Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computational Statistics & Data Analysis 47 (2004) 123–147
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Consistent estimation in an implicit quadratic measurement error model

Alexander Kukush[1], Ivan Markovsky*, Sabine Van Huffel
K.U. Leuven, ESAT-SISTA, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

Received 1 November 2002; received in revised form 30 October 2003; accepted 30 October 2003

Abstract

An adjusted least squares estimator is derived that yields a consistent estimate of the parameters of an implicit quadratic measurement error model. In addition, a consistent estimator for the measurement error noise variance is proposed. Important assumptions are: (1) all errors are uncorrelated identically distributed and (2) the error distribution is normal. The estimators for the quadratic measurement error model are used to estimate consistently conic sections and ellipsoids. Simulation examples, comparing the adjusted least squares estimator with the ordinary least squares method and the orthogonal regression method, are shown for the ellipsoid fitting problem.
© 2003 Elsevier B.V. All rights reserved.

Keywords: Adjusted least squares; Conic fitting; Consistent estimator; Ellipsoid fitting; Quadratic measurement error model

1. Introduction

A parameter estimation problem occurs when the relation among some observed variables $x_1,\ldots,x_n$ is described by a parameterized model. The parameters identify a unique model in a given *model class*, and the problem is to choose a model from the model class, given a set of observations $\{x^{(l)}\}_{l=1}^{m}$, where $x^{(l)} := [x_1^{(l)} \cdots x_n^{(l)}]^\top$ is the lth observed vector of variables. The model is selected according to certain performance criteria, specified later.

* Corresponding author. Tel.: +32-16-32-17-10; fax: +32-16-32-19-70.
E-mail address: ivan.markovsky@esat.kuleuven.ac.be (I. Markovsky).
[1] On leave from National Taras Shevchenko University, Vladimirskaya st. 64, 01601, Kiev, Ukraine.

0167-9473/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2003.10.022

We consider an *implicit quadratic model*, $x^\top Ax + b^\top x + d = 0$, A symmetric, relating the variables $x := [x_1 \cdots x_n]^\top$. It describes a *second order surface*

$$S(A,b,d) := \{x \in \mathbb{R}^n \colon x^\top Ax + b^\top x + d = 0\} \tag{1}$$

in $\mathbb{R}^n$. The model is called implicit because there is no difference between dependent and independent variables. The *parameters* are the symmetric matrix A, the vector b, and the scalar d and the model class is the set of all quadratic equations with an n-dimensional variable.

If $A = 0$ and $b \neq 0$, then surface (1) is a *hyperplane*, and if A is positive definite and $4d < b^\top A^{-1} b$, then (1) is an *elliptic surface*. The set $S(A,b,d)$ might be disconnected. Initially we will not make assumptions on the surface under estimation apart from the requirement of being a non-empty set. Later on we will specialize the results for the cases of conic section and ellipsoid estimation.

Without additional constraints imposed on the parameters, given a model in the model class, the model parameters are not unique: any multiple of a set of parameters defines the same model. This makes the quadratic model, parameterized by A, b, and d non-identifiable. To resolve the problem, we impose a *normalizing condition*, e.g., the parameters are assumed to satisfy the constraint

$$\|A\|_F^2 + \|b\|^2 + d^2 = 1. \tag{2}$$

With this normalizing condition, the parameters are unique up to a sign.

The vector of variables x is observed with additive error $e = [e_1 \cdots e_n]^\top$ and the error is described stochastically. The true value $\bar{x} = [\bar{x}_1 \cdots \bar{x}_n]^\top$ of the measured variables is assumed to satisfy the model for some unknown true values $\bar{A}$, $\bar{b}$, and $\bar{d}$ of the parameters. This assumption defines a *true model* in the model class. Models in which the variables are measured with additive noise $x = \bar{x} + e$ are called *measurement error models*. Thus the model considered in the paper is an implicit quadratic measurement error model.

The quadratic model is linear in the parameters, so that the linear least squares technique can be applied. This corresponds to estimation criterion:

$$\min_{A,b,d} \sum_{l=1}^{m} (x^{(l)\top} Ax^{(l)} + b^\top x^{(l)} + d)^2.$$

We will call the resulting estimator the *ordinary least squares* (OLS) estimator, in order to distinguish it from the adjusted least squares estimator, introduced later.

Due to the normalizing condition imposed on the parameters, the OLS problem is a quadratically constrained least squares problem and the necessary computation is to find the smallest eigenvalue/eigenvector of a self-adjoint and positive definite linear operator. The presence of measurement errors in all the covariates, however, makes the OLS estimator biased, see, e.g., Carroll et al. (1995).

Another approach is the *orthogonal regression* estimation. Let $\operatorname{dist}(x,S)$ be the distance from the point x to the set S. The orthogonal regression estimator is defined as a global solution of the following optimization problem:

$$\min_{A,b,d} \sum_{l=1}^{m} \operatorname{dist}(x^{(l)}, S(A,b,d))^2.$$

The non-linearity of the model with respect to the measurements, implies the inconsistency of this estimator as well, see the classical paper of Neyman and Scott (1948) and the discussion in (Fuller, 1987, p.250).

We assume that the measurement errors $e^{(1)},\dots,e^{(m)}$ are centered, uncorrelated among the measurement, and normally distributed, $e^{(l)} \sim \mathrm{N}(0,\bar{\sigma}^2 I)$ for all l, with noise variance $\bar{\sigma}^2$. We consider both cases, when $\bar{\sigma}^2$ is given, and when $\bar{\sigma}^2$ is unknown. The stochastic description of the measurement errors can be viewed as a model with parameter σ^2. Then the noise variance σ^2 is a nuisance parameter of the model.

Using the noise model assumptions, we apply an adjustment procedure, see Naidu (1990), that takes into account the quadratic structure of the model and corrects the OLS estimate appropriately. The resulting estimator, called an *adjusted least squares* (ALS) estimator, is consistent. Similar approach for consistent estimation is used in a bilinear model, see Kukush et al. (2003).

A nice feature of the ALS estimator is that its computation also requires the smallest eigenvector of a self-adjoint linear operator. This operator is obtained from the self-adjoint and positive definite operator used in the computation of the OLS estimator by applying the correction. If the measurement error variance $\bar{\sigma}^2$ is a priori known, we give the correction operator in terms of $\bar{\sigma}^2$. If however, $\bar{\sigma}^2$ is unknown, then it has to be estimated together with the model parameters. We propose a consistent procedure to estimate the unknown measurement error variance.

We use the ALS estimator, derived for the quadratic model, to solve the *conic fitting* and the *ellipsoid fitting* problems. In the ellipsoid fitting case, we obtain consistent estimators for the parameters A_e and c of the ellipsoid described by the quadratic model $(x-c)^\top A_e(x-c)=1$, with A_e positive definite.

We point out several papers in which the ellipsoid fitting problem is considered. Gander et al. (1994) consider algebraic and geometric fitting methods for circles and ellipses and note the inadequacy of the algebraic fit on some specific examples. Later on, the given examples are used as benchmarks for the algebraic fitting methods. Ellipsoid specific, as opposed to the more general conic fitting method is first proposed in Fitzgibbon et al. (1999). The method incorporates the ellipticity constraint into the normalizing condition and thus gives better results when an elliptic fit is desired. In Nievergelt (2001), a new algebraic fitting method is proposed that does not have as singularity the special case of a hyperplane fitting; if the best fitting manifold is affine the method coincides with the total least squares method. Geometric methods, minimizing the sum of absolute values of orthogonal deviations, are discussed in Nyquist (1988).

A statistical point of view on the ellipsoid fitting problem is taken in Kanatani (1994), Cabrera and Meer (1996), and Zhang (1997). Kanatani proposed an unbiased estimation method, called a renormalization procedure. He uses an adjustment similar to the one in the present paper but his approach of estimating the unknown noise variance is different. Moreover, the noise variance estimate proposed in Kanatani (1994) is still inconsistent; the bias is removed up to the first order approximation. We note, however, that in the context of the quadratic measurement error model, the notion of bias is inappropriate, see the discussion in Section 5.

Standard notation used in the paper is: $\mathbb{R}$ for the set of real numbers, $\mathbf{E}e$ for the expectation of the random variable e, $O_p(1)$ for a sequence of stochastically bounded random variables, $N(0, V)$ for the zero mean normal distribution in Euclidean space with variance–covariance matrix V, $\lambda_{\min}(\Psi)$ ($\lambda_{\max}(\Psi)$) for minimum (maximum) eigenvalue of the self-adjoint linear operator Ψ, $\|A\|_F$ for the Frobenius norm of the matrix A, and $\operatorname{dist}(x, y)$ is defined as $\|x - y\|$, where the norm is understood from the context. Throughout the paper $\mathbb{S}$ denotes the space of the $n \times n$ symmetric matrices. Specific notation is introduced in the text.

Section 2 defines the quadratic measurement error model. Sections 3 and 4 present, respectively, the OLS and the ALS estimators. In Section 5, we state the consistency of the ALS estimator with known noise variance, and in Section 6, we consider the noise variance estimation problem. The proofs of all results in Sections 5 and 6 are included in the Appendix. Sections 7 and 8 consider two special cases of the quadratic model estimation problem: conic section and ellipsoid fitting. Section 9 shows simulation examples for the ellipsoid fitting problem. Conclusions are given in Section 10.

2. Quadratic measurement error model

Let $\bar{A} \in \mathbb{S}$, $\bar{b} \in \mathbb{R}^n$, and $\bar{d} \in \mathbb{R}$ be such that the set $S(\bar{A}, \bar{b}, \bar{d})$, defined in (1), is non-empty and let the points $\bar{x}^{(1)}, \ldots, \bar{x}^{(m)}$, lie on the surface $S(\bar{A}, \bar{b}, \bar{d})$, i.e.,

$$\bar{x}^{(l)\top} \bar{A} \bar{x}^{(l)} + \bar{b}^\top \bar{x}^{(l)} + \bar{d} = 0, \quad \text{for } l = 1, \ldots, m. \tag{3}$$

The points $x^{(1)}, \ldots, x^{(m)}$, are *measurements* of the points $\bar{x}^{(1)}, \ldots, \bar{x}^{(m)}$, respectively, i.e.,

$$x^{(l)} = \bar{x}^{(l)} + e^{(l)}, \quad \text{for } l = 1, \ldots, m, \tag{4}$$

where $e^{(1)}, \ldots, e^{(m)}$ are the corresponding *measurement errors*. We make the following assumptions:

(1) the measurement errors $e^{(1)}, \ldots, e^{(m)}$ form a sequence of independent identically distributed random vectors, and
(2) the distribution of $e^{(l)}$, for all $l = 1, \ldots, m$, is normal $N(0, \bar{\sigma}^2 I_n)$.

