Low-rank approximation and its applications

Ivan Markovsky

University of Southampton

Exact line fitting

$$\text{the points } w_i = (x_i, y_i),\ i = 1, \dots, N \text{ lie on a line} \qquad (*)$$

$$\Updownarrow$$

$$\text{there is } (a,b,c) \neq 0, \text{ such that} \quad ax_i + by_i + c = 0, \text{ for } i = 1, \dots, N$$

$$\Updownarrow$$

$$\text{there is } (a,b,c) \neq 0, \text{ such that} \quad \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} = 0$$

$$\Updownarrow$$

$$\operatorname{rank}\left(\begin{bmatrix} x_1 & \cdots & x_N \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \leq 2 \qquad (**)$$

- restatement of problem $(*)$ as an equivalent problem $(**)$
- however, $(**)$ is a standard problem
- the solution generalizes to
  1. multivariable data (points in $\mathbb{R}^q$) fitting by an affine set
  2. time-series fitting by linear time-invariant dynamical models
  3. data fitting by nonlinear models

Exact conic section fitting

$$\text{the points } w_i = (x_i, y_i),\ i = 1, \dots, N \text{ lie on a conic section}$$

$$\Updownarrow$$

$$\text{there are } A = A^\top,\ b,\ c, \text{ at least one of them nonzero, such that}$$

$$w_i^\top A w_i + b^\top w_i + c = 0, \text{ for } i = 1, \dots, N$$

$$\Updownarrow$$

$$\text{there is } (a_{11}, a_{12}, a_{22}, b_1, b_2, c) \neq 0, \text{ such that}$$

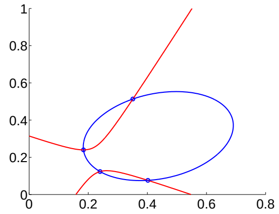
$$\begin{bmatrix} a_{11} & 2a_{12} & b_1 & a_{22} & b_2 & c \end{bmatrix} \begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} = 0$$

the points $w_i = (x_i, y_i)$, $i = 1, \dots, N$ lie on a conic section

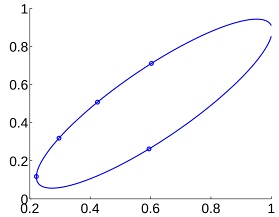
$$\Updownarrow$$

$$\operatorname{rank}\left(\begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix}\right) \leq 5$$

- $N < 4 \rightsquigarrow$ nonunique fit



- $N = 4$ (different points) $\rightsquigarrow$ unique fit



- $N > 4 \rightsquigarrow$ generically no conic section fits the data exactly

Exact fitting by linear homogeneous recurrence relations with constant coefficients

the sequence $w = (w_1, \dots, w_T)$ is generated by linear recurrence relations with lag $\leq \ell$

$$\Updownarrow$$

there is $a = (a_0, a_1, \dots, a_\ell) \neq 0$, such that

$$a_0 w_i + a_1 w_{i+1} + \dots + a_\ell w_{i+\ell} = 0, \text{ for } i = 1, \dots, T - \ell$$

$$\Updownarrow$$

there is $a = (a_0, a_1, \dots, a_\ell) \neq 0$, such that

$$a^\top \begin{bmatrix} w_1 & w_2 & \cdots & w_{T-\ell} \\ w_2 & w_3 & \cdots & w_{T-\ell+1} \\ \vdots & \vdots & & \vdots \\ w_{\ell+1} & w_{\ell+2} & \cdots & w_T \end{bmatrix} = a^\top \mathscr{H}_\ell(w) = 0$$

the sequence $w = (w_1, \dots, w_T)$ is a linear recursion with lag $\leq \ell$

$$\Updownarrow$$

$$\operatorname{rank}\left(\begin{bmatrix} w_1 & w_2 & \cdots & w_{T-\ell} \\ w_2 & w_3 & \cdots & w_{T-\ell+1} \\ \vdots & \vdots & & \vdots \\ w_{\ell+1} & w_{\ell+2} & \cdots & w_T \end{bmatrix}\right) \leq \ell$$

- $T \leq 2\ell \rightsquigarrow$ there is exact fit (independent of w)
- $T > 2\ell \rightsquigarrow$ generically there is no exact fit

Existence of greatest common divisor

the polynomials p and q have a GCD of degree $\geq \ell$

$\Updownarrow$

$\cdots$

$\Updownarrow$

rank(Sylvester matrix of p and q) $\leq m+n-\ell$

Data, model, and model class

	line fitting	conic section fitting	linear recurrence with lag $\leq \ell$	GCD
data	points (in $\mathbb{R}^2$)	points (in $\mathbb{R}^2$)	sequence	pair of polynomials
model	line (in $\mathbb{R}^2$)	conic section	autonomous LTI system	polynomials with nontrivial GCD
model class	$\{$lines (in $\mathbb{R}^2$)$\}$	$\{$conic sections$\}$	class of LTI systems	?

