Fitting algebraic curves to data

Ivan Markovsky

University of Southampton

Affine variety

consider system of p, q-variate polynomials

$$r_i(w_1, \dots, w_q) = 0, \quad i = 1, \dots, p \qquad \Longleftrightarrow \qquad R(w) = 0$$

the set of their real valued solutions

$$\mathscr{B} = \{ w \in \mathbb{R}^q \mid R(w) = 0 \}$$

is affine variety

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

$R(w) = 0$ is demoted to (kernel) representation of $\mathscr{B}$

Dimension of affine variety

image representation:

$$\mathscr{B} = \{ w \mid w = P(u), \text{ for all } u \in \mathbb{R}^g \}$$

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

affine variety of dimension one is called algebraic curve

Algebraic curves in 2D

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

$$x := w_1 \qquad \text{and} \qquad y := w_2$$

the set

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

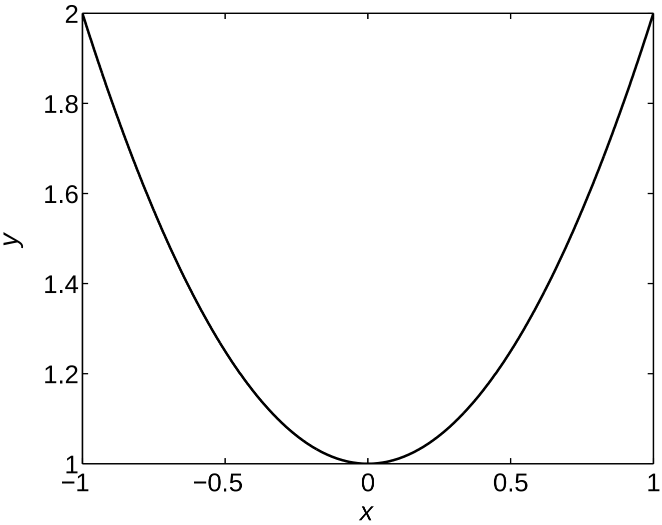
may be

- 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.*, $r(x,y) = x^2 + y^2 - 1$