Here $\bar{\sigma}^2 > 0$ is the *noise variance* and I_n is the $n \times n$ identity matrix.

The matrix $\bar{A} \in \mathbb{S}$ is the *true value* of the parameter A, while $\bar{b} \in \mathbb{R}^n$, and $\bar{d} \in \mathbb{R}^1$ are the true values of the parameters b and d, respectively. We assume that the true values of the parameters satisfy the normalizing condition (2).

3. Ordinary least squares estimator

The *elementary OLS cost function* is

$$q_{\text{ols}}(A, b, d; x) := (x^\top A x + b^\top x + d)^2 \quad \text{for all } A \in \mathbb{S},\ b \in \mathbb{R}^n,\ d \in \mathbb{R}.$$

It measures the discrepancy of a single measurement point x from the surface $S(A,b,d)$. The *OLS cost function*/ is the sum of the elementary cost function for all data points,

$$Q_{\mathrm{ols}}(A,b,d) := \sum_{l=1}^{m} q_{\mathrm{ols}}(A,b,d,x^{(l)}), \quad \text{for all } A \in \mathbb{S},\ b \in \mathbb{R}^n,\ d \in \mathbb{R}. \tag{5}$$

The *OLS estimator* $\hat{A}_{\mathrm{ols}}$, $\hat{b}_{\mathrm{ols}}$, $\hat{d}_{\mathrm{ols}}$ is defined as the global minimum point of (5), subject to the normalizing constraint (2).

We consider the parameter triple

$$\beta := (A,b,d) \in \mathbb{V}$$

as a vector in the Hilbert space $\mathbb{V} := \mathbb{S} \times \mathbb{R}^n \times \mathbb{R}$ with inner product

$$\langle (A_1,b_1,d_1),(A_2,b_2,d_2)\rangle := \operatorname{trace}(A_1^{\top}A_2) + b_1^{\top}b_2 + d_1d_2,$$
$$\text{for all } (A_1,b_1,d_1) \in \mathbb{V},\ (A_2,b_2,d_2) \in \mathbb{V}.$$

With this notation, the optimization problem, we want to solve, is

$$\min_{\beta} Q_{\mathrm{ols}}(\beta) \quad \text{s.t. } \langle \beta,\beta\rangle = 1. \tag{6}$$

The cost function in (6) is a quadratic form of β,

$$Q_{\mathrm{ols}}(\beta) = \langle \Psi_{\mathrm{ols}}\beta,\beta\rangle,$$

where Ψ_{ols} is a self-adjoint linear operator on $\mathbb{V}$. Therefore the global minimum point

$$\hat{\beta}_{\mathrm{ols}} := (\hat{A}_{\mathrm{ols}},\hat{b}_{\mathrm{ols}},\hat{d}_{\mathrm{ols}}) = \arg \min_{\langle\beta,\beta\rangle=1} Q_{\mathrm{ols}}(\beta)$$

is a normalized eigenvector of Ψ_{ols}, corresponding to the minimum eigenvalue $\lambda_{\min}(\Psi_{\mathrm{ols}})$. In order to find the operator $\Psi_{\mathrm{ols}} : \mathbb{V} \to \mathbb{V}$, we calculate the derivative $Q'_{\mathrm{ols}} = \mathrm{d}Q_{\mathrm{ols}}/\mathrm{d}\beta$.

The derivative of $q_{\mathrm{ols}}(\beta;x)$ with respect to β is

$$q'_{\mathrm{ols}}(\beta;x) = 2(x^{\top}Ax + b^{\top}x + d)(xx^{\top},x,1)$$
$$= 2\langle (xx^{\top},x,1),\beta\rangle\ (xx^{\top},x,1).$$

It defines a self-adjoint and positive semidefinite linear operator $\psi_{\mathrm{ols}}(x)$ on $\mathbb{V}$,

$$\psi_{\mathrm{ols}}(x)\beta := \langle (xx^{\top},x,1),\beta\rangle\ (xx^{\top},x,1) \quad \text{for all } \beta \in \mathbb{V}.$$

Thus

$$\Psi_{\mathrm{ols}} = \sum_{l=1}^{m} \psi_{\mathrm{ols}}(x^{(l)}). \tag{7}$$

Remark 1. The cost function Q_{ols} is quadratic in the parameters β, so that it has a matrix representation. For $A \in \mathbb{S}$, let $\mathrm{vec}_s(A)$ denotes the $n_A := n(n+1)/2$ dimensional

vector of the elements in the upper triangular part of A taken column-wise. There exists a matrix $M \in \mathbb{R}^{(n_A+n+1)\times(n_A+n+1)}$ associated with the operator Q_{ols}, such that

$$Q_{\mathrm{ols}}(\beta) = \begin{bmatrix} \mathrm{vec_s}(A) \\ b \\ d \end{bmatrix}^{\top} M \begin{bmatrix} \mathrm{vec_s}(A) \\ b \\ d \end{bmatrix} \quad \text{for all } A \in \mathbb{S},\ b \in \mathbb{R}^n,\ d \in \mathbb{R}. \tag{8}$$

Using the matrix representation (8), the whole derivation of the OLS estimator, and subsequently the one of the ALS estimator, can be carried out in linear algebra notation. We use the matrix representation approach in Markovsky et al. (2002), where the computation of the estimators is treated. In this paper, we use the abstract operator notion.

4. ALS estimator with known noise variance

The OLS estimator is readily computable but it is inconsistent. We propose an adjustment procedure, that defines a consistent estimator. The proposed approach is due to Kukush and Zwanzig (2002), and it is related to the method of corrected score functions, (see Carroll et al., 1995, Section 6.5). The model (3) is quadratic and similar adjustment for a *bilinear model*, arising in motion analysis, is proposed in Kukush et al. (2002)

We define the *elementary ALS cost function* $q_{\mathrm{als}}(\beta; x)$ by

$$\mathbf{E} q_{\mathrm{als}}(\beta; \bar{x} + e) = q_{\mathrm{ols}}(\beta; \bar{x}) \quad \text{for all } \beta \in \mathbb{V} \quad \text{and } \bar{x} \in \mathbb{R}^n, \tag{9}$$

where e is $\mathrm{N}(0, \bar{\sigma}^2 I_n)$ distributed. The *ALS cost function* is the sum of the elementary ALS cost functions for all data points

$$Q_{\mathrm{als}}(\beta) := \sum_{l=1}^{m} q_{\mathrm{als}}(\beta; x^{(l)}) \quad \text{for all } \beta \in \mathbb{V}.$$

The *ALS estimator* $\hat{\beta}_{\mathrm{als}}$ is defined as the global minimum point of the following optimization problem:

$$\min_{\beta} Q_{\mathrm{als}}(\beta) \quad \text{s.t.} \quad \langle \beta, \beta \rangle = 1. \tag{10}$$

The solution of (10) is described in the following theorem.

Theorem 2. *The ALS estimator* $\hat{\beta}_{\mathrm{als}}$ *is the normalized eigenvector of*

$$\Psi_{\mathrm{als}} := \sum_{l=1}^{m} \psi_{\mathrm{als}}(x^{(l)}),$$

corresponding to $\lambda_{\min}(\Psi_{\mathrm{als}})$, *where*

$$\psi_{\mathrm{als}}(x)\beta = (g_1(x)[A] + g_2(x)[b] + g_3(x)[d], \\ g_4(x)[A] + g_5(x)[b] + xd, g_6(x)[A] + x^{\top} b + d), \tag{11}$$

the functions g_s, $s=1,\dots,6$ *are defined by*

$$[g_1(x)[A]]_{pq}=\sum_{i,j=1}^{n} a_{ij}f_{ijpq}(x) \quad \textit{for all } A\in\mathbb{S}, \tag{12}$$

$$[g_2(x)[b]]_{pq}=\sum_{i=1}^{n} b_i f_{ipq}(x) \quad \textit{for all } b\in\mathbb{R}^n, \tag{13}$$

$$g_3(x)[d]=(xx^\top-\bar\sigma^2 I_n)d \quad \textit{for all } d\in\mathbb{R}, \tag{14}$$

$$[g_4(x)[A]]_{p}=\sum_{i,j=1}^{n} a_{ij}f_{ijp}(x) \quad \textit{for all } A\in\mathbb{S}, \tag{15}$$

$$g_5(x)[b]=(xx^\top-\bar\sigma^2 I_n)b \quad \textit{for all } b\in\mathbb{R}^n, \tag{16}$$

$$g_6(x)[A]=x^\top Ax-\bar\sigma^2\ \mathrm{trace}(A) \quad \textit{for all } A\in\mathbb{S}, \tag{17}$$

the functions f_{ijpq} *in* (12) *are defined by*

- *if all* i, j, p, q *are different, then* $f_{ijpq}(x)=x_ix_jx_px_q$;
- *if* $i=j=p$, $q\neq i$ *(with permutations), then* $f_{iiiq}(x)=x_qt_3(x_i)$;
- *if* $i=j=p=q$, *then* $f_{iiii}(x)=t_4(x_i)$;
- *if* $i=j$, $p=q$, $i\neq p$, *then* $f_{iipp}(x)=t_2(x_i)t_2(x_p)$;
- *if* $i=j$ *and* i, p, q *are different, then* $f_{iipq}(x)=x_px_qt_2(x_i)$;

the functions f_{ijp} *in* (13) *and* (15) *are defined by*

- *if* $i=p=q$, *then* $f_{iii}(x)=t_3(x_i)$;
- *if* $i=p$, $p\neq q$ *(with permutations), then* $f_{iiq}(x)=t_2(x_i)x_q$;
- *if* i,p,q *are different, then* $f_{ipq}(x)=x_ix_px_q$,

and the functions t_k, $k=1,\dots,4$ *are defined by*

$$t_1(\xi)=\xi,\quad t_2(\xi)=\xi^2-\bar\sigma^2,\quad t_3(\xi)=\xi^3-3\xi\bar\sigma^2,$$
$$\text{and}\quad t_4(\xi)=\xi^4-6\xi^2\bar\sigma^2+3\bar\sigma^4. \tag{18}$$

