DYSCO course on low-rank approximation and its applications

Behavioral approach

Ivan Markovsky

Vrije Universiteit Brussel

Plan

1. Introduction
2. Computational tools
3. Behavioral approach
4. System identification
5. Subspace methods
6. Generalizations

Outline

From $Ax = b$ to low-rank approximation

Low-rank $\leftrightarrow$ behavioral approach

Exercises

Linear static models representations

Linear time-invariant model representations

Solution methods

Outline

From $Ax = b$ to low-rank approximation

Low-rank $\leftrightarrow$ behavioral approach

Exercises

Linear static models representations

Linear time-invariant model representations

Solution methods

Line fitting $\overset{?}{\leftrightarrow}$ $Ax \approx b$

- classic problem: fit the points

$$\mathscr{D} := \{ d_1, \dots, d_N \} \in \mathbb{R}^2$$

by a line passing through the origin

- classic solution: find approx. solution $x \in \mathbb{R}$ of

$$\begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix} x = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}$$

where $d_i = (a_i, b_i)$

- the fitting line is

$$\mathscr{B}(x) := \{ (a,b) \in \mathbb{R}^2 \mid ax = b \} \qquad (*)$$

Line fitting $\overset{?}{\leftrightarrow}$ $Ax \approx b$

- another solution: find approx. solution $x' \in \mathbb{R}$ of

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} x' = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}$$

- the fitting line is

$$\mathscr{B}(x') := \{ (a, b) \in \mathbb{R}^2 \mid a = bx' \} \qquad (**)$$

- if exact solution exists, $x' = 1/x$, unless $x = 0$

- exceptions:
  - horizontal line ($x = 0$)
  - vertical line ($x' = 0$)

Approximation criteria

- solving approximately the overdetermined systems

$$Ax = B \quad \text{and} \quad A = Bx'$$

  where

$$A := \operatorname{col}(a_1, \ldots, a_{10}) \quad \text{and} \quad B := \operatorname{col}(b_1, \ldots, b_{10})$$

  in the least squares sense $\rightsquigarrow$ different solutions

- $\Longrightarrow$ the data modeling criterion depends on the (arbitrary) choice of model representation

Input/output representation

- $Ax = b$ and $A = bx'$ define lines in $\mathbb{R}^2$

$$\mathscr{B}(x) := \{\, (a, b) \in \mathbb{R}^2 \mid ax = b \,\} \qquad (*)$$

$$\mathscr{B}(x') := \{\, (a, b) \in \mathbb{R}^2 \mid a = bx' \,\} \qquad (**)$$

- $(*)$ and $(**)$ define a line $\mathscr{B}$ by functions
  - $a \mapsto b$ in $(*)$
  - $b \mapsto a$ in $(**)$

- input/output representations
  - in $(*)$, a is input, b is output   (a causes b)
  - in $(**)$, b is input, a is output   (b causes a)

Input/output representation

- we are interested in $\mathscr{M}$ — set of lines through 0
- separately $(*)$, $(**)$ don't represent all lines through 0
- I/O representation: any $\mathscr{B} \in \mathscr{M}$ is representable as

$$\mathscr{B} = \mathscr{B}(x, \Pi) = \{ \Pi \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] \mid ax = b \}$$

  for some $x \in \mathbb{R}$ and a permutation matrix Π

- link to system of linear equations

$$\mathscr{D} \subset \mathscr{B}(x, \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right]) \quad \iff \quad Ax = B$$

$$\mathscr{D} \subset \mathscr{B}(x, \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right]) \quad \iff \quad Bx' = A$$

Kernel representation

- any $\mathscr{B} \in \mathscr{M}$ can be represented as

$$\mathscr{B} = \ker(R) := \{\, d \in \mathbb{R}^2 \mid Rd = R_1 a + R_2 b = 0 \,\}$$

  for some nonzero vector $R \in \mathbb{R}^{1\times 2}$

- $Rd = 0$ defines a relation between a and b

- $\mathscr{D} \subset \ker(R)$ implies that

$$R \underbrace{\begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}}_{D} = 0$$

