

# NUMERICAL SPECTRAL FACTORIZATION BY KALMAN RECURSION

JARED A. MCBRIDE

ABSTRACT. In this note I describe a method presented by Kailath et al. in [1] and [2] for finding a spectral factor of the  $z$ -spectrum of a (wide-sense stationary) stochastic process under certain conditions. The method employs a version of Kalman filtering. As such I begin with a summary of relevant results from Kalman filtering. With those results in hand I present the method for finding spectral factors and conclude with examples.

## CONTENTS

1. Introduction	1
2. Kalman Filtering	2
2.1. Mathematical tools and context	2
2.2. the Kalman filter	4
2.3. Kalman filter recursions for the innovations	4
2.4. The Stationary Case and the Modeling Filter	8
3. Spectral Factorization by Kalman Filtering	11
3.1. Algorithm and input-output analysis	14
4. An Example	16
Appendix A. CKMS	16
Appendix B. Proof of Lemma 1	16
Appendix C. Illustration	19
References	19

## 1. INTRODUCTION

Given a wide-sense stationary (WSS) stochastic process  $X = (X_n, n = 0, 1, \dots)$  it's  $z$ -spectrum  $S_X(z)$  is defined by

$$S_X(z) = \sum_{n=-\infty}^{\infty} C_X[n]z^{-n},$$

where  $C_X[n] = \mathbb{E}X_nX_0^*$  is the autocovariance sequence of the process  $X^1$ . So that  $S_X(z)$  is also the  $z$ -transform of the autocovariance sequence. This function is real-valued on the unit circle since,  $C_X[-n] = C_X^*[n]$ , and by the Wiener-Khinchin theorem, it is, in fact, nonnegative definite on the unit circle<sup>2</sup>. Spectral factorization refers to a particular factorization in which

$$S_X(z) = L(z)L^*(z^{-*})$$

where  $L(z)$  is *minimum phase*, meaning both  $L(z)$  and  $L^{-1}(z) = 1/L(z)$  are analytic on and outside the unit circle. The function  $L$  will be referred to as a spectral factor of  $S_X$ . This factorization is not unique since for any spectral factor of  $S_X$ , right-multiplication by a unitary matrix produces another spectral factor of  $S_X$ .

In this note I describe a method, presented by Kailath et al. in [1] and [2, p. 336], for finding a spectral factor under certain conditions. The method employs a version of Kalman filtering, more specifically the Kalman recursion for finding the innovations process for a process with a finite, linear state-space model. The innovations for such a process will be defined below. Section 2 provides a summary of relevant results from Kalman filtering, together with some intuition relevant to the present context. With those results in hand, Section 3 presents the method for finding spectral factors. Then, in Section 4 are examples.

## 2. KALMAN FILTERING

The purpose of this section is to provide information from Kalman filtering theory that is pertinent to the development of the factorization algorithm to be discussed. To that end, the *summum bonum* of the section is the computation of a modeling filter for a given process. The associated whitening filter also bears significance in this work and will be briefly treated. I begin by introducing the mathematical tools that I will be using, the inner product, norm, and the concept of linear least-mean-squares estimate appropriate to the context. A result is then provided that is important to the discussion of the Kalman filter. We then develop the Kalman filter in some generality and apply it to the time-invariant state-space process. We then add the assumption of stationarity and conclude the section with a theorem for a modeling filter of a given time-invariant (finite) state-space process that is stationary.

**2.1. Mathematical tools and context.** Suppose we have a (discrete-time) stochastic process  $X = (X_i, i = 0, 1, 2, \dots)$ <sup>3</sup> with a finite state-space representation of the following form:

<sup>1</sup>A note on notation: the superscript  $*$  denotes the complex conjugate transpose (Hermitian transpose) operator for matrices and just the complex conjugate for real or complex numbers. The superscripts  $-*$  together is an abbreviation for the complex conjugate of the multiplicative inverse, i.e.

$$z^{-*} = \frac{1}{z^*}.$$

<sup>2</sup>Though what follows holds for scalar processes it also holds for vector valued processes. As such, I assume the vector case. Meaning, the covariance sequence is a sequence of square  $d \times d$ -matrices, where  $d$  is the dimension of  $X_0$ .

<sup>3</sup>From here till Section 2.4  $X$  need not be WSS.

$$\begin{aligned} (1a) \quad & \theta_{i+1} = F_i \theta_i + G_i u_i \\ (1b) \quad & X_i = H_i \theta_i + v_i \end{aligned}$$

where  $F_i \in \mathbb{C}^{m \times m}$ ,  $G_i \in \mathbb{C}^{m \times p}$ , and  $H_i \in \mathbb{C}^{d \times m}$  are known matrices, and  $u = (u_i)$ ,  $v = (v_i)$ , and  $\theta_0$  are random variables with the following property<sup>4</sup>

$$(1c) \quad \mathbb{E} \begin{pmatrix} \theta_0 \\ u_i \\ v_i \end{pmatrix} \begin{pmatrix} \theta_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi_0 & 0 & 0 & 0 \\ 0 & Q_i \delta_{ij} & S_i \delta_{ij} & 0 \\ 0 & S_i^* \delta_{ij} & R_i \delta_{ij} & 0 \end{pmatrix}$$

In the context here, I think of  $X$  as a given series of *observations*, they are directly measurable, and  $\theta = (\theta_i, i = 0, 1, 2, \dots)$  is the *state* of the modeled process (usually not directly measurable). Both of these series are sequences of random variables, along with  $u$  and  $v$ . The matrices  $F_i$ ,  $G_i$ ,  $H_i$ ,  $\Pi_0$ ,  $Q_i$ ,  $S_i$  and  $R_i$  are deterministic for all  $i$ .

Now, in the sequel there will be much discussion on linear least-mean-square estimators (LMSEs). A nice thing about LMSEs is that there is an orthogonality principle, namely the estimator error is orthogonal to the predictor<sup>5</sup>. What follows also affords such a principle. However, there is no Hilbert space that comfortably houses this theory. So, I will build up the context here.