Proof. Consider Eq. (9), which implicitly defines q_{als}. It is the following *deconvolution* problem:

$$\left(\frac{1}{2\pi\bar\sigma^2}\right)^{n/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} q_{\mathrm{als}}(\beta;\bar x+e)\prod_{i=1}^{n}\exp\left(-\frac{e_i^2}{2\bar\sigma^2}\right)\mathrm{d}e_1\cdots\mathrm{d}e_n=q_{\mathrm{ols}}(\beta;\bar x). \tag{19}$$

Since $q_{\mathrm{ols}}(\beta;\bar x)$ is quadratic in β, Eq. (19) holds for all β in $\mathbb{V}$, and the integral does not depend on β, q_{als} must be quadratic in β for all x. Thus

$$q_{\mathrm{als}}(\beta;x)=\langle\psi_{\mathrm{als}}(x)\beta,\beta\rangle \quad \text{for all } \beta\in\mathbb{V} \text{ and } x\in\mathbb{R}^{n\times 1},$$

where $\psi_{\mathrm{als}}(x)$ is a self-adjoint linear operator on $\mathbb{V}$, such that

$$\mathbf{E}\psi_{\mathrm{als}}(\bar{x}+e)=\psi_{\mathrm{ols}}(\bar{x}) \quad \text{for all } \bar{x}\in\mathbb{R}^n. \tag{20}$$

Then Q_{als} is also quadratic in β,

$$Q_{\mathrm{als}}(\beta)=\langle \Psi_{\mathrm{als}}\beta,\beta\rangle \quad \text{for all } \beta\in\mathbb{V},$$

where $\Psi_{\mathrm{als}}:=\sum_{l=1}^{m}\psi_{\mathrm{als}}(x^{(l)})$, and the *ALS estimator* $\hat{\beta}_{\mathrm{als}}$ is the normalized eigenvector of Ψ_{als}, corresponding to $\lambda_{\min}(\Psi_{\mathrm{als}})$.

Now, we describe *the* operator $\psi_{\mathrm{als}}(x)$ that solves (20). Solving a general deconvolution problem is not possible analytically. In our case, however, the normality assumption for the noise makes the problem tractable. Looking at the right-hand-side of (20),

$$\psi_{\mathrm{ols}}(x)\beta=(x^{\top}Ax+x^{\top}b+d)(xx^{\top},x,1) \quad \text{where } \beta=(A,b,d),$$

we see that the problem splits into six independent problems

$$\mathbf{E}\psi_{\mathrm{als}}(\bar{x}+e)=h_s(x)[\beta] \quad \text{for all } \bar{x}\in\mathbb{R}^n \text{ and for } s=1,\dots,6, \tag{21}$$

where $h_i(x)[\beta]$ are the summands in the expansion of $\psi_{\mathrm{ols}}(x)\beta$:

$$h_1(x)[A]:=xx^{\top}(x^{\top}Ax),\quad h_2(x)[b]:=xx^{\top}(x^{\top}b),\quad h_3(x)[d]:=xx^{\top}d,$$

$$h_4(x)[A]:=x(x^{\top}Ax),\quad h_5(x)[b]:=xx^{\top}b,\quad h_6(x)[A]:=x^{\top}Ax.$$

Let

$$g_1:\mathbb{S}\to\mathbb{S},\qquad g_2:\mathbb{R}^{n\times 1}\to\mathbb{S},\qquad g_3:\mathbb{R}\to\mathbb{S},$$

$$g_4:\mathbb{S}\to\mathbb{R}^{n\times 1},\quad g_5:\mathbb{R}^{n\times 1}\to\mathbb{R}^{n\times 1},\quad g_6:\mathbb{S}\to\mathbb{R},$$

be the solutions of (21), then the solution of (20) is given by (11).

Some of the functions g_s can be found by inspection. For example, the solution of the deconvolution equation for h_3 is (14). Similarly, the solution of the deconvolution equation for h_6 is (17). Due to the symmetry, $g_2=(g_4)^*$ and $g_3=(g_6)^*$, where g^* denotes the conjugate operator of g.

Next, we describe the other functions g_s. Let $A=[a_{ij}]$. Then

$$[h_1(x)[A]]_{pq}=\sum_{i,j=1}^{n}a_{ij}x_ix_jx_px_q \quad \text{for all } A\in\mathbb{S}.$$

Therefore the solution of the corresponding deconvolution problem is (12), where f_{ijpq} is a polynomial of the fourth order with the property

$$\mathbf{E}f_{ijpq}(\bar{x}+e)=\bar{x}_i\bar{x}_j\bar{x}_p\bar{x}_q. \tag{22}$$

The polynomials $t_k:\mathbb{R}\to\mathbb{R}$, $k=1,\dots,4$, defined in (18), have the property

$$\mathbf{E}t_k(\bar{\xi}+\tilde{\xi})=\bar{\xi}^k, \quad \text{for } k=1,\dots,4, \text{ and for all } \bar{\xi}\in\mathbb{R}, \text{ and } \tilde{\xi}\sim \mathrm{N}(0,\bar{\sigma}^2).$$

Then the functions f_{ijpq} defined in the theorem have the desired property (22). Similarly for

$$[h_2(x)[b]]_{pq}=\sum_{i=1}^{n}b_ix_ix_px_q,$$

the solution of the deconvolution problem (21) is (13), where f_{ipq} are defined in the theorem. Finally,

$$[h_4(x)[A]]_p = \sum_{i,j=1}^{n} a_{ij} x_i x_j x_p,$$

and the solution of the deconvolution problem (21) is (13). Thus the adjusted operator $\psi_{\text{als}}(x)$ in $\mathbb{V}$ is described thoroughly. □

Remark 3. If the given data is noise free, i.e., $\bar{\sigma} = 0$, then $x^{(l)} = \bar{x}^{(l)}$ for all l, and the solution of the deconvolution equation (19) is q_{ols}. In this case, the ALS estimator coincides with the OLS estimator.

5. Consistency of the ALS estimator

Let $n_\beta := \dim \mathbb{V} = n(n+1)/2 + n + 1 = (n+1)(n+2)/2$, and let

$$\lambda_1(\bar{\Psi}_{\text{ols}}/m) \geqslant \lambda_2(\bar{\Psi}_{\text{ols}}/m) \geqslant \cdots \geqslant \lambda_{n_\beta}(\bar{\Psi}_{\text{ols}}/m) = 0$$

be the eigenvalues of $\bar{\Psi}_{\text{ols}}/m$, where $\bar{\Psi}_{\text{ols}}$ is given in (7) with $\bar{x}^{(l)}$. We need the following assumptions:

(iii) There exists $m_0 \geqslant 1$ and $\varepsilon_0 > 0$, such that

$$\lambda_{n_\beta - 1}(\bar{\Psi}_{\text{ols}}/m) \geqslant \varepsilon_0 \quad \text{for all } m \geqslant m_0.$$

(iv) There exists a constant $\varepsilon_1 > 0$ and a number $\gamma \in [0, 1)$, such that

$$\frac{1}{m} \sum_{l=1}^{m} \|\bar{x}^{(l)}\|^6 \leqslant \varepsilon_1 m^\gamma \quad \text{for all } m \geqslant 1.$$

Assumption (iii) is a *contrast condition*, see the discussion in Kukush and Zwanzig (2002). Similar condition is used in Kukush et al. (2003). Assumption (vi) is a restriction from above. If $\bar{A}$ is positive definite and $4d < b^\top A^{-1} b$, then $S(\bar{A}, \bar{b}, \bar{d})$ is an elliptic surface, which is bounded, and (iv) holds with $\gamma = 0$.

Let

$$\operatorname{dist}(\beta_1, \beta_2) := \|\beta_1 - \beta_2\|_{\mathbb{V}}$$

and

$$\operatorname{dist}(\beta_1, \{\pm\beta_2\}) := \min\{\operatorname{dist}(\beta_1, \beta_2), \operatorname{dist}(\beta_1, -\beta_2)\}.$$

Theorem 4 (Strong consistency). *Assume that conditions* (i)–(iv) *hold. Then the ALS estimator* $\hat{\beta}_{\text{als}}$ *and the true value* $\bar{\beta} := (\bar{A}, \bar{b}, \bar{d})$ *satisfy the following convergence property*:

$$\operatorname{dist}(\hat{\beta}_{\text{als}}, \{\pm\bar{\beta}\}) \to 0 \quad \textit{as } m \to \infty, \textit{ a.s.}$$

Corollary 5. *Under the conditions* (i)–(iv),

$$\operatorname{dist}(\hat{\beta}_{\mathrm{als}}, \{\pm\bar{\beta}\}) = \frac{1}{m^{(1-\gamma)/2}}\, \mathrm{O_p}(1), \tag{23}$$

and for each $\delta > 0$,

$$\operatorname{dist}(\hat{\beta}_{\mathrm{als}}, \{\pm\bar{\beta}\}) m^{(1-\gamma)/2-\delta} \to 0 \quad \textit{as } m \to \infty, \textit{ a.s.} \tag{24}$$

Corollary 5 shows that for unbounded sequence $\{\bar{x}^{(l)},\ l \geqslant 1\}$, there may be a loss of order in the rate of convergence of the estimator. But if $\gamma = 0$, then the estimator is $\sqrt{m}$-*consistent*, i.e.,

$$\operatorname{dist}(\hat{\beta}_{\mathrm{als}}, \{\pm\bar{\beta}\}) = \mathrm{O_p}(1)/\sqrt{m}. \tag{25}$$

The statement of Theorem 4 is one of the main contributions of the paper. Adjustment procedures, similar to the one described in Section 4, already appeared in the literature; first proposed in Kanatani (1994) and later developed in Cabrera and Meer (1996) and Zhang (1997). In these papers, however, consistency of the ALS estimator is not proven. Instead the notion of unbiasedness is used, i.e., $\mathbf{E}\hat{\beta}_{\mathrm{als}} = \bar{\beta}$. In the present context, however, bias is not well defined for the reason that the expectation of the ALS estimator does not exist.

