

Consistent least squares fitting of ellipsoids

I. Markovsky¹, A. Kukush², S. Van Huffel¹

¹ ESAT, SCD-SISTA, K.U.Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium;
e-mail: {ivan.markovsky,sabine.vanhuffel}@esat.kuleuven.ac.be

² National Taras Shevchenko University, Vladimirskaya st. 64, 01033, Kiev, Ukraine;
e-mail: alexander.kukush@univ.kiev.ua

Received August 16, 2002 / Revised version received August 15, 2003 /
Published online April 8, 2004 – © Springer-Verlag 2004

Summary. A parameter estimation problem for ellipsoid fitting in the presence of measurement errors is considered. The ordinary least squares estimator is inconsistent, and due to the nonlinearity of the model, the orthogonal regression estimator is inconsistent as well, *i.e.*, these estimators do not converge to the true value of the parameters, as the sample size tends to infinity. A consistent estimator is proposed, based on a proper correction of the ordinary least squares estimator. The correction is explicitly given in terms of the true value of the noise variance.

Mathematics Subject Classification (2000): 65D15, 65D10, 15A63

1 Introduction

The main motivation for our work is the ellipsoid fitting problem—given a set of data points $\{x^{(l)}\}_{l=1}^m$, $x^{(l)} \in \mathbb{R}^n$, find an ellipsoid

$$(1) \quad \mathcal{E}(A_e, c) := \{x \in \mathbb{R}^n : (x - c)^\top A_e (x - c) = 1\}, \quad A_e > 0,$$

that “best matches” them. The freedom in the choice of the matching criterion gives rise to different estimation methods.

One approach, called *algebraic fitting*, is to solve the following optimization problem:

$$(2) \quad \min_{A_e, c} \sum_{l=1}^m ((x^{(l)} - c)^\top A_e (x^{(l)} - c) - 1)^2,$$

and to define the estimate as any global optimal point. We will refer to (2) as the *ordinary least squares* (OLS) method for the ellipsoid model.

Another approach, called *geometric fitting*, is to solve the optimization problem

$$(3) \quad \min_{A_e, c} \sum_{l=1}^m \left(\text{dist}(x^{(l)}, \mathcal{E}(A_e, c)) \right)^2,$$

where $\text{dist}(x, \mathcal{E})$ is the Euclidean distance from the point x to the set \mathcal{E} . In the statistical literature, (3) is called the *orthogonal regression* method.

We assume that all data points are noisy measurements $x^{(l)} := \bar{x}^{(l)} + \tilde{x}^{(l)}$ of some *true points* $\bar{x}^{(1)}, \dots, \bar{x}^{(m)}$ that lie on a *true ellipsoid* $\mathcal{E}(\bar{A}_e, \bar{c})$, i.e., the model is a *measurement error model* [CV99, CRS95]. The measurement errors $\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)}$ are centered, independent identically distributed, and the distribution is normal with variance-covariance matrix $\bar{\sigma}^2 I$, where $\bar{\sigma}^2$ is the *noise variance*.

Due to the quadratic nature of the ellipsoid model with respect to the measurement x , both the algebraic and the geometric fitting methods are inconsistent in a statistical sense, see the classical paper of [NS48] and the discussion in [Ful87, p. 250]. We propose a *consistent estimator*, called *adjusted least squares* (ALS) estimator, that is derived from the OLS cost function by applying a correction.

The OLS estimator, defined by (2), is a nonlinear least-squares problem. We use a computationally cheap, but suboptimal method to solve the optimization problem (2). The quadratic equation defining the ellipsoid model is “embedded” in the *quadratic equation*

$$(4) \quad x^\top A x + b^\top x + d = 0, \quad A > 0,$$

which is linear in the parameters A , b , and d , so that a linear least squares estimation is possible. For given estimates \hat{A} , \hat{b} , and \hat{d} , of the parameters in (4), assuming that $\hat{A} > 0$, the estimates of the original parameters in (2) are given by:

$$(5) \quad \hat{c} := -\frac{1}{2} \hat{A}^{-1} \hat{b} \quad \text{and} \quad \hat{A}_e := \frac{1}{\hat{c}^\top \hat{A} \hat{c} - \hat{d}} \hat{A}.$$

The necessary computation for the (suboptimal) OLS estimator involves finding an eigenvector associated with the smallest eigenvalue of a symmetric matrix. We use the same indirect approach to compute the ALS estimator.

The correction needed for the ALS estimator is given explicitly in terms of the noise variance $\bar{\sigma}^2$. We give an algorithm for ellipsoid fitting that imple-

ments the theoretical results. Its computational cost increases linearly with the sample size m . In a separate paper [KMV02a], we present the statistical properties of the estimator and treat the case when $\bar{\sigma}^2$ is unknown.

The orthogonal regression estimator, on the other hand, is computed via a local optimization method and scales worse with m and with the dimension n of the vector space. In addition, due to the nonconvexity of the cost function in (3), the computed solution depends on the supplied initial approximation. In degenerate cases, see [Nie02, p. 260–261], the global minimum of (3) is not unique, so that there are several “best” fitting ellipses.

We point out several papers in which the ellipsoid fitting problem is considered. Gander *et al.* [GGS94] 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. Fitting methods, specific for ellipsoids, as opposed to the more general conic sections are first proposed in [FPF99]. The methods incorporate the ellipticity constraint into the normalizing condition and thus give better results when an elliptic fit is desired. In [Nie01] 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. Numerical methods for the orthogonal fitting problem are developed in [Spä97a].

A statistical point of view on the ellipsoid fitting problem is taken in [Kan94] and [CM96]. 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 from the one presented in [KMV02a]. Moreover, the noise variance estimate proposed in [Kan94] is still inconsistent; the bias is removed up to the first order approximation.