The context requires the consideration of vectors of more than one size, since the observations and the state need not be the same dimension. Since it is possible to write solutions to the problems I consider in terms of covariance information a natural “inner product” may be defined by

$$(2) \quad \langle x, y \rangle = \mathbb{E} x y^* \quad \text{for all } x \in \mathbb{C}^d, y \in \mathbb{C}^m, \quad d, m \in \mathbb{N},$$

So, that  $\langle x, x \rangle \in \mathbb{C}^{d \times d}$  is the covariance of  $x$  and  $\langle x, y \rangle \in \mathbb{C}^{d \times m}$  is the cross covariance of  $x$  and  $y$ . I use the notation  $x \perp y$  if  $x$  and  $y$  are uncorrelated, meaning  $\langle x, y \rangle = 0 \in \mathbb{C}^{d \times m}$ .

Because this inner product<sup>6</sup> is matrix-valued, rather than adding structure to a linear space with a scalar *field* the underlining space must be a module, with a *ring* of (square) matrix-valued scalars. Because we are using linear combinations of variables of one dimension to estimate variables of a possibly different dimension, the natural choice of scalars are matrices of the appropriate size. This is an important point and leads me to this notation. Let

$$\text{span}_d \{y_i, i = 0, 1, \dots, N\} := \left\{ \sum_{i=0}^N A_i y_i \mid y_i \in \mathbb{C}^{m_i}, A_i \in \mathbb{C}^{d \times m_i}, i = 0, 1, \dots, N \right\}$$

<sup>4</sup>Throughout the note

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

<sup>5</sup>The predictor refers to the random variable with respect to which the estimator is a linear combination. In our case this will exclusively be the observations,  $X_i$

<sup>6</sup>If its arguments are restricted to the same linear space, it can be shown that  $\langle \cdot, \cdot \rangle$  satisfies the axioms of an inner product, provided positivity is replaced with positive definiteness

I would like us to think of the span as possibly set in a different space from the arguments.

There is a norm

$$(3) \quad \|x\| = \sqrt{\langle x, x \rangle}$$

associated with this inner product, defined in the usual way, using the unique (Hermitian) positive-semidefinite square root of the Hermitian positive-semidefinite matrix  $\langle x, x \rangle$ .

I have just introduced the context and notation with which we will be describing spectral factorization.

**2.2. the Kalman filter.** Now, the Kalman filter allows us to find, for some  $i > 0$ , the linear LMSE of the state  $\theta_{i+1}$  given all preceding observations  $X_j$  for  $j = 0, 1, 2, \dots, i$ . This quantity I will denote by  $\hat{\theta}_{i+1|i}$ . Since in this note I will only be concerned with linear LMSEs given the  $X_j$ 's, the subscript  $\cdot|_i$  will denote linear estimators with respect to and conditioned on all  $X_j$  for  $j \leq i$ . Consistent with this convention, when I write  $\hat{X}_{i|i-1}$  I mean the linear LMSE of  $X_i$  given  $X_j$  for  $j < i$ .

For the stake of rigor and completeness I have provided a lemma which describes the orthogonality principle in this context. For the sake of brevity I have placed the proof in Appendix B.

**Lemma 1.** *Given the inner product defined under (2) and the norm induced by it, under (3), let  $x \in C^d$  and  $y_i \in \mathbb{C}^m$  for  $i = 0, \dots, N$  be random variables. If  $\hat{x} \in \text{span}_d\{y_i, i = 0, 1, \dots, N\}$  is chosen such that*

$$(4) \quad (x - \hat{x}) \perp y_i \quad \text{for} \quad i = 0, 1, \dots, N,$$

*then  $\hat{x}$  is the linear LMSE for  $x$  given  $\{y_i, i = 0, \dots, N\}$ , meaning*

$$\mathbb{E}(x - \hat{x})(x - \hat{x})^* = \|x - \hat{x}\|^2 = \inf_{x' \in \text{span}_d\{y_i, \forall i\}} \|x - x'\|^2.$$

**2.3. Kalman filter recursions for the innovations.** Now, with these ideas in mind, we seek a LMSE of  $\theta_{i+1}$  restricted to a linear (matrix-valued coefficient) combination of observations,  $\{X_j, j < i\}$ . This means we seek  $\hat{\theta}_{i+1|i}$  such that,

$$(5) \quad \hat{\theta}_{i+1|i} \in \mathcal{L}_{d,i} := \text{span}_d\{X_i, i = 0, 1, \dots, N\}$$

$$(6) \quad \hat{\theta}_{i+1|i} = \underset{\hat{\theta} \in \mathcal{L}_{d,i}}{\text{argmin}} \|\theta_{i+1} - \hat{\theta}\|^2.$$

Borrowing from Hilbert space theory, we consider building an orthogonal basis for the space  $\mathcal{L}_{m,i} := \text{span}_m\{X_i, i = 0, 1, \dots, N\}$  (recall  $X_0 \in \mathbb{C}^m$ ),  $\{e_j, j = 1, 2, \dots, i\}$ . Then project  $\theta_{i+1|i}$  on to  $\mathcal{L}_{d,i}$  by the formula

$$(7) \quad \hat{\theta}_{i|k} = \sum_{j=0}^k \langle \theta_i, e_j \rangle R_{e,j}^{-1} e_j \quad k < i$$

where  $R_{e,i} = \langle e_i, e_i \rangle$ . We are not in a Hilbert space, so the claim that the estimator under (7) satisfies (6) will need to be verified. To that end, let us, first, build this orthogonal set  $\{e_j, j = 1, 2, \dots, i\}$ . I proceed in a way similar to the Gram-Schmidt process, but without

normalizing. Put  $e_0 = X_0$  and proceed for  $i > 0$  iteratively by

$$(8) \quad e_i = X_i - \sum_{j=0}^{i-1} \langle X_i, e_j \rangle R_{e,j}^{-1} e_j.$$

