Globally optimal least-squares misfit identification of multidimensional autonomous systems

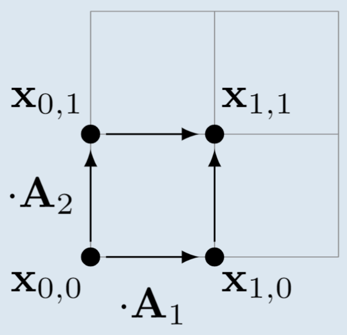
Lukas Vanpoucke and Bart De Moor
KU Leuven - Department of Electrical Engineering (ESAT-STADIUS)

Abstract

Previous work has shown that the globally optimal least-squares misfit identification problem for single output, autonomous, one-dimensional, discrete, linear, time-invariant systems corresponds to a polynomial optimization problem. The first order optimality conditions for such problems can be translated into a system of multivariate polynomial equations. In this specific case, the resulting system can be written as a special class of polynomial systems: a multiparameter eigenvalue problem (MEP). To solve such problems, we will use the block Macaulay method, based upon the well-understood language of linear algebra. This method performs better computationally than the standard Macaulay method for systems of polyomials. We aim to extend this methodology to the globally optimal least squares misfit identification for a multidimensional (mD) generalization of such systems.

Multidimensional state space models

We aim to identify autonomous, single output, commutative, linear state space models. For two-dimensional systems this corresponds to state space models of the form:



$$\mathbf{x}_{k+1,l} = \mathbf{A}_1 \cdot \mathbf{x}_{k,l}$$
$$\mathbf{x}_{k,l+1} = \mathbf{A}_2 \cdot \mathbf{x}_{k,l}$$
$$\hat{y}_{k,l} = \mathbf{C} \cdot \mathbf{x}_{k,l}$$

where $\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1$.

We consider the subclass where the commuting matrices $\mathbf{A}_1, \mathbf{A}_2$ are **diagonizable**. This implies that their eigenvalue decompositions have the following form, where $\mathbf{D}_1, \mathbf{D}_2$ are diagonal matrices:

$$\mathbf{A}_1 = \mathbf{V}\mathbf{D}_1\mathbf{V}^{-1}$$
$$\mathbf{A}_2 = \mathbf{V}\mathbf{D}_2\mathbf{V}^{-1}.$$

This is the **generic case** for commuting matrices.

Under this assumption one can choose $\mathbf{A}_1, \mathbf{A}_2$ to be **diagonal** without loss of generality. This allows one to **parameterize the output** by the eigenvalues: $\lambda_j^{(1)}, \lambda_j^{(2)}$ of $\mathbf{A}_1, \mathbf{A}_2$ and the initial conditions ξ_j:

$$\hat{y}_{k,l} = \sum_{j=1}^{n} \xi_j \left(\lambda_j^{(1)}\right)^k \left(\lambda_j^{(2)}\right)^l.$$

A vector containing the whole output sequence can then be written as the image of a matrix $\mathbf{\Lambda}$ containing the multivariate Vandermonde vectors associated with $(\lambda_j^{(1)}, \lambda_j^{(2)})$ as columns:

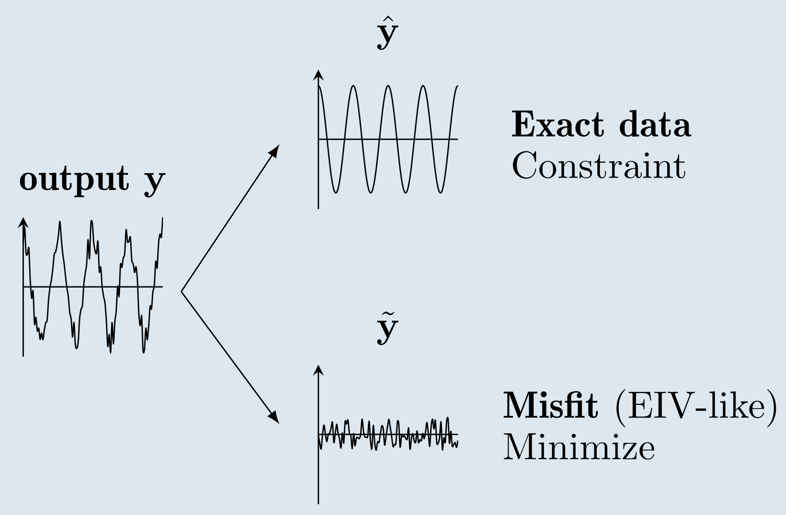
$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_{0,0} \\ \hat{y}_{1,0} \\ \hat{y}_{0,1} \\ \hat{y}_{2,0} \\ \hat{y}_{1,1} \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1^{(1)} & \dots & \lambda_n^{(1)} \\ \lambda_1^{(2)} & \dots & \lambda_n^{(2)} \\ \left(\lambda_1^{(1)}\right)^2 & \dots & \left(\lambda_n^{(1)}\right)^2 \\ \lambda_1^{(1)}\lambda_1^{(2)} & \dots & \lambda_n^{(1)}\lambda_n^{(2)} \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix} = \mathbf{\Lambda}\mathbf{x}_{0,0}.$$

Misfit identification

Using the parameterization above, we apply the least squares misfit identification framework to identify models from the given data:

1. Split the given output sequence $\mathbf{y}$ into an *exact* data sequence and a *misfit* data sequence:
$$\mathbf{y}_{k,l} = \hat{\mathbf{y}}_{k,l} + \tilde{\mathbf{y}}_{k,l}.$$
2. Constrain the *exact* data sequence to follow the predefined model parameterization exactly.
3. Find the parameters describing the exact data sequence closest to the given data.

This leads to an objective function and equality constraints that are both **multivariate polynomial**.



$$\min_{\lambda_j^{(1)}, \lambda_j^{(2)}, \xi_j} ||\tilde{\mathbf{y}}||_2^2$$
$$\text{s.t.} \begin{cases} \mathbf{y} = \tilde{\mathbf{y}} + \hat{\mathbf{y}} \\ \hat{\mathbf{y}} = \mathbf{\Lambda}\mathbf{x}_{0,0} \end{cases}$$

Multiparameter eigenvalue problems

The first order optimality conditions then correspond to a **system of multivariate polynomials**. Assuming real-valued and nonzero exact data, this can be rephrased as a **multiparameter eigenvalue problem** (MEP):

$$\begin{bmatrix} \mathbf{\Lambda}^T\mathbf{\Lambda} & \mathbf{\Lambda}^T\mathbf{y} \\ \left(\mathbf{\Lambda}^{(\lambda)}\right)^T\mathbf{\Lambda} & \left(\mathbf{\Lambda}^{(\lambda)}\right)^T\mathbf{y} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{0,0} \\ -1 \end{bmatrix} = \mathbf{0}.$$

Another example of a multiparameter eigenvalue problem is given below. The goal is to find the eigentuples (μ, ν) such that the polynomial matrix depending on these parameters has a nontrivial null space:

$$(\mathbf{A}_0 + \mu \cdot \mathbf{A}_1 + \nu \cdot \mathbf{A}_2)\mathbf{v} = \mathbf{0} \quad \text{with} \quad ||\mathbf{v}|| = 1.$$

To solve such systems, we apply the **block Macaulay method**, illustrated below. This relies on the key observation that the (block) multivariate Vandermonde vector associated with a solution (μ^*, ν^*) to a polynomial system lies in the null space of a so-called **(block) Macaulay matrix M**:

$$\begin{matrix} \cdot 1 \\ \cdot \mu \\ \cdot \nu \end{matrix} \underbrace{\begin{bmatrix} 1 & \mu & \nu & \mu^2 & \mu\nu & \nu^2 \\ \mathbf{A}_0 & \mathbf{A}_1 & \mathbf{A}_2 & & & \\ & \mathbf{A}_0 & & \mathbf{A}_1 & \mathbf{A}_2 & \\ & & \mathbf{A}_0 & & \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}}_{\mathbf{M}} \begin{bmatrix} 1 \cdot \mathbf{v} \\ \mu^* \cdot \mathbf{v} \\ \nu^* \cdot \mathbf{v} \\ (\mu^*)^2 \cdot \mathbf{v} \\ \nu^*\mu^* \cdot \mathbf{v} \\ (\mu^*)^2 \cdot \mathbf{v} \end{bmatrix} = \mathbf{0}.$$

Properties of the misfit sequence

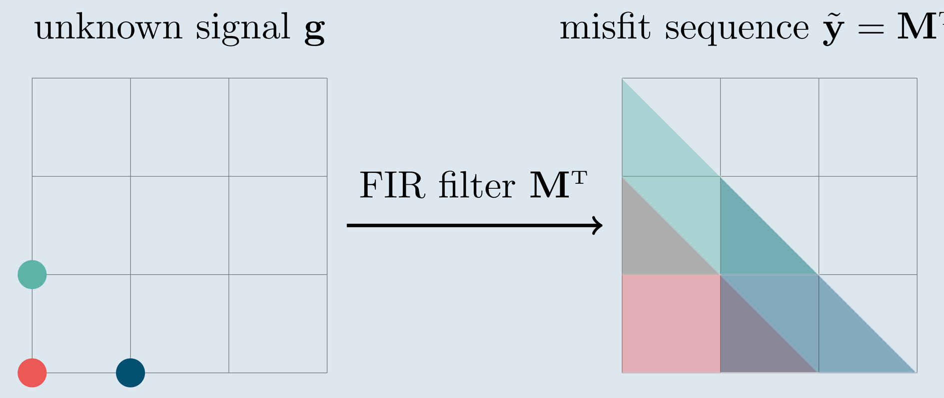
The first order optimality conditions imply that the misfit sequence is orthogonal to the vector space spanned by the multivariate Vandermonde columns of $\mathbf{\Lambda}$ and their derivatives to each $\lambda_j^{(i)}$, which compose $\mathbf{\Lambda}^{(\lambda)}$:

$$\tilde{\mathbf{y}} \perp \text{range}\left(\begin{bmatrix}\mathbf{\Lambda} & \mathbf{\Lambda}^{(\lambda)}\end{bmatrix}\right).$$

Multivariate Vandermonde vectors and their derivatives arise naturally in the null space of Macaulay matrices. The misfit $\tilde{\mathbf{y}}$ thus lies in the orthogonal complement, i.e. the row space of a Macaulay matrix $\mathbf{M}$:

$$\tilde{\mathbf{y}} \in \text{range}\left(\mathbf{M}^T\right).$$

This can be interpreted as the misfit $\tilde{\mathbf{y}}$ resulting from a **2D FIR filtering**. This is illustrated below.



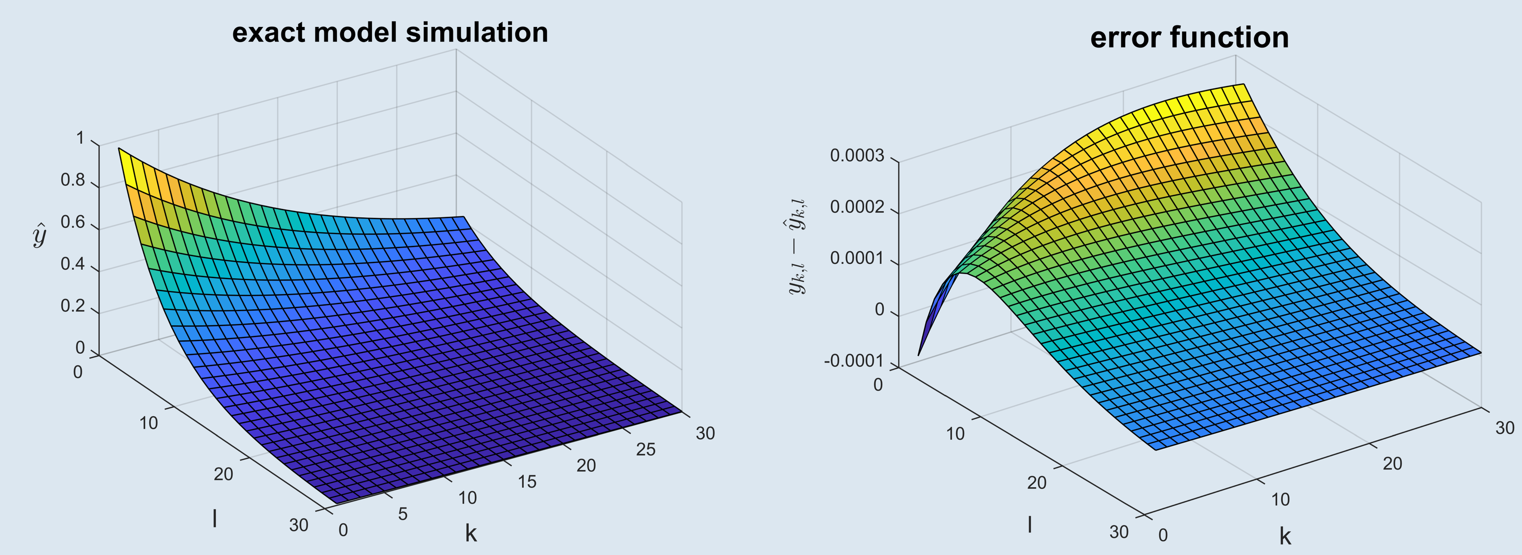
Numerical example

Proof of concept: consider the first order system:

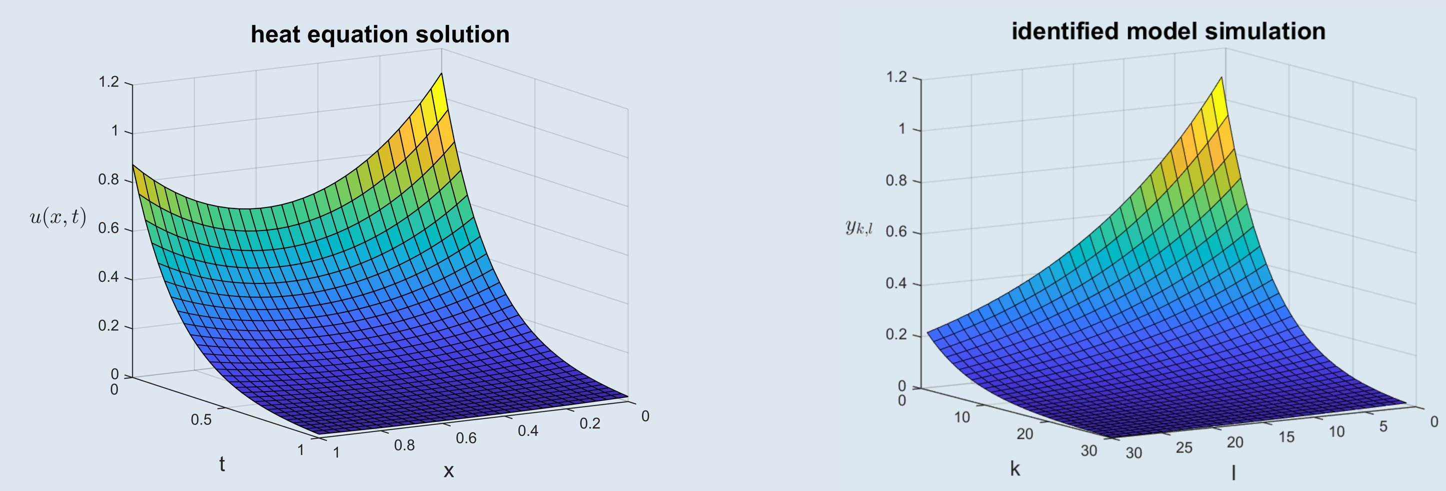
$$\mathbf{A}_1 = 0.95 \quad \mathbf{A}_2 = 0.85 \quad \mathbf{C} = 1 \quad \mathbf{x}_{0,0} = 1$$

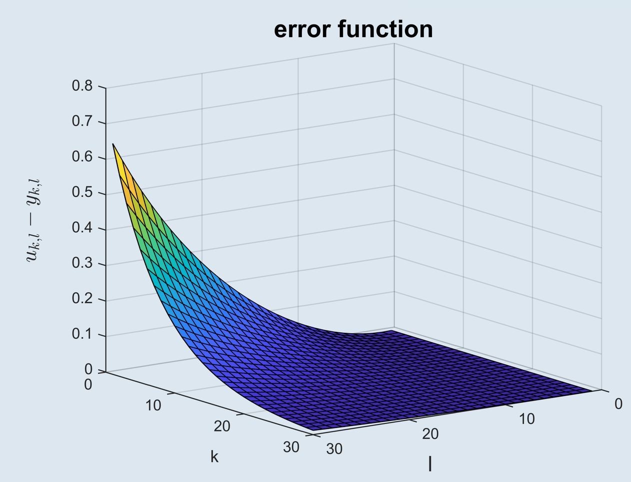
- Generate an exact data sequence. Only a small number of data points is used to keep the degree of the polynomials low as to make it computationally feasible.
- Using the orthogonality properties of the misfit, we generate a misfit sequence with norm 1. Adding this to the exact data, a signal of norm 2.15 is obtained.
- Solve the MEP for the stationary points and check the objective function for the global optimum.

Using this method, the original system, up to numerical errors, is recovered as the globally optimal model.



More practical example: The figure below depicts a solution to the homogeneous heat equation with two exponential modes. We apply the identification method to a subset of this data for a model of order one.





Judging from figures above, the identified model captures the time behaviour nicely. However, a first order model is unable to fully capture the second order spatial dynamics in x.

Applications and open problems

- This method can be used for **benchmarking** heuristic methods.
- The **degrees of the polynomials become large** quite quickly. This is computationally demanding and as such we can only use few data points to fit low order systems.
- A **parameterization in terms of the difference equations** gives rise polynomial systems of a degree independent of the number of data points.
- Such a parameterization can also help to **describe the structure of the optimal misfits more precisely**. In particular, we aim to generalize the Walsh-like property that arises in 1D misfit problems.
- Extension to **more general autonomous state space models**, i.e. with general commuting matrices.