SIMULTANEOUS CEPSTRAL AND COVARIANCE MATCHING FOR ARMA ESTIMATION WITH WHITENING*

ANDERS BLOMQVIST†

Abstract. Simultaneous cepstral and covariance matching provides a paradigm for ARMA estimation with attractive features: the estimates are unique and depend smoothly on the time series. In fact, it corresponds to a well-posed mathematical problem in the sense of Hadamard. A major drawback, however, is that simulations show that the estimates are not asymptotically efficient. Here we shall present a development of the paradigm based on whitening. We show that the uniqueness and smoothness are preserved while the statistically properties are potentially improved. A simulation study indicates that the estimates are asymptotically efficient.

Key words. cepstrum, well-posedness, convex program, ARMA estimation, Cramér-Rao bound

AMS subject classifications. 93B29, 30E05, 93E12, 90C25, 94A17

1. Introduction. Cepstral coefficients were introduced in [6] and have since mainly been used in signal processing, and speech processing in particular, for an alternative representation of AR models. Recently, Byrnes, Enqvist, and Lindquist in [10] proved the remarkable result that the combined window of covariances and cepstral coefficients provides a global parameterization of ARMA models. Moreover, they showed that the problem of going from these coefficients to the ARMA parameters is well-posed in the sense of Hadamard. Since both covariances and cepstral coefficients are directly computable from times series data this provided a well-posed approach to ARMA modeling which is in sharp contrast to the, for low-variance estimation, predominant maximum-likelihood methods. At that point the variance of the estimates was not studied.

As a prelude we shall consider a trivial example illustrating the Cepstral Covariance Matching (CCM) of [10, 15] while also revealing a shortcoming of the method.

Example 1.1 (ARMA estimation). Consider Figure 1.1. We assume, for the time being, that the measured scalar data $\{x_t\}_{t=1}^N$ is generated by feeding white noise, say Gaussian, with variance λ^2 through a stable, causal, minimum-phase linear filter of some known degree. That implies that we should model the normalized transfer function of the shaping filter as:

$$w(z) = \frac{\sigma(z)}{a(z)},$$

where a and σ are monic stable polynomials of some degree. Given the measurement we want to determine the best possible model of the filter according to some criterion.

Let us consider the very simple example $a(z) \equiv 1$, $\sigma(z) = 1 - \sigma_1^{-1} z$ for $-1 < \sigma_1 < 1$, and $\lambda = 1$. This corresponds to a Moving Average process of order one, MA(1).

In statistics the Maximum Likelihood (ML) is probably the most widely used estimator, see for instance [8]. It is the best possible estimator with respect to the statistical criteria. The counterpart in the engineering literature is the Prediction Error Method (PEM), which minimizes the prediction error and is widely used for off-line estimation, [25]. The PEM and the ML methods are equivalent when the driving

*This work has previously appeared in the doctoral thesis [2].

†Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, (anders.blomqvist@math.kth.se).

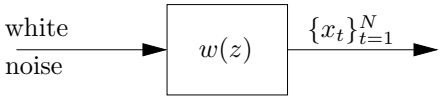

FIG. 1.1. *The shaping filter producing an ARMA process*

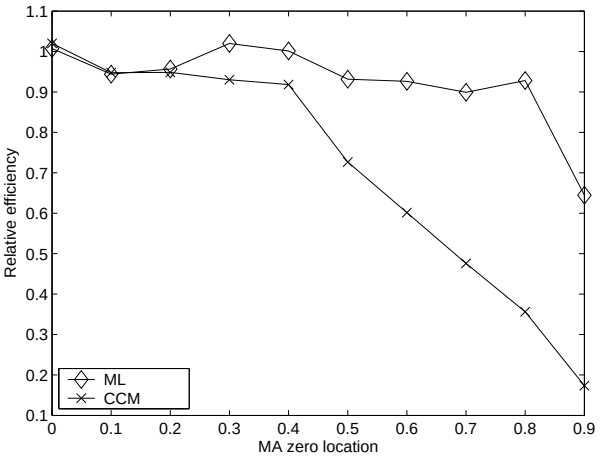

FIG. 1.2. *The estimated relative efficiency of the estimated zero of an MA(1) model using the ML and CCM estimators.*

white noise is Gaussian. These estimators are based on nonconvex optimization and therefore typically computationally demanding. Also they need to treat failure modes caused by the nonconvexity.

Now, for $0 \leq \sigma_1 \leq 0.9$ we generate a long data set, say $N = 1000$, and apply both the ML estimator `armax` in [24] and the CCM method of [15] with biased sample covariances and cepstral estimates based on a long AR model (length $L = 20$). The Cramér-Rao bound is known to be $N^{-1}(1-\sigma_1^2)$, see for instance [28, Chapter 5.2]. In Figure 1.2, the estimated *relative efficiency*, which is the ratio between the Cramér-Rao bound and the estimated variances, is plotted for each method based on a Monte Carlo simulation with 1000 realizations. The ML estimator is approximately efficient, that is, it has approximately relative efficiency 1, as expected from the theory. The CCM estimator seems to be efficient for a zero location close to origin but not otherwise, in that the relative efficiency is significantly less than one.

The main contribution of the paper is to generalize the CCM method to allow for frequency weighting/prefiltering. Thereby we can lower to the variance of the estimtes; in fact, a simulation example indicates that this might enable an asymptotically efficient estimator while maintaining the well-posedness.

The paper is organized as follows: in Section 2 we introduce the notation and define some smooth manifolds. In Section 3 we generalize the uniqueness result to our setting and in Section 4 we prove the smoothness of the parameterization. In Section 5 we construct an algorithm for solving the generalized problem which we in Section 6 apply to an example. Finally we give some concluding remarks.

2. Notation. In this section we shall give the main notation and define some manifold that will be used later. Denote the unit circle by $\mathbb{T}$. Let $\mathcal{C}$ be the set of not necessarily positive, continuous, real-valued functions on $\mathbb{T}$ and $\mathcal{C}_+$ the subset of

positive functions. We will consider functions in the usual $\mathcal{L}_2$ Hilbert space with the inner-product

$$\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(e^{-i\theta})}g(e^{i\theta})d\theta.$$

Now, restrict the consideration to finite degree real rational functions:

$$w(z) = \lambda\frac{\sigma(z)}{a(z)} = \lambda\frac{z^m + \sigma_1 z^{m-1} + \dots \sigma_m}{z^n + a_1 z^{n-1} + \cdots + a_n}. \tag{2.1}$$

In particular we will be interested in rational functions with all poles and zeros outside the unit circle. Let the *Schur region* $\mathcal{S}_n$ be the n-dimensional smooth manifold of monic polynomials with all roots outside the unit disc. For simplicity of notation we will identify this function space with the space of coefficients:

$$\mathcal{S}_n = \left\{a \in \mathbb{R}^n : z^n + a_1 z^{n-1} + \cdots + a_n \neq 0 \ \forall z \in \overline{\mathbb{D}}\right\}.$$

Normalized outer rational functions, that is with $\lambda = 1$, then belong to the direct product of two Schur regions

$$\mathcal{P}_{nm} := \mathcal{S}_n \times \mathcal{S}_m.$$

If the polynomials are of the same degree we simply write $\mathcal{P}_n$. Also, define the dense subset $\mathcal{P}^*_{nm}$ consisting of all coprime rational functions in $\mathcal{P}_{nm}$. The topology of $\mathcal{P}^*_n$ is fairly complicated. Firstly, note that the Schur region $\mathcal{S}_n$ in general is nonconvex. Secondly, the coprimeness assumption divides the space into $n+1$ connected components, see [7, 30]. Also, introduce the function space $\mathcal{L}_n$ consisting of all not necessarily monic polynomials of degree at most n.

The spectral density corresponding to the rational spectral factor $w(z)$ can be written

$$\Phi(z) = w(z)w^*(z) = \lambda^2\frac{\sigma(z)\sigma^*(z)}{a(z)a^*(z)} = \frac{P(z)}{Q(z)}, \tag{2.2}$$

where P and Q are *pseudo-polynomials* of degrees m and n defined as

$$P(z) := 1 + p_1/2(z + z^{-1}) + \cdots + p_m/2(z^m + z^{-m}),$$
$$Q(z) := q_0 + q_1/2(z + z^{-1}) + \cdots + q_n/2(z^n + z^{-n}).$$

We can generalize the pseudo-polynomials to be represented in some other function space:

$$\frac{P(z)}{Q(z)} = \frac{P(z)}{\tau(z)\tau^*(z)}\Big/\frac{Q(z)}{\tau(z)\tau^*(z)},$$

where we define

$$\tau(z) := \det(I - Az) = \tau_0 + \tau_1 z + \cdots + \tau_{n+1}z^{n+1}. \tag{2.3}$$

By taking $A = 0$ we recover the pseudo-polynomials. Also note, that if $\det A = 0$, $\tau_{n+1} = 0$ so $\tau(z)$ is of degree at most n. The generalized pseudo-polynomial can now be written

$$Q(z) = \sum_{k=0}^{n}\frac{q_k}{2}(G_k(z) + G_k^*(z)) = \frac{a(z)a^*(z)}{\tau(z)\tau^*(z)}, \tag{2.4}$$

for some basis functions $G_k(z)$ spanning the appropriate subspace and where $a(z)$ is a polynomial of degree n. We identify the space of pseudo-polynomials that are positive on the unit circle with the space of coefficients, given some basis functions:

$$\mathcal{Q}_+ = \left\{(q_0, q_1, \ldots, q_n) \in \mathbb{R}^{n+1} : Q(z) > 0, z \in \mathbb{T}\right\}.$$

We also define the subset for which the leading coefficient $q_0 = 1$ as $\mathcal{Q}_+^0$.

To represent transfer functions and their corresponding spectral densities we will use basis functions. A suitable framework for this, which can be interpreted as filter-banks, is given in [18, 19]. Let $A \in \mathbb{C}^{n+1\times n+1}$ and $B \in \mathbb{C}^{n+1}$. The pair (A, B) is called *reachable* if the reachability matrix

$$\Gamma := \begin{bmatrix} B & AB & \ldots & A^n B \end{bmatrix},$$

has full rank. If, in addition, A have all its eigenvalues in $\mathbb{D}$, we define the basis functions

$$\begin{bmatrix} G_0 \\ G_1 \\ \vdots \\ G_n \end{bmatrix} := G(z) := (I - Az)^{-1}B = B + Az(Iz - A)^{-1}B,$$

In particular, if $\det A = 0$ one basis function will be a constant. Clearly, the basis functions G_k will be analytic in $\mathbb{D}$. Define a set of basis functions as

$$\mathcal{G} := \left\{ G(z) = (I - Az)^{-1}B : \begin{array}{l} A \in \mathbb{C}^{n+1\times n+1},\ B \in \mathbb{C}^{n+1}, \\ \text{eig}(A) \subset \mathbb{D},\ (A, B) \text{ reachable},\ G_0(z) \equiv 1, \\ \langle G_0, G_k\rangle = \delta_{0k}, k = 1, \ldots, n+1, \end{array} \right\}. \quad (2.5)$$

For such basis function we define $\bar{G}$ by $G =: \begin{bmatrix} 1 & \bar{G}^T \end{bmatrix}^T$. Several classes of basis functions can be recovered by suitable choices of A and B, see for instance [18, 19, 2]. In particular we shall call $G_k(z) = z^k$ the *standard basis*.