Certainly,  $\text{span}_m\{e_j, j = 0, 1, \dots, i\} = \mathcal{L}_{m,i}$ . I will show orthogonality of the set  $\{e_j, j = 1, 2, \dots, i\}$  explicitly since matrix inner products are not very common. I proceed by induction. Observe that for  $i = 1$

$$\begin{aligned} \langle e_0, e_1 \rangle &= \langle e_0, X_1 - \langle X_1, e_0 \rangle R_{e,0}^{-1} e_0 \rangle \\ &= \langle e_0, X_1 \rangle - \langle e_0, \langle X_1, e_0 \rangle R_{e,0}^{-1} e_0 \rangle \\ &= \langle e_0, X_1 \rangle - \langle e_0, e_0 \rangle R_{e,0}^{-1} \langle e_0, X_1 \rangle = 0. \end{aligned}$$

Here the facts that  $R_{e,0}$  is hermitian and that  $\langle X_1, e_0 \rangle^* = \langle e_0, X_1 \rangle$  were used. Now, suppose for any  $i < k$  and  $j < i$  that  $\langle e_j, e_i \rangle = 0$ , and observe that for  $j < k$

$$\begin{aligned} \langle e_j, e_k \rangle &= \left\langle e_j, X_k - \sum_{l=0}^{k-1} \langle X_k, e_l \rangle R_{e,l}^{-1} e_l \right\rangle \\ &= \langle e_j, X_k \rangle - \sum_{l=0}^{k-1} \langle e_j, \langle X_k, e_l \rangle R_{e,l}^{-1} e_l \rangle \\ &= \langle e_j, X_k \rangle - \sum_{l=0}^{k-1} \langle e_j, e_l \rangle R_{e,l}^{-1} \langle e_l, X_k \rangle \\ &= \langle e_j, X_k \rangle - \langle e_j, e_j \rangle R_{e,j}^{-1} \langle e_j, X_k \rangle = 0. \end{aligned}$$

And so,  $\langle e_i, e_j \rangle = \mathbb{R}_{e,i} \delta_{ij}$  for all  $i, j \geq 0$ . A consequence of this orthogonality results together with Lemma 1 is the more intuitively useful equation

$$e_i = X_i - \hat{X}_{i|i-1} \quad \text{for } i > 0.$$

For this reason, the process  $e = (e_i, i = 0, 1, 2, \dots)$  is commonly referred to as the *innovations* associated with  $X$ . And now I can prove equation (7). Using Lemma 1 we need only show

$$\left( \theta_i - \sum_{l=0}^k \langle \theta_i, e_l \rangle R_{e,l}^{-1} e_l \right) \perp e_j \quad \text{for } j = 0, 1, \dots, k.$$

To that end consider the following for  $0 < j \leq k < i$

$$\begin{aligned} \left\langle e_j, \theta_i - \sum_{l=0}^k \langle \theta_i, e_l \rangle R_{e,l}^{-1} e_l \right\rangle &= \langle e_j, \theta_i \rangle - \sum_{l=0}^k \langle e_j, \langle \theta_i, e_l \rangle R_{e,l}^{-1} e_l \rangle \\ &= \langle e_j, \theta_i \rangle - \sum_{l=0}^k \langle e_j, e_l \rangle R_{e,l}^{-1} \langle e_l, \theta_i \rangle \\ &= \langle e_j, \theta_i \rangle - \langle e_j, e_j \rangle R_{e,j}^{-1} \langle e_l, \theta_i \rangle = 0. \end{aligned}$$

Thus we verify equation (7) by Lemma 1.

There are other orthogonality relations I would like to highlight. First noting that  $\theta_i \in \text{span}_m\{\theta_0, u_j, j < i\}$ , then by use of (1b) and the definition of  $e_i$  one can show the following.

$$(9a) \quad \langle \theta_i, u_j \rangle = 0 \quad \text{and} \quad \langle \theta_i, v_j \rangle = 0 \quad \text{for} \quad i \leq j;$$

$$(9b) \quad \langle X_i, u_j \rangle = 0 \quad \text{and} \quad \langle X_i, v_j \rangle = 0 \quad \text{for} \quad i < j;$$

$$(9c) \quad \langle X_i, u_j \rangle = S_i^* \quad \text{and} \quad \langle X_i, v_j \rangle = R_i \quad \text{for} \quad i = j;$$

$$(9d) \quad \langle e_i, u_j \rangle = 0 \quad \text{and} \quad \langle e_i, v_j \rangle = 0 \quad \text{for} \quad i < j;$$

$$(9e) \quad \langle e_i, u_j \rangle = S_i^* \quad \text{and} \quad \langle e_i, v_j \rangle = R_i \quad \text{for} \quad i = j.$$

Now, Kalman developed a recursion to cheaply find the innovations process associated with the observations, which we will derive here. I like the way Kailath put it, “It turns out that the recursive construction of the innovations combines nicely with the recursive evolution of the state variables to give a recursion for the innovations...”[2, p. 312].

