Fitting algebraic curves to data

Ivan Markovsky

University of Southampton

Affine variety

consider system of p, q-variate polynomials

$$r_i(d_{1\cdot},\dots,d_{q\cdot}) = 0, \quad i = 1,\dots,p \qquad \iff \qquad R(d) = 0$$

the set of their real valued solutions

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

is affine variety

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

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

Dimension of affine variety

image representation:

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

$\dim(\mathscr{B}) =:$ 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

$$x := d_{1\cdot} \qquad \text{and} \qquad y := d_{2\cdot}$$

the set

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

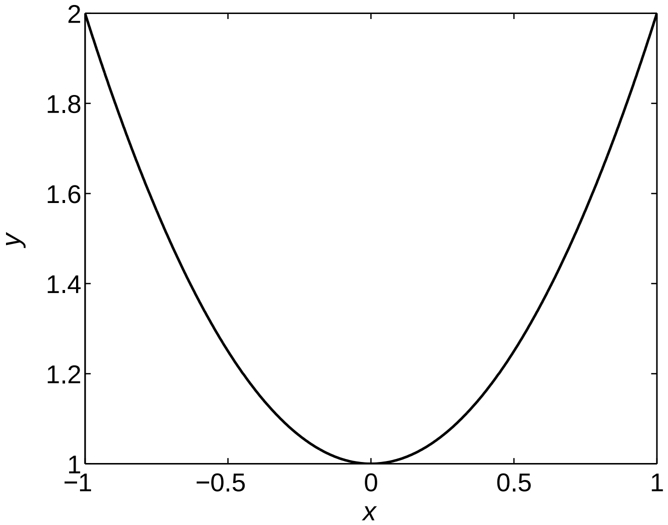
may be

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

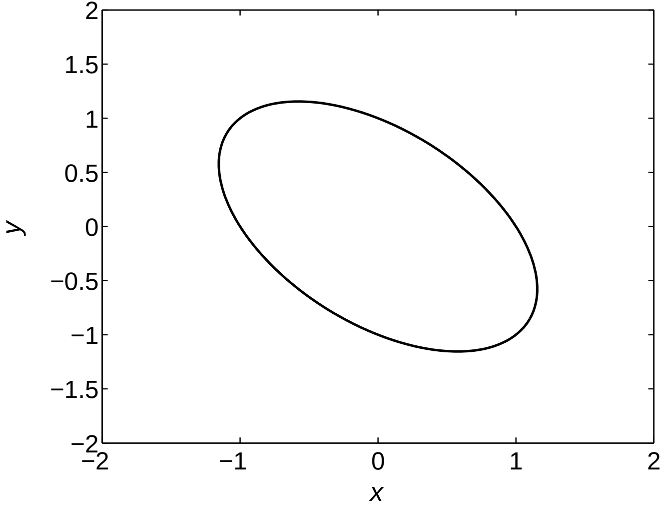
Examples

- 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 — $\mathscr{B} = \{(x,y) \mid (x^2+y^2)^3 - 4x^2y^2 = 0\}$

