Least Squares Realization of Complex-Valued Multiple-Output Autonomous LTI Systems via the Riemannian Singular Value Decomposition*

Sem Viroux, Sibren Lagauw, and Bart De Moor, *Life Fellow, IEEE*

***Abstract*—We present a generalization of least squares realization of multiple-output linear time-invariant models to the complex domain, thereby enabling broader applicability in areas such as signal processing, control, and system identification. We discuss the challenges inherent to optimization over complex variables and formulate the realization problem as a structured total least squares problem. A key contribution is the derivation of a complex-valued nonlinear generalized singular value decomposition. We adapt an existing heuristic algorithm that works over the real domain to solve the complex nonlinear generalized singular value decomposition. We validate our framework through illustrative examples and demonstrate its use in a nuclear magnetic resonance modeling problem and a direction of arrival estimation problem.**

***Index Terms*—Identification, optimization, numerical algorithms.**

I. Introduction

We consider complex-valued multiple-output autonomous linear time-invariant (LTI) causal lumped parameter dynamical models in discrete-time, with a pre-specified model order n. The output data $\widehat{\boldsymbol{y}}_i = [\hat{y}_{i,0} \ \cdots \ \hat{y}_{i,N-1}]^\mathsf{T} \in \mathbb{C}^N$ (with $N > n$) satisfy a difference equation of the form:

$$\hat{y}_{i,k+n} + a_1\hat{y}_{i,k+n-1} + \cdots + a_n\hat{y}_{i,k} = \boldsymbol{a}(z)\hat{y}_{i,k} = 0, \qquad (1)$$

where $\widehat{\boldsymbol{y}}_i$ is the ith output signals and $\boldsymbol{a}(z)$ is a degree n polynomial in the forward-shift operator z (i.e., $zy_{i,k} = y_{i,k+1}$), the coefficients of which correspond to the complex model parameters $\boldsymbol{a}_{\text{tr}} = [a_n \ a_{n-1} \ \cdots \ a_1]^\mathsf{T} \in \mathbb{C}^n$. The data $\widehat{\boldsymbol{y}}_i$ are said to be *model-compliant* since they satisfy the difference equation (1) exactly for all $k = 0, \ldots, N-n-1$.

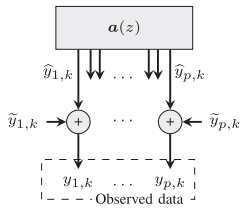


Fig. 1. Schematic overview of the autonomous LTI multiple-output model. The given output data $\boldsymbol{y}_i \in \mathbb{C}^N$ are modified using the so-called misfits $\widetilde{\boldsymbol{y}}_i \in \mathbb{C}^N$, so that the modified data $\widehat{\boldsymbol{y}}_i \in \mathbb{C}^N$ are model-compliant. The coefficients $\boldsymbol{a}_{\text{tr}} \in \mathbb{C}^n$ of the nth degree polynomial $\boldsymbol{a}(z)$ in the forward-shift operator z, are the unknown model parameters.

In the least squares (LS) realization problem, one is given p sequences of data $\boldsymbol{y}_i \in \mathbb{C}^N$ that do not comply with (1). The objective is to find the parameters $\boldsymbol{a}$ that minimize the 2-norm of the *misfits* $\widetilde{\boldsymbol{y}}_i = \boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i$, which is the modification needed to make the data $\boldsymbol{y}$ model-compliant (see Fig. 1):

$$\begin{aligned} \underset{\substack{\widehat{\boldsymbol{y}}_1,\ldots,\widehat{\boldsymbol{y}}_p \in \mathbb{C}^N \\ \boldsymbol{a}_{\text{tr}} \in \mathbb{C}^n}}{\text{minimize}} \quad & \frac{1}{2}\sum_{i=1}^{p} \|\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i\|_2^2 = \frac{1}{2}\sum_{i=1}^{p} \|\widetilde{\boldsymbol{y}}_i\|_2^2, \\ \text{subject to} \quad & \boldsymbol{a}(z)\hat{y}_{i,k} = 0, \ k = 0, \ldots, N-n-1. \end{aligned} \qquad (2a)$$

From a systems-theoretic perspective, the difference relation (1) is equivalent to the characteristic (or minimal) polynomial of a linear system, and thus relates subsequent samples of model-compliant output data from an autonomous LTI system or an LTI impulse response. The literature about estimating the coefficients of (1) from data–in a real-valued setting–is vast, see e.g., [1], [2] and references therein. The problem traces back to the work of Prony [3] and has since appeared in numerous formulations across signal processing, control, and system identification. The estimation problem (2) can be interpreted as a structured [4], [5] and/or constrained [6] total least squares (TLS) problem. This connection to low-rank matrix approximation was already implicit in the early system identification literature [7]. When a-priori statistical assumptions are imposed on the misfits, (2) naturally leads to a maximum likelihood setting, see e.g., dynamic errors in variables modeling [8] and the iterative quadratic maximum likelihood methods [9], [10].

Received 16 June 2025; revised 5 August 2025; accepted 17 August 2025. Date of publication 29 August 2025; date of current version 15 September 2025. This work was supported in part by the KU Leuven: Research Fund and several LRD bilateral industrial projects, under Project iBOF/23/064; in part by the Flemish Government agencies: EWI (Flanders AI Research Program), FWO; and in part by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme ERC Adv., under Grant 885682. The work of Sibren Lagauw was supported by the FWO Fundamental Research Ph.D. Fellowship under Grant 11K5625N. Recommended by Senior Editor T. Oomen. *(Sem Viroux and Sibren Lagauw contributed equally to this work.) (Corresponding author: Sem Viroux.)*

The authors are with the Center for Dynamical Systems, Signal Processing, and Data Analytics (STADIUS) and the Department of Electrical Engineering (ESAT), KU Leuven, 3001 Leuven, Belgium (e-mail: sem.viroux@kuleuven.be; sibren.lagauw@kuleuven.be; bart.demoor@kuleuven.be).

Digital Object Identifier 10.1109/LCSYS.2025.3604015

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

Systems theory and identification is, however, not inherently limited to a real-valued setting, and working with *complex-valued signals and model parameters* in (2) allows for a broader class of applications. Indeed, complex-valued signals arise naturally in numerous domains, including digital communication applications [11], nuclear magnetic resonance (NMR) modeling [12], rotor bearing systems identification [13], and direction of arrival (DOA) estimation in array processing [14]. Recently, complex-valued representations have also been explored in the neural network community [15]. For a rigorous treatment of signal processing of complex-valued data, we refer the reader to [16].

