Globally Optimal Least-Squares ARMA Model Identification Is an Eigenvalue Problem

Christof Vermeersch and Bart De Moor, *Fellow*, *IEEE & SIAM*

Abstract—We show that globally optimal least-squares identification of autoregressive moving-average (ARMA) models is an eigenvalue problem. The first-order optimality conditions of this identification problem constitute a system of multivariate polynomial equations, in which most variables appear linearly. This system is basically a multiparameter eigenvalue problem (MEP), which we solve by iteratively building a so-called block Macaulay matrix, the null space of which is block multi-shift-invariant. The set of all stationary points of the optimization problem, i.e., the n-tuples of eigenvalues and eigenvectors of the MEP, follows from a standard eigenvalue problem (EP) related to a multidimensional realization problem in that null space. At least one of these n-tuples corresponds to the global minimum of the original least-squares objective function. Contrary to existing heuristic techniques, this approach yields the globally optimal parameters of the ARMA model. We provide a numerical example to illustrate the new identification method.

Index Terms—Linear systems, identification, optimization.

I. Introduction

EIGENVALUE problems prevail in nature and science. More particularly, eigenvalues form the cornerstone of systems and control: They characterize stability, controllability, and observability of linear time-invariant (LTI) dynamical systems [1], arise in the steady-state solutions to LQR control and Kalman filtering problems [2], solve model reduction problems like modal approximation [3], etc. In this letter, we explore this intimate connection further and show that globally optimal least-squares identification of autoregressive moving-average models is an eigenvalue problem.

Autoregressive moving-average (ARMA) models regress an observed output sequence on its own lagged values and on a linear combination of unobserved, latent input samples [4]. In the statistical literature, this sequence of latent inputs is often assumed to be a white Gaussian process [5]. Although our results could be interpreted in an appropriate maximum likelihood framework, we refrain ourselves from those a priori assumptions. ARMA models emerge in a wide variety of domains [4], e.g., in modeling industrial processes, financial time series, or smart utility grid applications (electricity, water, etc.). Moreover, their model structure is an important building block for more sophisticated models [6], e.g., autoregressive moving-average models with exogenous inputs (ARMAX) and autoregressive integrated moving-average models (ARIMA). Already in 1927, Yule [7] proposed a pure autoregressive (AR) process, which only considers a regression of the output sequence on itself. The pure moving-average (MA) model was introduced simultaneously by Yule [7] and Slutzky [8]. There exists some dissonance about who was the first to combine these two models, but Whittle [9] and Walker [10] have often been cited as the founding fathers of ARMA modeling. Its popularization, however, was clearly thanks to the famous book by Box and Jenkins [4], who have really propagated these models as a useful tool in time series analysis.

Although numerous nonlinear identification techniques for ARMA models already exist, e.g., autocorrelation, penalty function, and innovation regression methods (see Choi [5], Ljung [6], or Brockwell and Davis [11]), most of them rely on non-convex numerical optimization and do not guarantee to find the globally optimal model parameters. Stochastic subspace methods, on the other hand, provide a geometric approach by means of projections, which work very good in practice, but are not known to be optimal in any sense (see for example Van Overschee and De Moor [12]). Batselier et al. [13] have already approached globally optimal prediction error method identification (and thus also the identification of ARMA models) as an eigenvalue problem. However, they have used the classical Macaulay matrix, which does not exploit the available structure in the problem and scales terribly with the number of output samples.

In this letter, we tackle and resolve this hiatus and find the globally optimal least-squares ARMA model parameters using the new block Macaulay matrix approach. The first-order optimality conditions of this identification problem constitute a system of multivariate polynomial equations, in which most variables appear linearly. This system is basically a multiparameter eigenvalue problem (MEP), which we solve by iteratively building the so-called block Macaulay matrix, the null space of which is block multi-shift-invariant. The set of all stationary points of the optimization problem, i.e., the n-tuples of eigenvalues and eigenvectors of the MEP, follows from a standard eigenvalue problem (EP) related to a multidimensional realization problem in that null space. At least one of the n-tuples corresponds to the global minimum of the original least-squares objective function and thus yields the globally optimal parameters of the ARMA model.

Our main contribution is two-fold: We claim and show that globally optimal least-squares identification of ARMA models is essentially an MEP, and we provide a new solution method for this type of problems based on the block Macaulay matrix.

The remainder of this letter proceeds as follows: Section II rigorously formulates the identification problem. In Section III, we propose a globally optimal least-squares approach to find the parameters of ARMA models. A numerical example is given in Section IV. We draw several conclusions from this work and point at challenges for future research in Section V.

