2019 18th European Control Conference (ECC)
Napoli, Italy, June 25-28, 2019

Least squares realization of LTI models is an eigenvalue problem

Bart De Moor, Fellow IEEE and SIAM

***Abstract*— We show how least squares optimal realization of autonomous linear time-invariant dynamical systems from given data, reduces to the solution of an eigenvalue problem. In this short paper, we can only schematically sketch the different steps: The first order optimality conditions result in a multi-parameter eigenvalue problem. The eigenvalue n-tuples are calculated from the null space of a quasi-Toeplitz block Macaulay matrix, which is shown to be multishift-invariant. This last property is then exploited via nD 'exact' realization theory, leading through several eigenvalue problems to the optimal model parameters.**

I. INTRODUCTION

Eigenvalue problems (EVP) are ubiquitous in nature and science. Specifically in systems and control, they are also omnipresent: in characterizing stability, controllability and observability of linear time-invariant (LTI) dynamical systems, but even so in the steady state versions of LQR control and the Kalman filter [9], the solutions of which follow from Algebraic Riccati Equations, which are Hamiltonian eigenvalue problems in disguise. Similarly, the solutions to their H_∞ counterparts derive from symplectic eigenvalue problems [9]. As a final example, model reduction in the Hankel norm is in essence an eigenvalue problem [12].

But *what about the dynamical models* themselves, when, as in system identification, they are estimated from observations on inputs and outputs? *In what sense are these identified models optimal*, whether they are obtained from subspace algorithms [15], prediction error methods [11] or errors-in-variables approaches [14]? Typically, the identification problem is formulated as a one- or two-step (weighted) least squares optimization problem, in which parametrized models act as constraints. 'Nonlinearities' occur in the interplay between the dynamic model parametrization and the assumptions by which data inaccuracies or unknown inputs are modelled. As a matter of fact, most model and data assumptions for LTI system identification lead to *nonlinear* least squares problems, the solution of which is tackled with iterative minimization algorithms, with typical challenges of finding good initial guesses, dealing with convergence issues and having no guarantee to converge to a global minimum.

However, one of the central observations that we will exploit, concerns the specific nature of these 'nonlinearities' in LTI system identification: They are all 'multivariate polynomial', and therefore in essence the solution reduces to an EVP. The close relation between roots of sets of multivariate polynomials on the one hand, and the algebraic EVP on the other hand, has been rediscovered many times in algebraic geometry (see e.g. [4] [13])[1]. In our work, we also have made the link with multidimensional realization theory [6], a relation that will be further exploited here too.

In this paper, we will concentrate on the so-called *least squares realization problem*: How to modify a data sequence y_k in a least squares sense into a sequence $\hat{y}$ that is the output of an LTI autonomous system of a given prespecified order n_a? Without going too much in detail, we will elaborate on the (non-trivial) steps that lead to the insight that this is an EVP as follows: In Section II, we formalize the least squares realization problem. First order optimality conditions are the subject of Section III, leading to some interesting system theoretic properties detailed in Section IV. The insight in Section V concerns the observation that the first order optimality conditions reduce to a multiparameter EVP. In Section VI, we show how to exploit multishift-invariance properties of certain subspaces, derived from the multiparameter EVP, to invoke nD-realization theory, that will ultimately deliver the optimizing parameters as eigenvalues of certain matrices. In the final Section, we offer some concluding remarks.

II. THE LEAST SQUARES REALIZATION PROBLEM

A. Data and model assumptions

Let $y = (y_0 \;\; y_1 \;\; \ldots y_{N-1})^T \in \mathbb{R}^N$ be a given data sequence. The least squares realization problem is to find $\hat{y} = (\hat{y}_0 \;\; \hat{y}_1 \;\; \ldots \hat{y}_{N-1})^T \in \mathbb{R}^N$ so that $\sigma^2 = \|y - \hat{y}\|_2^2 = \sum_{k=0}^{N-1}(y_k - \hat{y}_k)^2$ is minimized, with $\hat{y}$ the output of an LTI model of given, prespecified order n_a:

$$\hat{y}_k = CA^k x_0 \;, \tag{1}$$

with the unknown initial state $x_0 \in \mathbb{R}^{n_a}$, system matrix $A \in \mathbb{R}^{n_a \times n_a}$ and output vector $C \in \mathbb{R}^{1 \times n_a}$. We assume that the number of data points N is large enough (i.e. $N \geq 2n_a$, the minimal number of parameters to describe (1)) and that y itself is not the output of an LTI autonomous system[2] of order $n \leq n_a$. Let $\chi(A) = \det(\lambda I_{n_a} - A) = \lambda^{n_a} + \alpha_1 \lambda^{n_a - 1} + \ldots + \alpha_{n_a - 1}\lambda + \alpha_{n_a} = 0$ be the monic (with leading coefficient 1) characteristic equation of the matrix A. Then, from the Cayley-Hamilton Theorem, for every $k \geq 0$, $n_a + 1$ consecutive samples $\hat{y}_k$ satisfy the difference equation

$$\hat{y}_{k+n_a} + \alpha_1 \hat{y}_{k+n_a-1} + \ldots + \alpha_{n_a-1}\hat{y}_{k+1} + \alpha_{n_a}\hat{y}_k = 0, \tag{2}$$

The author is with the division ESAT-STADIUS of the Department of Electrical Engineering, KU Leuven, Belgium and holds the chair *Health Care System Quality and Accessibility*, endowed by 'CM Health Insurance'. E: bart.demoor@kuleuven.be, W: www.bartdemoor.be.