Given some spectral density $\Psi \in \mathcal{C}_+$ and some basis functions $G \in \mathcal{G}$ we define

$$r_k := \frac{1}{2\pi}\int_{-\pi}^{\pi} G_k(e^{i\theta})\Psi(e^{i\theta})\Phi(e^{i\theta})d\theta = \langle G_k, \Psi\Phi\rangle, \quad (2.6)$$

for the spectral density Φ. Note that we can think of Ψ as the density of a prefilter that is applied to the signal. For the special case with $\Psi \equiv 1$ and G as the standard basis the components r_k will be Fourier coefficients of the spectral density. In the setting of stochastic processes, these are exactly the covariances of the processes. For the special case of $\Psi \equiv 1$ and G such that $G_k(z) = 1/(z - z_k)$, the components r_k are interpolation values on the positive-real part of the spectral density in the poles of the basis: $f(z_k) = r_k$. We will call (2.6) *generalized prefiltered covariances*. The corresponding Pick matrix is given by

$$\Sigma = \frac{1}{2\pi}\int_{-\pi}^{\pi} G(e^{i\theta})\Psi(e^{i\theta})\Phi(e^{i\theta})G^*(e^{i\theta})d\theta,$$

and is related to the interpolation data matrix W via $\Sigma = WE + EW^*$, where E is the controllability Gramian of (A, B), see for instance [18, 19]. The covariance vector is

then given by $r = WB$. Likewise, given the reachable pair (A, B) and the covariance vector r there is a unique Pick matrix $\Sigma(r)$. In fact, using the representation

$$W = w_0 I + w_1 A + \cdots + w_n A^n,$$

the coefficients w_j are given by $w = \Gamma^{-1} r$. Hence we can determine W and then compute as $\Sigma = WE + EW^*$. We define the set of feasible generalized prefiltered covariances as

$$\mathcal{R}_n := \left\{ r \in \mathbb{C}^{n+1} : \Sigma(r) > 0 \right\}.$$

In this thesis we will also study moments of the logarithm of the spectral density defined as

$$c_k := \frac{1}{2\pi} \int_{-\pi}^{\pi} G_k(e^{i\theta}) \Psi(e^{i\theta}) \log \Big(\Psi(e^{i\theta}) \Phi(e^{i\theta}) \Big) d\theta = \langle G_k, \Psi \log (\Psi \Phi) \rangle . \tag{2.7}$$

For the special case with $\Psi \equiv 1$ and G as the standard basis, the components c_k will be Fourier coefficients of the logarithm of the spectral density. In signal processing and speech processing, in particular, these are called *cepstral coefficients*, see for instance [6, 27]. In signal processing, the cepstral coefficients have traditionally been considered as an alternative to covariances for parameterizing AR models. The basis G generalize the notion of cepstrum together with the prefiltering that Ψ represents. Therefore, we will call (2.7) the *generalized prefiltered cepstral coefficients*.

As we are interested in the Nevanlinna-Pick problem with degree constraint, it is instrumental to define the set of covariances and cepstral coefficients that corresponds to a rational density of degree n. We will slightly generalize the definitions in [10, 23], which do this in an implicit fashion. Let $\Psi \in \mathcal{C}_+$ and $G, H \in \mathcal{G}$ be given. Define

$$\mathcal{X}_{nm} := \left\{ (r, c) \in \mathbb{C}^{n+1} \times \mathbb{C}^n : \begin{array}{l} r \in \mathcal{R},\ \lambda \in \mathbb{R}_+,\ \sigma \in \mathcal{S}_m,\ a \in \mathcal{S}_n, \\ r_k = \left\langle H_k, \Psi \lambda^2 \dfrac{\sigma \sigma^*}{a a^*} \dfrac{\tau \tau^*}{t t^*} \right\rangle, k = 0, 1, \ldots, n \\ c_k = \left\langle G_k, \Psi \log \Psi \lambda^2 \dfrac{\sigma \sigma^*}{a a^*} \dfrac{\tau \tau^*}{t t^*} \right\rangle, k = 1, 2, \ldots, m \end{array} \right\}, \tag{2.8}$$

where $\tau = \det(I - A_H z)$ and $t = \det(I - A_G z)$. In many situations we will have $m = n$ and $G = H$; then we will denote the set $\mathcal{X}_n$.

REMARK 2.1. *The definition of $\mathcal{X}_n$ is implicit, making it as hard to check whether an element belongs to it, as to solve for the interpolating function. However, it will be of great theoretical value to define the set this way. For actual computation of an interpolant, there are ways to circumvent this difficulty as discussed and shown in the following chapters.*

Since spectral densities can be interpreted as distribution functions in the spectral domain we will adopt a discrepancy from statistics called *spectral Kullback-Leibler discrepancy* [22]:

DEFINITION 2.2 (Spectral Kullback-Leibler discrepancy). *Given two spectral densities $\Psi, \Phi \in \mathcal{C}_+$ with common zeroth moment, $\langle 1, \Psi \rangle = \langle 1, \Phi \rangle$, the spectral Kullback-Leibler discrepancy is given by*

$$\mathbb{S}(\Psi, \Phi) := \left\langle \Psi, \log \frac{\Psi}{\Phi} \right\rangle .$$

It is not symmetric in its arguments but jointly convex. It fulfills $\mathbb{S}(\Psi, \Phi) \geq 0$ with equality if and only if $\Psi = \Phi$, see for instance [20]. The spectral Kullback-Leibler discrepancy is a generalization of the entropy of a spectral density, which is recovered by taking $\Psi \equiv 1$.

3. Spectral Kullback-Leibler Approximation with Cepstral- and Covariance-Type Constraints. Here we are interested in the problem of finding spectral densities that fulfill conditions on their cepstra and second order moments. By assumptions on the interpolation data, we will ensure that there exists infinitely many solutions. Out of those, we will be interested in the particular solution that has the smallest spectral Kullback-Leibler discrepancy, with respect to a given density. Moreover, we will show that this spectral density is essentially unique.

Consider the following infinite dimensional approximation problem:

PROBLEM 3.1 (Kullback-Leibler Approximation). *Let $\Psi \in \mathcal{C}_+$ and $G, H \in \mathcal{G}$ be given. Assume that $(r, c) \in \mathcal{X}_{nm}$. Find any spectral density $\Phi \in \mathcal{C}_+$ that minimizes the spectral Kullback-Leibler discrepancy $\mathbb{S}(\Psi, \Phi)$ subject to the interpolation conditions*

$$r_k = \langle H_k, \Phi \rangle \qquad k = 0, \dots, n, \tag{3.1}$$

$$c_l = \langle G_l, \Psi \log \Phi \rangle \quad l = 1, \dots, m. \tag{3.2}$$

This Kullback-Leibler approximation problem is a generalization of the primal problem studied in [10, 15] in the style of [20]. In fact, in proving the theorem we shall follow these key references closely. Note that we let Ψ act as a frequency weighting of the log-spectrum of Φ.

The following theorem give the solution to the problem, its functional form, and conditions for a unique solution.

THEOREM 3.2. *The solution to Problem 3.1 is of the form $\Phi = \Psi \hat{P} \hat{Q}^{-1}$ where $\hat{P} \in \mathcal{Q}_+^0$ and $\hat{Q} \in \mathcal{Q}_+$. Moreover, if $(\hat{P}, \hat{Q})$ are coprime they are unique.*

A key feature of the theorem is the functional form of the solution, which can be interpreted as a complexity constraint. For instance, taking $m = 0$, G such that $G_k(z) = 1/(z - z_k)$, and $\Psi = \sigma\sigma^*/(\tau\tau^*)$ where $\sigma \in \mathcal{S}_n$ yields a complete parameterization of the Nevanlinna-Pick interpolation problem with degree constraint.

We will prove Theorem 3.2 using Lagrangian techniques. In fact, we will show that the dual, in mathematical programming sense, is

$$(\mathcal{D}) \quad \min_{(P,Q) \in \overline{\mathcal{Q}_+^0} \times \overline{\mathcal{Q}_+}} \underbrace{\langle Q, R \rangle - \langle P, \log R \rangle - \langle 1, P\Psi \rangle + \left\langle P\Psi, \log \frac{P\Psi}{Q} \right\rangle}_{=: \mathbb{J}(P,Q)}, \tag{3.3}$$

where $R(z) \in \mathcal{C}$ is any continuous function defined on $\mathbb{T}$, not necessarily positive, which fulfills the interpolation conditions (3.1) and (3.2).

As for the dual we will show the following also very central theorem, which will be the key in proving Theorem 3.2.

THEOREM 3.3. *The dual problem $(\mathcal{D})$ is a convex optimization problem and has a solution $(\hat{P}, \hat{Q})$ where $\hat{Q}$ is an interior point, that is $\hat{Q} \in \mathcal{Q}_+$. Any corresponding spectral density of the form $\Psi \hat{P} \hat{Q}^{-1}$ fulfills the interpolation conditions* (3.1)*. If in addition $\hat{P} \in \mathcal{Q}_+^0$ also the interpolation conditions (3.2) are satisfied. Moreover, if $(\hat{P}, \hat{Q})$ are coprime, they are unique.*

Next we shall prove the main theorems.

Proof of Theorem 3.2 First we form the Lagrangian

$$\begin{aligned}
&L(P,Q,\Phi)\\
&\quad := \langle \Psi, \log\Psi - \log\Phi\rangle - \sum_{k=0}^{n} q_k(r_k - \langle H_k, \Phi\rangle) + \sum_{l=1}^{m} p_l(c_l - \langle G_l\Psi, \log\Phi\rangle),\\
&\quad = -q^T r + p^T c + \left\langle \sum_{k=0}^{n} q_k H_k, \Phi \right\rangle - \left\langle \sum_{l=1}^{m} p_l G_l + 1, \Psi\log\Phi \right\rangle,\\
&\quad = -\langle Q, R\rangle + \langle P, \log R\rangle + \langle Q, \Phi\rangle - \langle P\Psi, \log\Phi\rangle,
\end{aligned}$$

where we have defined

$$\begin{aligned}
P &:= 1 + \frac{p_1}{2}(G_1 + G_1^*) + \cdots + \frac{p_m}{2}(G_m + G_m^*),\\
Q &:= q_0 + \frac{q_1}{2}(H_1 + H_1^*) + \cdots + \frac{q_n}{2}(H_n + H_n^*),
\end{aligned}$$

and, R as *any* function, not necessarily positive, on the circle, which fulfills the interpolation conditions (3.1) and (3.2). Here p_k and q_k are complex numbers except q_0 which is real.

The dual optimization problem then is

$$(\mathcal{D}) \quad \min_{(P,Q)\in\overline{\mathcal{Q}_+^0}\times\overline{\mathcal{Q}_+}} \; - \inf_{\Phi\in\mathcal{C}_+} L(P,Q,\Phi). \tag{3.4}$$

We get additional conditions on P and Q, by noting where the dual functional attains an infinite value. Firstly, $Q(z) \geq 0$ for all $z \in \mathbb{T}$ since otherwise the term $\langle Q, \Phi\rangle$ can be arbitrary large. Secondly, also $P(z) \geq 0$ for $z \in \mathbb{T}$ since otherwise $-\langle P\Psi, \log\Phi\rangle$ can be made arbitrarily large. These are all the requirements[1].