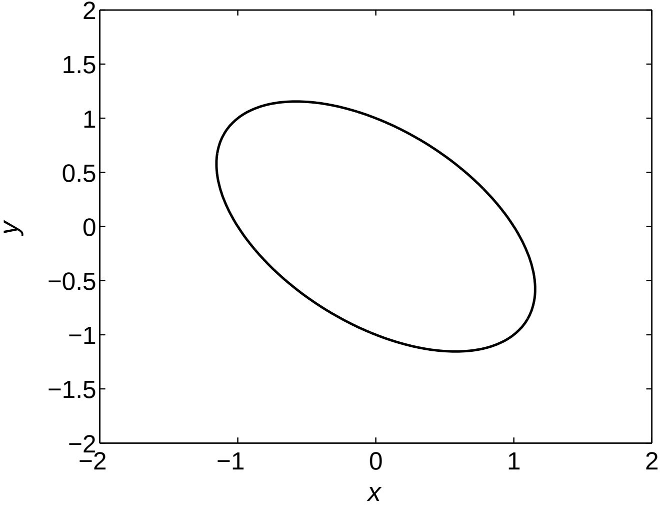
Examples

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

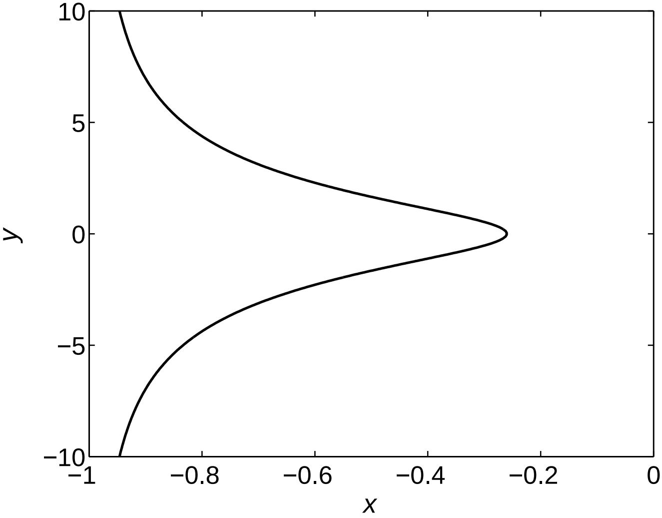
Parabola $y = x^2 + 1$



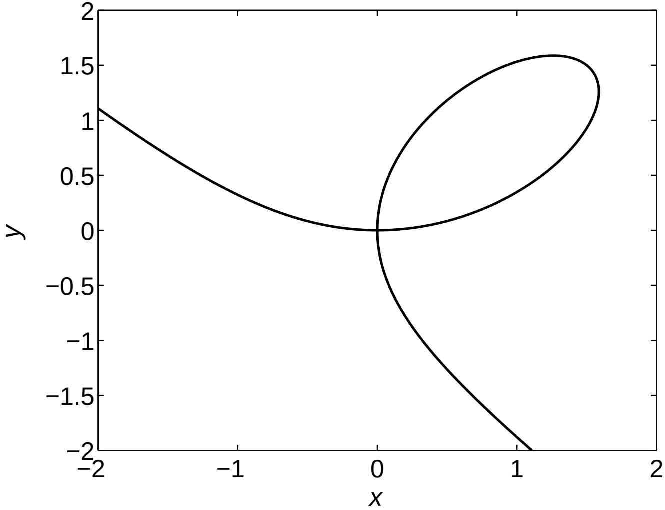
Ellipse $y^2 + xy + x^2 - 1 = 0$



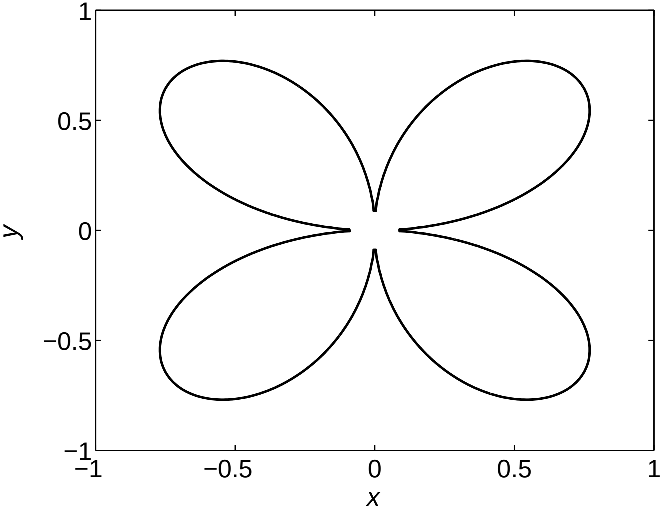
Cissoid $y^2(1+x) = (1-x)^3$



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



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



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)

The fitting problem

Given:

- data points $w_d = \{\, w_d(1), \dots, w_d(N) \,\}$
- set of candidate curves (model class) $\mathscr{M}$
- data-model distance measure $\operatorname{dist}(w_d, \mathscr{B})$

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

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{M} \quad \operatorname{dist}(w_d, \mathscr{B})$$

Algebraic vs geometric distance measures

geometric distance: $\operatorname{dist}(w_\mathrm{d}, \mathscr{B}) := \min_{\widehat{w} \subset \mathscr{B}} \|w_\mathrm{d} - \widehat{w}\|$

algebraic "distance": $\|R(w_\mathrm{d})\|$ where R defines kernel repr. of $\mathscr{B}$

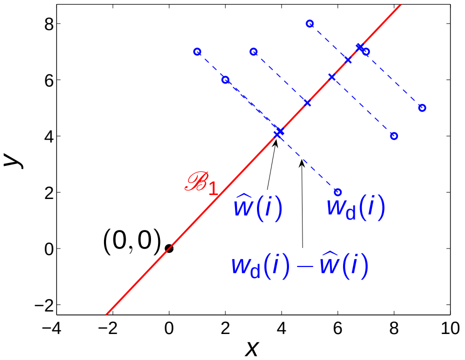
other interpretations:

- misfit vs latency

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

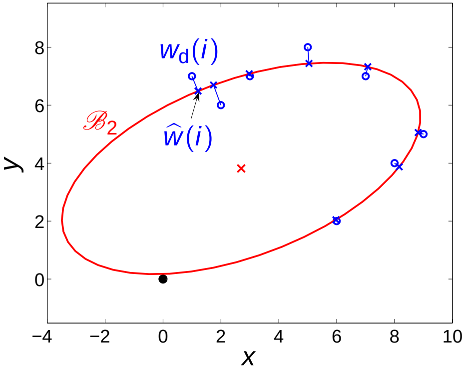
- algebraic $\leftrightarrow$ LS $\leftrightarrow$ ARMAX
- geometric $\leftrightarrow$ TLS/PCA $\leftrightarrow$ EIV SYSID

Example: geometric distance to a linear model



$$\mathrm{dist}(w_d, \mathscr{B}_1) = \min_{\widehat{w}(1),\dots,\widehat{w}(8)\in\mathscr{B}_1} \sqrt{\sum_{t=1}^{8} \|w_d(t) - \widehat{w}(t)\|_2^2} = 7.8865$$

Example: geometric distance to a quadratic model



$$\operatorname{dist}(w_d, \mathscr{B}_2) = \min_{\widehat{w}(1),\dots,\widehat{w}(8) \in \mathscr{B}_2} \sqrt{\sum_{t=1}^{8} \|w_d(t) - \widehat{w}(t)\|^2} = 1.1719$$

Kernel representation in 2D

$$r(w) = \sum_{k=1}^{n_\theta} \theta_k \phi_k(w) = \phi(w)\theta$$

linear in θ
nonlinear in w

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

$$q = 2, \quad d := \deg(r) = 2 \quad \rightsquigarrow \quad \phi(w) = \begin{bmatrix} x^2 & xy & x & y^2 & y & 1 \end{bmatrix}$$

$$d = 3 \rightsquigarrow \phi(w) = \begin{bmatrix} x^3 & x^2y^1 & x^2 & xy^2 & xy & x & y^3 & y^2 & y & 1 \end{bmatrix}$$

- $n_\theta = \binom{q+d}{d}$ — measure of complexity of $\mathscr{M}_d$

the degree d 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 $\mathscr{B}$

Algebraic curve fitting in $\mathbb{R}^2$

$$\text{minimize} \quad \text{over } \|\theta\|_2 = 1 \quad \sum_{i=1}^{N} \|r_\theta\big(w_\text{d}(i)\big)\|_2^2$$

$$\sum_{i=1}^{N} \|r_\theta\big(w_\text{d}(i)\big)\|_2^2 = \left\| \begin{bmatrix} \phi\big(w_\text{d}(1)\big) \\ \vdots \\ \phi\big(w_\text{d}(N)\big) \end{bmatrix} \theta \right\|_2^2 = \theta^\top \Phi^\top(w_\text{d}) \Phi(w_\text{d}) \theta = \theta^\top \Psi(w_\text{d}) \theta$$

algebraic curve fitting is eigenvalue problem

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

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

$$\begin{aligned} &\text{minimize} && \text{over } \widehat{\Phi} \text{ and } \theta \quad \|\Phi(w_\text{d}) - \widehat{\Phi}\|_\text{F} \\ &\text{subject to} && \operatorname{rank}(\widehat{\Phi}) \leq n_\theta - 1 \end{aligned}$$

Geometric distance

$$\text{minimize} \quad \text{over } \widehat{w} \subset \mathscr{B} \quad \|w_\mathrm{d} - \widehat{w}\|$$

let $\mathscr{B} = \{\, w \mid \phi(w)\theta = 0 \,\}$

$$\begin{aligned} \widehat{w} \subset \mathscr{B} \quad & \iff \quad \widehat{w}(i) \in \mathscr{B}, \quad \text{for } i = 1, \dots, N \\ & \iff \quad \phi\big(\widehat{w}(i)\big)\theta = 0, \quad \text{for } i = 1, \dots, N \\ & \iff \quad \Phi(\widehat{w})\theta = 0 \end{aligned}$$

the problem of computing the geometric distance is:

$$\text{minimize} \quad \text{over } \widehat{w} \quad \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \Phi(\widehat{w})\theta = 0$$