[1] It must have been known to Sylvester, who developed an elimination theory for variables in a set of multivariate polynomials, which reduces the whole problem to finding the roots of a polynomial in one variable only.

[2] This is a 'sufficiently richness' condition on the data, relative to the given order n_a. Without this assumption, the model (2) would be overparametrized.

so that

$$T_a\ \hat{y} = 0\ , \tag{3}$$

where $T_a \in \mathbb{R}^{(N-n_a)\times N}$ is a banded Toeplitz matrix with the coefficients of the characteristic polynomial reversed, i.e. its first row is $(\alpha_{n_a}\ \alpha_{n_a-1}\ \ldots\ \alpha_1\ 1\ 0 \ldots 0)$. The fact that $\chi(A)$ is monic implies that T_a is of full row rank: $\mathrm{rank}(T_a) = N - n_a$. Eq. (3) is a so-called 'kernel representation' in Willems's behavioral framework [17]: Only sequences $\hat{y}$ that are orthogonal to the row space of T_a, are 'compatible' with the model class (1).

B. Minimizing the misfit

We define the misfit $\tilde{y}$ via $y = \hat{y} + \tilde{y}$, so that $\sigma^2 = \|y - \hat{y}\|_2^2 = \|\tilde{y}\|_2^2 = \tilde{y}^T\tilde{y}$. From (3) we find that $T_a y = T_a\tilde{y}$. If T_a were known, this would be an underdetermined set of linear equations in the unknown misfit $\tilde{y}$. The unique minimum norm solution would then follow from the pseudo-inverse of T_a as

$$\tilde{y} = T_a^\dagger T_a y = T_a^T(T_aT_a^T)^{-1}T_a y\ . \tag{4}$$

The second equality follows from the fact that T_a is of full row rank, so that $T_aT_a^T$ is nonsingular and $T_a^\dagger = T_a^T(T_aT_a^T)^{-1}$. The matrix $\Pi_a = T_a^T(T_aT_a^T)^{-1}T_a$ is the orthogonal projector onto the row space of T_a. If T_a were known, the given data vector y could be decomposed into two mutually orthogonal vectors as

$$y = \hat{y} + \tilde{y} = (I_N - \Pi_a)y + \Pi_a y\ . \tag{5}$$

In general, there is an infinite number of orthogonal decompositions of the data vector y: Construct in $\mathbb{R}^N$ a sphere with center $y/2$ and radius $\|y/2\|_2$. Then, the set of all vectors $\hat{y}$ so that $y = \hat{y} + \tilde{y}$ and $\hat{y}^T\tilde{y} = 0$, is given by the equation of that sphere as $\|\hat{y} - y/2\|_2^2 = \|y/2\|_2^2$, from which it follows that $\hat{y}^T(y-\hat{y}) = 0$ and $\|y\|_2^2 = \|\hat{y}\|_2^2 + \|\tilde{y}\|_2^2$ (the last equation being Pythagoras's Theorem)[3]. But here, out of this infinite set of orthogonal projectors, we want to find that specific projector $\Pi_a = T_a^T(T_aT_a^T)^{-1}T_a$ that minimizes $\|\tilde{y}\|_2^2$ over the coefficients of $\chi(A)$.

III. First order optimality conditions

The least squares objective function can now be written as

$$\sigma^2 = \|\tilde{y}\|_2^2 = \|\Pi_a y\|_2^2 = y^T T_a^T(T_aT_a^T)^{-1}T_a y\ , \tag{6}$$

which is to be minimized over the coefficients $\alpha_i, i = 1,\ldots,n_a$ of the characteristic polynomial $\chi(A)$ of given degree n_a. We will use $D_a = T_aT_a^T$, which itself is a symmetric, positive definite, banded Toeplitz matrix.
The first order optimality conditions are, $\forall i = 1,\ldots,n_a$:

$$\begin{aligned}\frac{\partial\sigma^2}{\partial\alpha_i} \;=\; 0 \;=\; & 2y^T T_a^T D_a^{-1} T_a^{\alpha_i} y \\ & -y^T T_a^T D_a^{-1} D_a^{\alpha_i} D_a^{-1} T_a y,\end{aligned} \tag{7}$$

[3]This is a generalization to N dimensions of the ancient Thales's Theorem, stating that when a, b and c are distinct points on a circle where the line ac is the centerline diameter, the angle between ab and bc is a right one.