Standard notation used in the paper is: \mathbb{R} for the set of the real numbers, \mathbb{N} for the set of the natural numbers, $\mathbf{E}\tilde{x}$ for the expectation of the random variable \tilde{x} , $\mathbf{N}(0, V)$ for the zero mean normal distribution with variance-covariance matrix V , $\lambda_{\min}(A)$ ($\lambda_{\max}(A)$) for the minimum (maximum) eigenvalue of a symmetric matrix A , $\|x\|$ for the Euclidean norm of the vector x , and $\|A\|_F$ for the Frobenius norm of the matrix A . Throughout the paper \mathbb{S} denotes the space of the $n \times n$ symmetric matrices. Specific notation is introduced in the text.

The paper is structured as follows. Section 2 defines the quadratic measurement error model. The OLS and ALS estimators are defined in Sects. 3 and 4. Ellipsoid estimates are derived from the general quadratic model estimates in Sect. 5. An algorithms for ALS estimation is outlined in Sect. 6. Section 7 shows simulation examples and Sect. 8 gives conclusions.

2 Quadratic measurement error model

A *second order surface* in \mathbb{R}^n is the set

$$(6) \quad \mathcal{S}(A, b, d) := \{x \in \mathbb{R}^n : x^\top Ax + b^\top x + d = 0\},$$

where the symmetric matrix $A \in \mathbb{S}$, the vector $b \in \mathbb{R}^n$, and the scalar $d \in \mathbb{R}$ are *parameters* of the surface. If $A = 0$ and $b \neq 0$, then the surface (6) is a *hyperplane*, and if A is positive definite and $4d < b^\top A^{-1}b$, then (6) is an *ellipsoid*. Until Sect. 5, we will only assume that $\mathcal{S}(A, b, d)$ is a non empty set, but in Sect. 5, we will come back to the ellipsoid fitting problem, so that the parameters will be restricted.

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

$$(7) \quad \bar{x}^{(l)\top} \bar{A} \bar{x}^{(l)} + \bar{b}^\top \bar{x}^{(l)} + \bar{d} = 0, \quad \text{for } l = 1, \dots, m.$$

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

$$(8) \quad x^{(l)} = \bar{x}^{(l)} + \tilde{x}^{(l)}, \quad \text{for } l = 1, \dots, m,$$

where $\tilde{x}^{(1)}, \dots, \tilde{x}^{(m)}$ are the corresponding *measurement errors*. We assume that the measurement errors form an independent identically distributed sequence and the distribution of $\tilde{x}^{(l)}$, for all $l = 1, \dots, m$, is normal, zero mean, with variance-covariance matrix $\bar{\sigma}^2 I_n$, *i.e.*,

$$\mathbf{E} \tilde{x}^{(l_1)} \tilde{x}^{(l_2)\top} = 0, \quad \text{for } l_1, l_2 = 1, \dots, m, \quad l_1 \neq l_2,$$

and

$$\tilde{x}^{(l)} \sim \mathbf{N}(0, \bar{\sigma}^2 I_n), \quad \text{for } l = 1, \dots, m,$$

where $\bar{\sigma}^2 > 0$ is called the *noise variance*.

The matrix \bar{A} is the *true value* of the parameter A , while \bar{b} , and \bar{d} are the true values of the parameters b and d , respectively. Without additional constraint imposed on the parameters, for given second order surface $\mathcal{S}(A, b, d)$, the model parameters A, b , and d are not unique: $\mathcal{S}(\tau A, \tau b, \tau d)$ is the same surface for any real nonzero τ . 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 true values of the parameters are assumed to satisfy the constraint

$$(9) \quad \|\bar{A}\|_F^2 + \|\bar{b}\|^2 + \bar{d}^2 = 1.$$

Then the estimates are unique up to a sign.

Remark 1 (Invariance of the OLS and ALS estimators) As shown in [Boo79, p. 59, §4], [GGS94, p. 564, eq. (3.5)], and [Pra87, p. 147], the constraint (9) is not invariant under Euclidean transformations. As a result, the OLS estimator is not invariant under Euclidean transformations. Such a dependence on the coordinate system is undesirable. Suggestions for making the OLS estimator invariant can be found in [Nie01].

The following question arises. Are the ALS estimators derived with the constraint (9) invariant? If the noise variance is *fixed*, the answer is negative. However, if we are allowed to modify the noise variance after the transformation of the data, then the ALS estimator could be made invariant.

A modification of the noise variance that ensures invariance under Euclidean transformations is the noise variance estimation procedure derived in [KMV02a]. We demonstrate the invariance properties of the ALS estimator with estimated noise variance by a simulation example in Sect. 7. Rigorous proof of this property will be given elsewhere.

3 Ordinary least squares estimation

The OLS estimator for model (6) subject to the normalizing condition (9) is defined as a global minimum point of the following optimization problem:

$$(10) \quad \min_{A,b,d} \sum_{l=1}^m (x^{(l)\top} Ax^{(l)} + b^\top x^{(l)} + d)^2 \text{ s.t. } \begin{cases} A \text{ --- symmetric} \\ \|A\|_F^2 + \|b\|^2 + d^2 = 1. \end{cases}$$

The *OLS cost function* is

$$Q_{\text{ols}}(A, b, d) = \sum_{l=1}^m q_{\text{ols}}(A, b, d; x^{(l)}),$$

where the *elementary OLS cost function*,

$$q_{\text{ols}}(A, b, d; x) = (x^\top Ax + b^\top x + d)^2$$

measures the discrepancy of a single measurement point x from the surface $\mathcal{S}(A, b, d)$.

In order to derive the solution of (10), we introduce a parameter vector β containing all decision variables. Let $\text{vec}_s : \mathbb{S} \rightarrow \mathbb{R}^{(n+1)n/2}$ be an operator, a *symmetric matrix vectorizing operator*, that stacks the upper triangular part of A in a vector. The vector of decision variables is

$$(11) \quad \beta := [\text{vec}_s(A)^\top \ b^\top \ d]^\top.$$

It is an element of the *parameter space* \mathbb{R}^{n_β} , $n_\beta := (n + 1)n/2 + n + 1$.

