Polynomial and structured matrix methods in system theory and signal processing

Ivan Markovsky

University of Southampton

Outline

linear time-invariant dynamic systems	$\leftrightarrow$	polynomial matrices	$\leftrightarrow$	structured matrices

Linear time-invariant (LTI) models are used for

- simulation
- filtering/smoothing/prediction, and
- control

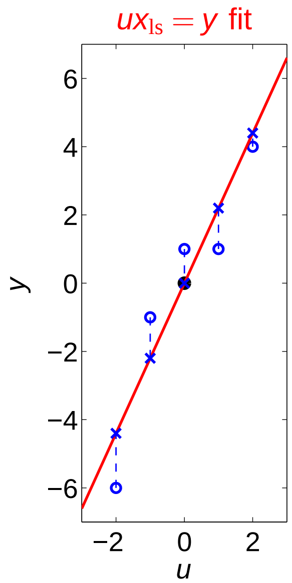
they are in the core of control theory and signal processing.

Polynomials are used for representation of LTI systems.

Structured matrix methods are used for

- deriving LTI models from data — system identification
- analysis and synthesis of LTI models.

What is a model?



Classic problem: Fit the points

$$w_1 = \begin{bmatrix} -2 \\ -6 \end{bmatrix},\ w_2 = \begin{bmatrix} -1 \\ -1 \end{bmatrix},\ \ldots,\ w_5 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

by a line passing through the origin.

Classic solution: Define $w_i =: \operatorname{col}(u_i, y_i)$ and solve the least squares problem

$$\operatorname{col}(u_1, \ldots, u_5)x = \operatorname{col}(y_1, \ldots, y_5).$$

The model is the line

$$\mathscr{B} := \{\, w = \operatorname{col}(u, y) \mid ux_{\text{ls}} = y \,\}$$

and not the equation $ux_{\text{ls}} = y$.

Linear static model: subspace $\mathscr{B} \subset \mathbb{R}^{\mathtt{w}}$

Representations of a linear static model $\mathscr{B} \subseteq \mathbb{R}^{\mathtt{w}}$:

- kernel $\mathscr{B} = \ker(R) = \{ w \mid Rw = 0 \}$ $\qquad ([R_{\mathrm{i}}\ R_{\mathrm{o}}] := R)$
- image $\mathscr{B} = \operatorname{image}(M) = \{ w = Mv \mid \text{for all } v \}$ $\qquad \left(\begin{bmatrix} M_{\mathrm{i}} \\ M_{\mathrm{o}} \end{bmatrix} := M\right)$
- input/output $\mathscr{B} = \mathscr{B}_{\mathrm{i/o}}(X) := \{ w = \operatorname{col}(u,y) \mid Xu = y \}$

$$\begin{tikzcd}
\mathscr{B}=\ker(R) \arrow[rr, leftrightarrow, "RM=0"] \arrow[dr, shift left, "X^\top=-R_{\mathrm{o}}^{-1}R_{\mathrm{i}}"] & & \mathscr{B}=\operatorname{image}(M) \arrow[dl, shift right, "X^\top=M_{\mathrm{o}}M_{\mathrm{i}}^{-1}"'] \\
& \mathscr{B}=\mathscr{B}_{\mathrm{i/o}}(X) \arrow[ul, shift left, "R=[X^\top\ -I]"] \arrow[ur, shift right, "M^\top=[I\ \ X]"'] &
\end{tikzcd}$$

Linear time-invariant dynamic models

Static model with $\mathtt{w}$ variables is a subset of $\mathbb{R}^{\mathtt{w}}$.

Dynamic model with $\mathtt{w}$ variables and time axis $\mathbb{T}$ is a subset of

the set of functions from $\mathbb{T}$ to $\mathbb{R}^{\mathtt{w}}$ $\quad (\mathbb{R}^{\mathtt{w}})^{\mathbb{T}} := \{ w \mid w : \mathbb{T} \mapsto \mathbb{R}^{\mathtt{w}} \}$.

$\mathbb{T} = \mathbb{R}$ — continuous time, $\qquad \mathbb{T} = \mathbb{Z}$ — discrete time

$\mathscr{B} \subset (\mathbb{R}^{\mathtt{w}})^{\mathbb{Z}}$ is linear if $\mathscr{B}$ is a subspace.

Define the shift operator $(\sigma w)(t) = w(t+1)$.

