Globally Optimal Parameter Estimation of Nonlinear Dynamical Models is an Eigenvalue Problem

43^{rd} Benelux Meeting on Systems and Control
Blankenberge, Belgium

Sarthak De
sarthak.de@esat.kuleuven.be

Bart De Moor
bart.demoor@esat.kuleuven.be

ESAT-STADIUS
KU Leuven, Belgium

March 27,2024

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Overview

Overview

1. Problem Statement
2. Output Difference Equation
3. Parameter Estimation
4. References

Model Class

- Single output autonomous discrete–time polynomial state–space models of the form Σ

$$\Sigma : \begin{cases} \hat{\boldsymbol{x}}_{k+1} = \boldsymbol{f}(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}) \\ \hat{y}_k = g(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}) \end{cases} \tag{1}$$

where $\hat{\boldsymbol{x}}_k \in \mathbb{R}^n$ are the state variables at instant $k : k \in \mathbb{Z}^+$, $\boldsymbol{\theta} \in \mathbb{R}^\ell$ are the model parameters. $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^\ell \to \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^\ell \to \mathbb{R}$ and $\boldsymbol{f}, g \in \mathbb{R}[\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}]$ where $\boldsymbol{f}, g \in \mathbb{R}[\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}]$ is a multivariate polynomial ring, $\hat{y} \in \mathbb{R}$ is the model output variable.

Model Class

- Single output autonomous discrete–time polynomial state–space models of the form Σ

$$\Sigma : \begin{cases} \hat{\boldsymbol{x}}_{k+1} = \boldsymbol{f}(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}) \\ \hat{y}_k = g(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}) \end{cases} \tag{1}$$

  where $\hat{\boldsymbol{x}}_k \in \mathbb{R}^n$ are the state variables at instant $k : k \in \mathbb{Z}^+$, $\boldsymbol{\theta} \in \mathbb{R}^\ell$ are the model parameters. $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}$ and $\boldsymbol{f}, g \in \mathbb{R}[\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}]$ where $\boldsymbol{f}, g \in \mathbb{R}[\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}]$ is a multivariate polynomial ring, $\hat{y} \in \mathbb{R}$ is the model output variable.
- Σ is identifiable, i.e., a model compliant trajectory $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_0 & \hat{y}_1 & \dots \hat{y}_{\mathsf{N}-1} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{\mathsf{N}}$ is generated uniquely by some $\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^\ell$, where Θ is an open neighborhood of $\mathbb{R}^\ell$

Parameter Estimation

- Given observed output sequence $\mathbf{y} = \begin{bmatrix} y_0 & y_1 & \dots y_{N-1} \end{bmatrix}^T \in \mathbb{R}^N$, find a model compliant trajectory $\hat{\mathbf{y}}$ generated by some $\boldsymbol{\theta}^*$, such that the observed data is 'closely' approximated.

Parameter Estimation

- Given observed output sequence $\mathbf{y} = \begin{bmatrix} y_0 & y_1 & \dots y_{N-1} \end{bmatrix}^T \in \mathbb{R}^N$, find a model compliant trajectory $\hat{\mathbf{y}}$ generated by some $\boldsymbol{\theta}^*$, such that the observed data is 'closely' approximated.
- Minimize

$$\begin{aligned} \min_{\boldsymbol{\theta},\hat{\mathbf{y}}} \quad & \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \\ \text{s.t.} \quad & \boldsymbol{\Phi}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = \mathbf{0} \end{aligned} \tag{2}$$

  where $\boldsymbol{\Phi}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = \mathbf{0}$, is a system of polynomial equations, such that $\mathcal{V}(\boldsymbol{\Phi})$, forms a manifold in $\mathbb{R}^\ell \times \mathbb{R}^N$ on which the model-compliant trajectory $\hat{\mathbf{y}}$ and the corresponding model parameters $\boldsymbol{\theta}$ lie.
- Opt. problem in (2) is non-convex

Parameter Estimation some history

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]
- In discrete-time models

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]
- In discrete-time models
  - output equation $\rightarrow$ prediction error model $\rightarrow$ recursive least squares $\rightarrow$ biased estimates (sub-optimal approach)

    [Billings and Voon, 1984]

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]
- In discrete-time models
  - output equation $\rightarrow$ prediction error model $\rightarrow$ recursive least squares $\rightarrow$ biased estimates (sub-optimal approach)

