A DATA DRIVEN ORTHONORMAL PARAMETERIZATION OF THE GENERALIZED ENTROPY MAXIMIZATION PROBLEM

ANDERS BLOMQVIST† AND BO WAHLBERG‡

Abstract. During the past years Lindquist and coworkers have formulated and studied the so called generalized maximum entropy problem. It generalizes the maximum entropy problem and thus includes problem like the Carathéodory extension problem and the Nevanlinna-Pick interpolation problem. A central ingredient in the theory is the dual (in mathematical programming sense) convex optimization problem. This paper studies a re-parameterization with orthonormal basis functions of that optimization problem. The basis functions can be chosen independent of the original problem; for instance they can be chosen to be data driven. In numerical examples it is shown that a sensible choice of basis functions can improve the numerical conditioning of the optimization problem.

1. Introduction

First we formulate the optimization problem and introduce some notation. The dual of the generalized maximum entropy problem in [6, 5] is, for $\boldsymbol{w} := \begin{bmatrix} w_0 & w_1 & \dots & w_n \end{bmatrix}^T$, the convex optimization problem

$$(P) \qquad \max_{\boldsymbol{q} \in \mathfrak{Q}^{(A,B)}} \mathbb{J}_\Psi(\boldsymbol{q}),$$

where

$$\mathbb{J}_\Psi(\boldsymbol{q}) := \boldsymbol{w}^T \boldsymbol{q} - \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi(e^{i\theta}) \log Q(e^{i\theta}) d\theta, \tag{1.1}$$

and Ψ is a coercive spectral density. Here Q is a trigonometric expansion of degree n for a set of basis functions given by the pair (A, B), $\boldsymbol{q}$ are the corresponding scalar coefficients and $\mathfrak{Q}^{(A,B)}$ is the space of feasible $\boldsymbol{q}$. More precisely, let $Q(z, z^{-1})$ be

$$\begin{aligned} Q &= G^* \Lambda G, \\ &= \sum_{k=0}^{n} \frac{q_k}{2} G_k + \frac{\overline{q_k}}{2} G_k^*, \\ &= \frac{\rho \rho^*}{\tau \tau^*}, \\ &= \frac{\boldsymbol{q}^T}{2} G + \frac{\boldsymbol{q}^*}{2} (G^*)^T, \end{aligned}$$

where $G(z) = (I - Az)^{-1} B$ is a vector representation of the basis functions with which the underlying interpolation conditions are given. Clearly, $\tau(z) = \det(I - Az)$. Here (A, B) is a controllable pair and all eigenvalues of A lie in the open unit disc. Here we will also make the simplifying assumption that $\det A = 0$. This corresponds to having one of the basis functions as a constant. This also makes τ a stable (Schur) polynomial of degree at most n. If desired, we can take the pair (A, B) so that basis functions G_k are orthonormal. The Hermitian matrix Λ is the dual variable in the Lagrangian relaxation of the spectral

† Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden.

‡ Department of Signal, Sensors and Systems, Royal Institute of Technology, SE-100 44 Stockholm, Sweden.

Kullback-Leibler approximation problem in [10]. The domain of feasible $\boldsymbol{q}$ is defined as, for given (A, B) and Q as above,

$$\mathfrak{Q}^{(A,B)} := \{\boldsymbol{q} \in \mathbb{C}^{n+1} : Q(z, z^{-1}) > 0, \forall z \in \{z \in \mathbb{C} : |z| = 1\}\}.$$

Despite being convex, (P) has shown to be relatively hard to solve due to numerical problems. This can be attributed to the logarithmic barrier term. The function value is finite but the gradient becomes unbounded when the solution tends towards the boundary of $\mathfrak{Q}^{(A,B)}$. Therefore, formulations in terms of the spectral factor has been proposed [9, 11, 2]. Those formulations are non-convex and need continuation methods to get convergence.

Motivated by these numerical issue this paper provides a framework for re-parameterizing the optimization problem by changing basis functions as well as basis function poles. In particular it is proposed to take these orthonormal, since they have proven to have advantageous numerical properties. Numerical examples indicate that a sensible choice of basis functions can improve the numerical conditioning.

2. Orthonormal basis functions

Finite Impulse Response (FIR) models are among the most basic tools in the area of signals and systems. The connections to orthogonal polynomials are well known. However, Infinite Impulse Response (IIR) models provide often much more compact descriptions and give improved performance in many cases. The corresponding orthogonal transfer function are related to orthogonal rational functions, which can be viewed as generalizations of orthogonal polynomials. The recent monograph [4] gives an excellent overview of the field of orthogonal rational functions. In parallel to the efforts in applied mathematics a very similar theory has been developed in the fields of control, signals and systems. See [3] for a thorough presentation of this work.