Next we will show that any stationary point of the map $\Phi \mapsto L(P,Q,\Phi)$ fulfills the complexity constraint $\Phi = \Psi\hat{P}\hat{Q}^{-1}$. Consider any feasible change of Φ:

$$\begin{aligned}
\delta L(P,Q,\Phi;\delta\Phi) &:= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(L(P,Q,\Phi+\varepsilon\delta\Phi) - L(P,Q,\Phi)\Big),\\
&= \langle Q, \delta\Phi\rangle - \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left\langle P\Psi, \underbrace{\log\frac{\Phi+\varepsilon\delta\Phi}{\Phi}}_{=\varepsilon\frac{\delta\Phi}{\Phi}+\text{h.o.t.}}\right\rangle,\\
&= \left\langle \delta\Phi, Q - \frac{P\Psi}{\Phi}\right\rangle.
\end{aligned}$$

Since we allow for all possible changes, any stationary point must satisfy the complexity constrain $\Phi = \Psi\hat{P}\hat{Q}^{-1}$. Evaluating the Lagrangian in the stationary point we have

$$\begin{aligned}
L(P,Q,\frac{P\Psi}{Q}) &= -\langle Q, R\rangle + \langle P, \log R\rangle + \left\langle Q, \frac{P\Psi}{Q}\right\rangle - \left\langle P\Psi, \log\frac{P\Phi}{Q}\right\rangle,\\
&= -\mathbb{J}(P,Q),
\end{aligned}$$

[1] One might believe that as for Q we also need to require that $P \leq 0$ since $\log\Phi$ can be made arbitrarily large. However, since for any fix $P > 0$ and $Q > 0$ the linear term $\langle Q, \Phi\rangle$ will dominate the logarithmic term $-\log\Phi$ for large $|\Phi|$, this is in fact not the case.

meaning that the dual problem in the Lagrangian relaxation is $(\mathcal{D})$ in (3.4).

Now, due to the definition of $\mathcal{X}_{nm}$ in (2.8), there exists at least one solution of the form $\Psi\hat{P}\hat{Q}^{-1}$ with $\hat{P} \in \mathcal{Q}_+^0$ and $\hat{Q} \in \mathcal{Q}_+$. Since the spectral Kullback-Leibler discrepancy is jointly convex we have that

$$L(\hat{P}, \hat{Q}, \hat{\Phi}) \leq L(\hat{P}, \hat{Q}, \Phi), \quad \forall\ \Phi \in \mathcal{C}_+. \tag{3.5}$$

However, for all Φ that fulfill the interpolation conditions (3.1) and (3.2), we have that

$$L(\cdot, \cdot, \Phi) = \mathbb{S}(\Psi, \Phi). \tag{3.6}$$

In particular $\hat{\Phi}$, again due to Theorem 3.3, fulfills the interpolation conditions. Therefore, combining (3.5) and (3.6) we have that

$$\mathbb{S}(\Psi, \hat{\Phi}) \leq \mathbb{S}(\Psi, \Phi), \quad \forall\ \Phi \in \mathcal{C}_+ \text{ satisfying (3.1) and (3.2)},$$

verifying the optimality of $\hat{\Phi}$. Appealing to Theorem 3.3 the solution is unique whenever $\hat{P}$ and $\hat{Q}$ are coprime. This concludes the proof of Theorem 3.2. □

Plugging that solution into the dual problem we get the dual $\mathcal{D}$ in (3.4). Next we will turn to the quite involved proof of Theorem 3.3. The proof is a fairly straightforward generalization of the corresponding proofs in [9, 11, 15].

Proof of Theorem 3.3 First we prove that the functional $\mathbb{J}(P, Q)$ is proper and bounded from below, that is, that inverse images of compact sets are compact in $\overline{\mathcal{Q}}_+^0 \times \overline{\mathcal{Q}}_+$. To this end, suppose that $(p^{(k)}, q^{(k)})$ is a sequence in $\mathbb{J}^{-1}((-\infty, \mu])$. To show that $\mathbb{J}^{-1}((-\infty, \mu])$ is compact it suffices to show that $(p^{(k)}, q^{(k)})$ has a subsequence that converges to a point in $\mathbb{J}^{-1}((-\infty, \mu])$.

First we show that $\overline{\mathcal{Q}}_+^0$ is compact. Clearly it is a closed subset of $\mathbb{R}^N$. We can factorize $P(z) = \lambda\sigma(z)\sigma^*(z)$ where $\sigma(z) \in \mathcal{S}_n$ and $p_0 = \lambda(1 + \sigma_1^2 + \cdots + \sigma_n^2)$. Clearly, the coefficients of $\sigma(z)$ are bounded and since $p_0 = 1$ also λ is bounded. Thus, also p_k for $k = 1, 2, \ldots, n$ are bounded which implies that $\overline{\mathcal{Q}}_+^0$ is bounded and hence compact. The compactness of $\overline{\mathcal{Q}}_+^0$ implies that $p^{(k)}$ has a convergent subsequence.

As for $q^{(k)}$ we can factor out the constant, $Q^{(k)}(z) = q_0^{(k)}\tilde{Q}^{(k)}(z)$ where $\tilde{q}^{(k)} \in \overline{\mathcal{Q}}_+^0$. Since $\overline{\mathcal{Q}}_+^0$ is compact it suffices to show that $q_0^{(k)}$ has a convergent subsequence. Now we can write the dual functional in (3.3) as

$$\mathbb{J}(P^{(k)}, Q^{(k)}) =: c_1^{(k)} q_0^{(k)} - c_2^{(k)} \log q_0^{(k)} - c_3^{(k)}, \tag{3.7}$$

where

$$\begin{aligned}
c_1^{(k)} &= \left\langle \tilde{Q}^{(k)}, R \right\rangle, \\
c_2^{(k)} &= \left\langle P^{(k)}\Psi, 1 \right\rangle, \\
c_3^{(k)} &= \left\langle P^{(k)}, \log R \right\rangle + \left\langle 1, P^{(k)}\Psi \right\rangle - \left\langle P^{(k)}\Psi, \log \frac{P^{(k)}\Psi}{\tilde{Q}^{(k)}} \right\rangle.
\end{aligned}$$

Clearly, since $P^{(k)}$ and $\tilde{Q}^{(k)}$ belong to $\overline{\mathcal{Q}}_+^0$ which is compact, $c_1^{(k)}$ and $c_2^{(k)}$ are positive and bounded away from positive infinity. Moreover, $c_3^{(k)}$ is bounded away from plus and minus infinity. Now, if $q_0^{(k)}$ would tend to 0 that second term in (3.7) would tend

to infinity and not stay inside $\mathbb{J}^{-1}((-\infty,\mu])$. Likewise, if $q_0^{(k)}$ would tend to positive infinity, the first term of (3.7) would tend to infinity. Thus we conclude that $q_0^{(k)}$ has a convergent subsequence and that $\mathbb{J}^{-1}((-\infty,\mu])$ is compact.

Since $\mathbb{J}$ is proper and defined on a closed, convex domain it attains a minimal point $(\hat{P},\hat{Q})$ there. Next we will show that $\hat{Q}$ is an interior point. We shall proceed as in [11]. First consider the directional derivative of $\mathbb{J}(P,Q)$ in any feasible direction $\left\{\delta P : P+\delta P \in \overline{\mathcal{Q}}_+^0\right\}$ and $\left\{\delta Q : Q+\delta Q \in \overline{\mathcal{Q}}_+\right\}$:

$$\begin{aligned}
&\delta\mathbb{J}(P,Q;\delta P,\delta Q)\\
&\quad:=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(\mathbb{J}(P+\varepsilon\delta P,Q+\varepsilon\delta Q)-\mathbb{J}(P,Q)\Big),\\
&\quad=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(\langle\varepsilon\delta Q,R\rangle-\langle\varepsilon\delta P,\log R\rangle-\langle 1,\varepsilon\delta P\Psi\rangle\\
&\qquad+\left\langle\varepsilon\delta P\Psi,\log\frac{(P+\varepsilon\delta P)\Psi}{Q+\varepsilon\delta Q}\right\rangle+\left\langle P\Psi,\log\frac{P+\varepsilon\delta P}{P}\right\rangle\\
&\qquad-\left\langle P\Psi,\log\frac{Q+\varepsilon\delta Q}{Q}\right\rangle\Big),\\
&\quad=\langle\delta Q,R\rangle-\langle\delta P,\log R\rangle-\langle\delta P,\Psi\rangle+\left\langle\delta P,\Psi\log\frac{P\Psi}{Q}\right\rangle\\
&\qquad+\left\langle P\Psi,\frac{\delta P}{P}\right\rangle-\left\langle P\Psi,\frac{\delta Q}{Q}\right\rangle,\\
&\quad=\left\langle\delta Q,R-\frac{P\Psi}{Q}\right\rangle-\left\langle\delta P,\log R-\Psi\log\frac{P\Psi}{Q}\right\rangle. \qquad (3.8)
\end{aligned}$$

For the moment, we will only study variations in $Q(z)$. Let $q\in\mathcal{Q}_+$ and $\overline{q}\in\partial\mathcal{Q}_+$ be arbitrary. Then $Q(z)$ is positive on the unit circle while $\overline{Q}(z)$ is nonnegative and equal to 0 for at least one $\theta_0\in[-\pi,\pi]$. Define $q_\lambda:=\overline{q}+\lambda(q-\overline{q})$ for $\lambda\in(0,1]$ where $\overline{q}$ corresponds to $\overline{Q}(z)$. Then also $Q_\lambda(z)$ is positive on the unit circle. Consider the directional derivative in (P,Q_λ) in the direction $\delta Q=\overline{Q}-Q$ and keeping P constant:

$$\delta\mathbb{J}(P,Q_\lambda;0,\overline{Q}-Q)=\left\langle\overline{Q}-Q,R-\frac{P\Psi}{Q_\lambda}\right\rangle=w^T(\overline{q}-q)-\left\langle P\Psi,\frac{\overline{Q}-Q}{Q_\lambda}\right\rangle. \qquad (3.9)$$

Now, note that

$$\frac{d}{d\lambda}\frac{\overline{Q}-Q}{Q_\lambda}=-\frac{\overline{Q}-Q}{Q_\lambda^2}\frac{dQ_\lambda}{d\lambda}=\left(\frac{\overline{Q}-Q}{Q_\lambda}\right)^2\geq 0,$$

and hence the integrand of the second term of (3.9) is a monotonically nondecreasing function of λ for all $z\in\mathbb{T}$. Thus the integrand tends pointwise on the unit circle to $(\overline{Q}-Q)/Q$ as $\lambda\to 0$. Since the $\{(\overline{Q}-Q)/Q_\lambda\}_\lambda$ is a Cauchy sequence in $\mathcal{L}_1(\mathbb{T})$ it converges almost everywhere to $(\overline{Q}-Q)/\overline{Q}$. However, since $(\overline{Q}-Q)/\overline{Q}$ has at least one pole on the unit circle it is not summable and

$$-\left\langle P\Psi,\frac{\overline{Q}-Q}{Q_\lambda}\right\rangle\to\infty,\quad\lambda\to 0.$$

Consequently, $\delta\mathbb{J}(P,Q_\lambda;0,\overline{Q}-Q)\to\infty$ as $\lambda\to 0$ for all $q\in\mathcal{Q}_+$ and $\overline{q}\in\partial\mathcal{Q}_+$. Hence, by [29, Lemma 26.2] $\mathbb{J}$ is an essentially smooth functional of Q and by [29,

Theorem 26.3] it is essentially strictly convex with respect to Q. Thus we have proven that there exists a minimizer $(\hat{P}, \hat{Q}) \in \overline{\mathcal{Q}}_+^0 \times \mathcal{Q}_+$.