where a superscript α_i denotes the partial derivative with respect to α_i, and we have used the matrix inverse derivative expression $\partial D_a^{-1}/\partial\alpha_i = -D_a^{-1}D_a^{\alpha_i}D_a^{-1}$. The observation that eqs. (7) are 'nonlinear' in the coefficients α_i, has led to a lot of heuristic algorithms in the past (see Section VIII). However, as $D_a^{-1} = \mathrm{adj}(D_a)/\det(D_a)$, where $\mathrm{adj}(D_a)$ is the adjugate of the matrix D_a (the matrix transposed with the cofactors of all elements of D_a), and because $\det(D_a) \neq 0$, these n_a equations (7) are equivalent to multivariate polynomials in the n_a unknowns α_i, after 'multiplying out' $\det(D_a)$. Their roots will contain all global and local minima and maxima, and all saddle points of the objective function (6).
Now, the relation between common roots of sets of multivariate polynomials on the one hand and the matrix EVP on the other hand, is (not so) well known (see e.g. [4] [13]). For an interpretation via multidimensional system realization theory, which we will also follow here, we refer to [6]. In any case, not obvious at first sight, eqs. (7) are equivalent to one or several eigenvalue problems, as we will elaborate on in the sequel.
Define the vector $f = D_a^{-1}T_a y \in \mathbb{R}^{N-n_a}$ and rewrite (6) as

$$\begin{pmatrix} D_a & T_a y \\ y^T T_a^T & \sigma^2 \end{pmatrix}\begin{pmatrix} f \\ -1 \end{pmatrix} = 0\ . \tag{8}$$

Taking partial derivatives with respect to all variables $\alpha_i, i = 1,\ldots,n_a$, and using the derivative chain rule, results in

$$\begin{aligned}&\begin{pmatrix} D_a^{\alpha_i} & T_a^{\alpha_i} y \\ y^T(T_a^{\alpha_i})^T & 0 \end{pmatrix}\begin{pmatrix} f \\ -1 \end{pmatrix} \\ &\qquad + \begin{pmatrix} D_a & T_a y \\ y^T T_a^T & \sigma^2 \end{pmatrix}\begin{pmatrix} f^{\alpha_i} \\ 0 \end{pmatrix} = 0.\end{aligned} \tag{9}$$

Eqs. (8) and (9) contain $(n_a+1)(N-n_a+1)$ equations, and the number of unknowns is the same ($(N-n_a)$ in f, 1 in σ, $n_a(N-n_a)$ in all f^{α_i} and n_a for all of the α_i).
Two observations can be made: Firstly, the last equation in (8) is the only one involving σ, because the last component of the last vector in (9) is 0. Secondly, we can easily recover the secular eqs. (7) from (9) by eliminating f^{α_i} from the first block rows in (9), plugging it in into the second block row and using $f = D_a^{-1}T_a y$ from (8).

IV. Some system theoretic properties

Before showing that (9) is a multiparameter EVP, let's first discuss some system theoretic properties.

A. Hankel matrices, shift invariance and realization theory

Eq. (3) can be rewritten as

$$T_a\hat{y} = \hat{Y}a = 0, \tag{10}$$

where $\hat{Y} \in \mathbb{R}^{(N-n_a)\times(n_a+1)}$ is a Hankel matrix generated from the sequence $\hat{y}_k$, and $a = (\alpha_{n_a}\ \alpha_{n_a-1}\ \ldots\ \alpha_1\ 1)^T \in \mathbb{R}^{(n_a+1)}$, contains the coefficients of $\chi(A)$ in reversed order. Eq. (10) expresses the well-known fact that a Hankel matrix, generated from a sequence $\hat{y}_k$ generated by the model (1),

has rank n_a. Notice that eqs. (2) and (10) are exactly the same, so that

$$T_a\Gamma = T_a \begin{pmatrix} C \\ CA \\ \cdots \\ CA^{N-1} \end{pmatrix} = 0 \,, \tag{11}$$

where Γ is the (extended) observability matrix. This equation shows that the row space of T_a and the column space of Γ are orthogonal and complementary. The column space $\mathbf{R}(\Gamma)$ of Γ is *shift-invariant*, meaning that $\mathbf{R}(\overline{\Gamma}) = \mathbf{R}(\underline{\Gamma})$, so that also

$$n_a = \mathrm{rank}(\Gamma) = \mathrm{rank}(\overline{\Gamma}) = \mathrm{rank}(\underline{\Gamma}) = \mathrm{rank}(\underline{\Gamma}\ \overline{\Gamma}), \tag{12}$$

where $\overline{\Gamma}$ and $\underline{\Gamma}$ are obtained from Γ by omitting its first, resp. last row. This property is sometimes called the *partial realization condition* and in fact follows directly from the Theorem of Cayley-Hamilton[4]. For single-output systems, the column space of Γ is uniquely determined by the eigenvalue spectrum of the system matrix A. It is obvious that $\underline{\Gamma}A = \overline{\Gamma}$, and, even though the choice of basis for the column space of Γ is not unique, this 'shift'-property is independent of the choice of basis as for any nonsingular matrix T, which transforms Γ into ΓT, we have:

$$(\underline{\Gamma}T)(T^{-1}AT) = (\overline{\Gamma}T). \tag{13}$$

Because the eigenvalues of A and $T^{-1}AT$ are the same (T is a similarity transformation), we can find the eigenvalues that characterize the shift-invariant column space of ΓT (independent of the specific choice of T), from the eigenvalues of

$$(T^{-1}AT) = (\underline{\Gamma}T)^{\dagger}(\overline{\Gamma}T) \,, \tag{14}$$

where a '$\dagger$' denotes the pseudo-inverse. There is also an interesting 'duality' between the banded Toeplitz structure of T_a, and the shift-invariance of Γ: The column space of Γ is uniquely determined by the eigenvalues of A, which are the zeros of the polynomial a from which T_a is constructed. Said in other words, the banded Toeplitz structure of T_a and the shift-invariance property of Γ go 'hand-in-hand', as expressed by (11) (a duality that will be worked out in more detail elsewhere).