Define the *symmetric Kronecker product* \otimes_s by

$$(12) \quad x^\top Ax = (x \otimes_s x)^\top \text{vec}_s(A), \quad \text{for all } x \in \mathbb{R}^n \text{ and } A \in \mathbb{S}.$$

We have for the elementary OLS cost function,

$$(13) \quad \begin{aligned} q_{\text{ols}}(\beta; x) &= (x^\top Ax + b^\top x + d)^2 \\ &= \left(\underbrace{\begin{bmatrix} (x \otimes_s x)^\top & x^\top & 1 \end{bmatrix}}_{y^\top} \begin{bmatrix} \text{vec}_s(A) \\ b \\ d \end{bmatrix} \right)^2 \\ &= (y^\top \beta)^2 = \beta^\top y y^\top \beta, \end{aligned}$$

and for the OLS cost function,

$$Q_{\text{ols}}(\beta) = \sum_{l=1}^m q_{\text{ols}}(\beta; x^{(l)}) = \sum_{l=1}^m ((y^{(l)\top} \beta)^2) = \|Y\beta\|^2 = \beta^\top Y^\top Y \beta,$$

where

$$y^{(l)} := \begin{bmatrix} x^{(l)} \otimes_s x^{(l)} \\ x^{(l)} \\ 1 \end{bmatrix}, \quad \text{for } l = 1, \dots, m, \quad \text{and} \quad Y := \begin{bmatrix} y^{(1)\top} \\ \vdots \\ y^{(m)\top} \end{bmatrix}.$$

Let $H \in \mathbb{R}^{n_\beta \times n_\beta}$ be a matrix, such that

$$(14) \quad \|H\beta\|^2 = \|A\|_F^2 + \|b\|^2 + d^2 \quad \text{for all } A \in \mathbb{S}, \quad b \in \mathbb{R}^n, \quad d \in \mathbb{R},$$

where β is defined in (11).

The OLS estimation problem (10) is equivalent to the following classical quadratically constrained least squares problem

$$(15) \quad \min_{\beta} \|Y\beta\|^2 \quad \text{s.t.} \quad \|H\beta\|^2 = 1.$$

The OLS estimator $\hat{\beta}_{\text{ols}}$ is $H^{-1}v_{\min}$, where v_{\min} is a normalized eigenvector of $H^{-T}Y^\top YH^{-1}$, corresponding to the smallest eigenvalue.

Remark 2 In order to avoid the computation of the Gram matrix $Y^\top Y$, one can obtain the solution from the singular value decomposition (SVD) of YH^{-1} . Let

$$(16) \quad \begin{aligned} YH^{-1} &= USV^\top, \quad \text{with } U^\top U = I, \quad V^\top V = I, \quad \text{and} \\ S &= \text{diag}(s_1, \dots, s_n), \quad s_1 \geq \dots \geq s_n \geq 0, \end{aligned}$$

then $\hat{\beta}_{\text{ols}}$ is $H^{-1}v_{\min}$, where v_{\min} is the last column of the matrix V .

Remark 3 The matrix H that ensures (14) is a diagonal matrix with diagonal elements equal to 1 or $\sqrt{2}$, where the latter correspond to the off-diagonal elements of A , see Remark 6. Since the normalizing condition (9) is arbitrary, however, we can choose any nonsingular matrix H in (15). Particularly simple is $H = I$. The OLS and ALS estimators depend on the normalizing condition but the ALS estimator is consistent for any non-degenerate normalizing condition, *i.e.*, for any full-rank matrix H .

Remark 4 Note that $\text{vec}_s(xx^\top) \neq x \otimes_s x$. One can verify that $x \otimes_s x = D \text{vec}_s(xx^\top)$, where D is a diagonal matrix with diagonal elements equal to 1 or 2; the latter corresponding to the off-diagonal elements of xx^\top appearing in the product $D \text{vec}_s(xx^\top)$, see Remark 6.

4 Adjusted least squares estimation

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 [KZ02], and it is related to the method of corrected score functions, see [CRS95, Sec. 6.5]. The model (7) is quadratic and a similar adjustment for a *bilinear model*, arising in motion analysis, is proposed in [KMV02b].

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

$$\min_{\beta} Q_{\text{als}}(\beta) \quad \text{s.t.} \quad \|H\beta\|^2 = 1,$$

where the *ALS cost function* Q_{als} is

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

Let $x = \bar{x} + \tilde{x}$, where \tilde{x} is normally distributed with zero mean and variance-covariance matrix $\bar{\sigma}^2 I$. The *elementary ALS cost function* q_{als} is defined by the following identity

$$(17) \quad \mathbf{E}q_{\text{als}}(\beta, \bar{x} + \tilde{x}) = q_{\text{ols}}(\beta, \bar{x}), \quad \text{for all } \beta \in \mathbb{R}^{n_{\beta}} \text{ and } \bar{x} \in \mathbb{R}^n.$$

We motivate the definition of the ALS cost function as follows.

$$\bar{Q}_{\text{ols}}(\beta) := \sum_{l=1}^m q_{\text{ols}}(\beta; \bar{x}^{(l)}), \quad \text{for all } \beta \in \mathbb{R}^{n_{\beta}},$$

has as a global minimum point the true value of the parameter vector

$$\bar{\beta} := [\text{vec}_s(\bar{A})^\top \quad \bar{b}^\top \quad \bar{d}]^\top.$$

Indeed, $\bar{Q}_{\text{ols}} \geq 0$ and by definition $\bar{Q}_{\text{ols}}(\bar{\beta}) = 0$. From

$$\mathbf{E}Q_{\text{als}} = \overline{Q}_{\text{ols}},$$

we see that, as the sample size grows, Q_{als} approximates $\overline{Q}_{\text{ols}}$. Provided that $\overline{Q}_{\text{ols}}$ has $\bar{\beta}$ as a unique global minimum (the *contrast condition* of [KMV02a]), the ALS estimator is strongly consistent.

