

Fitting algebraic curves to data

Ivan Markovsky

University of Southampton

1 / 33

Dimension of affine variety

image representation:

$$\mathcal{B} = \{d \mid d = P(\ell), \text{ for all } \ell \in \mathbb{R}^g\}$$

$\dim(\mathcal{B})$ =: minimum g in image representation of \mathcal{B}

affine variety of dimension one is called **algebraic curve**

3 / 33

Affine variety

consider system of p , q -variate polynomials

$$r_i(d_1, \dots, d_q) = 0, \quad i = 1, \dots, p \quad \Longleftrightarrow \quad R(d) = 0$$

the set of their real valued solutions

$$\mathcal{B} = \{d \in \mathbb{R}^q \mid R(d) = 0\}$$

is affine variety

of primary interest for data modeling is the set \mathcal{B} (the model)

$R(d) = 0$ is demoted to (kernel) **representation of \mathcal{B}**

2 / 33

Algebraic curves in 2D

in the special case $q = 2$, we use

$$x := d_1, \quad \text{and} \quad y := d_2.$$

the set

$$\mathcal{B} = \{(x, y) \in \mathbb{R}^2 \mid r(x, y) = 0\}$$

may be

- empty, e.g., $r(x, y) = x^2 + y^2 + 1$
- finite (isolated points), e.g., $r(x, y) = x^2 + y^2$, or
- infinite (curve), e.g., $r(x, y) = x^2 + y^2 - 1$

4 / 33

Examples

- **subspace** linear \mathcal{B} ($q \geq 2$, zeroth degree repr.)
- **conic section** second order algebraic curve in \mathbb{R}^2
- **cissoid** $\mathcal{B} = \{(x, y) \mid y^2(1+x) = (1-x)^3\}$
- **folium of Descartes** $\mathcal{B} = \{(x, y) \mid x^3 + y^3 - 3xy = 0\}$
- **four-leaved rose** $\mathcal{B} = \{(x, y) \mid (x^2 + y^2)^3 - 4x^2y^2 = 0\}$

5/33

The fitting problem

Given:

- data points $\mathcal{D} = \{d_1, \dots, d_N\}$
- set of candidate curves (model class) \mathcal{M}
- data-model distance measure $\text{dist}(d, \mathcal{B})$

find model $\hat{\mathcal{B}} \in \mathcal{M}$ that is as close as possible to the data:

minimize over $\mathcal{B} \in \mathcal{M}$ $\text{dist}(\mathcal{D}, \mathcal{B})$

12/33

Explicit vs implicit representations

- function $y = f(x)$ vs relation $(r(x, y) = 0)$ (mathematics)
- input/output vs kernel representation (system theory)
- regression vs EIV regression (statistics)
- functional vs structural models (statistics)

11/33

Algebraic vs geometric distance measures

geometric distance: $\text{dist}(d, \mathcal{B}) := \min_{\hat{d} \in \mathcal{B}} \|d - \hat{d}\|$

algebraic “distance”: $\|R(d)\|$ where R defines kernel repr. of \mathcal{B}

other interpretations:

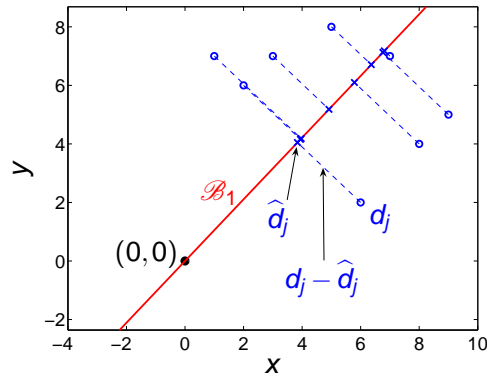
- misfit vs latency

P. Lemmerling and B. De Moor, Misfit versus latency, Automatica, 37:2057–2067, 2001

- algebraic \leftrightarrow LS \leftrightarrow ARMAX
- geometric \leftrightarrow TLS/PCA \leftrightarrow EIV SYSID

13/33

Example: geometric distance to a linear model



$$\text{dist}(\mathcal{D}, \mathcal{B}_1) = \min_{\hat{d}_1, \dots, \hat{d}_8 \in \mathcal{B}_1} \sqrt{\sum_{j=1}^8 \|d_j - \hat{d}_j\|_2^2} = 7.8865$$