We seek a recursion for  $\{e_j\}$  but start with finding a recursion for  $\hat{\theta}_{i|i-1}$  using (7),

$$\begin{aligned} \hat{\theta}_{i+1|i} &= \sum_{j=0}^i \langle \theta_{i+1}, e_j \rangle R_{e,j}^{-1} e_j \\ &= \left( \sum_{j=0}^{i-1} \langle \theta_{i+1}, e_j \rangle R_{e,j}^{-1} e_j \right) + \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1} e_i \\ &= \hat{\theta}_{i+1|i-1} + \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1} e_i. \end{aligned}$$

The first term of the last line undermines the recursion in  $\hat{\theta}_{i|i-1}$ ; however, we can remedy this by projecting equation (1a) on  $\mathcal{L}_{d,i-1}$  to get

$$\hat{\theta}_{i+1|i-1} = F_i \hat{\theta}_{i|i-1} + G_i \hat{u}_{i|i-1} = F_i \hat{\theta}_{i|i-1} + 0$$

by (9b). This provides

$$\hat{\theta}_{i+1|i} = F_i \hat{\theta}_{i|i-1} + \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1} e_i$$

The next thing to do is to make sure we can compute  $e_i$  from the  $\hat{\theta}_{i|i-1}$ 's. If we project equation (1b) on  $\mathcal{L}_{m,i-1}$  then

$$(10) \quad \hat{X}_{i|i-1} = H_i \hat{\theta}_{i|i-1} + \hat{v}_{i|i-1} = H_i \hat{\theta}_{i|i-1} + 0$$

also by (9b). So, that

$$e_i = X_i - \hat{X}_{i|i-1} = X_i - H_i \hat{\theta}_{i|i-1}.$$

From this point I will use the abbreviated notation  $\hat{\theta}_i = \hat{\theta}_{i|i-1}$ .

Putting this together, we get a recursion for the innovations

$$(11a) \quad \begin{cases} \hat{\theta}_{i+1} = F_i \hat{\theta}_i + K_i e_i, & \hat{\theta}_0 = 0 \\ e_i = X_i - H_i \hat{\theta}_i. \end{cases}$$

where  $K_i = \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1}$ .

Here then is the recursion we sought, which recursively generates the innovations associated with the process  $X$ . It turns out that for spectral factorization we do not need the innovations themselves; we only need the  $K_i$  and  $R_{e,i}$  and would therefore expect savings if these could be computed independently from the observations. In fact this is possible and I will demonstrate how this may be accomplished.

Let  $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$  and  $P_i = \langle \tilde{\theta}_i, \tilde{\theta}_i \rangle$ . This quantity is very useful in formulating an observation-free recursion. It is easy to see by subtracting (10) from (1b) that  $e_i = H_i \tilde{\theta}_i - v_i$ , and this is enough to put  $R_{e,i}$  in terms of  $P_i$ . Observe that because  $\hat{\theta}_i \in \mathcal{L}_{d,i}$ ,  $\hat{\theta}_i$  is orthogonal to  $v_i$  by (9b). And this yields,

$$R_{e,i} = H_i P_i H_i^* + R_i.$$

To get  $K_i$  in terms of only  $P_i$  and  $R_{e,i}$ , we start with  $K_i = \langle \theta_{i+1}, e_i \rangle R_{e,i}^{-1}$  and observe that by (1b)  $\langle \theta_{i+1}, e_i \rangle = F_i \langle \theta_i, e_i \rangle + G_i \langle u_i, e_i \rangle$ . Reflecting on both inner products in turn provides

$$\begin{aligned} \langle \theta_i, e_i \rangle &= \langle \theta_i, H \hat{\theta}_i - v_i \rangle \\ &= \langle \theta_i, \tilde{X}_i \rangle H_i^* - \langle \theta_i, v_i \rangle \\ &= \langle \hat{\theta}_i + \tilde{\theta}_i, \tilde{\theta}_i \rangle H_i^* + 0 \\ &= P_i H_i^* \end{aligned}$$

and  $\langle u_i, e_i \rangle = S_i$ , as we have seen in (9e). Substituting, we are able to find that

$$K_i = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1},$$

as desired. Now, with  $R_{e,i}$  and  $K_i$  in place, a recursion for  $P_i$  can be established by combining equations (1a) and (11a) and taking the covariance of both sides, a somewhat long calculation results in the recursion,

$$(12) \quad P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_i R_{e,i} K_i^*, \quad i \geq 0$$

which we initialize by

$$P_0 = \langle \hat{\theta}_0, \hat{\theta}_0 \rangle = \langle \theta_0, \theta_0 \rangle = \Pi_0.$$

We have just found a system of recursions for the  $K_i$  and  $R_{e,i}$ , which is independent of observations. It also provides a recursion for the error covariance of the state estimate (which is useful in its own right). The discussion so far can be summarized into a theorem

**Theorem 1.** *Consider the state-space model*

$$\begin{cases} \theta_{i+1} &= F_i \theta_i + G_i u_i \\ X_i &= H_i \theta_i + v_i \end{cases}$$

for  $i \geq 0$ , and satisfying (1c). The innovations process of  $X$  can be recursively computed using the equations

$$\begin{aligned} (13a) \quad & e_i = X_i - H_i \hat{\theta}_i, \quad i \geq 0, \\ (13b) \quad & \hat{\theta}_{i+1} = F_i \hat{\theta}_i + K_i e_i, \quad \hat{\theta}_0 = 0, \end{aligned}$$

where

$$(14a) \quad P_{i+1} = F_i P_i F_i^* + G_i Q_i G_i^* - K_i R_{e,i} K_i^*, \quad P_0 = \Pi_0,$$

$$(14b) \quad R_{e,i} = H_i P_i H_i^* + R_i,$$

$$(14c) \quad K_i = (F_i P_i H_i^* + G_i S_i) R_{e,i}^{-1}.$$

Here,  $P_i = \langle \tilde{\theta}_i, \tilde{\theta}_i \rangle$  where  $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$ .

**2.4. The Stationary Case and the Modeling Filter.** So far in this section, all we have assumed about  $X$  is that it is a discrete-time stochastic process that has a finite state-space representation defined by (1). But now it is time to assume conditions native to the problem at hand, which will be delineated shortly. These assumptions are afforded by the stationarity of the processes I work with. The plan is to realize stationarity and adjust the innovations process so as to produce both stationary innovations and stationary state estimators.

