REFERENCES

[1] R. R. Mohler, *Bilinear Systems, Volume II, Applications to Bilinear Control*. Upper Saddle River, NJ: Prentice-Hall, 1991.

[2] C. Bruni, G. Dipillo, and G. Koch, "Bilinear systems: an appealing class of nearly linear systems in theory and applications," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 334–348, 1974.

[3] M. Slemrod, "Stabilization of bilinear control systems with applications to nonconservative problems in elasticity," *SIAM J. Control Optim.*, vol. 16, pp. 131–141, 1978.

[4] J. P. Quinn, "Stabilization of bilinear systems by quadratic feedback controls," *J. Math. Anal. Applicat.*, vol. 75, pp. 66–80, 1980.

[5] E. P. Ryan and N. J. Buckingham, "On asymptotically stabilizing feedback control of bilinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 863–864, 1983.

[6] P. O. Gutman, "Stabilizing controllers for bilinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 917–921, 1981.

[7] M. S. Chen, "Exponential stabilization of a constrained bilinear system," *Automatica*, vol. 34, no. 8, pp. 989–992, 1998.

[8] M. E. Evans and D. N. P. Murthy, "Controllability of a class of discrete-time bilinear systems," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 78–83, Jan. 1977.

[9] Y. R. Hwang, "On the reachability of a class of continuous bilinear systems," *J. Chinese Inst. Eng.*, vol. 24, no. 5, pp. 635–639, 2001.

[10] F. Callier and C. A. Desoer, *Linear System Theory*. New York: Springer-Verlag, 1991.

[11] C. T. Chen, *Linear System Theory and Design*. New York: Holt, Rinehart, and Winston, 1984.

[12] P. Lu, "Closed-form control laws for linear time-varying systems," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 537–542, Mar. 2000.

[13] M. Vidyasagar, *Nonlinear Systems Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[14] S. Sastry, *Nonlinear Systems, Analysis, Stability, and Control*. New York: Springer-Verlag, 1999.

Orthogonal Rational Functions for System Identification: Numerical Aspects

Patrick Van gucht and Adhemar Bultheel

***Abstract*—Recently, there has been a growing interest in the use of orthogonal rational functions (ORFs) in system identification. There are many advantages over more classical techniques. Probably due to a known explicit expression for the basis functions when the orthogonality weight is uniformly equal to 1 (the so called Malmquist basis), the attention has been on the development of methods using this basis. However, for some discrete identification problems, this choice of the orthogonality weight may still lead to serious numerical problems due to the ill conditioning of the linear system of equations to be solved. In this note, we give an algorithm based on a more general system of ORF to overcome the numerical problem and which allows for a fast-order update of the estimate.**

***Index Terms*—Least squares, orthogonal rational functions (ORFs), system identification.**

Manuscript received January 31, 2001; revised November 5, 2001 and November 5, 2002. Recommended by Associate Editor E. Bai. This work was supported in part by the Fund for Scientific Research (FWO), project "CORFU: Constructive study of orthogonal functions," under Grant G.0184.02, and in part by the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Sciences, Technology and Culture, Grant IPA V/22. The scientific responsibility rests with the authors.

The authors are with the Department of Computer Science, Katholieke Universiteit Leuven, 3000 Leuven, Belgium (e-mail: Patrick.Vangucht@cs.kuleuven.ac.be).

Digital Object Identifier 10.1109/TAC.2003.809761

I. INTRODUCTION

The use of orthogonal rational functions (ORFs) in system identification, has been advertised by many authors in the recent literature. The advantages of representing a transfer function as a linear combination of ORF are numerous. However, in our opinion, some aspect has not been given proper attention. It concerns the choice of the appropriate inner product for the orthogonality. The ORF that are used in the literature are usually orthogonal with respect to a uniform measure on the complex unit circle or the real line.

For example, for discrete time systems, the polynomials orthogonal with respect to the Lebesgue measure on the unit circle are simply the monomials $\{z^k\colon\ k = 0, 1, \ldots\}$ (Fourier basis), while for a general measure they result in more general Szegő polynomials. The computation of the monomials is trivial since they are explicitly known. The general orthogonal polynomials with respect to an arbitrary measure are, however, also simple to compute. Given the moments of the orthogonality measure (e.g., the autocorrelation coefficients of the impulse response), they can be efficiently computed by the Levinson or Schur algorithm. The use of the Fourier basis is inappropriate in cases where the approximating model is computed for successive orders. It does not allow a recursive computation (updating an AR model of degree n to degree $n+1$ requires the recomputation of all the coefficients with respect to the Fourier basis) and it may result in a very bad numerical conditioning of the problem, which is manifested by very big rounding errors that may overrule the exact results completely.

