A doubly optimal ellipse fit

A. Al-Sharadqah[1], N. Chernov[2]

Abstract

We study the problem of fitting ellipses to observed points in the context of Errors-In-Variables regression analysis. The accuracy of fitting methods is characterized by their variances and biases. The variance has a theoretical lower bound (the KCR bound), and many practical fits attend it, so they are optimal in this sense. There is no lower bound on the bias, though, and in fact our higher order error analysis (developed just recently) shows that it can be eliminated, to the leading order. Kanatani and Rangarajan recently constructed an algebraic ellipse fit that has no bias, but its variance exceeds the KCR bound; so their method is optimal only relative to the bias. We present here a novel ellipse fit that enjoys both optimal features: the theoretically minimal variance and zero bias (both to the leading order). Our numerical tests confirm the superiority of the proposed fit over the existing fits.

Keywords: Errors-In-Variables regression, ellipse fitting, conic fitting, Cramer-Rao bound, bias reduction.

1 Introduction

Fitting geometric contours such as ellipses to observed points is a major task in computer vision, pattern recognition, and image processing applications [1, 3, 5, 8, 10, 12, 17, 22]. We denote the observed points by (x_1, y_1), ..., (x_n, y_n) and describe ellipses (conics) by quadratic equation

$$P(x, y; \mathbf{A}) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \tag{1}$$

[1]Department of Mathematics, University of Mississippi, University, MS 38677;aal-shara@olemiss.edu

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

where $\mathbf{A} = (A, B, C, D, E, F)^T$ is the vector of parameters to be estimated.

One can find the best fitting ellipse by minimizing geometric distances

$$\mathcal{F}_{\mathrm{g}}(\mathbf{A}) = \frac{1}{n}\sum_{i=1}^{n} d_i^2 \;\longrightarrow\; \min \tag{2}$$

where d_i is the orthogonal distance from the data point (x_i, y_i) to the ellipse $P(x, y; \mathbf{A}) = 0$. This method is called *geometric fit* or orthogonal distance regression (ODR). While highly praised in the literature for its accuracy [1, 17], it is often computationally burdensome – many practical minimization algorithms converge slowly or tend to diverge.

A popular alternative is to minimize simple algebraic distances

$$\mathcal{F}_{\mathrm{a}}(\mathbf{A}) = \tfrac{1}{n}\sum_{i=1}^{n}[P(x_i, y_i; \mathbf{A})]^2, \tag{3}$$

this method is called algebraic fit. We have $P(x_i, y_i; \mathbf{A}) = \mathbf{A}^T\mathbf{Z}_i$, where

$$\mathbf{Z}_i = (x_i^2, x_i y_i, y_i^2, x_i, y_i, 1)^T \tag{4}$$

is the "data" vector. Therefore

$$\mathcal{F}_{\mathrm{a}}(\mathbf{A}) = \mathbf{A}^T\mathbf{M}\mathbf{A}, \quad \text{where } \mathbf{M} = \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{M}_i = \tfrac{1}{n}\sum_{i=1}^{n}\mathbf{Z}_i\mathbf{Z}_i^T. \tag{5}$$

To avoid the unwanted solution $\mathbf{A} = 0$ one usually imposes a constraint $\mathbf{A}^T\mathbf{N}\mathbf{A} = 1$, where $\mathbf{N}$ is a certain symmetric matrix (which may or may not depend on the data points). The corresponding constrained minimization problem reduces to solving equation $\mathbf{M}\mathbf{A} = \lambda\mathbf{N}\mathbf{A}$, i.e., $\mathbf{A}$ is a generalized eigenvector for the matrix pencil $(\mathbf{M}, \mathbf{N})$; see [2, 12, 20]. Thus algebraic fit admits a fast non-iterative solution. It is, however, statistically inaccurate and often heavily biased, because algebraic distances $|P(x_i, y_i; \mathbf{A})|$ may be very different from geometric distances d_i. At best, the algebraic fit gives us a good initial guess for a subsequent iterative fitting procedure.

A popular improvement of the simple algebraic fit is to minimize "gradient weighted" algebraic distances:

$$\mathcal{F}_{\mathrm{w}}(\mathbf{A}) = \sum_{i=1}^{n}\frac{[P(x_i, y_i; \mathbf{A})]^2}{\|\nabla P(x_i, y_i; \mathbf{A})\|^2},$$

where $\nabla P = (\partial P/\partial x, \partial P/\partial y)$ denotes the gradient vector. The rational here is that $|P(x_i, y_i; \mathbf{A})|/\|\nabla P(x_i, y_i; \mathbf{A})\| = d_i + o(d_i)$, by Taylor expansion. This formula is also known as "Sampson error" (Sampson [23] was first to apply it to ellipses). By direct calculation $\|\nabla P(x_i, y_i; \mathbf{A})\|^2 = \mathbf{A}^T \mathbf{V}_i \mathbf{A}$, where

$$\mathbf{V}_i = \begin{bmatrix} 4x_i^2 & 2x_i y_i & 0 & 2x_i & 0 & 0 \\ 2x_i y_i & x_i^2 + y_i^2 & 2x_i y_i & y_i & x_i & 0 \\ 0 & 2x_i y_i & 4y_i^2 & 0 & 2y_i & 0 \\ 2x_i & y_i & 0 & 1 & 0 & 0 \\ 0 & x_i & 2y_i & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

is a symmetric matrix, hence

$$\mathcal{F}_{\mathrm{w}}(\mathbf{A}) = \sum_{i=1}^{n} \frac{\mathbf{A}^T \mathbf{M}_i \mathbf{A}}{\mathbf{A}^T \mathbf{V}_i \mathbf{A}} \tag{7}$$

is the sum of rational functions of $\mathbf{A}$. Note that $\mathcal{F}(\mathbf{A})$ is invariant under rescaling $\mathbf{A} \mapsto c\mathbf{A}$, thus no additional constraints on $\mathbf{A}$ are required, and we will simply set $\|\mathbf{A}\| = 1$. Differentiating (7) with respect to $\mathbf{A}$ gives

$$\mathcal{M}\mathbf{A} = \mathcal{L}\mathbf{A} \tag{8}$$

where we use the following notation:

$$\mathcal{M} = \sum_{i=1}^{n} \gamma_i^{-1} \mathbf{M}_i \qquad \text{with} \qquad \gamma_i = \mathbf{A}^T \mathbf{V}_i \mathbf{A} \tag{9}$$

and

$$\mathcal{L} = \sum_{i=1}^{n} \gamma_i^{-2} (\mathbf{A}^T \mathbf{M}_i \mathbf{A}) \mathbf{V}_i. \tag{10}$$

Thus $\mathbf{A}$ is a generalized eigenvector for the matrix pencil $(\mathcal{M}, \mathcal{L})$. We call the solution of (8) the *gradient weighted algebraic fit*, or *GRAF* for brevity.

Since both $\mathcal{M}$ and $\mathcal{L}$ depend on $\mathbf{A}$, Eq. (8) can only be solved by iterative procedures. Various practical solutions were proposed, including FNS (Fundamental Numerical Scheme) [8] and HEIV (Heteroscedastic Error-In-Variables) [22]. These schemes may occasionally diverge, but they are usually simpler and faster than the geometric fit (2).

The matrix $\mathcal{M}$ in (9) is usually of order 1, but the matrix $\mathcal{L}$ is small, $\mathcal{L} = \mathcal{O}(\sigma^2)$ where σ denotes the noise level (see Section 2). Hence $\mathcal{L}$ is of lesser significance, and some authors [4, 14, 16] propose to drop $\mathcal{L}$ and solve

$$\mathcal{M}\mathbf{A} = \mathbf{0}. \tag{11}$$

We call (11) the *reduced GRAF*. Other authors propose to replace $\mathcal{L}$ in (9) with some matrix $\mathcal{N}$ which may depend on $\mathbf{A}$, too. Then one finds $\mathbf{A}$ as a unit generalized eigenvector for the matrix pencil $(\mathcal{M}, \mathcal{N})$, i.e., by solving

$$\mathcal{M}\mathbf{A} = \lambda\mathcal{N}\mathbf{A}, \tag{12}$$

see surveys [4, 7] and Section 3. We call (12) *generalized GRAF*. We note that (11)–(12) do not minimize any specific function, so in case of multiple solutions the choice of the "right" one is not obvious. For (12), one usually chooses the eigenvector $\mathbf{A}$ corresponding to the smallest eigenvalue λ.