Given a set of possible complex poles $\{\xi_k,\ |\xi_k| < 1\}$, the so-called Takenaka-Malmquist functions

$$G_k(z) = \frac{\sqrt{1-|\xi_k|^2}}{z-\xi_k} \prod_{i=1}^{k-1} \left[\frac{1-\xi_i^* z}{z-\xi_i}\right],\ k = 1, 2, \dots \quad |\xi_k| < 1,$$

are orthonormal, i.e.,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} G_i(e^{i\theta})\overline{G_j(e^{-i\theta})}d\theta = \begin{cases} 1, & i = j, \\ 0 & \text{otherwise,} \end{cases}$$

A most useful special case is to repeat one fixed real pole $\xi_k = a,\ -1 < a < 1, \forall k$. This corresponds to the Laguerre functions.

As shown in [13] the use of rational orthogonal functions makes it possible to adapt the time constants of the functions while keeping most favorable properties of finite impulse response models. For example, if the poles of the system to be approximated are close to the unit circle, the corresponding basis functions should also have similar poles.

Gathering G_k in a vector, $G = \begin{bmatrix} G_0 & G_1 & \dots & G_n \end{bmatrix}^T$ we have a set of basis functions of the type discussed in the previous section. Corresponding state-space matrices (A, B), of course non-unique, can be determined as described in [3]. One way to derive a state-space representation of rational orthogonal functions is to study orthogonal state-space realizations of the all-pass transfer function

$$H(z) = \prod_{i=1}^{n} \frac{1-\xi_i^* z}{z-\xi_i},$$

Such realizations will satisfy

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^T = I,$$

and the corresponding input-to-state transfer functions $G(z) = (I - Az)^{-1}B$ will be orthonormal.

3. Orthonormal parameterization of the optimization problem

Next we will propose a change of basis functions from the ones associated with the interpolation data to new basis functions which just will be used as a computational tool. More precisely, suppose we choose a new controllable pair $(\tilde{A}, \tilde{B})$ with all eigenvalues of $\tilde{A}$ in the open unit disc. The new basis functions are the elements of $\tilde{G}(z) = (I - \tilde{A}z)^{-1}\tilde{B}$ and can be taken to be orthonormal. Define the new trigonometric expansion

$$\tilde{Q}(z, z^{-1}) := Q(z, z^{-1}) \frac{\tau\tau^*}{\tilde{\tau}\tilde{\tau}^*},$$

where naturally $\tilde{\tau} := \det(I - \tilde{A}z)$. We have the following proposition.

Proposition 3.1. *Let (A, B) and $(\tilde{A}, \tilde{B})$ be given as above. Then, for $\boldsymbol{q} \in \mathfrak{Q}^{(A,B)}$ and $\tilde{\boldsymbol{q}} \in \mathfrak{Q}^{(\tilde{A},\tilde{B})}$,*

$$\tilde{\boldsymbol{q}} = T\boldsymbol{q},$$

where T is the invertible matrix

$$\begin{aligned} T &:= \Gamma_{(\tilde{A},\tilde{B})}^{-T} L_{\tilde{\boldsymbol{\tau}}}^{-1} S(\tilde{\boldsymbol{\tau}})^{-1} S(\boldsymbol{\tau}) L_{\boldsymbol{\tau}}^T \Gamma_{(A,B)}^T, \\ r(z) &:= r_0 + r_1 z + \cdots + r_n z^n, \\ L_{\boldsymbol{r}} &:= \begin{bmatrix} r_0 & & & \\ r_1 & r_0 & & \\ \vdots & & \ddots & \\ r_n & r_{n-1} & \cdots & r_0 \end{bmatrix}, \\ S(\boldsymbol{r}) &:= \begin{bmatrix} r_0 & \cdots & r_{n-1} & r_n \\ r_1 & & r_n & \\ \vdots & \iddots & & \\ r_n & & & \end{bmatrix} + \begin{bmatrix} r_0 & r_1 & \cdots & r_n \\ & r_0 & & r_{n-1} \\ & & \ddots & \vdots \\ & & & r_0 \end{bmatrix}, \end{aligned}$$

and $\Gamma_{(A,B)}$ is the controllability matrix of the pair (A, B).

