Least-squares Globally Optimal Misfit Modelling for SISO Systems

41st Benelux Meeting on Systems and Control

Sibren Lagauw
sibren.lagauw@esat.kuleuven.be
Bart De Moor, *Fellow IEEE, SIAM*
bart.demoor@esat.kuleuven.be

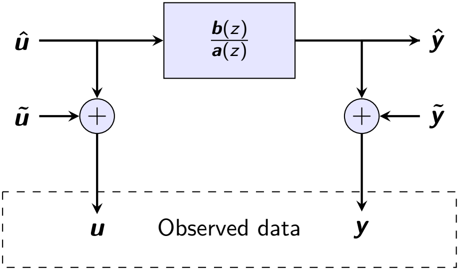
Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS),
Department of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium

July 5, 2022

STADIUS
Center for Dynamical Systems,
Signal Processing and Data Analytics

Table of Contents

Misfit modelling



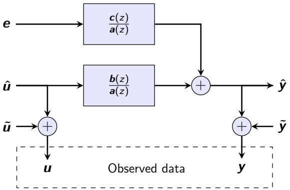
$$\min_{\tilde{\boldsymbol{u}},\tilde{\boldsymbol{y}},a_i,b_i} \quad \sigma^2 = ||\tilde{\boldsymbol{u}}||_2^2 + ||\tilde{\boldsymbol{y}}||_2^2 = \sum_{k=0}^{N-1}(u_k - \hat{u}_k)^2 + \sum_{k=0}^{N-1}(y_k - \hat{y}_k)^2,$$

$$\text{s.t.} \quad \hat{y}_{k+n} + a_1\hat{y}_{k+n-1} + \ldots + a_n\hat{y}_k = b_0\hat{u}_{k+n} + b_1\hat{u}_{k+n-1} + \ldots + b_n\hat{u}_k \quad \forall k \in \{0, \ldots, N-n-1\},$$

with $a_i, b_i \in \mathbb{R}, i = 1, \ldots, n$ the coefficients of $\boldsymbol{b}(z)/\boldsymbol{a}(z)$.

Misfit modelling: context

- Special case within misfit vs. latency framework *(Lemmerling and De Moor [4])*



- Subspace identification *(Van Overschee and De Moor [6])*
  - Linear algebra framework (QR, SVD)
  - Identification: Kalman state estimation + linear least-squares

- Behavioural systems theory *(Roorda [5], Willems [7])*
  - Set of difference equations $\rightarrow$ behaviour $\mathcal{B}$: $\hat{\boldsymbol{y}} \in \mathcal{B}$, $\tilde{\boldsymbol{y}} \in \mathcal{B}^{\perp}$
  - Identification: (Structured) total least squares *(De Moor [2])*

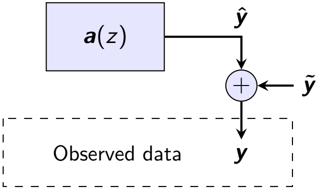
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

Misfit modelling: methodology

- Inner problem: find optimal misfit $\tilde{\boldsymbol{y}}^*$ for given model and data
  - Linear problem $\rightarrow$ orthogonal projection
  - Optimal misfit is structured
- Outer problem: find optimal model $(\boldsymbol{a}, \boldsymbol{b})^*$ for given data
  - Substitute $\boldsymbol{y}^*$ in objective function
  - Characterise optima $\rightarrow$ system of multivariate polynomial eqs.
  - Solve the polynomial root-finding problem

Table of Contents

Autonomous misfit modelling[1]



$$\min_{\boldsymbol{a},\hat{\boldsymbol{y}}} \quad \sum_{k=0}^{N-1} (y_k - \hat{y}_k)^2,$$
$$\text{s.t.} \quad \hat{\boldsymbol{y}}\, a(z) = 0.$$

- LS-criterion implies orthogonality:

$$\sum_{k=0}^{N-1} (y_k - \hat{y}_k)^2 \text{ is minimal } \iff \hat{\boldsymbol{y}} \perp \tilde{\boldsymbol{y}}$$

[1] See *(De Moor [3])* for an in-depth discussion.

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

Autonomous misfit modelling (inner)

- Canonical state-space realization $(\boldsymbol{A}, \boldsymbol{C})$ of $\boldsymbol{a}(z)$:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ -a_n & -a_{n-1} & \dots & \dots & -a_1 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}$$

- $\mathcal{O} \in \mathbb{R}^{(N \times n)}$ is backward shift-invariant:

$$\underline{\mathcal{O}} \cdot \boldsymbol{A} = \overline{\mathcal{O}}$$

with $\mathcal{O} = \left[\boldsymbol{C}^T, (\boldsymbol{CA})^T, \dots, (\boldsymbol{CA}^{N-1})^T \right]^T$.

$$\Rightarrow \boxed{\hat{\boldsymbol{y}} \in \text{range}\,(\mathcal{O})}$$

Autonomous misfit modelling (inner)

$$
\boldsymbol{T_a} = \begin{bmatrix} a_n & 0 & \ddots & 0 \\ a_{n-1} & a_n & \ddots & 0 \\ \vdots & a_{n-1} & \ddots & \vdots \\ a_1 & \vdots & \ddots & a_n \\ 1 & a_1 & \ddots & a_{n-1} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & a_1 \\ 0 & 0 & \ddots & 1 \end{bmatrix}
$$

- $\boldsymbol{T_a} \in \mathbb{R}^{(N \times N-n)}$
- $(\boldsymbol{T_a})^T \hat{\boldsymbol{y}} = 0$ and $\hat{\boldsymbol{y}} \in \text{range}\,(\mathcal{O})$