The accuracy of a statistical estimator $\hat{\mathbf{A}}$ of the parameter vector $\mathbf{A}$ can be described by its variance and bias. If the noise level is represented by σ (see precise definitions in Section 2), then

$$\mathsf{Var}(\hat{\mathbf{A}}) = \mathbb{E}\big[(\hat{\mathbf{A}} - \mathbb{E}(\mathbf{A}))(\hat{\mathbf{A}} - \mathbb{E}(\mathbf{A}))^T\big] = \sigma^2\mathbf{K} + \mathcal{O}(\sigma^4) \tag{13}$$

where $\mathbf{K}$ is a positive semidefinite matrix and

$$\mathrm{bias}(\hat{\mathbf{A}}) = \mathbb{E}(\hat{\mathbf{A}} - \tilde{\mathbf{A}}) = \sigma^2\mathbf{B} + \mathcal{O}(\sigma^4)$$

where $\mathbf{B}$ is a vector (here $\tilde{\mathbf{A}}$ denotes the true value of $\mathbf{A}$). The mean square error is given by

$$\begin{aligned} \mathrm{MSE}(\hat{\mathbf{A}}) &= \mathbb{E}\big[(\hat{\mathbf{A}} - \tilde{\mathbf{A}})(\hat{\mathbf{A}} - \tilde{\mathbf{A}})^T\big] \\ &= \mathsf{Var}(\hat{\mathbf{A}}) + \mathrm{bias}(\hat{\mathbf{A}})\mathrm{bias}(\hat{\mathbf{A}})^T \\ &= \sigma^2\mathbf{K} + \sigma^4\mathbf{B}\mathbf{B}^T + \cdots, \end{aligned} \tag{14}$$

hence $\mathbf{K}$ is the most important characteristic of the estimator $\hat{\mathbf{A}}$. It satisfies the Kanatani-Cramer-Rao (KCR) lower bound $\mathbf{K} \geq \mathbf{K}_{\min}$ (in the sense that $\mathbf{K} - \mathbf{K}_{\min}$ is a positive semidefinite matrix). In fact, $\mathbf{K}_{\min} = \tilde{\mathcal{M}}^-$ is the Moore-Penrose pseudoinverse of the matrix $\tilde{\mathcal{M}}$, which is the "true" version of the random matrix $\mathcal{M}$, as defined in Section 2 (see proofs in [6, 12]).

We remark that due to the indeterminacy of $\mathbf{A}$ (which is only determined up to a scalar multiple) and our setting $\|\mathbf{A}\| = 1$, all statistical variations

of $\mathbf{A}$ occur in the hyperplane orthogonal to $\tilde{\mathbf{A}}$, hence $\mathbf{K}$ is singular with $\mathrm{Ker}(\mathbf{K}) = \mathrm{span}(\tilde{\mathbf{A}})$. For the same reason, $\mathrm{Ker}(\tilde{\mathcal{M}}) = \mathrm{span}(\tilde{\mathbf{A}})$; see Section 2.

Now for the geometric fit (2) we have $\mathbf{K} = \mathbf{K}_{\min}$, i.e., the estimator is optimal to the leading order. The same is true for GRAF (7) and for generalized GRAF (12), i.e., they are optimal, too (see proofs in [6] and in Section 3). But this is *not* true for the algebraic fit (3). The corresponding matrix $\mathbf{K}$ is independent of $\mathbf{N}$ and strictly greater than $\mathbf{K}_{\min}$; see [19].

On the other hand, there is no lower bound for the bias leading term $\mathbf{B}$. In fact, in the context of the algebraic fit one can choose the constraint matrix $\mathbf{N}$ so that $\mathbf{B} = 0$; see [19].

This gives the best possible algebraic fit, though it is still not fully optimal, as its leading matrix $\mathbf{K}$ exceeds $\mathbf{K}_{\min}$.

Our goal is to find a matrix $\mathcal{N}$ for the generalized GRAF (12) so that its leading bias will vanish, i.e., $\mathbf{B} = 0$. At the same time its leading variance $\mathbf{K}$ will remain equal to $\mathbf{K}_{\min}$. Thus we will obtain a *doubly optimal* fit, which has the minimal possible variance, $\mathbf{K} = \mathbf{K}_{\min}$, and the minimal possible bias, $\mathbf{B} = 0$. To our best knowledge, it is the only fit combining both features.

Our numerical tests show that our fit is practically more accurate than some popular versions of GRAF. At the same time it is much faster and simpler than the geometric fit.

Lastly, we must admit that our expansion (14) should also contain terms of order σ^4 coming from $\mathsf{Var}(\mathbf{A})$, but those are of lesser significance. If one takes into account n, the number of data points, then those missing terms are of order σ^4/n, while $\mathrm{bias}(\hat{\mathbf{A}})\mathrm{bias}(\hat{\mathbf{A}})^T$ is of order σ^4; see [2] and [5, Chapter 6]. In all the relevant publications [2, 5, 18, 19, 20], the σ^4 terms coming from $\mathsf{Var}(\mathbf{A})$ are ignored, and we follow this tradition.

For the reader's convenience, we list our main symbols below, with references to an equation where they are defined:

$\mathbf{e}_{13}$, see (18)	$\tilde{\mathbf{G}}_1$, see (30)	$\tilde{\mathbf{G}}_2$, see (31)	$\tilde{\mathbf{G}}_1^*$, see (42)
$\mathbf{K}$, see (13)-(14)	$\mathcal{L}$, see (10)	$\mathbf{M}, \mathbf{M}_i$, see (5)	$\mathcal{M}$, see (9)
$\mathcal{N}$, see (12)	$\mathbf{R}$, see (28)	$\mathbf{V}_i$, see (6)	$\mathbf{Z}_i$, see (4)
γ_i, see (9)	λ, see (12)	ε_i, see (15)	δ_i, see (15)
$\tilde{\psi}_i$, see (32)	$\tilde{\mathbf{\Gamma}}_i$, see (33)		

2 Error Analysis of GRAF

It is a standard assumption in Errors-In-Variables (EIV) analysis that the observed points (x_i, y_i) are random perturbations of some unknown true points $(\tilde{x}_i, \tilde{y}_i)$ which lie on the (unknown) true conic, i.e., satisfy $P(\tilde{x}_i, \tilde{y}_i, \tilde{\mathbf{A}}) = 0$ for all $i = 1, \ldots, n$. In other words,

$$x_i = \tilde{x}_i + \delta_i \quad \text{and} \quad y_i = \tilde{y}_i + \varepsilon_i, \tag{15}$$

where δ_i and ε_i are small random errors (noise). The pairs $(\delta_i, \varepsilon_i)$ are commonly assumed to be independent normal vectors with zero mean.

We will further assume that the noise is homogeneous and isotropic, i.e., δ_i's and ε_i's are independent and have a common variance σ^2, i.e., $\mathsf{Cov}(\delta_i, \varepsilon_i) = 0$ and $\delta_i \sim \mathbb{N}(0, \sigma^2)$ and $\varepsilon_i \sim \mathbb{N}(0, \sigma^2)$. Now σ represents the noise level.