B. Beurling-Lax-Halmos: The misfit is structured

From (4), we get

$$\tilde{y} = T_a^T(T_aT_a^T)^{-1}T_a y = T_a^T f. \tag{15}$$

Using the forward shift operator z defined as $z(y_k) = y_{k+1}$, we can write the difference equation (2) as $a(z)\hat{y}_k = 0$, where $a(z)$ has the same coefficients and roots as $\chi(A)$. Eq. (15) implies that the misfit sequence $\tilde{y}_k$ is generated by a FIR filter, the coefficients of which are those of $a(z)$ in reversed order (which we denote by $a^{\mathrm{rev}}(z)$) and driven with the 'input' sequence f_k (appropriately padded with zeros) as $\tilde{y}_k = [(a^{\mathrm{rev}}(z))/z^{n_a}]f_k$. The zeros of this FIR filter are the inverses of the poles of the optimal approximating model $a(z)$. So we find that $\tilde{y}$ itself is generated by a linear system, the zeros of which are the inverses of the eigenvalues that characterize the shift-invariant column space of Γ. This can be seen as a finite-dimensional vector space version of the operator-theoretic Theorem of Beurling-Lax-Halmos (see e.g. [7] [12]).

[4]There is some technical complication when A is singular, case which we do not consider here.

C. Optimizing a metric and orthogonality

When dealing with least squares problems, there is always a lot of geometry going on (we have already given some examples using the Theorems of Pythagoras and Thales). One interpretation of (6) is that we try to find the optimal metric, represented by the nonnegative definite symmetric projection operator matrix Π_a, in the following sense: The set of all row spaces of T_a over all possible vectors a, is a manifold. We are looking for an optimal choice of a, so that the orthogonal decomposition of the data vector y as in (5), is such that the norm of $\tilde{y}$ as measured in the metric induced by Π_a, $\sigma^2 = \|\tilde{y}\|^2 = \tilde{y}^T\Pi_a\tilde{y}$, is minimized.

There are many other properties lurking in the background that require more attention. To give one example, let's concentrate on (9). From the first block row, we find $f^{\alpha_i} = -D_a^{-1}D_a^{\alpha_i}f + D_a^{-1}T_a^{\alpha_i}y$, and substituting this in the second block row $y^T(T_a^{\alpha_i})^Tf + y^TT_a^Tf^{\alpha_i} = 0$ and using $f = D_a^{-1}T_ay$, we find $2f^TT_a^{\alpha_i}y = f^TD_a^{\alpha_i}f$. Now, we use $D_a^{\alpha_i} = (T_aT_a^T)^{\alpha_i} = T_a^{\alpha_i}T_a^T + T_a(T_a^{\alpha_i})^T$ and (15) to find $f^TT_a^{\alpha_i}(y - \tilde{y}) = 0$. Writing this out for all α_i and using $\hat{y} = y - \tilde{y}$, we then obtain (illustrated here for $N = 6$ and $n_a = 2$):

$$f^T \begin{pmatrix} \hat{y}_0 & \hat{y}_1 \\ \hat{y}_1 & \hat{y}_2 \\ \hat{y}_2 & \hat{y}_3 \\ \hat{y}_3 & \hat{y}_4 \end{pmatrix} = 0 \,.$$

From (1) and (11), we deduce that f itself belongs to the left null space of an observability matrix, so that similarly to (15) and the reasoning that follows, f itself is the output of a FIR filter driven by an unknown signal g. Said in other words, the misfit $\tilde{y}$ is generated by filtering an unknown signal g twice through the same FIR filter, the zeros of which are the inverses of the eigenvalues that characterize the optimal shift-invariant column space of Γ: $f = (a^{rev}(z))^2g$ for some g. This is a finite-dimensional vector space version of what in the H_2-model reduction literature is known as Walsh's Theorem (see e.g. [12]).

V. Multiparameter eigenvalue problem

In (7), when $\det(D_a)$ is multiplied out, there are n_a unknowns α_i in the n_a multivariate polynomials, while in (8)-(9) there are $(n_a+1)(N-n_a+1)$ equations and unknowns, but here, the vectors f and f^{α_i} appear linearly. This is comparable to the ordinary EVP for a matrix $A \in \mathbb{R}^{n\times n}$, where both the formulations $Ax = x\lambda, \|x\| = 1$, and $\det(\lambda I_n - A) = 0$ are equivalent. The former contains $n+1$ equations and unknowns (x and λ) and the latter only 1 (λ). Indeed, the fact that the eigenvectors x appear linearly in the EVP, $Ax = x\lambda$, allows one to write $(\lambda I_n - A)x = 0$, which

only has a nontrivial solution $x \neq 0$ if and only if the characteristic polynomial $\chi(A) = \det(\lambda I_n - A) = 0$. We will now show how (8)-(9) is a multiparameter eigenvalue problem, where the vectors f and f^{α_i} are (parts of) eigenvectors, the α_i generate n_a-tuples $(\alpha_1, \ldots, \alpha_{n_a})$ of eigenvalues, and the secular equations (7) are the multiparameter generalizations of the characteristic equation in the ordinary EVP.