Geometric curve fitting

$$\text{minimize} \quad \text{over } \mathscr{B} \in \mathscr{M}_d \quad \operatorname{dist}(w_\mathrm{d}, \mathscr{B})$$

assuming that $N \geq n_\theta$, we have

$$\Phi(\widehat{w})\theta = 0, \ \theta \neq 0 \quad \iff \quad \operatorname{rank}\big(\Phi(\widehat{w})\big) \leq n_\theta - 1, \quad n_\theta := \binom{2+d}{d}$$

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

$$\begin{aligned} &\text{minimize} && \text{over } \widehat{w} \text{ and } \theta \quad \|w_\mathrm{d} - \widehat{w}\| \\ &\text{subject to} && \operatorname{rank}\big(\Phi(\widehat{w})\big) \leq n_\theta - 1 \end{aligned}$$

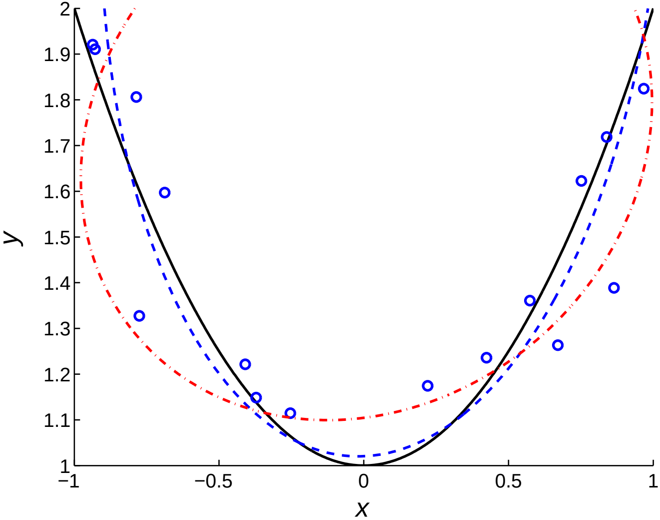
note: algebraic fitting is a relaxation of geometric fitting, obtained by removing the structure constraint

Comparison of algebraic and geometric fits

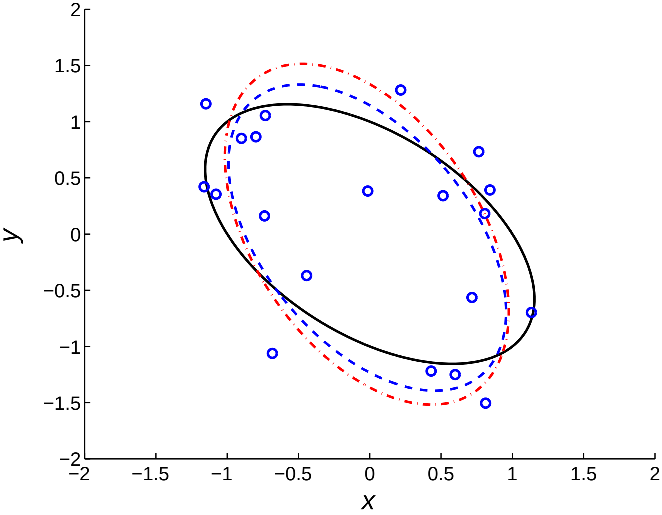
Simulation setup:

- true model $\bar{\mathscr{B}} = \{ w \mid \phi(w)\bar{\theta} = 0 \}$, $(q = 2, p = 1)$
- data points $w_{\mathrm{d}} = \bar{w} + \widetilde{w}$, $\bar{w} \subset \bar{\mathscr{B}}$, $\widetilde{w} \sim \mathrm{N}(0, \sigma^2 I)$
- algebraic fit — dashed dotted line
- geometric fit — dashed line

