Structured low-rank approximation approach to sum-of-exponentials

Ivan Markovsky

Objective: show alternative solution methods for sum-of-exponentials modeling

Model representations

Modeling algorithms

Generalizations of the problem

Model is set of signals

discrete-time sum-of-damped-exponentials model

$$\mathscr{B}_z = \left\{ \sum_{i=1}^{n} c_i \exp_{z_i} \;\middle|\; c \in \mathbb{C}^n \right\}, \qquad \exp_{z_i}(t) := z_i^t, \; t \in \mathbb{Z}$$

model complexity

$$n := \dim(\mathscr{B}_z) = \text{\# of exponents}$$

model class

$$\mathscr{L}_n := \{ \mathscr{B}_z \mid z \in \mathbb{C}^n \}$$

Model representation is equation

pole representation

$$\mathscr{B}_z = \{ \textstyle\sum_i c_i \exp_{z_i} \mid c \in \mathbb{C}^n \}$$

kernel representation

$$(\sigma y)(t) := y(t+1)$$

$$\mathscr{B}_R = \{ y \mid R_0 y + R_1 \sigma y + \cdots + R_n \sigma^n y = 0 \} =: \ker\big(R(\sigma)\big)$$

state-space representation

$$\mathscr{B}_{A,C} = \{ y \mid y = Cx,\ \sigma x = Ax \}$$

The representation parametrizes the model

representation	pole	kernel	state-space
model parameter	z	R	A, C
ini. condition	c	$y(-n+1), \dots, y(0)$	$x(0)$

given $\mathscr{B}$, z is unique, R and (A, C) are not unique

transitions among the representations are well understood

kernel and state space are more general than pole repr.
(polynomials $\times$ exponentials)

Modeling problem: find optimal model

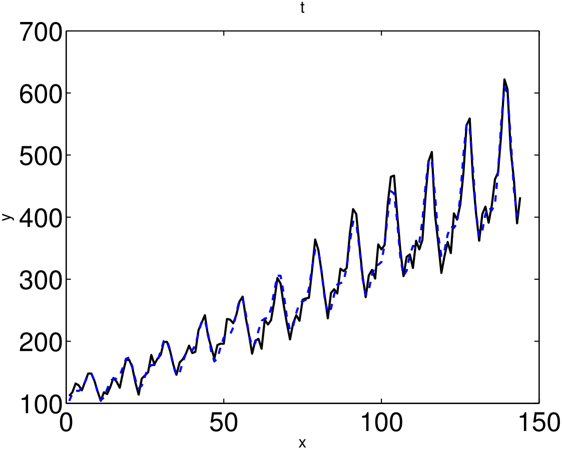
measurement error model

$$y = \bar{y} + \widetilde{y} \qquad \begin{array}{lcl} \bar{y} \in \bar{\mathscr{B}} \in \mathscr{L}_n & — & \text{true signal} \\ \widetilde{y} \sim N(0, \nu I) & — & \text{noise} \end{array}$$

maximum likelihood estimator

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{y} \text{ and } \widehat{\mathscr{B}} \quad \|y - \widehat{y}\| \\ \text{subject to} & \widehat{y} \in \widehat{\mathscr{B}} \in \mathscr{L}_n \end{array}$$

Example: airline passenger data 1949–1960



solid line — $T = 144$ data points
dashed — fit by $n = 6$ order model

Model validation problem: find optimal approximation of y in $\widehat{\mathscr{B}}$

$$\text{error}(y,\widehat{\mathscr{B}}) := \min_{\widehat{y}\in\widehat{\mathscr{B}}} \|y-\widehat{y}\|$$

likelihood of y, given $\widehat{\mathscr{B}}$

projection of y on $\widehat{\mathscr{B}}$

validation error of $\widehat{\mathscr{B}}$ on (new) data

fast algorithms: Kalman filter, displacement rank, ...

Summary

model ($\mathscr{B}_z$) $\neq$ representation ($\sum_i c_i \exp_{z_i}$)

define problems in terms of models (not representations)

- maximum likelihood estimator
- likelihood evaluation

use representation when solving the problem

Next . . .