Next, we derive an explicit expression for the ALS cost function Q_{als} . From (17) and (13), we have

$$\mathbf{E}q_{\text{als}}(\beta, x) = q_{\text{ols}}(\beta, \bar{x}) = \beta^\top \bar{y} \bar{y}^\top \beta =: \beta^\top \psi_{\text{ols}}(\bar{x}) \beta,$$

where

$$\bar{y} := [(\bar{x} \otimes_s \bar{x})^\top \bar{x}^\top \mathbf{1}]^\top \quad \text{and} \quad \psi_{\text{ols}}(\bar{x}) := \bar{y} \bar{y}^\top.$$

Thus the ALS elementary cost function q_{als} is quadratic in β ,

$$q_{\text{als}}(\beta; x) = \beta^\top \psi_{\text{als}}(x) \beta,$$

where

$$(18) \quad \mathbf{E}\psi_{\text{als}}(x) = \psi_{\text{ols}}(\bar{x}).$$

Under the normality assumption for the noise term \tilde{x} , (18) yields the following convolution equation:

$$\left(\frac{1}{2\pi\bar{\sigma}^2}\right)^{n/2} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi_{\text{als}}(\bar{x} + \tilde{x}) \prod_{i=1}^n \exp\left(-\frac{\tilde{x}_i^2}{2\bar{\sigma}^2}\right) d\tilde{x}_1 \cdots d\tilde{x}_n = \psi_{\text{ols}}(\bar{x}).$$

Solving for the unknown ψ_{als} is a *deconvolution problem*.

The deconvolution problem can be solved independently for the entries of ψ_{als} . The elements of the matrix $\psi_{\text{ols}}(\bar{x})$ are monomials of at most fourth order in \bar{x} . Consider the generic term

$$\eta_{\text{ols}}(\bar{x}) = \bar{x}_i \bar{x}_j \bar{x}_p \bar{x}_q, \quad \text{where } i, j, p, q \in \{0, 1, \dots, n\}$$

We formally set $\bar{x}_0 = 1$ and allow any of the indices to be zero, in order to allow η_{ols} to be of order less than four.

Let $r(s)$, $s = 1, \dots, n$, denotes the number of repetitions of the index s in the monomial $\bar{x}_i \bar{x}_j \bar{x}_p \bar{x}_q$. For example, let $n = 2$. In the monomial $\bar{x}_1 \bar{x}_2^3$, $r(1) = 1$ and $r(2) = 3$, and in the monomial \bar{x}_1^4 , $r(1) = 4$ and $r(2) = 0$.

The functions

$$(19) \quad \begin{aligned} t_0(\xi) &:= 1, & t_1(\xi) &:= \xi, & t_2(\xi) &:= \xi^2 - \bar{\sigma}^2, \\ t_3(\xi) &:= \xi^3 - 3\xi\bar{\sigma}^2, & \text{and} & & t_4(\xi) &:= \xi^4 - 6\xi^2\bar{\sigma}^2 + 3\bar{\sigma}^4. \end{aligned}$$

have the property

$$\mathbf{E}t_k(\bar{x}_s + \tilde{x}_s) = (\bar{x}_s)^k, \quad \text{for all } \bar{x}_s \in \mathbb{R} \text{ and } k = 0, 1, \dots, 4,$$

where $\tilde{x}_s \sim N(0, \bar{\sigma}^2)$. Thus the polynomial

$$(20) \quad \eta_{\text{als}}(x) := \prod_{s=1}^n t_{r(s)}(x_s),$$

has the property

$$\mathbf{E}\eta_{\text{als}}(x) = \bar{x}_i \bar{x}_j \bar{x}_p \bar{x}_q = \eta_{\text{ols}}(\bar{x}), \quad \text{for all } \bar{x} \in \mathbb{R}^n.$$

This shows that η_{als} is the desired solution. The matrix ψ_{als} is constructed element-wise in the described way.

The ALS cost function Q_{als} is quadratic in β ,

$$Q_{\text{als}}(\beta) = \beta^\top \Psi_{\text{als}} \beta, \quad \text{for all } \beta \in \mathbb{R}^{n\beta},$$

where

$$\Psi_{\text{als}} = \sum_{l=1}^m \psi_{\text{als}}(x^{(l)}).$$

Thus the function Q_{als} is described thoroughly.

Example 1 (The matrix ψ_{als} for the case $n = 2$) The model parameters are $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, and the scalar d . The parameter space is 6-dimensional with

$$\beta := [\text{vec}_s(A)^\top \quad b^\top \quad d]^\top = [a_{11} \quad a_{12} \quad a_{22} \quad b_1 \quad b_2 \quad d]^\top.$$

From (12), we have

$$[x_1 \quad x_2]^\top \otimes_s [x_1 \quad x_2]^\top = [x_1 x_1 \quad 2x_1 x_2 \quad x_2 x_2]^\top.$$

so that

$$y := [(x \otimes_s x)^\top \quad x^\top \quad 1]^\top = [x_1 x_1 \quad 2x_1 x_2 \quad x_2 x_2 \quad x_1 \quad x_2 \quad 1]^\top,$$

and

$$\psi_{\text{ols}}(x) = y y^\top = \begin{bmatrix} x_1^4 & 2x_1^3 x_2 & x_1^2 x_2^2 & x_1^3 & x_1^2 x_2 & x_1^2 \\ * & 4x_1^2 x_2^2 & 2x_1 x_2^3 & 2x_1^2 x_2 & 2x_1 x_2^2 & 2x_1 x_2 \\ * & * & x_2^4 & x_1 x_2^2 & x_2^3 & x_2^2 \\ * & * & * & x_1^2 & x_1 x_2 & x_1 \\ * & * & * & * & x_2^2 & x_2 \\ * & * & * & * & * & 1 \end{bmatrix},$$

with *'s indicating the symmetric elements.