14/33

Kernel representation in 2D

$$r(d) = \sum_{k=1}^{q_{\text{ext}}} \theta_k \phi_k(d) = \theta \phi(d) \quad \begin{array}{l} \text{linear in } \theta \\ \text{nonlinear in } d \end{array}$$

- θ — (row) vector of parameters
- $\phi(d)$ — vector of monomials, e.g.,

$$q=2, \quad n := \deg(r) = 2 \quad \rightsquigarrow \quad \phi(d) = [x^2 \quad xy \quad x \quad y^2 \quad y \quad 1]^\top$$

$$n=3 \rightsquigarrow \phi(d) = [x^3 \quad x^2y \quad x^2 \quad xy^2 \quad xy \quad x \quad y^3 \quad y^2 \quad y \quad 1]^\top$$

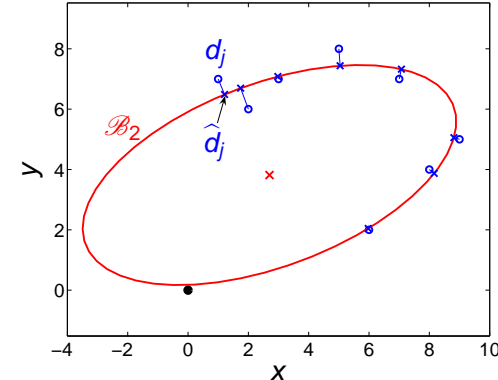
- $q_{\text{ext}} = \binom{q+n}{n}$ — measure of **complexity of \mathcal{M}_n**

the degree n is the only design parameter in the curve fitting prob.

- θ is nonunique, θ and $\alpha\theta$, for all $\alpha \neq 0$, define the same \mathcal{B}

16/33

Example: geometric distance to a quadratic model



$$\text{dist}(\mathcal{D}, \mathcal{B}_2) = \min_{\hat{d}_1, \dots, \hat{d}_8 \in \mathcal{B}_2} \sqrt{\sum_{j=1}^8 \|d_j - \hat{d}_j\|_2^2} = 1.1719$$

15/33

Algebraic curve fitting in \mathbb{R}^2

$$\text{minimize over } \|\theta\|_2 = 1 \quad \sum_{j=1}^N \|r_\theta(d_j)\|_2^2$$

$$\begin{aligned} \sum_{j=1}^N \|r_\theta(d_j)\|_2^2 &= \|\theta [\phi(d_1) \quad \dots \quad \phi(d_N)]\|_2^2 \\ &= \theta \Phi(\mathcal{D}) \Phi^\top(\mathcal{D}) \theta^\top = \theta \Psi(\mathcal{D}) \theta^\top \end{aligned}$$

algebraic curve fitting is eigenvalue problem

$$\text{minimize over } \|\theta\|_2 = 1 \quad \theta \Psi(\mathcal{D}) \theta^\top$$

or, equivalently, **(unstructured) low rank approximation problem**

$$\begin{aligned} &\text{minimize over } \hat{\Phi} \quad \|\Phi(\mathcal{D}) - \hat{\Phi}\|_F \\ &\text{subject to} \quad \text{rank}(\hat{\Phi}) \leq q_{\text{ext}} - 1 \end{aligned}$$

17/33

Geometric distance

$$\text{minimize over } \hat{\mathcal{D}} \subset \mathcal{B} \quad \left\| \underbrace{[d_1 \ \cdots \ d_N]}_D - \underbrace{[\hat{d}_1 \ \cdots \ \hat{d}_N]}_{\hat{D}} \right\|_F$$

$$\text{let } \mathcal{B} = \{d \mid \theta\phi(d) = 0\}$$

$$\hat{\mathcal{D}} \subset \mathcal{B} \iff \hat{d}_j \in \mathcal{B}, \text{ for } j = 1, \dots, N$$

$$\iff \theta\phi(\hat{d}_j) = 0, \text{ for } j = 1, \dots, N$$

$$\iff \theta\Phi(\hat{\mathcal{D}}) = 0$$

the problem of computing the geometric distance is:

$$\text{minimize over } \hat{\mathcal{D}} \quad \|D - \hat{D}\|_F \quad \text{subject to} \quad \theta\Phi(\hat{\mathcal{D}}) = 0$$

