51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

Bi-Orthonormal Basis Functions for Improved Frequency-Domain System Identification

Robbert van Herpen Tom Oomen Okko Bosgra

Abstract— **Frequency-domain identification algorithms are considered. The aim of this paper is to develop a new algorithm that i) converges to a minimum of the objective function, and ii) possesses optimal numerical properties. Hereto, recent results in instrumental variable system identification are exploited. In addition, a new bilinear form is proposed that leads to the novel introduction of bi-orthonormal polynomials in system identification. The combination of these aspects leads to the desired convergence properties in conjunction with optimal numerical conditioning. The results are supported by means of a simulation example.**

I. Introduction

Frequency-domain system identification [14] is of significant importance for a broad class of applications, since it enables i) straightforward data reduction, ii) straightforward combination of multiple data sets, iii) a direct estimation and use of nonparametric noise models, and iv) a direct connection to control-relevant identification criteria.

Common parametric identification techniques based on frequency-domain data involve a nonlinear least-squares problem. Here, the nonlinearity arises from the parametrization of the poles in the denominator polynomial. In [11], the nonlinear problem is solved using a single linear least-squares problem. However, this requires selection of an *a priori* unknown weighting function. The SK-algorithm [17] mitigates the effect of such weighting through iterations. In [1], [5], the SK-algorithm is generalized to multivariable systems. Yet, two aspects require further attention.

On the one hand, frequency-domain identification problems are typically numerically ill-conditioned. Several partial solutions exist, including i) frequency scaling [13], ii) amplitude scaling [8], iii) the use of Möbius transformations to recast continuous-time identification problems as discrete-time ones and *vice versa*, and iv) the use of orthonormal polynomials and orthonormal rational functions with respect to a continuous inner product, see, e.g., [10] and [12] for a relation to numerical properties. These partial solutions typically improve numerical conditioning, but may be insufficient to reliably solve complex frequency-domain identification problems. Thereto, in [9], an approach is presented that leads to optimal numerical conditioning of the SK-algorithm by using polynomials that are orthonormal with respect to a *data-based discrete inner product*, see [15], [19] for a definition and earlier results.

On the other hand, the fixed point of the SK-algorithm generally does not correspond to a (local) minimum of the nonlinear least-squares criterion, as shown in [20]. Consequently, the SK-algorithm is typically used as an initialization for subsequent Gauss-Newton iterations, see, e.g., [1], which guarantees convergence to a (local) minimum.

Recently, in [6, Sect. 3.5.3 and 3.5.8], the SK-algorithm has been reformulated to guarantee that the fixed point of the iterations corresponds to an optimum of the objective function, see also [3] for the multivariable case. This renders superfluous the commonly used Gauss-Newton iteration, enabling an increase of algorithm efficiency.

Although the result in [6] [3] potentially reduces the number of iterations in frequency-domain identification, a direct implementation of the algorithm in [3] exhibits poor numerical properties. This obstructs reliable and accurate computation of the optimal model. In fact, in this paper it is shown both theoretically and by means of a numerical example that the condition numbers associated with the linear systems of equations in [3] are quadratically larger than the conditioning of standard SK-iterations as encountered in [17], [1], [5]. In addition, the approach in [15], [19], [9] to optimally condition the SK-algorithm does not apply to the procedure in [3] due to the lack of an appropriate inner product.

The main contribution of this paper is to present a novel framework for frequency-domain system identification based on a nonlinear least-squares criterion that i) is efficient in the sense that Gauss-Newton iterations are rendered superfluous, ii) ensures optimal numerical conditioning. Essentially, the proposed solution exhibits the advantageous properties that are obtained in [3], while providing optimal numerical conditioning properties as in [15], [19], [9], though through a fundamentally different mechanism.

The key technical result that differs from earlier results in system identification is the use of new basis functions that satisfy a bi-orthonormality condition with respect to a certain bilinear form. These bi-orthonormal functions replace the commonly used orthonormal functions that are used in system identification, including those with respect to a continuous inner product in [10] and those orthonormal with respect to a data-based discrete inner product in [15], [19], [9]. The possibly indefinite bilinear form that is used to construct these bi-orthonormal functions replaces the inner product in the standard orthonormal case.

The paper is organized as follows. In Sect. II, the frequency-domain identification problem is posed and two iterative algorithms are compared with respect to their convergence and numerical properties. In Sect. III, bi-orthonormal polynomials are introduced in the context of system identification, where it is proven that they provide optimal numerical conditioning

The authors are with the Eindhoven University of Technology, Department of Mechanical Engineering, Control Systems Technology Group, Eindhoven, The Netherlands (email: R.M.A.v.Herpen@tue.nl). This research is supported by ASML, Veldhoven, The Netherlands.

for a class of algorithms with advantageous convergence properties. A numerical example is given in Sect. IV. Conclusions are provided in Sect. V.

To facilitate the exposition, attention is restricted to SISO continuous-time systems. The generalization to the multivariable and discrete-time situation is conceptually straightforward, yet, the full solution is beyond the scope of the present paper.

II. Parametric frequency-domain identification: Algorithms and analysis

In this section, the frequency-domain identification problem is formulated. Subsequently, two existing algorithms are analyzed with respect to their i) convergence properties, and ii) numerical properties.