The adjusted matrix ψ_{als} is $\psi_{\text{als}} = \psi_{\text{ols}} + \Delta\psi_{\text{als}}$, where the correction $\Delta\psi_{\text{als}}$ is

$$\begin{bmatrix} 3\bar{\sigma}^4 - 6\bar{\sigma}^2 x_1^2 & -6\bar{\sigma}^2 x_1 x_2 & \Delta\psi_{\text{als},13} & -3\bar{\sigma}^2 x_1 & -\bar{\sigma}^2 x_2 & -\bar{\sigma}^2 \\ * & \Delta\psi_{\text{als},22} & -6\bar{\sigma}^2 x_1 x_2 & -2\bar{\sigma}^2 x_2 & -2\bar{\sigma}^2 x_1 & 0 \\ * & * & 3\bar{\sigma}^4 - 6\bar{\sigma}^2 x_2^2 & -\bar{\sigma}^2 x_1 & -3\bar{\sigma}^2 x_2 & -\bar{\sigma}^2 \\ * & * & * & -\bar{\sigma}^2 & 0 & 0 \\ * & * & * & * & -\bar{\sigma}^2 & 0 \\ * & * & * & * & * & 0 \end{bmatrix},$$

and

$$\Delta\psi_{\text{als},13} = \bar{\sigma}^4 - \bar{\sigma}^2(x_1^2 + x_2^2), \quad \text{and} \quad \Delta\psi_{\text{als},22} = 4\bar{\sigma}^4 - 4\bar{\sigma}^2(x_1^2 + x_2^2).$$

The correction matrix $\Delta\psi_{\text{als}}$, without the fourth order terms in $\bar{\sigma}$, is derived in [Zha97, Sec. 7]. The derivation in [Zha97], however, is specialized for the two-dimensional case.

Remark 5 The recommended way of computing the OLS estimator is via the SVD of YH^{-1} . For the ALS estimator, however, we use the less accurate eigenvalue decomposition because the correction is derived for $\Psi_{\text{ols}} = Y^T Y$ and can not be determined for the factor Y .

5 Ellipsoid estimation

The ALS estimator β_{als} is derived for the general quadratic measurement error model (7–8). Now, we specialize it for the ellipsoid fitting problem, *i.e.*, we assume that the true surface belongs to the class of surfaces

$$(21) \quad \mathcal{C}(A_e, c) = \{x \in \mathbb{R}^n : (x - c)^\top A_e (x - c) = 1\}$$

for some true values $\bar{A}_e \in \mathbb{S}$, $\bar{A}_e > 0$ and \bar{c} of the parameters A_e and c . The equation defining $\mathcal{C}(\bar{A}_e, \bar{c})$ can be written as

$$x^\top \bar{A}_e x - 2(\bar{A}_e \bar{c})^\top x + \bar{c}^\top \bar{A}_e \bar{c} - 1 = 0,$$

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

$$x^\top (\bar{A}_e/\theta) x - 2(\bar{A}_e \bar{c}/\theta)^\top x + (\bar{c}^\top \bar{A}_e \bar{c} - 1)/\theta = 0.$$

Introduce the new parameters

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

As defined, \bar{A} , \bar{b} , and \bar{d} satisfy the normalizing condition (9).

We can go back to the original parameters \bar{A}_e and \bar{c} from \bar{A} , \bar{b} , and \bar{d} , that satisfy (9) by

$$(22) \quad \bar{c} = -\frac{1}{2}\bar{A}^{-1}\bar{b}, \quad \text{and} \quad \bar{A}_e = \frac{1}{\bar{c}^\top \bar{A} \bar{c} - \bar{d}} \bar{A}.$$

Note that $\theta = \bar{c}^\top \bar{A} \bar{c} - \bar{d}$ is nonzero. Let $\hat{A}, \hat{b}, \hat{d}$ be the ALS estimator of the parameters $\bar{A}, \bar{b}, \bar{d}$. The estimator of the parameters \bar{A}_e and \bar{c} is given by the transformation (5).

If the obtained estimate \hat{A}_e is indefinite, we impose a posteriori positive definiteness by the projection

$$(23) \quad \hat{A}_{e,2} := \sum_{i:\hat{\lambda}_i > 0} \hat{\lambda}_i \hat{v}_i \hat{v}_i^\top,$$

where $\hat{A}_e = \sum_{i=1}^n \hat{\lambda}_i \hat{v}_i \hat{v}_i^\top$ is the eigenvalue decomposition of \hat{A}_e . Indefinite estimate \hat{A}_e could be obtained because the estimator does not enforce the prior knowledge $\hat{A}_e > 0$. Clearly the two stage procedure— \hat{A}_e obtained on the first stage and $\hat{A}_{e,2}$ on the second stage—is suboptimal. Empirical results, however, suggest that the event of having the constraint $\hat{A}_e > 0$ active is rather rare. Typically it occurs for a small sample size with non-uniform data point distribution and for data with outliers. Due to $\bar{A}_e > 0$ and the consistency of the estimator \hat{A}_e , we expect that for large sample size $\hat{A}_e > 0$.

6 Algorithms for adjusted least squares estimation

In this section, we summarize the estimation procedure described above by giving an algorithm for its computation. MATLAB-like notation for indexing the elements of a matrix is used. For example, $A(i_1:i_2, j_1:j_2)$ stands for the submatrix of A obtained by selecting the elements with first index in the set $\{i_1, i_1 + 1, \dots, i_2\}$ and with second index in the set $\{j_1, j_1 + 1, \dots, j_2\}$.

Algorithm ALS (Adjusted least squares ellipsoid fitting)

Input: a matrix $X := [x^{(1)} \dots x^{(m)}] \in \mathbb{R}^{n \times m}$ and the noise variance $\bar{\sigma}^2$.

Output: the estimates \hat{A}_e, \hat{c} of the ellipsoid parameters.

- 1) Form the tensor $T \in \mathbb{R}^{5 \times n \times m}$

$$T(k, i, l) := t_k(X(i, l)), \quad \text{for } k = 0, \dots, 4, \\ i = 1, \dots, n, \text{ and } l = 1, \dots, m,$$

where the functions $t_k, k = 0, 1, 2, 3, 4$, are given in (19).