Suppose we draw N realizations of the measurement errors and compute the ALS estimates, for the corresponding data sets. Let $\hat{\beta}_{\mathrm{als},k}$ be the estimate for the kth data set. Then

$$\frac{1}{N}\sum_{k=1}^{N} \|\hat{\beta}_{\mathrm{als},k}\| \to \infty \quad \text{as } N \to \infty.$$

In the context of a linear measurement error model, the fact that $\mathbf{E}\hat{\beta}_{\mathrm{als}}$ does not exist is stated in (Fuller, 1987, Exercise 13, p. 28). It is proven for a multivariate linear measurement error model in the unpublished manuscript Cheng and Kukush (2001).

6. Consistent estimator in the case of unknown noise variance

Suppose we misspecified the noise variance. The true value is $\bar{\sigma}^2$ and we construct the operator $\Psi_{\sigma^2} := \Psi_{\mathrm{als}}$, regarding σ^2 to be the true value of that parameter. We study the difference

$$\mathbf{E}_{\bar{\sigma}^2}\psi_{\sigma^2}(\bar{x}+e) - \mathbf{E}_{\sigma^2}\psi_{\sigma^2}(\bar{x}+e),$$

where $\mathbf{E}_{\bar{\sigma}^2}$ and $\mathbf{E}_{\sigma^2}$ denote the expectation with $e \sim \mathrm{N}(0, \bar{\sigma}^2 I_n)$ and $e \sim \mathrm{N}(0, \sigma^2 I_n)$, respectively.

Consider the polynomials $t_k(\xi)$, $k = 2,3,4$, given in (18). Assuming σ^2 to be the true value of the noise variance, we have

$$t_2(\xi) = \xi^2 - \sigma^2,$$

which can be written as

$$t_2(\xi) = \xi^2 - \bar{\sigma}^2 + (\bar{\sigma}^2 - \sigma^2),$$

so that for $\xi = \bar{\xi} + \tilde{\xi}$, with $\tilde{\xi} \sim \mathrm{N}(0, \bar{\sigma}^2)$

$$\begin{aligned}\mathbf{E}_{\bar{\sigma}^2} t_2(\bar{\xi} + \tilde{\xi}) &= \mathbf{E}_{\bar{\sigma}^2}((\bar{\xi} + \tilde{\xi})^2 - \bar{\sigma}^2 + (\bar{\sigma}^2 - \sigma^2)) \\ &= \bar{\xi}^2 + (\bar{\sigma}^2 - \sigma^2) \quad \text{for all } \xi \in \mathbb{R}.\end{aligned}$$

Next, for the polynomial t_3, we have

$$\begin{aligned}t_3(\xi) &= \xi^3 - 3\xi\sigma^2 \\ &= \xi^3 - 3\xi\bar{\sigma}^2 + 3\xi(\bar{\sigma}^2 - \sigma^2),\end{aligned}$$

so that

$$\begin{aligned}\mathbf{E}_{\bar{\sigma}^2} t_3(\bar{\xi} + \tilde{\xi}) &= \mathbf{E}_{\bar{\sigma}^2}((\bar{\xi} + \tilde{\xi})^3 - 3(\bar{\xi} + \tilde{\xi})\bar{\sigma}^2 + 3(\bar{\xi} + \tilde{\xi})(\bar{\sigma}^2 - \sigma^2)) \\ &= \bar{\xi}^3 + 3\bar{\xi}(\bar{\sigma}^2 - \sigma^2).\end{aligned}$$

Finally for the polynomial t_4, we have

$$\begin{aligned}t_4(\xi) &= \xi^4 - 6\xi^2\sigma^2 + 3\sigma^4 \\ &= \xi^4 - 6\xi^2\bar{\sigma}^2 + 3\bar{\sigma}^4 + 6\xi^2(\bar{\sigma}^2 - \sigma^2) - 3(\bar{\sigma}^4 - \sigma^4),\end{aligned}$$

so that

$$\begin{aligned}\mathbf{E}_{\bar{\sigma}^2} t_4(\bar{\xi} + \tilde{\xi}) &= \bar{\xi}^4 + 6(\bar{\sigma}^2 - \sigma^2)\mathbf{E}_{\bar{\sigma}^2}(\xi + \tilde{\xi})^2 - 3(\bar{\sigma}^4 - \sigma^4) \\ &= \bar{\xi}^4 + 6(\bar{\sigma}^2 - \sigma^2)(\bar{\xi}^2 + \bar{\sigma}^2) - 3(\bar{\sigma}^4 - \sigma^4).\end{aligned}$$

Thus

$$\mathbf{E}_{\bar{\sigma}^2} t_k(\bar{\xi} + \tilde{\xi}) = \bar{x}_i^k + (\bar{\sigma}^2 - \sigma^2) z_k(\bar{\xi}; \bar{\sigma}^2 - \sigma^2), \quad \text{for } k = 2, 3, 4,$$

with

$$z_2(x; \bar{\sigma}^2 - \sigma^2) := 1, \quad z_3(x; \bar{\sigma}^2 - \sigma^2) := 3x,$$

and

$$\begin{aligned}z_4(x; \bar{\sigma}^2 - \sigma^2) &:= 6(x^2 + \bar{\sigma}^2) - 3(\bar{\sigma}^2 + \sigma^2) \\ &= 6x^2 + 3(\bar{\sigma}^2 - \sigma^2).\end{aligned}$$

For the polynomials f_{ijp}, defined in Section 4, we have

$$\mathbf{E}_{\bar{\sigma}^2} f_{ijp}(\bar{x} + e) = \bar{x}_i \bar{x}_j \bar{x}_p + (\bar{\sigma}^2 - \sigma^2) z_{ijp}(\bar{x}; \bar{\sigma}^2 - \sigma^2),$$

where z_{ijp} does not depend on $\bar{\sigma}^2 - \sigma^2$ or the dependence is linear, e.g., for $i = j \neq p$, we have

$$\begin{aligned}\mathbf{E}_{\bar{\sigma}^2} f_{iip}(\bar{x} + e) &= (\mathbf{E}_{\bar{\sigma}^2} t_2(\bar{x}_i + e_i))\bar{x}_p \\ &= (\bar{x}_i)^2 \bar{x}_p + (\bar{\sigma}^2 - \sigma^2)\bar{x}_p,\end{aligned}$$

then $z_{iip} = \bar{x}_p$. Similarly, the polynomials $z_{ijpq}(\bar{x}; \bar{\sigma}^2 - \sigma^2)$ are defined by

$$\mathbf{E}_{\bar{\sigma}^2} f_{ijpq}(\bar{x} + e) = \bar{x}_i \bar{x}_j \bar{x}_p \bar{x}_q + (\bar{\sigma}^2 - \sigma^2) z_{ijpq}(\bar{x}; \bar{\sigma}^2 - \sigma^2).$$

For the operator $\psi_{\bar\sigma^2}(x)$, which is constructed starting from the value σ^2, we have

$$\mathbf{E}_{\bar\sigma^2}\psi_{\bar\sigma^2}(\bar x+e)=\psi_{\mathrm{ols}}(\bar x)+(\bar\sigma^2-\sigma^2)z(\bar x;\bar\sigma^2-\sigma^2),$$

where $z=:(z_A,z_b,z_d)$ is a certain self-adjoint operator on $\mathbb{V}$, which either depends linearly on $\bar\sigma^2-\sigma^2$ or does not depend on $\bar\sigma^2-\sigma^2$. For example z_d does not depend on $\bar\sigma^2-\sigma^2$. Indeed

$$z_d\beta=\mathrm{trace}(A),$$

because for $x=\bar x+e$,

$$\mathbf{E}_{\bar\sigma^2}(g_5(x,A)+x^\top b+d)=(\bar x^\top A\bar x+x^\top b+d)+(\bar\sigma^2-\sigma^2)\ \mathrm{trace}(A).$$

Now, for the sum over m observations, the operator $Z(\bar\sigma^2-\sigma^2)$ is defined by

$$Z(\bar\sigma^2-\sigma^2):=\sum_{l=1}^{m}z(\bar x^{(l)};\bar\sigma^2-\sigma^2),$$

and then

$$\mathbf{E}_{\bar\sigma^2}\Psi_{\sigma^2}=\bar\Psi_{\mathrm{ols}}+(\bar\sigma^2-\sigma^2)Z(\bar\sigma^2-\sigma^2).$$

We need the following assumptions in order to estimate $\bar\sigma^2$:

(v) There exists a number $\kappa_0\in[1,n]$ and $\varepsilon_2>0$, such that

$$\frac{1}{m}\sum_{l=1}^{m}(\bar x_{\kappa_0}^{(l)})^2\leqslant\varepsilon_2\quad\text{for all } m\geqslant 1.$$

We define

$$F_m(\alpha):=\frac{(\bar\Psi_{\mathrm{ols}}+\alpha Z(\alpha))}{m}\quad\text{for } \alpha\in\mathbb{R}.$$

(vi) For each $v>0$ and $\varepsilon\in(0,v)$,

$$\liminf_{m\to\infty}\min_{|\alpha|\in[\varepsilon^2,v^2]}|\lambda_{\min}(F_m(\alpha))|>0.$$

We introduce the score function

$$U_m(\sigma^2):=\lambda_{\min}\left(\frac{\Psi_{\sigma^2}}{m}\right)\quad\text{for } 0\leqslant\sigma^2<\infty. \tag{26}$$

Lemma 6. *Assume that conditions* (i)–(iii) *and* (vi) *hold. Then with probability one*

$$U_m(0)\geqslant 0\quad\textit{and}\quad\lim_{\sigma^2\to\infty}U_m(\sigma^2)=-\infty.$$

We define an estimator $\hat\sigma^2$, as a random variable, such that

$$U_m(\hat\sigma^2)=0,\quad\text{a.s.} \tag{27}$$

The function U_m is continuous in $\sigma^2\in[0,\infty)$, and by Lemma 6 there exists a solution (may be not unique) of Eq. (27).

The reason in definitions (26) and (27) is as follows. For large m,

$$U_m(\bar{\sigma}^2) \approx \lambda_{\min}\left(\frac{\bar{\Psi}_{\text{ols}}}{m}\right) = 0.$$

On the other hand, for large m and misspecified noise variance $\sigma^2 \neq \bar{\sigma}^2$,

$$U_m(\sigma^2) \approx \lambda_{\min}\left(\frac{\bar{\Psi}_{\text{ols}} + (\bar{\sigma}^2 - \sigma^2)Z(\bar{\sigma}^2 - \sigma^2)}{m}\right).$$

By condition (vi), $U_m(\sigma^2)$ is asymptotically separated from 0. Thus we expect that the solution of (27) is close to $\bar{\sigma}^2$.