The true points $(\tilde{x}_i, \tilde{y}_i)$ can be thought of as fixed and treated as nuisance parameters; this is known as *functional model*. Alternatively $\tilde{x}$'s and $\tilde{y}$'s may be treated as realizations of some random variables; this is known as *structural model*. The functional model is more suitable for image processing applications [17], and we adopt it here. Our analysis is based on treating σ as a small parameter and expanding all variables into Taylor series; see a detailed presentation in [2] and [5, Chapter 6]. In particular, the random vector $\mathbf{Z}_i$ in (4) can be written as

$$\mathbf{Z}_i = \tilde{\mathbf{Z}}_i + \Delta_1 \mathbf{Z}_i + \Delta_2 \mathbf{Z}_i + \mathcal{O}_{\mathrm{P}}(\sigma^3), \tag{16}$$

where $\tilde{\mathbf{Z}}_i = (\tilde{x}_i^2, \tilde{x}_i\tilde{y}_i, \tilde{y}_i^2, \tilde{x}_i, \tilde{y}_i, 1)^T$ and the symbol $\mathcal{O}_{\mathrm{P}}$ denotes the order of magnitude in probability. The first order error $\Delta_1 \mathbf{Z}_i$ is a linear combination of δ_i's and ε_i's, and it is easy to see that

$$\Delta_1 \mathbf{Z}_i = \nabla \mathbf{Z}_i (\delta_i, \varepsilon_i)^T = \begin{bmatrix} 2\tilde{x}_i & \tilde{y}_i & 0 & 1 & 0 & 0 \\ 0 & \tilde{x}_i & 2\tilde{y}_i & 0 & 1 & 0 \end{bmatrix}^T \begin{bmatrix} \delta_i \\ \varepsilon_i \end{bmatrix} \tag{17}$$

where $\nabla \mathbf{Z}_i$ is the gradient matrix evaluated at the true points $(\tilde{x}_i, \tilde{y}_i)$. Accordingly, the variance matrix $\mathsf{Var}(\Delta_1 \mathbf{Z}_i)$ is equal to $\sigma^2 \tilde{\mathbf{V}}_i$, where $\tilde{\mathbf{V}}_i = \nabla \mathbf{Z}_i \nabla \mathbf{Z}_i^T$ is just the "true" version of the random matrix (6). The second order term $\Delta_2 \mathbf{Z}_i = (\delta_i^2, \delta_i\varepsilon_i, \varepsilon_i^2, 0, 0)^T$ has mean $\mathbb{E}(\Delta_2 \mathbf{Z}_i) = \sigma^2 \mathbf{e}_{13}$, where for brevity we denote

$$\mathbf{e}_{13} = (1, 0, 1, 0, 0, 0)^T \tag{18}$$

In our formulas, the tilde will always denote the true value of a random variable, obtained by substituting $\tilde{x}_i$ and $\tilde{y}_i$ for x_i and y_i (and $\tilde{\mathbf{A}}$ for $\mathbf{A}$). In particular, the minimal variance matrix $\mathbf{K}_{\min}$ in the KCR lower bound is $\mathbf{K}_{\min} = \tilde{\mathcal{M}}^-$, where $\tilde{\mathcal{M}}$ is the so constructed "true" version of the random matrix $\mathcal{M}$ defined by (9).

We will expand various functions of x_i's and y_i's into Taylor series up to quadratic terms around the true values $\tilde{x}_i$ and $\tilde{y}_i$. Since our noise is of order σ, we will always assume that σ is less than the radius of convergence. Our experiments reported in Section 5 indicate that for small σ the influence of higher order terms is negligible.

Now the Taylor expansion of $\gamma_i^{-1} = (\mathbf{A}^T\mathbf{V}_i\mathbf{A})^{-1}$ is

$$\gamma_i^{-1} = \tilde{\gamma}_i^{-1}\big[1 - \tilde{\gamma}_i^{-1}(\Delta_1\gamma_i + \Delta_2\gamma_i) + \tilde{\gamma}_i^{-2}(\Delta_1\gamma_i)^2\big] + \mathcal{O}_P(\sigma^3) \tag{19}$$

and that of $\mathbf{M}_i = \mathbf{Z}_i\mathbf{Z}_i^T$ is $\mathbf{M}_i = \tilde{\mathbf{M}}_i + \Delta_1\mathbf{M}_i + \Delta_2\mathbf{M}_i + \mathcal{O}_P(\sigma^3)$, where

$$\begin{aligned} \Delta_1\mathbf{M}_i &= \tilde{\mathbf{Z}}_i\,\Delta_1\mathbf{Z}_i^T + \Delta_1\mathbf{Z}\,\tilde{\mathbf{Z}}_i^T \\ \Delta_2\mathbf{M}_i &= \Delta_1\mathbf{Z}\,\Delta_1\mathbf{Z}_i^T + \Delta_2\mathbf{Z}_i\tilde{\mathbf{Z}}_i^T + \tilde{\mathbf{Z}}_i\,\Delta_2\mathbf{Z}_i^T. \end{aligned} \tag{20}$$

Thus expanding $\mathcal{M} = \sum_{i=1}^n \gamma_i^{-1}\mathbf{M}_i$ gives

$$\mathcal{M} = \tilde{\mathcal{M}} + \Delta_1^1\mathcal{M} + \Delta_1^2\mathcal{M} + \Delta_2^1\mathcal{M} + \Delta_2^2\mathcal{M} + \Delta_2^3\mathcal{M} + \mathcal{O}_{\mathrm{P}}(\sigma^3),$$

where $\Delta_1^1\mathcal{M}$ etc. denote various "parts" of $\Delta_1\mathcal{M}$ and $\Delta_2\mathcal{M}$ defined as follows:

$$\begin{aligned} \Delta_1^1\mathcal{M} &= \sum_{i=1}^n \tilde{\gamma}_i^{-1}\,\Delta_1\mathbf{M}_i, \quad \Delta_1^2\mathcal{M} = -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\Delta_1\gamma_i\,\tilde{\mathbf{M}}_i, \\ \Delta_2^1\mathcal{M} &= \sum_{i=1}^n \tilde{\gamma}_i^{-1}\,\Delta_2\mathbf{M}_i, \quad \Delta_2^2\mathcal{M} = -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\Delta_1\gamma_i\Delta_1\mathbf{M}_i, \\ \Delta_2^3\mathcal{M} &= \sum_{i=1}^n \tilde{\gamma}_i^{-2}\big[-\Delta_2\gamma_i + \tilde{\gamma}_i^{-1}(\Delta_1\gamma_i)^2\big]\tilde{\mathbf{M}}_i. \end{aligned}$$

Since $\mathbf{A}^T\mathbf{M}_i\mathbf{A} = (\mathbf{Z}_i^T\mathbf{A})^2$ and $\tilde{\mathbf{Z}}_i^T\tilde{\mathbf{A}} = 0$, we have $\mathbf{A}^T\mathbf{M}_i\mathbf{A} \sim \mathcal{O}_P(\sigma^2)$, and hence $\tilde{\mathcal{L}} = \Delta_1\mathcal{L} = \mathbf{0}$, thus $\mathcal{L} = \Delta_2\mathcal{L} + \mathcal{O}_{\mathrm{P}}(\sigma^3)$. In fact,

$$\Delta_2\mathcal{L} = \sum_{i=1}^n \tilde{\gamma}_i^{-2}\big[(\Delta_1\,\mathbf{Z}_i^T\tilde{\mathbf{A}})^2 + (\tilde{\mathbf{Z}}_i^T\Delta_1\,\mathbf{A})^2 + 2\tilde{\mathbf{Z}}_i^T\Delta_1\,\mathbf{A}\Delta_1\,\mathbf{Z}_i^T\tilde{\mathbf{A}}\big]\tilde{\mathbf{V}}_i. \tag{21}$$

Now the perturbation of equation (8) is

$$
\begin{aligned}
&(\tilde{\mathcal{M}} + \Delta_1\mathcal{M} + \Delta_2\mathcal{M} + \cdots)(\tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \Delta_2\,\mathbf{A} + \cdots) \\
&\qquad = (\Delta_2\mathcal{L} + \cdots)(\tilde{\mathbf{A}} + \Delta_1\mathbf{A} + \Delta_2\,\mathbf{A} + \cdots).
\end{aligned} \tag{22}
$$

Equating terms according to their order (in σ) gives us

$$
\begin{aligned}
\tilde{\mathcal{M}}\Delta_1\mathbf{A} + \Delta_1\mathcal{M}\tilde{\mathbf{A}} &= \mathbf{0}, \\
\tilde{\mathcal{M}}\Delta_2\mathbf{A} + \Delta_1\mathcal{M}\Delta_1\mathbf{A} + \Delta_2\mathcal{M}\tilde{\mathbf{A}} &= \Delta_2\mathcal{L}\tilde{\mathbf{A}}.
\end{aligned}
$$

Assuming that there are at least five distinct true points on the true ellipse, the kernel of $\tilde{\mathcal{M}}$ is one dimensional, i.e., $\text{kernel}(\tilde{\mathcal{M}}) = \text{span}(\tilde{\mathbf{A}})$.