Here are the further assumptions I impose on the process  $X$ ,

- the state-space representation for  $X$  is time-invariant, meaning, all the parameters are constant in time. So,

$$F_i = F, \quad G_i = G, \quad H_i = H, \quad Q_i = Q, \quad S_i = S \quad \text{and} \quad R_i = R \quad \text{for } i \geq 0$$

additionally I impose that  $F$  be stable (meaning its eigenvalues lie strictly in the unit circle) and  $R$  be strictly positive-definite,  $R > 0$ , also

- $\theta_i$  is in steady state, meaning  $\Pi_i = \Pi$  for  $i \geq 0$  where  $\Pi$  is the positive-semi-definite solution to the Lyapunov equation.

$$(15) \quad \Pi = F \Pi F^* + G Q G^*.$$

Since  $F$  is stable this has a unique solution.

A consequence of these assumptions is that  $X$  is wide-sense stationary, and its model can be written as

$$(16a) \quad \begin{cases} \theta_{i+1} = F \theta_i + G u_i \\ X_i = H \theta_i + v_i \end{cases}$$

where

$$(16c) \quad \mathbb{E} \begin{pmatrix} \theta_0 \\ u_i \\ v_i \end{pmatrix} \begin{pmatrix} \theta_0 \\ u_j \\ v_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi & 0 & 0 & 0 \\ 0 & Q \delta_{ij} & S \delta_{ij} & 0 \\ 0 & S^* \delta_{ij} & R \delta_{ij} & 0 \end{pmatrix}$$

It can be show by direct computation that the covariance function of  $X$ ,  $R_X(i-j)$ , can be written as follows [2, p. 267]

$$(17) \quad R_X(i-j) = \begin{cases} H F^{i-j-1} N & i > j, \\ R + H \Pi H^* & i = j, \\ N^* (F^*)^{j-i-1} H^* & i < j, \end{cases}$$

where  $N = F \Pi H^* + G S$ .



Under these conditions we now apply Theorem 1 and find that the innovations sequence associated with  $X$  has the following state-space representation

$$(18a) \quad e_i = X_i - H\hat{\theta}_i, \quad i \geq 0,$$

$$(18b) \quad \hat{\theta}_{i+1} = F\hat{\theta}_i + K_i e_i, \quad \hat{\theta}_0 = 0,$$

where now

$$(19a) \quad P_{i+1} = FP_iF^* + GQG^* - K_i R_{e,i} K_i^*, \quad P_0 = \Pi$$

$$(19b) \quad R_{e,i} = HP_iH^* + R$$

$$(19c) \quad K_i = (FP_iH^* + GS)R_{e,i}^{-1}$$

Rearranging (18) slightly we can write

$$(20a) \quad \begin{cases} \hat{\theta}_{i+1} = F\hat{\theta}_i + K_i e_i, & \hat{\theta}_0 = 0, \\ X_i = H\hat{\theta}_i + e_i, & i \geq 0. \end{cases}$$

Which provides an alternative way of modeling the process  $X$ , this time driven by the innovations process  $e$ . Said differently (20) defines a system  $\mathcal{L}$  which maps  $e$  to  $X$ . The system  $\mathcal{L}$  is sometimes referred to as a modeling filter since given the noise sequence  $e$  it will recover  $X$ , exactly.

Observe that  $\mathcal{L}$ , clearly causal, is also causally invertible since by algebra we can write,

$$(21a) \quad \begin{cases} \hat{\theta}_{i+1} = (F - KH)\hat{\theta}_i + K_i X_i, & \hat{\theta}_0 = 0, \\ e_i = -H\hat{\theta}_i + X_i, & i \geq 0. \end{cases}$$

This contrasts with the original model which took  $(\theta_0, u, v)$  to  $X^7$ , such is causal but not generally causally invertible.

Now, what can be said about the output  $\mathcal{L}w$ , for  $w = (w_i, i = 0, 1, 2, \dots)$  a noise sequence with  $\langle w_i, w_j \rangle = R_{e,i} \delta_{ij}$ ? Is it a stationary process with the same  $z$ -spectrum as  $X$ ? No, generally the output is not even stationary, since the model, found under (20), represents a time-variant system. This begs the question, does this model converge, in time, to a time-invariant model? It turns out that the system parameters  $P_i$ ,  $R_{e,i}$ , and  $K_i$  all converge as  $i \rightarrow \infty$ . Which suggests a time-invariant causal system that, when fed in a stationary input, will certainly produce a stationary output. A *modeling filter* for  $X$  is a causal, linear system which constructs the process  $X$  by passing a white-noise process through the filter. This is what we are seeking.

The convergence of  $P_i$  has been well studied [2, Sec. 14.5] and there are many conditions one can check to ensure that the limit of the recursion converges to the stabilizing solution of the appropriate Ricatti equation necessary to provide the causal and casually invertible modeling filter, namely the discrete algebraic Riccati equation (DARE), given by

$$(22) \quad P = FPF^* + GQG^* - (FPH^* + GS)(HPH^* + R)^{-1}(HPF^* + S^*G^*)$$

$$(23) \quad = FPF^* + GQG^* - KR_eK^*$$

<sup>7</sup>One thing it has done was remove the additional input variable, the initial state, setting it to zero

Since this theory is more problem dependent I will defer the matter of convergence to when we have a specific problem.

So, for now, assume that

$$\lim_{i \rightarrow \infty} P_i = P$$

So that  $P$  is the unique solution to

$$(24) \quad P = FPF^* + GQG^* - KR_eK^*$$

where  $R_e = \lim R_{e,i} = HPH + R$  and  $K = \lim K_i = (FPH^* + GS)R_e^{-1}$ , and that  $P$  is such that  $F - KH$  is stable. Let us, then, consider the system

$$L : w \mapsto Y$$

where  $w = (w_i, i \geq -\infty)$  with  $\langle w_i, w_j \rangle = R_e \delta_{ij}$  and  $Y = (Y_i, i \geq -\infty)$ <sup>8</sup> given by

$$(25a) \quad \begin{cases} \eta_{i+1} = F\eta_i + Kw_i, & \eta_{-\infty} = 0 \\ (25b) \quad Y_i = H\eta_i + w_i, & i \geq -\infty. \end{cases}$$

