Structured Low-Rank Approximation in Signal Processing

Ivan Markovsky

K.U.Leuven, ESAT-SISTA

Outline

Introduction

We will consider the following signal processing problems:

1. Linear prediction $\leftrightarrow$ sum-of-damped-exp. modeling
2. Harmonic retrieval $\leftrightarrow$ sum-of-exp. modeling
3. Deconvolution $\leftrightarrow$ FIR modeling
4. 2D deconvolution $\leftrightarrow$ image deblurring

Linear prediction problem

Future values of w are estimated as linear comb. of past values

$$w(t) = r_1 w(t-1) + r_2 w(t-2) + \cdots + r_n w(t-n) \qquad \text{(LP)}$$

r_i are the linear prediction coefficients

Given an observed signal w_d, how do we find the coefficients r_i?

There are many methods for doing this:

- Pisarenko, Prony, Kumaresan–Tufts methods
- subspace methods
- frequency domain methods
- maximum likelihood method

Sum-of-damped-exponentials model

Model the signal w as

$$w(t) = \sum_{i=1}^{n} a_i e^{d_i t} e^{\mathbf{i}(\omega_i t + \phi_i)} \qquad \text{(SDE)}$$

where a_i, d_i, ϕ_i, and ω_i are parameters of the model

a_i	—	amplitudes	d_i	—	dampings
ω_i	—	frequencies	ϕ_i	—	initial phases

For all $\{a_i, d_i, \omega_i, \phi_i\}$ there are r_i and $w(-n+1), \ldots, w(0)$, s.t. the solution of (LP) coincides with (SDE) and vice verse.

the LP problem $\iff$ modeling by (SDE)

Harmonic retrieval problem

Corresponds to modeling w as a sum-of-exponentials

$$w(t) = \sum_{i=1}^{n} a_i e^{\mathbf{i}(\omega_i t + \phi_i)}$$

A special sum-of-damped-exp. model, with dampings $d_i = 0$.

$\implies$ w satisfies the linear prediction (LP) equation

Moreover, a sum-of-exp. signal w satisfies the equation

$$w(t-n) = r_1 w(t-n+1) + r_2 w(t-n+2) + \cdots + r_n w(t) \quad \text{(LP')}$$

were r_i are the linear prediction coefficients.

(LP) — forward prediction (LP') — backward prediction

Deconvolution problem and FIR model

Given signals u and y, find a signals h, such that

$$y(t) = (h \star u)(t) := \sum_{\tau=T_{\text{ini}}}^{T_{\text{f}}} h(\tau) u(t-\tau)$$

Interpretation:

u	—	input	y	—	output
h	—	impulse response (of an FIR system)			

model y as the output of an FIR system with input u

$T_{\text{ini}} \geq 0 \implies$ causal system

2D deconvolution and image deblurring

Deconvolution for signals with two independent variables

$$y(t_1, t_2) = (h \star u)(t_1, t_2) := \sum_{\tau_1 = T_{\text{ini1}}}^{T_{\text{f1}}} \sum_{\tau_2 = T_{\text{ini2}}}^{T_{\text{f2}}} h(\tau_1, \tau_2) u(t_1 - \tau_1, t_2 - \tau_2)$$

Interpretation:

u	—	true image	y	—	blurred image
h	—	point spread function (PSF of an blurring operator)			

given a blurred image and PSF, find the true image

(The topic of the previous summer school organized in Monopoli.)

Outline

Signal processing

Low-rank approximation

Algorithms

Conclusions

Linear prediction problem

$w = \big(w(1),\dots,w(T)\big)$ sum-of-damped-exp. $\Longrightarrow$ w satisfies

$$r_0 w(t) + r_1 w(t+1) + \cdots + r_n w(t+n) = 0, \quad \text{for } t = 1,\dots,T-n$$

Written in a matrix form these equations are

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_n \end{bmatrix} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-n) \\ w(2) & w(3) & \cdots & w(T-n+1) \\ \vdots & \vdots & & \vdots \\ w(n+1) & w(n+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}_{n+1}(w)} = 0$$

which shows that the Hankel matrix $\mathscr{H}_{n+1}(w)$ is rank deficient

$$\operatorname{rank}\big(\mathscr{H}_{n+1}(w)\big) \le n$$

Harmonic retrieval problem

$w = \big(w(1),\dots,w(T)\big)$ sum-of-exp. $\Longrightarrow$ w satisfies

forward LP equation: $\quad r\,\mathscr{H}_{n+1}(w) = 0$

and backward LP equation

$$\begin{bmatrix} r_0 & r_1 & \cdots & r_n \end{bmatrix} \underbrace{\begin{bmatrix} w(n+1) & w(n+2) & \cdots & w(T) \\ \vdots & \vdots & & \vdots \\ w(2) & w(3) & \cdots & w(T-n+1) \\ w(1) & w(2) & \cdots & w(T-n) \end{bmatrix}}_{\mathscr{T}_{n+1}(w)} = 0$$