In this work, we extend the structured TLS framework previously developed for real-valued problems [4], [5] to the complex domain, resulting in a novel methodology for computing locally optimal solutions to the complex-valued LS realization problem (2). This constitutes our *main contribution*. In Section II, we address the key intricacies associated with optimization over complex decision variables. Characterizing first-order optimality leads to a nonlinear generalized singular value decomposition (coined Riemannian SVD) in the complex field, which is presented in Section III. Section IV introduces an extension of the variable metric inverse iteration algorithm to the complex domain. In Section V, we work out two numerical examples, drawn from NMR modeling and DOA estimation, to validate the proposed methodology and illustrate its performance. We conclude with remarks and future work in Section VI.

Notation: We denote scalars, vectors and matrices by lowercase, boldface lowercase and boldface uppercase letters, respectively. We use j to denote the imaginary unit $\sqrt{-1}$. The complex conjugate, transpose, and Hermitian transpose of a complex vector $\boldsymbol{x} = \boldsymbol{x}^{\mathrm{Re}} + \mathrm{j}\boldsymbol{x}^{\mathrm{Im}} \in \mathbb{C}^d$ are indicated by $\overline{\boldsymbol{x}}$, $\boldsymbol{x}^{\mathsf{T}}$, and $\boldsymbol{x}^{\mathsf{H}}$, respectively. Subscript i and k of a time series denote the output and time indices, respectively, e.g., $y_{i,k}$.

II. Optimization Over Complex Decision Variables

In this section we give a brief overview of the technicalities involved with equality-constrained optimization over complex decision variables. We first note that in the context of minimization over complex decision variables, the objective function must be real ($f : \mathbb{C}^d \to \mathbb{R} : \boldsymbol{x} \mapsto f(\boldsymbol{x}), d \in \mathbb{Z}_+$). Indeed, for the optimization problem to be well-defined, $f(\boldsymbol{x})$ must be such that there exists an ordering on its function values[1]: one has to be able to conclude whether a solution $\boldsymbol{x}_1$ performs *better* than another solution $\boldsymbol{x}_2$ by means of the objective function value, e.g., $f(\boldsymbol{x}_1) < f(\boldsymbol{x}_2)$.

The first-order necessary conditions for optimality (FONC) of an optimization problem are obtained via the first-order derivatives of the corresponding Lagrangian. The *Cauchy–Riemann equations* serve as a necessary and sufficient condition for a function to be *analytic*, which means that its derivative is well defined for all points in the complex domain. Since real-valued objective functions, except for $f(\boldsymbol{x}) = c \in \mathbb{R}$, do not satisfy the Cauchy–Riemann equations, they are not complex differentiable.

[1]Even though it would be possible to impose an ordering on complex numbers which allows for a convenient comparison between two complex numbers, concepts such as the gradient and higher order derivatives would lose their meaning. The problem is that, although the complex numbers can be ordered in various ways, they cannot be made into an ordered field with the usual addition and multiplication operators [17].

Thus, we need a manner of redefining differentiation of a real-valued function $f(\boldsymbol{x})$ over complex variables $\boldsymbol{x} \in \mathbb{C}^d$. The most straightforward way of doing this, is by decoupling the derivatives w.r.t. the real and the imaginary parts, such that the real-valued objective function $f(\boldsymbol{x})$ can be interpreted as a function of two independent real variables $f(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}})$. This approach makes the problem completely real-valued, allowing for the use of real differential calculus to formulate the FONC for optimality.

The aforementioned reformulation of the optimization problem over complex variables into a real-valued problem is, however, often impractical, since it tends to obscure or destroy the structure of the original complex-variable problem. An alternative approach involves the use of CR-calculus, also referred to as *Wirtinger derivatives* [17], which provides an equivalent but more elegant formulation of the FONC. This approach is based on the observation that a function $f(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}})$ that is analytic over the vector space $\mathbb{R}^{d\times 2}$, is also analytic over the vector space $\{(\boldsymbol{x}, \overline{\boldsymbol{x}})\} \subset \mathbb{C}^{d\times 2}$.

For such a *polyanalytic* function $f(\boldsymbol{x}, \overline{\boldsymbol{x}})$ the Wirtinger derivatives towards $\boldsymbol{x}$ and $\overline{\boldsymbol{x}}$ are defined via the chain rule:

$$\frac{\partial f(\boldsymbol{x}, \overline{\boldsymbol{x}})}{\partial \boldsymbol{x}} = \frac{\partial f\left(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}}\right)}{\partial \boldsymbol{x}^{\mathrm{Re}}} \frac{\partial \boldsymbol{x}^{\mathrm{Re}}}{\partial \boldsymbol{x}} + \frac{\partial f\left(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}}\right)}{\partial \boldsymbol{x}^{\mathrm{Im}}} \frac{\partial \boldsymbol{x}^{\mathrm{Im}}}{\partial \boldsymbol{x}},$$
$$\frac{\partial f(\boldsymbol{x}, \overline{\boldsymbol{x}})}{\partial \overline{\boldsymbol{x}}} = \frac{\partial f\left(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}}\right)}{\partial \boldsymbol{x}^{\mathrm{Re}}} \frac{\partial \boldsymbol{x}^{\mathrm{Re}}}{\partial \overline{\boldsymbol{x}}} + \frac{\partial f\left(\boldsymbol{x}^{\mathrm{Re}}, \boldsymbol{x}^{\mathrm{Im}}\right)}{\partial \boldsymbol{x}^{\mathrm{Im}}} \frac{\partial \boldsymbol{x}^{\mathrm{Im}}}{\partial \overline{\boldsymbol{x}}},$$

which in practice boils down to differentiation of $f(\boldsymbol{x}, \overline{\boldsymbol{x}})$ with respect to $\boldsymbol{x}$ (resp. $\overline{\boldsymbol{x}}$), while considering the variable $\overline{\boldsymbol{x}}$ (resp. $\boldsymbol{x}$) as a constant, i.e., $\boldsymbol{x}$ and $\overline{\boldsymbol{x}}$ are treated as independent variables for the purpose of differentiation.

