STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

KU LEUVEN FACULTY OF ENGINEERING SCIENCE

LEAST SQUARES OPTIMAL REALISATION OF SISO LTI SYSTEMS IS AN EIGENVALUE PROBLEM

Sibren Lagauw* and Bart De Moor*

KU Leuven, Department of Electrical Engineering (ESAT-STADIUS)

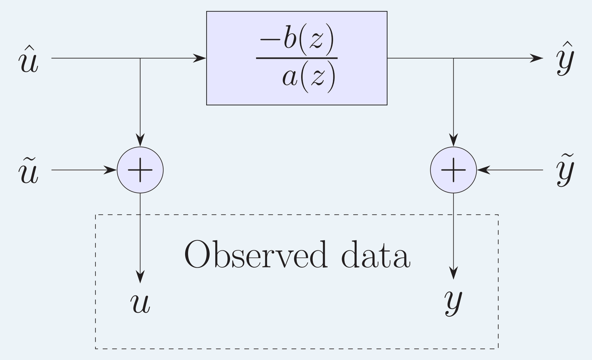
September, 2022

Abstract System identification translates time series data of dynamical systems into mathematical models. In least squares linear time-invariant (LTI) realisation problems, the observed data is modified in a least squares sense so that it satisfies linear dynamic relations, resulting in a 'structured' total least squares problem, for which several heuristic methods have been proposed in the literature [1, 3, 4]. Lately, it has been shown that the least squares optimal realisation problem for autonomous LTI systems is essentially equivalent to a (large) eigenvalue problem [2]. This new methodology, which exploits the shift-invariant structures present in the data and models, is deterministic in the sense that it guarantees to find the globally optimal solution, thereby eliminating the heuristics that are accustomed in the state of the art. In this poster, we describe the extension of these results to single-input single-output (SISO) LTI models, and show that also in this case the least squares optimal realisation problem can be solved using standard (numerical) linear algebra tools. We use Willems' behavioral modeling framework to circumvent the increase in theoretical complexity, leading to novel insights.

Problem formulation

The exact inputs $\hat{\boldsymbol{u}} = [\hat{u}_0, \ldots, \hat{u}_{N-1}]^{\mathrm{T}} \in \mathbb{R}^N$ and exact outputs $\hat{\boldsymbol{y}} = [\hat{y}_0, \ldots, \hat{y}_{N-1}]^{\mathrm{T}} \in \mathbb{R}^N$ of a SISO LTI model of order n satisfy a linear relation, such that:

$$\sum_{i=0}^{n} a_i \hat{y}_{k-i} + \sum_{i=0}^{n} b_i \hat{u}_{k-i} = 0, k = n, \ldots, N-1, \tag{1}$$



with $a_i, b_i \in \mathbb{R}$ $(i = 0, \ldots, n)$ the coefficients that make up the transfer function $H(z)$ of the model:

$$H(z) = \frac{-b(z)}{a(z)} = \frac{-(b_0 z^n + b_1 z^{n-1} + \cdots + b_{n-1} z + b_n)}{a_0 z^n + a_1 z^{n-1} + \cdots + a_{n-1} z + a_n}. \tag{2}$$

Given observed input-output signals $\boldsymbol{u} = [u_0, \ldots, u_{N-1}]^{\mathrm{T}}, \boldsymbol{y} = [y_0, \ldots, y_{N-1}]^{\mathrm{T}} \in \mathbb{R}^N$, we want to find $\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{y}}$ and $\boldsymbol{a} = [a_0, a_1, \ldots, a_n]$ and $\boldsymbol{b} = [b_0, \ldots, b_n]$, the coefficients of $a(z)$ and $b(z)$ respectively, such that:

$$J = \sum_{k=0}^{N-1} [(y_k - \hat{y}_k)^2 + (u_k - \hat{u}_k)^2] \tag{3}$$

is minimal, subject to (1) and $a_0 = 1$ to avoid the trivial solution. The difference between the observed and the exact inputs is called the input misfit $\tilde{\boldsymbol{u}} = \boldsymbol{u} - \hat{\boldsymbol{u}}$ (the misfit on the outputs $\tilde{\boldsymbol{y}}$ is defined similarly).