$\mathscr{B} \subset (\mathbb{R}^{\mathtt{w}})^{\mathbb{Z}}$ is time-invariant if $\sigma\mathscr{B} = \mathscr{B}$.

Polynomial representations of LTI systems

Theorem (J. C. Willems)

$\mathscr{B} \subset (\mathbb{R}^{\mathtt{w}})^{\mathbb{Z}}$ is LTI if and only if there is a polynomial matrix

$$R(z) = R_0 + R_1 z + \cdots + R_\ell z^\ell$$

such that

$$\mathscr{B} = \ker\big(R(\sigma)\big),$$

i.e., $\mathscr{B}$ is the kernel of a linear difference operator $R(\sigma)$.

$$\begin{aligned}
w \in \mathscr{B} \quad &\iff \quad R(\sigma)w = 0 \\
&\iff \quad R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \\
&\iff \quad R_0 w(t) + R_1 w(t+1) + \cdots + R_\ell w(t+\ell) = 0, \ \text{for all } t
\end{aligned}$$

Adding more structure: Input/output partitions

Define

$$\begin{bmatrix} Q & -P \end{bmatrix} := R\Pi, \quad \text{with} \quad P \text{ square}$$

where $\Pi \in \mathbb{R}^{\mathtt{w}\times\mathtt{w}}$ is a permutation matrix. Then

$$w = \Pi \begin{bmatrix} u \\ y \end{bmatrix}, \quad Q(\sigma)u = P(\sigma)y$$

$\det(P) \neq 0 \implies u$ is an "input" and y is an "output" of $\mathscr{B}$

Input/output partitions are not unique, however,

$\mathbf{p}(\mathscr{B}) := \dim(y) = \operatorname{rank}(R)$ and $\mathbf{m}(\mathscr{B}) := \dim(u) = \mathtt{w} - \mathbf{p}(\mathscr{B})$

are invariants of $\mathscr{B}$, called output and input cardinalities.

Input/output representation

$$\mathscr{B}_{\mathrm{i/o}}(\Pi, P, Q) := \{\, \Pi \operatorname{col}(u, y) \mid Q(\sigma)u = P(\sigma)y \,\} \qquad \text{(I/O)}$$

Theorem $\mathscr{B} \subset (\mathbb{R}^{\mathtt{w}})^{\mathbb{Z}}$ is LTI iff there is an I/O representation, *i.e.*,

there are Π, P, Q, such that $\mathscr{B} = \mathscr{B}_{\mathrm{i/o}}(\Pi, P, Q)$.

Generically, Π can be chosen to be the identity matrix I,

in which case we write $\mathscr{B}_{\mathrm{i/o}}(P, Q)$.

$\operatorname{lag}(\mathscr{B})$ — the smallest possible $\deg(R) = \deg(P)$,

such that $\ker\big(R(\sigma)\big) = \mathscr{B}$.

Controllability

Definition $\mathscr{B}$ is controllable if for all $w_1, w_2 \in \mathscr{B}$, $\exists\, w \in \mathscr{B}$, $\tau > 0$, such that $w_1(t) = w(t)$, for all $t < 0$ and $w_2(t) = w(t)$, for all $t \geq \tau$.

Think of w_1 as a given past traj. and w_2 as a desired future traj.

$\mathscr{B}$ controllable $\iff$ any given traj. can be steered to any desired trajectory

important condition for pole-placement, LQ, H_∞, ... control, *e.g.*,

controllability $\iff$ solvability of the state feedback pole-placement problem

Controllability test in terms of I/O representation

For numerically checking controllability of $\mathscr{B}$, we need to relate this property to the parameters of $\mathscr{B}$ in a particular representation.

Consider an I/O representation $\mathscr{B} = \mathscr{B}_{\mathrm{i/o}}(P, Q)$.

Theorem $\mathscr{B}_{\mathrm{i/o}}(P, Q)$ is controllable iff P and Q are coprime.

$\Longrightarrow$ checking controllability is a coprimness test problem for a pair of polynomial matrices.

