbol{\Omega}_2 = \frac{1}{n^2}\sum_{i=1}^{n}\big[\mathbb{E}(\Delta_2\mathbf{M}_i)\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbb{E}(\Delta_2\mathbf{M}_i)\big], \tag{A.26}$$

We firstly handel $\boldsymbol{\Omega}_1$. Recall $\tilde{\boldsymbol{\Pi}}_i = \tilde{\mathbf{z}}_i\check{\mathbf{e}}_1^T$ (so $\tilde{A}\tilde{\mathbf{z}}_i = \tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}}$), then using the definition of $\tilde{\Psi}_i$ gives

$$\begin{aligned}
\boldsymbol{\Omega}_1 &= \frac{1}{n^2}\sum_{i\neq j}^{n}\big[\mathbb{E}(\Delta_2\mathbf{M}_i)\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\mathbb{E}(\Delta_2\mathbf{M}_j)\big]\\
&= \frac{1}{n^2}\sum_{i\neq j}^{n}\big[\mathbb{E}((\Delta_2\mathbf{z}_i^T\tilde{\mathbf{A}})\tilde{\mathbf{z}}_i + (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})\Delta_1\mathbf{z}_i)\mathbb{E}((\Delta_2\mathbf{z}_j^T\tilde{\mathbf{A}})\tilde{\mathbf{z}}_j^T + (\Delta_1\mathbf{z}_j^T\tilde{\mathbf{A}})\Delta_1\mathbf{z}_j^T)\big]\\
&= \frac{\sigma^4}{n^2}\sum_{i\neq j}^{n}\big[(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}} + \tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(2\tilde{\boldsymbol{\Pi}}_j\tilde{\mathbf{A}} + \tilde{\mathbf{V}}_j\tilde{\mathbf{A}})^T\big]\\
&= \frac{\sigma^4}{n^2}\sum_{i\neq j}^{n}(\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\Psi}_j\tilde{\mathbf{A}})^T.
\end{aligned}$$

Note that $\boldsymbol{\Omega}_1$ is the most important term in MSE (because it is of order σ^4). Next, we verify the final form of $\boldsymbol{\Omega}_2$, which can be written in terms of $\mathbf{B}_i$, $\mathbf{C}_i$, and $\mathbf{D}_i$ for each $i = 1,\dots,n$. By Lemma A.1, we handel $\boldsymbol{\Omega}_2$.

$$\begin{aligned}
\boldsymbol{\Omega}_2 &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\Big[\big((\Delta_2\mathbf{z}_i^T\tilde{\mathbf{A}})\tilde{\mathbf{z}}_i + (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})\Delta_1\mathbf{z}_i\big)\big((\Delta_2\mathbf{z}_i^T\tilde{\mathbf{A}})\tilde{\mathbf{z}}_i^T + (\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}})\Delta_1\mathbf{z}_i^T\big)\Big]\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\Big[\mathbb{E}\big(\mathbf{B}_i + 2\mathcal{S}[\mathbf{C}_i] + \mathbb{E}(\mathbf{D}_i)\big]\\
&= \frac{\sigma^4}{n^2}\sum_{i=1}^{n}\big[2(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}})(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}})^T + 4\mathcal{S}[(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}})^T] + \tilde{\zeta}\tilde{\mathbf{V}}_i + 2(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})^T\big]\\
&= \frac{\sigma^4}{n^2}\sum_{i=1}^{n}\Big[2(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}} + \tilde{\mathbf{V}}_i\tilde{\mathbf{A}})(2\tilde{\boldsymbol{\Pi}}_i\tilde{\mathbf{A}} + \tilde{\mathbf{V}}_i\tilde{\mathbf{A}})^T + \tilde{\zeta}\tilde{\mathbf{V}}_i\Big]\\
&= \frac{\sigma^4}{n^2}\sum_{i=1}^{n}\Big[2(\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\Psi}_i\tilde{\mathbf{A}})^T + \tilde{\zeta}\tilde{\mathbf{V}}_i\Big].
\end{aligned}$$

Now if we combine $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, then we get

$$\mathbb{E}\big((\mathbf{R}_1\tilde{\mathbf{A}})(\mathbf{R}_1\tilde{\mathbf{A}})^T\big) = \frac{\sigma^4}{n^2}\sum_{i=1}^{n}\Big((\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\Psi}_i\tilde{\mathbf{A}})^T + \tilde{\zeta}\tilde{\mathbf{V}}_i + \sum_{j=1}^{n}(\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\Psi}_j\tilde{\mathbf{A}})^T\Big). \tag{A.27}$$

Lastly, we compute $\mathbb{E}((\mathbf{R}_1\tilde{\mathbf{A}})(\mathbf{R}_2\tilde{\mathbf{A}})^T)$.