Model representations

Modeling algorithms

Generalizations of the problem

Link to low-rank approximation

$$y \in \mathscr{B} \in \mathscr{L}_n$$

$$\Updownarrow$$

there is $R(z)$, such that $R(\sigma)y = 0$, *i.e.*,

$$R_0 y(t) + R_1 y(t+1) + \cdots + R_n y(t+n) = 0, \text{ for } t = 1, \ldots, T-n$$

$$\Updownarrow$$

there is $R = \begin{bmatrix} R_0 & R_1 & \cdots & R_n \end{bmatrix} \neq 0$, such that

$$R \begin{bmatrix} y(1) & y(2) & \cdots & y(T-n) \\ y(2) & y(3) & \cdots & y(T-n+1) \\ \vdots & \vdots & & \vdots \\ y(n+1) & y(n+2) & \cdots & y(T) \end{bmatrix} = 0$$

$y \in \mathscr{B} \in \mathscr{L}_n \iff$ rank deficient Hankel matrix

$$y \in \mathscr{B} \in \mathscr{L}_n$$

$$\Updownarrow$$

$$\operatorname{rank}\left(\begin{bmatrix} y(1) & y(2) & \cdots & y(T-n) \\ y(2) & y(3) & \cdots & y(T-n+1) \\ \vdots & \vdots & & \vdots \\ y(n+1) & y(n+2) & \cdots & y(T) \end{bmatrix}\right) \leq n$$

Hankel structured matrix

Sum-of-exponential modeling is equivalent to Hankel structured low-rank approximation

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{y} \text{ and } \widehat{\mathscr{B}} \quad \|y - \widehat{y}\| \\ \text{subject to} & \widehat{y} \in \widehat{\mathscr{B}} \in \mathscr{L}_n \end{array}$$

$$\Updownarrow$$

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{y} \quad \|y - \widehat{y}\| \\ \text{subject to} & \operatorname{rank}\big(\mathscr{H}_{n+1}(\widehat{y})\big) \leq n \end{array}$$

Three solution approaches:

nuclear norm heuristic

subspace methods

local optimization

Nuclear norm heuristic: replace rank by nuclear norm constraint

rank: number of nonzero singular values

nuclear norm $\|\cdot\|_*$: ℓ_1-norm of the singular values

minimization of the nuclear norm

- tends to increase sparsity $\implies$ reduce rank
- leads to a convex optimization problem

Nuclear norm minimization methods involve a hyper-parameter

$$\begin{array}{ll} \text{minimize} \quad \text{over } \widehat{y} & \|y - \widehat{y}\| \\ \text{subject to} & \|\mathscr{H}_{n+1}(\widehat{y})\|_* \leq \gamma \end{array}$$

$$\Updownarrow$$

$$\text{minimize} \quad \text{over } \widehat{y} \quad \alpha\|y - \widehat{y}\| + \|\mathscr{H}_{n+1}(\widehat{y})\|_*$$

γ/α — determines the rank of $\mathscr{H}_{n+1}(\widehat{y})$

we want $\alpha_{\text{opt}} = \max\{\, \alpha \mid \text{rank}\left(\mathscr{H}_{n+1}(\widehat{y})\right) \leq n \,\}$

α_{opt} can be found by bijection

Subspace methods $y \mapsto \mathscr{B}_R$ for exact data

find a basis R for the kernel of $\mathscr{H}_{n+1}(y)$

$$\underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_n \end{bmatrix}}_{R} \begin{bmatrix} y(1) & y(2) & \cdots & y(T-n) \\ y(2) & y(3) & \cdots & y(T-n+1) \\ \vdots & \vdots & & \vdots \\ y(n+1) & y(n+2) & \cdots & y(T) \end{bmatrix} = 0$$

interpret R as a polynomial

$$R(z) = R_0 + R_1 z + \cdots + R_n z^n$$

Subspace methods for noisy data (Prony's method in mathematics)

find an approximate kernel of $\mathscr{H}_{n+1}(y)$ by, *e.g.*,

$$\text{minimize} \quad \text{over } R \quad \|R\mathscr{H}_{n+1}(y)\|$$

equivalent to (unstructured) low-rank approximation

leads to an SVD solution method

Subspace methods $y \mapsto \mathscr{B}_{A,C}$ for exact data

1. rank revealing factorization

$$\mathscr{H}_L(y) = \underbrace{\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L+1} \end{bmatrix}}_{\mathscr{O}} \underbrace{\begin{bmatrix} x(0) & Ax(0) & A^2x(0) & \cdots & A^{T-L}x(0) \end{bmatrix}}_{\mathscr{C}}$$

2. shift equation

$$\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix} A = \begin{bmatrix} CA \\ CA^2 \\ \vdots \\ CA^{L} \end{bmatrix} \quad \Longleftrightarrow \quad \mathscr{O}(1{:}L-1,:)A = \mathscr{O}(2{:}L,:)$$