A. Frequency-domain identification problem

The following frequency-domain identification problem is considered in this paper. Let $P_o(s_k)$, $k = 1, \ldots, m$, be given measurements. Here, $s_k = j\omega_k$, $\omega_k \in \mathbb{R}$ denotes frequency. The goal is to determine a real-rational model

$$\hat{P}(s,\theta) = \frac{n(s,\theta)}{d(s,\theta)} \tag{1}$$

that minimizes the criterion

$$V(\theta) \;:=\; \left\| W\left(P_o(s) - \hat{P}(s,\theta)\right) \right\|_2^2 \tag{2}$$

$$= \sum_{k=1}^{m} \left[\left(P_o(s_k) - \hat{P}(s_k,\theta)\right)^* W(s_k)^* W(s_k) \left(P_o(s_k) - \hat{P}(s_k,\theta)\right) \right].$$

To solve (2), a polynomial basis for $n(s,\theta)$ and $d(s,\theta)$ needs to be selected. Hereto, the vector polynomial form

$$\begin{bmatrix} d(s,\theta) \\ n(s,\theta) \end{bmatrix} \;=\; \sum_{j=0}^{n} \varphi_j(s)\,\theta_j, \tag{3}$$

is considered, where $\varphi_j(s) \in \mathbb{R}^{2\times 2}[s]$ are 2×2 real-valued block polynomials, and $\theta_j \in \mathbb{R}^{2\times 1}$ are the corresponding coefficient vectors. A common choice is to parameterize $n(s,\theta)$ and $d(s,\theta)$ independently as a monomial basis, i.e.,

$$\varphi_j^{\mathrm{mon}}(s) = \begin{bmatrix} s^j & 0 \\ 0 & s^j \end{bmatrix}. \tag{4}$$

The cost criterion $V(\theta)$ in (2) is a nonlinear and typically non-convex function in θ, see (1) and (3). In the next sections, two iterative approaches are investigated to minimize (2).

B. SK-algorithm

The rationale behind the SK-algorithm is is to rewrite (2) as

$$V(\theta) \;:=\; \left\| W \tfrac{1}{d(s,\theta)} (P_o d(s,\theta) - n(s,\theta)) \right\|_2^2, \tag{5}$$

which motivates the following iterative algorithm.

Algorithm 1 (SK-iteration [17]) *Given $\theta^{\langle 0\rangle}$, iteratively solve for $\theta^{\langle i\rangle}$ the linear least-squares problem*

$$\left\| W \tfrac{1}{d(s,\theta^{\langle i-1\rangle})} \begin{bmatrix} P_o & -I \end{bmatrix} \begin{bmatrix} d(s,\theta^{\langle i\rangle}) \\ n(s,\theta^{\langle i\rangle}) \end{bmatrix} \right\|_2^2, \;\; i = 1,2,\ldots \tag{6}$$

The basic idea behind Alg. 1 is to iteratively compensate for the a priori unknown weighting $\frac{1}{d(s,\theta)}$ in (5).

To see that (6) is a linear least-squares problem indeed, observe that using (3) it can be rewritten in matrix form as

$$W_1 \;\Phi\; \theta^{\langle i\rangle} = b, \tag{7}$$

where

$$\theta^{\langle i\rangle} = \begin{bmatrix} \left(\theta_0^{\langle i\rangle}\right)^T & \left(\theta_1^{\langle i\rangle}\right)^T & \ldots & \left(\theta_{n-1}^{\langle i\rangle}\right)^T \end{bmatrix}^T, \tag{8}$$

$$\Phi \;=\; \begin{bmatrix} \varphi_0(s_1) & \varphi_1(s_1) & \ldots & \varphi_{n-1}(s_1) \\ \varphi_0(s_2) & \varphi_1(s_2) & \ldots & \varphi_{n-1}(s_2) \\ \vdots & \vdots & & \vdots \\ \varphi_0(s_m) & \varphi_1(s_m) & \ldots & \varphi_{n-1}(s_m) \end{bmatrix}, \tag{9}$$

$$W_1 \;=\; \mathrm{diag}\begin{pmatrix} w_{1,1} & \ldots & w_{1,m} \end{pmatrix}, \tag{10}$$

$$w_{1,k} \;=\; \tfrac{W(s_k)[P_o(s_k) \;\; -1]}{d(s_k,\theta^{\langle i-1\rangle})}, \tag{11}$$

$$b \;=\; W_1 \begin{bmatrix} (\varphi_n(s_1)\,\theta_n^{\langle i\rangle})^T & \ldots & (\varphi_n(s_m)\,\theta_n^{\langle i\rangle})^T \end{bmatrix}^T. \tag{12}$$

Here, $\theta_n^{\langle i\rangle}$ has been pre-specified such that $P(s,\theta)$ is strictly proper and $d(s,\theta)$ monic. Note that the matrices W and b in (7) are dependent on $\theta^{\langle i-1\rangle}$.

Next, the numerical properties of the essential computational step (7) are investigated. The numerical accuracy of the linear least-squares problem (7) is related to $\kappa(W_1\Phi)$, where

$$\kappa(\cdot) = \tfrac{\overline{\sigma}(\cdot)}{\underline{\sigma}(\cdot)}$$

is the condition number. A common choice, e.g., in [1], [5], regarding the polynomial basis $\varphi(s)$ in (3) is the monomial basis (4). In this case, the matrix Φ in (9) is a Vandermonde matrix, which is notoriously ill-conditioned for continuous-time systems. This prohibits the computation of an accurate solution to the least-squares problem in (7).

