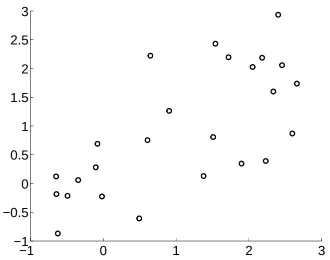
Estimation of an Ellipsoid from Observation of Points on its Boundary

Alexander Kukush, Ivan Markovsky, and Sabine Van Huffel
Department of Electrical Engineering, Katholieke Universiteit Leuven

Given some points in the plane,



what is the ellipsoid that best matches them?

Quadratic measurement error model

a second order surface in $\mathbb{R}^n$ is the set

$$S(A,b,d) := \{\, x \in \mathbb{R}^{n\times 1} \;:\; x^T A x + b^T x + d = 0 \,\}$$

where $A \in \mathbb{S}$, $b \in \mathbb{R}^{n\times 1}$, and $d \in \mathbb{R}$ are parameters of the surface
($\mathbb{S}$ is the set of $n \times n$ symmetric matrices)

special cases:

- if $A = 0$ and $b \neq 0$, then $S(0,b,d)$ is a hyperplane
- if $A = A_{\mathsf{e}} > 0$ and $4d < b^T A^{-1} b$, then $S(A_{\mathsf{e}}, b, d)$ is an elliptic surface

$$E(A_{\mathsf{e}}, c) := \{\, x \in \mathbb{R}^{n\times 1} \;:\; (x-c)^T A_{\mathsf{e}} (x-c) = 1 \,\}$$

$c \in \mathbb{R}^{n\times 1}$ is the center of the elliptic surface

let $\bar{x}^l$, for $l = 1,\dots,m$, lie on the surface $S(\bar{A}, \bar{b}, \bar{d})$, i.e.,

$$(\bar{x}^l)^T \bar{A} \bar{x}^l + \bar{b}^T \bar{x}^l + \bar{d} = 0, \qquad \text{for } l = 1,\dots,m$$

let $x^l \in \mathbb{R}^{n\times 1}$, $l = 1,\dots,m$ be observations of the points $\bar{x}^l$, $l = 1,\dots,m$, i.e.,

$$x^l = \bar{x}^l + \tilde{x}^l, \qquad \text{for } l = 1,\dots,m$$

$\tilde{x}^l$, $l = 1,\dots,m$ are measurement errors

assumption: $\{\tilde{x}^l\}_{l=1}^m$ form an i.i.d. sequence and $\tilde{x}^l \sim \mathrm{N}(0, \bar{\sigma}^2 I_n)$

$\bar{A} \in \mathbb{S}$, $\bar{b} \in \mathbb{R}^{n\times 1}$, and $\bar{d} \in \mathbb{R}$ are the true values of the parameters

normalizing condition: $||\bar{A}||_F^2 + ||\bar{b}||^2 + \bar{d}^2 = 1$

estimation problem: given $\{x^l\}_{l=1}^m$ and $\bar{\sigma}$, estimate $\bar{A}$, $\bar{b}$, and $\bar{d}$

Least squares estimation

the model is linear in the parameters $\Rightarrow$ ordinary least squares (OLS)

$q_{\mathsf{ols}}(A,b,d;x) := \left(x^T A x + b^T x + d - 1\right)^2$ — elementary OLS cost function

measures the discrepancy of the single measurement x from the model $S(A,b,d)$

$Q_{\mathsf{ols}}(A,b,d) := \sum_{l=1}^m q_{\mathsf{ols}}(A,b,d;x^l)$ — OLS cost function

$$\min_{A,b,d} Q_{\mathsf{ols}}(A,b,d) \quad \text{s.t.} \quad ||\bar{A}||_F^2 + ||\bar{b}||^2 + \bar{d}^2 = 1 \qquad \text{(OLS)}$$

quadratically constraint least squares problem

let $\mathsf{vec}_{\mathsf{s}}(A)$ be the vector of the elements in the lower triangular part of A

define $x \otimes_{\mathsf{s}} x$ by

$$x^T A x = (x \otimes x)^T \mathsf{vec}(A) =: (x \otimes_{\mathsf{s}} x)^T \mathsf{vec}_{\mathsf{s}}(A)$$