Note that $\|\mathbf{A}\| = \|\tilde{\mathbf{A}}\| = 1$, hence

$$
2\tilde{\mathbf{A}}^T\Delta_1\mathbf{A} + 2\tilde{\mathbf{A}}^T\Delta_2\mathbf{A} + \|\Delta_1\mathbf{A}\|^2 + \mathcal{O}_P(\sigma^3) = 0. \tag{23}
$$

Hence $\tilde{\mathbf{A}}^T\Delta_1\mathbf{A} = 0$, i.e., $\Delta_1\mathbf{A}$ is orthogonal to $\tilde{\mathbf{A}}$, and we can write

$$
\Delta_1\mathbf{A} = -\tilde{\mathcal{M}}^-\Delta_1\,\mathcal{M}\tilde{\mathbf{A}}. \tag{24}
$$

Recalling that $\tilde{\mathbf{Z}}_i^T\tilde{\mathbf{A}} = 0$ we rewrite (24) as

$$
\Delta_1\mathbf{A} = -\tilde{\mathcal{M}}^-\Delta_1^1\,\mathcal{M}\tilde{\mathbf{A}} = -\tilde{\mathcal{M}}^-\sum_{i=1}^{n}\tilde{\gamma}_i^{-1}(\tilde{\mathbf{A}}^T\Delta_1\tilde{\mathbf{Z}}_i)\tilde{\mathbf{Z}}_i. \tag{25}
$$

This gives us the variance of $\mathbf{A}$, to the leading order:

$$
\begin{aligned}
\mathsf{Var}(\Delta_1\mathbf{A}) &= \tilde{\mathcal{M}}^-\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\big[\tilde{\mathbf{A}}^T\mathbb{E}\big(\Delta_1\,\mathbf{Z}_i\Delta_1\,\mathbf{Z}_i^T\big)\tilde{\mathbf{A}}\big]\tilde{\mathbf{Z}}_i\tilde{\mathbf{Z}}_i^T\mathcal{M}^- \\
&= \sigma^2\tilde{\mathcal{M}}^-\sum_{i=1}^{n}\tilde{\gamma}_i^{-1}\tilde{\mathbf{M}}_i\mathcal{M}^- \\
&= \sigma^2\tilde{\mathcal{M}}^-, \qquad (26)
\end{aligned}
$$

where we used the identity $\tilde{\mathcal{M}}^-\tilde{\mathcal{M}}\tilde{\mathcal{M}}^- = \tilde{\mathcal{M}}^-$. Thus $\mathsf{Var}(\Delta_1\mathbf{A})$ achieves the KCR lower bound. Since (26) does not involve $\mathcal{L}$, it applies to the reduced GRAF (11) as well, in which case $\Delta_2\mathcal{L} = 0$. Thus both fits, the GRAF and the reduced GRAF, are statistically optimal in the sense of variance.

We now turn to the bias of $\mathbf{A}$. As (23) implies, $2\tilde{\mathbf{A}}^T\Delta_2\mathbf{A} + \|\Delta_1\mathbf{A}\|^2 = 0$, hence $\Delta_2\mathbf{A}$ is *not* orthogonal to $\tilde{\mathbf{A}}$. We decompose $\Delta_2\mathbf{A} = \Delta_2^{\|}\mathbf{A} + \Delta_2^{\perp}\mathbf{A}$ into the components parallel and orthogonal to $\tilde{\mathbf{A}}$. Then $\Delta_2^{\|}\mathbf{A} = -\frac{1}{2}\|\Delta_1\mathbf{A}\|^2\tilde{\mathbf{A}}$ and $\mathbb{E}(\Delta_2^{\|}\mathbf{A}) = -\frac{1}{2}\sigma^2(\operatorname{tr}\mathcal{M}^-)\tilde{\mathbf{A}}$, which really accounts for the curvature of the unit sphere $\|\mathbf{A}\| = 1$ rather than represents the bias of $\mathbf{A}$. Our goal will be to evaluate $\Delta_2^{\perp}\mathbf{A}$ and $\mathbb{E}(\Delta_2^{\perp}\mathbf{A})$, and we will suppress the $\perp$ sign for brevity.

Since $\Delta_1^2\mathcal{M}\tilde{\mathbf{A}} = \Delta_2^3\mathcal{M}\tilde{\mathbf{A}} = \mathbf{0}$, we have

$$\begin{aligned}\Delta_2\mathbf{A} &= \tilde{\mathcal{M}}^-\big(\Delta_2\mathcal{L}\tilde{\mathbf{A}} - \Delta_1\mathcal{M}\Delta_1\mathbf{A} - \Delta_2\mathcal{M}\tilde{\mathbf{A}}\big)\\ &= \tilde{\mathcal{M}}^-\big(\Delta_2\mathcal{L} - \mathbf{R}\big)\tilde{\mathbf{A}},\end{aligned} \tag{27}$$

where we denote for brevity

$$\mathbf{R} = \Delta_2^1\mathcal{M} + \Delta_2^2\mathcal{M} - \big(\Delta_1^1\mathcal{M} + \Delta_1^2\mathcal{M}\big)\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}. \tag{28}$$

For the reduced GRAF, we have $\Delta_2\mathcal{L} = \mathbf{0}$, hence $\Delta_2\mathbf{A}_{\mathrm{Red}} = -\tilde{\mathcal{M}}^-\mathbf{R}\tilde{\mathbf{A}}$. The following formula will be proved in Appendix:

$$\begin{aligned}\mathbb{E}(\mathbf{R})\tilde{\mathbf{A}} &= \sigma^2(\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2)\tilde{\mathbf{A}}\\ &= \sigma^2(\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2 + \tilde{\mathbf{G}}_2^T)\tilde{\mathbf{A}},\end{aligned} \tag{29}$$

where

$$\tilde{\mathbf{G}}_1 = \sum_{i=1}^{n}(\tilde{\gamma}_i^{-1} - \tilde{\gamma}_i^{-2}\tilde{\psi}_i)\tilde{\mathbf{V}}_i \tag{30}$$

and

$$\tilde{\mathbf{G}}_2 = \sum_{i=1}^{n}\big[\tilde{\gamma}_i^{-1}\tilde{\mathbf{Z}}_i\mathbf{e}_{13}^T - \tilde{\gamma}_i^{-2}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i + \tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{\Gamma}}_i - \tilde{\gamma}_i^{-2}\tilde{\mathbf{\Gamma}}_i\big]. \tag{31}$$

Here

$$\tilde{\psi}_i = \tilde{\mathbf{Z}}_i^T\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_i \tag{32}$$

and

$$\tilde{\mathbf{\Gamma}}_i = (\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}})\tilde{\mathbf{Z}}_i\tilde{\mathbf{a}}_i^T + (\tilde{\mathbf{A}}^T\tilde{\mathbf{S}}_i\tilde{\mathbf{A}})\tilde{\mathbf{Z}}_i\tilde{\mathbf{b}}_i^T, \tag{33}$$

where $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ are the first and the second columns of $\nabla\mathbf{Z}_i$, see (17)

(34)
$$\tilde{\mathbf{T}}_i = \begin{bmatrix} 8\tilde{x}_i & 2\tilde{y}_i & 0 & 2 & 0 & 0 \\ 2\tilde{y}_i & 2\tilde{x}_i & 2\tilde{y}_i & 0 & 1 & 0 \\ 0 & 2\tilde{y}_i & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \tilde{\mathbf{S}}_i = \begin{bmatrix} 0 & 2\tilde{x}_i & 0 & 0 & 0 & 0 \\ 2\tilde{x}_i & 2\tilde{y}_i & 2\tilde{x}_i & 1 & 0 & 0 \\ 0 & 2\tilde{x}_i & 8\tilde{y}_i & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now (29) gives us the bias of the reduced GRAF (11):

$$\mathbb{E}(\Delta_2\mathbf{A}_{\text{Red}}) = -\sigma^2\tilde{\mathcal{M}}^-(\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2)\tilde{\mathbf{A}} \tag{35}$$

To find the bias of GRAF (8) we need $\mathbb{E}(\Delta_2\mathcal{L})$, which was computed by Kanatani [16, pages 181–182]; it can also be easily derived from (21):

$$\mathbb{E}(\Delta_2\mathcal{L}) = \sigma^2\sum_{i=1}^{n}(\tilde{\gamma}_i^{-1} - \tilde{\gamma}_i^{-2}\tilde{\psi}_i)\tilde{\mathbf{V}}_i.$$

Thus adding $\tilde{\mathcal{M}}^-\mathbb{E}(\Delta_2\mathcal{L})\tilde{\mathbf{A}}$ to (35) will cancel $\tilde{\mathbf{G}}_1$, hence

$$\mathbb{E}(\Delta_2\mathbf{A}_{\text{GRAF}}) = -\sigma^2\tilde{\mathcal{M}}^-\tilde{\mathbf{G}}_2\tilde{\mathbf{A}} \tag{36}$$

We see that matrix $\mathcal{L}$ plays a productive role in (9) and should not be dropped. While the resulting estimator $\hat{\mathbf{A}}$ will have a minimal variance in either case, with or without $\mathcal{L}$, the use of $\mathcal{L}$ helps reduce its bias. As a result, GRAF is more accurate than its reduced version (11); this was observed numerically in [4].