    [Billings and Voon, 1984]
  - output equation $\rightarrow$ equation error cost function $\rightarrow$ total least squares $\rightarrow$ ignores non linear relations (sub-optimal approach)

    [Lu and Chon, 2003]

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]
- In discrete-time models
  - output equation $\rightarrow$ prediction error model $\rightarrow$ recursive least squares $\rightarrow$ biased estimates (sub-optimal approach)

    [Billings and Voon, 1984]
  - output equation $\rightarrow$ equation error cost function $\rightarrow$ total least squares $\rightarrow$ ignores non linear relations (sub-optimal approach)

    [Lu and Chon, 2003]
  - output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution)

    [Lu and Chon, 2003, Chon and Cohen, 1997]

Parameter Estimation some history

- In continuous-time models
  - polynomial state-space model $\rightarrow$ output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution) [Denis-Vidal et al., 2003, Verdiere, 2005]
- In discrete-time models
  - output equation $\rightarrow$ prediction error model $\rightarrow$ recursive least squares $\rightarrow$ biased estimates (sub-optimal approach)

    [Billings and Voon, 1984]
  - output equation $\rightarrow$ equation error cost function $\rightarrow$ total least squares $\rightarrow$ ignores non linear relations (sub-optimal approach)

    [Lu and Chon, 2003]
  - output equation $\rightarrow$ least squares cost function $\rightarrow$ iterative gradient based solver (sub-optimal solution)

    [Lu and Chon, 2003, Chon and Cohen, 1997]

Globally Optimal Parameter Estimation

polynomial state-space model $\rightarrow$ **output equation** $\rightarrow$ least-squares cost function $\rightarrow$ **system of polynomial equations** $\rightarrow$ **eigenvalue problem**

Overview

Output difference equation

The output difference equation

$$\Phi(\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k) = 0 \tag{3}$$

relates the consecutive samples of a model compatible output sequence, and is a multivariate polynomial equation, such that

$$\Phi \in I_\Sigma \cap \mathbb{R}\left[\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k\right] \tag{4}$$

where

$$\begin{aligned} I_\Sigma = \langle & \hat{y}_k - g(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}), \\ & \hat{y}_{k+1} - g(\boldsymbol{f}(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}), \boldsymbol{\theta}), \\ & \vdots \\ & \hat{y}_{k+n} - g(\boldsymbol{f}^n(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}), \boldsymbol{\theta}) > . \end{aligned} \tag{5}$$

note that $I_\Sigma \in \mathbb{R}\left[\hat{\boldsymbol{x}}_k, \boldsymbol{\theta}, \hat{y}_k, \ldots, \hat{y}_{k+n}\right]$

Output difference equation

Construction of Φ

Given a single-output autonomous discrete-time (DT) polynomial state-space model Σ of order n (as in (1)), there exists a unique polynomial output difference equation of minimal degree, denoted by $\Phi(\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k) = 0$, which is the generator of the elimination ideal $I_\Sigma|_n \in \mathbb{R}[\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k]$ where

$$\Phi \in \underbrace{I_\Sigma \cap \mathbb{R}\left[\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k\right]}_{I_\Sigma|_n}$$

Moreover, Φ is of the same model order n and encapsulates the dynamical behavior of Σ in a single equation.

Output difference equation
Example

Consider the discretized Lotka–Volterra model

$$\Sigma_{\mathrm{LV}} : \begin{cases} \hat{x}^{(1)}_{k+1} = \hat{x}^{(1)}_k (1 + b - p\hat{x}^{(2)}_k) \\ \hat{x}^{(2)}_{k+1} = \hat{x}^{(2)}_k (1 - d + p\hat{x}^{(1)}_k) \\ \hat{y}_k = \hat{x}^{(2)}_k \end{cases} \tag{6}$$

where the superscript over $\hat{x}^{(.)}_k$ indicates the component of the state-variable.

Output difference equation
Example

Consider the discretized Lotka–Volterra model

$$\Sigma_{\mathrm{LV}}: \begin{cases} \hat{x}^{(1)}_{k+1} = \hat{x}^{(1)}_{k}(1 + b - p\hat{x}^{(2)}_{k}) \\ \hat{x}^{(2)}_{k+1} = \hat{x}^{(2)}_{k}(1 - d + p\hat{x}^{(1)}_{k}) \\ \hat{y}_k = \hat{x}^{(2)}_{k} \end{cases} \tag{6}$$

where the superscript over $\hat{x}^{(.)}_{k}$ indicates the component of the state-variable. Here, $I_{\Sigma_{\mathrm{LV}}}$ can be generated using,

$$I_{\Sigma_{\mathrm{LV}}} = \begin{array}{c} < \hat{y}_k - \hat{x}^{(2)}_{k}, \\ \hat{y}_{k+1} - \hat{x}^{(2)}_{k}(1 - d + p\hat{x}^{(1)}_{k}), \\ \hat{y}_{k+2} - \hat{x}^{(2)}_{k}(1 - d + p\hat{x}^{(1)}_{k})(1 - d + p\hat{x}^{(1)}_{k}(1 + b - p\hat{x}^{(2)}_{k})) > \end{array} \tag{7}$$