Since  $\eta_i$  was initialized in the remote past (independent of  $w_i$ ) we have  $\langle \eta_i, w_i \rangle = 0$  for  $i > -\infty$  and  $\langle \eta_i, \eta_i \rangle = \Sigma$  for  $i > -\infty$  where

$$(25c) \quad \Sigma = F\Sigma F^* + KR_eK^*.$$

We want to check that the autocovariance of  $Y$  is the same as for  $X$  which is given in (17). This can be checked directly. For  $i > 0$

$$\begin{aligned} R_Y[i] &= \langle Y_i, Y_0 \rangle \\ &= \langle H\eta_i + w_i, H\eta_0 + w_0 \rangle \\ &= \langle HF\eta_{i-1} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\ &= \langle HF^2\eta_{i-2} + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\ &= \langle HF^i\eta_0 + HF^{i-1}Kw_0 + \dots + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 + w_0 \rangle \\ &= \langle HF^i\eta_0, H\eta_0 \rangle + \langle HF^{i-1}Kw_0 + \dots + HFKw_{i-2} + HKw_{i-1} + w_i, H\eta_0 \rangle \\ &\quad + \langle HF^i\eta_0, w_0 \rangle + \langle HF^{i-1}Kw_0 + \dots + HFKw_{i-2} + HKw_{i-1} + w_i, w_0 \rangle \\ &= HF^i\Sigma H^* + 0 + 0 + HF^{i-1}KR_e \\ &= HF^i\Sigma H^* + HF^{i-1}(FPH^* + GS) \\ &= HF^{i-1}(F(\Sigma + P)H^* + GS), \end{aligned}$$

a similar calculation for  $i < 0$  yields

$$R_Y[i] = (F(\Sigma + P)H^* + GS)^*(F^*)^{i-1}H^* = R_Y^*[-i],$$

<sup>8</sup>Notice the slight-of-hand, we now start our processes in the remote past and have them be bilaterally infinite. This is to obviate the need for the state variable to be initialized as an input in order for them to have a variance that obeys the Lyapunov equation below for each finite time

and for  $i = 0$

$$\begin{aligned}
R_Y(0) &= \langle Y_0, Y_0 \rangle \\
&= \langle H\eta_0 + w_0, H\eta_0 + w_0 \rangle \\
&= H\Sigma H^* + R_e \\
&= H\Sigma H^* + HPH^* + R \\
&= H(\Sigma + P)H^* + R
\end{aligned}$$

Now what is  $\Sigma + P$ , by there definitions

$$\begin{aligned}
\Sigma + P &= F\Sigma F^* + KR_eK^* + FPF^* + GQG^* - KR_eK^* \\
&= F(\Sigma + P)F^* + GQG^*.
\end{aligned}$$

Meaning that  $\Sigma + P$  solves the same discrete Lyapunov equation that  $\Pi$  solved in (15). We therefore conclude

$$\Pi = \Sigma + P$$

and that

$$(26) \quad R_Y[i] = \begin{cases} HF^{i-1}N & i > 0, \\ R + H\Pi H^* & i = 0, \\ N^*(F^*)^{-i-1}H^* & i < 0, \end{cases} = R_X[i]$$

recalling that  $N = F\Pi H^* + GS$ , as in equation (17). The system  $L$  maps *any* white-noise sequence with variance  $R_e$  to a stationary process with a given  $z$ -spectrum. This is what I refer to as a *weak* modeling filter.

Furthermore, observe that  $L$  (given by (25)) is causal and causally invertible, since, by simple algebra again,  $L^{-1}$  can be written as

$$\begin{aligned}
(27a) \quad & \begin{cases} \eta_{i+1} = (F - KH)\eta_i + KY_i, & \eta_{-\infty} = 0, \\ w_i = -H\eta_i + Y_i, & i \geq -\infty. \end{cases} \\
(27b) \quad &
\end{aligned}$$

where  $F - KH$  which is stable since  $P$  was a stabilizing solution to the Ricatti equation given under (24). Given  $Y$  we recover  $w$ , so,  $L^{-1}$  can be regarded as a whitening filter for  $Y$ .

I now summarize the above discussion into a theorem

**Theorem 2.** *Let  $X = (X_i; i \geq 0)$  be a stationary process with a time-invariant state-space representation of the form given in (16). Then, the LTI system  $L$  given by (25) is causal and causally invertible, and constructs a stationary process with the same autocovariance as  $X$  by passing in any white-noise process with variance  $R_e$ .*

### 3. SPECTRAL FACTORIZATION BY KALMAN FILTERING

In this section I explain how the above results from Kalman filtering may be employed in finding an approximate spectral factor of a given power spectrum ( $z$ -spectrum).

Given a  $z$ -spectrum  $S_X(z)$  of a WSS discrete-time process  $X_n \in \mathbb{C}^d$  for  $n \geq 0$ , if the decay of the covariance is sufficiently fast it is reasonable to truncate  $S_X(z)$  to a Laurent polynomial

$$\tilde{S}_X(z) = \sum_{n=-m}^m R_X[n]z^{-n}.$$

However, this isn't always the best way to approximate the spectrum. It is usual to apply some windowing function to the autocovariances. The Parzen windowing function  $\Lambda_m$ , given below, for some  $m > 0$  is an example.

$$\Lambda_m[n] = \begin{cases} 1 - 6\left(\frac{n}{m}\right)^2 + 6\left(\frac{|n|}{m}\right)^3, & |n| \leq m/2 \\ 2\left(1 - \frac{|n|}{m}\right)^3, & m/2 < |n| \leq m \\ 0, & |n| > m \end{cases}.$$

The Fourier transform  $\hat{\Lambda}_m(\omega)$  of the Parzen windowing function is

$$\hat{\Lambda}_m(\omega) = \frac{3}{4}m \left( \frac{\sin \pi\omega m/2}{\pi\omega m/2} \right)^4, \quad -\infty \leq \omega \leq \infty$$

Let  $R_{\text{win}}[n] = \Lambda_m[n]R_X[n]$  and

$$S_{\text{win}}(z) = \sum_{n=-m}^m R_{\text{win}}[n]z^{-n}.$$

The function  $S_{\text{win}}(z)$  is generally a better approximation of  $S_X(z)$  than  $\tilde{S}_X(z)$ .

