Adjusted least squares estimator for algebraic hypersurface fitting

Konstantin Usevich and Ivan Markovsky
Department ELEC, Vrije Universiteit Brussel

Centerparcs Heijderbos, 27 March 2014

33rd Benelux Meeting on Systems and Control
Algebraic curve fitting

Problem. Fit observed data points \( D = \{d^{(1)}, \ldots, d^{(N)}\} \subset \mathbb{R}^2 \), by a curve from a model class of curves \( \mathcal{P} \).

Example: class of conic sections
\[
\mathcal{P} = \left\{ \{d \in \mathbb{R}^2 \mid d^\top A d + b^\top d + c = 0\} \right\}
\]

hyperbola  
pair of lines  
ellipse
Algebraic curve fitting

Examples from:
N.Chernov, *Fitting geometric curves to observed data*, 2010:

Computer vision, medicine, nuclear physics, CAD, robotics ...

..., archaeology:

shape analysis of Greek stadia, pottery, *megalithic sites*
Introduction

From curves to algebraic hypersurfaces


This talk: degree 2 $\rightarrow$ higher degrees

Example: algebraic subspace clustering (Vidal, 2003)

$\begin{align*}
  d \in \mathcal{V}_1 \cup \cdots \cup \mathcal{V}_r \\
  \Leftrightarrow \\
  F(\theta, d) = 0 \\
  \text{union of subspaces} \\
  \text{algebraic hypersurface of degree } r
\end{align*}$
Outline of the talk

Algebraic hypersurface fitting

Adjusted least squares estimator

Conclusions
Outline of the talk

Algebraic hypersurface fitting

Adjusted least squares estimator

Conclusions
Algebraic hypersurfaces in $\mathbb{R}^q$

$F_A(\theta, d) = 0$ — implicit algebraic relation (hypersurface)

$F_A(\theta, d)$ is defined by $A = \{\alpha^{(1)}, \ldots, \alpha^{(n_\theta)}\} \subset \mathbb{Z}_+^q$, set of multidegrees

$\theta \in \mathbb{R}^{n_\theta}$

$F(\theta, d) = F_A(\theta, d) := \sum_{j=1}^{n_\theta} \theta_j \phi_j(d)$, where $\phi_j(d) := d_1^{\alpha_1^{(j)}} \cdots d_q^{\alpha_q^{(j)}}$


$F(\theta, [x, y]) = \theta_1 + \theta_2 y + \theta_3 x + \theta_4 x^2 + \theta_5 xy + \theta_6 y^2$

for conic section
Problem. Given points $\mathcal{D} = \{d^{(1)}, \ldots, d^{(N)}\}$, and $A = (\alpha^{(1)}, \ldots, \alpha^{(n_{\theta})})$ a set of multidegrees, find the best $\hat{\theta} \in \mathbb{R}^{n_{\theta}}$ such that

$$F(\hat{\theta}, d^{(j)}) \approx 0, j = 1, \ldots, N$$
Algebraic hypersurface fitting

Least squares vs. orthogonal regression

\[ F(\hat{\theta}, d^{(j)}) \approx 0 \]

- **Least squares fitting**

\[
\text{minimize} \sum_{j=1}^{N} (F(\hat{\theta}, d^{(j)}))^2
\]

- **Orthogonal regression**

\[
\text{minimize} \sum_{j=1}^{N} ||\hat{d}^{(j)} - d^{(j)}||^2_2 \\
\text{subject to} \quad F(\hat{\theta}, \hat{d}^{(j)}) = 0
\]
Algebraic hypersurface fitting

Least squares vs. orthogonal regression

\[ F(\hat{\theta}, d(j)) \approx 0 \]

Errors-in-variables model:
\[ d(j) = \bar{d}(j) + \tilde{d}(j), \; \tilde{d}(j) \sim N(0, \sigma^2 I) \]
\[ F(\bar{\theta}, \bar{d}(j)) = 0 \quad \text{— true hypersurface} \]

• Least squares fitting

\[
\text{minimize} \sum_{j=1}^{N} (F(\hat{\theta}, d(j)))^2
\]

• Orthogonal regression

\[
\text{minimize} \sum_{j=1}^{N} \|\hat{d}(j) - d(j)\|^2_2
\quad \text{subject to} \quad F(\hat{\theta}, \hat{d}(j)) = 0
\]
Least squares vs. orthogonal regression: estimation

\[ F(\hat{\theta}, d^{(j)}) \approx 0 \]