We should note that the bias of GRAF (36) is not always smaller than that of the reduced GRAF (35). Indeed, the vectors $\tilde{\mathbf{G}}_1\tilde{\mathbf{A}}$ and $\tilde{\mathbf{G}}_2\tilde{\mathbf{A}}$ depend on the true ellipse (specified by $\tilde{\mathbf{A}}$) and the location of the true points on it, and it may happen that the vector $\tilde{\mathbf{G}}_1\tilde{\mathbf{A}}$ partially cancels $\tilde{\mathbf{G}}_2\tilde{\mathbf{A}}$, so the sum $(\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2)\tilde{\mathbf{A}}$ is actually smaller than $\tilde{\mathbf{G}}_2\tilde{\mathbf{A}}$. However in general these two 6-dimensional vectors are loosely related to each other, so the cancelation is unlikely, i.e., in typical cases their sum is larger than each of them. We will adopt this principle throughout: every extra term (vector) in our formulas for the bias indicates that the resulting bias is larger.

3 Error analysis of generalized GRAF

Here we analyze the generalized GRAF (12). Applying matrix perturbation to $\mathcal{M}$, $\mathcal{N}$, $\mathbf{A}$, and λ in (12) gives

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

where terms of order σ^3 are omitted.

Setting $\sigma = 0$ gives $\tilde{\mathcal{M}}\tilde{\mathbf{A}} = \tilde{\lambda}\tilde{\mathcal{N}}\tilde{\mathbf{A}}$. We will assume that $\tilde{\mathcal{N}}\tilde{\mathbf{A}} \neq \mathbf{0}$, as otherwise the method is difficult to analyze. Since $\tilde{\mathcal{M}}\tilde{\mathbf{A}} = \mathbf{0}$, we get $\tilde{\lambda} = 0$. Now equating the terms of order σ gives

$$\tilde{\mathcal{M}}\Delta_1\mathbf{A} + \Delta_1\mathcal{M}\tilde{\mathbf{A}} = \Delta_1\lambda\tilde{\mathcal{N}}\tilde{\mathbf{A}} \tag{37}$$

Premultiplying (37) with $\tilde{\mathbf{A}}^T$ we get zero on the left hand side (note that $\tilde{\mathbf{A}}^T\Delta_1\mathcal{M}\tilde{\mathbf{A}} = 0$ because $\tilde{\mathbf{A}}^T\mathbf{Z}_i = 0$), hence $\Delta_1\lambda = 0$.

Therefore $\Delta_1\mathbf{A}$ is again given by (24)–(25), and its variance by (26). In particular, the generalized GRAF is statistically optimal (has a minimal variance) for any matrix $\mathcal{N}$.

Next we turn to its bias. The second order term $\Delta_2\mathbf{A}$ satisfies

$$\tilde{\mathcal{M}}\,\Delta_2\mathbf{A} = -(\Delta_2\mathcal{M}\,\tilde{\mathbf{A}} + \Delta_1\mathcal{M}\,\Delta_1\mathbf{A}) + \Delta_2\lambda\tilde{\mathcal{N}}\tilde{\mathbf{A}}. \tag{38}$$

We note a useful identity:

$$\begin{aligned}\mathbf{R}\tilde{\mathbf{A}} &= (\Delta_2^1\mathcal{M} + \Delta_2^2\mathcal{M})\tilde{\mathbf{A}} - (\Delta_1^1\mathcal{M} + \Delta_1^2\mathcal{M})\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\tilde{\mathbf{A}}\\ &= \Delta_2\mathcal{M}\tilde{\mathbf{A}} + \Delta_1\mathcal{M}\Delta_1\mathbf{A}.\end{aligned}$$

To find $\Delta_2\lambda$, we premultiply (38) with $\tilde{\mathbf{A}}^T$ and get $\tilde{\mathbf{A}}^T\tilde{\mathcal{M}} = \mathbf{0}$, hence

$$\Delta_2\lambda = \frac{\tilde{\mathbf{A}}^T\big(\Delta_2\mathcal{M}\tilde{\mathbf{A}} + \Delta_1\mathcal{M}\Delta_1\mathbf{A}\big)}{\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}}} = \frac{\tilde{\mathbf{A}}^T\mathbf{R}\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}}}, \tag{39}$$

thus $\Delta_2\mathbf{A}$ can be written as

$$\Delta_2\mathbf{A} = -\tilde{\mathcal{M}}^-\mathbf{R}\tilde{\mathbf{A}} + \tilde{\mathcal{M}}^-\frac{\tilde{\mathcal{N}}\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T}{\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}}}\mathbf{R}\tilde{\mathbf{A}}. \tag{40}$$

Recall that $-\tilde{\mathcal{M}}^-\mathbf{R}\tilde{\mathbf{A}} = \Delta_2\mathbf{A}_{\mathrm{Red}}$. Premultiplying (29) by $\tilde{\mathbf{A}}^T$ gives

$$\tilde{\mathbf{A}}^T\mathbb{E}(\mathbf{R})\tilde{\mathbf{A}} = \sigma^2\sum_{i=1}^{n}(\tilde{\gamma}_i^{-1} - \tilde{\gamma}_i^{-2}\tilde{\psi}_i)\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} = \sigma^2\sum_{i=1}^{n}(1 - \tilde{\gamma}_i^{-1}\tilde{\psi}_i)$$

Also note that

$$\sum_{i=1}^{n} \tilde{\gamma}_i^{-1}\tilde{\psi}_i = \sum_{i=1}^{n} \tilde{\gamma}_i^{-1}\tilde{\mathbf{Z}}_i^T\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_i = \sum_{i=1}^{n} \tilde{\gamma}_i^{-1}\mathrm{tr}(\tilde{\mathcal{M}}^-\tilde{\mathbf{M}}_i) = \mathrm{tr}(\tilde{\mathcal{M}}^-\tilde{\mathcal{M}}) = 5,$$

because $\mathrm{kernel}(\tilde{\mathcal{M}}) = \mathrm{span}(\tilde{\mathbf{A}})$ is one-dimensional. Thus

$$\mathbb{E}(\Delta_2\mathbf{A}) = \mathbb{E}(\Delta_2\mathbf{A}_{\mathrm{Red}}) + \sigma^2\tilde{\mathcal{M}}^-\frac{n-5}{\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}}}\tilde{\mathcal{N}}\tilde{\mathbf{A}}.$$

Thus the bias of the generalized GRAF may be even larger than that of the reduced GRAF. However, by a clever choice of the matrix $\mathcal{N}$ one can actually suppress it. In the next section we will construct $\mathcal{N}$ for which $\mathbb{E}(\Delta_2\mathbf{A}) = \mathbf{0}$.

We conclude this section by reviewing a particular version of the generalized GRAF proposed by Kanatani [11, 12] known as *renormalization scheme*; its constraint matrix is $\mathcal{N} = \sum_{i=1}^{n} \gamma_i^{-1}\mathbf{V}_i$. In this case $\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}} = n$, hence

$$\mathbb{E}(\Delta_2\mathbf{A}_{\mathrm{Ren}}) = -\sigma^2\tilde{\mathcal{M}}^-(\tilde{\mathbf{G}}_1^* + \tilde{\mathbf{G}}_2)\tilde{\mathbf{A}} \tag{41}$$

where

$$\tilde{\mathbf{G}}_1^* = \sum_{i=1}^{n}(5n^{-1}\tilde{\gamma}_i^{-1} - \tilde{\gamma}_i^{-2}\tilde{\psi}_i)\tilde{\mathbf{V}}_i, \tag{42}$$

which differs from $\tilde{\mathbf{G}}_1$ by an extra factor of $5n^{-1}$ in front of $\tilde{\gamma}_i^{-1}$. Thus for $n = 5$ the biases of the reduced and renormalization schemes coincide, but for larger n's the latter is smaller. At the same time (41) is larger than the bias of GRAF (36), because of $\tilde{\mathbf{G}}_1^*$.