Though, one must be careful when working with constrained optimization problems over complex variables [16], [17]. To construct the Lagrangian when using Wirtinger derivatives, the original constraint as well as its complex conjugate must be incorporated into the Lagrangian using a single set of complex-conjugated Lagrange multipliers.

III. Complex Riemannian SVD

A. Model-Compliance

Writing out (1) for all $k = 0, \ldots, N-n-1$, illustrates that constraint (2a) of the realization problem can be expressed as

$$\begin{bmatrix} \widehat{\boldsymbol{Y}}_1 \\ \widehat{\boldsymbol{Y}}_2 \\ \vdots \\ \widehat{\boldsymbol{Y}}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{\mathrm{tr}} \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{a}} & & \\ & \boldsymbol{T}_{\boldsymbol{a}} & \\ & & \ddots \\ & & \boldsymbol{T}_{\boldsymbol{a}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{y}}_1 \\ \widehat{\boldsymbol{y}}_2 \\ \vdots \\ \widehat{\boldsymbol{y}}_p \end{bmatrix} = \boldsymbol{0}, \tag{3}$$

where $\widehat{\boldsymbol{Y}}_i \in \mathbb{C}^{(N-n)\times(n+1)}$ is the Hankel matrix constructed from the elements of $\widehat{\boldsymbol{y}}_i$, and $\boldsymbol{T}_{\boldsymbol{a}} \in \mathbb{C}^{(N-n)\times N}$ is a banded-Toeplitz matrix, where the first row of $\boldsymbol{T}_{\boldsymbol{a}}$ is given by the model parameters appended with $N-n-1$ zeros: $[\boldsymbol{a}_{\mathrm{tr}}^{\mathsf{T}}\ 1\ 0\ \cdots\ 0]$, and all elements below the diagonal are zero. The subscript $\boldsymbol{a}$ denotes the vector $\boldsymbol{a} = [\boldsymbol{a}_{\mathrm{tr}}^{\mathsf{T}}\ 1]^{\mathsf{T}}$ generating the rows of $\boldsymbol{T}_{\boldsymbol{a}}$. Notice the equality between the Hankel-vector products and the (banded) Toeplitz-vector products in (3). This reformulation follows from the Hankel/Toeplitz structure of the involved matrices and will be used extensively throughout this manuscript.

Relation (3) illustrates that the model parameters $\boldsymbol{a}$ associated to a known sequence of model-compliant output data

can be retrieved via the null space of the model-compliant data block-Hankel matrix: $\widehat{\boldsymbol{Y}} = [\widehat{\boldsymbol{Y}}_1^{\mathsf{T}} \ \cdots \ \widehat{\boldsymbol{Y}}_p^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{p(N-n)\times(n+1)}$. Conversely, when the model parameters $\boldsymbol{a}$ are known, the *behavior* of the model[2] corresponds to the (pn)-dimensional null space of the block-diagonal matrix constructed from $\boldsymbol{T}_{\boldsymbol{a}}$.

The given data $\boldsymbol{y}$, however, are not compliant with the model class under consideration, such that the given data block-Hankel matrix $\boldsymbol{Y} = [\boldsymbol{Y}_1^{\mathsf{T}} \ \cdots \ \boldsymbol{Y}_p^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{p(N-n)\times(n+1)}$, where each $\boldsymbol{Y}_i$ is a Hankel matrix, is of full column rank[3].

The identification problem (2) corresponds to a structured low-rank approximation problem: for a given data block-Hankel matrix $\boldsymbol{Y}$ that is of full column-rank, i.e., containing data that is not compliant with the model class under consideration, one wants to identify the best approximating, in a (weighted) least squares sense, low-rank block-Hankel matrix $\widehat{\boldsymbol{Y}}$. This rank-deficiency $\operatorname{rank}(\widehat{\boldsymbol{Y}}) \leq n$ is enforced by the constraint (2a). If $\operatorname{rank}(\widehat{\boldsymbol{Y}}) = n$, the model parameters corresponding to the model-compliant data $\widehat{\boldsymbol{y}}$ are unique.

B. Data Optimality

We approach the realization problem (2) by first rewriting the constraints (2a) in terms of the Hankel matrices, as in (3), and constructing the Lagrangian, for which we rely on the concepts presented in Section II.

$$\mathcal{L}(\boldsymbol{a}_{\text{tr}}, \widehat{\boldsymbol{y}}, \boldsymbol{\ell}) = \sum_{i=1}^{p} \Big[(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i)^{\mathsf{H}} (\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i) + \boldsymbol{\ell}_i^{\mathsf{H}} \boldsymbol{T}_{\boldsymbol{a}} \widehat{\boldsymbol{y}}_i + \boldsymbol{\ell}_i^{\mathsf{T}} \overline{\boldsymbol{T}_{\boldsymbol{a}} \widehat{\boldsymbol{y}}_i} \Big],$$

where the vector $\boldsymbol{\ell} = [\boldsymbol{\ell}_1^{\mathsf{T}} \ \cdots \ \boldsymbol{\ell}_p^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{p(N-n)}$ contains the Lagrange multipliers. The system of equations that characterizes the first-order optimal solutions follows from equating the partial derivatives of the Lagrangian to zero:

$$\partial \mathcal{L} / \partial \overline{\boldsymbol{\ell}_i} = \boldsymbol{T}_{\boldsymbol{a}} \widehat{\boldsymbol{y}}_i = \boldsymbol{0}, \quad \forall i \in \{1, \ldots, p\} \tag{4}$$

$$\partial \mathcal{L} / \partial \overline{\widehat{\boldsymbol{y}}_i} = \widehat{\boldsymbol{y}}_i - \boldsymbol{y}_i + \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \boldsymbol{\ell}_i = \boldsymbol{0}, \quad \forall i \in \{1, \ldots, p\} \tag{5}$$

$$\partial \mathcal{L} / \partial \overline{\boldsymbol{a}_{\text{tr}}} = \sum_{i=1}^{p} \widehat{\boldsymbol{Y}}_i^{\mathsf{H}} \boldsymbol{\ell}_i = \boldsymbol{0}. \tag{6}$$

Since $\mathcal{L}$ is purely real, the partial derivatives w.r.t. $\boldsymbol{\ell}_i, \widehat{\boldsymbol{y}}_i$, and $\boldsymbol{a}_{\text{tr}}$ correspond to the complex conjugates of (4), (5), and (6), respectively.