The key idea in [15], [19], [9] to obtain a numerically reliable solution to (7) is to select a basis that is orthonormal with respect to the *data-dependent discrete inner product*

$$\langle \phi(s), \psi(s) \rangle := \sum_{k=1}^{m} \psi(s_k)^H \, w_{1,k}^H \, w_{1,k} \, \phi(s_k). \tag{13}$$

If $\varphi_j(s)$, $j = 0, 1, \ldots, n-1$ is chosen to be *orthonormal with respect to* (13), then $\kappa(W_1\Phi) = 1$. This can be verified directly from the normal equations[1] corresponding to (7):

$$\underbrace{(\Phi^H \; W_1^H \; W_1 \; \Phi)}_{=\, I_{2n}} \; \theta^{\langle i\rangle} = \Phi^H \; W_1^H \; b. \tag{14}$$

It is emphasized that the specific result for the lefthand side of (14) is enabled by the use of a polynomial basis that is orthonormal with respect to the data-specific inner product (13). In contrast, other common orthonormal basis functions in system identification are orthonormal with respect to a different, e.g., *continuous* inner product that is independent of the problem data W_1 is used, for instance $\int \psi(s)^H \phi(s)$. Such a choice commonly leads to $\kappa(W_1\Phi) \gg 1$.

C. IV-type algorithm

The fixed point of the SK-iteration in Alg. 1 is generally not an optimum of $V(\theta)$ in (2), as proven in [20] and exemplified in Sect. II-D. Recently, in [6] and [3], the SK-iteration has

[1] These normal equation need not be formulated explicitly, since an orthonormal basis with respect to the inner product (13) immediately carries the optimal approximant.

been reformulated to resolve this deficiency. The basic idea is to consider the first order necessary condition for optimality:

$$\frac{\partial V(\theta)}{\partial \theta} = 0. \tag{15}$$

Specifically, rewrite Criterion (2) as $V(\theta) = \varepsilon(s,\theta)^H \varepsilon(s,\theta)$, where $\varepsilon(s,\theta) := W(P_o(s) - \hat{P}(s,\theta))$. Then, (15) is recast as

$$\zeta^H(s,\theta)\,\varepsilon(s,\theta) \;=\; 0, \quad \text{where} \quad \zeta(s,\theta) := \frac{\partial \varepsilon(s,\theta)}{\partial \theta} = \frac{-\partial \hat{P}(s,\theta)}{\partial \theta}.$$

As a result, a (local) optimum of $V(\theta)$ is attained when

$$\sum_{k=1}^{m} \left[\frac{\partial \hat{P}(s_k,\theta)}{\partial \theta}\right]^H W(s_k)^* W(s_k)\,(P_o(s) - \hat{P}(s_k,\theta)) = 0. \tag{16}$$

As discussed in [3], the variable $\zeta(z,\theta)$ can be interpreted as an instrumental variable (IV), see [18] for an overview of instrumental variable methods in system identification.

Using the parametrization (3), it is seen that (16) is a nonlinear function in θ. In analogy to the SK-algorithm Alg. 1, the following iterative procedure is suggested in [3].

Algorithm 2 (IV-iteration [3]) *Given $\theta^{\langle 0\rangle}$, iteratively solve for $\theta^{\langle i\rangle}$ the linear system of equations*

$$\sum_{k=1}^{m} \left[\frac{\partial \hat{P}(s_k,\theta)}{\partial \theta}\right]^H \Bigg|_{\theta=\theta^{\langle i-1\rangle}} W(s_k)^* \cdot \frac{W(s_k)}{d(s_k,\theta^{\langle i-1\rangle})} \begin{bmatrix} P_o(s_k) & -1 \end{bmatrix} \begin{bmatrix} d(s_k,\theta^{\langle i\rangle}) \\ n(s_k,\theta^{\langle i\rangle}) \end{bmatrix} = 0. \tag{17}$$

The essential idea is that the fixed point of the iterative procedure in Alg. 2 satisfies (15), hence, corresponds to a (local) optimum of the nonlinear least-squares Criterion 2.

By observing that

$$\zeta(s,\theta) = -\frac{\partial \hat{P}(s,\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left[\begin{bmatrix} 0 & \frac{-1}{d(s,\theta)} \end{bmatrix} \begin{bmatrix} d(s,\theta) \\ n(s,\theta) \end{bmatrix}\right] = \frac{1}{d(s,\theta)} \begin{bmatrix} \hat{P}(s,\theta) & -1 \end{bmatrix} \begin{bmatrix} \varphi_0(s) & \varphi_1(s) & \dots & \varphi_n(s) \end{bmatrix},$$

the system of equations (17) can be written in matrix form as

$$(\Phi^H W_2^H W_1 \Phi)\, \theta^{\langle i\rangle} \;=\; \Phi^H W_2^H b, \tag{18}$$

where the variables in (18) are defined in (8)-(12), and

$$W_2 \;=\; \operatorname{diag}\begin{pmatrix} w_{2,1} & \dots & w_{2,m} \end{pmatrix}, \tag{19}$$

$$w_{2,k} = \frac{W(s_k)\begin{bmatrix} \hat{P}(s_k,\theta^{\langle i-1\rangle}) & -1 \end{bmatrix}}{d(s_k,\theta^{\langle i-1\rangle})}. \tag{20}$$

The numerical accuracy of the solution to (18) crucially depends on the condition number $\kappa(\Phi^H W_2^H W_1 \Phi)$, which is typically significantly worse than the condition number $\kappa(W_1 \Phi)$ corresponding to the SK algorithm, see (7). To support this, consider the situation where upon convergence of Alg. 2, $\hat{P}(s_k,\theta^\star) \approx P_o(s_k)$, $k = 1,\dots,m$, where $\theta^\star$ denotes the fixed point of the iteration (17). In this case, $W_2 \approx W_1$, see (10) and (19). Consequently,

$$\kappa(\Phi^H W_2^H W_1 \Phi) \;\approx\; \kappa(\Phi^H W_1^H W_1 \Phi) = \kappa(W_1 \Phi)^2. \tag{21}$$

As has been argued in Section II-B, the use of the monomial basis (4) leads to $\kappa(W_1 \Phi) \gg 1$. Consequently, (21) implies that using the monomial basis in Alg. 2, as is for instance assumed in [3], leads to a severely bad conditioning of (18).