Exactly the same situation occurs in the case of ORF. If the Lebesgue measure is used, an explicit form for the ORF is known. These ORF are attributed by Walsh [1, p. 224] to Takenaka [2] and Malmquist [3] (called TM basis in the rest of this note). The advantage of knowing the basis functions explicitly can be completely overshadowed by the disadvantages from a computational and numerical viewpoint. Models of successive degrees are not nested, so that a recursive update is impossible, and there can be a considerable loss of accuracy due to rounding errors for certain problem settings like for example frequency data in a narrow band and/or a high-degree model.

In this note, we want to show that by choosing an appropriate weight function in discrete time frequency domain identification, the ORF can still be computed efficiently, while the numerical conditioning becomes theoretically optimal.

Of course, given the same poles of the system and given the same objective function, the optimal model that we compute will theoretically be exactly the same as the one computed by any other method. It may also be that with respect to other optimality criteria, our computed solution is worse than the one computed by another method. We do not want to enter a polemic of what an optimality criterion should be, or whether computations should be time or frequency domain based. Our main point is that there is no reason why, from a numerical point of view, one should restrict the discussion to the TM basis. Although the theoretical analysis of the more general basis can be considerably more complicated, the numerics are not, so if another orthogonality weight is more appropriate, then it should be used. We illustrate our point with a very simple case of discrete frequency domain identification, but the idea is applicable in many other situations as well.

We will consider a discrete-time system

$$y(t) = G_0(q)u(t) + v(t), \qquad t \in \{0, \pm 1, \pm 2, \ldots\} \tag{1}$$

where the input $u(t)$ and output $y(t)$ are supposed to be in l^2 and the noise $v(t)$ is filtered white noise: $v(t) = H(q)e(t)$. This setting is standard [4].

The z-transform is denoted by capitals and it is defined using the $1/z$ convention. For example, the z-transform of the input is $U(z) = \sum_k u(k)z^{-k}$. We look for a function $\hat{G}_n = \sum_{l=0}^n \theta_l \phi_l$, where the $\{\theta_k\}$ are the expansion coefficients to be determined and the $\{\phi_k\}$ form an ORF basis. We want to find $\{\theta_k\}_{k=0}^n$ from frequency domain data. Thus, we suppose that $U(z)$ and $Y(z)$ are known for a certain set of measurement points on the unit circle, which we denote

$$\mathcal{T}_N = \{z_k = e^{i\omega_k},\, k = 1, \ldots, N\}. \tag{2}$$

We use Ω_N to denote the frequencies of the points in $\mathcal{T}_N$

$$\Omega_N = \{\omega_k,\, k = 1, \ldots, N,\, e^{i\omega_k} \in \mathcal{T}_N\}. \tag{3}$$

The most simple orthogonal functions are the monomials $z^{-k}, k \geq 0$ (polynomials in z^{-1}). These are orthogonal with respect to the inner product $\langle f, g\rangle = \int_{-\pi}^{\pi} f(e^{i\omega})\overline{g(e^{i\omega})}\,\mathrm{d}\lambda$, with $\mathrm{d}\lambda = (1/2\pi)\,\mathrm{d}\omega$ the (normalized) Lebesgue measure on the unit circle. (In the rest of this note λ will always refer to this measure.)

The simplest form of nonpolynomial ORF are the *Laguerre functions* obtained by orthogonalization of the basis $\{\zeta^k: k = 0, 1, 2, \ldots\}$ with $\zeta(z) = (1 - az)/(z - a)$, used by Lee [5] as early as 1933, but in the context of identification only since the late 1970s. Since then, also the two-parameter Laguerre models received much attention. These models were introduced much earlier by Kautz [6] in the 1950s. The Kautz basis repeats two (a complex conjugate pair of) poles instead of just one. For an historical survey and many references on the Kautz and Laguerre models, we refer to the introduction of [7]. In the 1990s, Van den Hof and Heuberger *et al.* introduced a finite number of poles that are repeated cyclically [8]–[10]. They also give a state-space description. Shortly after, Ninness *et al.* [11]–[15] used an arbitrary sequence of poles, which lead to the TM basis [1, pp. 224], the most general, and, in fact, also with the oldest roots (Takenaka [2] and Malmquist [3] papers dating back to the 1920s). For a state-space description, see [16].

The orthogonality considered in all these papers is with respect to the Lebesgue measure λ. A theory of ORF with respect to a more general measure μ has been given in [17].

II. Spaces of Rational Functions and Orthogonal Bases

We denote the unit circle in $\mathbb{C}$ as $\mathbb{T}$ its exterior as $\mathbb{E}$ and its interior as $\mathbb{D}$.