Errors-in-variables model:
\[ d^{(j)} = \bar{d}^{(j)} + \tilde{d}^{(j)}, \quad \tilde{d}^{(j)} \sim N(0, \sigma^2 I) \]
\[ F(\hat{\theta}, \bar{d}^{(j)}) = 0 \quad \text{true hypersurface} \]

- **Least squares estimator**
  \[
  \begin{array}{l}
  \text{minimize} \sum_{j=1}^{N} (F(\hat{\theta}, d^{(j)}))^2 \\
  \text{subject to} \quad F(\hat{\theta}, d^{(j)}) = 0
  \end{array}
  \]
  \( \hat{\theta} \) easy to compute
  \( \mathbb{E}(\hat{\theta}) < \infty \)
  biased, inconsistent

- **Orthogonal regression estimator**
  \[
  \begin{array}{l}
  \text{minimize} \sum_{j=1}^{N} \| \tilde{d}^{(j)} - d^{(j)} \|^2_2 \\
  \text{subject to} \quad F(\hat{\theta}, \tilde{d}^{(j)}) = 0
  \end{array}
  \]
  maximum likelihood
  hard to compute
  \( \not\exists \mathbb{E}(\hat{\theta}) \)
  asympt. biased, inconsistent
Outline of the talk

Algebraic hypersurface fitting

Adjusted least squares estimator

Conclusions
Least squares estimator: easy to compute

\[ F(\theta, d) := \sum_{j=1}^{n_{\theta}} \theta_j \phi_j(d), \quad \phi_j(d) \text{ — monomial} \]

\[ \hat{\theta}_{LS} = \arg\min_{\|\hat{\theta}\|_2=1} \sum_{j=1}^{N} (F(\hat{\theta}, d^{(j)}))^2 = \arg\min_{\|\hat{\theta}\|_2=1} \sum_{j=1}^{N} \|\hat{\theta}^\top \Phi(D)\|^2, \]

where

\[ \Phi(D) := \begin{bmatrix} \phi_1(d^{(1)}) & \cdots & \phi_1(d^{(N)}) \\ \vdots & \ddots & \vdots \\ \phi_{n_\theta}(d^{(1)}) & \cdots & \phi_{n_\theta}(d^{(N)}) \end{bmatrix} \text{ — multivariate Vandermonde matrix} \]

\[ \hat{\theta}_{LS} = \text{last eigenvector of } \Psi(D), \quad \Psi(D) := \Phi(D)\Phi^\top(D) \]
Adjusted least squares estimator

Closer look on $\Psi(D)$

$\hat{\theta}_{LS} =$ last e.v. of $\Psi(D)$, \quad $\Psi(D) := \Phi(D)\Phi^\top(D)$, \quad $\Phi$ — Vandermonde

Example: $F(\theta, [x\ y]) = \theta_1 + \theta_2 y + \theta_3 x + \theta_4 x^2 + \theta_5 x y + \theta_6 y^2$

$\mathcal{D} = \{[x_1\ y_1], \ldots, [x_N\ y_N]\}$

$\Psi(D) = \begin{bmatrix}
\sum_k 1 & \sum_k x_k & \sum_k y_k & \sum_k x_k^2 & \sum_k x_k y_k & \sum_k y_k^2 \\
\sum_k x_k & \sum_k x_k^2 & \sum_k x_k y_k & \sum_k x_k^3 & \sum_k x_k^2 y_k & \sum_k x_k y_k^2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\sum_k y_k^2 & \sum_k x_k y_k^2 & \sum_k y_k^3 & \sum_k x_k^2 y_k^2 & \sum_k x_k y_k^3 & \sum_k y_k^4
\end{bmatrix}$

quasi-Hankel matrix of moments
Least squares estimator: bias and adjustment

\[ \hat{\theta}_{LS}(D) = \text{last eigenvector of } \Psi(D), \quad \Psi(D) — \text{matrix of moments} \]

Errors-in-variables model: \( D = \overline{D} + \tilde{D} \) (i.e., \( d^{(k)} = \overline{d}^{(k)} + \tilde{d}^{(k)} \))

a) \( \tilde{d}^{(k)} = 0 \quad \Rightarrow \quad \Psi(D) = \Psi(\overline{D}) \quad \Rightarrow \quad \hat{\theta}_{LS}(D) = \overline{\theta} \)

b) \( \tilde{d}^{(k)} \sim N(0, \sigma^2 I) \quad \Rightarrow \quad E(\Psi(D)) \neq \Psi(\overline{D}) — \text{the source of bias} \)

Adjustment (bias-correction): (Kukush, Markovsky, Van Huffel, 2004)
Construct \( \Psi_{ALS,\sigma}(D) \), such that

\[
E(\Psi_{ALS,\sigma}(D)) = \Psi(\overline{D}) \quad \text{for any } \overline{D}
\]
Adjusted least squares estimator