$\Longrightarrow$ $r\big(\mathscr{H}_{n+1}(w) + \mathscr{T}_{n+1}(w)\big) = 0$, *i.e.*,

$$\operatorname{rank}\big(\mathscr{H}_{n+1}(w) + \mathscr{T}_{n+1}(w)\big) \le n$$

Deconvolution problem

1-dimensional case:

$$h \star u = y \quad \Longleftrightarrow \quad h\mathscr{T}_n(u) = y \quad \Longleftrightarrow \quad \operatorname{rank}\left(\begin{bmatrix} \mathscr{T}_n \\ y \end{bmatrix}\right) \le n$$

2-dimensional case:

$$h \star u = y \quad \Longleftrightarrow \quad \operatorname{rank}\left(\begin{bmatrix} \mathscr{S}_n \\ y \end{bmatrix}\right) \le n$$

where $\mathscr{S}$ is a block-Toeplitz, Toeplitz-block structured matrix

Optimal modeling: structured low-rank approximation

Given $w_{\mathrm{d}} = \big(w_{\mathrm{d}}(1),\dots,w_{\mathrm{d}}(T)\big)$, find $\widehat{w} = \big(\widehat{w}(1),\dots,\widehat{w}(T)\big)$ that is

1. sum-of-damped-exp, or sum-of-exp, or FIR
   $\iff$ rank $\big(\mathscr{S}(\widehat{w})\big) \le n$, for certain structures $\mathscr{S}$
2. as close as possible to w_{d}, *i.e.*, $\|w_{\mathrm{d}} - \widehat{w}\|$ is minimized

$$\boxed{\widehat{w}^* := \arg\min_{\widehat{w}} \|w_{\mathrm{d}} - \widehat{w}\| \quad \text{subject to} \quad \operatorname{rank}\big(\mathscr{S}(\widehat{w})\big) \le n}$$

In the first 3 appl., $\mathscr{S}(\cdot)$ is composed of blocks that are:

Hankel, Toeplitz, Hankel+Toeplitz, or unstructured.

$\rightsquigarrow$ efficient algorithms (computational complexity $O(T)$)

Multivariable and multidimensional systems

$y = \mathscr{T}(h)u = \mathscr{T}(u)h$

Toeplitz matrix times vector

$\Updownarrow$

$(u,y) \in \mathscr{B}(h)$ FIR sys. traj.	$\iff$	$y = h \star u$ convolution	$\iff$	$y(z) = h(z)u(z)$ polyn. multipl.

Multivariable case: block Toeplitz structure

multivariable systems	$\iff$	matrix valued time series	$\iff$	matrix valued polynomials

2D case: block Toeplitz–Toeplitz block structure

multidim. system	$\iff$	function of several indep. variables	$\iff$	polyn. of several var.

Outline

Unstructured low-rank approximation

$$\widehat{D}^* := \arg\min_{\widehat{D}} \|D - \widehat{D}\|_{\mathrm{F}} \quad \text{subject to} \quad \operatorname{rank}(\widehat{D}) \le n$$

Closed form solution: Let $D = U\Sigma V^\top$ be the SVD of D.

$$U =: \begin{array}{cc} n & * \\ \multicolumn{2}{c}{\begin{bmatrix} U_1 & U_2 \end{bmatrix}} \end{array}, \quad \Sigma =: \begin{array}{c} \begin{array}{cc} n & * \end{array} \\ \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \end{array} \begin{array}{c} n \\ * \end{array} \quad \text{and} \quad V =: \begin{array}{cc} n & * \\ \multicolumn{2}{c}{\begin{bmatrix} V_1 & V_2 \end{bmatrix}} \end{array}$$

An optimal low-rank approximate solution is

$$\widehat{D}^* = U_1 \Sigma_1 V_1^\top$$

It is unique if and only if $\sigma_n \ne \sigma_{n+1}$.

Structured low-rank approximation

No closed form solution is known for the general SLRA problem

$$\widehat{w}^* := \arg\min_{\widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \operatorname{rank}\big(\mathscr{S}(\widehat{w})\big) \leq n$$

NP-hard, consider solution methods based on local optimization

Representing the constraint in a kernel form, the problem is

$$\min_{r, rr^\top = 1} \Big(\min_{\widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad r\mathscr{S}(\widehat{w}) = 0 \Big)$$

Double minimization with bilinear equality constraint.

There is a matrix $G(r)$, such that $r\mathscr{S}(\widehat{w}) = 0 \iff \widehat{w}G(r) = 0$.

Variable projection vs. alternating projections

Two ways to approach the double minimization:

- Variable projections (VARPRO):
  solve the inner minimization analytically

$$\min_{r, rr^\top = 1} r\mathscr{S}(w_\mathrm{d})\big(G^\top(r)G(r)\big)^{-1}\mathscr{S}^\top(w_\mathrm{d})r^\top$$

$\rightsquigarrow$ a nonlinear least squares problem for r only.

- Alternating projections (AP):
  alternate between solving two least squares problems

VARPRO is globally convergent with a super linear conv. rate.

AP is globally convergent with a linear convergence rate.

Algorithmic details using the VARPRO approach

The structured low-rank approximation problem is equivalent to

$$\min_{r, rr^\top = 1} r\mathscr{S}(w_\mathrm{d})\big(G^\top(r)G(r)\big)^{-1}\mathscr{S}^\top(w_\mathrm{d})r^\top$$

To evaluate the cost function we need to solve for z

$$\big(G^\top(r)G(r)\big)z = \big(r\mathscr{S}(w_\mathrm{d})\big)$$

What special structure does $G^\top G$ have?

Banded Toeplitz for any $\mathscr{S} = \begin{bmatrix}\mathscr{S}_1 & \cdots & \mathscr{S}_q\end{bmatrix}$, where $\mathscr{S}_i$ is Toeplitz, Hankel, Toeplitz+Hankel, unstructured, or fixed.

Special case: sum-of-damped-exp. modeling

In the sum-of-damped-exp. modeling, the structure is

$$\mathscr{S}(w) = \mathscr{H}_{n+1}(w)$$

What matrix G satisfies

$$r\mathscr{H}_{n+1}(w) = 0 \quad \iff \quad wG(r) = 0$$

for all r and w? What is the structure of $G^\top G$?

Special case: sum-of-damped-exp. modeling

$$
\begin{bmatrix} r_0 & r_1 & \cdots & r_n \end{bmatrix}
\underbrace{\begin{bmatrix}
w(1) & w(2) & \cdots & w(T-n) \\
w(2) & w(3) & \cdots & w(T-n+1) \\
\vdots & \vdots & & \vdots \\
w(n+1) & w(n+2) & \cdots & w(T)
\end{bmatrix}}_{\mathscr{H}_{n+1}(w)}
$$

$$
= \begin{bmatrix} w_1 & w_2 & \cdots & w_T \end{bmatrix}
\underbrace{\begin{bmatrix}
r_0 & & & \\
r_1 & r_0 & & \\
\vdots & r_1 & \ddots & \\
r_n & \vdots & \ddots & r_0 \\
 & r_n & & r_1 \\
 & & \ddots & \vdots \\
 & & & r_n
\end{bmatrix}}_{G(r)}
$$

Special case: sum-of-damped-exp. modeling

Therefore,

$$
G^\top G = \begin{bmatrix}
r_0 & r_1 & \cdots & r_n & & & \\
 & r_0 & r_1 & \cdots & r_n & & \\
 & & \ddots & \ddots & & \ddots & \\
 & & & r_0 & r_1 & \cdots & r_n
\end{bmatrix}
\begin{bmatrix}
r_0 & & & \\
r_1 & r_0 & & \\
\vdots & r_1 & \ddots & \\
r_n & \vdots & \ddots & r_0 \\
 & r_n & & r_1 \\
 & & \ddots & \vdots \\
 & & & r_n
\end{bmatrix}
$$

(All missing elements are zeros.)