II. PROBLEM DEFINITION

A scalar autoregressive moving-average (ARMA) model combines a regression of the observed output variable $y_k \in \mathbb{R}$ on its own lagged values y_{k-i} with a linear combination of unobserved, latent inputs $e_{k-j} \in \mathbb{R}$ [4]:

$$\sum_{i=0}^{n_a} \alpha_i y_{k-i} = \sum_{j=0}^{n_c} \gamma_j e_{k-j}, \qquad (1)$$

where n_a and n_c are the orders of the autoregressive and moving-average part, respectively. The weighting factors $\alpha_i \in \mathbb{R}$, $i = 1, \ldots, n_a$, and $\gamma_j \in \mathbb{R}$, $j = 1, \ldots, n_c$, in the summations are the parameters of the ARMA model. To avoid indeterminacy and without loss of generality, we fix the leading parameters $\alpha_0 = \gamma_0 = 1$.

Given a data sequence of N observed output samples $y \in \mathbb{R}^N$ (not necessarily generated by an ARMA model), we want to find the parameters that satisfy the model structure of (1) and minimize the squared 2-norm of the unobserved, latent input vector $e \in \mathbb{R}^{N-n_a+n_c}$, on which we put no a priori unverifiable constraints (like for example whiteness or Gaussianity). For this problem, the model structure of (1) results in

$$T_a y = T_c e, \qquad (2)$$

where the two model matrices $T_a \in \mathbb{R}^{(N-n_a) \times N}$ and $T_c \in \mathbb{R}^{(N-n_a) \times (N-n_a+n_c)}$ are banded Toeplitz matrices of appropriate dimensions (the elements not shown are zero):

$$T_a = \begin{bmatrix} \alpha_{n_a} & \cdots & \alpha_1 & 1 & & & \\ & \alpha_{n_a} & \cdots & \alpha_1 & 1 & & \\ & & \ddots & & \ddots & \ddots & \\ & & & \alpha_{n_a} & \cdots & \alpha_1 & 1 \end{bmatrix}$$

$$T_c = \begin{bmatrix} \gamma_{n_c} & \cdots & \gamma_1 & 1 & & & \\ & \gamma_{n_c} & \cdots & \gamma_1 & 1 & & \\ & & \ddots & & \ddots & \ddots & \\ & & & \gamma_{n_c} & \cdots & \gamma_1 & 1 \end{bmatrix}.$$

This identification problem corresponds to a multivariate polynomial optimization problem in which we minimize the sum of squares of the latent inputs $\sigma^2 = \|e\|_2^2$, subject to the ARMA model structure of (2):

$$\min_{a,c,e} \|e\|_2^2$$
$$\text{subject to } T_a y = T_c e, \qquad (3)$$

where the unknown vectors $a \in \mathbb{R}^{n_a}$ and $c \in \mathbb{R}^{n_c}$ contain the parameters α_i and γ_j, respectively.

III. GLOBALLY OPTIMAL LEAST-SQUARES ARMA MODEL IDENTIFICATION

This section shows that globally optimal least-squares identification of ARMA models is an eigenvalue problem and provides an identification algorithm based on the new block Macaulay matrix. The first-order optimality conditions constitute a system of multivariate polynomial equations that defines the set of stationary points of the original least-squares objective function σ^2 (Subsection III-A). Subsection III-B shows that this system is basically a (nonlinear) multiparameter eigenvalue problem. Next, we solve this problem by iteratively building the so-called block Macaulay matrix (Subsection III-C) and exploiting the block multi-shift-invariant structure of its null space to set up a standard eigenvalue problem (Subsection III-D). At least one of the eigentuples corresponds to the globally optimal parameters of the ARMA model. Finally, Subsection III-E interprets this new algorithm in a system theoretic setting.

A. First-order optimality conditions

If the vectors a and c were known, (2) would be a set of underdetermined linear equations, the minimum norm solution of which is

$$e = T_c^\dagger T_a y,$$

where $T_c^\dagger$ is the pseudoinverse of the matrix T_c. This relationship between the unobserved, latent input vector e and the observed output vector y helps to remove the latent inputs from the least-squares objective function

$$\sigma^2 = \|e\|_2^2 = e^{\mathrm{T}} e = y^{\mathrm{T}} T_a^{\mathrm{T}} T_c^{\dagger \mathrm{T}} T_c^\dagger T_a y. \qquad (4)$$

Since the model matrix T_c is of full row rank, its pseudoinverse equals $T_c^\dagger = T_c^{\mathrm{T}} \left(T_c T_c^{\mathrm{T}} \right)^{-1}$. The objective function in (4) then reduces to

$$\sigma^2 = y^{\mathrm{T}} T_a^{\mathrm{T}} \left(T_c T_c^{\mathrm{T}} \right)^{-1} T_a y,$$

which has to be minimized over the parameters α_i and γ_j in a and c. To simplify the notation, we introduce the symmetric, positive definite, banded Toeplitz matrix $D_c = \left(T_c T_c^{\mathrm{T}} \right) \in \mathbb{R}^{(N-n_a) \times (N-n_a)}$, so that

$$\sigma^2 = y^{\mathrm{T}} T_a^{\mathrm{T}} D_c^{-1} T_a y.$$

The objective function σ^2 is clearly nonlinear in the parameters α_i and γ_j. Typically, in the literature, this type of problems is solved via numerical nonlinear optimization methods. However, these methods are heuristic and can get stuck in local optima. Therefore, we propose a new approach based on the null space of the so-called block Macaulay matrix, which starts from the first-order optimality conditions.