Let $\underline{\alpha} = \{\alpha_1, \alpha_2, \ldots\} \subset \mathbb{D}$ be an arbitrary sequence of preselected poles. The space of models of degree n at most is denoted as

$$\mathcal{L}_n = \left\{\frac{p_n(z)}{\pi_n(z)}: p_n \in \Pi_n,\, \pi_n(z) = \prod_{k=1}^n (z - \alpha_k)\right\}$$

where Π_n denotes the space of polynomials of degree less than or equal to n.

Let μ be a positive measure on $\mathbb{T}$. The Hilbert space $L^2(\mu)$ with inner product

$$\langle f, g\rangle_\mu = \int f(e^{i\omega})\overline{g(e^{i\omega})}\,\mathrm{d}\mu, \qquad f, g \in L^2(\mu) \tag{4}$$

is well known. We consider orthonormal basis functions $\{\phi_0, \phi_1, \ldots\}$ for $L^2(\mu)$ satisfying $\langle \phi_k, \phi_l\rangle_\mu = \delta_{kl}$, with δ_{kl} the Kronecker delta, and such that $\mathcal{L}_n = \mathrm{span}\{\phi_0, \phi_1, \ldots \phi_n\}, n = 0, 1, \ldots$. They may, e.g., be obtained by a Gram–Schmidt procedure applied to the basis of Blaschke products $B_n(z) = \prod_{l=1}^n \zeta_l(z), n = 1, 2, \ldots (B_0 = 1)$ with Blaschke factors $\zeta_l(z) = (1 - \overline{\alpha_l}z)/(z - \alpha_l)$. Note that if all $\alpha_k = 0$, then the space $\mathcal{L}_n$ reduces to Π_{-n}, the space of all polynomials in z^{-1} of degree less than or equal to n.

For convergence reasons it is important that the $L^2(\mu)$-closure of $\mathcal{L} = \cup_{n\geq 0}\mathcal{L}_n$ is $H^2(\mu)$, the space of all $L^2(\mu)$ functions for which $\int e^{-ik\omega}\overline{f(e^{i\omega})}\,\mathrm{d}\mu = 0$ for $k = 1, 2, \ldots$. When μ is the normalized Lebesque measure then the Blaschke condition, i.e., $\sum_{k=1}^\infty (1 - |\alpha_k|) = \infty$ is necessary and sufficient for density. For a more general μ, it is sufficient (but not necessary).

Theorem 1: If the Blaschke condition holds, then $\mathcal{L}$ is dense in $H^2(\mu)$ for any finite positive measure μon $\mathbb{T}$. If $\mu = \lambda$ (the Lebesgue measure) then the Blaschke condition is also necessary.

For a proof when the poles are all in $\mathbb{E}$, see [17]. The adaptation to the present situation is trivial (replace z by $1/z$).

The TM basis is obtained when $\mu = \lambda$, [1, p. 224].

Theorem 2 (TM Basis): If μ is the normalized Lebesgue measure, then the orthonormal basis is given by

$$\phi_0 = 1, \qquad \phi_k(z) = \frac{\sqrt{1 - |\alpha_k|^2}z}{z - \alpha_k} B_{k-1}(z), \qquad k = 1, 2, \ldots.$$

If all $\alpha_k = 0$, we obtain the Fourier basis $\phi_k(z) = z^{-k}$.

The TM basis is used, e.g., in [12] and several other places. For a state–space-based proof, see [16].

For an arbitrary measure μ, there is in general no explicit expression for these ORF, but the polynomial Szegő–Levinson recursion can be smoothly generalized. That allows for a very efficient computation of the ORF. It can also be found in [17] for poles in $\mathbb{E}$. For poles in $\mathbb{D}$, it reads (see [18]).

Theorem 3: Let ϕ_n be the normalized ORF with respect to a positive measure μ on $\mathbb{T}$. Define $\phi_n^*(z) = B_n(z)\overline{\phi_n(1/\overline{z})}$, then

$$\begin{bmatrix}\phi_n(z)\\ \phi_n^*(z)\end{bmatrix} = e_n \frac{z - \alpha_{n-1}}{z - \alpha_n}\begin{bmatrix}1 & \overline{L_n}\\ L_n & 1\end{bmatrix} \times \begin{bmatrix}\zeta_{n-1}(z) & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\phi_{n-1}(z)\\ \phi_{n-1}^*(z)\end{bmatrix}$$

where

$$L_n = -\frac{\left\langle \phi_k,\, \frac{1-\overline{\alpha_{n-1}}z}{z-\alpha_n}\,\phi_{n-1}\right\rangle_\mu}{\left\langle \phi_k,\, \frac{z-\alpha_{n-1}}{z-\alpha_n}\,\phi_{n-1}^*\right\rangle_\mu} \qquad \forall k \leq n-1 \tag{5}$$

$$e_n = \left(\frac{1 - |\alpha_n|^2}{1 - |\alpha_{n-1}|^2}\,\frac{1}{1 - |L_n|^2}\right)^{1/2}. \tag{6}$$

The initial conditions are $\phi_0(z) \equiv \langle 1, 1\rangle_\mu^{-1/2} \equiv \phi_0^*(z)$.