In order to eliminate $\boldsymbol{x}_k$ we will use consecutive Sylvester resultants

Sylvester Matrix and Resultants

Consider the system,

$$\begin{cases} f_1(x) = a_r x^r + a_{r-1}x^{r-1} + \ldots + a_0 = 0, \\ f_2(x) = b_s x^s + b_{s-1}x^{s-1} + \ldots + b_0 = 0 \end{cases}$$

which has common roots. Construct $\boldsymbol{Mk} = \mathbf{0}$ by multiplying $f_1(x)$ and $f_2(x)$ with powers of x s.t.,

$$\begin{matrix} s \text{ rows}: \\ \\ r \text{ rows}: \end{matrix} \left[\begin{array}{cccccccc} a_0 & a_1 & \ldots & a_r & & & & \\ & a_0 & a_1 & \ldots & a_r & & & \\ & & \ddots & \ddots & & \ddots & & \\ & & & a_0 & a_1 & \ldots & a_r & \\ \hline b_0 & b_1 & \ldots & b_s & & & & \\ & b_0 & b_1 & \ldots & b_s & & & \\ & & \ddots & \ddots & & \ddots & & \\ & & & b_0 & b_1 & \ldots & b_s & \end{array}\right] \begin{bmatrix} x^0 \\ x^1 \\ \vdots \\ x^{r+s-1} \end{bmatrix} = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \end{array}\right].$$

$\boldsymbol{M} \in \mathbb{R}^{(r+s)\times(r+s)}$ is the **Sylvester matrix** [Cox et al., 2015]

Sylvester Matrix and Resultants

Consider the system,

$$\begin{cases} f_1(x) = a_r x^r + a_{r-1}x^{r-1} + \ldots + a_0 = 0, \\ f_2(x) = b_s x^s + b_{s-1}x^{s-1} + \ldots + b_0 = 0 \end{cases}$$

which has common roots. Construct $\boldsymbol{Mk} = \mathbf{0}$ by multiplying $f_1(x)$ and $f_2(x)$ with powers of x s.t.,

$$\begin{array}{l} s \text{ rows}: \left\{ \vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}} \right. \\ r \text{ rows}: \left\{ \vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}} \right. \end{array} \left[\begin{array}{cccccccc} a_0 & a_1 & \ldots & a_r & & & & \\ & a_0 & a_1 & \ldots & a_r & & & \\ & & \ddots & \ddots & & \ddots & & \\ & & & a_0 & a_1 & \ldots & a_r \\ \hline b_0 & b_1 & \ldots & b_s & & & & \\ & b_0 & b_1 & \ldots & b_s & & & \\ & & \ddots & \ddots & & \ddots & & \\ & & & b_0 & b_1 & \ldots & b_s \end{array} \right] \begin{bmatrix} x^0 \\ x^1 \\ \vdots \\ x^{r+s-1} \end{bmatrix} = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ \hline 0 \\ 0 \\ \vdots \\ 0 \end{array} \right].$$

$\boldsymbol{M} \in \mathbb{R}^{(r+s)\times(r+s)}$ is the **Sylvester matrix** [Cox et al., 2015]

Sylvester Resultant

if $f_1(x)$ and $f_2(x)$ have a common root, then $\det \boldsymbol{M} = 0$

$$\mathrm{Res}(f_1, f_2, x) = \det(\mathrm{Syl}(f_1, f_2, x))$$

Output difference equation
Example

Let us first eliminate the state variable $\hat{x}_k^{(2)}$, First, lets consider

$$\text{Res} : \begin{cases} f_1(\hat{x}_k^{(2)}) = \hat{y}_k - \hat{x}_k^{(2)} \\ f_2(\hat{x}_k^{(2)}) = \hat{y}_{k+1} - (1 - d + p\hat{x}_k^{(1)})\hat{x}_k^{(2)} \end{cases}$$