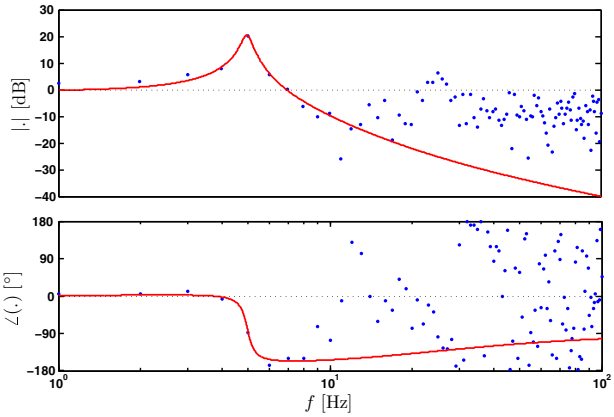

Fig. 1: Motivating example: measurements $P_o(s_k)$ (blue dot) and estimated model $\hat{P}(s,\theta)$ (red, solid).

Note that the result (21) is precisely the reason why (7) is numerically often solved using a QR-factorization of $(W_1\ \Phi)$, instead of explicitly forming the normal equations (14).

Summarizing, the use of a monomial basis in Alg. 2 is expected to lead to a poorly conditioned frequency-domain identification algorithm and thus an inaccurate solution. Although the use of orthonormal polynomials with respect to the data-based discrete inner product (13) might improve numerical conditioning, it generally will *not provide optimal conditioning*, in contrast to result (14) for the SK-algorithm. The fundamental reason is that the least-squares solution to (7) involves an *orthogonal* projection, whereas the solution to (18) involves an *oblique* projection, see [2] for a definition and interpretation. As a result, there is no underlying inner product of the form (13) for the system of equations (18).

In Sect. III, a new solution to (18) is proposed that leads to optimal numerical conditioning, i.e., $\kappa(\Phi^H W_2^H W_1 \Phi) = 1$.

D. Motivating example

In this section, the convergence and numerical properties of Alg. 1 and Alg. 2 are compared in a numerical example. Consider measurements $P_o(s_k), s_k = 1,\dots,m$ that are depicted in Fig. 1. A relatively simple low-order simulation example is considered to avoid excessively large condition numbers that obstruct convergence of the iterative procedures.

In Table I, converged results of the algorithms in [17], [3], and [9] are shown. Indeed, as explained in Sect. I, the fixed point of the SK-algorithm [17] is not a (local) minimum of $V(\theta)$, as $\left\| \frac{\partial V(\theta)}{\partial \theta}\big|_{\theta=\theta^\star_{\mathrm{SK}}} \right\|_2 > 0$. Hence, a suitable optimization routine, e.g., a Gauss-Newton iteration, has to be invoked to ensure convergence to a minimum. The IV-type algorithm in [3] does not suffer from this deficiency, as is evidenced by the numerical result in Table I. Specifically, the fixed point of the iteration $\theta^\star_{\mathrm{IV}}$ corresponds to an optimum, since the first order optimality condition $\frac{\partial V(\theta)}{\partial \theta}\big|_{\theta=\theta^\star_{\mathrm{IV}}} = 0$ holds. Indeed, this yields a slightly smaller criterion value, *i.e.*, $V(\theta^\star_{\mathrm{IV}}) < V(\theta^\star_{\mathrm{SK}})$.

TABLE I: MOTIVATING EXAMPLE

Alg.	basis	$V(\theta^\star)$	$\left\| \frac{\partial V(\theta)}{\partial \theta}\big\|_{\theta=\theta^\star} \right\|_2$	κ
1 (SK)	monomial (4)	30.47937	$1.97 \cdot 10^{-2}$	$8.09 \cdot 10^2$
1 (SK)	orthonormal (13)	30.47937	$1.97 \cdot 10^{-2}$	1.00
2 (IV)	monomial (4)	30.47901	$< 10^{-13}$	$6.56 \cdot 10^5$

Next, the numerical properties associated with both algorithms are considered. Using the monomial basis (4) in Alg. 1 leads to $\kappa(W_1\Phi) = 8.09 \cdot 10^2$. Importantly, by choosing a polynomial basis that is orthonormal with respect to the inner product (13), a linear system of equations is obtained that is *optimally conditioned* indeed, i.e., $\kappa(W_1\Phi) = 1$.

Although solving Alg. 2 with the monomial basis indeed yields an optimum of the criterion $V(\theta)$, it leads to $\kappa(\Phi^H W_2^H W_1 \Phi) = 6.56 \cdot 10^5$, which is approximately equal to $\kappa^2(W_1\Phi)$. This confirms validity of (21). In conclusion, ensuring that the fixed point of the iterative procedure is a local minimum of $V(\theta)$ seems to go at the expense of a significantly worse numerical conditioning. Unfortunately, the orthonormal polynomial theory, see, e.g., [9], that leads to optimal condition of the SK-iteration does not lead to optimal conditioning of IV-iteration in Alg. 2.

The aim of the next section is to develop an approach that achieves both the advantageous convergence properties related to Alg. 2 and optimal numerical properties.