The first-order optimality conditions of the objective function σ^2, $\forall i = 1, \ldots, n_a$ and $\forall j = 1, \ldots, n_c$, are:

$$\frac{\partial \sigma^2}{\partial \alpha_i} = y^{\mathrm{T}} T_a^{\mathrm{T}} D_c^{-1} T_a^{\alpha_i} y + y^{\mathrm{T}} T_a^{\alpha_i \mathrm{T}} D_c^{-1} T_a y = 0$$
$$\frac{\partial \sigma^2}{\partial \gamma_j} = -y^{\mathrm{T}} T_a^{\mathrm{T}} D_c^{-1} D_c^{\gamma_j} D_c^{-1} T_a y = 0, \qquad (5)$$

with the matrices $T_a^{\alpha_i} = \frac{\partial T_a}{\partial \alpha_i}$ and $D_c^{\gamma_j} = \frac{\partial D_c}{\partial \gamma_j}$. By introducing an auxiliary vector $f = D_c^{-1} T_a y \in \mathbb{R}^{N-n_a}$,

we partially linearize the optimization problem. The vectors $f^{\alpha_i} = D_c^{-1} T_a^{\alpha_i} y \in \mathbb{R}^{N-n_a}$ and $f^{\gamma_j} = -D_c^{-1} D_c^{\gamma_j} f \in \mathbb{R}^{N-n_a}$ are the partial derivatives of the vector f, with respect to the unknown parameters α_i and γ_j, respectively. With these definitions, we rewrite (5) and obtain

$$\frac{\partial \sigma^2}{\partial \alpha_i} = y^{\mathrm{T}} T_a^{\mathrm{T}} f^{\alpha_i} + y^{\mathrm{T}} T_a^{\alpha_i \mathrm{T}} f = 0$$
$$\frac{\partial \sigma^2}{\partial \gamma_j} = y^{\mathrm{T}} T_a^{\mathrm{T}} f^{\gamma_j} = 0. \tag{6}$$

Finally, the first-order optimality conditions in (6), together with the definitions of the vectors f, f^{α_i}, and f^{γ_j}, constitute the system of multivariate polynomial equations that defines the set of stationary points of the original least-squares objective function σ^2:

$$\begin{cases} y^{\mathrm{T}} T_a^{\mathrm{T}} f^{\alpha_i} + y^{\mathrm{T}} T_a^{\alpha_i \mathrm{T}} f = 0 & \forall i = 1, \dots, n_a \\ y^{\mathrm{T}} T_a^{\mathrm{T}} f^{\gamma_j} = 0 & \forall j = 1, \dots, n_c \\ D_c f^{\alpha_i} - T_a^{\alpha_i} y = 0 & \forall i = 1, \dots, n_a \\ D_c^{\gamma_j} f + D_c f^{\gamma_j} = 0 & \forall j = 1, \dots, n_c \\ D_c f - T_a y = 0 \end{cases} \tag{7}$$

At least one of the common roots of this system corresponds to the global minimum of the original multivariate optimization problem in (3), i.e., to the globally optimal least-squares parameters of the ARMA model.

B. Multiparameter eigenvalue problems

Equation (7) consists of $(N - n_a)(n_a + n_c + 1) + n_a + n_c$ cubic multivariate polynomial equations in $(N - n_a)(n_a + n_c + 1) + n_a + n_c$ variables, of which $(N - n_a)(n_a + n_c + 1)$ variables appear linearly in the problem, a structure that becomes more apparent when we isolate these linear variables:

$$\underbrace{\begin{bmatrix} I \otimes g^{\mathrm{T}} & 0 & \{g^{\alpha_i \mathrm{T}}\}_i & 0 \\ 0 & I \otimes g^{\mathrm{T}} & 0 & 0 \\ I \otimes D_c & 0 & 0 & \{g^{\alpha_i}\}_i \\ 0 & I \otimes D_c & \{D_c^{\gamma_j}\}_j & 0 \\ 0 & 0 & D_c & g \end{bmatrix}}_{A(\alpha_i, \gamma_j)} \underbrace{\begin{bmatrix} \{f^{\alpha_i}\}_i \\ \{f^{\gamma_j}\}_j \\ f \\ -1 \end{bmatrix}}_{z} = 0,$$

with the vectors $g = T_a y \in \mathbb{R}^{N-n_a}$ and $g^{\alpha_i} = T_a^{\alpha_i} y \in \mathbb{R}^{N-n_a}$. The operator $\otimes$ represents the Kronecker product and the curly brackets $\{M_i\}_i$ indicate a vertical stack of matrices M_i over the index i, e.g., for $i = 1, 2$, $\{M_i\}_i = \begin{bmatrix} M_1^{\mathrm{T}} & M_2^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$.