Image representation

- any $\mathscr{B} \in \mathscr{M}$ can be represented as

$$\mathscr{B} = \text{image}(P) := \{\, d = P\ell \mid \ell \in \mathbb{R} \,\}$$

for some vector $P \in \mathbb{R}^{2\times 1}$

- $\mathscr{D} \subset \text{image}(P)$ implies that

$$\begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix} = PL$$

for some $L \in \mathbb{R}^{1\times N}$

Summary

- common feature of the representations considered

$$\left.\begin{array}{llcl} \exists\ x \in \mathbb{R} & Ax = B & \Longrightarrow \\ \exists\ x' \in \mathbb{R} & Bx' = A & \Longrightarrow \\ \exists\ x \in \mathbb{R}, \Pi \text{ permut.} & \begin{bmatrix} x & -1 \end{bmatrix} \Pi D = 0 & \Longleftrightarrow \\ \exists\ R \in \mathbb{R}^{1\times 2}, R \neq 0 & RD = 0 & \Longleftrightarrow \\ \exists\ P \in \mathbb{R}^{2\times 1}, L \in \mathbb{R}^{1\times N} & D = PL & \Longleftrightarrow \end{array}\right\} \text{rank}(D) = 1$$

- representation free characterization of the exact data

$$\mathscr{D} \subset \mathscr{B} \text{ and}$$

$$\mathscr{B} \text{ is a line through } 0$$

$$\Updownarrow$$

$$\text{rank}(D) = 1$$

Low-rank approximation

- representation free formulation

- exact modeling problem:

$$\exists \text{ exact model for } \mathscr{D} \iff D \text{ is rank deficient}$$

- approximate modeling problem:

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{\mathscr{D}} \quad \text{error}(\mathscr{D}, \widehat{\mathscr{D}}) \\ \text{subject to} & \exists \text{ exact model for } \widehat{\mathscr{D}} \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{D} \quad \text{error}(D, \widehat{D}) \\ \text{subject to} & \widehat{D} \text{ is rank deficient} \end{array}$$

Generalizations

1. multivariable data fitting $\mathscr{U} = \mathbb{R}^q$
   - linear static model $\leftrightarrow$ subspace
   - model complexity $\leftrightarrow$ subspace dimension
   - rank(D) $\leftrightarrow$ upper bound on the model complexity

2. nonlinear static modeling
   - $\mathscr{D} \mapsto D$ — nonlinear function
   - nonlinearly structured low-rank approximation

3. linear time-invariant dynamical models
   - $\mathscr{D} \mapsto$ Hankel matrix D
   - Hankel structured low-rank approximation

4. nonlinear dynamic (2. + 3.)

Structure $\mathscr{S} \leftrightarrow$ Model class $\mathscr{M}$

unstructured	linear static
Hankel	scalar LTI
$q \times 1$ Hankel	q-variate LTI
$q \times N$ Hankel	N equal length traj.
mosaic Hankel	N general traj.
$\begin{bmatrix}\text{Hankel} & \text{unstructured}\end{bmatrix}$	finite impulse response
block-Hankel Hankel-block	2D linear shift-invariant

Related frameworks

- behavioral approach in systems and control theory
  - representation free: model $=$ set (the behavior)
  - no a priori separation of inputs and outputs
- errors-in-variables modeling
  - all variables are perturbed by noise
  - maximum likelihood estimation $\leftrightarrow$ LRA
- principal component analysis
  - another statistical setting for LRA
- factor analysis
  - factors $\leftrightarrow$ latent variables in image repr.

Outline

From $Ax = b$ to low-rank approximation

Low-rank $\leftrightarrow$ behavioral approach

Exercises

Linear static models representations

Linear time-invariant model representations

Solution methods

Behavioral approach

- Jan C. Willems (1939–2013)

Outline

Outline

Linear static model

"Good definition should formalize sensible intuition."
Jan Willems

- linear static model with q variables $=$ subspace of $\mathbb{R}^q$
- model complexity $\quad\leftrightarrow\quad$ subspace dimension
  (the more the model can fit, the less useful it is)
- linear static models with complexity at most $\mathtt{m}$ — $\mathscr{L}^q_{\mathtt{m},0}$
- any $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},0}$ admits kernel, image, and input/output representations

Representations

- kernel representation with parameter $R \in \mathbb{R}^{\mathtt{p} \times q}$

$$\ker(R) := \{\, d \mid Rd = 0 \,\}$$

- image representation with parameter $P \in \mathbb{R}^{q \times \mathtt{m}}$

$$\operatorname{image}(P) := \{\, d = P\ell \mid \ell \in \mathbb{R}^{\mathtt{m}} \,\}$$