$T = 2n+1$ samples suffice, $\quad L \in [n+1, T-n]$

Subspace methods for noisy data (Kung's algorithm in system theory)

do steps 1 and 2 approximately:

1. singular value decomposition of $\mathscr{H}_L(y)$
2. least squares solution of the shift equation

L is a hyper-parameter, that affects the solution $\widehat{\mathscr{B}}$

Local optimization using variable projections

"double" optimization

$$\min_{\widehat{\mathscr{B}} \in \mathscr{L}_n} \left(\min_{\widehat{y} \in \widehat{\mathscr{B}}} \|y - \widehat{y}\| \right)$$

"inner" minimization

$$\text{error}(y, \widehat{\mathscr{B}}) = \|\Pi_{\widehat{\mathscr{B}}} y\|$$

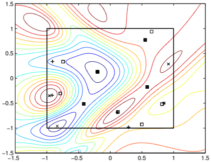
"outer" minimization

$$\min_{\widehat{\mathscr{B}} \in \mathscr{L}_n} \ \text{error}(y, \widehat{\mathscr{B}})$$

Parameter optimization problem

choosing kernel representation $\widehat{\mathscr{B}} = \mathscr{B}_R$

$$\min_{\widehat{\mathscr{B}} \in \mathscr{L}_n} \text{error}(y, \widehat{\mathscr{B}}) \quad \iff \quad \min_{R \neq 0} \text{error}(y, R)$$



optimization over Euclidean spaces

$$R \neq 0 \quad \iff \quad \begin{array}{c} R = \begin{bmatrix} x & 1 \end{bmatrix} \Pi \\ \Pi \text{ permutation} \end{array}$$

- Π fixed $\rightsquigarrow$ total least-squares
- Π can be changed during the optimization

Software `slra.github.io`

"low-level" SLRA package

- C++ implementation
- mosaic-Hankel structure
- element-wise weights

"high-level" IDENT package

- system identification
- unstable systems
- missing data and multiple data sets

Summary

representations lead to parameter optimization problems

three different optimization approaches

- convex relaxation
- subspace methods
- local optimization

variable projection is effective when $n \ll T$

Next . . .

Model representations

Modeling algorithms

Generalizations of the problem

Three generalizations

multiple experiments with missing values

models with inputs

nonlinear models

Using data from multiple experiments

for consistent estimation ($\widehat{\mathscr{B}} \to \bar{\mathscr{B}}$), T must go to infinity

however, long measurement is not possible in case of

- unstable system ($\bar{y}(t) \to \infty$)
- stable system ($\bar{y}(t) \to 0$)

data from N experiments: $y = \{ y^1, \dots, y^N \}$

y exact $\iff \operatorname{rank}\left(\mathscr{H}_{n+1}(y)\right) \leq n$

$$y \subset \mathscr{B} \in \mathscr{L}_n$$

$$\Updownarrow$$

$$y^k \in \mathscr{B} \in \mathscr{L}_n \quad \text{for all } k = 1, \ldots, N$$

$$\Updownarrow$$

$$\operatorname{rank}\Big(\underbrace{\big[\mathscr{H}_{n+1}(y^1) \ \cdots \ \mathscr{H}_{n+1}(y^N)\big]}_{\text{mosaic-Hankel matrix } \mathscr{H}_{n+1}(y)}\Big) \leq n$$

Dealing with exact and missing data values

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{y} \quad \|y - \widehat{y}\|_v \\ \text{subject to} & \operatorname{rank}\big(\mathscr{H}_{n+1}(\widehat{y})\big) \leq n \end{array}$$

weighted 2-norm approximation

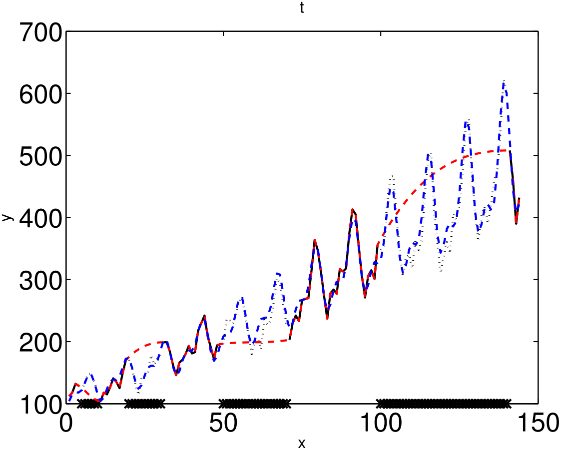
$$\|y - \widehat{y}\|_v := \sqrt{\sum_{k,t} v^k(t) \big(y^k(t) - \widehat{y}^k(t)\big)^2}$$