Contrary to 1-parameter eigenvalue problems (including the Jordan, Weierstrass and Kronecker canonical forms, or also eigenvalue problems polynomial in one variable), multiparameter eigenvalue problems are much less studied and described in the literature. Some references include [1] [3] [16].

We will omit the equation that defines σ^2 in (8) (the last row), as it is the only equation where σ appears, in which case eqs. (8) - (9) comprise $(n_a+1)(N-n_a)+n_a$ equations and unknowns. As an example, for $n_a = 2$, we then have

$$\begin{array}{c} \\ N-2 \\ N-2 \\ N-2 \\ 1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & N-2 & N-2 & N-2 \end{array} \\ \begin{pmatrix} T_a y & D_a & 0 & 0 \\ T_a^{\alpha_1} y & D_a^{\alpha_1} & D_a & 0 \\ T_a^{\alpha_2} y & D_a^{\alpha_2} & 0 & D_a \\ 0 & y^T (T_a^{\alpha_1})^T & y^T T_a^T & 0 \\ 0 & y^T (T_a^{\alpha_2})^T & 0 & y^T T_a^T \end{pmatrix} \end{array} z_0 = 0, \tag{16}$$

where $z_0 = (-1 \;\; f^T \;\; (f^{\alpha_1})^T \;\; (f^{\alpha_2})^T)^T$. The matrix in (16) is a function of the given data y and the unknown coefficients α_i, which appear quadratically in D_a and linearly in $D_a^{\alpha_i}$ and T_a. The pair (α_1, α_2) is called the eigenvalue-pair of the multiparameter eigenvalue problem. The eigenvector w contains f, f^{α_1} and f^{α_2}, which appear linearly in the equations. Notice that the matrix is rectangular, 1 more row than column for this $n_a = 2$ case, so that its rank deficiency is non-trivial.

For general n_a, there is an n_a-tuple $(\alpha_1, \ldots, \alpha_{n_a})$ of eigenvalues and an eigenvector containing the vectors $f, f^{\alpha_1}, \ldots, f^{\alpha_{n_a}}$. The matrix will be of size $(n_a(N-n_a)+n_a) \times (n_a(N-n_a)+1)$, so with more rows than columns. Its rank deficiency can be expressed by requiring that the determinants of certain submatrices (n_a of them) are simultaneously zero, which are exactly the 'secular' equations (7), replacing the one characteristic equation in the 1-parameter case $\chi(A) = 0$, with n_a equations in the n_a-variable case.

VI. Solving the multi-parameter eigenvalue problem

How to solve multiparameter eigenvalue problems like the ones we just derived? Most often in the literature, one resorts to nonlinear equations solvers to find solutions for (7), and then, obtain the eigenvectors by solving a set of (homogeneous) linear equations. In doing so, the problem is that, one needs appropriate initial guesses to start up the nonlinear iterations, one has to monitor the convergence behavior and it is not known beforehand how many solutions there might be.

Here, we outline a different approach. By multiplying each of the equations in (8)-(9) with monomials formed from the n_a unknown coefficients α_i, we create additional equations that will be grouped into a so-called block Macaulay matrix. We then show that the null space of this matrix is multishift invariant, a property that will be exploited to find the n_a unknowns α_i as the eigenvalues of certain matrices. We exploit the duality between the quasi-Toeplitz structure of this block Macaulay matrix and the multishift- invariances of its null space, for which we also describe the system theoretic structure that allows to identify the finite (affine) solutions and possible solutions at infinity.

For the sake of clarity, we now proceed with the simple cases of $n_a = 1$ and $n_a = 2$, but the results can perfectly be generalized for arbitrary $n_a > 2$.

A. Shift-invariant null space for $n_a = 1$

For $n_a = 1$, there is only 1 unknown α and the multiparameter eigenvalue problem reduces to one with 1 parameter only. In that case, eqs. (8)-(9) (without the equation for σ^2) can be combined to a 1-parameter eigenvalue problem that is polynomial in α as

$$(A_0 + A_1\alpha + A_2\alpha^2) z_0 = 0, \tag{17}$$

where $A_0, A_1, A_2 \in \mathbb{R}^{(2N-1)\times(2N-1)}$ and $z_0 = (-1 \; f^T \; (f^{\alpha})^T)^T$. A_0 contains all the coefficients of $\alpha^0 = 1$, A_1 those of α and A_2 those of α^2.

There are now two ways to proceed. The first one is to define $z_1 = z_0\alpha$ and then obtain the generalized EVP:

$$\begin{pmatrix} 0 & I_{2N-1} \\ A_0 & A_1 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = \begin{pmatrix} I_{2N-1} & 0 \\ 0 & -A_2 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \alpha.$$

Another way to proceed, is to start from (17), rewrite it as

$$\begin{pmatrix} A_0 & A_1 & A_2 \end{pmatrix} \begin{pmatrix} z_0 \\ z_0\alpha \\ z_0\alpha^2 \end{pmatrix} = 0, \tag{18}$$

and then create a null space with a shift-invariant structure. Hereto, generate a block Toeplitz matrix by multiplying (18) with increasing powers of α to get

$$\begin{pmatrix} A_0 & A_1 & A_2 & 0 & 0 & \ldots \\ 0 & A_0 & A_1 & A_2 & 0 & \ldots \\ 0 & 0 & A_0 & A_1 & A_2 & \ldots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \begin{pmatrix} z_0 \\ z_0\alpha \\ z_0\alpha^2 \\ z_0\alpha^3 \\ \vdots \end{pmatrix} = 0.$$