Outline

polynomial × polynomial $\iff$ Toeplitz matrix × vector

$$c(z) = a(z)b(z) \quad \iff \quad \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{\ell_c} \end{bmatrix} = \begin{bmatrix} a_0 & & & \\ a_1 & a_0 & & \\ \vdots & a_1 & \ddots & \\ a_{\ell_a} & \vdots & \ddots & a_0 \\ & a_{\ell_a} & & a_1 \\ & & \ddots & \vdots \\ & & & a_{\ell_a} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{\ell_b} \end{bmatrix}$$

$$\iff: \quad c = S_{\ell_b}(a)b \quad \iff \quad c = S_{\ell_a}(b)a$$

polynomial $c(z) \in \mathbb{R}[z]$, $\deg(c) = \ell_c \quad \longleftrightarrow \quad$ vector $c \in \mathbb{R}^{\ell_c+1}$

polynomial operations $\quad \longleftrightarrow \quad$ structured matrix operations

Example: computing the GCD of two polynomials

$p \in \mathbb{R}[z]$ and $q \in \mathbb{R}[z]$ have common divisor $c \in \mathbb{R}[z]$, $\deg(c) = \ell_c$ $\iff$ $\exists\, a \in \mathbb{R}[z]$, $\deg(a) = \ell_p - \ell_c$; $\exists\, b \in \mathbb{R}[z]$, $\deg(b) = \ell_q - \ell_c$ such that $p = ca$ and $q = cb$

$$\iff \quad qa - pb = 0$$

$$\iff \quad \begin{bmatrix} S_{\ell_a}(q) & S_{\ell_b}(p) \end{bmatrix} \begin{bmatrix} a \\ -b \end{bmatrix} = 0$$

$\iff \quad \begin{bmatrix} S_{\ell_a}(q) & S_{\ell_b}(p) \end{bmatrix}$ is rank deficient

($\begin{bmatrix} S_{\ell_a}(q) & S_{\ell_b}(p) \end{bmatrix}$ is $(\ell_p + \ell_q + 1 - \ell_c) \times (\ell_p + \ell_q + 2 - 2\ell_c)$)

Degree of GCD $\leftrightarrow$ rank of Sylvester matrix

Theorem The degree of the GCD c of p and q is equal to the rank deficiency of the Sylvester matrix $\begin{bmatrix} S_{\ell_p}(q) & S_{\ell_q}(p) \end{bmatrix}$, *i.e.*,

$$\deg(c) = \ell_p + \ell_q - \operatorname{rank}\left(\begin{bmatrix} S_{\ell_p}(q) & S_{\ell_q}(p) \end{bmatrix}\right).$$

computing the GCD's degree $\quad \longleftrightarrow \quad$ rank test of Sylvester structured matrix

rank test is numerically notoriously bad problem

Numerical rank of an unstructured matrix $A \in \mathbb{R}^{m \times n}$
minimal rank of a matrix $\widehat{A}$ in an ε-neighbourhood of A

For Sylvester structured matrix, however, we want to find the minimal rank of a Sylvester matrix in an ε-neighbourhood of A

Approx. GCD $\leftrightarrow$ Sylvester matrix low-rank approx.

Unstructured low-rank approximation:

$$\min_{\widehat{A}} \|A - \widehat{A}\| \quad \text{subject to} \quad \operatorname{rank}(\widehat{A}) \leq r.$$

Sylvester structured ℓ_c-rank approximation:

$$\varepsilon := \min_{\widehat{p},\widehat{q}} \|\operatorname{col}(p,q) - \operatorname{col}(\widehat{p},\widehat{q})\| \quad \text{subject to}$$
$$\operatorname{rank}\left(\begin{bmatrix} S_{\ell_p-\ell_c}(\widehat{q}) & S_{\ell_q-\ell_c}(\widehat{p}) \end{bmatrix}\right) \leq \ell_p + \ell_q - 2\ell_c + 1.$$

By construction $\widehat{p}$ and $\widehat{q}$ have GCD c of degree at least ℓ_c
$\rightsquigarrow$ approximate GCD of p and q.

backward error ε of c — the smallest size perturbation on p,q that makes c exact GCD

System theoretic meaning of approx. GCD

The LTI system $\mathscr{B}_{\mathrm{i/o}}(p,q)$ is controllable iff (p,q) are coprime.

In numerical linear algebra, yes/no questions ($\mathscr{B}$ contr./uncontr.) are replaced by quantitative measures (distance of $\mathscr{B}$ to uncontr.)

$$d(\mathscr{B}) := \min_{\widehat{\mathscr{B}}} \operatorname{dist}(\mathscr{B},\widehat{\mathscr{B}}) \quad \text{subject to} \quad \widehat{\mathscr{B}} \text{ is uncontrollable.}$$

With $p,\widehat{p}$ monic, $p,q,\widehat{p},\widehat{q}$ are unique (for given $\mathscr{B},\widehat{\mathscr{B}}$) and

$$\operatorname{dist}(\mathscr{B},\widehat{\mathscr{B}}) := \|\operatorname{col}(p,q) - \operatorname{col}(\widehat{p},\widehat{q})\|$$

becomes a property of the pair of systems $(\mathscr{B},\widehat{\mathscr{B}})$.