Special case: sum-of-damped-exp. modeling

$$
G^\top G = \begin{bmatrix}
\sum_{i=0}^n r_i r_i & \sum_{i=1}^n r_i r_{i-1} & \cdots & r_n r_0 & & & \\
\sum_{i=1}^n r_{i-1} r_i & \ddots & & & \ddots & & \\
\vdots & & \ddots & & & \ddots & \\
r_0 r_n & & & \ddots & & & r_n r_0 \\
 & \ddots & & & \ddots & & \vdots \\
 & & \ddots & & & \ddots & \sum_{i=1}^n r_i r_{i-1} \\
 & & & r_0 r_n & \cdots & \sum_{i=1}^n r_{i-1} r_i & \sum_{i=0}^n r_i r_i
\end{bmatrix}
$$

banded Toeplitz, bandwidth $2n+1$

Special case: sum-of-exp. modeling

In sum-of-exponentials modeling, the structure $\mathscr{S}$ is

$$\mathscr{S}(w) = \mathscr{H}_{n+1}(w) + \mathscr{T}_{n+1}(w)$$

What matrix G satisfies

$$r\left(\mathscr{H}_{n+1}(w) + \mathscr{T}_{n+1}(w)\right) = 0 \quad \iff \quad wG(r) = 0$$

for all r and w? What is the structure of $G^\top G$?

Special case: sum-of-exp. modeling

$$
\begin{bmatrix} r_0 & r_1 & \cdots & r_n \end{bmatrix} \left(\mathscr{H}_{n+1}(w) + \mathscr{T}_{n+1}(w)\right)
$$

$$
= \begin{bmatrix} w_1 & w_2 & \cdots & w_T \end{bmatrix} \underbrace{\left(\begin{bmatrix} r_0 & & & \\ r_1 & r_0 & & \\ \vdots & r_1 & \ddots & \\ r_n & \vdots & \ddots & r_0 \\ & r_n & & r_1 \\ & & \ddots & \vdots \\ & & & r_n \end{bmatrix} + \begin{bmatrix} r_n & & & \\ \vdots & r_n & & \\ r_1 & \vdots & \ddots & \\ r_0 & r_1 & & r_n \\ & r_0 & \ddots & \vdots \\ & & \ddots & r_1 \\ & & & r_0 \end{bmatrix} \right)}_{G(r)}
$$

Define

$$
\tilde{r}_i := r_i + r_{n-i}, \quad \text{for } i = 0, 1, \ldots, n
$$

Special case: sum-of-exp. modeling

Therefore,

$$
G^\top G = \begin{bmatrix} \tilde{r}_0 & \tilde{r}_1 & \cdots & \tilde{r}_n & & & \\ & \tilde{r}_0 & \tilde{r}_1 & \cdots & \tilde{r}_n & & \\ & & \ddots & \ddots & & \ddots & \\ & & & \tilde{r}_0 & \tilde{r}_1 & \cdots & \tilde{r}_n \end{bmatrix} \begin{bmatrix} \tilde{r}_0 & & & \\ \tilde{r}_1 & \tilde{r}_0 & & \\ \vdots & \tilde{r}_1 & \ddots & \\ \tilde{r}_n & \vdots & \ddots & \tilde{r}_0 \\ & \tilde{r}_n & & \tilde{r}_1 \\ & & \ddots & \vdots \\ & & & \tilde{r}_n \end{bmatrix}
$$

Special case: sum-of-exp. modeling

$$
G^\top G = \begin{bmatrix} \sum_{i=0}^{n} \tilde{r}_i \tilde{r}_i & \sum_{i=1}^{n} \tilde{r}_i \tilde{r}_{i-1} & \cdots & \tilde{r}_n \tilde{r}_0 & & & \\ \sum_{i=1}^{n} \tilde{r}_{i-1} \tilde{r}_i & \ddots & & & \ddots & & \\ \vdots & & \ddots & & & \ddots & \\ \tilde{r}_0 \tilde{r}_n & & & \ddots & & & \tilde{r}_n \tilde{r}_0 \\ & \ddots & & & \ddots & & \vdots \\ & & \ddots & & & \ddots & \sum_{i=1}^{n} \tilde{r}_i \tilde{r}_{i-1} \\ & & & \tilde{r}_0 \tilde{r}_n & \cdots & \sum_{i=1}^{n} \tilde{r}_{i-1} \tilde{r}_i & \sum_{i=0}^{n} \tilde{r}_i \tilde{r}_i \end{bmatrix}
$$

banded Toeplitz, bandwidth $2n+1$

Special case: FIR system modeling

In FIR system modeling, the structure $\mathscr{S}$ is

$$
\mathscr{S}(w) = \begin{bmatrix} \mathscr{H}_n(u) \\ y \end{bmatrix}
$$

What matrix G satisfies

$$
\begin{bmatrix} h & -1 \end{bmatrix} \begin{bmatrix} \mathscr{H}_n(u) \\ y \end{bmatrix} = 0 \quad \iff \quad wG(h) = 0
$$

for all h and $w = (u, y)$. What is the structure of $G^\top G$?