18/33

Bias corrected low rank approximation

assume that \mathcal{D} is generated by the **errors-in-variables model**

$$d_j = d_{0,j} + \tilde{d}_j, \quad \text{where } d_{0,j} \in \mathcal{B}_0 \text{ and } \tilde{d}_j \sim N(0, \sigma^2 I_q) \quad (\text{EIV})$$

- \mathcal{B}_0 is the “true” model
- $\mathcal{D}_0 := \{d_{0,1}, \dots, d_{0,N}\}$ is the true data, and
- $\tilde{\mathcal{D}} := \{\tilde{d}_1, \dots, \tilde{d}_N\}$ is the measurement noise

the estimate obtained by the algebraic fitting method is biased

define the matrices

$$\Psi := \Phi(\mathcal{D})\Phi^\top(\mathcal{D}) \quad \text{and} \quad \Psi_0 := \Phi(\mathcal{D}_0)\Phi^\top(\mathcal{D}_0)$$

we construct “corrected” matrix Ψ_c , such that $\mathbf{E}(\Psi_c) = \Psi_0$

20/33

Geometric curve fitting

$$\text{minimize over } \mathcal{B} \in \mathcal{M}_n \quad \text{dist}(\mathcal{D}, \mathcal{B})$$

assuming that $N \geq q_{\text{ext}}$, we have

$$\theta\Phi(\hat{\mathcal{D}}) = 0, \theta \neq 0 \iff \text{rank}(\Phi(\hat{\mathcal{D}})) \leq q_{\text{ext}} - 1, \quad q_{\text{ext}} := \binom{2+n}{n}$$

geometric curve fitting is **nonlinearly structured low rank approx.:**

$$\text{minimize over } \hat{\mathcal{D}} \text{ and } \theta \quad \|D - \hat{D}\|$$

$$\text{subject to} \quad \text{rank}(\Phi(\hat{\mathcal{D}})) \leq q_{\text{ext}} - 1$$

note: algebraic fitting is a relaxation of geometric fitting, obtained by removing the structure constraint

19/33

Hermite polynomials

the polynomials

$$h_0(x) = 1, \quad h_1(x) = x, \quad \text{and}$$

$$h_k(x) = xh_{k-1}(x) - (k-2)h_{k-2}(x), \quad \text{for } k = 2, 3, \dots$$

have the property

$$\mathbf{E}(h_k(x_0 + \tilde{x})) = x_0^k, \quad \text{where } \tilde{x} \sim N(0, \sigma^2) \quad (**)$$

21/33

Derivation of the correction

$$\Psi = \sum_{\ell=1}^N \phi(d_\ell) \phi^\top(d_\ell) = \sum_{\ell=1}^N [\phi_i(d_\ell) \phi_j(d_\ell)]$$

where the monomials ϕ_i are

$$\phi_k(d) = d_1^{n_{k1}} \dots d_q^{n_{kq}}, \quad \text{for } k = 1, \dots, q_{\text{ext}}$$

the (i, j) th element of Ψ is

$$\psi_{ij} = \sum_{\ell=1}^N d_{1\ell}^{n_{i1}+n_{j1}} \dots d_{q\ell}^{n_{iq}+n_{jq}} = \sum_{\ell=1}^N \prod_{k=1}^q (d_{0,k\ell} + \tilde{d}_{k\ell})^{n_{iq}+n_{jq}}$$

by (EIV), $\tilde{d}_{k\ell}$ are independent, zero mean, normally distributed

22/33

Comparison of algebraic, bias corrected, and geometric fits on simulation examples

Simulation setup: $q = 2, p = 1$

- true model $\mathcal{B}_0 = \{d \mid \theta_0 \phi(d) = 0\}$
- data points $d = d_0 + \tilde{d}, d_0 \in \mathcal{B}_0, \tilde{d} \sim N(0, \sigma^2 I)$
- algebraic fit — black dotted line
- bias corrected fit — **dashed dotted line**
- geometric fit — **dashed line**