Proof. For each $Q \in \mathfrak{Q}_+^{(A,B)}$ we have

$$Q(z, z^{-1}) = \frac{\boldsymbol{q}^T}{2} G + \frac{\boldsymbol{q}^*}{2} (G^*)^T = \frac{b}{\tau} + \frac{b^*}{\tau^*} = \frac{b\tau^* + b^*\tau}{\tau\tau^*} = \frac{d}{\tau\tau^*},$$

Note that both τ and $\tilde{\tau}$ are polynomials of degree at most n due to the assumption that $\det A = \det \tilde{A} = 0$. The coefficients of the pseudo-polynomial d are given by $S(\boldsymbol{\tau})\boldsymbol{b}$ where $\boldsymbol{b}$ is a vector with the coefficients of b. Noting that

$$a_0 + a_1 z + \cdots = \frac{b}{\tau} = \frac{\boldsymbol{q}^T}{2} G = \frac{\boldsymbol{q}^T}{2} (B + ABz + \cdots),$$

we have that $\boldsymbol{a} = \Gamma_{(A,B)}^T \boldsymbol{q}/2$. Comparing coefficients we also have that $\boldsymbol{b} = L_{\boldsymbol{\tau}} \boldsymbol{a}$. Combining the expressions, the coefficients of d is given by

$$S(\boldsymbol{\tau}) L_{\boldsymbol{\tau}} \Gamma_{(A,B)}^T \frac{\boldsymbol{q}}{2},$$

Now, since $\tilde{Q} = d/(\tilde{\tau}\tilde{\tau}^*)$ we have that

$$S(\boldsymbol{\tau})L_{\boldsymbol{\tau}}\Gamma_{(A,B)}^T\boldsymbol{q} = S(\tilde{\tau})L_{\tilde{\tau}}\Gamma_{(\tilde{A},\tilde{B})}^T\frac{\tilde{\boldsymbol{q}}}{2}.$$

All the included matrices are invertible: S by [7], L since A has all its eigenvalues in the open disc, and $\Gamma_{(A,B)}$ due to the controllability assumption. Thus, $\tilde{\boldsymbol{q}} = T\boldsymbol{q}$ with T invertible. □

With the new coordinates the functional in (1.1) becomes

$$\begin{aligned}\mathbb{J}_\Psi(\boldsymbol{q}) &= \boldsymbol{w}^T\boldsymbol{q} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{i\theta})\log Q(e^{i\theta})d\theta,\\ &= \underbrace{\boldsymbol{w}^TT\boldsymbol{q} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{i\theta})\log\tilde{Q}(e^{i\theta})d\theta}_{=:J_\Psi(\tilde{\boldsymbol{q}})} + \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{i\theta})\log\frac{\tau(e^{i\theta})\overline{\tau(e^{-i\theta})}}{\tilde{\tau}(e^{i\theta})\overline{\tilde{\tau}(e^{-i\theta})}}d\theta.\end{aligned}$$

We get the optimization problem

$$(\tilde{P}) \qquad \max_{\tilde{\boldsymbol{q}}\in\mathfrak{Q}^{(\tilde{A},\tilde{B})}} J_\Psi(\tilde{\boldsymbol{q}}).$$

The new problem is of the same form as the old but, depending on the choice of $(\tilde{A},\tilde{B})$, it may have better numerical properties close to the unique optimum. To analyze this we can study the condition number of the Hessian matrix at the optimum. This makes sense since the the Hessian is a continuous function and its ill-conditioning leads to numerical problems in Newton-type algorithms. Next we will study a first order example.

Example 1 (ARMA(1,1))**.** An ARMA(1,1) model is determined by its pole $-1 < a < 1$ and its zero $-1 < b < 1$. We will compare the condition number of the Hessian at the optimal point, i.e., the point corresponding to the true ARMA model, for two sets of basis functions. The default basis functions and proposed basis function are taken as

$$G = \begin{bmatrix}1\\ z^{-1}\end{bmatrix}, \quad \tilde{G} = \begin{bmatrix}1\\ \frac{\sqrt{1-p^2}}{z-p}\end{bmatrix},$$

where $-1 < p < 1$ is the pole of the proposed orthonormal basis function. Note that also the default basis functions are orthonormal.

First we take $a = 0.495$ and $b = 0$ which corresponds to an AR(1) model. The condition numbers are for different choices of the basis function polynomial p are plotted in Figure 3.1. We note that the proposed basis functions give a smaller condition number when the p is in the neighborhood of the model's pole but significantly higher in other regions.

Next we plot the ratio of the condition numbers for default and proposed and for different choices of true model's pole in Figure 3.2. A value *larger* than one means that the proposed parameterization leads to a smaller condition number. We see that the closer to the unit circle the true pole is the more we can improve the condition number by choosing an appropriate basis function pole.