Now that we have approximated, by a Laurent polynomial, the  $z$ -spectrum we seek to factor, we can construct a process  $Y = (Y_i, i = 0, 1, 2, \dots)$  with a finite, time-invariant, state-space model whose  $z$ -spectrum is exactly  $S_{\text{win}}(z)$ , that is  $S_Y(z) = S_{\text{win}}(z)$  [1, p. 488]. We then take the state-space model and using Theorem 2, compute the weak modeling filter.

The process  $Y$  is constructed using the following finite, time-invariant, state-space model:

$$\begin{aligned} (28a) \quad & \theta_{i+1} = F\theta_i + Gv_i \\ (28b) \quad & Y_i = H\theta_i + u_i \end{aligned}$$

Where<sup>9</sup>

$$(28c) \quad \mathbb{E} \begin{pmatrix} \theta_0 \\ v_i \\ u_i \end{pmatrix} \begin{pmatrix} \theta_0 \\ v_j \\ u_j \\ 1 \end{pmatrix}^* = \begin{pmatrix} \Pi & 0 & 0 & 0 \\ 0 & R\delta_{ij} & S\delta_{ij} & 0 \\ 0 & S^*\delta_{ij} & Q\delta_{ij} & 0 \end{pmatrix}.$$

<sup>9</sup>Recall  $d$  is the dimension of the observation vectors and  $m$  is the degree of the Laurent polynomial approximation of the given  $z$ -spectrum

The parameters  $F$ ,  $H$ , and  $N$  (think of the  $N$  from equation (17)), are defined explicitly by

$$F = \begin{pmatrix} 0 & & & & \\ I & 0 & & & \\ & I & 0 & & \\ & & \ddots & \ddots & \\ & & & I & 0 \end{pmatrix} \in \mathbb{C}^{md \times md},$$

$$H = (0 \quad \dots \quad 0 \quad I) \in \mathbb{C}^{d \times md},$$

and

$$N = \begin{pmatrix} R_{\text{win}}[m] \\ R_{\text{win}}[m-1] \\ \vdots \\ R_{\text{win}}[1] \end{pmatrix} \in \mathbb{C}^{md \times d}.$$

The remaining quantities are less important, almost auxiliary. With the exception of  $\Pi$ , they have no bearing on the covariance data of the process (recall equation (17)). The existence of  $\Pi$  is guaranteed by the fact that  $F$  is stable. It's value never needs to be determined, as we will see. For completeness if one were interested in actually realizing the process  $Y$  from this statespace form the remaining variables must satisfy the following conditions

$$\Pi = F\Pi F^* + GQG^*$$

$$GS = N - F\Pi H^*$$

$$R = R_{\text{win}}[0] - H\Pi H^*$$

Now, provided of all the above can be satisfied it can be show that  $Y$  is stationary and its autocovariance is

$$C_Y[i] = \begin{cases} HF^{i-1}N & i > 0 \\ H\Pi H^* + R & i = 0 \\ NF^{*(-i-1)}H^* & i < 0 \end{cases}$$

$$= \begin{cases} 0 & i > m \\ R_{\text{win}}[i] & m \geq i \geq -m \\ 0 & i < -m \end{cases}$$

So, that  $S_Y(z) = S_{\text{win}}(z) \approx S_X(z)$  as desired.

In  $Y$ , we have a stationary process with a time-invariant state-space representation, given under (28). We now invoke Theorem 2, and assert that the system  $M : w = (w_i, i > -\infty) \mapsto Y' = (Y'_i, i > -\infty)$  given by

$$(29a) \quad \begin{cases} \eta_{i+1} = F\eta_i + Kw_i, & \eta_{-\infty} = 0 \\ (29b) \quad Y'_i = H\eta_i + w_i, & i \geq -\infty, \end{cases}$$

is causal and causally invertible and constructs a stationary process  $Y'$  with the same autocovariance of  $Y$  by passing in any white-noise process with variance  $R_e$ . This system  $M$  constitutes a weak modeling filter for  $Y$ . Observe that from System (29) it can be shown

that

$$\begin{aligned}
 Y'_i &= HF\eta_{i-1} + HKw_{i-1} + Kw_i \\
 &= HF^2\eta_{i-2} + HF^2Kw_{i-2} + HKw_{i-1} + Kw_i \\
 &= HF^{m+1}\eta_{i-m-1} + HF^mKw_{i-m-1} + HF^{m-1}Kw_{i-m} + \cdots + HKw_{i-1} + Kw_i \\
 &= 0 + 0 + \sum_{j=1}^m HF^{j-1}Kw_{i-j} + w_i \\
 (30) \quad &=: \sum_{j=0}^m K_j^\# w_{i-j}
 \end{aligned}$$

where

$$K_j^\# = \begin{cases} K_{m-j+1} & 1 \leq j \leq m \\ I & j = 0 \end{cases}$$

The notation views  $K \in C^{md \times d}$  as a column vector with matrix entries,  $(d \times d)$ -blocks. So in words, when  $K \in C^{md \times d}$  is viewed as a block column vector with  $d \times d$ -blocks,  $K^\#$  is the block-reverse of  $K$  appended on top by the  $d \times d$  identity matrix.