In this way, we can recursively construct larger and larger block Toeplitz matrices. From a certain recursion step on, the nullity, which is the dimension of the null space, will stabilize. It counts the total number of solutions, including the ones at infinity (This is equivalent to Bezout's Theorem in Algebraic Geometry, see [4]). One can then always find a column compression, as explained in the next Section VII, that separates the affine (=finite) roots in α from those at infinity. The resulting 'affine' subspace is shift-invariant and one can then apply realization theory on the affine part as in (13) and (14) to find all affine α's as the eigenvalues that characterize this shift-invariant subspace.

B. Shift invariant null space for $n_a = 2$

We start from the multiparameter EVP (16). Define the vector $z_0^T = (-1 \;\; f^T \;\; (f^{\alpha_1})^T \;\; (f^{\alpha_2})^T \;\;)^T$, and matrices A_{ij} that contain the coefficients of the monomials $\alpha_1^i \alpha_2^j$ with $i, j = 0, 1, 2$. Write (16) as a multiparameter EVP:

$$(A_{00}+A_{10}\alpha_1+A_{01}\alpha_2+A_{20}\alpha_1^2+A_{11}\alpha_1\alpha_2+A_{02}\alpha_2^2)z_0 = 0.$$

Next, we 'enlarge' this multiparameter eigenvalue problem by multiplying it with monomials in α_1, α_2 of increasing degree (a first recursion with α_1 and α_2, a second recursion with α_1^2, $\alpha_1\alpha_2$, α_2^2, a third recursion with α_1^3, $\ldots$, etc.) so that we get

$$\begin{array}{c} \begin{array}{ccccccccc} 1 & \alpha_1 & \alpha_2 & \alpha_1^2 & \alpha_1\alpha_2 & \alpha_2^2 & \alpha_1^3 & \ldots \end{array} \\ \begin{pmatrix} A_{00} & A_{10} & A_{01} & A_{20} & A_{11} & A_{02} & 0 & \ldots \\ 0 & A_{00} & 0 & A_{10} & A_{01} & 0 & A_{20} & \ldots \\ 0 & 0 & A_{00} & 0 & A_{10} & A_{01} & 0 & \ldots \\ 0 & 0 & 0 & A_{00} & 0 & 0 & A_{10} & \ldots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \end{array}$$
$$\times (z_0^T \;\; z_0^T\alpha_1 \;\; z_0^T\alpha_2 \;\; z_0^T\alpha_1^2 \;\; z_0^T\alpha_1\alpha_2 \;\; z_0^T\alpha_2^2 \;\; z_0^T\alpha_1^3 \;\; \ldots)^T = 0. \tag{19}$$

We get a quasi-Toeplitz block Macaulay matrix, the null space of which consists of vectors that are block multishift-invariant, a property that we elaborate on in the next section.

VII. nD realization theory in multi-shift invariant subspaces

It can be shown (see e.g. [6]) that there is a non-trivial null space of the quasi-Toeplitz block Macaulay matrix, the nullity of which will 'stabilize' after a sufficient number of recursions (provided the solution set of the multiparameter EVP is zero-dimensional, i.e. the corresponding variety has dimension 0). This nullity reveals the total number of eigenvalue-pairs (α_1, α_2) of the multi-parameter EVP, including solutions at infinity. Let the matrix Γ represent the null space of the quasi-Toeplitz block Macaulay matrix (e.g. obtained from its SVD). Then, it can be shown that there is always a non-singular column compression T from the right, that transforms Γ as follows:

$$\begin{array}{rcl} \Gamma T & = & \begin{array}{cc} \overset{n_1}{} & \overset{n_2}{} \end{array} \\ & & \left(\begin{array}{cc} \Gamma_R & \Gamma_S \end{array}\right) \\ & = & \left(\begin{array}{cc} C_R & 0 \\ \hline C_RA_1 & 0 \\ C_RA_2 & 0 \\ \hline C_RA_1^2 & 0 \\ C_RA_1A_2 & 0 \\ C_RA_2^2 & 0 \\ \hline \vdots & \vdots \\ \hline C_RA_1^{n_1-1} & 0 \\ C_RA_1^{n_1-2}A_2 & 0 \\ \vdots & \vdots \\ C_RA_2^{n_1-1} & 0 \\ \hline\hline C_RA_1^{n_1} & 0 \\ \vdots & \vdots \\ \hline\hline \vdots & * \\ \vdots & * \\ \hline \vdots & * \\ \vdots & * \end{array}\right) \end{array} \tag{20}$$

This specific column compression T can be found following several rank tests and SVDs (details omitted). T will split ΓT into Γ_R ('R' from regular) and Γ_S ('S' from singular), which together form the observability matrix of a singular multi-dimensional system, of order $n = n_1 + n_2$, where n_1 is the number of affine (finite) pairs of eigenvalues (α_1, α_2), and n_2 is the number of eigenvalue pairs at infinity.