denote $\beta := [\mathsf{vec}_{\mathsf{s}}(A)^T \; b^T \; 1]^T$, we have

$$Q_{\mathsf{ols}}(\beta) = \sum_{l=1}^m \Bigg(\underbrace{\left[(x^l \otimes_{\mathsf{s}} x^l)^T \; (x^l)^T \; 1\right]}_{f^l} \begin{bmatrix} \mathsf{vec}_{\mathsf{s}}(A) \\ b \\ 1 \end{bmatrix} \Bigg)^2 = \Bigg|\Bigg| \underbrace{\begin{bmatrix} f^1 \\ \vdots \\ f^m \end{bmatrix}}_{F} \beta \Bigg|\Bigg|^2$$

the (OLS) problem becomes

$$\min_{\beta} ||F\beta||^2 \quad \text{s.t.} \quad ||\beta||^2 = 1$$

and the solution is obtained from the SVD of F (or from the EVD of $F^T F$)

take the right singular vector corresponding to the smallest singular value

the OLS estimator $\hat{\beta}_{\mathsf{ols}}$ is inconsistent

Adjusted least squares estimation

define an auxiliary function $q_{\mathsf{als}}(\beta; x)$, such that

$\mathbf{E}\, q_{\mathsf{als}}(A,b,d;\bar{x}+\tilde{x}) = q_{\mathsf{ols}}(A,b,d;\bar{x})$ for all $A \in \mathbb{S}$, $b \in \mathbb{R}^{n\times 1}$, $d \in \mathbb{R}$, $\bar{x} \in \mathbb{R}^{n\times 1}$

the ALS estimator is a normalized eigenvector of $\Psi_{\mathsf{als}} := \sum_{i=1}^m q'_{\mathsf{als}}(x_i)$ corresponding to the minimum eigenvalue

we have

$$\Psi_{\mathsf{als}} = \Psi_{\mathsf{ols}} + \Delta\Psi_{\mathsf{als}}, \qquad \Psi_{\mathsf{ols}} := \sum_{i=1}^m q'_{\mathsf{ols}}(x_i)$$

for an appropriate correction $\Delta\Psi_{\mathsf{als}}$

we show the necessary correction for the two-dimensional case ($n = 2$)

in the space $\mathbb{R}^{(n+1)n/2+n+1}$, the operator Ψ_{ols} is represented by the matrix $F^T F$, i.e., a sum of matrices of the type (symmetric elements are shown with $*$)

$$\begin{bmatrix}
x_1^4 & 2x_1^3x_2 & x_1^2x_2^2 & x_1^3 & x_1^2x_2 & x_1^2 \\
* & 4x_1^2x_2^2 & 2x_1x_2^3 & 2x_1^2x_2 & 2x_1x_2^2 & 2x_1x_2 \\
* & * & x_2^4 & x_1x_2^2 & x_2^3 & x_2^2 \\
* & * & * & x_1^2 & x_1x_2 & x_1 \\
* & * & * & * & x_2^2 & x_2 \\
* & * & * & * & * & 1
\end{bmatrix}$$

the necessary correction matrix for the one above is

$$\begin{bmatrix}
3\bar{\sigma}^4 - 6\bar{\sigma}^2x_1^2 & -6\bar{\sigma}^2x_1x_2 & \bar{\sigma}^4 - \bar{\sigma}^2(x_1^2+x_2^2) & -3\bar{\sigma}^2x_1 & -\bar{\sigma}^2x_2 & -\bar{\sigma}^2 \\
* & 4\bar{\sigma}^4 - 4\bar{\sigma}^2(x_1^2+x_2^2) & -6\bar{\sigma}^2x_1x_2 & -2\bar{\sigma}^2x_2 & -2\bar{\sigma}^2x_1 & 0 \\
* & * & 3\bar{\sigma}^4 - 6\bar{\sigma}^2x_2^2 & -\bar{\sigma}^2x_1 & -3\bar{\sigma}^2x_2 & -\bar{\sigma}^2 \\
* & * & * & -\bar{\sigma}^2 & 0 & 0 \\
* & * & * & * & -\bar{\sigma}^2 & 0 \\
* & * & * & * & * & 0
\end{bmatrix}$$

because of the linearity the total correction is the sum of the corrections for all x^l

thus the operator Ψ_{als} is represented by $F^T F + \sum_{l=1}^m$ (correction matrix for x^l)