The matrix A is only a function of the known output samples y_k and the unknown parameters α_i and γ_j. This system of multivariate polynomial equations is basically a (nonlinear) multiparameter eigenvalue problem (MEP), where the nonlinear variables (the parameters α_i and γ_j) constitute the $(n_a + n_c)$-tuples of eigenvalues and the linear variables (the vectors f, f^{α_i}, and f^{γ_j}) generate the eigenvectors z. In order to support this claim, we rewrite the system $Az = 0$ as

$$\left(A_1 - \sum_{\omega \neq 1} A_\omega \omega \right) z = 0, \tag{8}$$

with the matrices A_ω (e.g., A_1 or A_{α_1}) containing the coefficients of the monomial $\omega = \alpha_1^{k_1} \dots \alpha_{n_a}^{k_{n_a}} \gamma_1^{l_1} \dots \gamma_{n_c}^{l_{n_c}}$ with

degrees k_i and l_j in the matrix A. Consequently, it is easy to restate (8) in the classical form of an MEP (see for example Atkinson [14] or De Moor [15]), namely

$$A_1 z = \sum_{\omega \neq 1} A_\omega \omega z. \tag{9}$$

Indeed, this additive structure closely resembles that of a standard eigenvalue problem (EP), where $Az = \lambda z$ (with eigenvalues λ) is equal to $(A - I\lambda) z = 0$, which has a nontrivial solution, i.e., an eigenvector $z \neq 0$, if and only if the characteristic polynomial $\chi(A) = \det(A - I\lambda) = 0$.

To solve this type of problems, we introduce in the next subsection the so-called block Macaulay matrix, which iteratively linearizes the MEP [15].

C. Null space of the block Macaulay matrix

The next step in the identification procedure is to tackle the MEP of (9) and to find the stationary points of the original least-squares objective function σ^2 as the $(n_a + n_c)$-tuples of eigenvalues and eigenvectors of this MEP.

For didactic reasons, we start our exposition with a first-order ARMA(1,1) model, which has only two unknown parameters, i.e., α_1 and γ_1. Then, the system of (7) consists of $3N - 1$ cubic multivariate polynomial equations in $3N - 1$ variables, of which $3N - 3$ variables, i.e., f, f^{α_1}, and f^{γ_1}, appear linearly in the problem. We can rewrite this MEP as

$$\begin{bmatrix} y^{\mathrm{T}} T_a^{\mathrm{T}} & 0 & y^{\mathrm{T}} T_a^{\alpha_1 \mathrm{T}} & 0 \\ 0 & y^{\mathrm{T}} T_a^{\mathrm{T}} & 0 & 0 \\ D_c & 0 & 0 & T_a^{\alpha_1} y \\ 0 & D_c & D_c^{\gamma_1} & 0 \\ 0 & 0 & D_c & T_a y \end{bmatrix} \begin{bmatrix} f^{\alpha_1} \\ f^{\gamma_1} \\ f \\ -1 \end{bmatrix} = 0, \tag{10}$$

where the parameters constitute the 2-tuples of eigenvalues and the linear variables generate the eigenvectors z. To solve the MEP, we introduce the matrices $A_\omega \in \mathbb{R}^{(3N-1) \times (3N-2)}$, as in (8), which contain the coefficients of the monomials $\omega = \alpha_1^k \gamma_1^l$ with degrees k and l in the matrix A, and obtain

$$(A_1 + A_{\alpha_1} \alpha_1 + A_{\gamma_1} \gamma_1 + A_{\alpha_1^2} \alpha_1^2 + A_{\alpha_1 \gamma_1} \alpha_1 \gamma_1 \\ + A_{\gamma_1^2} \gamma_1^2) z = 0. \tag{11}$$