As one can see in (20), there are three block row zones in the matrix ΓT (separated by the double lines in (20)). These zones can be determined when one starts checking the ranks of submatrices of ΓT, from the top to the bottom, by including more and more blocks in the rank test. For a sufficiently large number of recursions in monomials of α_1 and α_2, there will be 3 block row zones:

Zone I: the 'regular'-zone: The rank increases with at least 1 per block, up to the block where the block row degree is $n_1 - 1$;

Zone II: 'Mind-the-gap'-zone: The rank does not increase anymore (this is a generalization of Cayley-Hamilton for pairs of matrices). In this zone, all rows are linearly dependent on some of the rows in Zone I.

Zone III: 'A-bout-du-souffle'-zone: The rank increases again per block, until the rank of ΓT equals the total nullity of the block Macaulay matrix. The zeros at infinity can be found from the structures and properties in Zone III (details omitted).

It can be shown that these observability matrices Γ_R and Γ_S are generated by a singular 2D state space system of the form (already converted in what could be called the Weierstrass Canonical Form for 2D commutative linear systems):

$$\begin{aligned} x^R_{k+1,l} &= A_1 x^R_{k,l}, \\ x^S_{k-1,l} &= E_1 x^S_{k,l}, \\ x^R_{k,l+1} &= A_2 x^R_{k,l}, \\ x^S_{k,l-1} &= E_2 x^S_{k,l}, \\ y_{k,l} &= C_R x^R_{k,l} + C_S x^S_{k,l}. \end{aligned}$$

Here, for the case $n_a = 2$, $x^R_{k,l}$ is the regular part of the state, governed by two discrete indices k and l. The 2D-grid state propagation over increasing k is modelled by A_1, while that over increasing l is governed by A_2. Together they form the dynamics of the regular state $x^R_{k,l}$. Here, $A_1, A_2 \in \mathbb{R}^{n_1 \times n_1}$ commute: $A_1A_2 = A_2A_1$ (Intuitively, they should, as $x^R_{k+1,l+1}$ can be reached from $x^R_{k,l}$ in 2 different ways as $x^R_{k+1,l+1} = A_1 x^R_{k,l+1} = A_1A_2x^R_{k,l} = A_2x^R_{k+1,l} = A_2A_1x^R_{k,l}$, which should hold for arbitrary $x^R_{k,l}$). The singular part of the state is $x^S_{k,l}$, which propagates backward both in k and l via the matrices E_1 and E_2, where $E_1, E_2 \in \mathbb{R}^{n_2 \times n_2}$ also commute: $E_1E_2 = E_2E_1$ and in addition, E_1 and E_2 are nilpotent, i.e. when powered up, from a certain power on, we get a zero matrix. In (20), we do not explicitly show the backward propagation structure of the singular observability matrix Γ_S: we have replaced it in zone III by '*'-s, as it is not relevant for our present discussion. In any case, in Γ_S there will be only nonzero elements in Zone III: starting from the bottom we get blocks with

increasing powers of $E_1^i E_2^j$ that die out in Zone II because of the nilpotency (hence the name 'a-bout-du-souffle'). The Jordan structure of E_1 and E_2 reveals the eigenvalue-pair structure at infinity, but we will not discuss that here any further. Finally, there are the output matrices $C_R \in \mathbb{R}^{l_R \times n_1}$, $C_S \in \mathbb{R}^{l_S \times n_2}$ with l_R and l_S the specific numbers of outputs that follow from the multiparameter EVP (19). The 'mind-the-gap' zone will only appear for a sufficiently large block Macaulay matrix: From a certain number of recursions on, all rows in Zone II will be linearly dependent on rows in Zone I, and due to the nilpotency of E_1 and E_2, there will only be zero rows of Γ_S in Zone II. As the block Macaulay matrix keeps on growing, the 'gap' between Zone I and Zone III becomes wider and wider.

The column space of the regular observability matrix Γ_R is a *block multishift-invariant subspace*. Denote by Γ_1 the submatrix of Γ_R that contains the first block rows up to degree $n_1 - 1$ (so all block rows of Zone I). One can now verify that (recall that A_1 and A_2 commute):

$$\Gamma_1 \, A_1 \; = S_1 \Gamma_R \;\text{ and }\; \Gamma_1 \, A_2 \; = S_2 \Gamma_R, \qquad (21)$$

where the selector matrix S_1 selects the block rows $(2, 4, 5, 7, 8, 9, \ldots)$ and the selector matrix S_2 selects the block rows $(3, 5, 6, 8, 9, 10, \ldots)$. So, from the column compressed ΓT, we can now find A_1 and A_2 by exploiting this multishift-invariance property (21) to find

$$A_1 = \Gamma_1^{\dagger}(S_1 \Gamma_R) \;\text{ and }\; A_2 = \Gamma_1^{\dagger}(S_2 \Gamma_R) \; .$$

The eigenvalue pairs (α_1, α_2) then follow from the eigenvalues α_1 of A_1 and α_2 of A_2.