- input/output representation

$$\mathscr{B}_{\mathrm{i/o}}(X, \Pi) := \{\, d = \Pi \begin{bmatrix} u \\ y \end{bmatrix} \mid u \in \mathbb{R}^{\mathtt{m}},\ y = X^\top u \,\}$$

with parameters $X \in \mathbb{R}^{\mathtt{m} \times \mathtt{p}}$ and permutation matrix Π

Nonuniqueness of the parameters

- columns of P are generators of the model $\mathscr{B}$
- rows of R are annihilators of $\mathscr{B}$
- the parameters R and P are not unique due to
  - addition of linearly dependent generators/annihilators
  - change of basis transformation

$$\ker(R) = \ker(UR), \qquad \text{for all } U \in \mathbb{R}^{\mathrm{p}\times\mathrm{p}}, \det(U) \neq 0$$

$$\operatorname{image}(P) = \operatorname{image}(PV), \qquad \text{for all } V \in \mathbb{R}^{\mathrm{m}\times\mathrm{m}}, \det(V) \neq 0$$

- the smallest number of generators $\mathrm{m} := \dim(\mathscr{B})$
- max. number of annihilators $\mathrm{p} := q - \dim(\mathscr{B})$

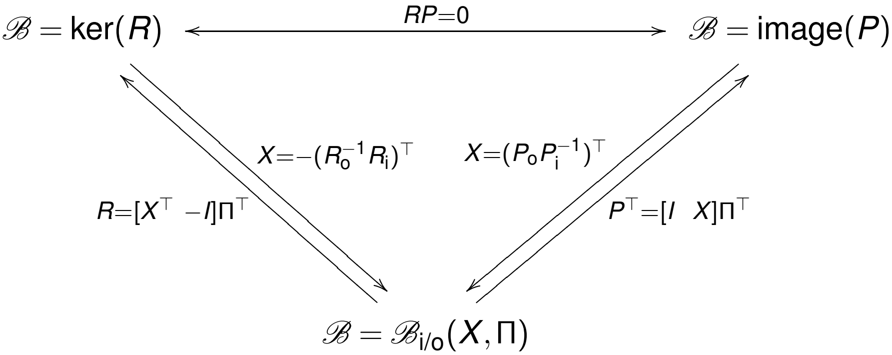
Inputs and outputs

- input is a "free" variable

$$\Pi\,\mathrm{col}(u,y) \in \mathscr{B} \text{ and } u \text{ input} \quad \Longleftrightarrow \quad u \in \mathbb{R}^{\mathtt{m}}$$

- output is bound by input and model
- fact: $\mathtt{m} := \dim(\mathscr{B})$ — number of inputs
- $\mathtt{p} := q - \mathtt{m}$ — number of outputs
- generically any I/O partition is possible
- choosing a partition amounts to choosing full rank $\mathtt{p} \times \mathtt{p}$ submatrix of R or full rank $\mathtt{m} \times \mathtt{m}$ submatrix of P

Transition among representations



$$\Pi^\top P =: \begin{bmatrix} P_i \\ P_o \end{bmatrix} \begin{matrix} \mathtt{m} \\ \mathtt{p} \end{matrix} \quad \text{and} \quad R\Pi =: \begin{matrix} \mathtt{m} & \mathtt{p} \\ [R_i & R_o] \end{matrix}$$

(for details, see Section 2.1)

Outline

From $Ax = b$ to low-rank approximation

Low-rank $\leftrightarrow$ behavioral approach

Exercises

Linear static models representations

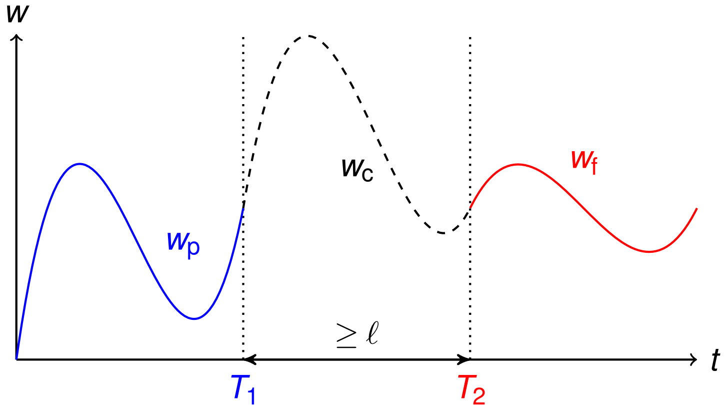
Linear time-invariant model representations