Direct-form Lagrangian

With $\boldsymbol{l} = [l_0, \ldots, l_{N-n-1}] \in \mathbb{R}^{N-n}$ and $\lambda \in \mathbb{R}$ lagrange multipliers, the Lagrangian of the optimization problem described in the previous block is given as:

$$\mathcal{L}(\hat{\boldsymbol{y}}, \hat{\boldsymbol{u}}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{l}, \lambda) = \sum_{k=0}^{N-1} [(y_k - \hat{y}_k)^2 + (u_k - \hat{u}_k)^2] + \sum_{k=n}^{N-1} l_{k-n} \left(\sum_{i=0}^{n} a_i \hat{y}_{k-i} + \sum_{i=0}^{n} b_i \hat{u}_{k-i} \right) + \lambda(a_0 - 1).$$

Consider the partial derivatives of the Lagrangian to the exact outputs $\hat{y}_k$:

$$\frac{\partial \mathcal{L}(\ldots)}{\partial \hat{y}_k} = \begin{cases} (-1)(y_k - \hat{y}_k) + \sum_{i=0}^{k} a_{n-i} l_{k-i} & (0 \leq k \leq n) \\ (-1)(y_k - \hat{y}_k) + \sum_{i=0}^{n} a_{n-i} l_{k-i} & (n \leq k \leq N-n-1) \\ (-1)(y_k - \hat{y}_k) + \sum_{i=n-(N-1-k)}^{k} a_{n-i} l_{k-i} & (N-n-1 \leq k \leq N-1) \end{cases}.$$

Equating each line to zero shows that the misfits are obtained from a convolution operation:

$$\boldsymbol{y} - \hat{\boldsymbol{y}} = \tilde{\boldsymbol{y}} = \boldsymbol{T}_{\boldsymbol{a_r}} \boldsymbol{l}, \tag{4}$$

with $\boldsymbol{T}_{\boldsymbol{a_r}} \in \mathbb{R}^{N \times (N-n)}$ defined as (empty cells represent zero submatrices):

$$\boldsymbol{T}_{\boldsymbol{a_r}} = \begin{bmatrix} a_n & & & & & \\ a_{n-1} & a_n & & & & \\ \vdots & \ddots & \ddots & & & \\ a_0 & \ldots & a_{n-1} & a_n & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & a_0 & a_1 & \ldots & a_n \\ & & & \ddots & \ddots & \vdots \\ & & & & a_0 & a_1 \\ & & & & & a_0 \end{bmatrix}$$

We use the subscript $\boldsymbol{a_r}$ because the columns of $\boldsymbol{T}_{\boldsymbol{a_r}}$ are forward shifts of the vector $\boldsymbol{a_r} = [a_n, \ldots, a_1, 1]^{\mathrm{T}}$. Equivalently, with $\boldsymbol{T}_{\boldsymbol{b_r}}$ defined similarly as $\boldsymbol{T}_{\boldsymbol{a_r}}$, the partial derivatives to $\hat{u}_k$ lead to:

$$\tilde{\boldsymbol{u}} = \boldsymbol{T}_{\boldsymbol{b_r}} \boldsymbol{l}. \tag{5}$$

Equating the partial derivatives of $\mathcal{L}$ with respect to the lagrange multipliers $\boldsymbol{l}$ to zero results in the constraints of the optimization problem (1), i.e., the linear relation between $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{u}}$:

$$\left[\frac{\partial \mathcal{L}(\ldots)}{\partial l_i} \right]_{i=0,\ldots,N-n-1} = 0 \iff \boldsymbol{T}_{\boldsymbol{a_r}}^{\mathrm{T}} \hat{\boldsymbol{y}} + \boldsymbol{T}_{\boldsymbol{b_r}}^{\mathrm{T}} \hat{\boldsymbol{u}} = \boldsymbol{0} \iff \boldsymbol{T}_{\boldsymbol{a_r}}^{\mathrm{T}} \hat{\boldsymbol{y}} = -\boldsymbol{T}_{\boldsymbol{b_r}}^{\mathrm{T}} \hat{\boldsymbol{u}}. \tag{6}$$

In the z-domain this corresponds to the transfer function of the exact data model from (2).

Minimization of the least squares misfit objective (I)

Optimal misfit Define $\hat{\boldsymbol{w}} = [\hat{\boldsymbol{w}}_0^{\mathrm{T}}, \ldots, \hat{\boldsymbol{w}}_{N-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2N}$ with $\hat{\boldsymbol{w}}_k = [\hat{y}_k, \hat{u}_k]^{\mathrm{T}}$ as the behavior [5] of the linear model ($\boldsymbol{w}$ and $\tilde{\boldsymbol{w}}$ are defined similarly) and write out the $N - n$ recursions from (1) to get:

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

from which it is clear that the optimal $\hat{\boldsymbol{w}}$ must be perpendicular to the row space of this 'double shifted' banded Toeplitz matrix, the transpose of which we will denote with $\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}} \in \mathbb{R}^{(N-n)\times(2N)}$:

$$\hat{\boldsymbol{w}} \perp \mathrm{range}\left(\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}\right).$$

Assuming that the model parameters a_i, b_i are known, the minimal norm $\tilde{\boldsymbol{w}}$ can be found from the orthogonal projection of $\boldsymbol{w}$ onto the column space of $\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}$, giving:

$$\tilde{\boldsymbol{w}} = (\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}})^{\dagger} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{w} = \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}} (\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}})^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{w}.$$

Minimization of the least squares misfit objective (II)

Optimal model Similarly as in [2], the objective function (3) can be rewritten as:

$$J = ||\tilde{\boldsymbol{w}}||_2^2 = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}} (\boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}})^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{w}$$

With $\boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}} = \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}$, the first order optimality conditions of the realization can be written as:

$$\begin{cases} \dfrac{\partial J}{\partial a_i} = 2\boldsymbol{w}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{a_i} \boldsymbol{w} - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{a_i} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}} \boldsymbol{w}, & i = 1, \ldots, n, \\ \dfrac{\partial J}{\partial b_i} = 2\boldsymbol{w}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{b_i} \boldsymbol{w} - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{\mathrm{T}} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{b_i} \boldsymbol{D}_{\boldsymbol{a_r},\boldsymbol{b_r}}^{-1} \boldsymbol{T}_{\boldsymbol{a_r},\boldsymbol{b_r}} \boldsymbol{w}, & i = 0, \ldots, n, \end{cases}$$

where the superscript x_i denotes the partial derivative wrt. x_i. This can be shown to be equivalent to a multiparameter eigenvalue problem, the eigentuples of which characterize all the stationary points of the realization problem [2].

Isometric behavioral modeling

Consider a state-space representation for the n-th order SISO system $(\boldsymbol{A}, \boldsymbol{B'}, \boldsymbol{C}, \boldsymbol{D'})$ mapping inputs $\hat{u}_k$ to outputs $\hat{y}_k$, with $\boldsymbol{A} \in \mathbb{R}^{n\times n}, \boldsymbol{B'} \in \mathbb{R}^{n\times 1}, \boldsymbol{C} \in \mathbb{R}^{1\times n}$ and $\boldsymbol{D'} \in \mathbb{R}$. A behavioral embedding of the SISO model can be taken as:

$$\begin{aligned} \hat{\boldsymbol{x}}_{k+1} &= \boldsymbol{A}\hat{\boldsymbol{x}}_k + \boldsymbol{B}\hat{v}_k \\ \hat{\boldsymbol{w}}_k &= \underbrace{\begin{bmatrix} \boldsymbol{0}_m \\ \boldsymbol{C'} \end{bmatrix}}_{\boldsymbol{C}} \hat{\boldsymbol{x}}_k + \underbrace{\begin{bmatrix} \boldsymbol{I}_m \\ \boldsymbol{D'} \end{bmatrix}}_{D} \hat{v}_k, \quad k = 0, \ldots, N-1. \end{aligned} \tag{7}$$

This model maps auxiliary inputs $\hat{\boldsymbol{v}} \in \mathbb{R}^N$, representing the degree of freedom at each point in time, to the exact behavior $\hat{\boldsymbol{w}} = [\hat{\boldsymbol{w}}_0^{\mathrm{T}}, \ldots, \hat{\boldsymbol{w}}_{N-1}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2N}$ (set $\hat{v}_k = \hat{u}_k$ to get $\hat{\boldsymbol{w}}_k = [\hat{u}_k, \hat{y}_k]^{\mathrm{T}}$). If the roots of $a(z)$ lie strictly within the unit circle, we can, without loss of generality, assume that this behavioral state-space model is isometric [4]. Using the Lagrange multipliers $\boldsymbol{\lambda}_k \in \mathbb{R}^n, \boldsymbol{\mu}_k \in \mathbb{R}^2, k = 0, \ldots, N-1$, write the Lagrangian of the realization problem as:

$$\mathcal{L}(\hat{\boldsymbol{w}}, \hat{\boldsymbol{x}}_k, \boldsymbol{\lambda_k}, \boldsymbol{\mu}_k, \hat{v}_k) = \tfrac{1}{2}\textstyle\sum_{k=0}^{N-1} ||\boldsymbol{w}_k - \hat{\boldsymbol{w}}_k||^2 + \sum_{k=0}^{N-1} \boldsymbol{\lambda}_k^{\mathrm{T}}(\hat{\boldsymbol{x}}_{k+1} - \boldsymbol{A}\hat{\boldsymbol{x}}_k - \boldsymbol{B}\hat{v}_k) + \sum_{k=0}^{N-1} \boldsymbol{\mu}_k^{\mathrm{T}}(\hat{\boldsymbol{w}}_k - \boldsymbol{C}\hat{\boldsymbol{x}}_k - \boldsymbol{D}\hat{v}_k)$$

Equating all partial derivatives to zero allows one to derive that the optimal misfit $\tilde{\boldsymbol{w}}$ is generated by the *misfit model*:

$$\begin{bmatrix} \boldsymbol{\lambda}_k \\ \tilde{\boldsymbol{w}}_k \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \tilde{\boldsymbol{B}} \\ \boldsymbol{C} & \tilde{\boldsymbol{D}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}_{k-1} \\ \tilde{v}_k \end{bmatrix}, \quad k = 0, \ldots, N-1, \tag{8}$$

where the matrices $\tilde{\boldsymbol{B}}$ and $\tilde{\boldsymbol{D}}$ come from a unitary completion of the exact model, such that:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \tilde{\boldsymbol{B}} \\ \boldsymbol{C} & \boldsymbol{D} & \tilde{\boldsymbol{D}} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \tilde{\boldsymbol{B}} \\ \boldsymbol{C} & \boldsymbol{D} & \tilde{\boldsymbol{D}} \end{bmatrix} = \boldsymbol{I} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \tilde{\boldsymbol{B}} \\ \boldsymbol{C} & \boldsymbol{D} & \tilde{\boldsymbol{D}} \end{bmatrix} \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} & \tilde{\boldsymbol{B}} \\ \boldsymbol{C} & \boldsymbol{D} & \tilde{\boldsymbol{D}} \end{bmatrix}^{\mathrm{T}}.$$

Optimality properties

Orthogonality Multiply (6) from the left with $\boldsymbol{l}^{\mathrm{T}}$ and use (4) and (5) to show that in the stationary points of the optimization problem, the exact data and the misfits are orthogonal wrt. each other:

$$\begin{aligned} &\boldsymbol{l}^{\mathrm{T}}(\boldsymbol{T}_{\boldsymbol{a_r}}^{\mathrm{T}} \hat{\boldsymbol{y}} + \boldsymbol{T}_{\boldsymbol{b_r}}^{\mathrm{T}} \hat{\boldsymbol{u}}) = 0 \\ \iff &\tilde{\boldsymbol{y}}^{\mathrm{T}} \hat{\boldsymbol{y}} + \tilde{\boldsymbol{u}}^{\mathrm{T}} \hat{\boldsymbol{u}} = 0 \end{aligned} \tag{9}$$

Structured misfit If we assume that the input is bounded and that the zeros of $a(z)$ lie within the unit circle, then use (2), the fact that generically $\hat{U}(z^{-1}) \neq 0$ and exploit the orthogonality (9) to derive that:

$$\begin{aligned} &\langle \tilde{Y}(z), \frac{-b(z)}{a(z)} \hat{U}(z) \rangle + \langle \tilde{U}(z), \hat{U}(z) \rangle = 0 \\ \iff &\oint_{|z|=1} \left(\tilde{Y}(z) \frac{-b(z^{-1})}{a(z^{-1})} + \tilde{U}(z) \right) \hat{U}(z^{-1}) \mathrm{d}z = 0 \\ \iff &\tilde{Y}(z) = \frac{a(z^{-1})}{b(z^{-1})} \tilde{U}(z) = \frac{a_r(z)}{b_r(z)} \tilde{U}(z), \end{aligned}$$

where $a_r(z)$ and $b_r(z)$ are the 'coefficient reversed' polynomials of $a(z)$ and $b(z)$. The same relation between $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{u}}$ is found when one constructs the transfer matrices of the state-space models in (7) and (8):

$$\hat{\boldsymbol{H}}(z) = \left[c\frac{a(z)}{d(z)} \quad -c\frac{b(z)}{d(z)} \right]^{\mathrm{T}}, \quad \tilde{\boldsymbol{H}}(z) = \left[c\frac{b_r(z)}{d(z)} \quad c\frac{a_r(z)}{d(z)} \right]^{\mathrm{T}},$$

where c is an arbitrary constant and $d(z)$ so that $d(z)d_r(z) = c^2(a_r(z)a(z) + b_r(z)b(z))$.

References

[1] Bart De Moor. Dynamic total linear least squares. In *Proc. of the 10th IFAC Symposium on System Identification*, volume 3, pages 159 – 164, Copenhagen, Denmark, July 1994.

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

[3] Bart De Moor and Berend Roorda. l_2-optimal linear system identification: Structured total least squares for SISO systems. In *Proc. of the 33rd IEEE Conference on Decision and Control (CDC)*, pages 2874–2879, Lake Buena Vista, Florida, USA, December 1994.

[4] Berend Roorda. *Global Total Least Squares - A method for the construction of open approximate models from vector time series*. PhD thesis, Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands, 1995.

[5] Jan 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.

* KU Leuven, Dept. of Elec. Engineering (ESAT), Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS), Kasteelpark Arenberg 10, 3001 Leuven, Belgium (sibren.lagauw@esat.kuleuven.be, bart.demoor@esat.kuleuven.be).

This work was supported in part by the KU Leuven: Research Fund (project C16/15/059), and several Leuven Research and Development bilateral industrial projects, in part by Flemish Government agencies: FWO (EOS project G0F6718N (SeLMA), and PhD grants SB/1SA1319N, SB/1S93918, SB/1S1319N), EWI (Flanders AI Research Program), and in part by the European Commission (EU Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant) under grant 885682).