In fact, for $n = 5$ all our fits return the same ellipse (the one interpolating the data points), so their biases are the same. Indeed, by simple algebra one can see that $n = 5$ implies $\tilde{\gamma}_i = \tilde{\psi}_i$, hence $\tilde{\mathbf{G}}_1 = \tilde{\mathbf{G}}_1^* = \mathbf{0}$.

On the other hand, for large n, we have $\tilde{\mathcal{M}} = \mathcal{O}(n)$, hence $\tilde{\mathcal{M}}^- = \mathcal{O}(n^{-1})$ and $\psi_i = \mathcal{O}(n^{-1})$. This implies $\tilde{\mathbf{G}}_1^* = \mathcal{O}(1)$, while $\tilde{\mathbf{G}}_1 = \mathcal{O}(n)$. Thus the renormalization scheme suppresses $\tilde{\mathbf{G}}_1$ to a matrix of a smaller order, while GRAF eliminates it completely.

4 Doubly optimal generalized GRAF

We now turn to our main goal – choosing a constraint matrix $\mathcal{N}$ for the generalized GRAF (12) so that $\mathbb{E}(\Delta_2\mathbf{A}) = \mathbf{0}$.

Recall that $\Delta_2\mathbf{A}$ is given by (40), in which only the matrix $\mathbf{R}$ is random. Thus we need to find $\mathcal{N}$ so that

$$\frac{\tilde{\mathbf{A}}^T\mathbb{E}(\mathbf{R})\tilde{\mathbf{A}}}{\tilde{\mathbf{A}}^T\tilde{\mathcal{N}}\tilde{\mathbf{A}}}\tilde{\mathcal{N}}\tilde{\mathbf{A}} = \mathbb{E}(\mathbf{R})\tilde{\mathbf{A}}. \tag{43}$$

This is true if and only if the vector $\tilde{\mathcal{N}}\tilde{\mathbf{A}}$ is parallel to $\mathbb{E}(\mathbf{R})\tilde{\mathbf{A}}$. We recall that $\mathbb{E}(\mathbf{R})\tilde{\mathbf{A}} = \sigma^2(\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2)\tilde{\mathbf{A}}$ and choose $\mathcal{N}$ so that $\tilde{\mathcal{N}} = \tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2$. The matrix $\mathcal{N}$ can be simply constructed by replacing the true values $(\tilde{x}_i, \tilde{y}_i)$'s with the observations (x_i, y_i)'s and $\tilde{\mathbf{A}}$ with $\mathbf{A}$ in $\tilde{\mathbf{G}}_1 + \tilde{\mathbf{G}}_2$, which is given by (29)–(31). Thus we set $\mathcal{N} := \mathbf{G}_1 + \mathbf{G}_2$, where

$$\begin{aligned}
\mathbf{G}_1 &= \sum_{i=1}^{n}(\gamma_i^{-1} - \gamma_i^{-2}\psi_i)\mathbf{V}_i \\
\mathbf{G}_2 &= \sum_{i=1}^{n}\big[\gamma_i^{-1}\mathbf{Z}_i\mathbf{e}_{13}^T - \gamma_i^{-2}\mathbf{M}_i\mathcal{M}^-\mathbf{V}_i + \gamma_i^{-3}\mathbf{M}_i\mathcal{M}^-\mathbf{\Gamma}_i - \gamma_i^{-2}\mathbf{\Gamma}_i\big].
\end{aligned} \tag{44}$$

One can think of $\mathbf{G}_k$, $k = 1, 2$, as a "random realization" of $\tilde{\mathbf{G}}_k$.

Note that $\mathbf{G}_1$ is symmetric but $\mathbf{G}_2$ is not. For solving the generalized eigenvalue problem (12) it is desirable to have a symmetric constraint matrix $\mathcal{N}$, and one can be defined by $\mathcal{N} := \mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_2^T$. This also implies (43) due to (29).

Next we need to choose a particular eigenvector $\mathbf{A}$ solving (12). It is a common practice to take the one corresponding to the smallest (closest to zero) eigenvalue λ; see [2, 18, 19, 20]. The rational here is that the true value of λ equals zero; in fact we proved that $\lambda = \mathcal{O}(\sigma^2)$.

Lastly we summarize our algorithm:

1. Set $k = 0$ and choose an initial guess $\mathbf{A}$ (see below).
2. Compute matrices $\mathcal{M}$ and $\mathcal{N}$ given by equations (9) and (44).
3. Solve the generalized eigenvalue problem $\mathcal{M}\mathbf{A}' = \lambda\mathcal{N}\mathbf{A}'$ and take a unit vector $\mathbf{A}'$ corresponding to the smallest (closest to zero) λ.
4. if $\|\mathbf{A} + \mathbf{A}'\|_2 < \|\mathbf{A} - \mathbf{A}'\|_2$, reset $\mathbf{A}' = -\mathbf{A}'$;
5. If $\|\mathbf{A} - \mathbf{A}'\|_2$ is small enough, terminate the procedure, return $\mathbf{A}$.
6. Update $\mathbf{A} = \mathbf{A}'$ and go back to Step 2.

For the initial guess, we use the best algebraic fit (called HyperLS) developed recently by Kanatani and Rangarajan [18].

5 Numerical test

We have tested our algorithm against three other schemes described before: GRAF, the renormalization method, and the reduced scheme.

We used ellipse $x^2/a^2 + y^2/b^2 = 1$ with semiaxes $a = 10$ and $b = 2$, placed $n = 30$ true points according to

$$\tilde{x}_i = a\cos\varphi_i, \qquad \tilde{y}_i = b\sin\varphi_i, \qquad \varphi_i = \pi(i - 0.5)/n$$

for $i = 1, \ldots, n$, so that all the points are on the upper half of the ellipse, and added Gaussian noise at level σ, which ranged from 0.0005 to 0.045 (when $\sigma > 0.045$, some fits start returning hyperbolas or diverging). For each σ, we generated $N = 10^7$ random samples and fit ellipses by each of the tested scheme. All our schemes are iterative and we used Taubin's algebraic ellipse fit [24] to initialize them.

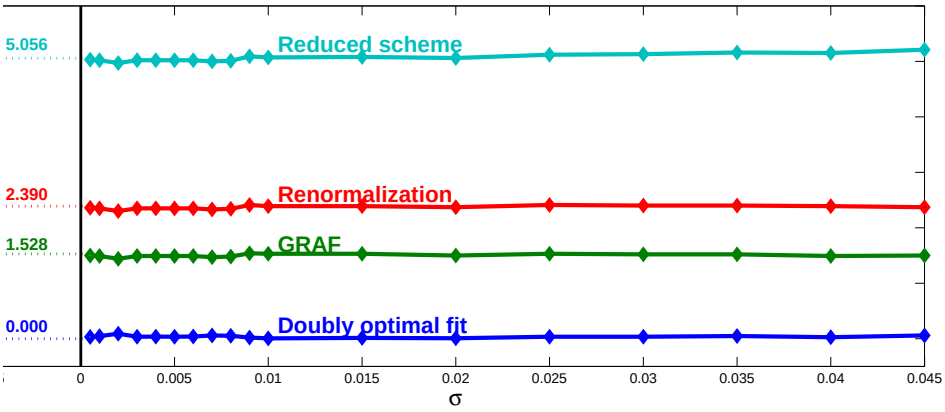

Figure 1: The normalized bias, $\|\mathcal{B}\|/\sigma^2$, plotted versus σ for the reduced scheme, the renormalization method, the GRAF, and our doubly optimal fit. As $\sigma \to 0$, each normalized bias converges to a limit value printed on the left and marked by a horizontal dotted line (for each fit separately).