Solution methods

Dynamical models

- observations are trajectories: functions $\mathscr{T} \mapsto \mathbb{R}^q$
- universal set: $(\mathbb{R}^q)^{\mathscr{T}}$ — set of functions
- the time axis $\mathscr{T}$ is $\mathbb{Z}$ (discrete) or $\mathbb{R}$ (continuous)
- dynamic model $\mathscr{B}$ is a subset of $(\mathbb{R}^q)^{\mathscr{T}}$
- linearity: $w, v \in \mathscr{B} \implies \alpha w + \beta v \in \mathscr{B}, \; \forall\, \alpha, \beta$
- shift operator: $(\sigma^\tau w)(t) := w(t+\tau)$, for all $t \in \mathscr{T}$
- time-invariance: $\sigma^\tau \mathscr{B} = \mathscr{B}$, for all $\tau \in \mathscr{T}$

Controllability



for all w_p, $w_f \in \mathscr{B}$, $\exists$ w_c, such that $w_p \wedge w_c \wedge w_f \in \mathscr{B}$

Complexity of an LTI model

- static model $\mathscr{B} \in \mathscr{L}^q_{\mathtt{m},0}$ — complexity $= \mathtt{m}$
  (increasing $\mathtt{m}$ requires increasing # of var. q)

- LTI dynamic model has two aspects:
  - multivariable — number of inputs $\mathtt{m}$
  - dynamics — time memory span ℓ

- complexity of an LTI model is ordered pair $(\mathtt{m}, \ell)$

- notation:
  - $\mathscr{L}^q$ — all LTI models with q variables
  - $\mathscr{L}^q_{\mathtt{m}}$ — at most $\mathtt{m}$ inputs
  - $\mathscr{L}^q_{\mathtt{m},\ell}$ — complexity bounded by $(\mathtt{m}, \ell)$

Restriction of the behavior on an interval

- $w_{\rm p} \wedge w_{\rm f}$ — concatenation of $w_{\rm p}$ and $w_{\rm f}$

$$\mathscr{B}|_T := \{\, w \in (\mathbb{R}^q)^T \mid \exists\, w_{\rm p},\ w_{\rm f},\ \text{such that } w_{\rm p} \wedge w \wedge w_{\rm f} \in \mathscr{B} \,\}$$

- for $\mathscr{B} \in \mathscr{L}^q$ and $T > 0$, $\mathscr{B}|_T$ is a subspace

$$\dim(\mathscr{B}|_T) \;\leq\; T\mathtt{m} + \mathtt{p}\ell$$

- complexity of $\mathscr{B} \sim \dim(\mathscr{B}|_T)$
- therefore, $(\mathtt{m}, \ell)$ specifies the complexity

Representations

- kernel representation with par. $R(z) \in \mathbb{R}^{g \times q}[z]$

$$\ker(R) = \{\, w \mid R(\sigma)w = R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \,\}$$

- image representation with par. $P(z) \in \mathbb{R}^{q \times g}[z]$

$$\operatorname{image}(P) = \{\, w = P(\sigma)v \mid \text{for some } v \,\}$$

- input/state/output representation

$$\begin{aligned}\mathscr{B}(A,B,C,D,\Pi) := \{\, & w = \Pi \operatorname{col}(u,y) \mid \\ & \exists\, x,\ \text{such that } \sigma x = Ax + Bu \text{ and } y = Cx + Du \,\}\end{aligned}$$

(default $\Pi = I$, in which case it is skipped)

- any $\mathscr{B} \in \mathscr{L}^q$ admits kernel and I/S/O representations
- any controllable $\mathscr{B} \in \mathscr{L}^q$ admits image representation
- lag of $\mathscr{B}$ — minimal ℓ, for which kernel repr. exists
- minimal $\text{row dim}(R) =$ number of outputs
- minimal $\text{col dim}(P) =$ number of inputs

(for details, see Section 2.2)