Next, we construct the so-called block Macaulay matrix, which extends the classical Macaulay matrix (see for example Dreesen et al. [16], [17]) and exploits the structure of MEPs. The initial block Macaulay matrix starts with (11) and has degree $d_0 = 3$ (because of the cubic polynomials). In a first iteration, we multiply (11) with shifts of first total degree (α_1 and γ_1). Subsequently, in a second iteration, we use shifts of second total degree (α_1^2, $\alpha_1 \gamma_1$, and γ_1^2). We continue these iterations with monomials of increasing total degrees until the quasi-Toeplitz block Macaulay matrix M reaches the desired degree $d \geq d^*$ (see below how we define d^*):

$$\underbrace{\begin{bmatrix} A_1 & A_{\alpha_1} & A_{\gamma_1} & A_{\alpha_1^2} & A_{\alpha_1 \gamma_1} & A_{\gamma_1^2} & \cdots \\ 0 & A_1 & 0 & A_{\alpha_1} & A_{\gamma_1} & 0 & \cdots \\ 0 & 0 & A_1 & 0 & A_{\alpha_1} & A_{\gamma_1} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} z \\ z\alpha_1 \\ z\gamma_1 \\ z\alpha_1^2 \\ z\alpha_1\gamma_1 \\ z\gamma_1^2 \\ \vdots \end{bmatrix}}_{K} = 0.$$

For simplicity, let us assume that all solutions are affine and simple (see below for solutions at infinity and with multiplicity larger than one). Then, the vectors k in the multivariate block Vandermonde basis K of the null space, one for every solution, span the (right) null space of the block Macaulay matrix M. It is clear that the null space of this matrix M has a special structure. We will exploit this structure in the next subsection and show how it yields the solutions of this quadratic MEP.

But first, we treat the general $n_a > 1$ and $n_c > 1$ case. The block Macaulay matrix for general-order ARMA(n_a, n_c) models extends the structure of the above-described matrix. Let us reorder again the system of (7) as in (10), but this time for more than two parameters. The vector z has now a more general structure, namely

$$\begin{bmatrix} f^{\alpha_1 \mathrm{T}} & \cdots & f^{\alpha_{n_a} \mathrm{T}} & f^{\gamma_1 \mathrm{T}} & \cdots & f^{\gamma_{n_c} \mathrm{T}} & f^{\mathrm{T}} & -1 \end{bmatrix}^{\mathrm{T}}.$$

We also introduce matrices A_ω, as in (8), and obtain

$$\left(A_1 + A_{\alpha_1} \alpha_1 + A_{\alpha_2} \alpha_2 + \cdots + A_{\alpha_1^2} \alpha_1^2 + \cdots \right) z = 0. \quad (12)$$

The parameters α_i and γ_j generate the $(n_a + n_c)$-tuples of eigenvalues and the vectors z are again the eigenvectors of this nonlinear MEP. The block Macaulay matrix starts initially with (12) and grows in every iteration with shifts of increasing total degrees. These iterations with monomials of increasing total degrees generate a special structure in the null space, which we will now exploit to solve the MEP.

D. Multidimensional realization theory in block multi-shift-invariant subspaces

In this subsection, we will use the special structure of the null space of the block Macaulay matrix, which we call block multi-shift-invariant, to solve the MEP and obtain the parameters of the ARMA model. This block multi-shift-invariant structure of the null space will lead to the formulation of an EP that yields the $(n_a + n_c)$-tuples of eigenvalues and eigenvectors of the MEP. This approach is similar to the eigenvalue problem formulation of Stetter [18] in multivariate polynomial system solving and the shift trick of Ho and Kalman [19] in systems and control.

Recall that, for simplicity, we consider an MEP with only m_a affine and simple solutions. We will cover the more general problem below. In the previous subsection, the null space of the block Macaulay matrix appeared to have a block multi-shift-invariant structure. This structure can easily be understood using the multivariate block Vandermonde basis K: If we select a block[1] of a multivariate block Vandermonde vector k and multiply (or shift) it by one of the eigenvalues, i.e., the parameters α_i and γ_j, we find another block of that same vector k. Hence, the null space is block multi-shift-invariant[2]. Note that this structure is a property of the null space as a vector space and not of the specific basis [17].

[1] We define a block as the rows corresponding to the eigenvector z or one of its shifts (e.g., $z\alpha_1^2$). A degree block, on the other hand, is the collection of all blocks of the same total degree d (e.g., $z\alpha_1^2$, $z\alpha_1\gamma_1$, and $z\gamma_1^2$).

[2] Contrary to the classical Macaulay matrix, every shift during the construction of the block Macaulay matrix adds a block of rows to the null space. The null space is consequently block multi-shift-invariant instead of multi-shift-invariant and exploits the available structure of the MEP.