III. Reliable frequency-domain identification: A bi-orthonormal polynomial approach

In this section, a reliable implementation of Alg. 2 is presented that relies on a novel solution for optimal numerical conditioning of (18). As observed in Sect. II-C, (18) corresponds to an *oblique projection*. As a result, the inner product (13) that is inherently related to the orthogonal projection associated with (14), does not yield optimal conditioning here. The main contribution of this section is the introduction of *bi-orthonormal polynomials* in system identification, which enables optimal numerical conditioning of (18).

The outline of this section is as follows. In Sect. III-A, a new bilinear form is considered that replaces the inner product (13). This bilinear form is shown to be intimately related to bi-orthonormal polynomials. In Sect. III-B, it is shown that these polynomials in turn relate to an inverse eigenvalue problem for a tri-diagonal matrix. In Sect. III-C, it is confirmed that this enables the efficient computation of bi-orthonormal polynomials by exploiting a three-term recurrence relation. Finally, in Sect. III-D, it is shown that the new bi-orthonormal polynomials lead to optimal numerical conditioning of (18) indeed.

A. Bi-Orthonormal polynomials

The following data-dependent bilinear form is considered.

Definition 3 *Given m distinct nodes λ_k, $k = 1, \ldots, m$. Let $w_{1,k}$, $w_{2,k} \in \mathbb{C}^{1\times\alpha}$ be corresponding nonzero weights. Then, for α-dimensional block polynomials $\pi(\lambda), \varrho(\lambda) \in \mathbb{R}^{\alpha\times\alpha}[\lambda]$, the following bilinear form is considered:*

$$\langle \pi(\lambda), \varrho(\lambda) \rangle := \sum_{k=1}^{m} \varrho(\lambda_k)^H \, w_{2,k}^H \, w_{1,k} \, \pi(\lambda_k) \,. \qquad (22)$$

It is emphasized that the bilinear form (22) is *not* an inner product. Indeed, in contrast to the positive definite form (13), the bilinear form (22) is generally indefinite. Associated with (22) is the following definition of bi-orthonormal polynomials.

Definition 4 *Consider polynomials $\pi_k(\lambda), \varrho_\ell(\lambda)$, $k, \ell = 1, \ldots, m$, where $\pi_j(\lambda), \varrho_j(\lambda)$ are of degree j. Then $\pi_k(\lambda), \varrho_\ell(\lambda)$ are bi-orthonormal polynomials (BPs) with respect to the bilinear form* (22) *if $\langle \pi_k, \varrho_\ell \rangle = \delta_{k\ell}$.*

For the SISO frequency-domain identification problem presented in Sect. II, in Def. 3, $\alpha = 2$ and $\lambda_k = s_k = j\omega_k$, where $\omega_k \in \mathbb{R}$. In the remainder of this section, however, the following simplifying assumption is imposed to facilitate the exposition and derivation of the algorithm in Sect. III-C.

Assumption 5 *Throughout Sect. III, $\alpha = 1$ and $\lambda_k \in \mathbb{R}$, $k = 1, \ldots, m$, i.e., m distinct nodes on the real line are considered. In addition, it is assumed that $w_{1,k}$, $w_{2,k} \in \mathbb{R}\backslash\{0\}$.*

Note that the theory presented in this paper can directly be generalized to α-dimensional block polynomials and arbitrary nodes $\lambda_k \in \mathbb{C}$ along similar lines.

Since nonzero weights $w_{1,k}$, $w_{2,k}$ are assumed in Assumption 5, it is always possible to find BPs as defined in Def.4. This follows from an analysis of the moment matrices associated to (22), *cf.* [4, Sect. 4.2], [16, Lem. 1]. Moreover, the BPs satisfy three-term-recurrence relations.

Lemma 6 *The BPs $\pi_j(\lambda)$, $\varrho_j(\lambda)$ in Def. 4 satisfy the three-term-recurrence relations:*

$$\pi_j(\lambda) = ((\lambda - \alpha_j)\,\pi_{j-1}(\lambda) - \beta_{j-1}\,\pi_{j-2}(\lambda))\,/\,\gamma_j, \quad (23)$$

$$\varrho_j(\lambda) = ((\lambda - \alpha_j)\,\varrho_{j-1}(\lambda) - \gamma_{j-1}\,\varrho_{j-2}(\lambda))\,/\,\beta_j, \quad (24)$$

$j = 1, \ldots, m-1$. The recursions are initialized with $\pi_{-1} = \varrho_{-1} = 0$ and $\pi_0 = \varrho_0 = 1/\sigma_0$. Here, $\sigma_0 > 0$, α_j, β_j, and γ_j are pre-determined real recursion coefficients.

The gist of constructing *data-dependent* biorthonormal polynomial bases is the efficient and reliable derivation of the recursion coefficients σ_0, α_j, β_j, and γ_j from given problem data. In the next section, the problem of deriving these parameters from given nodes λ_k with corresponding weights $w_{1,k}$ and $w_{2,k}$ is posed as an inverse eigenvalue problem.

B. An inverse eigenvalue problem for tri-diagonal matrices

Starting from the nodes and weights, the three-term-recursion coefficients can be obtained by transforming an initial matrix containing the nodes and weights into tri-diagonal form under similarity.