Since $\hat{Q}$ is an interior point the stationarity condition must be satisfied there. Taking $\delta Q = H_k + H_k^*$ and $\delta P = 0$ in (3.8) yields the stationarity condition

$$r_k = \langle H_k, R \rangle = \langle H_k, \Phi \rangle \quad k = 0, \dots, n.$$

If in addition $\hat{P}$ is an interior point, and thus a stationary point, (3.8) also yields

$$c_l = \langle G_l, \log R \rangle = \langle G_l, \Psi \log \Phi \rangle \quad l = 1, \dots, m.$$

We need to show that the optimal point is unique whenever $\hat{P}$ and $\hat{Q}$ are coprime. Consider the second variation:

$$\begin{aligned}
&\delta^2 \mathbb{J}(P, Q; \delta P, \delta Q) \\
&\quad := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} (\delta \mathbb{J}(P + \varepsilon \delta P, Q + \varepsilon \delta Q; \delta P, \delta Q) - \mathbb{J}(P, Q; \delta P, \delta Q)), \\
&\quad = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\left\langle \delta Q, R - \frac{(P + \varepsilon \delta P)\Psi}{Q + \varepsilon \delta Q} \right\rangle - \left\langle \delta P, \log R - \Psi \log \frac{(P + \varepsilon \delta P)\Psi}{Q + \varepsilon \delta Q} \right\rangle \right. \\
&\qquad \left. - \left\langle \delta Q, R - \frac{P\Psi}{Q} \right\rangle + \left\langle \delta P, \log R - \Psi \log \frac{P\Psi}{Q} \right\rangle \right), \\
&\quad = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\left\langle \delta Q, \frac{Q(P + \varepsilon \delta P)\Psi - (Q + \varepsilon \delta Q)P\Psi}{(Q + \varepsilon \delta Q)Q} \right\rangle \right. \\
&\qquad \left. + \left\langle \delta P, \Psi (\log \frac{P + \varepsilon \delta P}{P} - \log \frac{Q + \varepsilon \delta Q}{Q}) \right\rangle \right), \\
&\quad = \left\langle \delta Q, \frac{P \delta Q \Psi - \delta P Q \Psi}{Q^2} \right\rangle + \left\langle \delta P, \Psi \left(\frac{\delta P}{P} - \frac{\delta Q}{Q} \right) \right\rangle, \\
&\quad = \left\langle \frac{\Psi}{PQ^2}, (\delta P Q - P \delta Q)^2 \right\rangle \geq 0.
\end{aligned}$$

Therefore, the dual functional $\mathbb{J}$ is convex. The second variation is zero only when $P\delta Q - \delta P Q$, that is,

$$\frac{P}{Q} = \frac{\delta P}{\delta Q}.$$

However, this is impossible if $\hat{P}$ and $\hat{Q}$ are coprime, since $p_0 \equiv 1$ implies that $\delta p_0 = 0$. Thus, $\mathbb{J}$ is strictly convex at $(\hat{P}, \hat{Q})$ if they are coprime, and the optimal point is indeed unique. That concludes the proof of Theorem 3.3. □

The statement of Problem 3.1 might appear intractable since, as stated in Remark 2.1, there is no available test for checking whether a point (r, c) belongs to $\mathcal{X}_{nm}$. The benefit of this formulation is the direct parameterization of ARMA models, on which we will elaborate more in the next section. Also, seen as an ARMA estimator and considering the asymptotical statistical properties of the parameters, the case when $(\hat{r}, \hat{c}) \notin \mathcal{X}_{nm}$ can easily be taken care of. In fact, we can take $(\hat{r}^N, \hat{c}^N) = (\hat{r}^{N-1}, \hat{c}^{N-1})$ whenever the N^{th} estimate falls outside $\mathcal{X}_{nm}$ and the initial estimate $(\hat{r}^0, \hat{c}^0)$ arbitrary in $\mathcal{X}_{nm}$. This will not affect the asymptotic behavior.

Yet, as a practical procedure in ARMA estimation and robust control, this is indeed an issue. Following [14, 15] we will study a *regularized* dual optimization

problem, where we introduce a barrier-like term which will force the optimal point into the interior of the feasible region, that is, also with respect to the numerator pseudo-polynomial. More precisely, consider the problem

$$(\mathcal{D}_\lambda) \qquad \min_{(P,Q)\in\overline{\mathcal{Q}_+^0}\times\overline{\mathcal{Q}_+}} \mathbb{J}(P,Q) - \lambda \langle 1, \log P\rangle\,, \tag{3.10}$$

where $\lambda > 0$. Repeating the arguments in the proof of Theorem 3.3 one can readily show that the additional term, $-\lambda\langle 1, \log P\rangle$ is functional, which is proper and bounded from above, and whose derivative tends to negative infinity when the P tend to the boundary of $\mathcal{Q}_+^0$. Therefore the functional will still be proper and bounded from above so there exist a solution. Also, a parallel discussion with respect to P rules out the possibility to have a boundary solution. The first order variation (3.8) now becomes

$$\delta\mathbb{J}(P,Q;\delta P,\delta Q) = \left\langle \delta Q, R - \frac{P\Psi}{Q}\right\rangle + \left\langle \delta P, \log R - \Psi\log\frac{P\Psi}{Q} - \frac{\lambda}{P}\right\rangle.$$

At the stationary point we will therefore not quite match the cepstral estimate, but rather the modified estimate:

$$c_l = \langle G_l, \Psi\log\Phi\rangle - \lambda\left\langle 1, \frac{1}{P}\right\rangle \quad l = 1,\dots,m. \tag{3.11}$$

In fact, we have the following result.

THEOREM 3.4. *Let $(r,c) \in \mathcal{R}_n \times \mathbb{C}^n$. Then the regularized dual problem $(\mathcal{D}_\lambda)$ is a convex optimization problem and has an interior point solution $(\hat{P},\hat{Q}) \in \mathcal{Q}_+^0 \times \mathcal{Q}_+$. Any corresponding spectral density of the form $\Psi\hat{P}\hat{Q}^{-1}$ fulfills the interpolation conditions* (3.1) *and* (3.11).

As for the special case studied in [15], we recover the original problem with $\lambda = 0$. When $\lambda \to \infty$ the regularization term tend to infinity unless $P \to 1$. Therefore, as argued in [15], the maximum entropy solution is recovered when $\lambda = \infty$. These properties make λ a natural choice for deformation parameter in a numerical continuation method, see [1], and the algorithm of [15] is based on this observation.

REMARK 3.5. *The theorems of this section are generalizations of the results in [10, 20]. In fact, taking $\Psi = 1$ and G as the standard basis yields Theorem 5.1 of [10] while taking $m = 0$, that is no cepstral interpolation, yields Theorem 5 of [20].*

4. A Family of Global Coordinatizations of $\mathcal{P}_n^*$. In this section we shall show that the normalized generalized prefiltered covariances and generalized prefiltered cepstral coefficients provide a coordinatization of stable miniphase real rational functions of fixed degree for each choice of prefilter and each choice of basis functions. Hence we get a family of coordinatizations of $\mathcal{P}_n^*$ with the standard covariances and correlation coefficients as one member. By spectral factorization it is also a coordinatization of positive real functions of bounded degree. Note that in this section we only treat the real case rather than the complex case in Section 3. Thereby, all functions in $\mathcal{C}_+$ are real, the matrices (A,B) are real making $G \in \mathcal{G}$ real, and all interpolation data (r,c) is real.

We shall treat the normalized case, that is when $a,b,\sigma \in \mathcal{S}_n$ and where we normalize the covariance-type interpolation conditions to $r = r/r_0$. This will reduce the dimension of the problem by one and simplify the overall analysis somewhat. It can be perceived as a counterpart of analytically reducing the innovation variance in Maximum Likelihood ARMA estimation. Also see the discussion in [10, p. 29].

Since all functions are scalar in this section we write

$$\langle G, \Phi \rangle = \begin{bmatrix} \langle G_0, \Phi \rangle \\ \vdots \\ \langle G_n, \Phi \rangle \end{bmatrix},$$

which is a slight abuse of notation. However, it simplifies the presentation considerably.

THEOREM 4.1. *Let $\Psi \in \mathcal{C}_+$ and $G \in \mathcal{G}$ be given. The corresponding generalized prefiltered normalized covariances $r_1, r_2, \ldots, r_n$ and the generalized prefiltered cepstral coefficients $c_1, c_2, \ldots, c_n$ provide a smooth coordinatization of $\mathcal{P}_n^*$.*

The theorem states that the map

$$\begin{array}{rccc} F & : & \mathcal{P}_n^* & \to & \mathcal{X}_n, \\ & & (a, \sigma) & \mapsto & (r, c), \end{array} \tag{4.1}$$

where r and c are the generalized filtered normalized covariances and cepstral coefficients in (2.6) and (2.7), respectively, is a diffeomorphism. The normalization means that $r_0 = 1$ and we have taken $r = (r_1, \ldots, r_n)$. A direct consequence of the theorem is

COROLLARY 4.2. *The map F is a homeomorphism and $\mathcal{X}_n$ has the same topological properties as $\mathcal{P}_n^*$.*

REMARK 4.3. *Theorem 4.1 is a generalization of [10, Theorem 3.1] which is recovered by taking $\Psi \equiv 1$ and G as the standard basis. Our proof has the same structure as that of [10] but introducing Ψ render some technical difficulties.*

The rest of this section is devoted to the proof of Theorem 4.1. One might believe that F is a diffeomorphism as a direct consequence of some global inverse function theorem, such as Hadamard's global inverse function theorem [21]. However, the rather complicated topology of $\mathcal{P}_n^*$ and $\mathcal{X}_n$, see Section 2, make such global theorems not applicable. Instead, we will perform a global analysis of two foliations of the manifold $\mathcal{P}_n$ in order to prove that F is a *local* diffeomorphism at each point of $\mathcal{P}_n^*$. In fact we will prove:

THEOREM 4.4. *The map F if a local diffeomorphism on $\mathcal{P}_n^*$.*

In order to prove Theorem 4.4, we will study two sets of submanifolds of $\mathcal{P}_n$. In fact, they both form n-dimensional foliations of $\mathcal{P}_n$. For $k = 0, \ldots, n$, define the maps

$$\begin{array}{rccc} \xi_k & : & \mathcal{P}_n & \to & \mathbb{R}, \\ & & (a, \sigma) & \mapsto & \left\langle G_k, \Psi \dfrac{\sigma\sigma^*}{aa^*} \right\rangle. \end{array} \tag{4.2}$$

Normalization with the zeroth generalized prefiltered covariance gives $\eta : \mathcal{P}_n \to \mathbb{R}^n$ with components $\eta_k = \xi_k/\xi_0$, $k = 1 \ldots n$. The normalization makes η a map to the *generalized prefiltered correlation coefficients*. We have that $\mathcal{R}_n = \eta(\mathcal{P}_n)$, where $\mathcal{R}_n \subset \mathbb{R}^n$ with the previously described normalization. Given $r \in \mathcal{R}_n$ define the first set of submanifolds as the subsets of $\mathcal{P}_n$ matching r, that is,

$$\mathcal{P}_n(r) := \eta^{-1}(r).$$

As for the second foliation, we define the map $\zeta : \mathcal{P}_n \to \mathbb{R}^n$ to the cepstral coefficients:

$$\begin{array}{rccc} \zeta_k & : & \mathcal{P}_n & \to & \mathbb{R}, \\ & & (a, \sigma) & \mapsto & \left\langle G_k, \Psi \log \dfrac{\sigma\sigma^*}{aa^*} \right\rangle, \end{array} \tag{4.3}$$

for $k = 1 \dots n$. The set of feasible cepstra is given by $\mathcal{C}_n = \zeta(\mathcal{P}_n)$. Now, given $c \in \mathcal{C}_n$, define the second set of submanifolds as the subsets of $\mathcal{P}_n$ with cepstra c, that is,

$$\mathcal{P}_n(c) := \zeta^{-1}(c).$$

First we will state and prove a preliminary result is regarding a linear map. Let $\phi \in \mathcal{S}_n$. Consider the linear map from the vector space of polynomials of degree at most $n-1$:

$$\begin{array}{rcccl} \vartheta_\phi & : & \mathcal{L}_{n-1} & \to & U \subset \mathbb{R}^n, \\ & & u & \mapsto & \left\langle \bar{G}, \dfrac{T(\phi)u}{\phi\phi^*} \right\rangle. \end{array}$$

The map is invertible, generalizing [10, Lemma 4.1]. However, we can not directly generalize the proof since $\langle \Psi, T(\phi)u/(\phi\phi^*) \rangle$ is, in general, nonzero for nonconstant Ψ.