  $\Rightarrow \text{null}\,\left((\boldsymbol{T_a})^T\right) = \text{range}\,(\mathcal{O})$
- $\text{range}\,(\boldsymbol{T_a}) \perp \text{null}\,\left((\boldsymbol{T_a})^T\right)$

$$
\text{range}\,(\boldsymbol{T_a}) \oplus \text{range}\,(\mathcal{O}) = \mathbb{R}^N \quad \text{with} \quad \text{range}\,(\boldsymbol{T_a}) \perp \text{range}\,(\mathcal{O})
$$

$$
\Rightarrow \boxed{\tilde{\boldsymbol{y}} = \boldsymbol{T_a}\left((\boldsymbol{T_a})^T \boldsymbol{T_a}\right)^{-1} (\boldsymbol{T_a})^T \boldsymbol{y}}
$$

Autonomous misfit modelling (inner)

- For $N \to \infty$: discrete Beurling–Lax theorem[2]:

$$\text{range}\,(\boldsymbol{T_a}) \oplus \text{range}\,(\mathcal{O}) = \ell_2$$

  - range $(\boldsymbol{T_a})$ a forward shift-invariant subspace $\mathcal{M}$ of ℓ_2
  - range $(\mathcal{O})$ a backward shift-invariant subspace $\mathcal{M}^\perp$ of ℓ_2

- Continuous Beurling–Lax via isometric misfit model[3]:
  - Orthogonal decomposition of $\mathcal{H}_2$
  - *All-pass* completion of $\boldsymbol{a}(z)$ as inner-function:

$$H(z) = \frac{\boldsymbol{a_r}(z)}{\boldsymbol{a}(z)} = \frac{1 + a_1 z + \ldots + a_n z^n}{z^n + \cdots + a_{n-1} z + a_n}$$

[2] The original paper by Beurling *(Beurling [1])* derives the theorem for continuous functions in $\mathcal{H}_2$.

[3] $Y(z)$ denotes the Z-transform of $\boldsymbol{y}$, $\tilde{Y}(z)$ and $\hat{Y}(z)$ are defined similarly.

Autonomous misfit modelling (outer)

- First-order necessary conditions for optimality:

$$\min_{\boldsymbol{a}} \quad \sigma^2 = ||\tilde{\boldsymbol{y}}||_2^2 = \boldsymbol{y}^T \boldsymbol{T_a}(\boldsymbol{T_a}^T \boldsymbol{T_a})^{-1} \boldsymbol{T_a} \boldsymbol{y}$$
$$\Rightarrow \quad \frac{\partial \sigma^2}{\partial a_i} = 0, \quad \forall i = 1, \ldots, n,$$

  is a system of n polynomial equations in n variables.

- Common roots characterize all the stationary points

  $\longrightarrow$ select globally optimal solution(s)

- Walsh's theorem

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

Table of Contents

SISO misfit modelling

$$\underbrace{\begin{bmatrix} a_n & -b_n & a_{n-1} & -b_{n-1} & \dots & a_1 & -b_1 & 1 & -b_0 & 0 & \dots & 0 & 0 \\ 0 & 0 & a_n & -b_n & a_{n-1} & -b_{n-1} & \dots & a_1 & -b_1 & 1 & b_0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & 0 & \dots & & a_n & -b_n & a_{n-1} & -b_{n-1} & \dots & a_1 & -b_1 & 1 & b_0 \end{bmatrix}}_{\boldsymbol{T_{a,b}}} \underbrace{\begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \\ \vdots \\ \hat{w}_{N-1} \end{bmatrix}}_{\hat{\boldsymbol{w}}} = \mathbf{0},$$

where:

$$\hat{\boldsymbol{w}} = \begin{bmatrix} \hat{y}_0 & \hat{u}_0 & \hat{y}_1 & \hat{u}_1 & \dots & \hat{y}_N & \hat{u}_N \end{bmatrix}.$$

- $\boldsymbol{T_{a,b}} \in \mathbb{R}^{(2N-2n \times 2N)}$
- Treat as autonomous model of order $2n$?
  - Optimal misfit: $\tilde{\boldsymbol{w}} = \boldsymbol{T_{a,b}} \left((\boldsymbol{T_{a,b}})^T \boldsymbol{T_{a,b}} \right)^{-1} (\boldsymbol{T_{a,b}})^T \boldsymbol{w}$
  - Optimal model: $\sigma^2 = ||\tilde{\boldsymbol{w}}||_2^2 = \boldsymbol{w}^T \boldsymbol{T_{a,b}} (\boldsymbol{T_{a,b}^T} \boldsymbol{T_{a,b}})^{-1} \boldsymbol{T_{a,b}} \boldsymbol{w}$

Questions?

Table of Contents I

¡beamer¿

[1] A. Beurling. On two problems concerning linear transformations in Hilbert space. *Acta Mathematica*, 81:239 – 255, 1949.

[2] B. De Moor. Total least squares for affinely structured matrices and the noisy realization problem. *IEEE Transactions on Signal Processing*, 42(11):3104 – 3113, 1994.

[3] B. De Moor. Least squares optimal realisation of autonomous LTI systems is an eigenvalue problem. *Communications in Information and Systems*, 20(2):163 – 207, 2020.

[4] P. Lemmerling and B. De Moor. Misfit versus latency. *Automatica*, 37:2057 – 2067, 2001.

[5] B. Roorda. Algorithms for global total least square modelling of finite multivariable time series. *Automatica*, 1995.

[6] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory – Implementation – Applications*. Kluwer Academic Publishers, 1996.

[7] J. C. Willems. From time series to linear system, Part I: Finite dimensional linear time invariant systems; Part II: Exact modelling; Part III: Approximate modelling. *Automatica*, 22/23(5,6,1):561 – 580, 675 – 694, 87 – 115, 1986, 1987.

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics