Globally optimal least-squares misfit identification of multidimensional autonomous systems

Lukas Vanpoucke and Bart De Moor

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

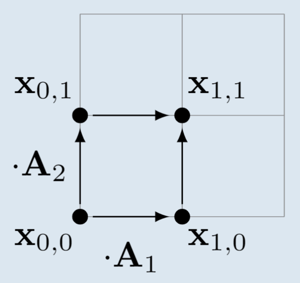
Abstract

Previous work has shown that the globally optimal least-squares misfit identification problem for single output, autonomous, one-dimensional, discrete, linear systems corresponds to a polynomial optimization problem [4]. The first order optimality conditions for such problems can be translated into a system of multivariate polynomials. In this specific case, the resulting system can be written as a special class of polynomial systems: a multiparameter eigenvalue problem (MEVP) [3]. 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 for globally optimal least squares misfit identification to general multidimensional (mD), autonomous, discrete, single output, linear systems.

Overdertermined state space models

We aim to identify of so-called overdetermined [2], autonomous, single output, linear state space models. For two-dimensional systems this corresponds to state space models of the form:

state transition diagram:



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

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

Any finite dimensional, autonomous, discrete, linear system can be written in this form [5]. We consider the subclass where the **eigendecompositions** of the commuting matrices $\mathbf{A}_1, \mathbf{A}_2$ **share the same eigenvectors**.

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

Under this assumption one can choose $\mathbf{A}_1, \mathbf{A}_2$ to be **diagonal**, which allows for the following **parametrization of the output**, with the eigenvalues $\lambda_j^{(1)}, \lambda_j^{(2)}$ of $\mathbf{A}_1, \mathbf{A}_2$ and the initial conditions ξ_j as parameters:

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

Note: this last assumption only excludes systems with multiple modes and as such, the model class fully covers the **generic** case of discrete, autonomous, linear mD systems.

Misfit identification

The **general idea** of least squares misfit identification [6]:

- Split the given output sequence $\mathbf{y}$ into an "exact" data sequence and a and a "misfit" data sequence:
$$\mathbf{y}_{k,l} = \hat{\mathbf{y}}_{k,l} + \tilde{\mathbf{y}}_{k,l}.$$
- Constrain the "exact" data sequence to follow the predefined model parametrization exactly.
- Minimize the 2-norm of the "misfit" vector. This leads to the constrained optimization problem:

$$\min_{\lambda_j^{(1)},\lambda_j^{(2)},\xi_j,c_j} ||\tilde{\mathbf{y}}||_2^2 = ||\mathbf{y}-\hat{\mathbf{y}}|| = \sum_{k=0}^{N_1}\sum_{k=0}^{N_2}(y_{k,l}-\hat{y}_{k,l})^2$$

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

By substituting in the constraint, a multivariate **polynomial objective function** is obtained.

First order optimality conditions

Using Wirtinger calculus [1] and assuming the data and misfit vectors to be real, one can show that the stationary points correspond to the solution of the following **multivariate polynomial system**:

$$0 = \mathbf{x}_{0,0}{}^{\mathsf{T}}\mathbf{\Lambda}^{\mathsf{T}}\frac{\delta\mathbf{\Lambda}}{\delta\lambda_{i,k}}\mathbf{x}_{0,0} - \mathbf{y}^{\mathsf{T}}\frac{\delta\mathbf{\Lambda}}{\delta\lambda_{i,k}}\mathbf{x}_{0,0} \qquad i = 1...n,\ \ k = 1,...m$$
$$0 = \mathbf{x}_{0,0}{}^{\mathsf{T}}\mathbf{\Lambda}^{\mathsf{T}}\mathbf{\Lambda}\frac{d\mathbf{x}_{0,0}}{d\xi_i} - \mathbf{y}^{\mathsf{T}}\mathbf{\Lambda}\frac{d\mathbf{x}_{0,0}}{d\xi_i} \qquad i = 1,...n.$$

To solve this system practically using the algorithm in [7], we first **rewrite it as an MEVP**, beacause the **Block Macaulay method is computationally less expensive** compared to the standard Macaulay method. For this purpose, we assume $\mathbf{x}_{0,0} \neq \mathbf{0}$, since these solutions correspond to maximizing solutions.

Properties of the misfit sequence

The first order optimality conditions imply the misfit to be orthogonal to the mD observability matrix of the identified model:

$$\tilde{\mathbf{y}} \perp \text{range}(\mathbf{\Lambda})$$

It can also be worked out that the optimality conditions also imply the following orthogonality conditions:

$$\tilde{\mathbf{y}} \perp \text{range}\left(\frac{\delta\mathbf{\Lambda}}{\delta\lambda_{i,k}}\right) \qquad \forall i,k.$$

- These orthogonality constraints **impose structure on the misfit sequence**, hence it cannot be considered as simple statistical noise.
- Useful for the **construction of test problems** a stationary points at a predefined location. Such problems can be used to test software implementations.

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 MEVP 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.

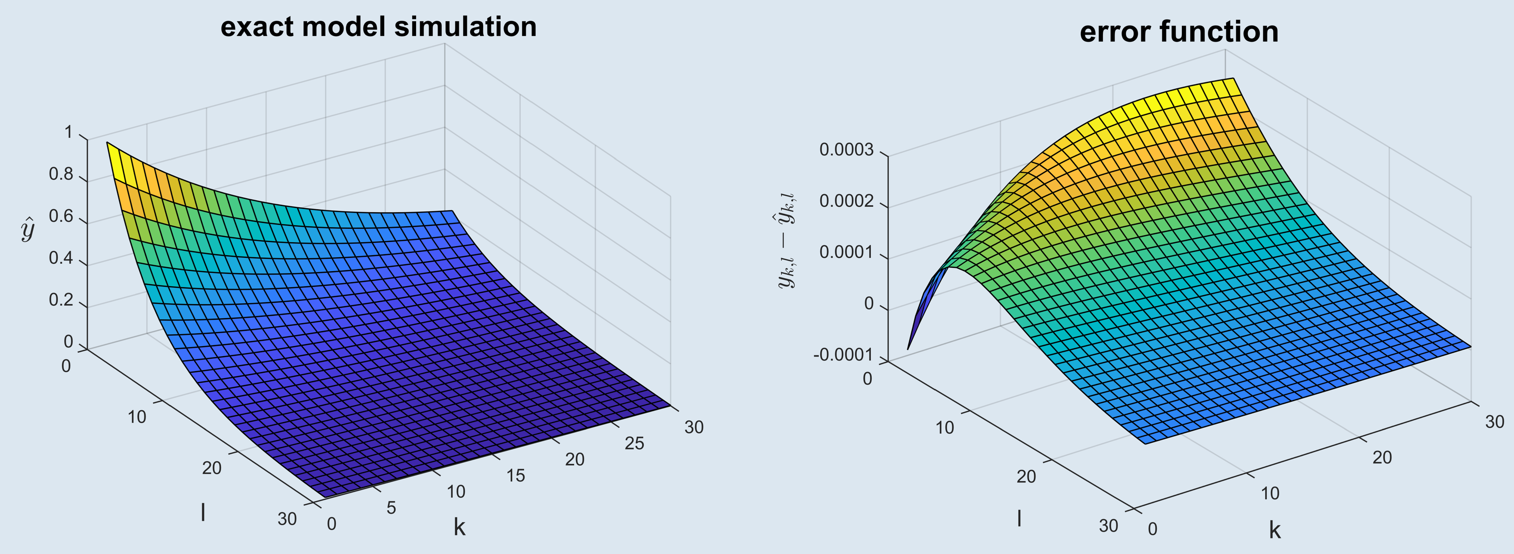


Fig. 1: Identification process: exact data sequence (left) and the resulting error function (right).

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

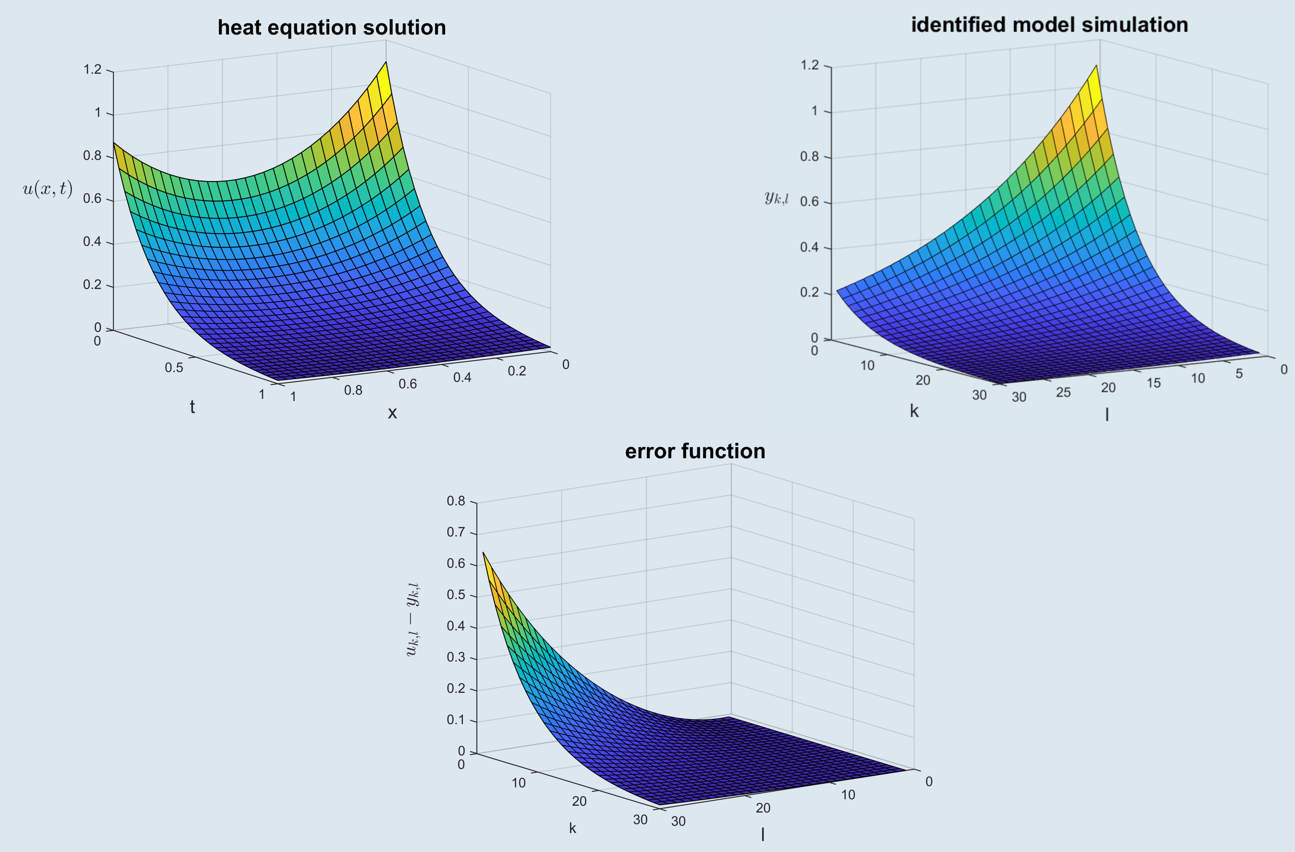


Fig. 2: Identification PDE data: heat equation solution (left), simulation of the identified model based on this data (right), error function (under).

Judging from Figure 2, 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.

Open problems and applications

- 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 **minimal parameterization in terms of the difference equations** is better suited for systems with multiplicities and gives rise polynomial systems of a degree independent on the number of data points.
- This method can be used for e.g. **benchmarking** heuristic methods.

References

[1] Tulay Adali and Peter J. Schreier. "Optimization and Estimation of Complex-Valued Signals: Theory and applications in filtering and blind source separation". In: *IEEE Signal Processing Magazine* 31.5 (2014), pp. 112–128.

[2] Joseph A Ball and Victor Vinnikov. "Overdetermined multidimensional systems: state space and frequency domain methods". In: *Mathematical systems theory in biology, communications, computation, and finance.* 2003, pp. 63–119.

[3] Katrien De Cock and Bart De Moor. "Multiparameter Eigenvalue Problems and Shift-invariance". In: *IFAC-PapersOnLine, IFAC-PapersOnLine, Part of Special Issue: 24th International Symposium on Mathematical Theory of Networks and Systems (MTNS2020)* 54.9 (2021), pp. 159–165.

[4] Bart De Moor. "Least squares optimal realisation of autonomous LTI systems is an eigenvalue problem". In: *Communications in Information and Systems* 20.2 (2020), pp. 163–207.

[5] E. Fornasini, P. Rocha, and S. Zampieri. "State Space Realization of 2-D Finite-Dimensional Behaviours". In: *SIAM Journal on Control and Optimization* 31.6 (1993), pp. 1502–1517.

[6] Philippe Lemmerling and Bart De Moor. "Misfit versus latency". In: *Automatica* 37.12 (2001), pp. 2057–2067.

[7] Christof Vermeersch and Bart De Moor. "Two complementary block Macaulay matrix algorithms to solve multiparameter eigenvalue problems". In: (2022). KU Leuven, Leuven, Belgium, Tech. Rep.

Acknowledgements: 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, and 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).