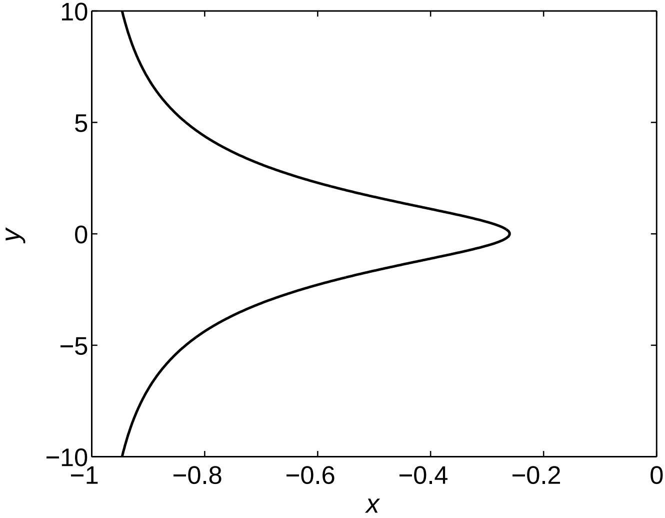
Parabola $y = x^2 + 1$



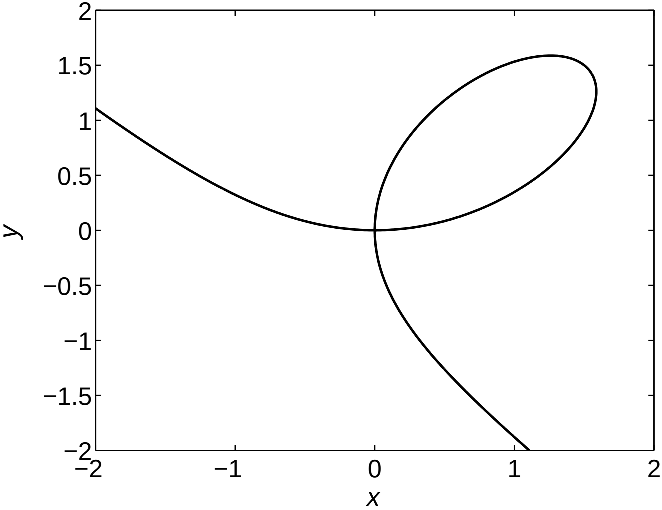
Ellipse $y^2 + xy + x^2 - 1 = 0$



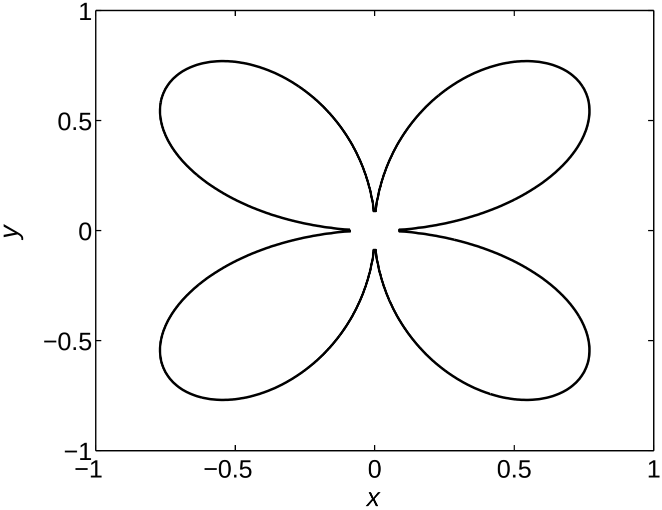
Cissoid $y^2(1+x) = (1-x)^3$



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

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

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

Algebraic vs geometric distance measures

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

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

other interpretations:

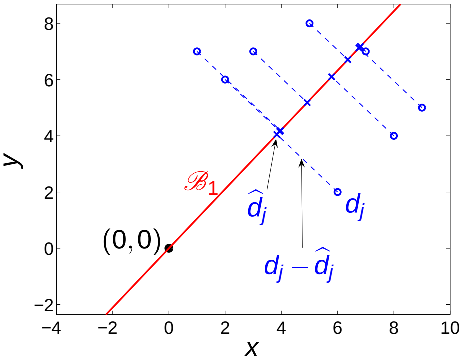
- misfit vs latency

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

- algebraic $\leftrightarrow$ LS $\leftrightarrow$ ARMAX

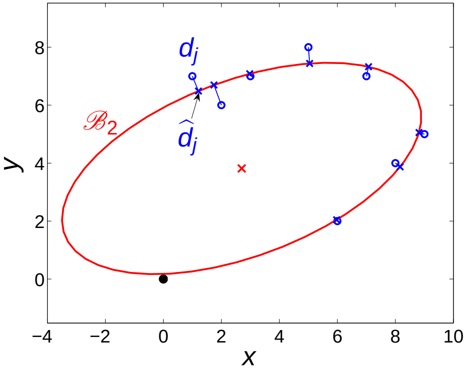
- geometric $\leftrightarrow$ TLS/PCA $\leftrightarrow$ EIV SYSID