Example 1: To clarify, one could apply this rationale to the simple first-order ARMA(1,1) model. Take a multivariate block Vandermonde vector k of degree $d = 2$, i.e.,

$$k = \begin{bmatrix} z \\ z\alpha_1 \\ z\gamma_1 \\ z\alpha_1^2 \\ z\alpha_1\gamma_1 \\ z\gamma_1^2 \end{bmatrix},$$

and multiply the rows of the first three blocks by α_1. The multiplied rows form again three blocks of the same vector:

$$\begin{bmatrix} z \\ z\alpha_1 \\ z\gamma_1 \end{bmatrix} \xrightarrow{\alpha_1} \begin{bmatrix} z\alpha_1 \\ z\alpha_1^2 \\ z\alpha_1\gamma_1 \end{bmatrix}.$$

This can alternatively be written, using row selection matrices S_1 and S_2, as $S_1 k \alpha_1 = S_2 k$, with (I is the identity matrix)

$$S_1 = \begin{bmatrix} I & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \end{bmatrix}$$

and

$$S_2 = \begin{bmatrix} 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \end{bmatrix}.$$

The multiplication does not have to be with a simple eigenvalue. Any polynomial $g(\alpha_i, \gamma_j)$ in the given eigenvalues results in a valid multiplication (or shift). This multiplicative relationship can be repeated for every column of the basis K, i.e., for all m_a affine solutions, which yields

$$S_1 K D_g = S_2 K, \quad (13)$$

where D_g is a diagonal matrix with as its elements the evaluations of the polynomial $g(\alpha_i, \gamma_j)$ in the different solutions. We recognize in (13) a generalized EP, with as the matrix of eigenvectors the identity matrix. Remark that in order to ensure that this generalized EP is not degenerate, the matrix $S_1 K$ needs to be of full column rank, which requires the selection matrix S_1 to include m_a linearly independent rows. Therefore, we have to increase the degree of the block Macaulay matrix at least until its nullity equals m_a (degree d^*). Since, in practice, the solutions are not known in advance, the multivariate block Vandermonde basis K cannot be used and we work with a numerical basis Z, obtained for example via the singular value decomposition. There exists a relationship between both bases, namely $K = ZT$, with $T \in \mathbb{R}^{m_a \times m_a}$ a non-singular matrix, which reduces (13) to

$$(S_1 Z) T D_g = (S_2 Z) T,$$

and rephrases the MEP as a standard EP:

$$T D_g T^{-1} = (S_1 Z)^\dagger (S_2 Z).$$

The matrix of eigenvectors T relates both bases, by means of $K = ZT$, and can be used to find the multivariate block Vandermonde basis K. From this basis K, we can determine the $(n_a + n_c)$-tuples of eigenvalues and eigenvectors of the MEP, and thus we obtain, via this EP, the stationary points of the original least-squares objective function σ^2.
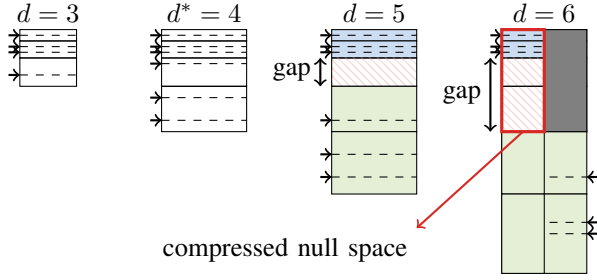
Fig. 1. The null space of the block Macaulay matrix grows as its degree increases. At a certain degree d^*, the nullity stabilizes at the Bézout number m_b. From that degree on, some linearly independent rows (corresponding to the affine solutions) stabilize, while the other linearly independent rows (corresponding to the solutions at infinity) move to higher degree blocks. A gap separates these rows. The influence of the solutions at infinity can be removed via a column compression. The affine procedure can then be applied straightforwardly on the compressed null space.

Solutions at infinity and with multiplicity larger than one: Due to algebraic relationships among the coefficients of the polynomials, solutions at infinity can emerge [17]. The total number of solutions in the projective space (both affine and at infinity) is given by the Bézout number $m_b = \prod_{i=1}^n d_i$, where d_i is the total degree of every polynomial (see for example Cox et al. [20]). After sufficient iterations, the nullity of the block Macaulay matrix stabilizes at the Bézout number (degree d^*). In that null space, we find not only linearly independent rows that remain at a certain position (corresponding to the affine solutions), but also linearly independent rows that move to higher degree blocks if the degree d of the block Macaulay matrix increases (corresponding to the solutions at infinity). This behavior actually helps to separate the affine solutions from the solutions at infinity and to remove the influence of the solutions at infinity via a column compression of the null space (see for example Fig. 1 and [17]). Afterwards, the above-described affine procedure can be applied straightforwardly.

In addition to solutions at infinity, multiple solutions may also occur. Dreesen [16] explains how the above-mentioned algorithm deals with these multiplicities.