2) Define the vectors $\mathbf{1}, \mathbf{i} \in \mathbb{R}^{n+1}$ by

$$\mathbf{1} := [1 \cdots 1 \ 1]^\top, \quad \mathbf{i} := [1 \cdots n \ 0]^\top,$$

and form the matrix $M \in \mathbb{R}^{n_\beta \times 2}$, $n_\beta := (n+1)n/2 + n + 1$,

$$M := [\text{vec}_s(\mathbf{i}\mathbf{1}^\top) \ \text{vec}_s(\mathbf{1}\mathbf{i}^\top)].$$

We use M to find the indices of \bar{x} in the entries of $\psi_{\text{ols}}(\bar{x})$. Note that $(M(p, 1), M(p, 2))$ are the indices of \bar{x} in the p -th entry of $\bar{y} := \bar{x} \otimes_s \bar{x}$. Recall that $\psi_{\text{ols}}(\bar{x}) := \bar{y}\bar{y}^\top$. Thus the indices of \bar{x} in the (p, q) -th entry of $\psi_{\text{ols}}(\bar{x})$ are $(M(p, 1), M(p, 2), M(q, 1), M(q, 2))$.

3) Define a binary operator $==$ by

$$(i_1 == i_2) := \begin{cases} 1, & \text{if } i_1 = i_2 \\ 0, & \text{otherwise} \end{cases}, \quad \text{for all } i_1, i_2 \in \mathbb{R}.$$

Form the tensor $R \in \mathbb{R}^{n_\beta \times n_\beta \times n}$,

$$R(p, q, i) = (M(p, 1) == i) + (M(p, 2) == i) \\ + (M(q, 1) == i) + (M(q, 2) == i),$$

for all $q \geq p$ and $i = 1, \dots, n$.

Note that $R(p, q, i)$ is the number of repetitions of the index i in the (p, q) -th entry of $\psi_{\text{ols}}(\bar{x})$. Thus in terms of the function r , defined in Sect. 4, the tensor R stores $(r(1), \dots, r(n))$ for the entries of $\psi_{\text{ols}}(\bar{x})$.

4) Compute

$$\eta_{\text{als}}(p, q) = \sum_{l=1}^m \prod_{i=1}^n T(R(p, q, i), i, l), \quad \text{for all } q \geq p.$$

This step corresponds to the correction (20) from Sect. 4.

In the construction of Ψ_{als} , however, we have to take into account the presence of factors 2 and 4 in some of the entries of $\psi_{\text{ols}}(\bar{x})$, see Example 1. Up to now we have computed the corrections for monomials of the type $\bar{x}_i \bar{x}_j \bar{x}_p \bar{x}_q$, but there are constants 2 and 4 coming from the symmetric Kronecker product $\bar{x} \otimes_s \bar{x}$, see Remark 4.

5) Form the set \mathcal{D} of the indices of the vector $\text{vec}_s(A)$, corresponding to the off-diagonal elements of A ,

$$\mathcal{D} = \{1, \dots, (n+1)n/2\} - \{i(i+1)/2 : i = 1, \dots, n\}.$$

Here $\mathcal{D}_1 - \mathcal{D}_2$ denotes the set difference of the sets \mathcal{D}_1 and \mathcal{D}_2 . Note that $\{i(i+1)/2 : i = 1, \dots, n\}$ are the indices of $\text{vec}_s(A)$, corresponding to the diagonal elements of A .

6) Form the symmetric matrix Ψ_{als} by

$$\Psi_{\text{als}}(p, q) := \begin{cases} 4\eta_{\text{als}}(p, q) & , \text{ if } p \in \mathcal{D} \text{ and } q \in \mathcal{D}, \\ 1\eta_{\text{als}}(p, q) & , \text{ if } p \notin \mathcal{D} \text{ and } q \notin \mathcal{D}, \\ 2\eta_{\text{als}}(p, q) & , \text{ otherwise,} \end{cases}$$

for all $q \geq p$, and

$$\Psi_{\text{als}}(p, q) := \Psi_{\text{als}}(q, p), \quad \text{for all } q < p.$$

7) Find an eigenvector $\hat{\beta}_{\text{als}}$ associated with the smallest eigenvalue of Ψ_{als} .

8) Normalize $\hat{\beta}_{\text{als}}$,

$$\hat{\beta}_{\text{als}} := \hat{\beta}_{\text{als}} / \|\hat{\beta}_{\text{als}}\|.$$

9) Reconstruct the estimates \hat{A} , \hat{b} , and \hat{d} from the vector $\hat{\beta}_{\text{als}}$,

$$\begin{aligned} \hat{A} &:= \text{vec}_s^{-1}(\hat{\beta}_{\text{als}}(1 : n(n+1)/2)), \\ \hat{b} &:= \hat{\beta}_{\text{als}}(n(n+1)/2 + 1 : n_\beta - 1), \\ \hat{d} &:= \hat{\beta}_{\text{als}}(n_\beta), \end{aligned}$$

where $\text{vec}_s^{-1} : \mathbb{R}^{n(n+1)/2} \rightarrow \mathbb{S}$, forms a symmetric matrix out of the vector of the elements in its upper triangular part.

10) The estimates of the ellipsoid parameters \bar{A}_e and \bar{c} are obtained by (5).

11) If $\hat{A}_e \leq 0$, project \hat{A} on the positive definite cone by (23).