Example: geometric distance to a linear model



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

Example: geometric distance to a quadratic model



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

Kernel representation in 2D

$$r(d) = \sum_{k=1}^{q_{\text{ext}}} \theta_k \phi_k(d) = \theta \phi(d)$$

linear in θ
nonlinear in d

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

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

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

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

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 $\mathscr{B}$

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

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

$$\sum_{j=1}^{N} \|r_\theta(d_j)\|_2^2 = \left\| \theta \begin{bmatrix} \phi(d_1) & \cdots & \phi(d_N) \end{bmatrix} \right\|_2^2$$

$$= \theta \Phi(\mathscr{D}) \Phi^\top(\mathscr{D}) \theta^\top = \theta \Psi(\mathscr{D}) \theta^\top$$

algebraic curve fitting is eigenvalue problem

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

or, equivalently, (unstructured) low rank approximation problem

$$\begin{aligned} &\text{minimize} && \text{over } \widehat{\Phi} \quad \|\Phi(\mathscr{D}) - \widehat{\Phi}\|_{\mathrm{F}} \\ &\text{subject to} && \operatorname{rank}(\widehat{\Phi}) \leq q_{\mathrm{ext}} - 1 \end{aligned}$$

Geometric distance

$$\text{minimize} \quad \text{over } \widehat{\mathscr{D}} \subset \mathscr{B} \quad \Big\| \underbrace{\begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}}_{D} - \underbrace{\begin{bmatrix} \widehat{d}_1 & \cdots & \widehat{d}_N \end{bmatrix}}_{\widehat{D}} \Big\|_{\mathrm{F}}$$

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

$$\begin{aligned} \widehat{\mathscr{D}} \subset \mathscr{B} \quad &\iff \quad \widehat{d}_j \in \mathscr{B}, \quad \text{for } j = 1, \ldots, N \\ &\iff \quad \theta\phi(\widehat{d}_j) = 0, \quad \text{for } j = 1, \ldots, N \\ &\iff \quad \theta\Phi(\widehat{\mathscr{D}}) = 0 \end{aligned}$$

the problem of computing the geometric distance is:

$$\text{minimize} \quad \text{over } \widehat{\mathscr{D}} \quad \|D - \widehat{D}\|_{\mathrm{F}} \quad \text{subject to} \quad \theta\Phi(\widehat{\mathscr{D}}) = 0$$

Geometric curve fitting

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

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

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

geometric curve fitting is nonlinearly structured low rank approx.:

$$\begin{aligned} &\text{minimize} \quad \text{over } \widehat{\mathscr{D}} \text{ and } \theta \quad \|D - \widehat{D}\| \\ &\text{subject to} \quad \text{rank}\left(\Phi(\widehat{\mathscr{D}})\right) \leq q_{\text{ext}} - 1 \end{aligned}$$

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

Bias corrected low rank approximation

assume that $\mathscr{D}$ is generated by the errors-in-variables model

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

- $\mathscr{B}_0$ is the "true" model
- $\mathscr{D}_0 := \{ d_{0,1}, \ldots, d_{0,N} \}$ is the true data, and
- $\widetilde{\mathscr{D}} := \{ \widetilde{d}_1, \ldots, \widetilde{d}_N \}$ is the measurement noise

the estimate obtained by the algebraic fitting method is biased

define the matrices

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

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

Hermite polynomials

the polynomials

$$h_0(x) = 1, \quad h_1(x) = x, \quad \text{and}$$
$$h_k(x) = xh_{k-1}(x) - (k-2)h_{k-2}(x), \quad \text{for } k = 2, 3, \ldots$$

have the property

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

Derivation of the correction

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

where the monomials ϕ_i are

$$\phi_k(d) = d_{1\cdot}^{n_{k1}} \cdots d_{q\cdot}^{n_{kq}}, \qquad \text{for} \quad k = 1, \ldots, q_{\text{ext}}$$

the (i,j)th element of Ψ is

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

by (EIV), $\widetilde{d}_{k\ell}$ are independent, zero mean, normally distributed then, by the property $(**)$ of the Hermite polynomials

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

has the desired property

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

the corrected matrix Ψ_c is an even polynomial in σ

$$\Psi_\mathrm{c}(\sigma^2) = \Psi_{\mathrm{c},0} + \sigma^2 \Psi_{\mathrm{c},1} + \cdots + \sigma^{2n_\psi} \Psi_{\mathrm{c},n_\psi}$$

the estimate $\widehat{\theta}$ is in the null space of $\Psi_\mathrm{c}(\sigma^2)$, *i.e.*, $\Psi_\mathrm{c}(\sigma^2)\widehat{\theta} = 0$