For a state–space-based proof, see [19].

III. Identification Problem

Let G_0 be the stable rational transfer function of the system so that (1) in the z-domain becomes

$$Y(z) = G_0(z)U(z) + V(z) \tag{7}$$

with $U(z), Y(z) \in H^2(\lambda)$ and $V(z) = H(z)E(z)$.

First, we have to choose the model poles. These are of course a set of nonlinear parameters that should be chosen in an optimal way. There are many approaches to solve this highly nonlinear and nontrivial problem, that we do not discuss in this paper. We suppose they are given.

We are trying to identify the parameters $\underline{\theta} = [\theta_0, \ldots, \theta_n]^T$ in the approximate model

$$\hat{G}_n(z, \underline{\theta}) = \sum_{l=0}^n \theta_l \phi_l(z). \tag{8}$$

Remark: The approximation $\hat{G}_n$ of the transfer function G is proper, but not strictly proper. A strictly proper transfer function can be modeled by redefining $\mathcal{L}_n$ as the span of the Blaschke products $B_k(z)/z$ and some other small changes to the theory, but we do not go into detail here.

Given the experimental frequency data $\{U(z_k), Y(z_k)\}_{k=1}^N$ at the points $z_k \in \mathcal{T}_N$, we want to minimize $\sum_{k=1}^N |E_n(z_k, \underline{\theta})K(z_k)|^2$ where

$$E_n(z, \underline{\theta}) = Y(z) - \hat{Y}(z, \underline{\theta}) = Y(z) - \left(\sum_{l=0}^n \theta_l \phi_l(z)\right) U(z) \tag{9}$$

and $|K(z_k)|^2$ are weights that are given. This setting is used, e.g., in [20] with $K(z_k) = W_k/U(z_k)$. The weights $|K(z_k)|^2$ can, e.g., be used to pay more attention on some frequencies.

IV. Why a More General Measure μ?

For simplicity of the presentation, we put the minimization problem from above in vector formulation. Thus, we want to minimize $\|\mathbf{\Lambda E}\| = (\mathbf{E}^H\mathbf{\Lambda}^H\mathbf{\Lambda E})^{1/2}$ with $\mathbf{E} = \mathbf{Y} - \hat{\mathbf{Y}}$ (here and in the rest of this paper we use bold face letters like $\mathbf{Y}$ to denote column vectors of values $Y(z_k)$, $k = 1, \ldots, N$ in the measurement points $z_k = e^{i\omega_k} \in \mathcal{T}_N$) and $\mathbf{\Lambda} = \mathrm{diag}(K(z_1), K(z_2), \ldots, K(z_N))$.

It is our aim to show that the TM basis of Theorem 2 may lead to an ill-conditioned problem. Therefore, we recall from numerical linear algebra (see [21, p. 311]).

Theorem 4: Consider the linear least squares problem $\min_{\mathbf{X}} \|\mathbf{F}\|$ with $\mathbf{F} = A\mathbf{X} - \mathbf{B}$. Denote the 2-condition number by $\kappa(A)$ (this is the ratio of the largest over the smallest nonzero singular value of A) and the machine precision by ϵ_M (this ϵ_M is approximately 10^{-16} in IEEE double precision arithmetic). Suppose that $\tilde{\mathbf{X}}$ is the solution computed by the QR method and $\mathbf{X}$ is the exact solution, then

$$\frac{\|\tilde{\mathbf{X}} - \mathbf{X}\|}{\|\mathbf{X}\|} \le 2\gamma\kappa(A)\epsilon_M + \gamma\kappa^2(A)\,\frac{\|\mathbf{F}\|}{\|A\|\,\|\mathbf{X}\|}\,\epsilon_M$$

with γ slowly growing with the dimension.

If $\kappa(A) > 10^{16}$, all significant digits may be lost in the computation. We have an optimal conditioning $\kappa(A) = 1$ if the columns of A are orthonormal: $A^H A = I$. So, here is our main theorem

Theorem 5: The Jacobian of the least squares problem $\min_{\underline{\theta}} \|\mathbf{\Lambda E}(\underline{\theta})\|$, previously described, will have an optimal condition number when the basis functions ϕ_k are orthogonal with respect to a discrete measure μ taking the values $|U(z_k)K(z_k)|^2$ at the points $z_k \in \mathcal{T}_N$.