For each estimator $\hat{\mathbf{A}}$ we computed its bias by $\mathcal{B} = \frac{1}{N}\sum \hat{\mathbf{A}} - \tilde{\mathbf{A}}$ and then plotted the "normalized" bias $\|\mathcal{B}\|/\sigma^2$ versus σ. Recall that $\mathcal{B} = \mathcal{O}(\sigma^2)$, so the ratio $\|\mathcal{B}\|/\sigma^2$ converges to a constant as $\sigma \to 0$, i.e.,

$$\lim_{\sigma\to 0} \frac{\|\mathcal{B}\|}{\sigma^2} = b,$$

where $b \geq 0$ is different for each method. The theoretically computed values of b are $b = 5.056$ for the reduced scheme, $b = 2.390$ for the renormalization method, $b = 1.528$ for GRAF, and $b = 0$ for our doubly optimal fit. These values are also marked on our plot.

The plot shows a remarkable stability of the bias for each method, it hardly changes over the range $0 < \sigma \leq 0.045$. Note that the bias of the reduced scheme slightly increases as σ approaches 0.045, but other biases remain virtually constant. This indicates that the higher order corrections (from the terms of order σ^3) are insignificant.

Our experiment clearly demonstrates the superiority of the doubly optimal fit over others in this group.

We note that some other ways for reducing bias were proposed in the literature. Certain implementations of the GRAF [15, 22] include bias-reduction options, but our paper provides a first full proof of the elimination of the bias to the leading order.

We emphasize that our analysis is based on the assumption that n is fixed and $\sigma \to 0$, which is common in computer vision applications (see [13] or [5, Sect. 2.4]). One may want to adopt a more traditional statistical approach, i.e., fix $\sigma > 0$ and let $n \to \infty$. It is known, however, that none of the standard fitting algorithms (not even the geometric fit (2)) would be statistically consistent. More precisely, the estimate $\hat{\mathbf{A}}$ would converge (in probability and almost surely), as $n \to \infty$, to a limit $\mathbf{A}^*$ different from the true value $\tilde{\mathbf{A}}$. This fact is proved under general conditions in [9], see also [5, Sect. 6.10]. When the sample size n is large and σ is not small, one should use special *consistent* ellipse fits; see [21].

Acknowledgement. We are grateful to K. Kanatani and to anonymous referees for many useful remarks. N. C. is partially supported by NSF grant DMS-0969187.

Appendix

Here we prove (29). Let $\mathbf{\Lambda}_i = \big(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\tilde{\mathbf{A}}\big)\Delta_1^1\mathcal{M}$. Then

$$\begin{aligned}\mathbb{E}\big(\mathbf{\Lambda}_i\big) &= \sum_{j,k=1}^n \tilde{\gamma}_j^{-1}\tilde{\gamma}_k^{-1}\mathbb{E}\Big[\big(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-(\Delta_1\mathbf{Z}_j^T\tilde{\mathbf{A}})\tilde{\mathbf{Z}}_j\big)\big(\Delta_1\,\mathbf{Z}_k\tilde{\mathbf{Z}}_k^T + \tilde{\mathbf{Z}}_k\Delta_1\,\mathbf{Z}_k^T\big)\Big],\\ &= \sum_{j,k=1}^n \tilde{\gamma}_j^{-1}\tilde{\gamma}_k^{-1}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j)\Big[\mathbf{E}_{kj}\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_k^T + \tilde{\mathbf{Z}}_k\tilde{\mathbf{A}}^T\mathbf{E}_{jk}\Big].\end{aligned}$$

where $\mathbf{E}_{kj}\colon = \mathbb{E}(\Delta_1\,\mathbf{Z}_k\Delta_1\,\mathbf{Z}_j^T) = \sigma^2\tilde{\mathbf{V}}_k$ if $k = j$ and $\mathbf{0}$ otherwise, so

$$\mathbb{E}\big(\mathbf{\Lambda}_i\big) = 2\sigma^2\sum_{j=1}^n \tilde{\gamma}_j^{-2}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j)\mathcal{S}\big[\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_j^T\big]. \tag{45}$$

where $\mathcal{S}(\cdot)$ denotes the symmetrization of a matrix, i.e., $\mathcal{S}(\mathbf{B}) = (\mathbf{B}+\mathbf{B}^T)/2$. Next, since $\gamma_i = \mathbf{A}^T\mathbf{V}_i\mathbf{A}$, we have

$$\Delta_1\gamma_i = 2(\Delta_1\mathbf{A}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}) + \tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}}, \tag{46}$$

where $\Delta_1\mathbf{V}_i = \tilde{\mathbf{T}}_i\delta_i + \tilde{\mathbf{S}}_i\varepsilon_i$, with $\tilde{\mathbf{T}}_i$ and $\tilde{\mathbf{S}}_i$ being defined in (34).

Now we integrate (28) term by term. First we compute

$$\mathbf{I}_1\colon = \mathbb{E}(\Delta_1^1\mathcal{M}\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}) = \sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\big(\Delta_1\mathbf{M}_i\tilde{\mathcal{M}}^-\Delta_1\mathbf{M}_i\big),$$

which expands to

$$\sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\Big(\big(\Delta_1\,\mathbf{Z}_i\tilde{\mathbf{Z}}_i^T + \tilde{\mathbf{Z}}_i\Delta_1\,\mathbf{Z}_i^T\big)\tilde{\mathcal{M}}^-\big(\Delta_1\,\mathbf{Z}_i\tilde{\mathbf{Z}}_i^T + \tilde{\mathbf{Z}}_i\Delta_1\,\mathbf{Z}_i^T\big)\Big).$$

By direct inspection, $\mathbb{E}\big(\Delta_1\,\mathbf{Z}_i^T\tilde{\mathcal{M}}^-\Delta_1\,\mathbf{Z}_i\big) = \sigma^2\mathrm{tr}(\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i)$, hence

$$\mathbf{I}_1 = \sigma^2\sum_{i=1}^n \tilde{\gamma}_i^{-2}\big[2\mathcal{S}[\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{M}}_i] + \mathrm{tr}\,(\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i)\tilde{\mathbf{M}}_i + \tilde{\psi}_i\tilde{\mathbf{V}}_i\big]. \tag{47}$$

We will also need an auxiliary formula below that follows from (25):

$$\mathbb{E}(\Delta_1\mathbf{A}\Delta_1\mathbf{Z}_i^T) = -\sigma^2\tilde{\gamma}_i^{-1}\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_i\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i \tag{48}$$

and hence $\mathbb{E}(\Delta_1\mathbf{Z}_i\Delta_1\mathbf{A}^T) = -\sigma^2\tilde{\gamma}_i^{-1}\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_i^T\tilde{\mathcal{M}}^-$. Next we compute

$$\mathbf{I}_2 \colon = \mathbb{E}(\Delta_1^2\mathcal{M}\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}) = -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\big(\Delta_1\gamma_i\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\big). \tag{49}$$

Using $\Delta_1^1\mathcal{M} = \sum\tilde{\gamma}_i^{-1}\Delta_1\mathbf{M}_i$, then (46), (45), and (48) gives

$$\begin{aligned}
\mathbf{I}_2 &= -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\Big[\big(2(\Delta_1\mathbf{A}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}) + \tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}}\big)\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\Big] \\
&= -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\big(2(\Delta_1\mathbf{A}^T\tilde{\mathbf{V}}_i\tilde{\mathbf{A}})\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\big) \\
&\quad -\sum_{i=1}^n \tilde{\gamma}_i^{-2}\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}})\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\Delta_1^1\mathcal{M}\big) \\
&= 4\sigma^2\sum_{i,j=1}^n \tilde{\gamma}_i^{-2}\tilde{\gamma}_j^{-2}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j)\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\mathcal{S}[\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_j^T] \\
&\quad -\sum_{i=1}^n \tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}})\Delta_1\mathbf{M}_i\big).
\end{aligned}$$