The problem of computing $d\big(\mathscr{B}_{\mathrm{i/o}}(p,q)\big)$ is equivalent to computing an approximate GCD of (p,q) with degree 1.

Optimal modelling

Given a time series

$$w_{\mathrm{d}} = \big(w_{\mathrm{d}}(1),\dots,w_{\mathrm{d}}(T)\big)$$

find another time series

$$\widehat{w} = \big(\widehat{w}(1),\dots,\widehat{w}(T)\big)$$

that is

1. as close as possible to w_{d},
2. an exact trajectory of an LTI system $\widehat{\mathscr{B}}$,
3. $\widehat{\mathscr{B}}$ is of bounded complexity.

$$\boxed{\min_{\widehat{\mathscr{B}},\widehat{w}} \underbrace{\|w_{\mathrm{d}} - \widehat{w}\|}_{1} \quad \text{subject to} \quad \underbrace{\widehat{w} \in \widehat{\mathscr{B}}}_{2}, \quad \underbrace{\operatorname{lag}(\widehat{\mathscr{B}}) \le \ell,\ \mathbf{m}(\widehat{\mathscr{B}}) \le \mathtt{m}}_{3}}$$

Optimal modelling $\leftrightarrow$ structured low-rank appr.

Define

$$\mathscr{H}_\ell(\widehat{w}) := \begin{bmatrix} \widehat{w}(1) & \widehat{w}(2) & \cdots & \widehat{w}(T-\ell) \\ \widehat{w}(2) & \widehat{w}(3) & \cdots & \widehat{w}(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ \widehat{w}(\ell+1) & \widehat{w}(\ell+2) & \cdots & \widehat{w}(T) \end{bmatrix}$$

then

$$\operatorname{rank}\big(\mathscr{H}_\ell(\widehat{w})\big) \le \ell \mathtt{p} \quad \Longrightarrow \quad \widehat{w} \in \widehat{\mathscr{B}},\ \operatorname{lag}(\widehat{\mathscr{B}}) \le \ell,\ \mathbf{m}(\widehat{\mathscr{B}}) \le \mathtt{m}$$

so that

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

Applications of structured low-rank approximation in signal processing

We will consider the following signal processing problems:

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

Linear prediction problem

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

$$w(t) = p_1 w(t-1) + p_2 w(t-2) + \cdots + p_\ell w(t-\ell) \qquad \text{(LP)}$$

p_i are the linear prediction coefficients

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

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

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 p_i and $w(-\ell+1), \dots, w(0)$, s.t. the solution of (LP) coincides with (SDE) and vice verse.

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

Linear prediction problem as low-rank approx.

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

$$p_0 w(t) + p_1 w(t+1) + \cdots + p_\ell w(t+\ell) = 0, \quad \text{for } t = 1, \dots, T-\ell$$

Written in a matrix form these equations are

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_\ell \end{bmatrix} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}_\ell(w)} = 0$$

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

$$\operatorname{rank}\big(\mathscr{H}_\ell(w)\big) \le \ell$$

Harmonic retrieval problem

Corresponds to modelling w as a sum-of-exponentials

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

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

$$w(t-\ell) = p_1 w(t-\ell+1) + p_2 w(t-\ell+2) + \cdots + p_\ell w(t) \quad \text{(LP')}$$

were p_i are the linear prediction coefficients.

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

Harmonic retrieval problem as low-rank appr.

$w = \big(w(1),\dots,w(T)\big)$ sum-of-exponentials $\implies$ w satisfies

forward LP equation: $p\mathscr{H}_\ell(w) = 0$

and backward LP equation

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_\ell \end{bmatrix} \underbrace{\begin{bmatrix} w(\ell+1) & w(\ell+2) & \cdots & w(T) \\ \vdots & \vdots & & \vdots \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ w(1) & w(2) & \cdots & w(T-\ell) \end{bmatrix}}_{\mathscr{T}_\ell(w)} = 0$$

$\implies p\big(\mathscr{H}_\ell(w) + \mathscr{T}_\ell(w)\big) = 0$, *i.e.*,

$$\operatorname{rank}\big(\mathscr{H}_\ell(w) + \mathscr{T}_\ell(w)\big) \leq \ell$$

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

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.