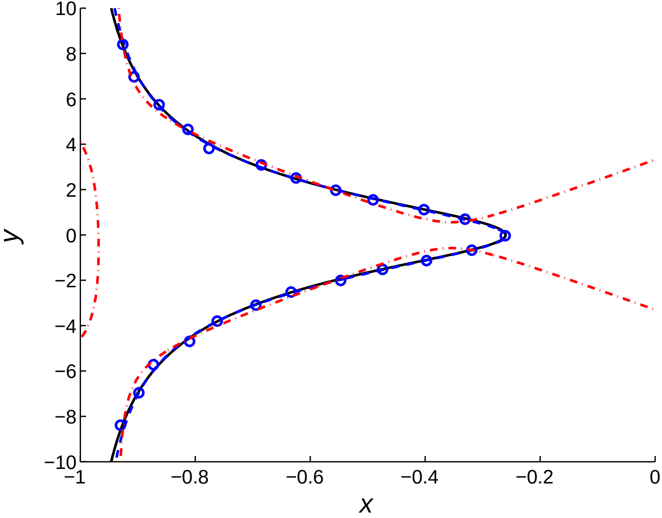
Parabola $y = x^2 + 1$



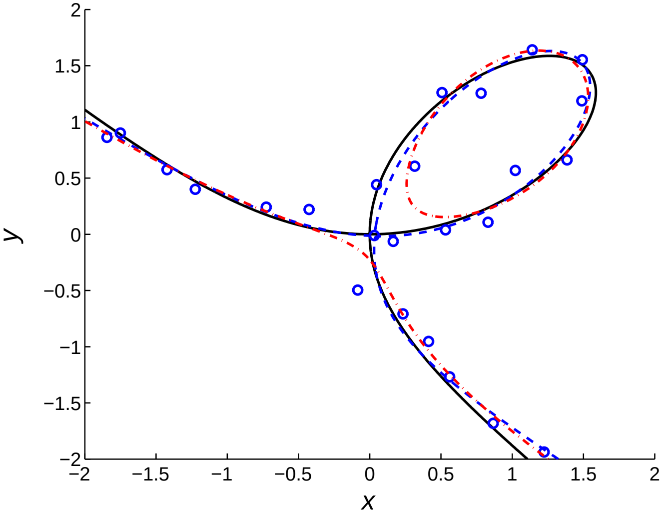
Ellipse $y^2 + xy + x^2 - 1 = 0$



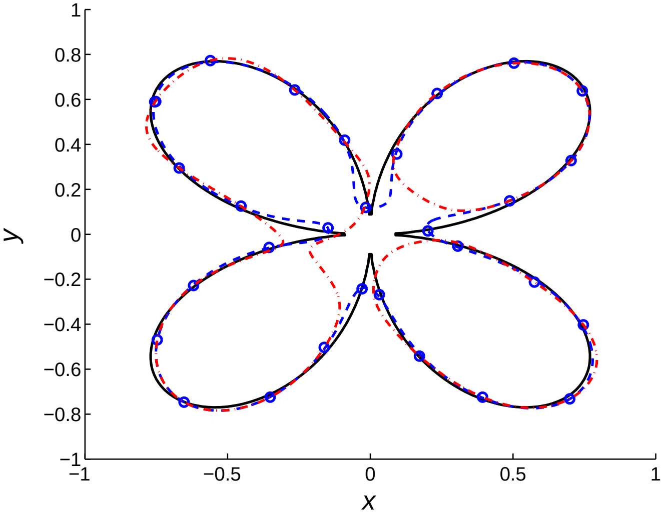
Cissoid $y^2(1+x) = (1-x)^3$



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



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



new application of structured low-rank approximation
the first I know of with nonlinear structure

To-do list:

- Robust and efficient optimization methods
- Noniterative (subspace-type) methods
- Generalize to nD (vector polynomials)
- Link to linear system identification
- Link to related curve fitting methods, *e.g.*, principal curves
- Statistical properties
- Impact on applications

Questions?