Finally, we let the zero, b, vary as well. In Figure 3.3 the ratio of the default and proposed condition numbers, but now letting the model's zero be 0.99. We note a similar pattern: we can improve the numerical conditioning, especially when the true model's pole is close to the unit circle, by choosing the basis function pole correspondingly. However, the improvement decreases drastically when the zero is close to the unit disc and in the same frequency region as the model's pole.

As seen in the example the best choice of basis function poles depends on the true model. One idea then would be to choose the basis function poles depending on the current

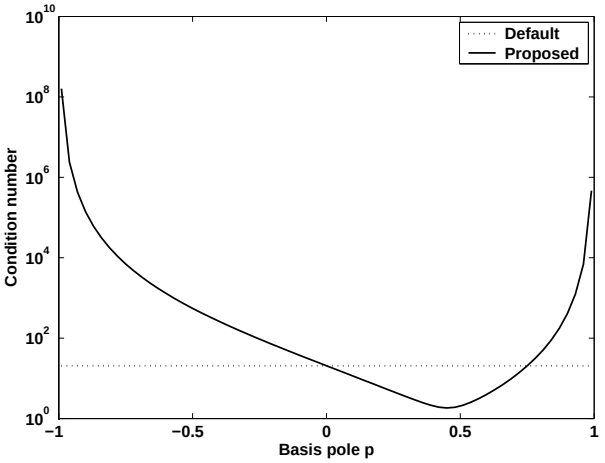

FIGURE 3.1. The condition numbers for AR(1) processes as a function of the basis function's pole location.

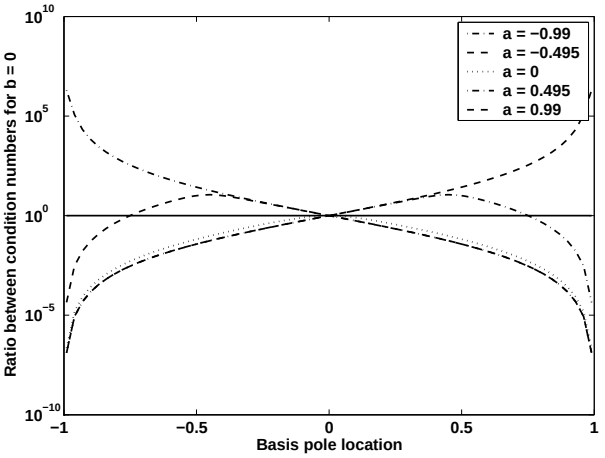

FIGURE 3.2. The ratio of the default and proposed condition numbers for AR(1) processes as a function of the basis function's pole location.

iteration point, which converges to the best estimate. That would yield a *data driven* parameterization.

4. Orthonormal Parameterization of The Positive Real Equations

An alternative approach for computing the unique minimizer of the generalized entropy maximization problem was proposed in [1]. Relying on the uniqueness of the solution it was recognized that one instead can solve what we might call the *positive real equations*. Given a spectral density $\Phi = \rho\rho^*/(\alpha\alpha^*) = f + f^*$ where f is the positive real function β/α we have the positive real equations as

$$\alpha\beta^* + \beta\alpha^* = \rho\rho^* = \delta,$$

where δ is the so called dissipation polynomial. A key observation is that any second moment conditions on the density, that is covariance-type conditions, can be written as a *linear* relation between the coefficients of α and β. In fact, letting bold letter denote the

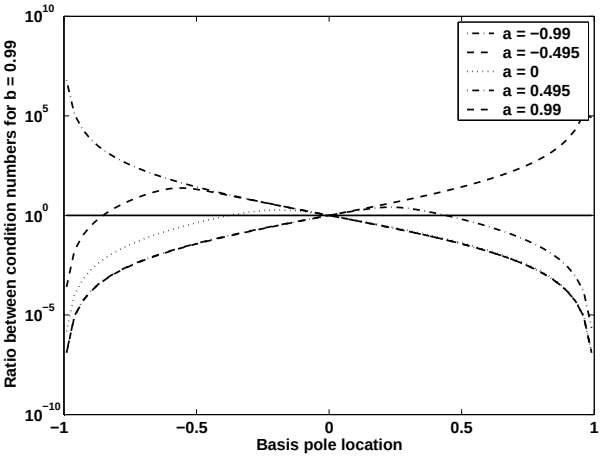

FIGURE 3.3. The ratio of the default and proposed condition numbers for ARMA(1,1) processes with different pole locations as a function of the basis function's pole location.

corresponding coefficient vectors, we have

$$S(\boldsymbol{\alpha})K\boldsymbol{\alpha} = \boldsymbol{\delta}, \tag{4.1}$$

where K is the constant matrix relating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and $S(\cdot)$ is the operator defined in Proposition 3.1. These equations have the advantage that they continuously extend to the boundary of the feasible domain (where the gradient of the optimization problem grows infinite). A homotopy in terms of the dissipation polynomial has proven to be computationally efficient and numerically robust for solving the nonlinear equations.