Lemma 7 *Define the node matrix $\Lambda \in \mathbb{R}^{m\times m}$ and weight vectors $\widetilde{W}_1, \widetilde{W}_2 \in \mathbb{R}^{m\times 1}$, see also* (11) *and* (20)*, as:*

$$\Lambda := \mathit{diag}(\lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m)\,,$$

$$\widetilde{W}_1 := \begin{bmatrix} w_{1,1} & w_{1,2} & \ldots & w_{1,m} \end{bmatrix}^T,$$

$$\widetilde{W}_2 := \begin{bmatrix} w_{2,1} & w_{2,2} & \ldots & w_{2,m} \end{bmatrix}^T.$$

There exists a similarity transformation matrix $Q \in \mathbb{R}^{m\times m}$ such that:

$$\left[\begin{array}{c|c} 1 & \\ \hline & Q^{-1} \end{array}\right] \left[\begin{array}{c|c} 0 & \widetilde{W}_2^T \\ \hline \widetilde{W}_1 & \Lambda \end{array}\right] \left[\begin{array}{c|c} 1 & \\ \hline & Q \end{array}\right] = \left[\begin{array}{c|cc} 0 & \sigma_0 & 0_{1,m-1} \\ \hline \sigma_0 & & \\ 0_{m-1,1} & & T \end{array}\right] \quad (25)$$

where $T \in \mathbb{R}^{m\times m}$ is a tri-diagonal matrix of the form:

$$T \;=\; \begin{bmatrix} \alpha_1 & \beta_1 & & \dots & 0 \\ \gamma_1 & \alpha_2 & \ddots & & \vdots \\ & \ddots & \ddots & \ddots & \\ \vdots & & \ddots & \ddots & \beta_{m-1} \\ 0 & \dots & & \gamma_{m-1} & \alpha_m \end{bmatrix}, \tag{26}$$

and $\sigma_0 = \sqrt{\widetilde{W}_2^T \widetilde{W}_1}$ *and* $\gamma_1, \dots, \gamma_{m-1}$ *are strictly positive.*

A proof of Lemma 7 follows along similar lines as in [16]. Note that the lower right component of (26) represents an inverse eigenvalue problem for the tri-diagonal matrix T. An algorithm to solve problems of this type is given in, e.g., [16], [7, p. 504]. Implementations with enhanced numerical stability can be derived along the lines of [15], [19], [9].

In the next section, it is confirmed that the inverse eigenvalue problem (26) connects to the three-term-recurrence relations for bi-orthonormal polynomials.

C. Three-term-recurrences for bi-orthonormal polynomials

Since Q in (25) is a similarity transformation matrix, the columns of Q and Q^{-T} are bi-orthonormal by definition.

Definition 8 *Let Q be given in Lemma 7. Define the matrix:*

$$P \;:=\; Q^{-T}\,. \tag{27}$$

Now, let Q and P be decomposed into individual columns:

$$Q \;=\; \begin{bmatrix} q_1, q_2, \dots, q_m \end{bmatrix}, \quad P \;=\; \begin{bmatrix} p_1, p_2, \dots, p_m \end{bmatrix},$$

By definition, Q and P are bi-orthonormal, i.e., $P^T Q = I$.

The tridiagonal similarity transformation in Lemma 7 induces *three-term-recurrence* relations between the columns of Q and P, respectively, as is formalized next.

Lemma 9 *Let Λ, $\widetilde{W}_1$, $\widetilde{W}_2$ be defined in Lemma 7. The columns q_k, p_k in Def. 8 satisfy the three-term-recurrence relations:*

$$\Lambda\, q_k \;=\; \beta_{k-1}\, q_{k-1} \;+\; \alpha_k\, q_k \;+\; \gamma_k\, q_{k+1}\,, \tag{28}$$

$$\Lambda\, p_k \;=\; \gamma_{k-1}\, p_{k-1} \;+\; \alpha_k\, p_k \;+\; \beta_k\, p_{k+1}\,. \tag{29}$$

The recursions are initialized with:

$$q_1 \;=\; \widetilde{W}_1 / \sqrt{\widetilde{W}_2^T \widetilde{W}_1}\,, \tag{30}$$

$$p_1 \;=\; \widetilde{W}_2 / \sqrt{\widetilde{W}_2^T \widetilde{W}_1}\,. \tag{31}$$

Proof: *By virtue of* (27)*, the tri-diagonal decomposition* (25) *requires $P^T \Lambda\, Q \;=\; T$, hence:*

$$\Lambda\, Q \;=\; Q\, T\,, \tag{32}$$

$$\Lambda^T P \;=\; P\, T^T\,. \tag{33}$$

Evaluating the k^{th} column of (32) *and* (33) *yields* (28) *and* (29)*. The specific initialization vectors q_1 and p_1 in* (30)–(31) *enforce the desired zeroes in the first row and column of* (25)*.*

Lemma 9 gives three-term-recurrence relations for *bi-orthonormal vectors* q_j and p_j. Using the same recursion coefficients, the recurrence relations (23)–(24), as introduced in Sect. III-A, yield a related set of *polynomials*. In the following lemma, this connection between the BPs $\{\pi(\lambda), \varrho(\lambda)\}$ in Def. 4 and the columns of $\{Q, P\}$ in Def. 8 is made explicit.

Lemma 10 *Consider weight matrices $W_1, W_2 \in \mathbb{R}^{m\times m}$ as in* (10) *and* (19)*, respectively, where under Assumption 5 $w_{1,k}, w_{2,k} \in \mathbb{R}$ are scalar. Denote with $\Pi_j, R_j \in \mathbb{R}^{m\times 1}$ the columns obtained after evaluation of $\pi_j,\ \varrho_j, j = 0, \dots, m-1$ at the considered nodes λ_k,* i.e.*:*

$$\Pi_j \;:=\; \begin{bmatrix} \pi_j(\lambda_1) & \pi_j(\lambda_2) & \dots & \pi_j(\lambda_m) \end{bmatrix}^T, \tag{34}$$

$$R_j \;:=\; \begin{bmatrix} \varrho_j(\lambda_1) & \varrho_j(\lambda_2) & \dots & \varrho_j(\lambda_m) \end{bmatrix}^T. \tag{35}$$