Orthogonal regression

define the distance from a point $x \in \mathbb{R}^n$ to a set $Y \subset \mathbb{R}^n$ by

$$\text{dist}(x, Y) := \min_{y \in Y} ||x - y||$$

the elementary orthogonal regression cost function is the distance from the data point x to the second order surface $S(A, b, d)$

$$q_{\text{orth}}(A, b, d; x) := \text{dist}\big(x, S(A, b, d)\big)$$

the orthogonal regression cost function is the sum of these distances for all data points

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

the orthogonal regression estimator is defined as a global minimum point of Q_{orth} subject to the normalizing condition

$$\min_{A,b,d} Q_{\text{orth}}(A, b, d) \quad \text{s.t.} \quad ||A||_F^2 + ||b||^2 + d^2 = 1$$

in general this is a non-convex optimization problem

- expensive to solve
- no guarantee that global minimum is found

the orthogonal regression cost function has geometric meaning:
sum of distances from the data points to the estimated surface

Q_{orth} is intuitively appealing cost function but defines an inconsistent estimator

advantage: gives good results for small sample size (only a few data points)

for small sample size ALS estimator is unstable

Ellipsoid estimation with ALS

suppose that the true surface belongs to the class of surfaces

$$E(A_e, c) = \{\, x \in \mathbb{R}^{n \times 1} : (x - c)^T A_e (x - c) = 1 \,\}, \quad \text{with } A_e > 0$$

the defining equation is quadratic in x but nonlinear in the parameters

$$x^T A_e x - (A_e c)^T x + c^T A_e c - 1 = 0$$

we normalize by dividing with $\lambda := \sqrt{||A_e||_F^2 + ||A_e c||^2 + (c^T A_e c - 1)^2}$

$$x^T (A_e/\lambda) x - (A_e c/\lambda)^T x + (c^T A_e c - 1)/\lambda = 0$$

and define the new parameters

$$A := \frac{A_e}{\lambda}, \quad b := -\frac{A_e c}{\lambda}, \quad d := \frac{c^T A_e c - 1}{\lambda}$$

that satisfy the normalizing condition

we can renew c and A_e from A, b, d by

$$c = -A^{-1} b, \qquad \text{and} \qquad A_e = \frac{1}{c^T A c - d} A$$

the estimator of the true parameters $\bar{A}_e$ and $\bar{c}$ is

$$\hat{c} = -(\hat{A})^{-1} \hat{b} \qquad \text{and} \qquad \hat{A}_e = \frac{1}{\hat{c}^T \hat{A} \hat{c} - \hat{d}} \hat{A}$$

for small sample size $\hat{A}_e$ might not be positive definite; we do the following additional step, let

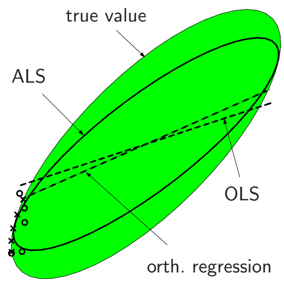
$$\hat{A}_e = \sum_{i=1}^{n} \hat{\lambda}_i \hat{v}_i^T \hat{v}_i$$

be the EVD of $\hat{A}_e$, then we redefine $\hat{A}_e$ by $\hat{A}_e := \sum_{i:\hat{\lambda}_i \geq 0} \hat{\lambda}_i \hat{v}_i^T \hat{v}_i$

Simulation examples with known center

the results depend on the error realization; for example:

ALS works, OLS and orthogonal regression fail



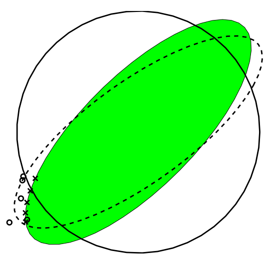
$$\bar{A} = \begin{bmatrix} 1.0000 & -0.7500 \\ -0.7500 & 1.0000 \end{bmatrix}$$