24/33

then, by the property $(**)$ of the Hermite polynomials

$$\phi_{c,ij} := \sum_{\ell=1}^N \prod_{k=1}^q h_{n_{iq}+n_{jq}}(d_{k\ell})$$

has the desired property

$$\mathbf{E}(\psi_{c,ij}) = \sum_{\ell=1}^N \prod_{k=1}^q d_{0,k\ell}^{n_{iq}+n_{jq}} =: \psi_{0,ij}$$

the corrected matrix Ψ_c is an even polynomial in σ

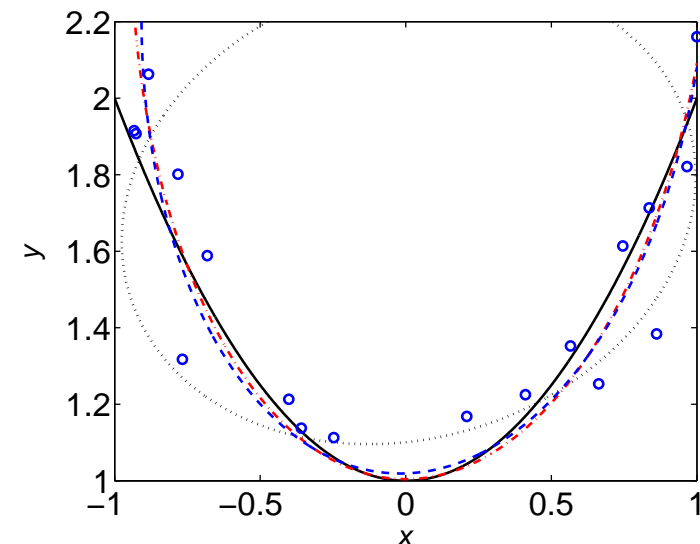
$$\Psi_c(\sigma^2) = \Psi_{c,0} + \sigma^2 \Psi_{c,1} + \dots + \sigma^{2n_\Psi} \Psi_{c,n_\Psi}$$

the estimate $\hat{\theta}$ is in the null space of $\Psi_c(\sigma^2)$, i.e., $\Psi_c(\sigma^2) \hat{\theta} = 0$

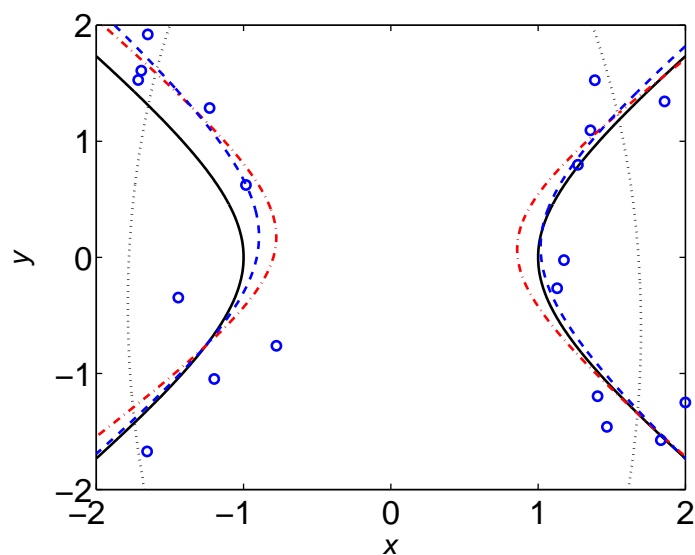
computing simultaneously σ and θ is a **polynomial EVP**