LEMMA 4.5. *The linear map ϑ_ϕ is a bijection.*

Proof. Start with injectivity by supposing that $\vartheta_\phi u = 0$. Then

$$\left\langle G_k, \Psi \frac{T(\phi)u}{\phi\phi^*} \right\rangle = 0, \quad k = 1, 2, \dots n.$$

By symmetry this also holds for $k = -1, -2, \cdots - n$. Therefore

$$\left\langle G_k \frac{\tau\tau^*}{\phi\phi^*}, \frac{\phi\phi^*}{\tau\tau^*} \Psi \frac{T(\phi)u}{\phi\phi^*} \right\rangle = 0, \quad k = \pm 1, \pm 2, \cdots \pm n.$$

Now let $\hat{G} \in \mathcal{G}$ be a set of basis functions corresponding to $(\hat{A}, \hat{B})$ such that $\phi = \det(I - \hat{A}z)$. Since $\langle 1, G_k \rangle =< 1, \hat{G}_k >= 0$ we then have that

$$\left\langle \hat{G}_k, \frac{\phi\phi^*}{\tau\tau^*} \Psi \frac{T(\phi)u}{\phi\phi^*} \right\rangle = 0, \quad k = \pm 1, \pm 2, \cdots \pm n.$$

Now, since

$$\frac{T(\phi)u}{\phi\phi^*} = \frac{u}{\phi} + \frac{u^*}{\phi^*},$$

with u/ϕ strictly proper, taking an appropriate linear combination we have

$$\left\langle \frac{T(\phi)u}{\phi\phi^*}, \frac{\phi\phi^*}{\tau\tau^*} \Psi \frac{T(\phi)u}{\phi\phi^*} \right\rangle = \left\| \frac{T(\phi)u}{\phi\tau} w \right\|^2 = 0,$$

where w is the spectral factor of Ψ. Hence $T(\phi)u = 0$ and by the invertibility of T, see for instance [13, Lemma 2.1], we also have $u = 0$. Hence ϑ_ϕ is injective. Being a linear map between vector spaces of the same real dimension, it is also surjective, and hence bijective. □

Now, we can prove the first major result regarding the submanifolds of $\mathcal{P}_n$.

PROPOSITION 4.6. *The manifolds $\mathcal{P}_n(c)$ are smooth n-manifolds and their tangent space $T_{(a,\sigma)}\mathcal{P}_n(c)$ consists of those $(u, v) \in \mathcal{L}_{n-1} \times \mathcal{L}_{n-1}$ for which*

$$\left\langle G_k, \Psi \frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle G_k, \Psi \frac{T(a)u}{aa^*} \right\rangle, \tag{4.4}$$

for $k = 1 \dots n$. Moreover the connected components of the n-manifolds $\{\mathcal{P}_n(c) : c \in \mathcal{C}_n\}$ form the leaves of a foliation of $\mathcal{P}_n$.

Proof. The tangent vector of $\mathcal{P}_n(c)$ at (a, σ), $T_{(a,\sigma)}\mathcal{P}_n(c)$ are the vectors in the kernel of the Jacobian of ζ at (a, σ). For $u, v \in \mathcal{L}_{n-1}$:

$$D_{(u,v)}\zeta(a,\sigma) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\zeta(a + \varepsilon u, \sigma + \varepsilon v) - \zeta(a,\sigma)\right).$$

Applying the calculation

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\log(\sigma + \varepsilon v) - \log \sigma\right) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \log\left(1 + \epsilon \frac{v}{\sigma}\right) = \frac{v}{\sigma},$$

we get that

$$\begin{aligned} D_{(u,v)}\zeta(a,\sigma) &= \left\langle \bar{G}, \Psi\left(\frac{v}{\sigma} + \frac{v^*}{\sigma^*} - \frac{u}{a} - \frac{u^*}{a^*}\right)\right\rangle, \\ &= \left\langle \bar{G}, \Psi\left(\frac{T(\sigma)v}{\sigma\sigma^*} - \frac{T(a)u}{aa^*}\right)\right\rangle. \end{aligned} \tag{4.5}$$

Thus we have proven that the tangent space is given by (4.4). Since both the maps ϑ_a and ϑ_σ are bijective linear maps by Lemma 4.5, the tangent space is of dimension n. Hence the rank of $\mathrm{Jac}(\zeta)|_{(a,\sigma)}$ is then full for all feasible (a, σ). By the implicit function theorem $\mathcal{P}_n(c)$ are then smooth n-manifolds.

Since $\mathrm{Jac}(\zeta)|_{(a,\sigma)}$ is full rank, ζ is a submersion, and hence the connected components of the n-manifolds $\{\mathcal{P}_n(c) : c \in \mathcal{C}_n\}$ form the leaves of a foliation of $\mathcal{P}_n$. □

Now we have a parallel statement for $\mathcal{P}_n(r)$, which we will prove in a similar fashion. Here, the normalization makes the analysis somewhat more involved but the generalization from [10] is more direct.

PROPOSITION 4.7. *The manifolds $\mathcal{P}_n(r)$ are smooth n-manifolds and their tangent space $T_{(a,\sigma)}\mathcal{P}_n(r)$ consists of those $(u, v) \in \mathcal{L}_{n-1} \times \mathcal{L}_{n-1}$ for which*

$$\left\langle G_k, \Psi \frac{T(\sigma)v}{aa^*} \right\rangle = \left\langle G_k, \Psi \frac{\sigma\sigma^*}{aa^*} \frac{T(a)u}{aa^*} \right\rangle + \varphi(a,\sigma,u,v) \left\langle G_k, \Psi \frac{\sigma\sigma^*}{aa^*} \right\rangle, \tag{4.6}$$

for $k = 0 \dots n$ and where

$$\varphi(a,\sigma,u,v) := D_{(u,v)} \log \xi_0(a,\sigma) = \frac{D_{(u,v)}\xi_0(a,\sigma)}{\xi_0(a,\sigma)}.$$

Moreover the connected components of the n-manifolds $\{\mathcal{P}_n(r) : r \in \mathcal{R}_n\}$ form the leaves of a foliation of $\mathcal{P}_n$.

Proof. Again we compute the directional derivative of η at $(a, \sigma) \in \mathcal{P}_n$ in the direction $(u, v) \in \mathcal{L}_{n-1} \times \mathcal{L}_{n-1}$. We have

$$D_{(u,v)}\eta_k(a,\sigma) = \frac{1}{\xi_0(a,\sigma)} D_{(u,v)}\xi_k(a,\sigma) - \frac{\xi_k(a,\sigma)}{\xi_0(a,\sigma)^2} D_{(u,v)}\xi_0(a,\sigma), \tag{4.7}$$

where

$$D_{(u,v)}\xi_k(a,\sigma) = \left\langle G_k, \Psi\left(\frac{T(\sigma)v}{aa^*} - \frac{\sigma\sigma^*}{aa^*}\frac{T(a)u}{aa^*}\right)\right\rangle. \tag{4.8}$$

Multiplying (4.7) with $\xi_0(a,\sigma) = r_0 > 0$ we get the the kernel of $\mathrm{Jac}(\eta)|_{(a,\sigma)}$ to consist of all $(u,v) \in \mathcal{L}_n \times \mathcal{L}_n$ such that

$$\left\langle G_k, \Psi \frac{T(\sigma)v}{aa^*} \right\rangle = \left\langle G_k, \Psi \frac{\sigma\sigma^*}{aa^*} \frac{T(\sigma)v}{aa^*} \right\rangle + \varphi(a,\sigma,u,v) \left\langle G_k, \Psi \frac{\sigma\sigma^*}{aa^*} \right\rangle, \tag{4.9}$$

for $k = 1 \dots n$. Since $\eta_0 = \xi_0/\xi_0 = 1$ (4.9) trivially also holds for $k = 0$. This establishes (4.6). Next we will prove that the tangent space is n-dimensional for all $(a,\sigma) \in \mathcal{P}_n$. Let p be a polynomial of degree n defined by $p(z) := v(z) + \varphi a(z)$. Then the tangent equations can be written as

$$\Pi p = \Upsilon u,$$

where the linear operators $\Pi : \mathcal{L}_n \to \mathbb{R}^{n+1}$ and $\Upsilon : \mathcal{L}_{n-1} \to \mathbb{R}^{n+1}$ are given by

$$\Pi p := \left\langle G, \Psi \frac{\sigma\sigma^*}{aa^*} \frac{T(\sigma)p}{aa^*} \right\rangle \text{ and } \Upsilon u := \left\langle G, \Psi \frac{T(\sigma)u}{aa^*} \right\rangle.$$

To see this, note that $T(a)a/(aa^*) = 2$. Now, Π is in fact injective. Assume that $\Pi p = 0$. By changing basis functions from G to some $\tilde{G} \in \mathcal{G}$ associated with $(\tilde{A}, \tilde{B})$ such that $\det(I - \tilde{A}z) = a(z)$ we have that, for some nonsingular U, that

$$\begin{aligned} &\Pi p = U \left\langle \tilde{G}, \frac{aa^*}{\tau\tau^*} \Psi \frac{\sigma\sigma^*}{aa^*} \frac{T(\sigma)p}{aa^*} \right\rangle = 0, \\ \Rightarrow \quad &\left\langle \tilde{G}_k, \Psi \frac{\sigma\sigma^*}{\tau\tau^*} \frac{T(\sigma)p}{aa^*} \right\rangle = 0, \ k = 0, \pm 1, \cdots \pm n. \end{aligned}$$

Now, taking appropriate linear combinations we have that

$$0 = \left\langle \frac{T(\sigma)p}{aa^*}, \Psi \frac{\sigma\sigma^*}{\tau\tau^*} \frac{T(\sigma)p}{aa^*} \right\rangle = \left\| w \frac{\sigma T(\sigma)p}{a\tau} \right\|^2,$$

where w is the spectral factor of Ψ. Hence $T(\sigma)p = 0$, implying that $p = 0$. Thus Π is injective. Then we have $p = \Pi^{-1}\Upsilon v$.

Since the leading coefficient of p is $\varphi/2$, this defines an affine map $L : \mathcal{L}_{n-1} \to \mathcal{L}_{n-1}$ sending u to $v := \Pi^{-1}\Upsilon u - \varphi a/2$. Then $T_{(a,\sigma)}\mathcal{P}_n(r)$ consists of those $(u,v) \in \mathcal{L}_{n-1} \times \mathcal{L}_{n-1}$ such that $v = Lu$ which hence is n dimensional. Therefore the rank of $\mathrm{Jac}(\eta)|_{(a,\sigma)}$ is full for all $(a,\sigma) \in \mathcal{P}_n$ so that $\mathcal{P}_n(r)$ are smooth n-manifolds by the implicit function theorem.

As in the proof of Proposition 4.6, η is a submersion and the claim follows. □

Next we shall study the intersection of the tangent spaces $T_{(a,\sigma)}\mathcal{P}_n(c)$ and $T_{(a,\sigma)}\mathcal{P}_n(r)$. Whenever the intersection is a unique point, the submanifolds $\mathcal{P}_n(c)$ and $\mathcal{P}_n(r)$ are complementary and provide a coordinatization.