Now recall that $\tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}} = (\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}})\delta_i + (\tilde{\mathbf{A}}^T\tilde{\mathbf{S}}_i\tilde{\mathbf{A}})\varepsilon_i$. Also note that $\mathbb{E}(\delta_i\Delta_1\mathbf{M}_i) = 2\sigma^2\mathcal{S}[\tilde{\mathbf{a}}_i\tilde{\mathbf{Z}}_i^T]$ and $\mathbb{E}(\varepsilon_i\Delta_1\mathbf{M}_i) = 2\sigma^2\mathcal{S}[\tilde{\mathbf{b}}_i\tilde{\mathbf{Z}}_i^T]$, where $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ denote the first and second columns of $\nabla\mathbf{Z}_i$. Thus

$$\begin{aligned}
\mathbb{E}\big((\tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}})\Delta_1\mathbf{M}_i\big) &= 2\sigma^2(\tilde{\mathbf{A}}^T\tilde{\mathbf{T}}_i\tilde{\mathbf{A}})\mathcal{S}[\tilde{\mathbf{a}}_i\tilde{\mathbf{Z}}_i^T] + 2\sigma^2(\tilde{\mathbf{A}}^T\tilde{\mathbf{S}}_i\tilde{\mathbf{A}})\mathcal{S}[\tilde{\mathbf{b}}_i\tilde{\mathbf{Z}}_i^T] \\
&= 2\sigma^2\mathcal{S}[\tilde{\mathbf{\Gamma}}_i],
\end{aligned} \tag{50}$$

where $\tilde{\mathbf{\Gamma}}_i$ is defined in (33). These facts imply

$$\begin{aligned}
\mathbf{I}_2 &= 4\sigma^2\sum_{i,j=1}^n \tilde{\gamma}_i^{-2}\tilde{\gamma}_j^{-2}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j)\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\mathcal{S}[\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_j^T] \\
&\quad -2\sigma^2\sum_{i=1}^n \tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\mathcal{S}[\tilde{\mathbf{\Gamma}}_i].
\end{aligned} \tag{51}$$

Next we use (20) to compute

$$\mathbb{E}(\Delta_2^1\mathcal{M}) = \sum_{i=1}^n \tilde{\gamma}_i^{-1}\mathbb{E}(\Delta_2\mathbf{M}_i) = \sigma^2\sum_{i=1}^n \tilde{\gamma}_i^{-1}\Big(\tilde{\mathbf{V}}_i + 2\mathcal{S}\big[\mathbf{e}_{13}\tilde{\mathbf{Z}}_i^T\big]\Big). \tag{52}$$

Lastly, we compute

$$\begin{aligned}\mathbb{E}(\Delta_2^2\mathcal{M}) &= -\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\mathbb{E}(\Delta_1\gamma_i\Delta_1\mathbf{M}_i)\\ &= -\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\mathbb{E}\Big(\big(2(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\Delta_1\mathbf{A}) + \tilde{\mathbf{A}}^T\Delta_1\mathbf{V}_i\tilde{\mathbf{A}}\big)\Delta_1\mathbf{M}_i\Big).\end{aligned}$$

We handle the first term by (48) and the second by (50), thus

$$\mathbb{E}(\Delta_2^2\mathcal{M}) = 2\sigma^2\sum_{i=1}^{n}\Big(2\tilde{\gamma}_i^{-3}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_i)\mathcal{S}[\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_i^T] - \tilde{\gamma}_i^{-2}\mathcal{S}[\tilde{\mathbf{\Gamma}}_i]\Big) \tag{53}$$

This completes the integration of (28).

To prove (29) we multiply (47), (51), (52), and (53) by $\tilde{\mathbf{A}}$. First,

$$\mathbf{I}_1\tilde{\mathbf{A}} = \sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\big[\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} + \tilde{\psi}_i\tilde{\mathbf{V}}_i\tilde{\mathbf{A}}\big], \tag{54}$$

Also note a useful identity $\mathcal{S}[\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}\tilde{\mathbf{Z}}_j^T]\tilde{\mathbf{A}} = \tilde{\mathbf{Z}}_j(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_j\tilde{\mathbf{A}}) = \tilde{\gamma}_j\tilde{\mathbf{Z}}_j$, as well as $\tilde{\mathbf{\Gamma}}_i^T\tilde{\mathbf{A}} = \mathbf{0}$. Now we have

$$\begin{aligned}\mathbf{I}_2\tilde{\mathbf{A}} &= 2\sigma^2\sum_{i,j=1}^{n}\tilde{\gamma}_i^{-2}\tilde{\gamma}_j^{-1}(\tilde{\mathbf{A}}^T\tilde{\mathbf{V}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j)\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{Z}}_j - \sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{\Gamma}}_i\tilde{\mathbf{A}}\\ &= 2\sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\Big[\sum_{j=1}^{n}\tilde{\gamma}_j^{-1}\tilde{\mathbf{M}}_j\Big]\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} - \sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{\Gamma}}_i\tilde{\mathbf{A}}\\ &= 2\sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-2}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} - \sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-3}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{\Gamma}}_i\tilde{\mathbf{A}},\end{aligned} \tag{55}$$

where we used the identity $\tilde{\mathcal{M}}^-\tilde{\mathcal{M}}\tilde{\mathcal{M}}^- = \tilde{\mathcal{M}}^-$. Next

$$\mathbb{E}(\Delta_2^1\mathcal{M})\tilde{\mathbf{A}} = \sigma^2\sum_{i=1}^{n}\tilde{\gamma}_i^{-1}(\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} + \tilde{\mathbf{Z}}_i\mathbf{e}_{13}^T\tilde{\mathbf{A}}). \tag{56}$$

Lastly, we apply the above useful identities again and compute

$$\mathbb{E}(\Delta_2^2\mathcal{M})\tilde{\mathbf{A}} = \sigma^2\sum_{i=1}^{n}\Big(2\tilde{\gamma}_i^{-2}\tilde{\mathbf{M}}_i\tilde{\mathcal{M}}^-\tilde{\mathbf{V}}_i\tilde{\mathbf{A}} - \tilde{\gamma}_i^{-2}\tilde{\mathbf{\Gamma}}_i\tilde{\mathbf{A}}\Big) \tag{57}$$

Combining (54), (55), (56), and (57) completes the proof of (29) (note that $\tilde{\mathbf{G}}_2^T\tilde{\mathbf{A}} = \mathbf{0}$). Note that some terms in (55) and (57) cancel each other.

References

[1] S. J. Ahn. *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space*. LNCS, 3151, Springer, Berlin, 2004.

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

[3] F. L. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, 9:56–71, 1979.

[4] N. Chernov. On the convergence of fitting algorithms in computer vision. *J. Math. Imag. Vision*, 27:231–239, 2007.

[5] N. Chernov. *Circular and linear regression: Fitting circles and lines by least squares*, volume 117. Chapman & Hall/CRC, Boca Raton, 2010.

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

[7] W. Chojnacki, M. J. Brooks, and A. van den Hengel. Rationalising the renormalisation method of Kanatani. *J. Math. Imaging & Vision*, 14:21–38, 2001.

[8] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *IEEE Trans. Pattern Analysis Machine Intelligence*, 22:1294–1303, 2000.

[9] I. Fazekas, A. Kukush, and S. Zwanzig. Correction of nonlinear orthogonal regression estimator. *Ukr. Math. J.*, 56:1101–1118, 2004.

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

[11] K. Kanatani. Statistical bias of conic fitting and renormalization. *IEEE Trans. Pattern Analysis Machine Intelligence*, 16:320–326, 1994.

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

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

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

[15] K. Kanatani. Ellipse fitting with hyperaccuracy. In *Proc. 9th European Conf. Computer Vision (ECCV 2006)*, volume 1, pages 484–495, 2006.

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

[17] K. Kanatani. Optimal estimation. In K. Ikeuchi, editor, *Encyclopedia of Computer Vision*, Berlin, 2011. Springer.

[18] K. Kanatani and P. Rangarajan. Improved algebraic methods for circle fitting. *Electr. J. Statist.*, 3:1075–1082, 2009.

[19] K. Kanatani and P. Rangarajan. Hyper least squares fitting of circles and ellipses. *Comput. Stat. Data Anal.*, 55:2197–2208, 2011.

[20] K. Kanatani, P. Rangarajan, Y. Sugaya, and H. Niitsuma. HyperLS and its applications. *IPSJ Trans. Computer Vision Appl.*, 3, 2011.

[21] A. Kukush, I. Markovsky, and S. Van Huffel. Consistent estimation in an implicit quadratic measurement error model. *Comput. Statist. Data Anal.*, 47:123–147, 2004.

[22] Y. Leedan and P. Meer. Heteroscedastic regression in computer vision: Problems with bilinear constraint. *Intern. J. Comp. Vision*, 37:127–150, 2000.

[23] P. D. Sampson. Fitting conic sections to very scattered data: an iterative refinement of the Bookstein algorithm. *Comp. Graphics Image Proc.*, 18:97–108, 1982.

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