Special case: FIR system modeling

$$\begin{bmatrix} h_1 & \cdots & h_n & -1 \end{bmatrix} \underbrace{\begin{bmatrix} u(1) & u(2) & \cdots & u(T-n+1) \\ \vdots & \vdots & & \vdots \\ u(n) & u(n+1) & \cdots & u(T) \\ y(1) & u(2) & \cdots & y(T-n+1) \end{bmatrix}}_{\mathrm{col}(\mathscr{H}_n(u),y)}$$

$$= \begin{bmatrix} u_1 & \cdots & u_T & | & y_1 & \cdots & y_{T-n+1} \end{bmatrix} \underbrace{\left[\begin{array}{cccc} h_1 & & & \\ \vdots & h_1 & & \\ h_n & \vdots & \ddots & \\ & h_n & & h_1 \\ & & \ddots & \vdots \\ & & & h_n \\ \hline & & -I & \end{array}\right]}_{G(h)}$$

Special case: FIR system modeling

Therefore,

$$G^\top G = \left[\begin{array}{ccccc|c} h_1 & \cdots & h_n & & & \\ & h_1 & \cdots & h_n & & -I \\ & & \ddots & & \ddots & \\ & & & h_1 & \cdots \; h_n & \end{array}\right] \left[\begin{array}{cccc} h_1 & & & \\ \vdots & h_1 & & \\ h_n & \vdots & \ddots & \\ & h_n & & h_1 \\ & & \ddots & \vdots \\ & & & h_n \\ \hline & & -I & \end{array}\right]$$

Special case: FIR system modeling

$$G^\top G = \begin{bmatrix} 1+\sum_{i=1}^{n} h_i h_i & \sum_{i=2}^{n} h_i h_{i-1} & \cdots & h_n h_1 & & & \\ \sum_{i=2}^{n} h_{i-1} h_i & \ddots & & & \ddots & & \\ \vdots & & \ddots & & & \ddots & \\ h_1 h_n & & & \ddots & & & h_n h_1 \\ & \ddots & & & \ddots & & \vdots \\ & & \ddots & & & \ddots & \sum_{i=1}^{n} h_i h_{i-1} \\ & & & h_1 h_n & \cdots & \sum_{i=2}^{n} h_{i-1} h_i & 1+\sum_{i=1}^{n} h_i h_i \end{bmatrix}$$

banded Toeplitz, bandwidth $2n-1$

Outline

- Signal processing
- Low-rank approximation
- Algorithms
- **Conclusions**

What are the issues?

model class

$\mathscr{M}_n$

$\downarrow$

modeling problem

$w_\mathrm{d} \mapsto \widehat{w} \in \mathscr{M}_n$

$\downarrow$

solution method

(parameterization of $\mathscr{M}_n$, optimization method, ...)

$\downarrow$

algorithms

$\downarrow$

software

For examples ...

$$\mathscr{M}_n = \{\, w \in \mathbb{R}^T \mid w \text{ is a sum of } \leq n \text{ damped-exp.} \,\}$$

$\downarrow$

$$\min_{\widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \widehat{w} \in \mathscr{M}_n$$

$\downarrow$

parameterize $\mathscr{M}_n$ by

$\{\, a_i, d_i, \omega_i, \phi_i \,\}_{i=1}^n$ or $\{\, r_i \,\}_{i=1}^n, w(-n+1), \dots, w(0)$

$\downarrow$

exploit matrix structure for efficiency

$\rightsquigarrow$ Cholesky factorization of banded Toeplitz matrix

$\downarrow$

use the SLICOT function `MB02GD`, MINPACK, ...

The sequence

model class

$\mathscr{M}$

$\Downarrow$

structure in the problem

$\mathscr{S}$

$\Downarrow$

structure in the algorithm

G

is common.

Main point: $w \in \mathscr{M}_n \iff \operatorname{rank}\big(\mathscr{S}(w)\big) \leq n$

An modeling problem

$$\min_{\widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \widehat{w} \in \mathscr{M}_n$$

is equivalent to structured low-rank approximation problem

$$\min_{\widehat{w}} \|w_\mathrm{d} - \widehat{w}\| \quad \text{subject to} \quad \operatorname{rank}\big(\mathscr{S}(w)\big) \leq n$$

for certain structure $\mathscr{S}$ that depends on $\mathscr{M}$.

Often $\mathscr{S}$ is composed of Hankel/Toeplitz, Hankel+Toeplitz, and unstr. blocks $\rightsquigarrow$ an algorithm with comput. complexity $O(T)$

$\Longrightarrow$ one algorithm solves efficiently a variety of problems

Thank you