we can construct the Sylvester matrix as,

$$\underbrace{\begin{bmatrix} \hat{y}_k & -1 \\ \hat{y}_{k+1} & -(1 - d + p\hat{x}_k^{(1)}) \end{bmatrix}}_{\text{Syl}(f_1, f_2, \hat{x}_k^{(2)})} \begin{bmatrix} 1 \\ \hat{x}_k^{(2)} \end{bmatrix} = \mathbf{0}$$

$$\text{Res}(f_1, f_2, \hat{x}_k^{(2)}) = \det(\text{Syl}(f_1, f_2, \hat{x}_k^{(2)})) = \hat{y}_k(1 - d + p\hat{x}_k^1) - \hat{y}_{k+1}$$

Output difference equation
Example

Now, lets consider

$$\text{Res}: \begin{cases} f_2(\hat{x}_k^{(2)}) = \hat{y}_{k+1} - (1 - d + p\hat{x}_k^{(1)})\hat{x}_k^{(2)} \\ f_3(\hat{x}_k^{(2)}) = \hat{y}_{k+2} + f_{31}(b, d, p, \hat{x}_k^{(1)})\hat{x}_k^{(2)} + f_{32}(b, d, p, \hat{x}_k^{(1)})(x\hat{(2)}_k)^2 \end{cases}$$

we can construct the Sylvester matrix as,

$$\underbrace{\begin{bmatrix} \hat{y}_{k+1} & -(1 - d + p\hat{x}_k^{(1)}) & 0 \\ 0 & \hat{y}_{k+1} & -(1 - d + p\hat{x}_k^{(1)}) \\ \hat{y}_{k+2} & f_{31}(b, d, p, \hat{x}_k^{(1)}) & f_{32}(b, d, p, \hat{x}_k^{(1)}) \end{bmatrix}}_{\text{Syl}(f_2, f_3, \hat{x}_k^{(2)})} \begin{bmatrix} 1 \\ \hat{x}_k^{(2)} \\ (\hat{x}_k^{(2)})^2 \end{bmatrix} = \mathbf{0}$$

$$\text{Res}(f_2, f_3, \hat{x}_k^{(2)}) = \det(\text{Syl}(f_2, f_3, \hat{x}_k^{(2)}))$$

Output difference equation

Observe

$$\mathrm{Res}(f_1, f_2, \hat{x}_k^{(2)}), \mathrm{Res}(f_2, f_3, \hat{x}_k^{(2)}) \in \mathbb{R}[b, d, p, x_k^{(1)}, \hat{y}_k, \hat{y}_{k+1}, \hat{y}_{k+2}]$$

eliminate $\hat{x}_k^{(1)}$ by computing,

$$\Phi = \mathrm{Res}(\mathrm{Res}(f_1, f_2, \hat{x}_k^{(2)}), \mathrm{Res}(f_2, f_3, \hat{x}_k^{(2)}), \hat{x}_k^{(1)}) = 0$$

which is,

$$\hat{y}_k^2 \hat{y}_{k+1}(pd - p) + \hat{y}_k \hat{y}_{k+1}^2 p - \hat{y}_k \hat{y}_{k+1}(bd - b) - \hat{y}_{k+1}^2(b+1) + \hat{y}_k \hat{y}_{k+2} = 0$$

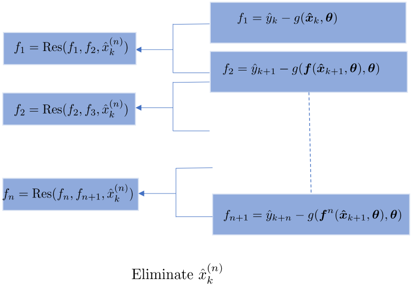
Summary: Output difference equation

$$f_1 = \hat{y}_k - g(\hat{\boldsymbol{x}}_k, \boldsymbol{\theta})$$

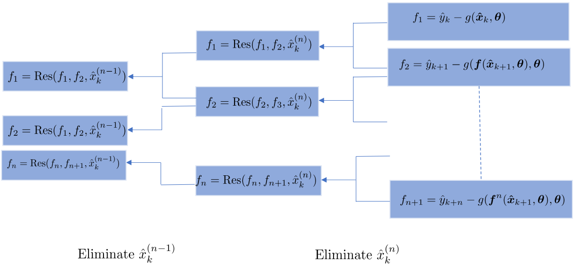
$$f_2 = \hat{y}_{k+1} - g(\boldsymbol{f}(\hat{\boldsymbol{x}}_{k+1}, \boldsymbol{\theta}), \boldsymbol{\theta})$$

$$f_{n+1} = \hat{y}_{k+n} - g(\boldsymbol{f}^n(\hat{\boldsymbol{x}}_{k+1}, \boldsymbol{\theta}), \boldsymbol{\theta})$$