$$\mathbb{E}\big((\mathbf{R}_1\tilde{\mathbf{A}})(\mathbf{R}_2\tilde{\mathbf{A}})^T\big) = \frac{1}{n^3}\sum_{i,j,k=1}^{n}\mathbb{E}\big(\Delta_2\mathbf{M}_i\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T\Delta_1\mathbf{M}_j\tilde{\mathbf{M}}^-\Delta_1\mathbf{M}_k\big) \tag{A.28}$$

$\Delta_2\mathbf{M}_i$ is a quadratic form of the random vector $\mathbf{h}$, also $\Delta_1\mathbf{M}_i$ is a linear combination of $\mathbf{h}$, hence the expectation of all terms having unequal indices j and k is zero. A more detailed analysis, after excluding the case $i = j$, shows $\mathbb{E}((\mathbf{R}_1\tilde{\mathbf{A}})(\mathbf{R}_2\tilde{\mathbf{A}})^T)$

$$\begin{aligned}
&= \frac{2}{n^3}\sum_{i\neq j}^{n}\mathbb{E}\big[(\Delta_2\mathbf{z}_i^T\tilde{\mathbf{A}})\,\tilde{\mathbf{z}}_i + \Delta_1\mathbf{z}_i\Delta_1\mathbf{z}_i^T\tilde{\mathbf{A}}\big]\mathbb{E}\big[(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j)\,\tilde{\mathbf{z}}_j^T\tilde{\mathbf{M}}^-\mathcal{S}[\Delta_1\mathbf{z}_j\tilde{\mathbf{z}}_j^T]\big] \\
&= \frac{\sigma^2}{n^3}\sum_{i\neq j}^{n}\big[2\tilde{A}\tilde{\mathbf{z}}_i + \tilde{\mathbf{V}}_i\tilde{\mathbf{A}}\big]\big[\mathbb{E}\big(\tilde{\mathbf{A}}^T\Delta_1\mathbf{z}_j\Delta_1\mathbf{z}_j^T\tilde{\mathbf{M}}^-\tilde{\mathbf{M}}_j\big) + \tilde{\psi}_j\tilde{\mathbf{A}}^T\mathbb{E}\big(\Delta_1\mathbf{z}_j\Delta_1\mathbf{z}_j^T\big)\big] \\
&= \frac{\sigma^4}{n^3}\sum_{i\neq j}^{n}(\tilde{\Psi}_i\tilde{\mathbf{A}})\big(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_j\tilde{\mathbf{M}}^-\tilde{\mathbf{M}}_j + \tilde{\psi}_j\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_j\big) \\
&= \frac{\sigma^4}{n^3}\sum_{i\neq j}^{n}(\tilde{\Psi}_i\tilde{\mathbf{A}})\big(\tilde{\eta}_j\tilde{\mathbf{z}}_j + \tilde{\psi}_j\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}\big)^T.
\end{aligned}$$

Finally, substituting the formal expressions of (A.27) and (A.28) in equation (A.24) gives (3.3). Now $\mathrm{cov}(\Delta_2\mathbf{A})$ follows immediately if we subtract the outer product of $\mathrm{Bias}(\Delta_2\mathbf{A})$ represented by

$$\frac{\sigma^4}{n^2}\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\sum_{i,j=1}^{n}\Big((\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\Psi}_j\tilde{\mathbf{A}})^T - \frac{2}{n}\mathcal{S}\big[(\tilde{\Psi}_i\tilde{\mathbf{A}})(\tilde{\psi}_j\tilde{\mathbf{V}}_j\tilde{\mathbf{A}} + \tilde{\eta}_j\tilde{\mathbf{z}}_j)^T\big]\Big)\mathbf{S}_{\tilde{\mathbf{N}}}^T\tilde{\mathbf{M}}^-$$

from $\tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\tilde{\mathbf{\Phi}}\mathbf{S}_{\tilde{\mathbf{N}}}\tilde{\mathbf{M}}^-$. This gives $\mathrm{cov}(\Delta_2\mathbf{A}) = \tilde{\mathbf{M}}^-\mathbf{S}_{\tilde{\mathbf{N}}}\tilde{\mathbf{\Phi}}_2\mathbf{S}_{\tilde{\mathbf{N}}}^T\tilde{\mathbf{M}}^-$, with $\tilde{\mathbf{\Phi}}_2$ given in (3.22). □

References

[1] Al-Sharadqah, A. and Chernov, N. (2009). Error analysis for circle fitting algorithms. *Electronic Journal of Statistics.* **3** 886–911. MR2540845

[2] Al-Sharadqah, A. and Chernov, N. (2011). Statistical analysis of curve fitting methods in errors-in-variables models. *Theor. Probability and Math. Statist.* **84** 4–17 MR2857412

[3] Al-Sharadqah, A. and Chernov, N. (2012). A doubly optimal ellipse fit. *Computational Statistics and Data Analysis* **56** 2771–2781. MR2915161

[4] Al-Sharadqah, A. and Chernov, N. and Kanatani, K. and Sugaya, Y. (2012). Renormalization returns: Hyper-renormalization and its applications. In *European Conference on Computer Vision (ECCV 2012)*, **3**, 385–398, Firenze, Italy, October 7–13.