Unstructured low-rank approximation

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

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

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

An optimal low-rank approximate solution is

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

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

Structured low-rank approximation

No closed form solution is known for the general problem

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

($\mathscr{S} : \mathbb{R}^\bullet \mapsto \mathbb{R}^{\bullet\times\bullet}$ defines the structure, *e.g.*, Hankel, Sylvester, ...)

NP-hard, consider solution methods based on local optimisation

Representing the constraint in a kernel form, the problem is

$$\min_{p,\ pp^\top=1} \left(\min_{\widehat{w}} \|w_{\mathrm{d}} - \widehat{w}\| \quad \text{subject to} \quad p\mathscr{S}(\widehat{w}) = 0 \right)$$

Double minimisation with bilinear equality constraint.

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

Variable projection vs. alternating projections

Two ways to approach the double minimisation:

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

$$\min_{p,\ pp^\top=1} p\mathscr{S}(w_{\mathrm{d}})\left(G^\top(p)G(p)\right)^{-1}\mathscr{S}^\top(w_{\mathrm{d}})p^\top$$

  $\rightsquigarrow$ a nonlinear least squares problem for p 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

$$\min_{p,\ pp^\top=1} p\mathscr{S}(w_{\mathrm{d}})\left(G^\top(p)G(p)\right)^{-1}\mathscr{S}^\top(w_{\mathrm{d}})p^\top$$

To evaluate the cost function we need to solve the linear system

$$\left(G^\top(p)G(p)\right)x = \left(p\mathscr{S}(w_{\mathrm{d}})\right)^\top$$

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

Fact: For Sylvester, Toeplitz, Hankel, Toeplitz+Hankel $\mathscr{S}$, $G^\top(p)G(p)$ is banded-Toeplitz

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

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

$$\mathscr{S}(w) = \mathscr{H}_\ell(w)$$

What matrix G satisfies

$$p\mathscr{H}_\ell(w) = 0 \quad \Longleftrightarrow \quad wG(p) = 0$$

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

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

$$\begin{bmatrix} p_0 & p_1 & \cdots & p_\ell \end{bmatrix} \underbrace{\begin{bmatrix} w(1) & w(2) & \cdots & w(T-\ell) \\ w(2) & w(3) & \cdots & w(T-\ell+1) \\ \vdots & \vdots & & \vdots \\ w(\ell+1) & w(\ell+2) & \cdots & w(T) \end{bmatrix}}_{\mathscr{H}_\ell(w)}$$

$$= \begin{bmatrix} w_1 & w_2 & \cdots & w_T \end{bmatrix} \underbrace{\begin{bmatrix} p_0 & & & \\ p_1 & p_0 & & \\ \vdots & p_1 & \ddots & \\ p_\ell & \vdots & \ddots & p_0 \\ & p_\ell & & p_1 \\ & & \ddots & \vdots \\ & & & p_\ell \end{bmatrix}}_{G(p)},$$

Note that

$$G(p) = S_{T-\ell}(p)$$

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

Therefore,

$$G^\top G = \begin{bmatrix} p_0 & p_1 & \cdots & p_\ell & & & \\ & p_0 & p_1 & \cdots & p_\ell & & \\ & & \ddots & \ddots & & \ddots & \\ & & & p_0 & p_1 & \cdots & p_\ell \end{bmatrix} \begin{bmatrix} p_0 & & & \\ p_1 & p_0 & & \\ \vdots & p_1 & \ddots & \\ p_\ell & \vdots & \ddots & p_0 \\ & p_\ell & & p_1 \\ & & \ddots & \vdots \\ & & & p_\ell \end{bmatrix}$$

(All missing elements are zeros.)