Now, we prove that the estimator σ^2 is bounded in m, a.s.

Lemma 7. *Assume that conditions* (i)–(v) *hold. Then*

$$\sup_{m \geqslant 1} \hat{\sigma}^2_m < \infty, \quad \text{a.s.}$$

Lemma 8. *Assume that conditions* (i)–(vi) *hold. Then*

$$\hat{\sigma}^2 \to \bar{\sigma}^2, \quad \text{as } m \to \infty, \text{ a.s.},$$

where $\bar{\sigma}^2$ is the true value of the parameter σ^2.

(vii) There exists an ε_3, such that $1/m \sum_{l=1}^{m} \|\vec{x}^{(l)}\|^2 \leqslant \varepsilon_3$, for all $m \geqslant 1$.

With unknown $\bar{\sigma}^2$, the ALS estimator $\hat{\beta}$ is defined as a normalized vector satisfying $\hat{\Psi}_{\sigma^2}\hat{\beta} = 0$.

Theorem 9. *Let $\bar{\sigma}^2$ be unknown. Assume* (i)–(iv), (vi), and (vii). *Then*

$$\operatorname{dist}(\hat{\beta}, \{\pm\bar{\beta}\}) \to 0 \quad \text{as } m \to \infty, \text{ a.s.}$$

7. Fitting conic sections

Now, we suppose that the true surface belongs to the class of surfaces

$$C(A_c, c) = \{x \in \mathbb{R}^n \colon (x - c)^\top A_c (x - c) = 1\} \tag{28}$$

for some true values $\bar{A}_c$ and $\bar{c}$ of the parameters A_c and c. Here A_c ("c" stands for conic) is a non-singular symmetric $n \times n$ matrix, and $c \in \mathbb{R}^n$ is the center of the surface. The equation defining $C(A_c, c)$ can be written as

$$x^\top A_c x - 2(A_c c)^\top x + c^\top A_c c - 1 = 0$$

or, with $\theta := (\|A_c\|_F^2 + \|2A_c c\|^2 + (c^\top A_c c - 1)^2)^{1/2}$,

$$x^\top (A_c/\theta) x - 2(A_c c/\theta)^\top x + (c^\top A_c c - 1)/\theta = 0.$$

Define the new parameters

$$A := \frac{A_c}{\theta}, \quad b := -2\frac{A_c c}{\theta} \quad \text{and} \quad d := \frac{c^\top A_c c - 1}{\theta}.$$

As defined, A, b, and d satisfy the normalizing condition (2).

We can renew the original parameters A_c and c from A, b, and d, that satisfy (2) by

$$c = -\frac{1}{2} A^{-1} b, \quad \text{and} \quad A_c = \frac{1}{c^\top A c - d} A. \tag{29}$$

Note that $\theta = c^\top A c - d$ is non-zero.

Now suppose that we observe the points $x^1, \ldots, x^m$, given in (4). The true values $\bar{x}^1, \ldots, \bar{x}^m$, satisfy

$$(\bar{x}^{(l)} - \bar{c})^\top \bar{A}_c (\bar{x}^{(l)} - \bar{c}) = 1 \quad \text{for } l = 1, \ldots, m. \tag{30}$$

We rewrite (30) in the form

$$\bar{x}^{(l)\top} \bar{A} \bar{x}^{(l)} + 2\bar{b}^\top \bar{x}^{(l)} + \bar{d} = 0 \quad \text{for } l = 1, \ldots, m,$$

where

$$\bar{A} = \frac{\bar{A}_c}{\theta}, \quad \bar{b} = -2\frac{\bar{A}\bar{c}}{\theta}, \quad \bar{d} = \frac{\bar{c}^\top \bar{A} \bar{c} - 1}{\theta},$$

and

$$\theta = \sqrt{||\bar{A}_c||_F^2 + ||2\bar{A}_c\bar{c}||^2 + (\bar{c}^\top \bar{A}_c \bar{c} - 1)^2}.$$

Let the noise variance $\bar{\sigma}^2$ be unknown. Assume (i)–(vi). Then

$$\operatorname{dist}(\hat{\beta}, \{\pm\bar{\beta}\}) \to 0 \quad \text{as } m \to \infty, \text{ a.s.}, \tag{31}$$

where $\hat{\beta} := (\hat{A}, \hat{b}, \hat{d})$ is the ALS estimator of the parameters $\bar{A}$, $\bar{b}$, and $\bar{d}$. The estimator of the parameters $\bar{A}_c$ and $\bar{c}$ is

$$\hat{c} = -\frac{1}{2} \hat{A}^{-1} \hat{b} \quad \text{and} \quad \hat{A}_c = \frac{1}{\hat{c}^\top \hat{A} \hat{c} - \hat{d}} \hat{A}. \tag{32}$$

Under (i)–(vi), the estimators are well defined for $m \geqslant m_0(\omega)$, a.s., and

$$\|\hat{A}_c - \bar{A}_c\|_F^2 + \|\hat{c} - \bar{c}\|^2 \to 0 \quad \text{as } m \to \infty, \text{ a.s.}$$

Indeed from (31), we have, see the formulae in (29), that

$$\hat{c} = -\frac{1}{2} \hat{A}^{-1} \hat{b} \to -\frac{1}{2} \bar{A}^{-1} \bar{b} = \bar{c} \quad \text{as } m \to \infty, \text{ a.s.}$$

And

$$\hat{A}_c = \frac{1}{\hat{c}^\top \hat{A} \hat{c} - \hat{d}} \hat{A} \to \frac{1}{\bar{c}^\top \bar{A} \bar{c} - \bar{d}} \bar{A} = \bar{A}_c.$$

It is important here that $\bar{A}_c$ is non-singular. If $\hat{A}$ is singular, then the estimators $\hat{A}_c$ and $\hat{c}$ are not defined, and if $\hat{A}$ is non-singular but $\hat{b}^\top (\hat{A})^{-1} \hat{b} = \hat{d}$, then $\hat{A}_c$ is not defined. But that does not happen for $m \geqslant m_0(\omega)$, a.s.

8. Estimation of ellipsoid

We specialize the case described in Section 7 for elliptical surface. Let in (28) $A_c = A_e$, where A_e ("e" stands for elliptic) is a positive definite matrix. Then $C(A_c, c)$ is an elliptical surface. The true value $\bar{A}_e$ of the parameter A_e is positive definite.

We can improve estimator (32) in this case. The problem is that A_c can be indefinite. We do the following additional step. Let

$$\hat{A}_c = \sum_{i=1}^{n} \hat{\lambda}_i \hat{v}_i \hat{v}_i^\top$$

be the EVD of $\hat{A}_c$, given in (32). Then we set

$$\hat{A}_e := \sum_{i\,:\,\hat{\lambda}_i > 0} \hat{\lambda}_i \hat{v}_i \hat{v}_i^\top .$$

The estimator $\hat{A}_e$ is positive semidefinite. Moreover as $\bar{A}_e$ is positive definite now, we have

$$\|\hat{A}_e - \bar{A}_e\|_F \leqslant \|\hat{A}_c - \bar{A}_e\|_F,$$

and the estimator $\hat{A}_e$ is a strongly consistent estimator of $\bar{A}_e$, i.e., $\hat{A}_e \to \bar{A}_e$, as $m \to \infty$, a.s.

9. Simulation examples

In this section, we show simulation examples for the ellipsoid fitting problem. The aim is to illustrate the consistency results of the paper and to compare the ALS estimator with the OLS and the orthogonal regression estimators. All experiments are carried out in the environment of MATLAB.

Define the (truncated) *average relative errors* of estimation by

$$\bar{e}_A := \frac{1}{N} \sum_{k=1}^{N} \min\left(\frac{\|\hat{A}_{e,k} - \bar{A}_e\|_F}{\|\bar{A}_e\|_F}, 1 \right), \quad \bar{e}_c := \frac{1}{N} \sum_{k=1}^{N} \min\left(\frac{\|\hat{c}_k - \bar{c}\|}{\|\bar{c}\|}, 1 \right),$$

and

$$\bar{e}_\sigma := \frac{1}{N} \sum_{k=1}^{N} \min\left(\frac{|\hat{\sigma}_k - \bar{\sigma}|}{\bar{\sigma}}, 1 \right),$$

where $\hat{A}_{e,k}$, $\hat{c}_k$, and $\hat{\sigma}_k$ are the estimates obtained on the kth repetition of the estimation experiment. In each repetition, different noise realization is used. The reason for using the truncated average of the relative errors of estimation is that the expectation of the relative errors does not exist, see the discussion in Section 5. We have selected the truncation level to 100%.

In Fig. 1, we show *asymptotic plots* of $\bar{e}_A$, $\bar{e}_c$, and $\bar{e}_\sigma$ as a function of the sample size m. The true data points $\bar{x}^{(l)}$ are equidistantly spaced on the boundary of the ellipsoid and the noise variance is $\bar{\sigma}^2 = 0.36$. In the experiment $N = 1000$ repetitions are used