[5] Anderson, T. W. (1976). Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometric. *J. R. Statist. Soc. B.* **38** 1–36. MR0411025
[6] Anderson, T. W. and Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient. *J. R. Statist. Soc. B.* **44** 52–62. MR0655374
[7] Atieg, A. and Watson, G. A. (2004). Fitting circular arcs by orthogonal distance regression. *Appl. Numer. Anal. Comput. Math.* **1** 66–76. MR2168317
[8] Chan, N. N. (1965). On circular functional relationships. *J. R. Statist. Soc. B.* **27** 45–56. MR0189163
[9] Chan, Y. T. and Thomas, S. M. (1995). Cramer-Rao lower bounds for estimation of a circular arc center and its radius. *Graph. Models Image Proc.* **57** 527–532.
[10] Chernov, N. (2010). Circular and linear regression: Fitting circles and lines by least squares, volume 117. *Chapman & Hall/CRC, Boca Raton*. MR2723019
[11] Chernov, N. and Lesort, C. (2004). Statistical efficiency of curve fitting algorithms. *Comp. Stat. Data Anal.* **47** 713–728. MR2101548
[12] Chernov, N. and Lesort, C. (2005). Least squares fitting of circles. *J. Math. Imag. Vision* **23** 239–251. MR2181705
[13] Chernov, N. and Sapirstein, P. (2008). Fitting circles to data with correlated noise. *Comput. Statist. Data Anal.* **52** 5328–5337. MR2526598
[14] Cooke, T. (2010). A fast automatic ellipse detector. *Digital Image Computing: Techniques and Applications (DICTA).*
[15] Delogne, P. (1972). Computer optimization of Deschamps' method and error cancellation in reflectometry. *Proc. IMEKO-Symp. Microwave Measurement (Budapest)*, 117–123.
[16] Fuller, W. A. (1987). Measurement Error Models. *Wiley & Son, New York*. MR0898653
[17] Joseph, S. H. (1994). Unbiased least-squares fitting of circular arcs. *Graph. Mod. Image Process* **56** 424–432.
[18] Kadane, J. B. (1970). Testing overidentifying restrictions when the disturbances are small. *J. Amer. Statist. Assoc.* **65** 182–185.
[19] Kanatani, K. (1998). Cramer-Rao lower bounds for curve fitting. *Graph. Mod. Image Process* **60** 93–99.
[20] Kanatani, K. (2004). For geometric inference from images: What kind of statistical model is necessary? *Syst. Comp. Japan* **35** 1–9.
[21] Kanatani, K. (2005). Optimality of maximum likelihood estimation for geometric fitting and the KCR lower bound. *Memoirs Fac. Engin. Okayama Univ.* **39** 63–70.
[22] Kanatani, K. (2008). Statistical optimization for geometric fitting: Theoretical accuracy bound and high order error analysis. *Int. J. Computer Vision* **80** 167–188.
[23] Kanatani, K. and Rangarajan, P. (2009). Improved algebraic methods for circle fitting. *Electr. J. Statist.* **3** 1075–1082. MR2557129

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

[25] Kilambi, S. and Tipton, S. (2011). Development of a laser scan inspection tool for coiled tubing. *SPE/ICoTA Coiled Tubing & Well Intervention Conference and Exhibition*.

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

[27] Pratt, V. (1987). Direct least-squares fitting of algebraic surfaces. *Computer Graphics* **21** 145–152. MR0987653

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

[29] Späth, H. (1996). Least-squares fitting by circles. *Computing* **57** 179–185. MR1407351

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

[31] 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 Analysis Machine Intelligence* **13** 1115–1138.

[32] Teixidó, M. (2012). Two-dimensional radial laser scanning for circular marker detection and external mobile robot tracking. *Sensors* **12** 16482–16497.

[33] Wang, T., Li, B. and Wu, T. (2010). Study on pre-correction method for led chip declination in sorting system. *International Conference on Mechatronics and Automation* 1110–1114.

[34] Wu, T. (2010). Study on auto-path planning according to grade priority for sorting dies. *International Conference on Machine Learning and Cybernetics (ICMLC)* **3**.

[35] Yin, S. J. and Wang, S. G. (2004). Estimating the parameters of a circle by heteroscedastic regression models. *J. Statist. Planning Infer.* **124** 439–451. MR2081496

[36] Zelniker, E. and Clarkson, V. (2006). Maximum-likelihood estimation of circle parameters via convolution. *IEEE Trans. Image Proc.* **15** 865–876. MR2479910

[37] Zelniker, E. and Clarkson, V. (2006). A statistical analysis of the Delogne-Kåsa method for fitting circles. *Digital Signal Proc.* **16** 498–552.

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

[39] Zhao, Y. and Xiong, C. (2011). Prediction-based geometric feature extraction for 2nd laser scanner. *Robotics and Autonomous Systems* **59(6)** 402–409.