E. System theoretic interpretations

Just as for the classical Macaulay matrix (see Dreesen et al. [16], [17]), a basis of the null space of the block Macaulay matrix (after stabilization, i.e., $d > d^*$) can also be interpreted as a multidimensional observability matrix Γ. In that setting, it is possible to view globally optimal least-squares ARMA model identification as an exact multidimensional realization problem in that null space. To clarify, we consider the block column echelon basis H. Dreesen et al. [17] showed that H can always be computed via a transformation of the numerical basis Z. The null space consists of three zones and this basis allows for a natural system theoretic interpretation of these zones:

$$H = \begin{bmatrix} H_R^{(1)} & 0 \\ H_R^{(2)} & 0 \\ \times & H_S \end{bmatrix} \begin{matrix} \leftarrow \text{regular zone} \\ \leftarrow \text{gap zone} \\ \leftarrow \text{singular zone} \end{matrix}$$

The three zones can be discovered by checking the rank of the basis row-wise from the top to the bottom:

1) Zone I (regular zone): In the first zone, the rank increases with at least one per degree block (i.e., all the rows corresponding to monomials of the same total degree), up to the number of affine roots m_a.
2) Zone II (gap zone): Then, the rank does not increase anymore (a generalization of Cayley–Hamilton for sets of commuting matrices). All the rows in this zone are linearly dependent on some rows of the first zone. There is a so-called gap of linearly dependent rows.
3) Zone III (singular zone): Finally, the rank starts to increase again, with at least one per degree block, until it reaches the nullity m_b of the block Macaulay matrix.

Example 2: Let us specify this block column echelon basis H of the null space for a simple first-order ARMA(1,1) model:

$$H = \begin{bmatrix} C_R & 0 \\ C_R A_{\alpha_1} & 0 \\ C_R A_{\gamma_1} & 0 \\ C_R A_{\alpha_1}^2 & 0 \\ C_R A_{\alpha_1} A_{\gamma_1} & 0 \\ C_R A_{\alpha_1}^2 & 0 \\ C_R A_{\alpha_1}^3 & 0 \\ \vdots & \vdots \\ C_R A_{\alpha_1}^n & 0 \\ \vdots & \vdots \\ \times & C_S E_{\alpha_1}^{m-1} \\ \vdots & \vdots \end{bmatrix}$$

The regular columns of this block column echelon basis H, i.e., the m_a left-most columns corresponding to the affine solutions, and the singular columns, i.e., the remaining $(m_b - m_a)$ right-most columns corresponding to solutions at infinity, determine two observability matrices Γ_R and Γ_S, which are generated by a multidimensional descriptor system, as described by Dreesen et al. [17]. Contrary to the null space of the classical Macaulay matrix, C_R and C_S are output matrices instead of vectors, which is a block extension of the observability matrix in [17]. In this multidimensional observability matrix, we find the multidimensional realization problems that yield us the EPs to solve the MEP. Indeed, if one selects two blocks of H, for example $S_{\alpha_1} H = C_R A_{\alpha_1}$ and $S_{\alpha_1,\gamma_1} H = C_R A_{\alpha_1} A_{\gamma_1}$, we recognize the multiplicative relationship, which is a property of the null space,

$$S_{\alpha_1} H A_{\gamma_1} = S_{\alpha_1,\gamma_1} H,$$

or

$$A_{\gamma_1} = (S_{\alpha_1} H)^\dagger (S_{\alpha_1,\gamma_1} H).$$

The eigenvalues of all these system matrices A_ω correspond to the different $(n_a + n_c)$-tuples of eigenvalues of the MEP.

IV. NUMERICAL EXAMPLE

In order to illustrate the new algorithm presented in this letter, this section provides a numerical proof-of-concept. We show that the block Macaulay matrix approach is able to identify the globally optimal least-squares parameters of an ARMA model.

Method	α_1	γ_1	$\|e\|_2^2$
Original parameters	0.5000	0.5000	0.0431
System id. toolbox	-0.8929	0.9474	0.0150
Block Macaulay approach	-0.7307	-0.6078	0.0053

Consider a first-order ARMA(1,1) model with parameters $\alpha_1 = 0.5$ and $\gamma_1 = 0.5$. This model generates an output sequence

$$y_k = e_k + \gamma_1 e_{k-1} - \alpha_1 y_{k-1}.$$

Let us consider a sequence of $N = 5$ output samples

$$y = \begin{bmatrix} 0.2000 \\ 0.0739 \\ 0.0935 \\ 0.0812 \\ 0.0838 \end{bmatrix},$$

which are generated by the ARMA model and a random latent input (without any a priori assumptions).

Then, the system of (7) consists of 14 polynomial equations in 14 variables, of which 12 variables appear linearly in the problem. A block Macaulay matrix of degree $d = 30$ suffices to find the gap in its null space. Since the first zone contains 48 linearly independent rows, the system of multivariate polynomial equations (or the MEP) has $m_a = 48$ affine solutions. From these linearly independent rows, we construct the EP $(S_1 Z)^\dagger (S_2 Z)$, which yields the 48 affine solutions, i.e., the different 2-tuples of the MEP. The affine solution with the smallest sum of squares of the latent inputs $\|e\|_2^2$ corresponds to the globally optimal least-squares ARMA model parameters for this given vector of output samples y. These parameters result in a smaller $\|e\|_2^2$ than the original parameters and the solution found by the armax function of the MATLAB system identification toolbox[3] (see Table I).