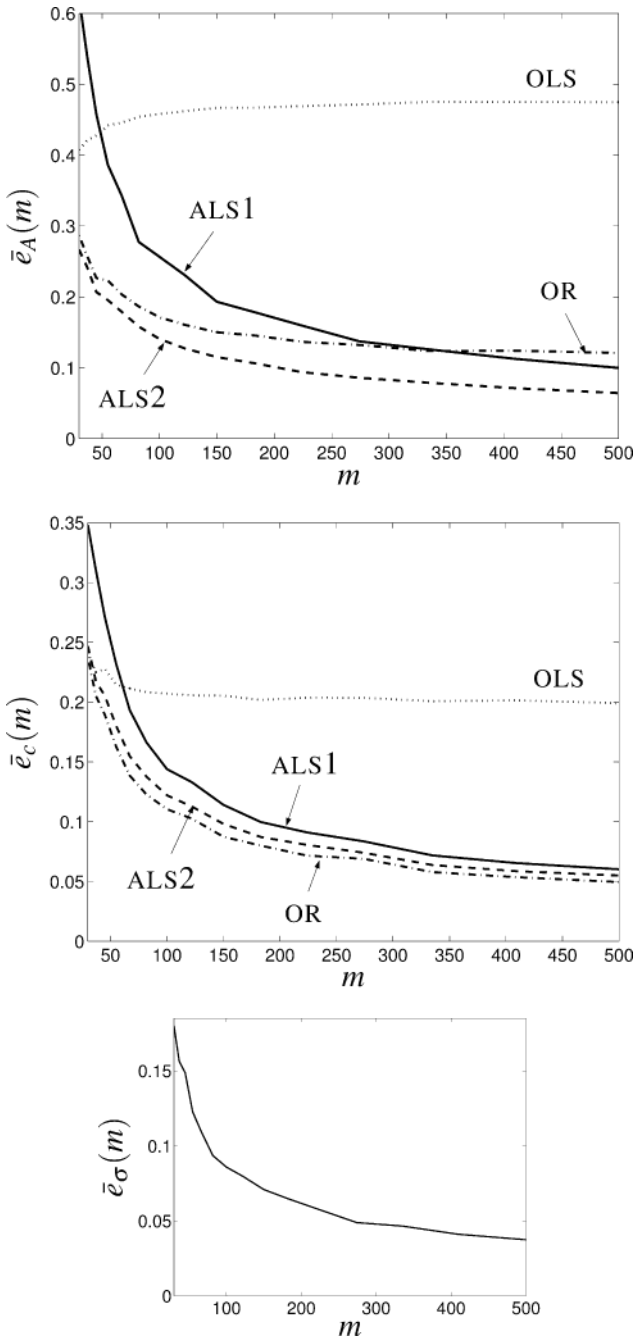

Fig. 1. Asymptotic plots of the average relative errors of estimation as a function of the sample size. OLS—ordinary least squares, OR—orthogonal regression, ALS1—ALS estimator with known $\vec{\sigma}^2$, and "ALS2"—ALS estimator with estimated noise variance.

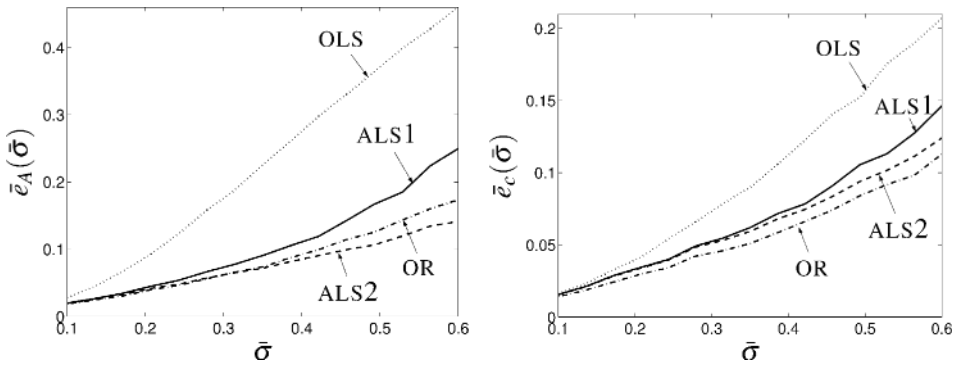

Fig. 2. Average relative error of estimation as a function of the noise standard deviation $\bar{\sigma}$. OLS—ordinary least squares, OR—orthogonal regression, ALS1—ALS estimator with known $\bar{\sigma}^2$, and "ALS2"—ALS estimator with estimated noise variance.

for each value of m. The initial approximation for the computation of the orthogonal regression estimator is the OLS estimate.

The OLS estimator is clearly biased and the error of the ALS estimator is $\sqrt{m}$-consistent. Note that the ALS estimator with unknown true noise variance (ALS2) performs consistently better than the ALS estimator with known true noise variance (ALS1).

Fig. 2 shows the relative errors of estimation as a function of the noise standard deviation. The setting of the experiment is as before. The noise standard deviation is increased from 0.1 to 0.6 and the sample size is fixed to $m=100$ data points. The initial approximation for the computation of the orthogonal regression estimator is again the OLS estimate.

The last experiment shows the performance of the estimators on a test example from Gander et al. (1994). The example is used in Gander et al. (1994) to illustrate the inadequacy of the algebraic fitting method and to show the advantage of the orthogonal regression method.

Given are data points only; even if they are generated with a true model, we do not know it. For this reason the comparison is visual. Fig. 3, left, shows the data points with the estimated ellipses superimposed on them. The dashed line represents the OLS estimate, the dashed–dotted lines, the orthogonal regression estimates (when initial approximation is the OLS estimate and the ALS estimate), and the solid line, the ALS estimate. The data points are marked with circles ($\circ$) and the centers of the estimated ellipses are marked with crosses ($\times$).

The orthogonal regression estimator is influenced by the initial approximation. Using the OLS estimate as initial approximation, the optimization algorithm (MATLAB's `fminsearch` function) converges to a local minimum. The resulting estimate is the dashed–dotted ellipse closer to the OLS estimate. Using the ALS estimate as initial approximation, the algorithm finds the global minimum point, which corresponds to the dashed–dotted ellipse closer to the ALS estimate. Although the sample size is only

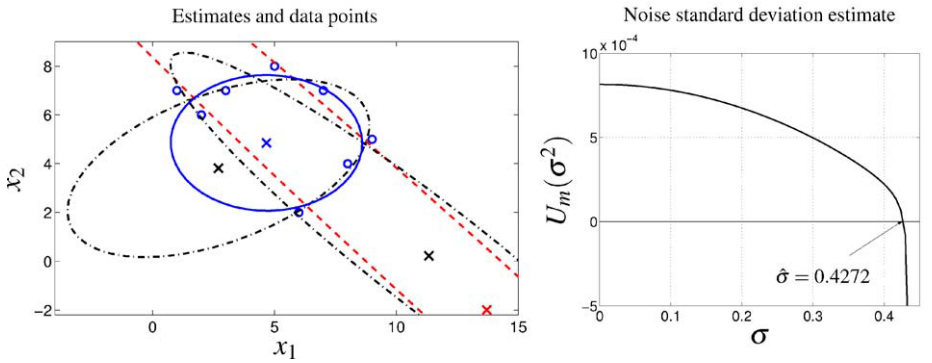

Fig. 3. Test example from Gander et al. (1994). Dashed line—OLS estimate, dashed–dotted lines—orthogonal regression estimates (with initial approximation, the OLS estimate and the ALS estimate), solid line—ALS estimate, ○—data points, ×—centers of the estimated ellipses.

$m = 8$ data points, the ALS estimator gives good estimate and is comparable with the orthogonal regression estimate, corresponding to the global minimum point.

Fig. 3, right, shows the functions U_m used for the estimation of the noise variances. From the given data, we compute an upper bound of the true noise standard deviation

$$v := \left(\frac{1}{n} \left(\frac{1}{m} \sum_{l=1}^{m} \|x_c^{(l)}\|^2 - \min_{1 \leqslant l \leqslant m} \|x_c^{(l)}\|^2 \right) \right)^{1/2} \quad \text{where } x_c^{(l)} := x^{(l)} - \frac{1}{m} \sum_{k=1}^{m} x^{(k)},$$

and use a bisection method, see Gill et al. (1999), to find a zero of U_m in the interval $\sigma \in [0, v]$. For the example there is a unique zero in the interval $[0, v]$, which corresponds to the noise standard deviation estimate.

10. Conclusions

We have presented a consistent estimator for the parameters of an implicit quadratic measurement error model. The method used is the adjustment procedure is due to Kukush and Zwanzig (2002). We give conditions, under which the estimator is strongly consistent. The adjustment needs the true noise variance. We show, however, a procedure to estimate the noise variance. This procedure defines a consistent estimator of the model parameters with unknown noise variance. The quadratic model is used for the conic section and ellipsoid fitting problems. We give simulation results for the ellipsoid estimation that illustrate the consistency of the ALS estimator.

The results are derived under the assumption that the measurement errors are normal. They can be generalized, however, for homogeneous errors described with one parameter density function. A procedure, similar to the one presented in the paper, can be derived for the estimation of the noise parameter. An open question is how

the normalizing conditions for the parameters affects the efficiency of the estimator. In particular, what is the optimal normalizing condition in terms of efficiency.

Acknowledgements

A. Kukush is supported by a postdoctoral research fellowship of the Belgian office for Scientific, Technical and Cultural Affairs, promoting Scientific and Technical Collaboration with Central and Eastern Europe. S. Van Huffel is a full professor and I. Markovsky is a research assistant with the Katholieke Universiteit Leuven. I. Markovsky is supported by a K.U. Leuven doctoral scholarship.

This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction (IUAP Phase V-22), initiated by the Belgian State, Prime Minister's Office—Federal Office for Scientific, Technical and Cultural Affairs, of the Concerted Research Action (GOA) projects of the Flemish Government MEFISTO-666 (Mathematical Engineering for Information and Communication Systems Technology), of the IDO/99/03 and IDO/02/009 projects (K.U. Leuven) "Predictive computer models for medical classification problems using patient data and expert knowledge", of the FWO projects G.0078.01 and G.0270.02.

Appendix.