THEOREM 4.8. *The tangent spaces $T_{(a,\sigma)}\mathcal{P}_n(r)$ and $T_{(a,\sigma)}\mathcal{P}_n(c)$ are complementary in $\mathcal{P}_n^*$. The dimension of $\Theta := T_{(a,\sigma)}\mathcal{P}_n(r) \cap T_{(a,\sigma)}\mathcal{P}_n(c)$ is the degree of the greatest common divisor.*

Proof. First consider the equations for $T_{(a,\sigma)}\mathcal{P}_n(c)$:

$$\left\langle G_k, \Psi \frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle G_k, \Psi \frac{T(a)u}{aa^*} \right\rangle, \ k = \pm 1, \cdots \pm n, \tag{4.10}$$

which can be written

$$\left\langle G_k \frac{\tau\tau^*}{\sigma\sigma^*}, \frac{\sigma\sigma^*}{\tau\tau^*} \Psi \frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle G_k \frac{\tau\tau^*}{\sigma\sigma^*}, \frac{\sigma\sigma^*}{\tau\tau^*} \Psi \frac{T(a)u}{aa^*} \right\rangle, \ k = \pm 1, \cdots \pm n.$$

Now let $\hat{G} \in \mathcal{G}$ be a set of basis functions corresponding to $(\hat{A}, \hat{B})$ such that $\sigma = \det(I - \hat{A}z)$. Since $\langle 1, G_k \rangle =< 1, \hat{G}_k >= 0$ we then have that

$$\left\langle \hat{G}_k, \frac{\sigma\sigma^*}{\tau\tau^*}\Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \hat{G}_k, \frac{\sigma\sigma^*}{\tau\tau^*}\Psi\frac{T(a)u}{aa^*} \right\rangle, \ k = \pm 1, \cdots \pm n.$$

Now, since

$$\frac{T(\sigma)v}{\sigma\sigma^*} = \frac{v}{\sigma} + \frac{v^*}{\sigma^*},$$

with v/σ strictly proper, taking an appropriate linear combination we have

$$\left\langle \frac{T(\sigma)v}{\sigma\sigma^*}, \frac{\sigma\sigma^*}{\tau\tau^*}\Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \frac{T(\sigma)v}{\sigma\sigma^*}, \frac{\sigma\sigma^*}{\tau\tau^*}\Psi\frac{T(a)u}{aa^*} \right\rangle,$$

that is

$$\left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(a)u}{aa^*} \right\rangle.$$

Taking linear combinations of (4.10) corresponding to $T(\sigma)v/(\tau\tau^*)$ yields

$$\left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(a)u}{aa^*} \right\rangle + \left\langle 1, \frac{T(\sigma)v}{\tau\tau^*} \right\rangle \left\langle \Psi, \frac{T(\sigma)v}{\sigma\sigma^*} - \frac{T(a)u}{aa^*} \right\rangle.$$

Since $T(\sigma)v/(\tau\tau^*)$ is a density for nonzero v, combining the expressions we have

$$\left\langle \Psi, \frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \Psi, \frac{T(a)u}{aa^*} \right\rangle,$$

on $T_{(a,\sigma)}\mathcal{P}_n(c)$. Hence, we have that

$$\left\langle G_k, \Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle G_k, \Psi\frac{T(a)u}{aa^*} \right\rangle,$$

for $k = 0, \pm 1, \cdots \pm n$ on $T_{(a,\sigma)}\mathcal{P}_n(c)$. The equations describing $T_{(a,\sigma)}\mathcal{P}_n(r)$ are

$$\left\langle G_k, \Psi\frac{T(\sigma)v}{aa^*} \right\rangle = \left\langle G_k, \Psi\frac{\sigma\sigma^*}{aa^*}\frac{T(\sigma)v}{aa^*} \right\rangle + \varphi(a,\sigma,u,v)\left\langle G_k, \Psi\frac{\sigma\sigma^*}{aa^*} \right\rangle,$$

for $k = 0, \pm 1, \cdots \pm n$. Now, taking appropriate linear combinations we have that

$$\left\langle \Psi, \frac{T(\sigma)v}{\tau\tau^*} \right\rangle = \left\langle \Psi, \frac{\sigma\sigma^*}{\tau\tau^*}\frac{T(a)u}{aa^*} \right\rangle,$$
$$\left\langle \Psi, \frac{T(\sigma)v}{\tau\tau^*} \right\rangle = \left\langle \Psi, \frac{\sigma\sigma^*}{\tau\tau^*}\frac{T(a)u}{aa^*} \right\rangle + \varphi\left\langle \Psi, \frac{\sigma\sigma^*}{\tau\tau^*} \right\rangle.$$

We conclude that $\varphi = 0$ on Θ.

Thus on Θ we have

$$\left\langle G_k, \Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle G_k, \Psi\frac{T(a)u}{aa^*} \right\rangle, \qquad k = 0, \pm 1, \cdots \pm n,$$
$$\left\langle G_k, \Psi\frac{T(\sigma)v}{aa^*} \right\rangle = \left\langle G_k, \Psi\frac{\sigma\sigma^*}{aa^*}\frac{T(\sigma)v}{aa^*} \right\rangle, \ k = 0, \pm 1, \cdots \pm n.$$

Again taking appropriate linear combinations we have that

$$\left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(\sigma)v}{\sigma\sigma^*} \right\rangle = \left\langle \frac{T(\sigma)v}{\tau\tau^*}, \Psi\frac{T(a)u}{aa^*} \right\rangle,$$
$$\left\langle \frac{T(a)u}{\tau\tau^*}, \Psi\frac{T(\sigma)v}{aa^*} \right\rangle = \left\langle \frac{T(a)u}{\tau\tau^*}, \Psi\frac{\sigma\sigma^*}{aa^*}\frac{T(a)u}{aa^*} \right\rangle.$$

Again letting w be the spectral factor of Ψ and defining

$$f_1 = \frac{T(\sigma)v}{\tau\sigma^*}w \text{ and } f_2 = \frac{T(a)u\,\sigma}{\tau aa^*}w,$$

the equations can be written $\|f_1\|^2 = \langle f_1, f_2\rangle$ and $\langle f_1, f_2\rangle = \|f_2\|^2$. By the parallelogram law we then have

$$\|f_1 - f_2\| = \|f_1\|^2 + \|f_2\|^2 - 2\langle f_1, f_2\rangle = 0.$$

Hence $f_1 = f_2$ on the unit circle implying that

$$\frac{v}{\sigma} + \frac{v^*}{\sigma^*} = \frac{T(\sigma)v}{\sigma\sigma^*} = \frac{T(a)u}{aa^*} = \frac{u}{a} + \frac{u^*}{a^*},$$

on the unit circle. Being real polynomials this need to hold also for the positive real part. Moreover, it clearly holds for $u = v = 0$ so now consider the nontrivial case. We can write the equation as

$$\frac{v}{u} = \frac{\sigma}{a}.$$

This has no solution if (a, σ) are coprime, which establishes that $T_{(a,\sigma)}\mathcal{P}_n(r)$ and $T_{(a,\sigma)}\mathcal{P}_n(c)$ are complementary in $\mathcal{P}_n^*$. On the other hand, if they have a common factor of degree d, then u and v can be any polynomials of degree $n-1$ with a common factor of degree at least $d-1$, hence defining a vector space of dimension d establishing the rest of the claim. □

We summarize with the proofs of the main theorems.

Proof of Theorem 4.4 Since $T_{(a,\sigma)}\mathcal{P}_n(r)$ and $T_{(a,\sigma)}\mathcal{P}_n(c)$ are complementary in $\mathcal{P}_n^*$ by Theorem 4.8, the kernels of $\mathrm{Jac}(\eta)|_{(a,\sigma)}$ and $\mathrm{Jac}(\zeta)|_{(a,\sigma)}$ are complementary at any $(a, \sigma) \in \mathcal{P}_n^*$. Hence the Jacobian of the joint map F is full rank. By the implicit function theorem F is a local diffeomorphism on $\mathcal{P}_n^*$. □

Proof of Theorem 4.1 First we prove that the map F is a bijection as a consequence of Theorem 3.2 in the previous section. Take $H = G$. Let $Q(z) = \lambda_1 a(z)a^*(z)$ and $P(z) = \lambda_2\sigma(z)\sigma^*(z)$ where λ_1 and λ_2 are taken so that $p_0 = 1$ and $r_0 = 1$. Since P and Q are coprime F is a bijection by Theorem 3.2. Together with Theorem 4.4 this implies that F is a diffeomorphism. □

5. Simultaneously Solving the Cepstral and Covariance Equations. In this section we will study the scalar problem also allowing for the cepstral-type conditions. Since we have generalized the theory of [10], the numerical algorithms of [14, 15] are no longer applicable. Here we will construct a homotopy with respect to the covariances and cepstral interpolation data from some initial values to some desired values. By choosing the initial values so that the covariance-type conditions are satisfied, we can construct a homotopy which lies in the connected sub-manifold $\mathcal{P}_n(r)$. Thereby, we can solve the inverse problem of going from the interpolation data to a

model by solving a set of ordinary differential equations. The well-posedness shown in Section 4 is critical in motivating the here proposed algorithm.

We shall consider the case of $m = n$ being the default. Also, the smoothness properties in Section 4 were derived for this case. However, a generalization to the arbitrary (n, m) seems within reach. Moreover we will assume that the prefilter density is rational of degree n. In this section we shall index the polynomial coefficients in decreasing powers of the variable.

Assume that the prefilter density Ψ is given by

$$\Psi = \frac{\hat{a}\hat{a}^*}{\hat{\sigma}\hat{\sigma}^*},$$

where $\hat{a}$ and $\hat{\sigma}$ are of order n. Typically, for ARMA estimation, we will take $(\hat{a}, \hat{\sigma})$ as a preliminary estimate of the ARMA model. We also choose some basis function $\tilde{G} \in \mathcal{G}$. The map ξ, defined by (4.2), then has the components

$$\begin{aligned} \xi_k \quad : \quad \mathcal{P}_n \quad &\to \quad \mathbb{R}, \\ (a, \sigma) \quad &\mapsto \quad \left\langle \tilde{G}_k, \frac{\hat{a}\hat{a}^*}{\hat{\sigma}\hat{\sigma}^*} \frac{\sigma\sigma^*}{aa^*} \right\rangle, \end{aligned}$$

for $k = 0 \dots n$. The normalized coefficients are as before given by $\eta_k = \xi_k/\xi_0$ for $k = 0 \dots n$. Likewise, for the cepstral-type equations, we have the map ζ defined in (4.3) with components

$$\begin{aligned} \zeta_k \quad : \quad \mathcal{P}_n \quad &\to \quad \mathbb{R}, \\ (a, \sigma) \quad &\mapsto \quad \left\langle \tilde{G}_k, \frac{\hat{a}\hat{a}^*}{\hat{\sigma}\hat{\sigma}^*} \log \frac{\sigma\sigma^*}{aa^*} \right\rangle, \end{aligned}$$

for $k = 1 \dots n$. The map F is given by (4.1). Let r and c be some given data normalized so that $r_0 = 1$. As noted in Section 3, generic data might not belong to $\mathcal{X}_n$. To circumvent this issue we study some regularization of the problem. Here we consider the regularization terms

$$s_k(\sigma) := \left\langle \tilde{G}_k, \frac{\hat{a}\hat{a}^*}{\hat{\sigma}\hat{\sigma}^*} \frac{1}{\sigma\sigma^*} \right\rangle, \ k = 1 \dots m,$$

for the corresponding cepstral equations.

We will consider a particular choice of basis function, namely orthonormal basis functions $G = (I - Az)^{-1}B$ and so that $\det(I - Az) = \hat{\sigma}(z)$ and another set of orthonormal basis function $\tilde{G} = (I - \tilde{A}z)^{-1}\tilde{B}$ such that $\det(I - \tilde{A}z) = \hat{a}(z)$. By a trivial change of basis functions, see [2, 5] we can instead consider the simplified maps with components

$$\begin{aligned} \xi_k(a, \sigma) &= \left\langle G_k, \frac{\sigma\sigma^*}{aa^*} \right\rangle, \\ \zeta_k(a, \sigma) &= \left\langle G_k, \log \frac{\sigma\sigma^*}{aa^*} \right\rangle, \\ s_k(\sigma) &= \left\langle G_k, \frac{1}{\sigma\sigma^*} \right\rangle, \end{aligned}$$

for $k = 0, 1, \dots n$.