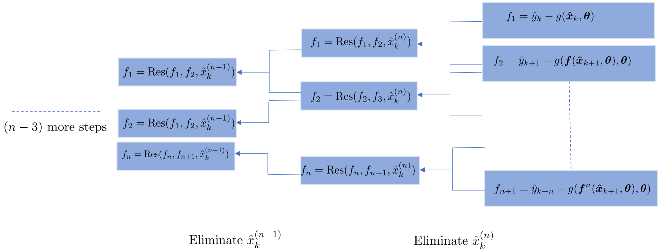
Summary: Output difference equation



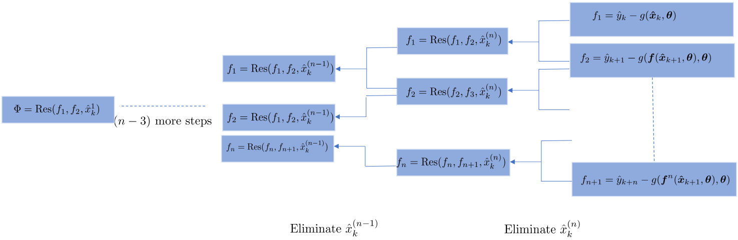
Summary: Output difference equation



Summary: Output difference equation



Summary: Output difference equation



Manifold of the Model Compliant Data

Given model compliant data,

$$\hat{\mathbf{y}} = \left[\hat{y}_0, \ldots, \hat{y}_{\text{N-1}}\right]^{\mathsf{T}} \in \mathbb{R}^{\mathsf{N}}$$

the output equation $\Phi(\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k) = 0$ is satisfied by all $\hat{y}_k$ where $k \in \mathbb{Z}^+$.

Manifold of the Model Compliant Data

Given model compliant data,

$$\hat{\mathbf{y}} = \left[\hat{y}_0, \ldots, \hat{y}_{\text{N-1}}\right]^{\mathsf{T}} \in \mathbb{R}^{\mathsf{N}}$$

the output equation $\Phi(\boldsymbol{\theta}, \hat{y}_{k+n}, \ldots, \hat{y}_k) = 0$ is satisfied by all $\hat{y}_k$ where $k \in \mathbb{Z}^+$. Consider the system of equations,

$$\boldsymbol{\Phi}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = \begin{bmatrix} \phi(\boldsymbol{\theta}, \hat{y}_0, \ldots, \hat{y}_n) \\ \phi(\boldsymbol{\theta}, \hat{y}_1, \ldots, \hat{y}_{n+1}) \\ \vdots \\ \phi(\boldsymbol{\theta}, \hat{y}_{\mathsf{N}-n-1}, \ldots, \hat{y}_{\mathsf{N}-1}) \end{bmatrix} = \mathbf{0}, \qquad (8)$$

where $\boldsymbol{\Phi} \in \mathbb{R}\left[\boldsymbol{\theta}, \hat{\mathbf{y}}\right]$ $\boldsymbol{\Phi} : \mathbb{R}^{\mathsf{N}+\ell} \rightarrow \mathbb{R}^{\mathsf{N}-n}$. $\mathcal{V}(\boldsymbol{\Phi})$ describes the positive dimensional variety over which the model compatible data and the associated model parameter lie.

Overview

Parameter Estimation as system of polynomial equations

Minimize the misfit as

$$\begin{aligned} \min_{\theta, \hat{\mathbf{y}}} \quad & \frac{1}{2}\|\tilde{\mathbf{y}}\|_2^2 = \frac{1}{2}\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 , \\ \text{s.t.} \quad & \boldsymbol{\Phi}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = 0 \end{aligned}$$

Parameter Estimation as system of polynomial equations

Minimize the misfit as

$$\begin{aligned} \min_{\theta, \hat{\mathbf{y}}} \quad & \frac{1}{2} \|\tilde{\mathbf{y}}\|_2^2 = \frac{1}{2} \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 \,, \\ \text{s.t.} \quad & \mathbf{\Phi}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = 0 \end{aligned}$$