Equating the partial derivatives of the Lagrangian w.r.t. the model-compliant data to zero (cfr. (5)) illustrates that for each output i, the misfits $\widetilde{\boldsymbol{y}}_i$ lie in the row space of the complex conjugate of the Toeplitz matrix $\boldsymbol{T}_{\boldsymbol{a}}$

$$-(\boldsymbol{y}_i - \widehat{\boldsymbol{y}}_i) + \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \boldsymbol{\ell}_i = \boldsymbol{0} \iff \widetilde{\boldsymbol{y}}_i = \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \boldsymbol{\ell}_i. \tag{7}$$

Combining this with the constraints (3) and $\boldsymbol{y}_i = \widehat{\boldsymbol{y}}_i + \widetilde{\boldsymbol{y}}_i$ gives

$$\boldsymbol{T}_{\boldsymbol{a}} \widehat{\boldsymbol{y}}_i = \boldsymbol{T}_{\boldsymbol{a}} (\boldsymbol{y}_i - \widetilde{\boldsymbol{y}}_i) = \boldsymbol{0} \iff \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{y}_i = \boldsymbol{T}_{\boldsymbol{a}} \widetilde{\boldsymbol{y}}_i \iff \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{y}_i = \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \boldsymbol{\ell}_i.$$

[2]This is the set of all model-compliant data sequences.

[3]Remark that the matrix of given data $\boldsymbol{Y}$ has a non-trivial null space by construction when $p(N-n) < n+1$, such that $N \geq \lceil n + (n+1)/p \rceil$ is a necessary condition for the identification problem to be well-defined.

Writing this out for all outputs $i = 1, \ldots, p$ gives

$$\begin{bmatrix} \boldsymbol{T}_{\boldsymbol{a}} & & & \\ & \boldsymbol{T}_{\boldsymbol{a}} & & \\ & & \ddots & \\ & & & \boldsymbol{T}_{\boldsymbol{a}} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{y}}_1 \\ \widehat{\boldsymbol{y}}_2 \\ \vdots \\ \widehat{\boldsymbol{y}}_p \end{bmatrix} = \boldsymbol{D}_{\boldsymbol{a}} \begin{bmatrix} \boldsymbol{\ell}_1 \\ \vdots \\ \boldsymbol{\ell}_p \end{bmatrix}, \tag{8}$$

where the square matrix $\boldsymbol{D}_{\boldsymbol{a}} \in \mathbb{C}^{p(N-n)\times p(N-n)}$ corresponds to a block-diagonal matrix of p blocks:

$$\boldsymbol{D}_{\boldsymbol{a}} = \operatorname{blockdiag}\left(\boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}}, \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}}, \ldots, \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \right).$$

Using (3) we can rewrite (8) as,

$$\begin{bmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_p \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_{\text{tr}} \\ 1 \end{bmatrix} = \boldsymbol{D}_{\boldsymbol{a}} \boldsymbol{\ell}, \tag{9}$$

where $\boldsymbol{\ell} = [\boldsymbol{\ell}_1^{\mathsf{T}}, \ldots, \boldsymbol{\ell}_p^{\mathsf{T}}]^{\mathsf{T}}$. It follows from (8) that the optimal Lagrange multipliers $\boldsymbol{\ell}_i$ correspond to

$$\boldsymbol{\ell}_i = \left(\boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \right)^{-1} \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{y}_i$$

such that the optimal misfits $\widetilde{\boldsymbol{y}}_i$ are obtained via the orthogonal projection of the given data $\boldsymbol{y}_i$ onto $\operatorname{row}(\boldsymbol{T}_{\boldsymbol{a}})$:

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \left(\boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \right)^{-1} \boldsymbol{T}_{\boldsymbol{a}} \boldsymbol{y}_i.$$

C. Model Optimality

Optimality of the model parameters $\boldsymbol{a}$ is obtained when we equate the partial derivatives of the Lagrangian w.r.t. the model-parameters (6) to zero

$$\sum_{i=1}^{p} \widehat{\boldsymbol{Y}}_i^{\mathsf{H}} \boldsymbol{\ell}_i = \sum_{i=1}^{p} \left(\boldsymbol{Y}_i^{\mathsf{H}} - \widetilde{\boldsymbol{Y}}_i^{\mathsf{H}} \right) \boldsymbol{\ell}_i = \boldsymbol{0}$$

$$\iff \begin{bmatrix} \boldsymbol{Y}_1^{\mathsf{H}} & \cdots & \boldsymbol{Y}_2^{\mathsf{H}} \end{bmatrix} \begin{bmatrix} \boldsymbol{\ell}_1 \\ \vdots \\ \boldsymbol{\ell}_p \end{bmatrix} = \sum_{i=1}^{p} \boldsymbol{T}_{\boldsymbol{\ell}_i} \overline{\widetilde{\boldsymbol{y}}_i}.$$

Use now that $\overline{\widetilde{\boldsymbol{y}}_i} = \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{T}} \overline{\boldsymbol{\ell}_i}$ which follows from the complex conjugate of (7) and $\boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{T}} \overline{\boldsymbol{\ell}_i} = \boldsymbol{T}_{\boldsymbol{\ell}_i}^{\mathsf{H}} \boldsymbol{a}$, such that

$$\begin{bmatrix} \boldsymbol{Y}_1^{\mathsf{H}} & \cdots & \boldsymbol{Y}_2^{\mathsf{H}} \end{bmatrix} \boldsymbol{\ell} = \boldsymbol{D}_{\boldsymbol{\ell}} \boldsymbol{a}, \tag{10}$$

where the square matrix $\boldsymbol{D}_{\boldsymbol{\ell}} \in \mathbb{C}^{(n+1)\times(n+1)}$ corresponds to

$$\boldsymbol{D}_{\boldsymbol{\ell}} = \sum_{i=1}^{p} \boldsymbol{T}_{\boldsymbol{\ell}_i} \boldsymbol{T}_{\boldsymbol{\ell}_i}^{\mathsf{H}}.$$