Continue the sequences

$(\quad 1, \; 2, \quad 3, \quad 5, \; 8, \; 13, \; ? \;)$

$(\; -5, \; 5, \quad 0, \quad 5, \; 5, \; 10, \; ? \;)$

$(\quad 1, \; 0, \; -1, \; -1, \; 0, \quad 1, \; ? \;)$

An algorithm for continuation of a sequence

Input: $w = (w_1, \ldots, w_T)$

1: $\ell := 1$

2: **while** rank$\big(\mathscr{H}_\ell(w)\big) = \ell + 1$ **do**

3: $\quad \ell := \ell + 1$

4: **end while**

5: compute nonzero vector a in the left null space of $\mathscr{H}_\ell(w)$

Output: $w_{T+1} = -\dfrac{1}{a_\ell}(a_0 w_{T-\ell+1} + a_1 w_{T-\ell+2} + \cdots + a_{\ell-1} w_T)$

Abstract setting for data modeling

- data space $\mathscr{U}$

  examples: $\mathbb{R}^q$, $(\mathbb{R}^q)^T$, $\mathbb{R}[z] \times \mathbb{R}[z]$, $\{\,\text{true}, \text{false}\,\}$

- data $\mathscr{D} = \{\,\mathscr{D}_1, \dots, \mathscr{D}_N\,\} \subset \mathscr{U}$

  $\mathscr{D}_i \in \mathscr{U}$ — observation, relalization, or outcome

- model $\mathscr{B} \subset \mathscr{U}$

  an exclusion rule, declares what outcomes are possible

- model class $\mathscr{M} \subset 2^{\mathscr{U}}$

Exact vs approximate models

- $\mathscr{B}$ is an exact model for $\mathscr{D}$ if $\mathscr{D} \subset \mathscr{B}$

  otherwise $\mathscr{B}$ is an approximate model for $\mathscr{D}$

- $\mathscr{B} = \mathscr{U}$ is a (trivial) exact model for any $\mathscr{D} \subset \mathscr{U}$

  $\rightsquigarrow$ we want nontrivial model

  $\rightsquigarrow$ notion of model complexity

- any model is approximate model for any data set

  $\rightsquigarrow$ we need to quantify the approximation accuracy

  $\rightsquigarrow$ notion of model accuracy (w.r.t. the data)

Model complexity

- the "smaller" a model is the more powerful/useful it is
- the "bigger" a model is the more complex it is
- we prefer simple models over complex ones
- exact modeling problem:

  find the least complex model that fits the data exactly

Linear model complexity

- a linear model $\mathscr{B}$ is a subspace of $\mathscr{U}$ ($\mathscr{U}$ is a vector space)
- the complexity of $\mathscr{B}$ is defined as its dimension
- in the linear case

$$\mathscr{D} \subset \mathscr{B} \qquad \Longrightarrow \qquad \text{span}(\mathscr{D}) \subset \mathscr{B}$$

  and the rank of the data matrix is $\leq \dim(\mathscr{B})$

- $\text{span}(\mathscr{D})$ — the smallest linear model, consistent with $\mathscr{D}$

Model accuracy

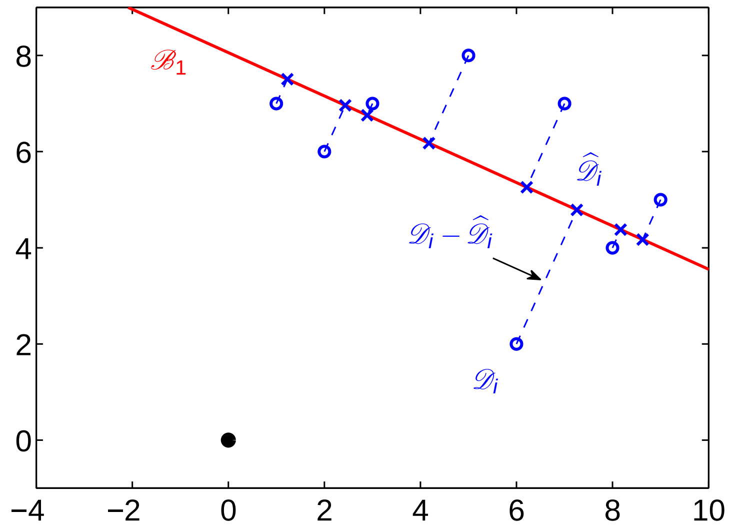
- let $\mathscr{U}$ be a normed vector space with norm $\|\cdot\|$
- the distance between the data $\mathscr{D}$ and a model $\mathscr{B}$ is

$$\operatorname{dist}(\mathscr{D},\mathscr{B}) := \min_{\widehat{\mathscr{D}}\subset\mathscr{B}} \|\mathscr{D}-\widehat{\mathscr{D}}\|$$