Proofs of the statements

Proof of Theorem 4. Under assumption (iv),

$$\frac{1}{m}\sum_{l=1}^{m} f_{ijpq}(\vec{x}^{(l)} + e^{(l)}) - \frac{1}{m}\sum_{l=1}^{m} \mathbf{E}\, f_{ijpq}(\vec{x}^{(l)} + e^{(l)}) \to 0 \quad \text{as } m \to \infty, \text{ a.s.}$$

To show this, we will restrict our attention to the most unfavorable case $i = j = p = q$. Then $f_{iiii}(x) = x_i^4 - 6x_i^2\bar{\sigma}^2 + 3\bar{\sigma}^4$ and we have to show that

$$\frac{1}{m}\sum_{l=1}^{m} ((x_i^{(l)})^4 - 6(x_i^{(l)})^2\bar{\sigma}^2 + 3\bar{\sigma}^4) - \frac{1}{m}\sum_{l=1}^{m} (\vec{x}_i^{(l)})^4 \to 0 \quad \text{as } m \to \infty, \text{ a.s.} \tag{33}$$

Now, $x_i^{(l)} = \vec{x}_i^{(l)} + e_i^{(l)}$. Then

$$\begin{aligned}(x_i^{(l)})^4 &- 6(x_i^{(l)})^2\bar{\sigma}^2 + 3\bar{\sigma}^4 - (\vec{x}_i^{(l)})^4 \\ &= 4(\vec{x}_i^{(l)})^3 e_i^{(l)} + 6(\vec{x}_i^{(l)})^2((e_i^{(l)})^2 - \bar{\sigma}^2) \\ &\quad + 4\vec{x}_i^{(l)}(e_i^{(l)})^3 - 12\vec{x}_i^{(l)}e_i^{(l)} + ((e_i^{(l)})^4 - 6e_i^{(l)}\bar{\sigma}^2 + 3\bar{\sigma}^4).\end{aligned} \tag{34}$$

Here the most unfavorable summand is $4(\vec{x}_i^{(l)})^3 e_i^{(l)}$. We consider

$$\eta_m = \frac{1}{m}\sum_{l=1}^{m} (\vec{x}_i^{(l)})^3 e_i^{(l)}.$$

By the Rosenthal inequality, see Rosenthal (1970), we have that

$$\mathbf{E}|\eta_m|^{2+\rho} = \frac{1}{m^{2+\rho}}\,\mathbf{E}\left|\sum_{l=1}^{m}(\bar{x}_i^{(l)})^3 e_i^{(l)}\right|^{2+\rho}$$

$$\leqslant \frac{1}{m^{2+\rho}}\left(\sum_{l=1}^{m}(\bar{x}_i^{(l)})^6\right)^{(2+\rho)/2} C(\rho,\bar{\sigma}),$$

for arbitrary $\rho > 0$, where the constant $C(\rho,\bar{\sigma})$ depends only on ρ and $\bar{\sigma}$. Next, by (iv) we have

$$\mathbf{E}|\eta_m|^{2+\rho} = \text{const}\,\frac{1}{m^{1+\rho/2}}\left(\frac{1}{m}\sum_{l=1}^{m}\|\bar{x}^{(l)}\|^6\right)^{1+\rho/2}$$

$$\leqslant \text{const}\,\frac{1}{m^{(1+\rho/2)(1-\gamma)}}.$$

We choose and fix ρ large enough in order to have the inequality $(1+\rho/2)(1-\gamma) > 1$. Then $\sum_{m=1}^{\infty}\mathbf{E}|\eta_m|^{2+\rho} < \infty$, therefore by the Chebyshev inequality and the Borel–Cantelli lemma, see Papoulis (1991), $\eta_m \to 0$, as $m \to \infty$, a.s. In a similar way the other summands of (34) being averaged for $l = 1,\ldots,m$, tend to zero as $m \to \infty$, a.s., e.g., for

$$\tau_m := \frac{1}{m}\sum_{l=1}^{m}(\bar{x}_i^{(l)})^2((e_i^{(l)})^2 - \bar{\sigma}^2)$$

we have

$$\mathbf{E}|\tau_m|^{2+\rho} \leqslant \text{const}\,\frac{1}{m^{1+\rho/2}}\left(\frac{1}{m}\sum_{l=1}^{m}\|\bar{x}^{(l)}\|^4\right)^{1+\rho/2}$$

$$\leqslant \text{const}\,\frac{1}{m^{1+\rho/2}}\left(\frac{1}{m}\sum_{l=1}^{m}\|\bar{x}^{(l)}\|^6\right)^{4/6(1+\rho/2)}$$

$$\leqslant \text{const}\,\frac{1}{m^{(1+\rho/2)(1-2\gamma/3)}} \quad \text{for } \rho > 0,$$

and the inequality $(1+\rho/2)(1-2\gamma/3) > 1$, which holds for large ρ, implies $\tau_m \to 0$, as $m \to \infty$, a.s. Thus (33) holds.

But $\mathbf{E}f_{ijpq}(\bar{x}^{(l)} + e^{(l)}) = \bar{x}_i^{(l)}\bar{x}_j^{(l)}\bar{x}_p^{(l)}\bar{x}_q^{(l)}$, therefore

$$\frac{1}{m}\sum_{l=1}^{m} f_{ijpq}(\bar{x}^{(l)} + e^{(l)}) - \frac{1}{m}\sum_{l=1}^{m}\bar{x}_i^{(l)}\bar{x}_j^{(l)}\bar{x}_p^{(l)}\bar{x}_q^{(l)} \to 0 \quad \text{as } m \to \infty, \text{ a.s.} \tag{35}$$

Similar convergence holds for $f_{ijp}(\bar{x}^{(l)})$. This implies that for

$$\bar{\Psi}_{\mathrm{ols}} := \sum_{l=1}^{m} \psi_{\mathrm{ols}}(\bar{x}),$$

we have

$$\left\| \frac{1}{m}\,\Psi_{\mathrm{als}} - \frac{1}{m}\,\bar{\Psi}_{\mathrm{ols}} \right\| \to 0, \quad \text{as } m \to \infty, \text{ a.s.} \tag{36}$$

Let

$$\lambda_1(\Psi_{\mathrm{als}}/m) \geqslant \lambda_2(\Psi_{\mathrm{als}}/m) \geqslant \cdots \geqslant \lambda_{n_\beta}(\Psi_{\mathrm{als}}/m)$$

be the eigenvalues of Ψ_{als}/m. Suppose that $\|\Psi_{\mathrm{als}}/m - \bar{\Psi}_{\mathrm{ols}}/m\| \leqslant \varepsilon$, and $\varepsilon < \varepsilon_0$, where ε_0 comes from assumption (iii). Then

$$\left|\lambda_{n_\beta}(\Psi_{\mathrm{als}}/m) - \lambda_{n_\beta}(\bar{\Psi}_{\mathrm{ols}}/m)\right| \leqslant \left\| \frac{1}{m}\,\Psi_{\mathrm{als}} - \frac{1}{m}\,\bar{\Psi}_{\mathrm{ols}} \right\| \leqslant \varepsilon,$$

therefore $|\lambda_{n_\beta}(\Psi_{\mathrm{als}}/m)| \leqslant \varepsilon$. By making use of the perturbation theorems of eigenvectors, as given in Wedin (1972) and Davis and Kahan (1970), we have for the corresponding normalized eigenvectors $\hat{\beta}_{\mathrm{als}}$ and $\bar{\beta}$ that

$$\begin{aligned} \mathrm{dist}(\hat{\beta}_{\mathrm{als}}, \{\pm\bar{\beta}\}) &\leqslant \frac{\varepsilon}{\lambda_{n_\beta-1}(\bar{\Psi}_{\mathrm{ols}}/m) - \lambda_{n_\beta}(\bar{\Psi}_{\mathrm{als}}/m)} \\ &\leqslant \frac{\varepsilon}{\varepsilon_0 - \varepsilon} =: L(\varepsilon), \end{aligned} \tag{37}$$

and $\lim_{\varepsilon\to 0} L(\varepsilon) = 0$. This relation and convergence (36) prove the statement. □

Proof of Corollary 5. The convergence of (35) was studied in the proof of Theorem 4. Consider for the most unfavorable summands

$$\eta_m := \frac{1}{m}\sum_{l=1}^{m} (\bar{x}_i^{(l)})^3\, e_i^{(l)}.$$

It was shown that for each $\rho > 0$, there is a constant ε_ρ, that depends only on ρ, for which

$$\mathbf{E}|\eta_m|^{2+\rho} \leqslant \frac{\varepsilon_\rho}{m^{(2+\rho)(1-\gamma)/2}}. \tag{38}$$

Therefore

$$|\eta_m|^{2+\rho} m^{(2+\rho)(1-\gamma)/2} = \mathrm{O_p}(1),$$

and

$$\eta_m = \frac{1}{m^{(1-\gamma)/2}}\,\mathrm{O_p}(1).$$

The other summands in (35), which have expectation zero, also satisfy this relation. Then

$$\left\| \frac{1}{m}\,\Psi_{\mathrm{als}} - \frac{1}{m}\,\bar{\Psi}_{\mathrm{ols}} \right\| = \frac{1}{m^{(1-\gamma)/2}}\,\mathrm{O_p}(1),$$

and from (37), we have

$$\operatorname{dist}(\hat{\beta}_{\mathrm{als}}, \{\pm\bar{\beta}\}) = \left\| \frac{1}{m} \Psi_{\mathrm{als}} - \frac{1}{m} \bar{\Psi}_{\mathrm{ols}} \right\| \mathrm{O_p}(1)$$

$$= \frac{1}{m^{(1-\gamma)/2}} \mathrm{O_p}(1).$$

Now, we show (24). From (38), we have

$$\mathbf{E}(m^{\varepsilon}|\eta_m|^{2+\rho}) \leqslant \frac{\varepsilon_\gamma}{m^{(2+\rho)(1-\gamma)/2-\varepsilon}}, \quad \text{for } \varepsilon > 0.$$

For large enough ρ, we have

$$(2+\rho)(1-\gamma)/2 - \varepsilon > 1,$$

and then by the Chebyshev inequality and the Borel–Cantelli lemma, see Papoulis (1991),

$$m^{\varepsilon}|\eta_m|^{2+\rho} \to 0 \quad \text{as } m \to \infty, \text{ a.s.},$$

and

$$|\eta_m| m^{\varepsilon/(2+\rho)} \to 0 \quad \text{as } m \to \infty, \text{ a.s.},$$

when $\varepsilon < (2+\rho)(1-\gamma)/2 - 1$. Fix $0 < \delta < (1-\gamma)/2$. Then

$$\varepsilon/(2+\rho) = (1-\gamma)/2 - 2/(2+\rho),$$

and

$$2/(2+\rho) = \delta \quad \text{for } \rho = 2/\delta - 2.$$

Then

$$|\eta_m| m^{(1-\gamma)/2-\delta} \to 0 \quad \text{as } m \to \infty, \text{ a.s.},$$

and

$$\left\| \frac{1}{m} \Psi_{\mathrm{als}} - \frac{1}{m} \bar{\Psi}_{\mathrm{ols}} \right\| m^{(1-\gamma)/2-\delta} \to 0 \quad \text{as } m \to \infty, \text{ a.s.}$$

Then (37) implies statement (24). □

Proof of Lemma 6. We have

$$\Psi_{\sigma^2} = \Psi_{\mathrm{ols}} \quad \text{for } \sigma^2 = 0,$$

therefore

$$U_m(0) = \lambda_{\min}\left(\frac{\Psi_{\mathrm{ols}}}{m}\right) \geqslant 0.$$

(In practice, for noisy observations and for $m > n_\beta$, Ψ_{ols} is strictly positive definite operator, thus $U_m(0) > 0$.)