D. Complex Riemannian SVD

Combining (9) with (10) leads to the following system:

$$\boldsymbol{Y}\boldsymbol{a} = \boldsymbol{D}_{\boldsymbol{a}} \boldsymbol{\ell}, \quad \boldsymbol{Y}^{\mathsf{H}} \boldsymbol{\ell} = \boldsymbol{D}_{\boldsymbol{\ell}} \boldsymbol{a}.$$

Substituting $\boldsymbol{\ell}$ via $\boldsymbol{\ell} = \tau \boldsymbol{u}$, with $\tau \in \mathbb{R}$ such that $\boldsymbol{u}^{\mathsf{H}} \boldsymbol{D}_{\boldsymbol{a}} \boldsymbol{u} = 1$ and $\boldsymbol{D}_{\boldsymbol{\ell}} = \tau^2 \boldsymbol{D}_{\boldsymbol{u}}$, we obtain the set of equations:

$$\boldsymbol{Y}\boldsymbol{a} = \boldsymbol{D}_{\boldsymbol{a}} \boldsymbol{u} \tau, \boldsymbol{a} = \begin{bmatrix} \boldsymbol{a}_{\text{tr}} \\ 1 \end{bmatrix}, \quad \boldsymbol{Y}^{\mathsf{H}} \boldsymbol{\ell} = \boldsymbol{D}_{\boldsymbol{u}} \boldsymbol{a} \tau, \boldsymbol{u}^{\mathsf{H}} \boldsymbol{D}_{\boldsymbol{a}} \boldsymbol{u} = 1. \tag{11}$$

This is a nonlinear generalized SVD with $\boldsymbol{D_a}$ and $\boldsymbol{D_u}$ positive definite weighting matrices, quadratic in $\boldsymbol{a}$ and $\boldsymbol{u}$, respectively. On the basis of our analysis, we conclude that the original real Riemannian SVD (RiSVD) generalizes to the complex plane by replacing the transposes with Hermitian conjugates. The value of τ^2 is equal to the objective function, i.e., the squared 2-norm of the misfits, just as in [4].

In other words, the identification problem (2) can be solved in a locally optimal way by computing the triplet $(\boldsymbol{a}, \tau, \boldsymbol{u}) \in \mathbb{C}^{n+1} \times \mathbb{R} \times \mathbb{C}^{p(N-n)}$ that solves the RiSVD (11) with smallest possible τ^2. The corresponding model-compliant data $\widehat{\boldsymbol{y}}$ can be obtained from the triplet $(\boldsymbol{a}, \tau, \boldsymbol{u})$ via

$$\begin{bmatrix} \widehat{\boldsymbol{y}}_1 \\ \vdots \\ \widehat{\boldsymbol{y}}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_p \end{bmatrix} - \tau \begin{bmatrix} \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} & & \\ & \ddots & \\ & & \boldsymbol{T}_{\boldsymbol{a}}^{\mathsf{H}} \end{bmatrix} \boldsymbol{u}. \tag{12}$$

IV. An Inverse-Iteration Method

To solve for a minimizing triplet $(\boldsymbol{a}, \tau, \boldsymbol{u})$ of the real-valued RiSVD, the standard variable metric inverse iteration algorithm can be used [4]. With a few minor adjustments, this algorithm generalizes to the complex-valued RiSVD (11): All transposes in the original algorithm should be changed to Hermitian transposes; and to ensure uniqueness of equivalent obtained triplets $\boldsymbol{a}^{(k)}$ and $\boldsymbol{u}^{(k)}$ are projected such that $a_0^{(k)} = 1$ (line 13 and 14). Algorithm 1 shows the modified algorithm. Note that all variables in Algorithm 1 are complex.

Like alternating least squares (ALS) methods, this algorithm iteratively solves the two equations in the first column of (11) by linearizing them freezing $\boldsymbol{D_u}$ and $\boldsymbol{D_a}$ for the current values of $\boldsymbol{u}$ and $\boldsymbol{a}$. In each subsequent iteration, the metrics $\boldsymbol{D_u}$ and $\boldsymbol{D_a}$ are updated with the new values of the intermediate solutions $\boldsymbol{u}^{(k)}$ and $\boldsymbol{a}^{(k)}$ and the same (linearized) equations are solved again. The algorithm splits $\boldsymbol{u}$ into two orthogonal components $\boldsymbol{Q}_1\boldsymbol{z}$ and $\boldsymbol{Q}_2\boldsymbol{w}$. The output of the algorithm depends on the initial guesses $\boldsymbol{u}^{(k)}$ and $\boldsymbol{a}^{(k)}$, which makes this a heuristic approach. Convergence is tested by measuring the difference between subsequent values $\tau^{(k)}$.

As for computational complexity, the main cost of the algorithm is the matrix inversion[4] of the symmetric positive definite $[p(N-n)-(n+1)] \times [p(N-n)-(n+1)]$ matrix $\boldsymbol{Q}_2^{\mathsf{H}}\boldsymbol{D_a}\boldsymbol{Q}_2$, which is computed in each iteration (line 6). A single QR decomposition of $\boldsymbol{Y}$ is required at initialization, with negligible cost compared to the iterative steps.

V. Numerical Experiments

A. Nuclear Magnetic Resonance Modeling

Complex signals are prevalent in nuclear magnetic resonance (NMR) modeling, where they allow storing phase information obtained via quadrature detection. We closely follow an experiment described in [12], where a signal produced by an nth-order autonomous single-output system is modeled as:

$$y_{\text{exact},k} = \sum_{m=1}^{n} A_m e^{\mathrm{j}(\varphi_0+\varphi_m)} e^{(-d_m+\mathrm{j}2\pi f_m)t_k} \tag{13}$$

[4]While symmetry and positive definiteness can be exploited, such implementation considerations fall outside the scope of this letter.