23/33

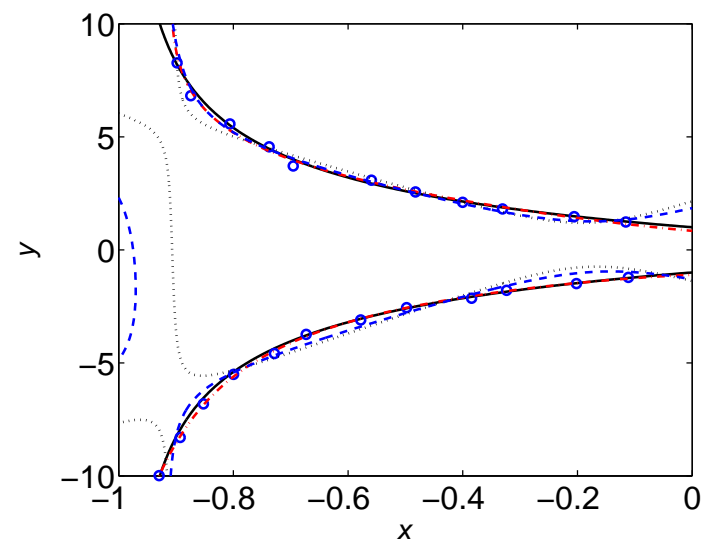
Parabola $y = x^2 + 1$



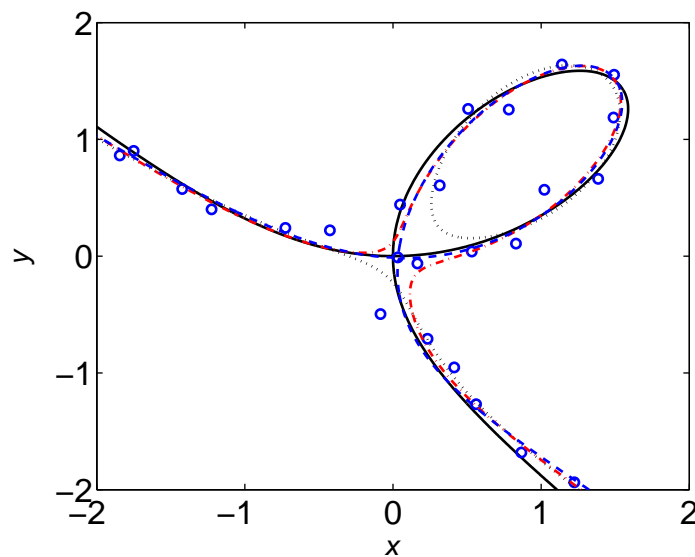
25/33

Hyperbola $x^2 - y^2 - 1 = 0$ 

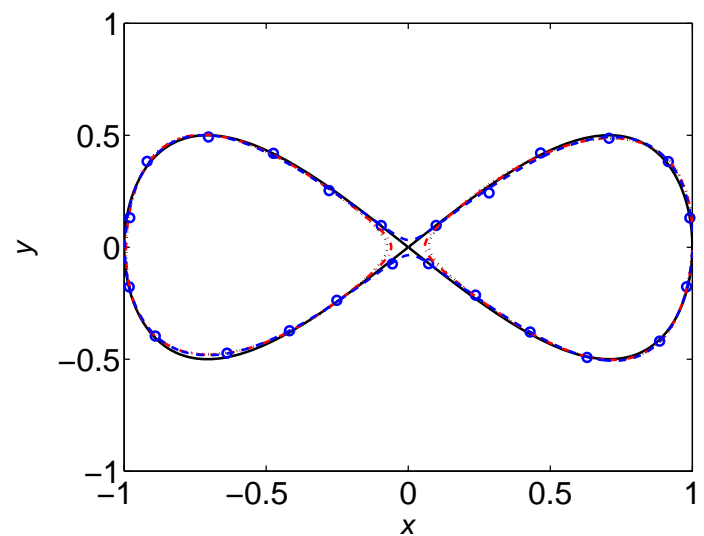
26/33

Cisoid $y^2(1+x) = (1-x)^3$ 

27/33

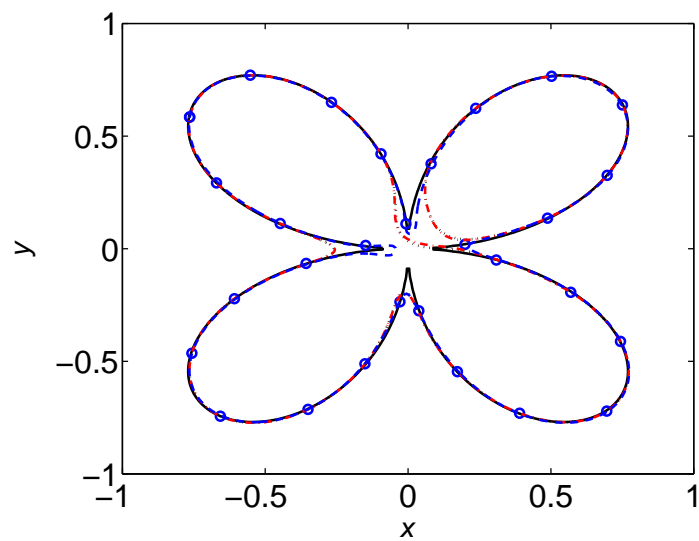
Folium of Descartes $x^3 + y^3 - 3xy = 0$ 

28/33

Eight curve $y^2 - x^2 + x^4 = 0$ 

29/33

Rose $(x^2 + y^2)^3 - 4x^2y^2 = 0$



“Special data” example