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

$$G^\top G = \begin{bmatrix} \sum_{i=0}^{\ell} p_i p_i & \sum_{i=1}^{\ell} p_i p_{i-1} & \cdots & p_\ell p_0 & & & \\ \sum_{i=1}^{\ell} p_{i-1} p_i & \ddots & & & \ddots & & \\ \vdots & & \ddots & & & \ddots & \\ p_0 p_\ell & & & \ddots & & & p_\ell p_0 \\ & \ddots & & & \ddots & & \vdots \\ & & \ddots & & & \ddots & \sum_{i=1}^{\ell} p_i p_{i-1} \\ & & & p_0 p_\ell & \cdots & \sum_{i=1}^{\ell} p_{i-1} p_i & \sum_{i=0}^{\ell} p_i p_i \end{bmatrix}$$

banded Toeplitz, bandwidth $2\ell+1$

Special case: approximate GCD

Remind the following equivalences:

$$
\begin{array}{c}
p \in \mathbb{R}[z] \text{ and } q \in \mathbb{R}[z] \\
\text{have common divisor} \\
c \in \mathbb{R}[z],\ \deg(c) = \ell_c
\end{array}
\iff
\begin{array}{c}
\exists\, a \in \mathbb{R}[z],\ \deg(a) = \ell_p - \ell_c \\
\exists\, b \in \mathbb{R}[z],\ \deg(b) = \ell_q - \ell_c \\
\text{such that } p = ca \text{ and } q = cb
\end{array}
$$

$$\iff \quad qa - pb = 0$$

$$\iff \quad \begin{bmatrix} S_{\ell_a}(q) & S_{\ell_b}(p) \end{bmatrix} \begin{bmatrix} a \\ -b \end{bmatrix} = 0$$

$$\iff \quad \begin{bmatrix} S_{\ell_a}(q) & S_{\ell_b}(p) \end{bmatrix} \text{ is rank deficient}$$

We have the following equivalent problems for approx. GCD

Sylvester structured low-rank approximation:

$$\min_{\widehat{p},\widehat{q}} \left\| \operatorname{col}(p,q) - \operatorname{col}(\widehat{p},\widehat{q}) \right\|_2 \quad \text{subject to}$$

$$\operatorname{rank}\left(\begin{bmatrix} S_{\ell_p-\ell_c}(\widehat{q}) & S_{\ell_q-\ell_c}(\widehat{p}) \end{bmatrix} \right) \leq \ell_p + \ell_q - 2\ell_c + 1.$$

$$\Updownarrow$$

Adding auxiliary variables a and b:

$$\min_{\widehat{p},\widehat{q},a,b} \left\| \begin{bmatrix} p \\ q \end{bmatrix} - \begin{bmatrix} \widehat{p} \\ \widehat{q} \end{bmatrix} \right\|_2 \quad \text{subject to} \quad \begin{bmatrix} S_{\ell_p-\ell_c}(\widehat{q}) & S_{\ell_q-\ell_c}(\widehat{p}) \end{bmatrix} \begin{bmatrix} a \\ -b \end{bmatrix} = 0$$

$$\text{and} \quad \| \operatorname{col}(a,b) \| = 1 \quad (\text{ or } b \text{ monic})$$

$$\Updownarrow$$

Adding auxiliary variables a, b, and c:

$$\min_{\widehat{p},\widehat{q},a,b,c} \left\| \begin{bmatrix} p \\ q \end{bmatrix} - \begin{bmatrix} \widehat{p} \\ \widehat{q} \end{bmatrix} \right\|_2 \quad \text{subject to} \quad \begin{bmatrix} \widehat{p} \\ \widehat{q} \end{bmatrix} = \begin{bmatrix} S_{\ell_c}(a) \\ S_{\ell_c}(b) \end{bmatrix} c \qquad (*)$$

Theorem Problem $(*)$ is equivalent to

$$\min_{c \in \mathbb{R}^{\ell_c}} \operatorname{trace}\left(\begin{bmatrix} p^\top \\ q^\top \end{bmatrix} \left(I - S\left(\left[\begin{smallmatrix} c \\ 1 \end{smallmatrix}\right]\right) \left(S^\top\left(\left[\begin{smallmatrix} c \\ 1 \end{smallmatrix}\right]\right) S\left(\left[\begin{smallmatrix} c \\ 1 \end{smallmatrix}\right]\right) \right)^{-1} S^\top\left(\left[\begin{smallmatrix} c \\ 1 \end{smallmatrix}\right]\right) \right) \begin{bmatrix} p & q \end{bmatrix} \right).$$