Algorithm 1: Variable Metric Inverse Iteration Algorithm on $\mathbb{C}$, Based on [4]

Input: Initial guess $\boldsymbol{a}^{(0)}$, $\boldsymbol{u}^{(0)}$;
$\boldsymbol{Q} \in \mathbb{C}^{p(N-n)\times p(N-n)}, \boldsymbol{R} \in \mathbb{C}^{p(N-n)\times(n+1)} \leftarrow \mathrm{QR}(\boldsymbol{Y})$

1 $\boldsymbol{Q}_1, \boldsymbol{Q}_2 \leftarrow \boldsymbol{Q}_{[1:n+1]}, \boldsymbol{Q}_{[n+2:p(N-n)]}$ /* subscript denotes a range of columns */
2 $k \leftarrow 0$
3 **repeat**
4 $\quad k \leftarrow k+1$
5 $\quad \boldsymbol{z}^{(k)} \leftarrow \boldsymbol{R}^{-\mathsf{H}} \boldsymbol{D}_{\boldsymbol{u}^{(k-1)}} \boldsymbol{a}^{(k-1)}$
6 $\quad \boldsymbol{w}^{(k)} \leftarrow -\left(\boldsymbol{Q}_2^{\mathsf{H}} \boldsymbol{D}_{\boldsymbol{a}^{(k-1)}} \boldsymbol{Q}_2\right)^{-1} \left(\boldsymbol{Q}_2^{\mathsf{H}} \boldsymbol{D}_{\boldsymbol{a}^{(k-1)}} \boldsymbol{Q}_1\right) \boldsymbol{z}^{(k)}$
7 $\quad \boldsymbol{u}^{(k)} \leftarrow \boldsymbol{Q}_1 \boldsymbol{z}^{(k)} + \boldsymbol{Q}_2 \boldsymbol{w}^{(k)}$
8 $\quad \boldsymbol{u}^{(k)} \leftarrow \frac{\boldsymbol{u}^{(k)}}{\|\boldsymbol{u}^{(k)}\|_2}$
9 $\quad \boldsymbol{a}^{(k)} \leftarrow \boldsymbol{R}^{-1} \boldsymbol{Q}_1^{\mathsf{H}} \boldsymbol{D}_{\boldsymbol{a}^{(k-1)}} \boldsymbol{u}^{(k)}$
10 $\quad \tau^{(k)} \leftarrow \left\|\boldsymbol{a}^{(k)}\right\|_2$
11 $\quad \boldsymbol{a}^{(k)} \leftarrow \frac{\boldsymbol{a}^{(k)}}{\tau^{(k)}}$
12 **until** *convergence*;
13 $\boldsymbol{u}^{(k)} \leftarrow \boldsymbol{u}^{(k)} a_0^{(k)}$
14 $\boldsymbol{a}^{(k)} \leftarrow \boldsymbol{a}^{(k)} / a_0^{(k)}$

TABLE I
Exact Parameter Values of the Simulated NMR Signal Used in (13) in Experiment 1, as Also Used in [12, Table II]

m	f_m [Hz]	d_m [rad/s]	A_m [1]	φ_m [rad]
1	-1379	208	6.1	$15\pi/180$
2	-685	256	9.9	$15\pi/180$
3	-271	197	6.0	$15\pi/180$
4	353	117	2.8	$15\pi/180$
5	478	808	17.0	$15\pi/180$

where $t_k = kT_\mathrm{s}$ with T_s the sampling period. The frequencies f_m and damping factors d_m can be retrieved via the roots z_m of the characteristic polynomial $\boldsymbol{a}(z)$ of the LTI model.

Experiment 1: The performance of the method is tested using simulated data obtained from a model with parameters A_m, φ_m, d_m, and f_m as shown in Table I, and $\varphi_0 = 0$. Note that the frequencies f_4 and f_5 lie close to each other, making the estimation problem quite ill conditioned.

The model is evaluated for $N = 128$ data points, uniformly sampled with sampling time $T_\mathrm{s} = 10^{-4}$ s. The data sequence is perturbed with complex zero mean Gaussian noise $\epsilon_k \in \mathbb{C}$ with variance $4\nu^2$, i.e., where

$$\epsilon_k = \epsilon_k^{\mathrm{Re}} + \mathrm{j}\epsilon_k^{\mathrm{Im}}, \qquad \epsilon_k^{\mathrm{Re}}, \epsilon_k^{\mathrm{Im}} \sim \mathcal{N}\left(0, \nu^2\right). \tag{14}$$

The best total least squares (TLS) estimator of the model parameters $\boldsymbol{a}_{\mathrm{TLS}}$ is used to determine an initial guess for the inverse iteration algorithm (Algorithm 1): (1) $\boldsymbol{a}^{(0)}$ is taken as the TLS estimate $\boldsymbol{a}_{\mathrm{TLS}}$; (2) $\boldsymbol{u}^{(0)}$ is then calculated as $\boldsymbol{u}^{(0)} = (\boldsymbol{T}_{\boldsymbol{a}^{(0)}}^{\mathsf{H}})^{\dagger}(\boldsymbol{y} - \widehat{\boldsymbol{y}}_{\mathrm{TLS}})$, where $\widehat{\boldsymbol{y}}_{\mathrm{TLS}}$[5] is the model-compliant data as estimated using TLS. The method is evaluated across 200 independent noisy signal realizations $\boldsymbol{y}_i$ at varying noise levels ν, using: a single sensor (p=1); two sensors (p=2); and four sensors (p=4). For p=2 and p=4, the data is grouped into sets of two and four realizations, respectively, resulting in 100 and 50 executions of the algorithm.

Fig. 2 shows the distribution of the misfits and the relative errors on the hard to estimate frequencies f_4 and f_5 and

[5] Note that $\widehat{\boldsymbol{y}}_{\mathrm{TLS}}$ is not a critical point of (2), since it disregards the Hankel structure of the given data matrix $\boldsymbol{Y}$ when reducing its rank.

damping factors d_4 and d_5. Each of the obtained solutions is a local minimum of the realization problem (2).

It is noticeable that for larger noise levels, the given data $\boldsymbol{y}$ stray further away from the *model-compliant data space*. Since this space is merely a small subspace of the large ambient space $\mathbb{C}^{pN}$ and $\widetilde{\boldsymbol{y}}$ corresponds to the residual of the orthogonal projection of $\boldsymbol{y}$ onto this subspace, it is not unexpected to observe $\|\widetilde{\boldsymbol{y}}\|_2^2$ to increase with ν.

When the number of observations p increases, both $\|\widetilde{\boldsymbol{y}}\|_2^2$ and the accuracy of the estimated parameters improve on average and the variance of the estimates across multiple independent runs/noise realizations decreases. These results empirically support the intuition that concurrently using multiple independent sequences enhances estimation robustness. We observe that the runtime of Algorithm 1 scales approximately as $\mathcal{O}(p^3)$ for a single execution. However, when the total amount of data is fixed, increasing p reduces the number of independent runs proportionally, resulting in an overall computational cost that grows roughly quadratically with p.