Proof: With the notation previously introduced, we get

$$\begin{aligned}\|\mathbf{\Lambda E}\|^2 &= \sum_{k=1}^{N}\left|\left(G(z_k) - \hat{G}_n(z_k, \underline{\theta})\right)U(z_k)K(z_k)\right|^2\\ &= \|G - \hat{G}_n(\cdot, \underline{\theta})\|_\mu = \left\|G - \sum_{l=0}^{n}\phi_l\theta_l\right\|_\mu\\ &= \|\mathbf{\Lambda Y} - W\Phi\underline{\theta}\| = \|\mathbf{\Upsilon} - \Psi\underline{\theta}\|\end{aligned}$$

where μ is as claimed, $\mathbf{\Upsilon} = \mathbf{\Lambda Y}$ and $W = \mathrm{diag}(U(z_1)K(z_1), U(z_2)K(z_2), \ldots, U(z_N)K(z_N))$. The Jacobian is the matrix Ψ with (k, l)-element $\psi_l(z_k)$ where $\psi_l = KU\phi_l$. The condition number is optimal if $\Psi^H\Psi = \Phi^H W^H W\Phi = I$, i.e., if $\langle\phi_i, \phi_j\rangle_\mu = \delta_{i,j}$. ■

Note that if all the weights $|K(z_k)|^2$ are equal and the spectrum of the input signal $|U(z)|^2$, with $z \in \mathcal{T}_N$ is flat, then the weighting will not be of great importance, but if the weights are not equal or the spectrum is not flat, then one can expect numerical problems.

Example 1: Choose as basis the Laguerre functions with pole $\alpha = 0.5$. For the input $u(t) = 1$, $t = 0, \ldots, 50$ with the same number of frequency data equidistant in the interval $[-\pi/6, \pi/6]$ and no extra weighting ($\mathbf{\Lambda} = I$), the least squares problem for an approximation order 20, has a condition number of order $O(10^9)$. Taking 300 samples and an approximation order 33 gives a condition number $O(10^{16})$. This means that we can loose all digits in the computation of $\underline{\theta}$ due to rounding errors. This simple example has a very oscillating spectrum, which results in the numerical instability.

Thus, using a basis orthogonal with respect to the Lebesgue measure does not solve the numerical problems. A data-dependent measure μ, should be used instead.

V. Identification Algorithm

Suppose we are given the frequency response of the input and output $\{U(e^{i\omega_k}), Y(e^{i\omega_k})\}_{k=1}^N$ and the weights $\{|K(e^{i\omega_k})|^2\}_{k=1}^N$ in the frequencies $\omega_k \in \Omega_N$. With μ as in Theorem 5, the inner product (4) becomes (denote $W = UK$)

$$\langle f, g\rangle_\mu = \sum_{k=1}^{N} f(e^{i\omega_k})W(e^{i\omega_k})\overline{W(e^{i\omega_k})g(e^{i\omega_k})}. \qquad (10)$$

By multiplying the recurrence relation of Theorem 3 with $W = UK$, we get a recurrence for the functions $\psi_l = W\phi_l$. Since the recurrence is to be evaluated for vectors of function values, the computations can be executed very efficiently in Matlab.

In theory, these columns are orthogonal so that the least squares solution is $\underline{\theta}^{LS} = \Psi^H\mathbf{\Upsilon}$. In practice however they will be nearly orthogonal so that a least squares solution by, e.g., QR decomposition or even singular value decomposition, is very fast. Viewed in this way, our method may be considered as a prewhitening of the problem which in theory is perfect, but may deviate slightly for numerical reasons.

Increasing the order of the approximant requires just one more step of the recurrence relation to compute an additional column for Ψ and, in theory, one extra θ-coefficient needs to be computed since the ones already computed do not change. Practically, they may change a little bit when the new least squares problem is solved.

The recursive computation of orthogonal columns has the well-known drawback that after a while orthogonality may be lost. A reorthogonalization may be needed when the approximation order is high. Hence, the extra QR factorization.

For frequency domain identification using the TM basis and cyclically repeated poles, see, e.g., [20]. This basis is however not orthogonal with respect to the discrete inner product of the space in which the norm is taken. The consequences are that evaluating the explicit expressions for the ORF is somewhat cheaper, but because of the ill conditioning of the problem, most of the accuracy may be lost to numerical rounding errors and increasing the order will result in a whole θ-vector to be changed.

However, the asymptotic analysis of the parameter estimate found in [20] can be smoothly carried over to our approach. If we make the following assumptions that are also made in [20]

Assumption 1: The Blaschke condition holds. From Theorem 1, we find that G_0 allows a general Fourier series expansion

$$G_0(z) = \sum_{k=0}^{n}\theta_k^*\phi_k(z) + Z(z)$$

with $Z(z) = \sum_{k=n+1}^{\infty}\theta_k^*\phi_k(z)$. We denote $\underline{\theta}^* = [\theta_0^* \cdots \theta_n^*]^T$.

Assumption 2: The frequency domain data is calculated by a DFT in N frequencies $\omega_k \in \{2\pi l/t, l = 1, \ldots, t\}$ from length t time domain data, which is periodic, having r periods of length t_0, with $t = r \cdot t_0$, $r \ge 1$.

Now, we can prove an analogue of [20, Prop. 5.1].

Theorem 6: Consider Assumptions 1 and 2. Denote $W = UK$. If $n \le N \le t_0 \le \infty$, then

1) $\mathrm{plim}_{t\to\infty}\underline{\theta}^{LS} = \hat{\underline{\theta}}$, where $\hat{\underline{\theta}} = \underline{\theta}^* + \mathbf{\Psi}^H W\mathbf{Z}$;
2) for $t \to \infty$, the stochastic variable $\sqrt{t}(\underline{\theta}^{LS} - \hat{\underline{\theta}})$ tends to a normal distribution $\mathcal{N}(0, P)$, where the asymptotic covariance matrix P is determined by its scalar elements

$$\begin{aligned}P_{ik} &= t_0\sum_{l=1}^{N}\Psi_{il}\Psi_{kl}^*\,\frac{\Phi_v(\omega_l)}{\Phi_u(\omega_l)}\\ &= t_0\sum_{l=1}^{N}\overline{\phi_{i-1}(z_l)}\phi_{k-1}(z_l)|W(z_l)|^4\,\frac{\Phi_v(\omega_l)}{\Phi_u(\omega_l)}\\ &\qquad i, k = 1, \ldots, n+1\end{aligned}$$

where Φ_u and Φ_v denote the spectrum of the input and the noise, respectively.

Additionally

$$\lim_{t\to\infty} t \cdot \text{cov}(\underline{\theta}^{LS}) = P.$$

The proof is similar to the one in [20].

Remark: Similar techniques can be used in the time domain. If the model is $y(t) = \sum_k \phi_k(q)u(t) + d(t)$ where q is the shift operator: $qu(t) = u(t-1)$ and $d(t)$ is zero mean unit white noise, then an estimator is obtained from data in the points $t = t_1, t_2, \ldots, t_N$ by minimizing $E_n(t, \underline{\theta}) = y(t) - \sum_{k=0}^{n} \theta_k \psi_k(t)$ with $\psi_k(t) = \phi_k(q)u(t)$ in least squares norm. Thus, replacing the frequency domain quantities by their time domain equivalent, basically the same problems occur with the same type of solution using ORF. However, since now orthogonality is with respect to summation over the t_k which are all on the real line, we need the three-term recurrence relation for the ORF which generalizes the classical three-term recurrence relation for orthogonal polynomials on the real line. For details, see [17, Ch. 11].

VI. Numerical Example: A Robot Arm

Suppose we are given measured frequency data of the transfer function $\{G(e^{i\omega_k})\}$ and the variances $\{\sigma_k\}$ on these measurements.

We are interested in an approximation $\hat{G}_n$ of G that minimizes the following cost function:

$$\sum_{k=1}^{N} \left| \frac{G(e^{i\omega_k}) - \hat{G}_n(e^{i\omega_k})}{\sigma_k} \right|^2 .$$

This means that, considering (9), we choose the weights $K(z_k) = 1/\sigma_k U(z_k)$. We search for the least squares solution $\hat{G}_n$.

We have experimental data from a robot arm problem. The length of the signal is $N = 100$. The transfer function and the variances are shown in Fig. 1. Note that due to the great uncertainty around the peaks, every approximation will have trouble matching these peaks.

We do not know the exact location of the poles, but the problem has been a test case for several methods which lead to an approximate location of the optimal poles near $-0.0174 \pm 0.9318i$, $0.8768 \pm 0.4763i$, and $0.9733 \pm 0.2265i$.

We use the poles $0.9 \pm 0.3i$ and $0.5 \pm 0.4i$, repeated seven times, to obtain an approximation of degree 28. The condition number of the system matrix in the TM case is $O(10^{16})$ meaning that it is numerically rank deficient, while in our approach the condition number is almost perfect: $1 + O(10^{-10})$.

The relative error $\|Y - \hat{Y}\|/\|Y\|$ is 0.0790 in the TM case and 0.0644 in the ORF case, which is an improvement of more than 20%.

In Fig. 1, we have plotted the transfer function and the approximations. As expected, the two peaks are not very well approximated.