Proof (assuming $\ell_p = \ell_q = \ell$): Rewrite the constraint of $(*)$ as

$$\begin{bmatrix} \widehat{p} \\ \widehat{q} \end{bmatrix} = \begin{bmatrix} S_{\ell_c}(a) \\ S_{\ell_c}(b) \end{bmatrix} c \quad \iff \quad \begin{bmatrix} \widehat{p} & \widehat{q} \end{bmatrix} = S_{\ell-\ell_c}(c) \begin{bmatrix} a & b \end{bmatrix}$$

so that

$$\left\| \begin{bmatrix} p \\ q \end{bmatrix} - \begin{bmatrix} \widehat{p} \\ \widehat{q} \end{bmatrix} \right\|_2 = \left\| \begin{bmatrix} p & q \end{bmatrix} - S_{\ell-\ell_c}(c) \begin{bmatrix} u & v \end{bmatrix} \right\|_{\mathrm{F}}$$

$(*)$ becomes an ordinary least-squares problem in u, v

$\rightsquigarrow$ closed form expression in c

Notes on the theorem

- eliminates in $(*)$ the var. $\widehat{p}$, $\widehat{q}$, a, b and the constraint
- gives standard nonlinear least squares problem
- locally optimal solution can be obtained by any local optimization method (*e.g.*, Levenberg–Marquardt)
- the optimization variable is a vector of dimension ℓ_c
- cost function evaluations: solve a structured LS problem
- exploiting structure, comput. complexity per iteration $O(n)$

Suboptimal initial approximations

can be computed from unstructured low rank approx. (SVD) of

1. Sylvester matrix $S(p,q)$
2. Bezout matrix $B(p,q)$
3. Hankel matrix $H(h)$
4. Balanced model reduction

$$B(p,q) := \begin{bmatrix} p_1 & \cdots & p_n \\ \vdots & \iddots & \\ p_n & & 0 \end{bmatrix} \begin{bmatrix} q_0 & \cdots & q_{n-1} \\ & \ddots & \vdots \\ 0 & & q_{n-1} \end{bmatrix} - \begin{bmatrix} q_1 & \cdots & q_n \\ \vdots & \iddots & \\ q_n & & 0 \end{bmatrix} \begin{bmatrix} p_0 & \cdots & p_{n-1} \\ & \ddots & \vdots \\ 0 & & p_{n-1} \end{bmatrix}$$

$$H(h) := \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & h_3 & \iddots & h_{n+1} \\ \vdots & \iddots & \iddots & \vdots \\ h_n & h_{n+1} & \cdots & h_{2n} \end{bmatrix}, \qquad \frac{q(z)}{p(z)} = \sum_{t=0}^{\infty} h_t z^{-t-1}$$

Adapted version for approximate GCD with known multiplicity structure

Problem: Find an optimal approx. GCD that has the form

$$c(z) = (z - z_1)^{\ell_1} \cdots (z - z_k)^{\ell_k}, \qquad (**)$$

where $\ell_1, \dots, \ell_k$ are given and $z_1, \dots, z_k$ are to-be-determined.

More specifically, we add $(**)$ as a constraint in $(*)$.

The local optimization method becomes:

$$\min_{z \in \mathbb{C}^k} \operatorname{trace}\left(\begin{bmatrix} p^\top \\ q^\top \end{bmatrix} \Big(I - S(\mathbf{c}(z)) \big(S^\top(\mathbf{c}(z)) S(\mathbf{c}(z)) \big)^{-1} S^\top(\mathbf{c}(z)) \Big) \begin{bmatrix} p & q \end{bmatrix} \right)$$

where $\mathbf{c} : \mathbb{C}^k \to \mathbb{R}^{\ell_1 + \cdots + \ell_k + 1}$ is the mapping $z \mapsto c$, given by $(**)$.

Conclusions

- Mathematical models are subsets of the data space.
  Linear models are subspaces.
- LTI model representations: kernel, image, input/output.
  The parameter (identifying the model) is a polynomial matrix.
- Polynomial multiplication $\leftrightarrow$ Toeplitz matrix $\times$ vector
  Approximate GCD computation $\leftrightarrow$ Sylvester low-rank approx.
  Approximate LTI modelling $\leftrightarrow$ Hankel low-rank approx.

Conclusions

- Applications in signal processing
  - Linear prediction $\leftrightarrow$ sum-of-damped-exp. modelling,
  - Harmonic retrieval $\leftrightarrow$ sum-of-exp. modelling,
  - deconvolution $\leftrightarrow$ FIR modelling.
- Algorithms based on local optimisation
  Efficient cost function evaluation exploiting the structure.

LTI systems	$\leftrightarrow$	polynomials	$\leftrightarrow$	structured matrices
system theory	$\leftrightarrow$	abstract algebra	$\leftrightarrow$	numerical linear algebra