It is further clear from Fig. 2 that for larger noise levels ν the parameter f_4 is not correctly estimated. This is due to the RiSVD method overfitting on the noise and consequently interpreting the contributions of f_4 and f_5 as contributions done by only a single frequency. Overfitting on the noise is a natural consequence of the objective function (2).

B. Direction of Arrival Estimation

In a second experiment we relate the autonomous realization problem to a direction of arrival (DOA) estimation problem. Estimating the DOA is a common problem in radar, sonar, and wireless communication applications [14], [18].

Consider a setup of p sensors, evenly distributed along a line as illustrated in Fig. 3. A signal $s(t)$ is transmitted and received by the sensor array under an angle ϑ, where $s(t)$ is modeled as a sum of n constant amplitude complex exponentials:

$$s(t) = \sum_{m=1}^{n} R_m e^{\mathrm{j}\psi_m} e^{\mathrm{j}2\pi f_m t}.$$

With $R_m, \psi_m, f_m \in \mathbb{R}$. The signal source is assumed to be far enough such that the DOA ϑ is virtually the same for each sensor. The ith antenna receives a phase-shifted discrete-time version of $s(t)$ perturbed by zero mean Gaussian measurement noise:

$$y_{i,k} = \sum_{m=1}^{n} R_m e^{\mathrm{j}\psi_m} e^{\mathrm{j}2\pi f_m (kT_\mathrm{s}+\varphi_0+\varphi_i)} + \epsilon_{i,k}. \tag{15}$$

We assume φ_0 to be zero. The phase-shift of the ith sensor is then given by:

$$\varphi_i = i\frac{\Delta\cos(\vartheta)}{c} \tag{16}$$

where c is the speed of light and Δ is the distance between two successive sensors. After calculating the best fitting model parameters $\boldsymbol{a}$ corresponding to the received data $\boldsymbol{y}_i$ (15), the corresponding $\widehat{R}_m$, $\widehat{\psi}_m$ and $\widehat{f}_m$ can be determined and consequently ϑ can be estimated using (16).

Experiment 2: We consider a setup of $p = 4$ sensors, where the distance between two successive sensors is $\Delta = 250\,\mathrm{m}$ and the (unknown) DOA $\vartheta = \pi/4$. We generate 200 different signals by randomly generating 200 sets of $\boldsymbol{R}, \boldsymbol{f} \in \mathbb{R}^n$ and $\boldsymbol{\psi} = \boldsymbol{0}$ and evaluate the model (15) for $N = 128$ samples, where $T_\mathrm{s} = 1\,\mu\mathrm{s}$. To ensure the frequencies f_m being sufficiently distinct, they are drawn from a normal distribution $\mathcal{N}(\frac{m-1}{2nT_\mathrm{s}}, \frac{1}{(2nT_\mathrm{s})^2})$. The amplitudes R_m are drawn from a uniform distribution on the range $[5, 15]$. The experiment is performed for different signal to noise ratios $\mathrm{SNR} = 0, 2, \ldots, 10$ dB. For each SNR, zero mean Gaussian noise with variance $4\nu^2$ (14) is added, such that $\mathrm{SNR} = 10\log_{10}(P_\mathrm{avg}/4\nu^2)$ where $P_\mathrm{avg} = \|y_\mathrm{exact}\|_2^2/pN$ is the average power of the noiseless signal. The resulting data is gathered in a vector $\boldsymbol{y}$ and structured in a Hankel matrix $\boldsymbol{Y}$. Using Algorithm 1, the optimal RiSVD

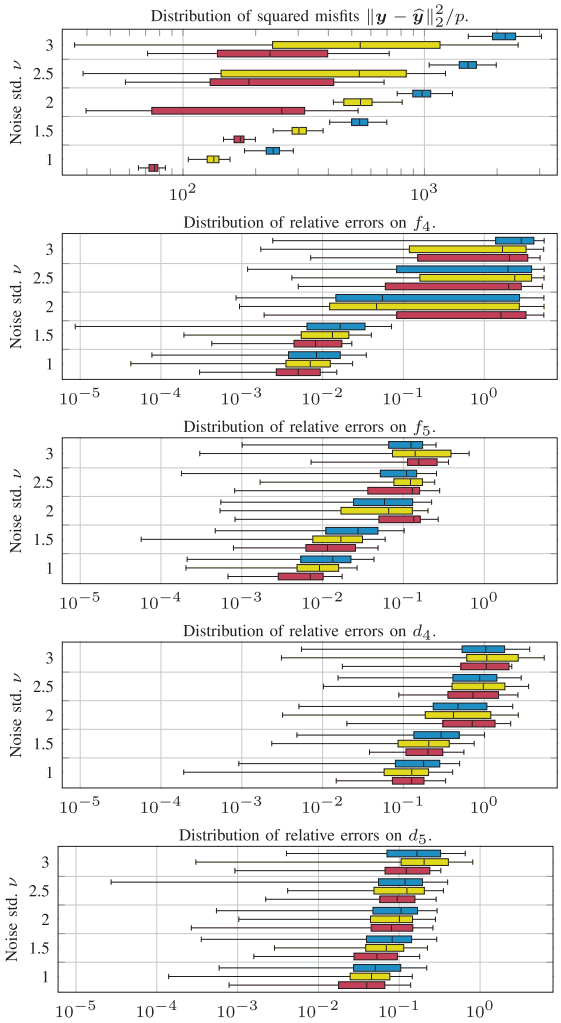


Fig. 2. Distribution of the squared 2-norm of the misfits and relative errors on f_4, f_5, d_4, and d_5 for 200 noisy realizations for various noise levels and various number of sensors p, corresponding to Experiment 1. Legend: $p = 1$, $p = 2$, $p = 4$.

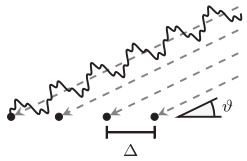


Fig. 3. A sensor array consisting of $p = 4$ sensors on a line receiving the same signal that arrives in an angle ϑ w.r.t. the sensor array. Each sensor is represented by a dot. The distance between two successive sensors is Δ.