Nonuniqueness of I/S/O representation

- choice of an input/output partition
- redundant states (nonminimality of representation)
- change of state space basis

$$\mathscr{B}(A, B, C, D) = \mathscr{B}(T^{-1}AT, T^{-1}B, CT, D),$$

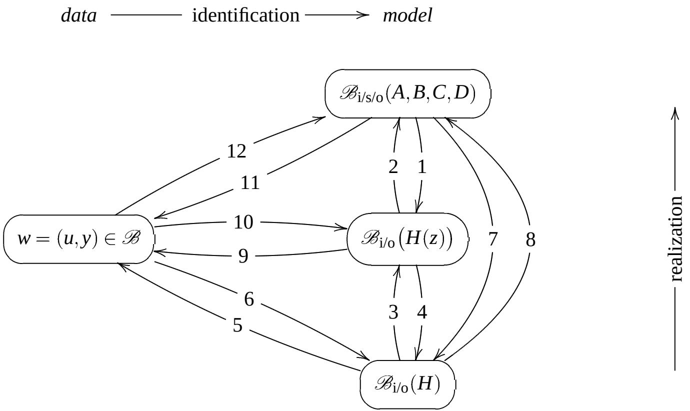
for any nonsingular matrix $T \in \mathbb{R}^{\mathrm{n} \times \mathrm{n}}$

- minimal representation $\implies$ smallest n = order of $\mathscr{B}$

Transition among representations

- using different representations is a powerful idea
- problems are trivial, given suitable representations
- *cf.*, matrix factorizations in numerical linear algebra
- the problem becomes to transform representations

Links among I/O model representations



1. $H(z)=C(Iz-A)^{-1}B+D$
2. realization of a transfer function
3. Z or Laplace transform of $H(t)$
4. inverse transform of $H(z)$
5. convolution $y_{\text{d}}=H\star u_{\text{d}}$
6. exact identification
7. $H(0)=D$, $H(t)=CA^{t-1}B$ (discrete-time), $H(t)=Ce^{At}B$ (continuous-time), for $t>0$
8. realization of an impulse response
9. simulation with input u_{d} and $x(0)=0$
10. exact identification
11. simulation with input u_{d} and $x(0)=x_{\text{ini}}$
12. exact identification

Outline

Problems with analytic solutions

- unstructured, unweighted $\qquad (\|\cdot\|_F := \|\mathrm{vec}(\cdot)\|_2)$

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{D} \quad \|D - \widehat{D}\|_F \\ \text{subject to} & \operatorname{rank}(\widehat{D}) \leq r \end{array} \qquad \text{(LRA)}$$

- unstructured, with left/right weighting matrices

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{D} \quad \|W_l(D - \widehat{D})W_r\|_F \\ \text{subject to} & \operatorname{rank}(\widehat{D}) \leq r \end{array}$$

- circulant structure

Truncated SVD

Theorem

Let $D = U\Sigma V^\top$ be the (thin) SVD of $D \in \mathbb{R}^{q\times N}$ and define

$$U =: \begin{matrix} & r & q-r & \\ & [U_1 & U_2] & q \end{matrix}, \quad \Sigma =: \begin{matrix} & r & q-r & \\ & \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} & \begin{matrix} r \\ q-r \end{matrix} \end{matrix}, \quad V =: \begin{matrix} & r & q-r & \\ & [V_1 & V_2] & N \end{matrix}$$

An optimal low-rank approximation (a solution of (LRA)) is

$$\widehat{D}^* = U_1\Sigma_1 V_1^\top, \qquad \widehat{\mathscr{B}}^* = \ker(U_2^\top) = \operatorname{image}(U_1).$$

It is unique if and only if $\sigma_r \neq \sigma_{r+1}$.

- $\widehat{\mathscr{B}}^*$ depends only on the left singular vectors

- in general
  - structures other than circular
  - norms other than 2-norm
  - weights other than "left/right" multiplication of $D - \widehat{D}$

  lead to hard non-convex optimization problems

- there are many (heuristic) solution methods

Overview of algorithms

- global solution methods
  - SDP relaxations of rational function min. problem
  - systems of polynomial equations (computer algebra)
  - branch-and-bound, simulating annealing, ...
- local optimization methods
  - variable projections
  - alternating projections
  - variations (parameterization $+$ optimization method)
- convex relaxations / multistage methods
  - subspace methods
  - nuclear norm heuristic

Summary

- linear static model $=$ subspace
- model representations
  - input/output (a function, system $AX = B$)
  - kernel (implicit function, relation)
  - image (introduces latent variables)
- representation invariant problem formulation $\rightsquigarrow$ LRA
- different representations $\rightsquigarrow$ different solution methods

". . . most linear resistors let us treat current as a function of voltage or voltage as a function of current, since R is neither zero nor infinite. But in the two limiting cases - the short circuit and the open circuit - that's not true. To fit these cases neatly in a unified framework, we shouldn't think of the relation between current and voltage as defining a function. It's just a relation!"

John Baez