REMARK 5.1. *Another obvious choice of basis function is to take $\tilde{G}$ as the standard basis defined. This typically enable fast evaluation of the functions ξ and ζ as well as their derivatives. However, it is our experience that the method presented here yields better numerical scaling.*

Since the manifold $\mathcal{P}_n^*$, or equivalently $\mathcal{X}_n$, is known to have $n+1$ connected components, see, for instance, [7, 30], we need to be somewhat careful in designing a homotopy from a known solution to the desired solution. We will use the following result proven by Byrnes and Lindquist:

COROLLARY 5.2. *[13, Corollary 5.5] The submanifolds $\mathcal{P}_n(r)$ are connected.*

Therefore, we wish to start in a point (a_0, σ_0) which fulfills the covariances type conditions. Supposing that the prefilter density corresponds to a model, which is close to the desired, a natural initial point is to take $\sigma_0 = \hat{\sigma}$. Then, to find an a_0 such that the covariance-type conditions hold, we can use the algorithm of [4]. Note that in general this $a_0 \neq \hat{a}$. Then we can construct a homotopy from $F(a_0, \sigma_0) = (r, c_0)$ to the desired values $F(a, \sigma) = (r, c)$ which stays in one component of $\mathcal{P}_n^*$ as

$$\begin{aligned} h \quad : \quad [0,1] \times \mathcal{P}_n \quad &\to \quad U \subset \mathbb{R}^{2n}, \\ (\mu, a, \sigma) \quad &\mapsto \quad F(a,\sigma) - \varepsilon \begin{bmatrix} 0 \\ s(\sigma) \end{bmatrix} - \mu \begin{bmatrix} r_0 \\ c \end{bmatrix} - (1-\mu) \begin{bmatrix} r_0 \\ c_0 \end{bmatrix}. \end{aligned} \tag{5.1}$$

The initial point is given by $h(0, a_0, \sigma_0) = 0$ and the desired solution is given by the nonlinear equation $h(1, a, \sigma) = 0$. We have a *trajectory* defined by

$$\{(a,\sigma) \in \mathcal{P}_n(r) : h(\mu, a, \sigma) = 0, \mu \in [0,1]\}.$$

Since $\mathcal{P}_n(r)$ is a subset of $\mathcal{P}_n^*$, we have by Theorem 4.1 that F restricted to $\mathcal{P}_n(r)$ is a diffeomorphism onto its image. Hence it has a full rank Jacobian there, and 0 is a regular value of the homotopy h. Thus we will get a smooth curve from the initial point to the solution. In particular we have no turning point, bifurcations, and the curve is of finite length.

We define the initial value problem as in [1], by differentiating h with respect to μ. Let $x := (a, \sigma)$.

$$\frac{\partial h}{\partial x}\frac{\partial x}{\partial \mu} + \frac{\partial h}{\partial \mu} = 0.$$

Here

$$\frac{\partial h}{\partial x} = \frac{dF}{dx} - \varepsilon \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{ds}{d\sigma} \end{bmatrix},$$

so F being diffeomorphic implies that $\partial h/\partial x$ is full rank for all x for some $\varepsilon > 0$. Hence we get the initial value problem as

$$\begin{cases} \dfrac{dx}{d\mu} &= -\left(\dfrac{\partial h}{\partial x}\right)^{-1} \dfrac{\partial h}{\partial \mu}, \\ x(0) &= 0. \end{cases} \tag{5.2}$$

To solve the initial value problem one can apply some predictor-corrector method, see [1], or some other ordinary differential equation solver. For the predictor-corrector method we have the Euler step as

$$v(\mu, x) := -\left(\frac{\partial h}{\partial x}\right)^{-1} \frac{dh}{d\mu}, \tag{5.3}$$

ALGORITHM 1. *Cepstral- and Covariance Matching*
Set $a \leftarrow a^{ME}$, $\sigma \leftarrow \hat{\sigma}$, and $\mu \leftarrow 0$
while $\mu < 1$
  begin Predictor step
    Determine Euler direction v by (5.3)
    Set $d\mu \leftarrow c_3(\mu) \leq 1 - \mu$.
    while $x + d\mu v \notin \mathcal{P}_n^*$
      Set $d\mu \leftarrow d\mu/2$
    end while
    Set $x \leftarrow x + d\mu v$
  end Predictor step
  begin Corrector step
    while $\max\{h_k(\mu, x)\} > c_4(\mu)$
      Determine Newton step v by (5.4)
      Set $d\nu \leftarrow 1$
      while $x + d\nu v \notin \mathcal{P}_n^*$
        Set $d\nu \leftarrow d\nu/2$
      end while
      Set $x \leftarrow x + d\nu v$
    end while
  end Corrector step
end while

and Newton steps as

$$x_{k+1} - x_k = -\left(\frac{\partial h}{\partial x}\right)^{-1} h(\mu, x). \tag{5.4}$$

Expressions for computing the partial derivatives are direct, see [2].

Algorithm 1 is a predictor-corrector algorithm for solving the cepstral and covariance equations simultaneously. Here $c_3(\mu)$ is some function of μ determining the step size in the predictor step. With small increments we follow the trajectory closely, but that increases the number of steps. How to choose $c_3(\mu)$ is therefore a trade-off. The function $c_4(\mu)$ affects the accuracy in the corrector step, that is how close to the trajectory we need to be before taking a new predictor step. The value $c_4(1)$ gives the accuracy of the final solution. To test whether $x \in \mathcal{P}_n^*$, we compute the Schur parameters of a and σ and check whether they are less than one in modulus.

6. ARMA Estimation. Next we shall consider how the generalized simultaneous cepstral and covariance matching can be applied to ARMA estimation. As seen from Example 1.1 in the introduction, simultaneously matching a direct estimate of the covariances and cepstral coefficients is not expected to yield a statistically efficient estimate. However, we observed that the estimator seemed to be efficient, or close to efficient, when the MA zero was close to origin. Then, the covariances and the cepstral coefficients were computed for data that was close to white noise. Here we will explore this by trying to prefilter the data to make it close to white noise. This is sometimes called *whitening*[2].

[2]In particular one can interpret the Prediction Error Method, PEM, as whitening the data. By feeding the data reversely through the current model estimate the prediction errors are obtained. The PEM estimate is the minimizer of the sum of squared prediction errors.

Consider Figure 6.1. Here $\{x_t\}_{t=1}^N$ is the measured data in the same way as in Example 1.1. Assume that we have a preliminary estimate of $w(z)$, say $\tilde{w}(z)$. By choosing $\psi(z) = \tilde{w}^{-1}(z)$ the filtered data $\{y_t\}_{t=1}^N$ will be closer to white noise if the preliminary estimate were decent. Now, estimate the biased covariances and cepstral coefficients of the filtered data. Given these, we formulate the Kullback-Leibler approximation problem, see Problem 3.1, with $\Psi = \psi\psi^*$. From Theorem 3.2 we know that there is a unique solution of the form $\Phi = \Psi\hat{P}\hat{Q}^{-1}$. Since the prefilter contributes multiplicatively with Ψ, we can ignore that factor and our estimate of the ARMA model will be the stable, miniphase spectral factor of $\hat{P}\hat{Q}^{-1}$.

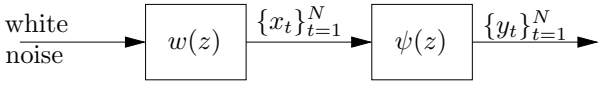

FIG. 6.1. *The prefiltering for ARMA estimation.*

We can iteratively change the prefilter ψ by using the new estimates of the model. In Procedure 1 we propose one possible such scheme, which uses Algorithm 1. Being iterative, it is not clear whether the procedure converges. However, we will not study this issue in this paper.

PROCEDURE 1. *Prefiltered Cepstral-Covariance Matching, CCM(f)*
Set $\hat{a} \leftarrow 1$ and $\hat{\sigma} \leftarrow 1$
for $k = 1, \ldots, 5$
  Set $\varepsilon \leftarrow c_5(k)$
  Set $\Psi \leftarrow \hat{a}\hat{a}^*/(\hat{\sigma}\hat{\sigma}^*)$
  Estimate r and c from original data
  Determine a and σ using Algorithm 1
end for

No matter how interesting, it is beyond the scope of this paper to include an exhaustive statistical analysis of proposed CCM(f) estimator. However, we shall prove one fairly immediate result, which ought to be a key result in any statistical analysis of the estimator.

THEOREM 6.1. *Let $\Psi = (\tau\tau^*)/(\sigma\sigma^*) \in \mathcal{Q}_+$ and $G \in \mathcal{G}$ such that $\det(I - Az) = \tau(z)$ be given. Also let a time series of length N generated from an ARMA process with parameters Θ_0 be given. Assume that some estimation procedure estimates the generalized prefiltered covariances and cepstral coefficients Ξ such that*

$$\Xi \text{ is } AN\left(\Xi_0, N^{-1}U\right), \quad U = \begin{pmatrix} U_1 & U_2 \\ U_2^T & U_3 \end{pmatrix},$$

where Ξ_0 are the parameters corresponding to Θ_0. Then

$$\Theta \text{ is } AN(\Theta_0, N^{-1}W),$$

where W is the covariance matrix

$$W = \left[\frac{\partial F}{\partial \Theta}\right]^{-1} \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} U_1 & U_2 \\ U_2^T & U_3 \end{pmatrix} \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix}^T \left[\frac{\partial F}{\partial \Theta}\right]^{-T} \Bigg|_{\Theta_0},$$

where F is the map defined in (4.1) *and D is given by*

$$D := r_0^{-1} \begin{bmatrix} -r_1/r_0 & 1 & 0 & \cdots & 0 \\ -r_2/r_0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ -r_n/r_0 & 0 & 0 & & 1 \end{bmatrix}. \tag{6.1}$$

Proof. The Jacobian of the map mapping the generalized prefiltered covariances to the normalized ditto, that is, $\begin{bmatrix} r_0 & r_1 & \dots & r_n \end{bmatrix}^T \mapsto \begin{bmatrix} r_1/r_0 & r_2/r_0 & \dots & r_n/r_0 \end{bmatrix}^T$, is given by (6.1). By Theorem 4.1 the map F has an everywhere invertible Jacobian. Since F is a diffeomorphism and the diagonal elements of U are nonzero, so are the diagonal elements of W. Hence the claim follows by Proposition 6.4.3 in [8]. □

The theorem tells us that the CCM(f) estimates inherit the statistical properties from the estimates of the covariances and cepstral coefficients. The underlying reason is of course the smoothness of the map F discussed in detail in Section 4. Thus, if we can estimate the covariances and the cepstral coefficients statistically efficiently, then we automatically have a statistically efficient estimate of the ARMA model. Also, consistently estimated cepstral coefficients and covariences yield consistent estimates of ARMA models.

The joint distribution of generalized prefiltered covariances and cepstral coefficients is unknown, also in the case of unfiltered coefficients with the standard basis. Yet, Theorem 6.1 serves as a conceptual tool in understanding the following example, where we will study the idea presented above for ARMA(n,n) models. This is believed to be a generic example, in comparison to the overly simplified example in the introduction with a simple real zero. However, we acknowledge that higher order models are more useful in practice – not the least in term of numerical issues.

Example 6.2 (ARMA(n,n)). Consider Figure 6.1. First take the true model to have poles in $0.5e^{\pm 2i}$ and zeros in $0.98e^{\pm i}$. The corresponding density is plotted in Figure 6.2. Note that the zero close to the unit circle creates a frequency region with low magnitude of the spectrum. This makes the identification harder. Also compare to the case in Example 1.1 when the simple zero tended towards the circle.