The Lagrangian is,

$$\mathcal{L}(\boldsymbol{\theta}, \hat{\mathbf{y}}) = \frac{1}{2} ||\mathbf{y} - \hat{\mathbf{y}}||_2^2 + \boldsymbol{\lambda}^\mathsf{T} \mathbf{\Phi} \tag{9}$$

Parameter Estimation as system of polynomial equations

Minimize the misfit as

$$\min_{\theta,\hat{\mathbf{y}}} \quad \frac{1}{2}\|\tilde{\mathbf{y}}\|_2^2 = \frac{1}{2}\|\mathbf{y}-\hat{\mathbf{y}}\|_2^2\,,$$
$$\text{s.t.} \quad \mathbf{\Phi}(\boldsymbol{\theta},\hat{\mathbf{y}}) = 0$$

The Lagrangian is,

$$\mathcal{L}(\boldsymbol{\theta},\hat{\mathbf{y}}) = \frac{1}{2}||\mathbf{y}-\hat{\mathbf{y}}||_2^2 + \boldsymbol{\lambda}^{\mathsf{T}}\mathbf{\Phi} \tag{9}$$

The associated FONCs are,

$$\partial\mathcal{L}/\partial\hat{\mathbf{y}} = -(\mathbf{y}-\hat{\mathbf{y}}) + \left(\frac{\partial\mathbf{\Phi}}{\partial\hat{\mathbf{y}}}\right)^{\mathsf{T}}\boldsymbol{\lambda} = 0, \tag{10}$$

$$\partial\mathcal{L}/\partial\boldsymbol{\theta} = \left(\frac{\partial\mathbf{\Phi}}{\partial\boldsymbol{\theta}}\right)^{\mathsf{T}}\boldsymbol{\lambda} = 0 \tag{11}$$

$$\partial\mathcal{L}/\partial\boldsymbol{\lambda} = \mathbf{\Phi} = 0 \tag{12}$$

Here, $\boldsymbol{\lambda}\in\mathbb{R}^{\mathsf{N}-n}$, and (10)- (12) is a square system of polynomial equations with $(\mathsf{N})+\ell+(\mathsf{N}-n)$ equations.

Parameter Estimation as system of polynomial equations

Consider the FONC in (11),

$$\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}} \boldsymbol{\lambda} = 0$$

here, $\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{(N-n)\times \ell}$

Parameter Estimation as system of polynomial equations

Consider the FONC in (11),

$$\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}} \boldsymbol{\lambda} = 0$$

here, $\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{(\mathsf{N}-n)\times \ell}$ we know from [Nõmm and Moog, 2016], if the model Σ is identifiable then,

$$\operatorname{rank}\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}}\right) = \ell \tag{13}$$

Parameter Estimation as system of polynomial equations

Consider the FONC in (11),

$$\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}} \boldsymbol{\lambda} = 0$$

here, $\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}} \in \mathbb{R}^{(\mathsf{N}-n)\times \ell}$ we know from [Nõmm and Moog, 2016], if the model Σ is identifiable then,

$$\text{rank}\left(\frac{\partial \mathbf{\Phi}}{\partial \boldsymbol{\theta}}\right) = \ell \tag{13}$$

and since $\boldsymbol{\lambda} \in \text{null}\left(\frac{\partial \phi}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}}$, we can write,

$$\boldsymbol{\lambda} = \boldsymbol{V}(\boldsymbol{\theta}, \hat{\mathbf{y}})\boldsymbol{c}$$

where, $\boldsymbol{V} \in \mathbb{R}^{(\mathsf{N}-n)\times(\mathsf{N}-n-\ell)}$ and $\boldsymbol{c} \in \mathbb{R}^{N-n-\ell}$. $\boldsymbol{V}$ is the basis of the nullspace of $\left(\frac{\partial \phi}{\partial \boldsymbol{\theta}}\right)^{\mathsf{T}}$ and the components of $\boldsymbol{V}$, $v_{ij} \in \mathbb{R}[\boldsymbol{\theta}, \hat{\mathbf{y}}]$

Parameter Estimation as system of polynomial equations

The FONCs can be re-written as,

$$-(\mathbf{y} - \hat{\mathbf{y}}) + \left(\frac{\partial \boldsymbol{\Phi}}{\partial \hat{\mathbf{y}}}\right)^{\mathsf{T}} \boldsymbol{V} \boldsymbol{c} = \mathbf{0} \quad (14)$$

$$\boldsymbol{\Phi} = 0 \quad (15)$$

Here, (14)- (15) is a square system with $\mathsf{N} + (\mathsf{N} - n)$ equations