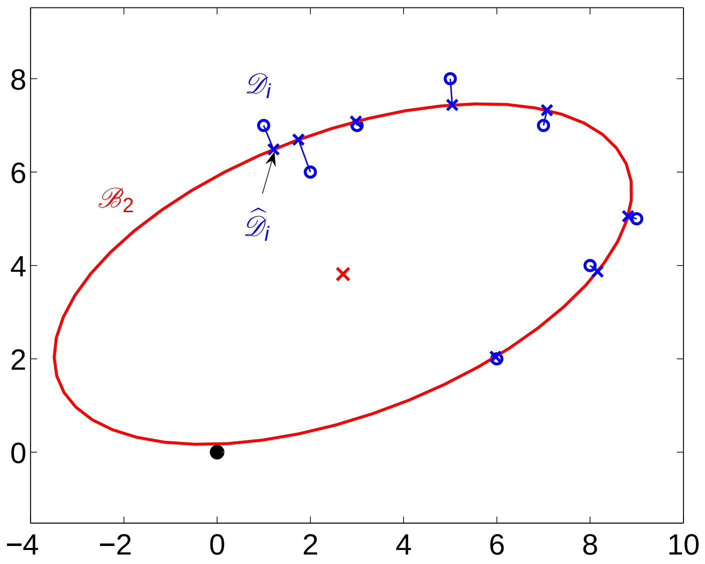
- approximate modeling problem:

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

Example: $\mathscr{U}=\mathbb{R}^2$, Euclidean norm, $\mathscr{M}=\{\text{lines}\}$



Example: $\mathscr{U}=\mathbb{R}^2$, Euclidean norm, $\mathscr{M}=\{\text{ellipses}\}$



Approximate modeling $\iff$ Low-rank approximation

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

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } \widehat{\mathscr{D}}\subset\mathscr{B}\in\mathscr{M} \quad \|\mathscr{D}-\widehat{\mathscr{D}}\|$$

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } \widehat{\mathscr{D}} \quad \|\mathscr{D}-\widehat{\mathscr{D}}\| \quad \text{subject to}$$
$$\text{rank}\big(\text{matrix constructed of } \widehat{\mathscr{D}}\big) \le \text{complexity of } \mathscr{M}$$

- in general, nonconvex optimization problem
- different matrix structures occur in the applications

Approximate line fitting in $\mathbb{R}^2$

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

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } \widehat{x}_i, \widehat{y}_i, i = 1, \ldots, N \quad \sum_{i=1}^{N} \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} \widehat{x}_i \\ \widehat{y}_i \end{bmatrix} \right\|_2^2$$

$$\text{subject to} \quad \text{rank} \left(\begin{bmatrix} \widehat{x}_1 & \cdots & \widehat{x}_N \\ \widehat{y}_1 & \cdots & \widehat{y}_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \leq 2$$

can be solved using the singular value decomposition

Approximate conic section fitting in $\mathbb{R}^2$

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

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } \widehat{x}_i, \widehat{y}_i, i = 1, \ldots, N \quad \sum_{i=1}^{N} \left\| \begin{bmatrix} x_i \\ y_i \end{bmatrix} - \begin{bmatrix} \widehat{x}_i \\ \widehat{y}_i \end{bmatrix} \right\|_2^2$$

$$\text{subject to} \quad \text{rank} \left(\begin{bmatrix} \widehat{x}_1^2 & \cdots & \widehat{x}_N^2 \\ \widehat{x}_1\widehat{y}_1 & \cdots & \widehat{x}_N\widehat{y}_N \\ \widehat{x}_1 & \cdots & \widehat{x}_N \\ \widehat{y}_1^2 & \cdots & \widehat{y}_N^2 \\ \widehat{y}_1 & \cdots & \widehat{y}_N \\ 1 & \cdots & 1 \end{bmatrix} \right) \leq 5$$

requires iterative optimization methods and initial approx.

Conclusions

- common pattern in data modeling

  data is exact for a model of bounded complexity

  $$\Updownarrow$$

  matrix constructed from the data is rank deficient

- exact modeling $\approx$ rank computation
- approximate modeling problem

  modify the data as little as possible in order to
  make it exact for a model in the given model class

- the problem is equivalent to low-rank approximation

Conclusions

- seemingly unrelated problems (line, ellipse, dynamic model fitting, GCD computation) are reduced to one problem
- theory, algorithms, and related software for the generic problem have impact on many applied problems
- postdoc position available from January 2011

Questions?