VII. Conclusion

We have given an identification algorithm for the frequency domain. The main advantage of this approach, where we incorporate a weight which depends on the input signal, is that the least squares problem becomes well conditioned and is straightforward. This results in faster algorithms, which are numerically stable.

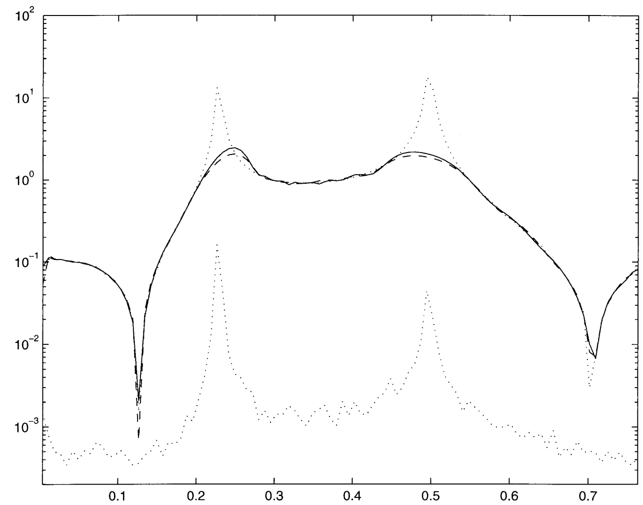

Fig. 1. Plot of the simulated system $\boldsymbol{G}$ (dotted line), the approximating system found by the TM basis (dashed line) and the one found by our approach (solid line). The lowest graph is a rescaled version of the variance on the measured frequency data.

Acknowledgment

The authors would like to thank the referees for valuable suggestions which have improved the presentation of this note considerably.

References

[1] J. L. Walsh, "Interpolation and approximation," in *Amer. Math. Soc. Colloq. Publ.*, vol. 20, Providence, RI, 1960.

[2] S. Takenaka, "On the orthogonal functions and a new formula of interpolation," *Jpn. J. Math.*, vol. 2, pp. 129–145, 1925.

[3] F. Malmquist, "Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de poits," in *Proc. C.R. 6ième Cong. Math. Scand.*, (Kopenhagen 1925), Copenhagen 1926, Gjellerups, pp. 253–259.

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

[5] Y. W. Lee, "Synthesis of electrical networks by means of the Fourier transforms of Laguerre functions," *J. Math. Phys.*, vol. 11, pp. 83–113, 1933.

[6] W. H. Kautz, "Transient synthesis in the time domain," *IRE Trans. Circuits Theory*, vol. 1, pp. 29–39, 1954.

[7] B. Wahlberg and P. M. Mäkilä, "On approximation of stable linear dynamical systems using Laguerre and Kautz functions," *Automatica*, vol. 32, no. 5, pp. 693–708, 1996.

[8] P. S. C. Heuberger, P. M. J. Van den Hof, and O. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 451–465, Mar. 1995.

[9] Z. Szabó, J. Bokor, and F. Schipp, "Identification of rational approximate models in $\boldsymbol{H}_\infty$ using generalized orthogonal basis," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 153–159, Jan. 1999.

[10] P. M. J. Van den Hof, P. S. C. Heuberger, and J. Bokor, "System identification with generalized orthogonal basis functions," *Automatica*, vol. 31, pp. 1821–1834, 1995.

[11] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," in *Proc. IEEE Conf. Decision Control*, Orlando, FL, Dec. 1994, pp. 3388–3393.

[12] B. Ninness and F. Gustafsson, "A unified construction of orthogonal bases for system identification," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 515–522, Apr. 1997.

[13] H. Akçay and B. Ninness, "Rational basis functions for robust identification from frequency and time domain measurements," *Automatica*, vol. 34, no. 9, pp. 1101–1117, 1998.

[14] B. Ninness, "Frequency domain estimation using orthonormal bases," in *Proc. 13th IFAC World Congr.*, vol. 8, San Francisco, CA, 1996, pp. 381–386.

[15] N. F. D. Ward and J. R. Partington, "Robust identification in the disc algebra using rational wavelets and orthonormal basis functions," *Int. J. Control*, vol. 64, pp. 409–423, 1996.
[16] P. Van gucht and A. Bultheel, "State-space realization and orthogonal rational functions," Dept. Computer Science, K. U. Leuven, Leuven, Belgium, Tech. Rep. TW292, Sept. 1999.
[17] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåastad, *Orthogonal Rational Functions*. Cambridge, U.K.: Cambridge Univ. Press, 1999, vol. 5 of Cambridge Monographs on Applied and Computational Mathematics.
[18] P. Van gucht and A. Bultheel, "Computing orthogonal rational functions analytic outside the unit disc," Dept. Computer Science, K.U. Leuven, Leuven, Belgium, Tech. Rep. TW312, Sept. 2000.
[19] ——, "State space representation for arbitrary orthogonal rational functions," *Syst. Control Lett.*, 2003, to be published.
[20] D. K. de Vries and P. van den Hof, "Frequency domain identification with generalized orthonormal basis functions," *IEEE Trans. Automat. Contr.*, vol. 43, pp. 656–669, May 1998.
[21] G. W. Stewart, *Matrix Algorithms. Basic Decompositions*. Philadelphia, PA: SIAM, 1998.