Given a time series consisting of the measurements x_t with sample lengths $N = 200, 400, \dots, 12800$ we try to identify the filter. We compare three estimators. As reference we use the Maximum-Likelihood estimator implemented in `armax` in [24]. The estimator will be denote ML. The second estimator is the Cepstral-Covariance Matching estimator without prefiltering and with the standard basis, as presented in [10, 15], though computed by an implementation of Algorithm 1. More precisely, we use the standard biased sample covariances. For estimation of the cepstral coefficients we first estimate long AR models of orders $L = 10, 15, \dots, 40$ for the different sample sizes, respectively. The corresponding ARMA models are computed using an implementation of Algorithm 1 with the default prefilter $\Psi \equiv 1$. We will call the estimator CCM. The last estimator is based on Procedure 1 where we recursively estimate new, generalized, covariances and cepstral coefficients. To estimate the generalized prefiltered covariances we apply the input-to-state framework and estimate the state-covariance. We use the least-squares approach suggested in [19, p. 34] to estimate a feasible estimate of the interpolation data matrix W; see also the discussion in [3]. For the generalized cepstrum, we again estimate long AR models of orders $L = 10, 15, \dots, 40$. From the AR model we can directly compute the generalized prefiltered cepstral coefficients.

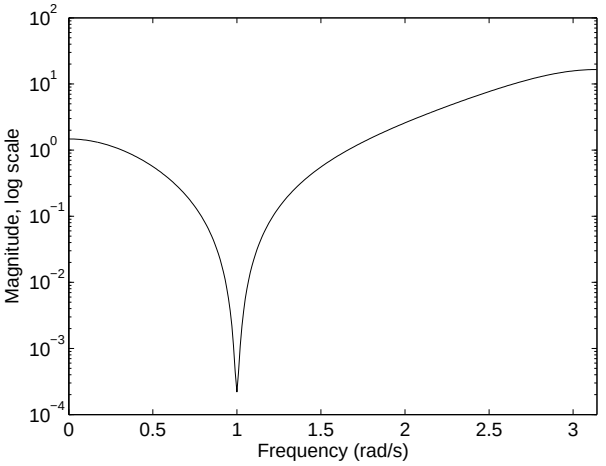

FIG. 6.2. *The spectrum of the AR(2,2) process.*

TABLE 6.1
The mean and standard deviation in the estimation of the ARMA parameters for $N = 12800$.

	True	ML		CCM(f)		CCM	
		Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
a_1	0.000	−0.000	0.010	0.000	0.009	0.009	0.013
a_2	−0.250	−0.249	0.010	−0.250	0.010	−0.244	0.014
σ_1	−1.070	−1.067	0.003	−1.065	0.004	−1.008	0.007
σ_2	0.980	0.974	0.006	0.972	0.006	0.884	0.007

First we make a statistical comparison of the parameter estimate. Since the variance is asymptotical decoupled from estimating the other ARMA parameters, see for instance [28], we will only compare the parameters $\begin{bmatrix}\sigma_1 & \sigma_2 & a_1 & a_2\end{bmatrix}$. In Table 6.1 the estimated means and variances of the parameter estimates for the different methods using a Monte Carlo simulation with 500 runs is given. We note that all methods seem to be unbiased. Moreover, the variance for the CCM(f) estimator is approximately the same as for the ML estimator, which in turn is approximately efficient. Meanwhile, the unfiltered estimator, CCM, does not seem to be efficient. Thus the example indicates that the prefiltering seem to make the CCM method approximately asymptotically efficient.

Another way of comparing the estimators is to compute some error measure of each estimate and then by the Monte Carlo simulation estimate what the mean error is. In Figures 6.3 and 6.4 the estimated prediction error and Kullback-Leibler discrepancy relative to the true model are plotted as a function of the sample size. The prediction errors are computed by feeding white Gaussian noise with 100 samples through the filters and then estimating the prediction errors. The Kullback-Leibler discrepancy used is $\mathbb{S}(\Phi_{true}, \Phi_{est})$ where Φ_{true} and Φ_{est} are the normalized densities of the true and estimated densities, respectively. We again note that, independent of comparison method, the CCM(f) seem to have asymptotical error similar to ML while CCM's error seems larger. This highlights the reason for introducing the prefiltering.

REMARK 6.3. *An observation made in Example 1.1 is that the high order AR based cepstral estimate seems to be asymptotically efficient for white noise. This ex-*

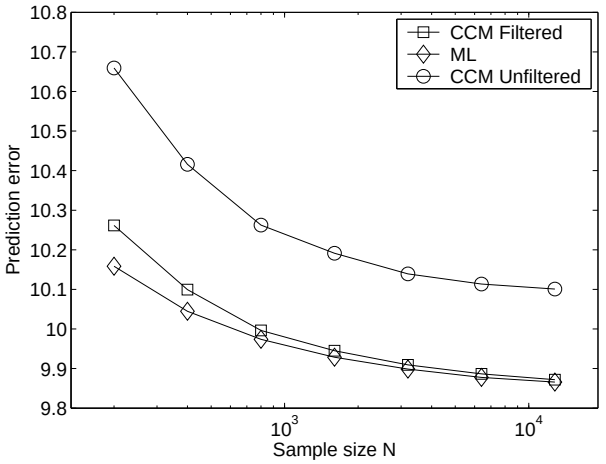

FIG. 6.3. *The prediction error as a function of the sample length for the ARMA(2,2) model.*

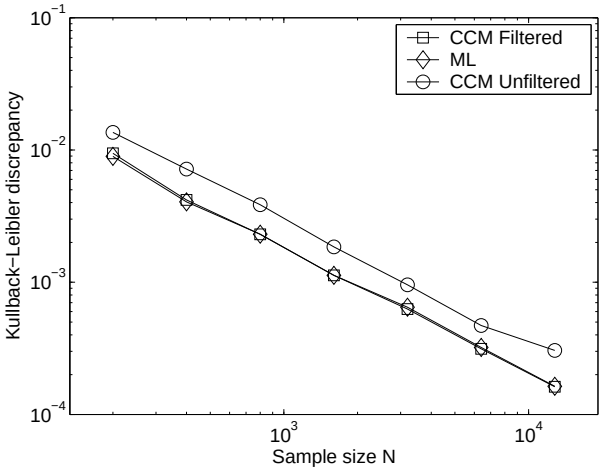

FIG. 6.4. *The Kullback-Leibler discrepancy as a function of the sample length for the ARMA(2,2) model.*

ample indicates that this might generalize to other ARMA models. Therefore it can be in place to conjecture, that if the prefilter is the inverse of the true model, the covariance and cepstral estimates are asymptotically efficient and hence so the corresponding ARMA estimates by Theorem 6.1. Furthermore, the example indicates that using a recursive estimation scheme for determining the prefilter might also constitute an asymptotically efficient estimator.

7. Conclusions. We have generalized the cepstral-covariance coordinatization of ARMA processes of Byrnes *et al.* in [10, 15] to include frequency weighting/prefiltering. A simulation example indicates that this might yield an asymptotically efficient estimator while maintaining the well-posedness of problem.

We have shown that the statistical properties ARMA parameters are inherited from the covariances and cepstral coefficients. These depend on how the coefficients are computed and is partly unknown; some results can be found in [16, 26]. This

topic requires further attention.

We propose an algorithm which for a given prefilter computes the estimates utilizes the algorithm in [4]. Thereby, instead of solving the convex optimization problem, the set of nonlinear moment equations are considered. Similar problems occur elsewhere, see for instance [12, 17], and the numerical aspects are not fully understood.

A procedure for iteratively estimating a suitable prefilter is used in the simulation example. It serves the purpose of illustrating the prefiltering idea but is not robust and reliable enough to be a practical algorithm.

REFERENCES

[1] E. L. Allgower and K. Georg. *Numerical Continuation Methods*. Springer-Verlag, 1990.

[2] A. Blomqvist. *A Convex Optimization Approach to Complexity Constrained Analytic Interpolation with Applications to ARMA Estimation and Robust Control*. PhD thesis, Royal Institute of Technology, 2005.

[3] A. Blomqvist and G. Fanizza. Identification of rational spectral densities using orthonormal basis functions. In *The proceedings of Symposium on System Identification*, Rotterdam, The Netherlands, 2003.

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

[5] A. Blomqvist and B. Wahlberg. A data driven orthonormal parameterization of the generalized entropy maximization problem. In *The proceedings of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems*, 2004.

[6] B. P. Bogert, J. R. Healy, and J. W. Tukey. The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. In M. Rosenblatt, editor, *Prooceedings of the Symposium on Time Series Analysis*, pages 209–243. John Wiley and Sons, 1963.

[7] R. W. Brockett. Some geometric questions in the theory of linear systems. *IEEE Trans. Automat. Control*, 21(4):449–455, 1976.

[8] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods, Second Edition*. Springer Series in Statistics. Springer, 1991.

[9] C. I. Byrnes, P. Enqvist, and A. Lindquist. Cepstral coefficient, covariance lags and pole-zero models for finite data strings. *IEEE Trans. Signal Processing*, 49(4):677–693, April 2001.

[10] C. I. Byrnes, P. Enqvist, and A. Lindquist. Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach. *SIAM J. Contr. and Optimiz.*, 41(1):23–59, 2002.

[11] 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.

[12] C. I. Byrnes, S. V. Gusev, and A. Lindquist. From finite covariance windows to modeling filters: A convex optimization approach. *SIAM Review*, 43(4):645–675, 2001.

[13] C. I. Byrnes and A. Lindquist. On the duality between filtering and Nevanlinna-Pick interpolation. *SIAM J. Contr. and Optimiz.*, 39(3):757–775, 2000.

[14] P. Enqvist. *Spectral Estimation by Geometric, Topological and Optimization Methods*. PhD thesis, Royal Institute of Technology, Stockholm, Sweden, 2001.

[15] P. Enqvist. A convex optimization approach to ARMA(n,m) model design from covariance and cepstral data. *SIAM J. Contr. and Optimiz.*, 43(3):1011–1036, 2004.

[16] Y. Ephraim and M. Rahim. On second-order statistics and linear estimation of cepstral coefficients. *IEEE Trans. Signal Processing*, 7(2):162–176, 1999.

[17] T. T. Georgiou. Solution of the general moment problem via a one-parameter imbedding. Submitted to IEEE Trans. Automat. Control.

[18] T. T. Georgiou. Signal estimation via selective harmonic amplification: MUSIC, Redux. *IEEE Trans. Signal Processing*, 48(3):780–790, March 2000.

[19] T. T. Georgiou. Spectral estimation via selective harmonic amplification. *IEEE Trans. Automat. Control*, 46(1):29–42, 2001.

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

[21] J. Hadamard. Sur les correspondances ponctuelles. In *Oeuvres, Editions du Centre Nationale de la Researche Scientifique*, pages 383–384. Paris, 1968.

[22] S. Kullback. *Information Theory and Statistics*. Wiley Publications in Statistics. John Wiley & Sons, 1959.
[23] A. Lindquist. Notes on covariance lags and cepstral coefficients. 2002.
[24] L. Ljung. Systems identification toolbox. Mathworks MATLAB toolbox.
[25] L. Ljung. *System Identification, Theory for the User*. Prentice Hall, 1999.
[26] N. Merhav and C.-H. Lee. On the asymptotic statistical behavior of emperical cepstral coefficients. *IEEE Trans. Signal Processing*, 41(5):1990–1993, May 1993.
[27] A. V. Oppenheim and R. W. Shafer. *Digital Signal Processing*. Prentice Hall, London, 1975.
[28] B. Porat. *Digital Processing of Random Signals, Theory & Methods*. Prentice Hall, 1994.
[29] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princton, NJ, 1970.
[30] G. Segal. The topology of spaces of rational functions. *Acta Mathematica*, 143:39–72, 1979.