with element-wise weights

$$\begin{array}{lll} v^k(t) \in (0, \infty) & \text{if } y^k(t) \text{ is noisy} & \text{approximate } y^k(t) \\ v^k(t) = 0 & \text{if } y^k(t) \text{ is missing} & \text{interpolate } y^k(t) \\ v^k(t) = \infty & \text{if } y^k(t) \text{ is exact} & \widehat{y}^k(t) = y^k(t) \end{array}$$

Example: airline passenger data 1949–1960

`[5:10 20:30 50:70 100:140]` are missing

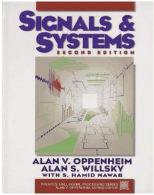


piecewise cubic interpolation, 6th order LTI model

In control and signal processing, models may have inputs

the model maps input u to output y

$$u \longrightarrow \boxed{\text{model}} \longrightarrow y$$



the model is specified by a function $y = f(u)$

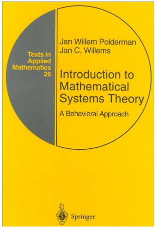
how to incorporate inputs in the sum-of-exp. model?

Behavioral definition of model

a model is a subset

$$\mathscr{B} = \{\, \widehat{w} \mid g(\widehat{w}) = 0 \text{ holds} \,\}$$

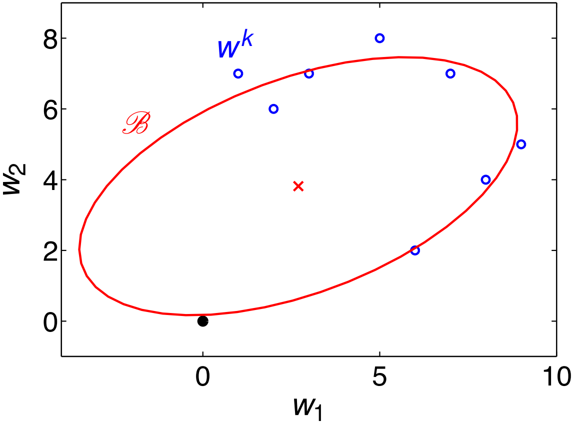
represented by relation



$\widehat{y} = f(\widehat{u})$ is special case of $g(\widehat{w}) = 0$ $\quad (g(\widehat{u}, \widehat{y}) = \widehat{y} - f(\widehat{u}))$

Nonlinear (static quadratic) model

$\mathscr{B}$ — model: conic section
$\mathscr{M}$ — model class: all conic sections



Conic section fitting

the points $(x_1, y_1), \ldots, (x_N, y_N)$ lie on a conic section

$$\Updownarrow$$

there are $A = A^\top$, b, c, at least one of them nonzero, s.t.

$$\begin{bmatrix} x_i & y_i \end{bmatrix} A \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} x_i & y_i \end{bmatrix} b + c = 0, \quad \text{for } i = 1, \ldots, N$$

$$\Updownarrow$$

there is $\theta = \begin{bmatrix} a_{11} & a_{12} & a_{22} & b_1 & b_2 & c \end{bmatrix} \neq 0$, such that

$$\theta \begin{bmatrix} x_1^2 & \cdots & x_N^2 \\ x_1 y_1 & \cdots & x_N y_N \\ x_1 & \cdots & x_N \\ y_1^2 & \cdots & y_N^2 \\ y_1 & \cdots & y_N \\ 1 & \cdots & 1 \end{bmatrix} = 0$$

Conic section fitting $\iff$ rank deficiency

the points $(x_1, y_1), \ldots, (x_N, y_N)$ lie on a conic section

$$\mathscr{B}(\theta) = \{\, w \mid w^\top A w + w^\top b + c = 0 \,\}$$

$$\Updownarrow$$

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

Conclusion

considering alternative representations of the model

- poles
- kernel
- state-space

allows us to unify different solution methods

- nuclear norm
- subspace (Prony, Kung, . . .)
- local optimization

and generalize the sum-of-exponentials problem to

- data from multiple experiments with fixed/missing values
- models with inputs
- nonlinear models

Scope of structured low-rank approximation