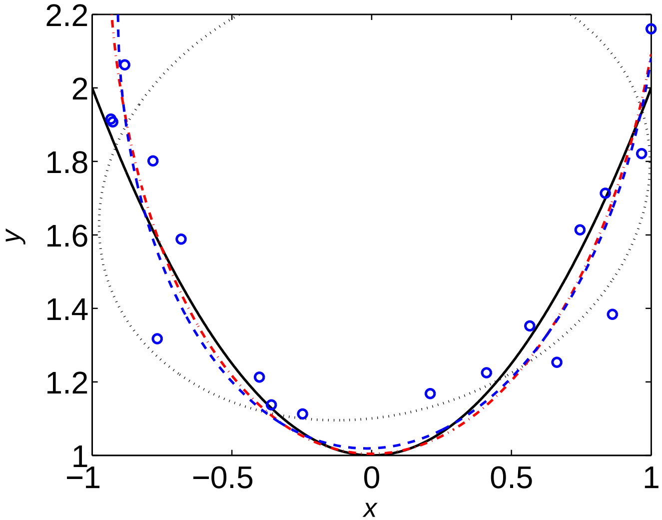
computing simultaneously σ and θ is a polynomial EVP

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

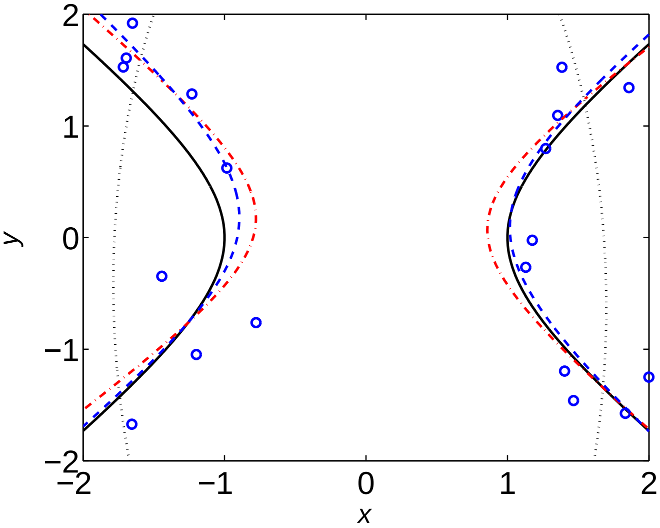
Simulation setup: $q = 2$, $p = 1$

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

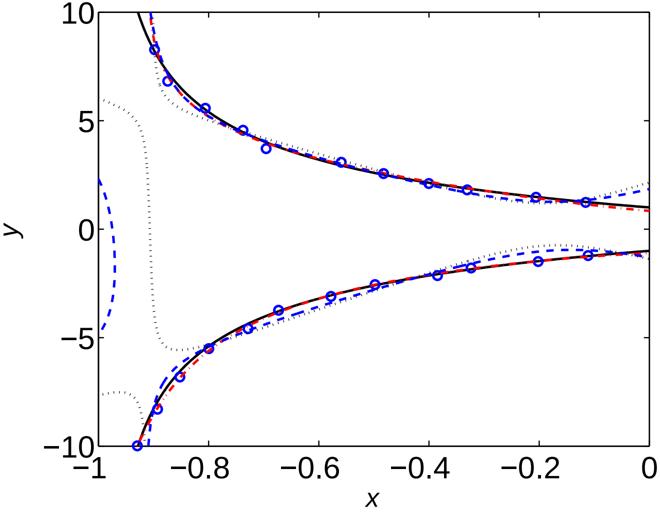
Parabola $y = x^2 + 1$



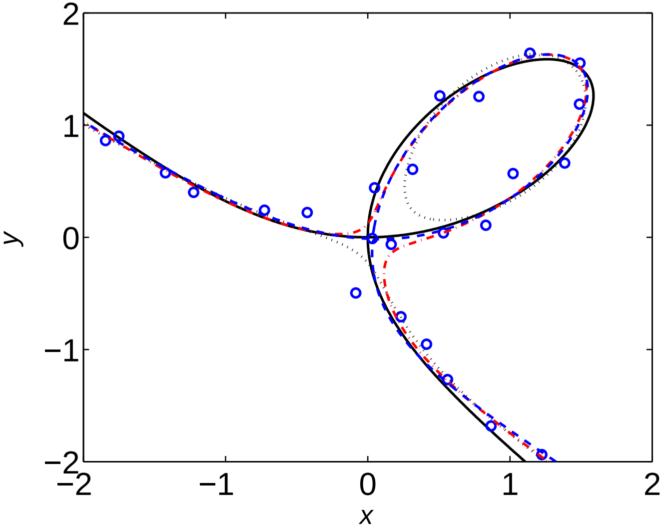
Hyperbola $x^2 - y^2 - 1 = 0$



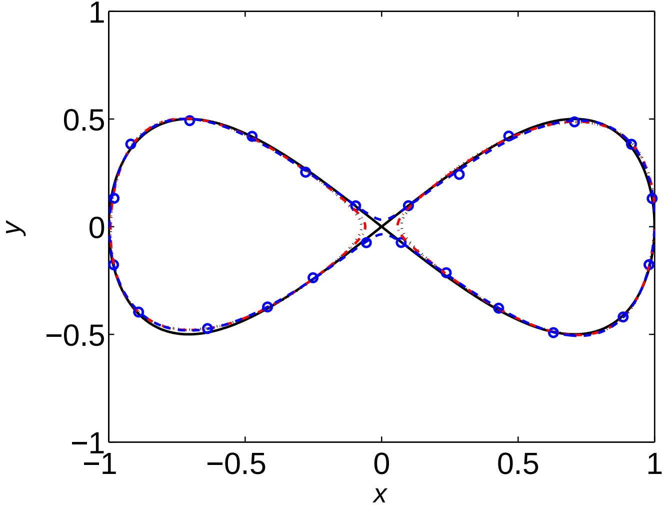