Now introduce the unit vector $h \in \mathbb{V}$,

$$h := (0, e_1, 0),$$

where $e_1 \in \mathbb{R}^{n\times 1}$ is $e_1 := [1\ 0 \cdots 0]^\top$. We have for the scalar product in $\mathbb{V}$,

$$\left\langle \frac{1}{m}\, \Psi_{\sigma^2} h, h \right\rangle = \left\langle \frac{1}{m}\, \Psi_{\bar\sigma^2} h, h \right\rangle + \bar\sigma^2 - \sigma^2,$$

and

$$\left\langle \frac{1}{m}\, \Psi_{\sigma^2} h, h \right\rangle \to -\infty \quad \text{as } \sigma^2 \to \infty, \text{ a.s.}$$

But

$$\lambda_{\min}\left(\frac{1}{m}\, \Psi_{\sigma^2}\right) \leqslant \left\langle \frac{1}{m}\, \Psi_{\sigma^2} h, h \right\rangle,$$

and this implies the second statement in Lemma 6. □

Proof of Lemma 7. Let κ_0 be the constant from condition (v) and let $h_0 \in \mathbb{V}$ be the unit vector

$$h_0 = (0, e_{\kappa_0}, 0),$$

where $e_{\kappa_0} \in \mathbb{R}^n$ is $e_{\kappa_0} := [0 \cdots 0\ 1\ 0 \cdots 0]^\top$, with 1 on the κ_0th position. From the definition of $\hat\sigma^2$, we have

$$0 \leqslant \left\langle \frac{1}{m}\, \Psi_{\hat\sigma^2} h_0, h_0 \right\rangle = \left\langle \frac{1}{m}\, \Psi_{\bar\sigma^2} h_0, h_0 \right\rangle + \bar\sigma^2 - \hat\sigma^2,$$

and, see (36),

$$\hat\sigma^2 \leqslant \bar\sigma^2 + \left\langle \frac{1}{m}\, \bar\Psi_{\text{ols}}\ h_0, h_0 \right\rangle + \text{o}(1) \quad \text{as } m \to \infty, \text{ a.s.}$$

Then by condition (v),

$$\begin{aligned} \hat\sigma^2 &\leqslant \bar\sigma^2 + \frac{1}{m} \sum_{l=1}^{m} (\bar x_{\kappa_0}^{(l)})^2 + \text{o}(1) \\ &\leqslant \bar\sigma^2 + \varepsilon_2 + \text{o}(1). \qquad \square \end{aligned}$$

Proof of Lemma 8. It can be shown that for each $v > 0$,

$$E_m(v, \omega) := \sup_{0 \leqslant \sigma^2 \leqslant v^2} \left\| \frac{1}{m}\, \Psi_{\sigma^2}(\omega) - F_m(\bar\sigma^2 - \sigma^2) \right\| \to 0 \quad \text{as } m \to \infty, \text{ a.s.} \tag{39}$$

We fix $\omega \in \Omega$, for which $\hat\sigma_m^2(\omega)$ is bounded in m, and for which the convergence (39) holds for every $v \in \mathbb{N}$. The sequence $\{\hat\sigma_m^2(\omega),\ m \geqslant 1\}$ belongs to the interval $[0, v^2]$. Here $v = v(\omega) \in \mathbb{N}$, and we assume that $v > \bar\sigma$. We have from (27) and (39) that

$$E_m(v, \omega) \geqslant |\lambda_{\min}(F_m(\bar\sigma^2 - \hat\sigma^2))|. \tag{40}$$

Consider any convergent subsequence $\{\hat{\sigma}^2_{m(k)}(\omega),\ k \geqslant 1\}$, say $\hat{\sigma}^2_{m(k)}(\omega) \to \sigma^2_\infty$, as $k \to \infty$. Suppose that $\sigma^2_\infty \neq \bar{\sigma}^2$. Then for certain $m_1 = m_1(\omega)$ and $\delta = \delta(\omega) > 0$, we have for all $m(k) \geqslant m_1$, that

$$|\lambda_{\min}(F_{m(k)}(\bar{\sigma}^2 - \hat{\sigma}^2))| \geqslant \min_{\delta^2 \leqslant |\alpha| \leqslant v^2} |\lambda_{\min}(F_{m(k)}(\alpha))|. \tag{41}$$

But from (40) and (41), we have for $m(k) \geqslant m_1$, that

$$\min_{\delta^2 \leqslant |\alpha| \leqslant v^2} |\lambda_{\min}(F_{m(k)}(\alpha))| \leqslant E_{m(k)}(v, \omega) \to 0 \quad \text{as } k \to \infty.$$

This contradicts assumption (vi). Therefore $\sigma^2_\infty = \bar{\sigma}^2$. Thus each convergent subsequence of $\{\hat{\sigma}^2_m(\omega),\ m \geqslant 1\}$ converges to $\bar{\sigma}^2$, therefore $\hat{\sigma}^2_m(\omega) \to \bar{\sigma}^2$, as $m \to \infty$. We fixed ω from the set Ω_0 of probability one, therefore $\hat{\sigma}^2_m \to \bar{\sigma}^2$, a.s. □

Proof of Theorem 9. The proof is similar to the proof of Theorem 2 of Kukush et al. (2003). Due to the quadratic structure of Ψ_{σ^2} with respect to σ^2 and due to (vii), we have for each $v > 0$,

$$\sup_{m \geqslant 1} \sup_{\substack{0 \leqslant \sigma_1^2 \leqslant v^2,\ 0 \leqslant \sigma_2^2 \leqslant v^2 \\ |\sigma_1^2 - \sigma_2^2| \leqslant \delta}} \left\| \frac{1}{m} \Psi_{\sigma_1^2} - \frac{1}{m} \Psi_{\sigma_2^2} \right\| \to 0 \quad \text{as } \delta \to \infty, \text{ a.s.}$$

This means that the functions $\{\Psi_{\sigma^2}/m,\ \sigma^2 \in [0, v^2];\ m \geqslant 1\}$ are equicontinuous, a.s. Therefore, see Lemma 8, for $v^2(\omega) := \sup_{m \geqslant 1} \hat{\sigma}^2_m(\omega)$ we have

$$\begin{aligned}
\left\| \frac{1}{m} \Psi_{\hat{\sigma}^2} - \frac{1}{m} \bar{\Psi}_{\text{ols}} \right\| &\leqslant \left\| \frac{1}{m} \Psi_{\hat{\sigma}^2} - \frac{1}{m} \Psi_{\bar{\sigma}^2} \right\| + \left\| \frac{1}{m} \Psi_{\bar{\sigma}^2} - \frac{1}{m} \bar{\Psi}_{\text{ols}} \right\| \\
&\leqslant \sup_{m \geqslant 1} \sup_{\substack{0 \leqslant \sigma^2 \leqslant v^2(\omega) \\ |\sigma^2 - \bar{\sigma}^2| \leqslant |\hat{\sigma}^2_m - \bar{\sigma}^2|}} \left\| \frac{1}{m} \Psi_{\sigma^2} - \frac{1}{m} \Psi_{\bar{\sigma}^2} \right\| \\
&\quad + \left\| \frac{1}{m} \Psi_{\bar{\sigma}^2} - \frac{1}{m} \bar{\Psi}_{\text{ols}} \right\| \to 0 \quad \text{as } m \to \infty, \text{ a.s.}
\end{aligned}$$

Then like in the proof of Theorem 4, we obtain that $\text{dist}(\hat{\beta}, \{\pm\bar{\beta}\}) \to 0$, as $m \to \infty$, a.s. □

References

Cabrera, J., Meer, P., 1996. Unbiased estimation of ellipses by bootstrapping. IEEE Trans. Pattern Anal. Mach. Intell. 18 (7), 752–756.

Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement error in nonlinear models. no. 63. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, London/Boca Raton.

Cheng, C., Kukush, A., 2001. Nonexistence of the first moment of the adjusted least squares estimator in multivariate errors-in-variables model, unpublished manuscript.

Davis, C., Kahan, W.M., 1970. The rotation of eigenvectors by a perturbation III. SIAM J. Numer. Anal. 7, 1–46.

Fitzgibbon, A., Pilu, M., Fisher, R., 1999. Direct least-squares fitting of ellipses. IEEE Trans. Pattern Anal. Mach. Intell. 21 (5), 476–480.

Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.

Gander, W., Golub, G.H., Strebel, R., 1994. Fitting of circles and ellipses: Least-squares fitting of circles and ellipses. BIT 34 (4), 558–578.

Gill, P.E., Murray, M., Wright, M.H., 1999. Practical Optimization. Academic Press, NY, USA.

Kanatani, K., 1994. Statistical bias of conic fitting and renormalization. IEEE Trans. Pattern Anal. Mach. Intell. 16 (3), 320–326.

Kukush, A., Zwanzig, S., 2002. On consistent estimators in nonlinear functional EIV models. In: Van Huffel, S., Lemmerling, P. (Eds.), Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications. Kluwer, USA, pp. 145–155.

Kukush, A., Markovsky, I., Van Huffel, S., 2002. Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. Comput. Statist. Data Anal. 41 (1), 3–18.

Kukush, A., Markovsky, I., Van Huffel, S., 2003. Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. Metrika 57 (3), 253–285.

Markovsky, I., Kukush, A., Van Huffel, S., 2002. Consistent least squares fitting of ellipsoids. Technical Report 02–116, Department of EE, K.U. Leuven, (Numer. Math. accepted for publication (2004)).

Naidu, L.K., 1990. An adjusted linear estimator. Comput. Statist. Data Anal. 10 (2), 143–151.

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

Nievergelt, Y., 2001. Hyperspheres and hyperplanes fitting seamlessly by algebraic constrained total least-squares. Linear Algebra Appl. 331, 43–59.

Nyquist, H., 1988. Least orthogonal absolute deviations. Comput. Statist. Data Anal. 6 (4), 361–367.

Papoulis, A., 1991. Probability, Random Variables, and Stochastic Processes. McGraw-Hill, USA.

Rosenthal, H.P., 1970. On the subspaces of $L^p(p > 2)$ spanned by sequences of independent random variables. Israel J. Math. 8, 273–303.

Wedin, P.A., 1972. Perturbation bounds in connection with the singular value decomposition. BIT 12 (1), 99–111.

Zhang, Z., 1997. Parameter estimation techniques: a tutorial with application to conic fitting. Image Vision Comput. J. 15 (1), 59–76.