Adjustment procedure: some details

Hermite polynomials: orthogonal polynomials w.r.t. \( w(x) = ce^{-\frac{x^2}{2\sigma^2}} \)

\[
\begin{align*}
h_0(x) &= 1, \\
h_1(x) &= x, \\
h_k(x) &= xh_{k-1}(x) - (k - 1)\sigma^2 h_{k-2}(x).
\end{align*}
\]

For any \( a \in \mathbb{R} \) and \( \xi \sim N(0, \sigma^2) \)

\[
\mathbb{E}(h_k(a + \xi)) = a^k
\]

\( \rightarrow \) in the moment matrix \( \Psi(\mathcal{D}) \), replace each monomial by a product of Hermite polynomials

\[
\mathbb{E}(\Psi_{ALS,\sigma}(\mathcal{D})) = \Psi(\mathcal{D}) \quad \text{for any } \mathcal{D}
\]
Adjusted least squares estimator

1. Variant with known $\sigma$

$$\hat{\theta}_{ALS,\sigma} := \arg\min \hat{\theta}^T \Psi_{ALS,\sigma} \hat{\theta} \quad \text{s.t.} \quad ||\hat{\theta}||_2 = 1$$

2. Variant with unknown $\sigma$

$$\hat{\theta}_{ALS} := \arg\min \hat{\theta}^T \Psi_{ALS,\hat{\sigma}} \hat{\theta} \quad \text{— polynomial eigenvalue problem}$$

Both estimators are consistent ($N \to \infty$) (Shklyar, 2008), (Kukush, Markovsky, Van Huffel, 2004-06).
Adjusted least squares estimator

1. Variant with known $\sigma$

$$\hat{\theta}_{ALS,\sigma} := \arg\min_{\|\hat{\theta}\|_2=1} \hat{\theta}^\top \Psi_{ALS,\sigma} \hat{\theta}$$

--- last eigenvector of $\Psi_{ALS,\sigma}$

2. Variant with unknown $\sigma$

$$\hat{\theta}_{ALS} := \arg\min_{\|\hat{\theta}\|_2=1, \hat{\sigma} \geq 0} \hat{\theta}^\top \Psi_{ALS,\hat{\sigma}} \hat{\theta}$$

--- polynomial eigenvalue problem

- Both estimators are consistent ($N \to \infty$) (Shklyar, 2008), (Kukush, Markovsky, Van Huffel, 2004-06)
- $\not\mathbb{E}(\hat{\theta})$
- $\hat{\theta}_{ALS}$ works for small data sets
- Better to estimate $\sigma$ even if it is known
Example: hyperbola

\[ y^2 - x^2 - 1 = 0, \ N = 30, \ \sigma = 0.3 \]

black: \( \overline{\theta} \), blue: \( d^{(k)} \), green: \( \hat{\theta}_{LS} \), red: \( \hat{\theta}_{ALS, \sigma} \), orange: \( \hat{\theta}_{ALS} \)
Invariance of the estimators

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\theta}_{LS}$</th>
<th>$\hat{\theta}_{ALS,\sigma}$</th>
<th>$\hat{\theta}_{ALS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation-invariant</td>
<td>$+^*$</td>
<td>$+^*$</td>
<td>$+^*$</td>
</tr>
<tr>
<td>Translation/homothety-invariant</td>
<td>$-$</td>
<td>$-$</td>
<td>$+^*$</td>
</tr>
</tbody>
</table>

blue: $d^{(k)}$, green: $\hat{\theta}_{LS}$, red: $\hat{\theta}_{ALS,\sigma}$, orange: $\hat{\theta}_{ALS}$
Outline of the talk

Algebraic hypersurface fitting

Adjusted least squares estimator

Conclusions
Conclusions

Summary:

<table>
<thead>
<tr>
<th>Consistency</th>
<th>degree = 2</th>
<th>degree &gt; 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Kukush, Markovsky,</td>
<td>(Shklyar, 2008)</td>
<td></td>
</tr>
<tr>
<td>Van Huffel, Shklyar)</td>
<td>2004-2006</td>
<td>(Markovsky, 2012), this talk</td>
</tr>
<tr>
<td>Invariance</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Open questions (TODO list):

- Why $\hat{\theta}_{ALS}$ works better? (with unknown $\sigma$)
- What to do with $\mathbb{E}(\hat{\theta})$?
- Can we enforce structure on $\hat{\theta}$? (factorizable, ...)
- What happens in dynamic problems?
Thank you for your attention!