What we will propose here is the possibility of putting the equations into a more general framework allowing for other orthonormal basis functions than the currently used $(\{z^k\}_{k=0}^n)$. More generally, let a, b, and r be rational functions:

$$a = \frac{\alpha}{\tau} = \boldsymbol{a}^T G,\ b = \frac{\beta}{\tau} = \boldsymbol{b}^T G,\ \text{and } r = \frac{\rho}{\tau} = \boldsymbol{r}^T G,$$

where G are taken as above. We get an orthonormal parameterization whenever the basis functions $G(z)$ are taken orthonormal. Again we can pick the poles of G to incorporate information about the pole location of α and β.

To extract the coefficient equations we can integrate with the basis functions. These equations can be simplified significantly due to the orthonormality by applying the results of [12, 8].

Remark 4.1. Another approach is to transform the standard equations in (4.1) for the change of bases. The basis for the default basis functions is given by

$$A = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Given a new, desired orthonormal basis $(\tilde{A}, \tilde{B})$, by Proposition 3.1 we have that

$$\boldsymbol{\delta} = T\boldsymbol{d}.$$

The relation between $\boldsymbol{\alpha}$ and $\boldsymbol{a}$ is given by

$$\boldsymbol{\alpha} = L_{\boldsymbol{\tau}} \Gamma_{(\tilde{A},\tilde{B})}^T \boldsymbol{a}.$$

Therefore the corresponding equations for an orthonormal basis is given by

$$S(L_{\boldsymbol{\tau}}\Gamma^T_{(\tilde{A},\tilde{B})}\boldsymbol{a})KL_{\boldsymbol{\tau}}\Gamma^T_{(\tilde{A},\tilde{B})}\boldsymbol{a} = T\boldsymbol{d}.$$

5. Conclusions

A framework for re-parameterization of the generalized maximum entropy problem using a general basis function representation has been proposed. The poles of the new basis functions can be chosen arbitrarily in the stable region. In particular, these basis functions can be taken orthonormal. In small-scale numerical examples it has been shown that the conditioning of the Hessian is affected in a advantageous way for a sensitive choice of basis functions.

References

[1] A. Blomqvist, G. Fanizza, and R. Nagamune. Computation of bounded degree Nevanlinna-Pick interpolants by solving nonlinear equations. pages 4511–4516, 2003. The proceedings of the 42nd IEEE Conference on Decision and Control.

[2] A. Blomqvist and R. Nagamune. Optimization-based computation of analytic interpolants of bounded degree. Preprint, 2004.

[3] J. Bokor, P. Heuberger, B. Ninness, T. Oliveira e Silva, P. Van den Hof, and B. Wahlberg. Modelling and identification with orthogonal basis functions. In *Workshop Notes, 14:th IFAC World Congress, Workshop nr 6*, Beijing, PRC, 7 1999.

[4] A. Bultheel, P. González-Vera, and O. Njåstad. *Orthogonal Rational Functions*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 1999.

[5] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[6] C. I. Byrnes, S. V. Gusev, and A. Lindquist. A convex optimization approach to the rational covariance extension problem. *SIAM J. Contr. and Optimiz.*, 37(1):211–229, 1998.

[7] C. I. Byrnes, A. Lindquist, and Y. Zhou. On the nonlinear dynamics of fast filtering algorithms. *SIAM J. Contr. and Optimiz.*, 32(3):744–789, 1994.

[8] T.J. de Hoog, Z. Szabó, P.S.C. Heuberger, P.M.J. Van den Hof, and J. Bokor. Minimal partial realization from generalized orthonormal basis function expansions. *Automatica*, 38(4):655–669, 2002.

[9] P. Enqvist. A homotopy approach to rational covariance extension with degree constraint. *Int. J. Applied Mathematics and Computer Science*, 11(5):1173–1201, 2001.

[10] T. T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Information Theory*, 49(11):2910–2917, November 2003.

[11] R. Nagamune. A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 48(1):113–117, 2003.

[12] Z. Szabó, P. S. C. Heuberger, J. Bokor, and P. M. J. Van den Hof. Extended Ho-Kalman algorithm for systems represented in generalized orthonormal bases. *Automatica J. IFAC*, 36(12):1809–1818, 2000.

[13] B. Wahlberg. Orthonormal basis functions models: A transformation analysis. *SIAM Review*, pages 689 – 705, 11 2003.