The columns q_{j+1}, p_{j+1} in Def. 8 and the BPs $\pi_j(\lambda), \varrho_j(\lambda)$ in Def. 4 are related as follows:

$$q_{j+1} \;=\; W_1\, \Pi_j\,,$$

$$p_{j+1} \;=\; W_2\, R_j\,.$$

Proof: *Follows by rewriting* (28)–(29)*, initialized with* (30)–(31)*, as:*

$$q_{k+1} = \left((\Lambda - \alpha_k)\, q_k \;-\; \beta_{k-1}\, q_{k-1}\right) / \gamma_k,$$

$$p_{k+1} = \left((\Lambda - \alpha_k)\, p_k \;-\; \gamma_{k-1}\, p_{k-1}\right) / \beta_k,$$

and comparing with (23)–(24)*, with $\pi_0 = \varrho_0 = 1/\sqrt{\widetilde{W}_2^T \widetilde{W}_1}$.*

The result in Lemma 10 implies that (23)–(24), with recursion coefficients obtained from the inverse eigenvalue decomposition (25), indeed yield bi-orthonormal polynomials in the sense of Def. 4, hence, yield optimal conditioning.

D. Bi-orthonormal polynomials lead to optimally conditioned frequency-domain identification

The following result constitutes the main result of this section.

Lemma 11 *Let $\pi_j(\lambda), \varrho_j(\lambda), j = 0, \dots, n-1$ be polynomials obtained by applying* (23)–(24) *with recursion coefficients from the decomposition* (25)*. Then, π_j, ϱ_j are bi-orthonormal with respect to the bilinear form* (22)*, i.e.,*

$$R^T W_2 W_1\, \Pi \;=\; I_n\,, \tag{36}$$

$\Pi := [\Pi_0, \dots, \Pi_{n-1}]$, $R := [R_0, \dots, R_{n-1}]$, see (34)–(35)*.*

Proof: *Since Q in Lemma 7 is a similarity transformation matrix, Q and P are biorthonormal,* i.e.*, $P^T Q = I$,* cf. *Def. 8. Using Lemma 10, this biorthonormality condition can be rewritten as* (36)*. Since Π and R are evaluations of $\pi(\lambda)$ and $\varrho(\lambda)$ at the considered set of nodes $\lambda_k, k = 1, \dots, m$,* (36) *implies that π_j, ϱ_j are biorthonormal in the sense of* (22)*.*

Returning to the linear system of equations corresponding to Alg. 2, bi-orthonormal polynomials with respect to the bilinear form (22) lead to the following system of equations:

$$(R^T W_2 W_1\, \Pi)\theta^{\langle i \rangle} = R^T W_2 b. \tag{37}$$

By virtue of the result (36), $\kappa(R^T W_2 W_1\, \Pi) = 1$, leading to an optimal numerical conditioning as desired.

From an identification perspective, the essential difference of (37) compared to (18) is that two sets of basis functions are used. One is used to parameterize the model and the other to parameterize the instrumental variables ζ. In this respect, the choice of basis related to R in (37) can be interpreted as a linear transformation of the instruments ζ. In fact, the solution is invariant under such transformations, *cf.* [18, Section 3.1].

In the next section, the reason for optimal conditioning of Alg. 2 by using bi-orthonormal polynomials is supported with a brief example.

IV. Example

In this section, the results of Sect. III are illustrated.

Measurements *Consider the* 7^{th} *order system* $P_o(\lambda) = \frac{n_o(\lambda)}{d_o(\lambda)} = \frac{(\lambda-400)(\lambda-250)(\lambda-50)(\lambda+10)(\lambda+50)(\lambda+250)(\lambda+400)+4\cdot 10^{16}}{(\lambda-300)(\lambda-200)(\lambda-100)(\lambda)(\lambda+100)(\lambda+200)(\lambda+300)+2.5\cdot 10^{16}}$.

A 300*-point perturbed response vector* $\tilde{P}_o(\lambda_k)$ *is generated:*

$$\tilde{P}_o(\lambda_k) \;=\; P_o(\lambda_k) + \nu(\lambda_k),$$
$$\lambda_k = \{-299, -297, \ldots, 297, 299\}, \quad \textit{where}$$

$\nu(\lambda_k)$ *is a zero mean random noise with variance* $\sigma_\nu^2 = 0.5$.

Approximation of P_o by a rational transfer function requires simultaneous optimization of the numerator and denumerator polynomials, *cf.* Alg. 2. This in turn demands for $\alpha = 2$-dimensional vector polynomials in Def. 3. For the sake of simplicity, however, $n_o(\lambda)$ is fixed, enabling the use of scalar basis polynomials, i.e., $\alpha = 1$. Thus, Assumption 5 holds.

Model *A* 7^{th} *order model* $\hat{P}$ *with exact numerator polynomial* n_o *is considered:*

$$\hat{P}(\lambda,\theta) \;=\; \frac{n_o(\lambda)}{d(\lambda,\theta)}, \qquad d(\lambda,\theta) = \sum_{j=0}^{n} \varphi_j(\lambda)\,\theta_j,$$

with basis polynomials $\varphi_j(\lambda) \in \mathbb{R}[\lambda]$ *and coefficients* $\theta_j \in \mathbb{R}$.

Reformulating (17) in Alg. 2 using a *monomial basis* yields $\kappa(\Psi_{\text{mon}}^T W_2^T W_1 \Psi_{\text{mon}}) = 5.51 \cdot 10^{32}$ in (18). This is due to the large dispersion in the amplitude of the basis functions on the considered domain, *cf.* Fig. 2. In contrast, the use of *bi-orthonormal basis polynomials* yields $\kappa(R^T W_2^T W_1 \Pi) = 1.00$ in (37), due to the fact that the basis functions are selected specific to the data of the identification problem at hand, see Fig. 3 for an illustration.