Remark 6 If a general quadratic model is estimated, the normalizing condition is given as prior knowledge, see Remark 3. If an ellipsoid is estimated, however, the normalizing condition is arbitrary. In Algorithm ALS, we set $H = I$, which corresponds to a normalizing condition

$$(\text{vec}_s(A))^\top \text{vec}_s(A) + \|b\|^2 + d^2 = 1.$$

The matrix H corresponding to the normalizing condition (9) is

$$H = \begin{bmatrix} \sqrt{D} & & \\ & I_n & \\ & & 1 \end{bmatrix},$$

where D is a diagonal matrix with diagonal elements

$$D_{ii} = \begin{cases} 2 & , \text{ if } i \in \mathcal{D}, \\ 1 & , \text{ otherwise.} \end{cases}$$

Remark 7 (Known blocks of the matrix Ψ_{als}) Algorithm ALS can be improved by setting certain elements of Ψ_{als} in advance and not via following the general adjustment procedure. Consider a block partitioning of the matrices ψ_{ols} , ψ_{als} , and Ψ_{als} according to the partitioning of the vector

$$\left[(x \otimes_s x)^\top \mid x^\top \mid 1 \right]^\top,$$

e.g., for ψ_{ols} , denote

$$\psi_{\text{ols}} =: \begin{bmatrix} \psi_{\text{ols},11} & \psi_{\text{ols},12} & \psi_{\text{ols},13} \\ * & \psi_{\text{ols},22} & \psi_{\text{ols},23} \\ * & * & \psi_{\text{ols},33} \end{bmatrix}.$$

All elements of ψ_{ols} are monomials in x ; moreover all elements of:

- $\psi_{\text{ols},11}(x)$ are of fourth order,
- $\psi_{\text{ols},12}(x)$ are of third order,
- $\psi_{\text{ols},13}(x)$ and $\psi_{\text{ols},22}(x)$ are of second order,
- $\psi_{\text{ols},23}(x)$ are of first order, and
- the scalar $\psi_{\text{ols},33}(x) = 1$ is independent of x .

For the blocks of order zero and one, there is no correction applied in the formation of the matrix ψ_{als} . The correction for the elements of the blocks of order two is $-\bar{\sigma}^2 I_n$. Thus for the corresponding blocks of ψ_{als} , we have

$$\begin{aligned} \psi_{\text{als},22}(x) &= xx^\top - \bar{\sigma}^2 I_n, & \psi_{\text{als},23}(x) &= x, \\ \psi_{\text{als},13}(x) &= x \otimes_s x - \text{vec}_s(\bar{\sigma}^2 I_n), & \psi_{\text{als},33}(x) &= 1. \end{aligned}$$

Finally, the corresponding blocks of Ψ_{als} are

$$\begin{aligned} \Psi_{\text{als},22} &= \sum_{l=1}^m x^{(l)} x^{(l)\top} - m\bar{\sigma}^2 I_n, & \Psi_{\text{als},23} &= \sum_{l=1}^m x^{(l)}, \\ \Psi_{\text{als},13} &= \sum_{l=1}^m x^{(l)} \otimes_s x^{(l)} - \text{vec}_s(m\bar{\sigma}^2 I_n), & \Psi_{\text{als},33} &= m, \end{aligned}$$

and only the upper triangular part of the block $\Psi_{\text{als},11}$ and the block $\Psi_{\text{als},12}$ need be computed in steps 4) and 6) of Algorithm ALS.

7 Simulation examples

We show the ALS, OLS, and orthogonal regression (OR) estimates for a test example from the literature [GG94], called “special data”. It is designed to illustrate the inadequacy of the algebraic fitting method and to show the advantage of the OR method.

Only data points are given; even if they are generated with a true model, we do not know it. For this reason the comparison is visual. Since the noise variance needed for the ALS estimator is unknown, we estimate it via the procedure proposed in [KMV02a].

Figure 1, shows the data points with the estimated ellipses superimposed on them. The OR estimator is computed by a general purpose optimization algorithm (MATLAB's `fmincon` function). The cost function is evaluated as explained in [Zha97, Sec. 5.2].

For the first test example, see Fig. 1, left, the OR estimator is influenced by the initial approximation. Using the OLS estimate as initial approximation, the optimization algorithm 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 obtained estimate is the dashed-dotted ellipse closer to the ALS estimate. Next we will consider the better of the two OR estimates.

Although the sample size is only $m = 8$ data points, the ALS estimator gives good estimates that are comparable with the OR estimate. The value of the OR cost function (see (3)) is 3.2531 for the OLS estimator, 1.6284 for the ALS estimator, and 1.3733 for the OR estimate. The ALS estimator is less than 19% suboptimal. Moreover, the volume of the OR estimate is 62.09 square units, while the volume of the ALS estimate is 34.37 square units, which is nearly twice as small. Visually (as well as in other senses) "smaller" estimates are preferable.

In a second example, taken from [Spä97b], the ALS estimate is close to the OR estimate, see Fig. 1, right. In terms of the OR cost function, the ALS estimate is less than 25% suboptimal. The volume of the ALS estimate is comparable with that of the OR estimate.

Figure 2 illustrates the invariance properties of the ALS estimator with estimated noise variance. The data used is again the "special data" from

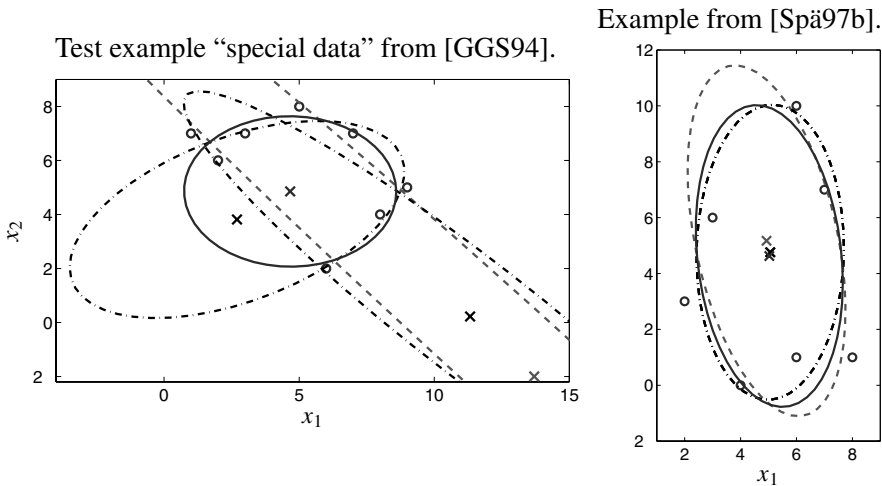


Fig. 1. Test examples. dashed—OLS, dashed-dotted—OR, solid—ALS, ○—data points, ×—centers of the estimated ellipses