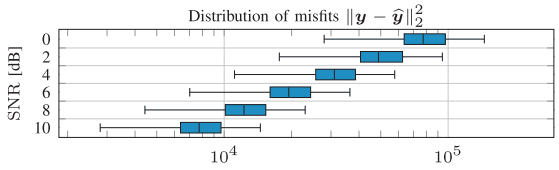


Fig. 4. Distribution of misfits for 200 different noisy realizations for various noise levels corresponding to Experiment 2.

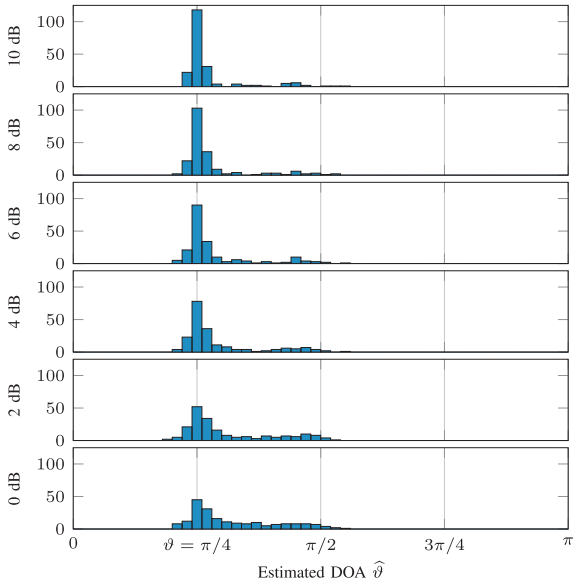


Fig. 5. Histograms of the estimated DOA $\hat{\vartheta}$ for 200 different noisy signals for SNR $= 0, 2, \ldots, 10$ dB corresponding to Experiment 2.

triplet $(\boldsymbol{u}, \tau, \boldsymbol{a})$ is determined with corresponding model-compliant data $\widehat{\boldsymbol{y}}$ via (12). The parameter estimates $\widehat{R}_m, \widehat{\psi}_m, \widehat{f}_m$ corresponding to the model-compliant data $\widehat{\boldsymbol{y}}$ can be retrieved via the model parameters $\boldsymbol{a}$, and subsequently the DOA ϑ can be estimated.

Fig. 4 shows the distribution of the misfits. Similarly as in Experiment 1, we observe increasing misfits for increasing noise levels. Fig. 5 shows histograms of the estimated DOA $\hat{\vartheta}$ of the 200 different noisy signals for each of the various noise levels SNR $= 0, 2, \ldots, 10$ dB. The histograms are constructed using bins of width $\pi/50$ rad. The estimated DOA shows a peak around the true DOA $\vartheta = \pi/4$ for all noise levels, where the peak is more pronounced for smaller noise levels. This experiment shows that for smaller noise levels the RiSVD method is able to estimate ϑ within a reasonable accuracy, and even for larger noise levels the estimations still show a peak near the true DOA ϑ.

VI. Conclusions and Future Work

We show that the (locally) optimal solutions of the complex LS realization problem follow from a nonlinear generalized singular value decomposition over the complex field and that with few minor adjustments the standard variable metric inverse iteration algorithm can be used to solve for minimizing singular triplets. The numerical experiments illustrate the use of the realization algorithm for two experimental setups.

While the focus of this work is theoretical–extending the structured TLS framework to the complex domain–an important direction for future research is to assess the practical performance of the proposed approach in comparison with the state of the art. We have not aimed to establish competitiveness in this regard, but rather to provide a rigorous foundation for further methodological exploration.

References

[1] B. D. Moor, "Least squares optimal realization of autonomous LTI systems is an eigenvalue problem," *Commun. Inf. Syst.*, vol. 20, no. 2, pp. 163–207, 2020.
[2] S. Ando, "Frequency-domain Prony method for autoregressive model identification and sinusoidal parameter estimation," *IEEE Trans. Signal Process.*, vol. 68, pp. 3461–3470, 2020.
[3] C. Prony, F. M. Gaspard, and R. De Baron, "Essai experimental et analytique, ect," *J. de l'Ecole Polytechnique de Paris*, vol. 1, pp. 24–76, Jun. 1795.
[4] B. De Moor, "Structured total least squares and L_2 approximation problems," *Linear Algebra Appl.*, vols. 188–189, pp. 163–205, Jul. 1993.
[5] B. De Moor, "Structured total least squares for Hankel matrices," in *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds., Boston, MA, USA: Springer, 1997, pp. 243–258, doi: 10.1007/978-1-4615-6281-8_13.
[6] T. J. Abatzoglou, J. M. Mendel, and G. A. Harada, "The constrained total least squares technique and its applications to harmonic superresolution," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1070–1087, May 1991.
[7] M. Levin, "Estimation of a system pulse transfer function in the presence of noise," *IEEE Trans. Autom. Control*, vol. AC-9, no. 3, pp. 229–235, Jul. 1964.
[8] T. Söderström, *Errors-in-Variables Methods in System Identification* (Communications and Control Engineering). Cham, Switzerland: Springer, 2018.
[9] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-34, no. 5, pp. 1081–1089, Oct. 1986.
[10] P. Lemmerling, L. Vanhamme, S. Van Huffel, and B. De Moor, "IQML-like algorithms for solving structured total least squares problems: A unified view," *Signal Process.*, vol. 81, no. 9, pp. 1935–1945, 2001.
[11] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice Hall, 2005.
[12] S. Vanhuffel, H. Chen, C. Decanniere, and P. Vanhecke, "Algorithm for time-domain NMR data fitting based on total least squares," *J. Magn. Reson.*, vol. 110, no. 2, pp. 228–237, Oct. 1994.
[13] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach.*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.
[14] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-37, no. 7, pp. 984–995, Jul. 1989.
[15] C. Lee, H. Hasegawa, and S. Gao, "Complex-valued neural networks: A comprehensive survey," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 8, pp. 1406–1426, Aug. 2022.
[16] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[17] K. Kreutz-Delgado. "The complex gradient operator and the $\mathbb{CR}$-calculus." 2009. [Online]. Available: https://arxiv.org/abs/0906.4835
[18] A. M. Molaei, B. Zakeri, S. M. H. Andargoli, M. A. B. Abbasi, V. Fusco, and O. Yurduseven, "A comprehensive review of direction-of-arrival estimation and localization approaches in mixed-field sources scenario," *IEEE Access*, vol. 12, pp. 65883–65918, 2024.