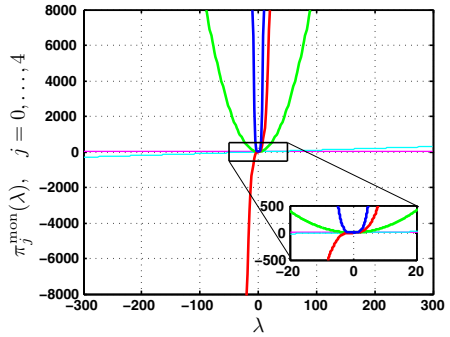

Fig. 2: First 5 monomial basis functions $\varphi_j^{\text{mon}}(\lambda), j = 0, \ldots, 4$.

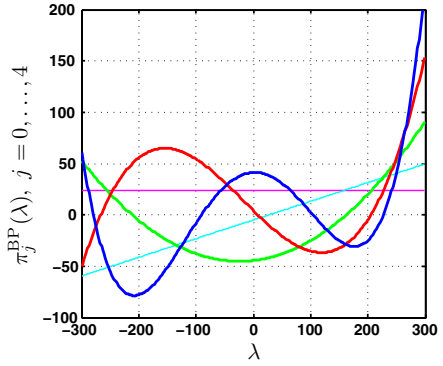

Fig. 3: First 5 basis functions $\pi_j^{\text{BP}}(\lambda), j = 0, \ldots, 4$, which are bi-orthonormal with $\varrho_j^{\text{BP}}(\lambda)$.

V. Conclusions

In this paper, a frequency-domain system identification algorithm is considered that has favorable convergence properties, in the sense that the fixed point of the iterative procedure is an optimum of the identification criterion. It is shown that a reliable implementation of the algorithm hinges on the choice of basis polynomials. The key novel result of this paper is the introduction of bi-orthonormal polynomials in system identification to achieve optimal numerical conditioning. This is illustrated with an example.

Acknowledgement

The authors thank Rogier Blom for an interesting exchange of ideas.

References

[1] D. S. Bayard. High-order multivariable transfer function curve fitting: Algorithms, sparse matrix methods and experimental results. *Automatica*, 30(9):1439–1444, 1994.

[2] R. T. Behrens and L. L. Scharf. Signal processing applications of oblique projection operators. *IEEE Trans. Signal Processing*, 42(6):1413–1424, 1994.

[3] R. S. Blom and P. M. J. Van den Hof. Multivariable frequency domain identification using IV-based linear regression. *Proc. 49th IEEE Conf. Dec. Cont., Atlanta, GA, USA*, pages 1148–1153, 2010.

[4] C. Brezinski. *Padé-Type Approximation and General Orthogonal Polynomials*. Birkhäuser Verlag, Basel, Switzerland, 1980.

[5] R. A. de Callafon, D. de Roover, and P. M. J. Van den Hof. Multivariable least squares frequency domain identification using polynomial matrix fraction descriptions. *Proc. 35th Conf. Dec. Cont., Kobe, Japan*, pages 2030–2035, 1996.

[6] S. G. Douma. *From data to performance: System identification uncertainty and robust control design*. PhD thesis, Technische Universiteit Delft, 2006.

[7] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, USA, 1989.

[8] R. G. Hakvoort and P. M. J. Van den Hof. Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability. *International Journal of Control*, 60(5):809–825, 1994.

[9] R. van Herpen, T. Oomen, and O. Bosgra. Numerically reliable frequency-domain estimation of transfer functions: A computationally efficient methodology. *16th IFAC Symp. Sys. Id., Brussels, Belgium*, pages 595–600, 2012.

[10] P. S. C. Heuberger, P. M. J. Van den Hof, and B. Wahlberg. *Modelling and Identification with Rational Orthogonal Basis Functions*. Springer Verlag, New York, NY, USA, 2005.

[11] E. C. Levy. Complex-curve fitting. *IRE Trans. Automatic Control*, 4(1):318–321, 1959.

[12] B. Ninness and H. Hjalmarsson. Model structure and numerical properties of normal equations. *IEEE Trans. Circuits and Systems*, 48(4):425–437, 2001.

[13] R. Pintelon and I. Kollár. On the frequency scaling in continuous-time modelling. *IEEE Trans. Instrument. Meas.*, 54(1):318–321, 2005.

[14] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach*. IEEE Press, New York, NY, USA, 2001.

[15] L. Reichel, G. S. Ammar, and W. B. Gragg. Discrete least squares approximation by trigonometric polynomials. *Mathematics of Computation*, 57(195):273–289, 1991.

[16] Y. Saad. The Lanczos biorthogonalization algorithm and other oblique projection methods for solving large unsymmetric systems. *SIAM Journal on Numerical Analysis*, 19(3):485–506, 1982.

[17] C. K. Sanathanan and J. Koerner. Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. Automatic Control*, 8(1):56–58, 1963.

[18] T. Söderström and P. Stoica. Instrumental variable methods for system identification. *Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin*, 1983.

[19] M. Van Barel and A. Bultheel. Orthonormal polynomial vectors and least squares approximation for a discrete inner product. *Electronic Trans. Numerical Analysis, Kent State University*, 3:1–23, 1995.

[20] A. H. Whitfield. Asymptotic behaviour of transfer function synthesis methods. *International Journal of Control*, 45(3):1083–1092, 1987.