Cissoid $y^2(1+x) = (1-x)^3$



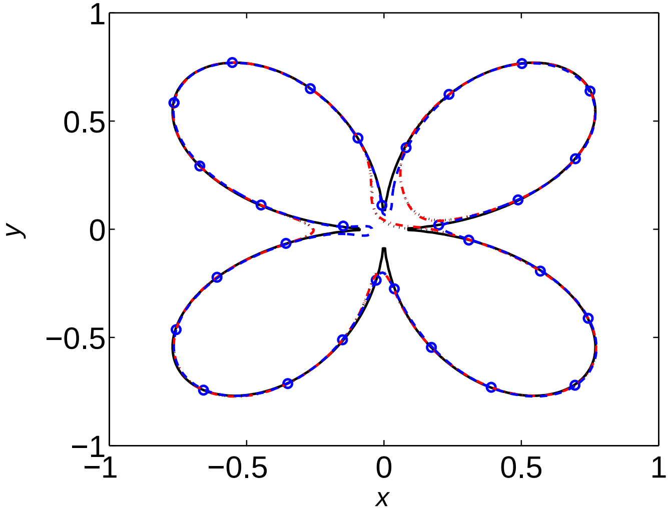
Folium of Descartes $x^3 + y^3 - 3xy = 0$



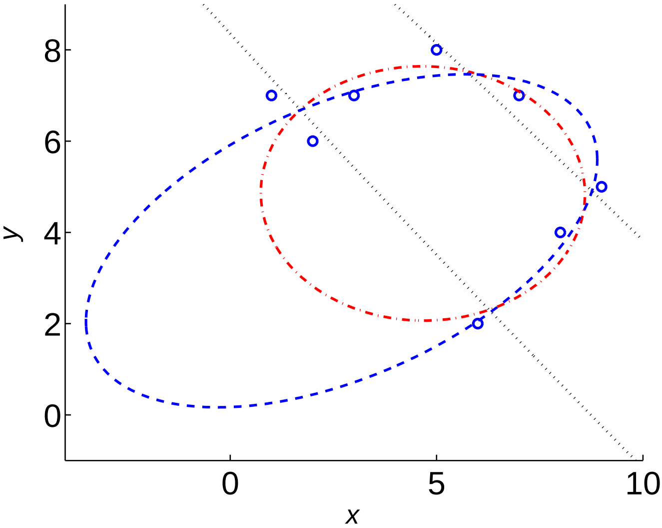
Eight curve $y^2 - x^2 + x^4 = 0$



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



“Special data” example



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

To-do list:

- Robust and efficient optimization 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?