Correction to "Robust D-Stability for Linear Uncertain Discrete Time-Delay Systems"

Te-Jen Su, Yun-Chu Chen, Wen-Jye Shyr, and Boi-Wei Wang

The authors have found an error in [1, Th. 1] and a typographical error in [1, eq. (10)]. The corrections are given as follows.

Let us treat a linear uncertain discrete time-delay system

$$x(k+1)=\sum_{i=0}^{\ell}(A_i+\Delta A_i)x(k-i) \qquad (1)$$

where A_i is an $n\times n$ matrix, ℓ is a positive integer, and ΔA_i is an $n\times n$ perturbed matrix with

$$|A_i|_m \le P_iV_i, \qquad (i=1,2,\ldots,\ell) \qquad (2a)$$

$$|\Delta A_i|_m \le Q_iU_i, \quad (i=0,1,2,\ldots,\ell) \qquad (2b)$$

Manuscript received April 23, 2002; revisd August 15, 2002. Recommended by Associate Editor A. Datta.
T.-J. Su, Y.-C. Chen, and B.-W. Wang are with the Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung 807, Taiwan, R.O.C.
W.-J. Shyr is with the Department of Industrial Education, National Changhua University of Education, Paisa Village, Changhua 500, Taiwan, R.O.C.
Digital Object Identifier 10.1109/TAC.2003.809766

where P_i and Q_i are positive real numbers and V_i and U_i are the nonnegative matrices representing highly structured information on A_i and ΔA_i, respectively.

[1, Th. 2] and [1, Ex. 3] are corrected as follows.

Theorem 2: All the poles of the perturbed system (1) will remain in the circle $D(\alpha,r)$ if all the poles of A_0 lie within a circle $D(\alpha,r)$, and the following inequality is satisfied:

$$r\left[\frac{H\left(G_0(K)\right)\left(\sum_{i=1}^{\ell}P_iV_i+\sum_{i=0}^{\ell}Q_iU_i\right)}{r}\right]<1 \qquad (9)$$

where $H(G_0(K))$ has the same definition in Lemma 3, and $G_0(K)$ is the pulse response sequence matrix of the system

$$G_0(z)=(zI-\overline{A}_0)^{-1}$$

where

$$\overline{A}_0=\frac{(A_0-\alpha I)}{r}.$$

Example 3: Consider the perturbed discrete time-delay system with a single delay of 1 sampling periods in the states. That is

$$x(k+1)=A_0x(k)+A_1x(k-1)+\sum_{i=0}^{1}\Delta A_ix(k-i)$$

where

$$A_0=\begin{bmatrix}0.54 & 0.05\\ 0 & 0.5\end{bmatrix} \quad A_1=\begin{bmatrix}0.02 & 0\\ 0.01 & -0.015\end{bmatrix}.$$

The eigenvalues of A_0 are 0.5 and 0.54. We let $D(\alpha,r) = D(0.3,0.6)$.

By setting $U_0=U_1=V_1=\begin{bmatrix}1&1\\1&1\end{bmatrix}$, we find $P_1=0.02$ and $Q_0+Q_1<0.0916$.

Let

$$\Delta A_0=\begin{bmatrix}0.05 & 0.06\\ 0.06 & 0.04\end{bmatrix} \quad \Delta A_1=\begin{bmatrix}0.02 & 0.02\\ 0.03 & 0.05\end{bmatrix}.$$

The eigenvalues of the system are 0.4926, 0.7432, -0.0189, and -0.0869; all poles lie within the circle $D(0.3,0.6)$.

[1, eq. (10)] is corrected as follows:

$$\frac{\left(\sum_{i=1}^{l}\|A_i\|+\sum_{i=0}^{l}\|\Delta A_i\|\right)}{r}<\rho\left[\frac{(A_0-\alpha I)}{r}\right]. \qquad (10)$$

References

[1] T. J. Su and W. J. Shyr, "Robust D-stability for linear uncertain discrete time-delay systems," *IEEE Trans. Automat. Contr.*, vol. 39, pp. 425–439, Feb. 1994.