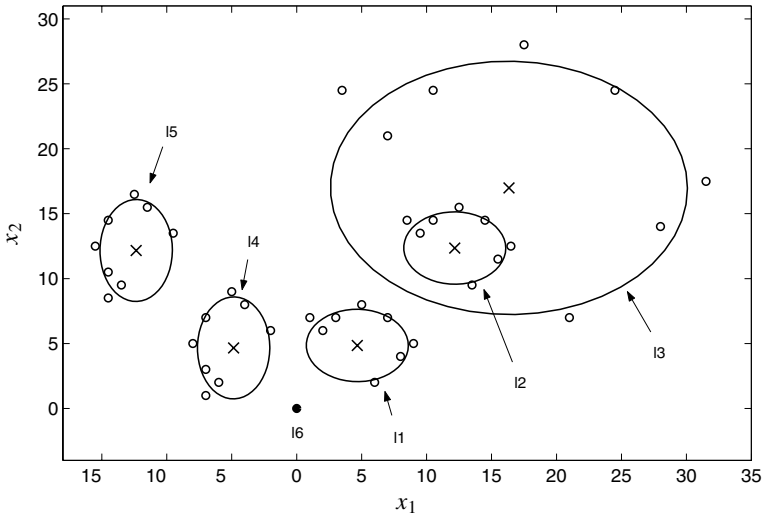


Fig. 2. ALS estimates of the original, translated, rotated, scaled, and translated and rotated data points. \circ —data points, \times —centers of the estimated ellipses, \bullet —point (0, 0)

[GGS94]. The figure shows translated, rotated, scaled, and translated and rotated data points with the corresponding ALS estimates.

8 Conclusions

The OLS estimation of the ellipsoid parameters from noisy measurements of points on its boundary is a nonlinear least squares problem. An indirect, suboptimal approach was used that transforms the ellipsoid model to a general quadratic model and applies linear least squares estimation. Due to the measurement errors, however, the ordinary least squares estimator is inconsistent.

Assuming that the measurement errors are normally distributed, a correction is derived that uses the true measurement error variance and adjusts the OLS cost function, so that the resulting ALS estimator is consistent. Algorithm for the necessary computation is outlined.

The ALS estimator is illustrated via simulation examples. Compared with the orthogonal regression estimator, it has the advantage of being cheaper to compute and independent of initial approximation. The computational efficiency is crucial for higher dimensional ellipsoid fitting and for problems with large sample size.

Acknowledgements. 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. 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. This paper presents research results of the Belgian Programme on Interuniversity Poles of Attraction (IUP 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 project (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.

References

- [Boo79] Bookstein, F.L.: Fitting conic sections to scattered data. *Comput. Graphics and Image Processing* **9**, 59–71 (1979)
- [CM96] Cabrera, J., Meer, P.: Unbiased estimation of ellipses by bootstrapping. *IEEE Trans. Pattern Anal. Machine Intell.* **18**(7), 752–756 (1996)
- [CRS95] Carroll, R.J., Ruppert, D., Stefanski, L.A.: *Measurement Error in Nonlinear Models*. Number 63 in *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, 1995
- [CV99] Cheng, C., Van Ness, J.W.: *Statistical regression with measurement error*. London: Arnold, 1999
- [FPF99] Fitzgibbon, A., Pilu, M., Fisher, R.: Direct least-squares fitting of ellipses. *IEEE Trans. Pattern Anal. Machine Intell.* **21**(5), 476–480 (1999)
- [Ful87] Fuller, W.A.: *Measurement error models*. New York: Wiley, 1987
- [GGS94] Gander, W., Golub, G., Strebel, R.: Fitting of circles and ellipses: Least squares solution. *BIT* **34**, 558–578 (1994)
- [Kan94] Kanatani, K.: Statistical bias of conic fitting and renormalization. *IEEE Trans. Pattern Anal. Machine Intell.* **16**(3), 320–326 (1994)
- [KMV02a] Kukush, A., Markovsky, I., Van Huffel, S.: Consistent estimation in an implicit quadratic measurement error model. To appear in *Comp. Stat. Data Anal.* 2004
- [KMV02b] Kukush, A., Markovsky, I., Van Huffel, S.: Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. *Comp. Stat. Data Anal.* **41**(1), 3–18 (2002)
- [KZ02] Kukush, A., Zwanzig, S.: On consistent estimators in nonlinear functional EIV models. In: S. Van Huffel and P. Lemmerling, (eds.), *Total least squares and errors-in-variables modeling: Analysis, Algorithms and Applications*, Kluwer, 2002, pp. 145–155
- [Nie01] Nievergelt, Y.: Hyperspheres and hyperplanes fitting seamlessly by algebraic constrained total least-squares. *Linear Algebra and Its Appl.* **331**, 43–59 (2001)
- [Nie02] Nievergelt, Y.: A finite algorithm to fit geometrically all midrange lines, circles, planes, spheres, hyperplanes, and pyperspheres. *Numer. Math.* **91**, 257–303 (2002)
- [NS48] Neyman, J., Scott, E.L.: Consistent estimates based on partially consistent observations. *Econometrica* **16**(1), 1–32 (1948)
- [Pra87] Pratt, V.: Direct least-squares fitting of algebraic surfaces. *ACM Comput. Graphics* **21**(4), 145–152 (1987)
- [Spä97a] Späth, H.: Least-squares fitting of ellipses and hyperbolas. *Comput Stat.* **12**, 329–341 (1997)

- [Spä97b] Späth, H.: Orthogonal least squares fitting by conic sections. In: S. Van Huffel, (ed.), Recent advances in Total least squares techniques and errors-in-variables modeling, SIAM, 1997
- [Zha97] Zhang, Z.: Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Comput. J.* **15**(1), 59–76 (1997)