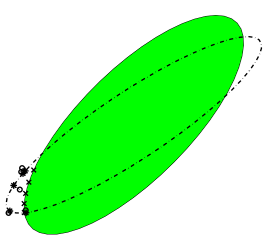
$$\hat{A}_{\text{ALS}} = \begin{bmatrix} 1.5740 & -1.6606 \\ -1.6606 & 2.4552 \end{bmatrix}$$

$$\hat{A}_{\text{OLS}} = \begin{bmatrix} 0.3289 & 0.3464 \\ 0.3464 & -0.6118 \end{bmatrix}$$

$$\hat{A}_{\text{ORTH}} = \begin{bmatrix} 0.1038 & 0.7480 \\ 0.7480 & -1.2612 \end{bmatrix}$$

ALS fails, OLS and orthogonal regression work





$$\hat{A}_{\text{ALS}} = \begin{bmatrix} 0.378 & -0.000 \\ 0.000 & 0.378 \end{bmatrix}, \; \hat{A}_{\text{OLS}} = \begin{bmatrix} 0.914 & -0.913 \\ -0.913 & 1.511 \end{bmatrix}, \; \hat{A}_{\text{ORTH}} = \begin{bmatrix} 1.588 & -2.047 \\ -2.047 & 3.313 \end{bmatrix}$$

Simulation examples with known center, cont.

we need to look at the average performance of the estimator

average performance for $m = 5$ points

for 1000 repetitions of the estimation (with different noise realization),
define the average relative error of estimation by

$$e := \frac{1}{N} \sum_{i=1}^{N} \frac{||\hat{A} - A_0||_F}{||A_0||_F}$$

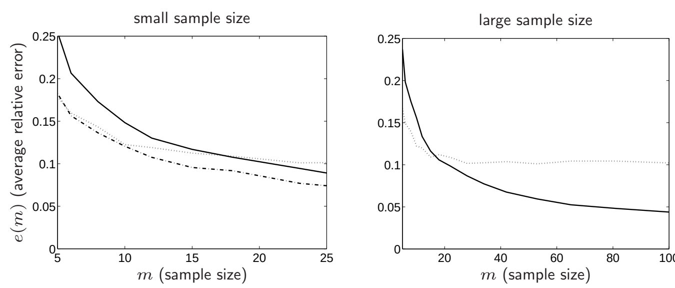
for the experiment shown above

$$e_{\text{ALS}} = 0.2375, \qquad e_{\text{OLS}} = 0.1715, \qquad e_{\text{ORTH}} = 0.1782$$

so ALS is indeed worse then OLS and orthogonal regression

asymptotic average performance

plot average performance as a function of the sample size

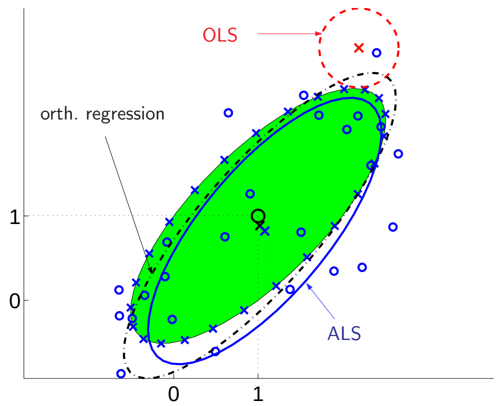


solid — ALS, dotted — OLS, dashed-dotted — orthogonal regression

OLS is biased, ALS is consistent

Simulation examples with unknown center

true ellipsoid, OLS, ALS, and orth. regression estimates



$$A_{\text{als}} = \begin{bmatrix} 1.146 & -0.739 \\ -0.739 & 0.877 \end{bmatrix},\ A_{\text{orth}} = \begin{bmatrix} 1.220 & -0.899 \\ -0.899 & 0.966 \end{bmatrix},\ c_{\text{als}} = \begin{bmatrix} 1.078 \\ 0.821 \end{bmatrix},\ c_{\text{orth}} = \begin{bmatrix} 1.020 \\ 0.886 \end{bmatrix}$$

relative errors: $e(A_{\text{als}}) = 0.108$, $e(A_{\text{orth}}) = 0.173$, $e(c_{\text{als}}) = 0.137$, $e(c_{\text{orth}}) = 0.081$

asymptotic average error of estimation