V. CONCLUSIONS AND FUTURE WORK

In this letter, we showed that globally optimal least-squares identification of ARMA models is essentially an eigenvalue problem and proposed a first usable algorithm to find these optimal parameters. This new approach translates the ARMA identification problem, via the first-order optimality conditions, into a system of multivariate polynomial equations, in which most variables appear linearly. This system is basically a (nonlinear) multiparameter eigenvalue problem (MEP) and we showed that we can solve this type of problems via an exact multidimensional realization problem in the block multi-shift-invariant null space of the so-called block Macaulay matrix. The n-tuples of eigenvalues and eigenvectors of this MEP correspond to the set of all stationary points of the optimization problem, of which at least one gives us the globally optimal

parameters of the ARMA model. We provided a proof-of-concept with a numerical example in which we identified the parameters of an ARMA(1,1) model using this new approach.

As the orders of the model and the number of observed output samples increase, the set of stationary points grows rapidly. Hence, solving the MEP and evaluating all solutions rapidly becomes impractical. Therefore, one of our current research efforts is to adapt the algorithm so that it only calculates the optimal n-tuple of eigenvalues. Furthermore, in future work, we will also report rigorously on the properties of this new block Macaulay matrix. Even though we did not tackle large practical problems, the mathematical claim that globally optimal least-squares identification of ARMA models is an MEP and the proposed solution approach for these MEPs are important contributions to the field of systems and control.

REFERENCES

[1] T. Kailath, *Linear Systems*, ser. Prentice Hall Information and System Sciences Series. Englewood Cliffs: Prentice Hall, 1980.

[2] T. Kailath, S. H. Ali, and H. Babak, *Linear Estimation*, ser. Prentice Hall Information and System Sciences Series. Upper Saddle River: Prentice Hall, 2000.

[3] A. C. Antoulas, *Approximation of Large-Scale Dynamical Systems*, ser. Advances in Design and Control. Philadelphia: Society for Industrial and Applied Mathematics, 2005.

[4] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., ser. Holden-Day Series in Time Series Analysis, E. Robinson, Ed. Oakland: Holden-Day, 1976.

[5] B. Choi, *ARMA Model Identification*, ser. Springer Series in Statistics, J. Gani and C. C. Heyde, Eds. New York: Springer-Verlag, 1992.

[6] L. Ljung, *System Identification: Theory for the User*, 2nd ed., ser. Prentice Hall Information and System Sciences Series, T. Kailath, Ed. Upper Saddle River: Prentice Hall, 1999.

[7] U. G. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions of the Royal Society of London*, vol. 226, no. 636-646, pp. 267–298, 1927.

[8] E. Slutzky, "The summation of random causes as the source of cyclic processes," *Econometrica*, vol. 5, no. 2, pp. 105–146, 1937.

[9] P. Whittle, "Hypothesis testing in time series analysis," Ph.D. dissertation, Uppsala University, Uppsala, 1951.

[10] A. M. Walker, "Note on a generalization of the large sample goodness of fit test for linear autoregressive schemes," *Journal of the Royal Statistical Society*, vol. 12, no. 1, pp. 102–107, 1950.

[11] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 1991.

[12] P. Van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Dordrecht: Kluwer Academic Publishers, 1996.

[13] K. Batselier, P. Dreesen, and B. De Moor, "Prediction error method identification is an eigenvalue problem," in *Proc. of the 16th IFAC Symposium on System Identification*, Brussels, 2012, pp. 221–226.

[14] F. V. Atkinson, *Multiparameter Eigenvalue Problems*, ser. Mathematics in Science and Engineering, R. Bellman, Ed. New York: Academic Press, 1972.

[15] B. De Moor, "Least-squares realization of LTI models is an eigenvalue problem," Internal Report 18-140, KU Leuven, Leuven, 2018. Accepted for presentation at and publication in the proceedings of ECC 2019.

[16] P. Dreesen, "Back to the roots: Polynomial system solving using linear algebra," Ph.D. dissertation, KU Leuven, Leuven, 2013.

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

[18] H. J. Stetter, *Numerical Polynomial Algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 2004.

[19] B. L. Ho and R. E. Kalman, "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik*, vol. 14, no. 1-12, pp. 545–548, 1966.

[20] D. A. Cox, J. Little, and D. O'Shea, *Using Algebraic Geometry*, 2nd ed., ser. Graduate Texts in Mathematics. New York: Springer, 2005, vol. 185.

[3]The armax function minimizes the prediction errors in order to find the model parameters. It uses a nonlinear optimization algorithm as described in the book of Ljung [6, Chapter 7].