VIII. Concluding remarks

In this short paper, we could only schematically sketch an outline of some major results: Least squares optimal realization of observed data is basically an EVP. It is surprising that a (difficult) nonlinear problem in a 1D system theoretic setting, can in principle be solved exactly as an EVP, in an n_a-dimensional system theoretic setting. The required steps involve writing the first order optimality conditions as a multiparameter EVP, next, using recursions to generate a block Macaulay matrix, finding in its null space the regular part that is multishift invariant. Then use n_a-dimensional realization theory to calculate matrices $A_1, \ldots, A_{n_a}$, the eigenvalues of which will generate the n_a-tuples $(\alpha_1, \ldots, \alpha_{n_a})$, one of which corresponds to the global minimum. In doing so, we also presented a new solution method for multiparameter EVP.

Of course, many more details will be discussed elsewhere, but let's make some final observations. This work is a nice combination of several disciplines, like numerical linear algebra, system theory in one and more dimensions, (commutative) algebraic geometry, operator theory, etc. Many heuristic algorithms have been described in the literature for heuristically 'solving' the least squares realization problem, often as an Alternating Projection Method: Iterative Quadratic Maximum Likelihood (see e.g. [10]), Steiglitz-McBride (see e.g. [12]), Cadzow's iteration (see e.g. [2]), the Riemannian SVD (see [5]), plain numerical optimization (like in PEM in [11]), weighted null space fitting [8], etc. We hope that our results will shed some new light in understanding these heuristic approaches (e.g. the nature of their 'fixed points' if and to which they converge). As a matter of fact, there are numerous applications where the underlying model is of the form (1) (e.g. moment problems, impulse response realization, 'shape-from-moment' problems, nD texture modelling, signal processing approaches like MUSIC and ESPRIT, etc...), for which the results presented here are highly relevant. The fact that also new algorithms to obtain all solutions to multiparameter eigenvalueproblems can be derived using the ideas established here, forms another exciting perspective.

Acknowledgments: Work funded by *FWO*: EOS Project no 30468160 (SeLMA), PhD grants; *VLAIO*: Projects (COT.2018.018), PhD grants (HBC2017.0539), Industrial Projects (HBC.2018.0405); *Eur. funding*: EU H2020-SC1-2016-2017 Grant No.727721: MIDAS Meaningful Integration of Data, Analytics and Services; *KU Leuven*: C16/15/059, C32/16/013, C24/18/022.

References

[1] Atkinson F., Mingarelli A. *Multiparameter Eigenvalue Problems; Sturm-Liouville Theory*. CRC Press, Taylor and Francis Group, 2011, 279 pp.

[2] Cadzow, J. *Signal enhancement: a composite property mapping algorithm*. IEEE Trans. on Acoustics, Speech, and Signal Processing, 36(2) p. 4962, 1988.

[3] Carmichael R.D. *Boundary Value Problems and Expansion Problems.. Part I: Algebraic Basis of the Theory*: Amer. J. Math, 43, p. 69-101, 1921; *Part II: Formulation of Various Transcendental Problems*: Amer. J. Math, 43, p. 232-270, 1921; *Part III: Oscillatory, companion and expansion problems*: Amer. J. Math, 44, p. 129-152, 1922.

[4] Cox D., Little J., O'Shea D. *Ideals, Varieties and Algorithms*, 3rd Ed., Springer, 2007, 551 pp.

[5] De Moor B. *Total least squares for affinely structured matrices and the noisy realization problem*. IEEE Trans. on Signal Processing. 42(11) p. 3104 - 3113, 1994.

[6] Dreesen P., Batselier K., De Moor B. *Multidimensional realisation theory and polynomial system solving.* International Journal of Control, vol. 91, no. 12, 2018, pp. 2692-2704.,

[7] Foias C., Frazho A. *The Commutant Lifting Approach to Interpolation Problems.* Birkhauser, Operator Theory: Advances and Applications, vol.44, 1990, 632 pp.

[8] Galrinho M. *System Identification with Multi-step Least-Squares Methods.* PhD Thesis, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, 2018, 307 pp.

[9] Kailath T., Sayed A., Hassibi B. *Linear estimation*. Prentice Hall Information and System Science Series, Prentice Hall, 2000, 854 pp.

[10] Lemmerling P., Vanhamme L., Van Huffel S., De Moor B. *IQML-like algorithms for solving structured total least squares problem*, Signal Processing, vol. 81, 2001, p. 1935-1945.

[11] Ljung L. System identification: Theory for the user. Prentice Hall Information and System Sciences Series, 1999, 2nd Ed.

[12] Regalia P.A. *Adaptive IIR Filtering in Signal Processing and Control.* Marcel Dekker Inc., 1995, 678 pp.

[13] Stetter H. *Numerical polynomial algebra.* SIAM, 2004, 470 pp.

[14] Soderstrom T. *Errors-in-variables Methods in System Identification.* Springer, Communications and Control Engineering, 2018.

[15] Van Overschee P., De Moor B. *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. Kluwer Academic Publishers, 1996, 254 pp. (see also ftp://ftp.esat.kuleuven.be/pub/stadius/ida/reports/96-26a.pdf).

[16] Volkmer H. *Multiparameter Eigenvalue Problems and Expansion Theorems.* Lecture Notes in Mathematics, Springer-Verlag, 1988, 157 pp.

[17] Willems J. *From time series to linear system*, Automatica, *Part I. Finite dimensional linear time invariant systems*, Vol. 22, p. 561-580, 1986; *Part II. Exact modelling*, Vol. 22, p. 675-694, 1986; *Part III. Approximate modelling* Vol. 23, p. 87-115, 1987.