Parameter Estimation: Example $n = 1$

Let's first consider a first order model with one model parameter

$$\Sigma_1 : \begin{cases} \hat{x}_{k+1} = \theta \hat{x}_k^3 \\ \hat{y}_k = \hat{x}_k \end{cases} \tag{16}$$

The output difference equation is given as,

$$\phi(\hat{y}_k, \hat{y}_{k+1}, \theta) = \hat{y}_{k+1} - \theta \hat{y}_k^3 = 0 \tag{17}$$

The parameter in θ is globally
identifiable [Nõmm and Moog, 2016], thus we can expect it to a find a unique minimizer

Parameter Estimation: Example $n = 1$

Given $\mathbf{y} = \begin{bmatrix} 1.00685 & 0.59511 & 0.02801 \end{bmatrix}^\mathsf{T}$ we can write (10)- (12) for Σ_1 as,

$$\begin{cases} \left(\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} - \begin{bmatrix} 1.00685 \\ 0.59511 \\ 0.02801 \end{bmatrix} \right) + \begin{bmatrix} -3\theta \hat{y}_0^2 & 0 \\ 1 & -3\theta \hat{y}_1^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} -\hat{y}_0^3 & -\hat{y}_1^3 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} \hat{y}_1 - \theta \hat{y}_0^3 \\ \hat{y}_2 - \theta \hat{y}_1^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{cases}$$

Parameter Estimation: Example $n = 1$

Given $\mathbf{y} = \begin{bmatrix} 1.00685 & 0.59511 & 0.02801 \end{bmatrix}^{\mathsf{T}}$ we can write (10)- (12) for Σ_1 as,

$$\begin{cases} \left(\begin{bmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} - \begin{bmatrix} 1.00685 \\ 0.59511 \\ 0.02801 \end{bmatrix} \right) + \begin{bmatrix} -3\theta\hat{y}_0^2 & 0 \\ 1 & -3\theta\hat{y}_1^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} -\hat{y}_0^3 & -\hat{y}_1^3 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \begin{bmatrix} \hat{y}_1 - \theta\hat{y}_0^3 \\ \hat{y}_2 - \theta\hat{y}_1^3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \end{cases}$$

Solving (25) using `HomotopyContinuation.jl` [Breiding and Timme, 2018], we find the globally optimal solution $(\boldsymbol{\theta}, \hat{y}_0, \hat{y}_1, \hat{y}_2) = (0.5194, 1.0228, 0.5558, 0.0891)$

Parmeter Estimation: Example $n = 1$

There exists a partial linear structure in the system (25), such that the system can be written as a multiparameter eigenvalue problem (MEVP) of the form

$$\underbrace{\begin{bmatrix} \hat{y}_0 - 1.00685 & -3\theta\hat{y}_0^2 & 0 \\ \hat{y}_1 - 0.59511 & 1 & -3\theta\hat{y}_1^2 \\ \hat{y}_2 - 0.02801 & 0 & 1 \\ 0 & -\hat{y}_0^3 & -\hat{y}_1^3 \\ \hat{y}_1 \quad -\theta\hat{y}_0^3 & 0 & 0 \\ \hat{y}_2 \quad -\theta\hat{y}_1^3 & 0 & 0 \end{bmatrix}}_{\mathcal{M}(\theta,\hat{\mathbf{y}})} \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \mathbf{0}$$

here, (23) is 4-parameter 4th degree MEVP. Using `MacaulayLab` [Vermeersch and De Moor, 2022] we find the same globally optimal solution.

Parameter Estimation: Example n=1

We can incorporate the identifiability rank condition which allows us to write $\begin{bmatrix}\lambda_1 & \lambda_2\end{bmatrix}^\mathsf{T} = \begin{bmatrix}\hat{y}_1^3 & -\hat{y}_0^3\end{bmatrix}^\mathsf{T} c$ such that the resulting system of equation is,

$$\begin{cases} \left(\begin{bmatrix}\hat{y}_0\\ \hat{y}_1\\ \hat{y}_2\end{bmatrix} - \begin{bmatrix}1.00685\\ 0.59511\\ 0.02801\end{bmatrix}\right) + \begin{bmatrix}-3\theta\hat{y}_0^2 & 0\\ 1 & -3\theta\hat{y}_1^2\\ 0 & 1\end{bmatrix} \begin{bmatrix}\hat{y}_1^3\\ -\hat{y}_0^3\end{bmatrix} c = \begin{bmatrix}0\\ 0\\ 0\end{bmatrix} \\ \begin{bmatrix}\hat{y}_1 - \theta\hat{y}_0^3\\ \hat{y}_2 - \theta\hat{y}_1^3\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \end{cases}$$

`HomotopyContinuation.jl` yields the same globally optimal solution, however the equivalent MEVP is of degree 6 which makes solving the EP using `MacaulayLab` inefficient

Parameter Estimation: Example n=1

We will now consider a special case where $\mathrm{N} = 2n + l$. Since we can write $\hat{\mathbf{y}} = \boldsymbol{f}_{\text{comp}}(\boldsymbol{x}_0, \boldsymbol{\theta})$, satisfies (15), we can substitute it in (14), resulting in a smaller system of equations $2n + \ell$ equations in $2n + \ell$ variables. For the cubic model we are already in the situation where $\mathrm{N} = 2n + \ell = 3$, the FONCs reduce to,

$$\left\{ \left(\begin{bmatrix} \hat{x}_0 \\ \theta \hat{x}_0^3 \\ \theta^4 \hat{x}_0^9 \end{bmatrix} - \begin{bmatrix} 1.00685 \\ 0.59511 \\ 0.02801 \end{bmatrix} \right) + \begin{bmatrix} -3\theta \hat{x}_0^2 & 0 \\ 1 & -3\theta^2 \hat{x}_0^6 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta^3 \hat{x}_0^9 \\ -\hat{x}_0^3 \end{bmatrix} c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right.$$

The solution is $(0.5194, 1.0228)$. Note, that the resulting system of equations is of degree 16. The final question is a Numerical one, is it better to work with more equations of lower degree OR less equations of higher degree.

Parameter Estimation: Lotka–Volterra ($n = 2$)

Consider we are given $\mathrm{N} = 6$ measured sequence from the Lotka–Volterra model (6).

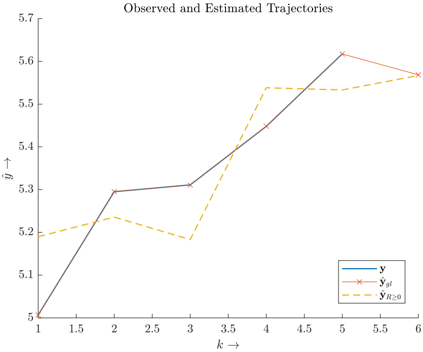


Figure: Estimation of $\mathrm{N} = 6$ datapoints where $\ell = 3$ requires a maximum 13 equations, which are of degree 5.

Overview

References I

Billings, S. A. and Voon, W. S. F. (1984).
Least squares parameter estimation algorithms for non-linear systems.
International Journal of Systems Science, 15(6):601–615.

Breiding, P. and Timme, S. (2018).
Homotopycontinuation. jl: A package for homotopy continuation in julia.
In *Mathematical Software–ICMS 2018: 6th International Conference, South Bend, IN, USA, July 24-27, 2018, Proceedings 6*, pages 458–465. Springer.

Chon, K. and Cohen, R. (1997).
Linear and nonlinear arma model parameter estimation using an artificial neural network.
IEEE Transactions on Biomedical Engineering, 44(3):168–174.

References II

Cox, D., Little, J., and O'Shea, D. (2015).
Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra.

Undergraduate Texts in Mathematics. Springer International Publishing.

Denis-Vidal, L., Joly-Blanchard, G., and Noiret, C. (2003).
System identifiability (symbolic computation) and parameter estimation (numerical computation).
Numerical Algorithms, 34:283–292.

Lu, S. and Chon, K. (2003).
Nonlinear autoregressive and nonlinear autoregressive moving average model parameter estimation by minimizing hypersurface distance.
IEEE Transactions on Signal Processing, 51(12):3020–3026.

References III

Nõmm, S. and Moog, C. H. (2016).
Further results on identifiability of discrete-time nonlinear systems.
Automatica, 68:69–74.

Verdiere, Nathalie, D.-V. L. J.-B. G. D. D. (2005).
Identifiability and estimation of pharmacokinetic parameters for the ligands of the macrophage mannose receptor.
International Journal of Applied Mathematics and Computer Science, 15(4):517–526.

Vermeersch, C. and De Moor, B. (2022).
Two complementary block macaulay matrix algorithms to solve multiparameter eigenvalue problems.
Linear Algebra and its Applications, 654:177–209.