From equation (30), since  $w$  is white-noise it is clear that system  $M$  is actually nothing more than a VMA( $m$ ) (model) process. It can be shown that the impulse response of  $M$  is  $\{K_i^\#\}$  (zero everywhere but  $i = 0, 1, \dots, m$ ), and that

$$Y'_i = \sum_{j=-\infty}^{\infty} K_j^\# w_{i-j} = (K^\# * w)_i.$$

Which means

$$S_X(z) \approx S_{\text{win}}(z) = S_Y(z) = S_{Y'}(z) = S_{K^\# * w}(z) = \bar{L}(z)R_e\bar{L}^*(z^{-*})$$

where

$$\bar{L}(z) = \sum_{n=-\infty}^{\infty} K_n^\# z^{-n},$$

is the  $z$ -transform of  $K^\#$  and the transfer function of  $M$ . To get  $L(z)$  take  $L(z) = \bar{L}(z)R_e^{-1/2}$ . This is tantamount to letting  $\ell = (K_i^\# R_e^{-1/2}, i = 0, 1, \dots, m)$ , for  $\ell$ , the impulse response of the system associated with  $L(z)$ . And with that we have,

$$S_X(z) \approx S_{\text{win}}(z) = L(z)L^*(z^{-*})$$

as promised.

**3.1. Algorithm and input-output analysis.** I organized the above procedure as Algorithm 1. The algorithm takes in the first  $m + 1$  nonnegative lags of the windowed autocovariance sequence. Observe that the initial data is not iterate but defines a parameter in an iterative scheme that is initialized by a constant matrix, the iterate here is the covariance of the state estimation error  $P$ .

Intuitively, we can think of the iterations starting with the causal part of the approximated  $z$ -spectrum, denoted and defined as

$$\{S_{\text{win}}(z)\}_+ = \sum_{n=0}^m R_{\text{win}}[n]z^{-n}$$

The algorithm iteratively refines corrections to this the coefficients are optimal linear predictions of the next state variable given the current innovation. To explain, recall that

$$K_i = K_{p,i}R_{e,i}^{-1} = \langle \theta_{i+1}, e_i \rangle.$$

Theses coefficients, as (30) illustrates, are the essence of the model filter, and they are recursively obtained from the statespace model which includes the covariance data of the original timeseries.

---

**Algorithm 1** Spectral Factorization by direct Kalman filter

---

1: **procedure** KALMANSPECFAC( $C_{\text{win}}[0], C_{\text{win}}[1], \dots, C_{\text{win}}[m]$ )

**Initialize variables**

2:      $F \leftarrow \begin{pmatrix} 0 & & & \\ I & 0 & & \\ & \ddots & \ddots & \\ & & I & 0 \end{pmatrix}$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)

3:      $H \leftarrow (0 \quad \dots \quad 0 \quad I)$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)

4:      $G \leftarrow I_{md \times md}$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (need not be used at all)

5:      $Q \leftarrow I_{md \times md}$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)

6:      $R \leftarrow C_{\text{win}}[0] - mI$   $\triangleright \in \mathbb{C}^{d \times d}$

7:      $S \leftarrow \text{col}(C_{\text{win}}[m-i], i = 0, 1, \dots, m-1)$   $\triangleright \in \mathbb{C}^{dm \times d}$

8:      $P \leftarrow \begin{pmatrix} I & & & \\ & 2I & & \\ & & \ddots & \\ & & & mI \end{pmatrix}$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)

**Run Kalman recursion**

9:     Assign  $N > 0$

10:    **for**  $i = 1$  to  $N$  **do**

11:        $R_e = R + HPH^*$

12:        $K = FPH^* + GS$

13:        $P = FPF^* + GQG^* - KR_e^{-1}K^*$

**Extract spectral factor coefficients**

14:      $K \leftarrow K/R_e$

15:      $R_e \leftarrow \sqrt{R_e}$

16:      $\ell \leftarrow \text{zeros}(d, d, m+1)$

17:     **for**  $i = 0$  to  $m$  **do**

18:        $\ell[:, :, m-i+1] \leftarrow K[di+1 : d(i+1), :]R_e$

**return**  $\ell$

---

## 4. AN EXAMPLE

Consider that an exponentially-correlated random process  $X$  with  $C_X[z] = a^{|z|}$  for  $0 < a < 1$ . The  $z$ -spectrum can easily computed as being

$$S_x(z) = \frac{1 - a^2}{(1 - az^{-1})(1 - az)},$$

with a spectral factor of  $L(z) = \frac{\sqrt{1-a^2}}{1-az^{-1}}$ . Now let us use Algorithm 1. The input data. Is simply the first  $m = 50$  terms of the autocovariance sequence  $a = 0.9$ , together with the number of iterations  $N = 0, 1, 2, 4$ .

Figure 1 show the causal part of the  $z$ -spectrum together with the analytic and approximated spectral factor.

## APPENDIX A. CKMS

CKMS, named for Chandrasekhar, Kailath, Morf, and Sidhu who in various ways and times contributed to its development, is an efficient algorithm for computing the Kalman filter recursion for innovations. I describe this briefly now. Kailath, Morf and Sidhu (1973) observed that though  $P_i$  is full rank the change for one step to the next  $\delta P_i := P_{i+1} - P_i$  can have low rank. So write,

$$\delta P_i = L_i M_i L_i^*$$

It can be shown that if  $\delta P_0 = L_0 M_0 L_0^*$  with  $M_0$  hermitian, nonsingular, and of size  $\alpha \times \alpha$ , then for  $i > 0$ ,  $\delta P_i = L_i M_i L_i^*$  with  $M_i$  also hermitian, nonsingular, and of size  $\alpha \times \alpha$ . Exploiting this fact Kailath et al. provided the following theorem [1, p. 409].

**Theorem 3** (The Fast (CKMS) Kalman Recursions ). *The  $K_i$  and  $R_{e,i}$  from the Kalman recursion above can be recursively computed by the following set of coupled recursions, for  $i \geq 0$*

$$\begin{aligned} K_{i+1} &= K_i - F L_i R_{r,i}^{-1} L_i^* H^* \\ L_{i+1} &= F L_i - K_i R_{e,i}^{-1} H L_i \\ R_{e,i+1} &= R_{e,i} - H L_i R_{r,i}^{-1} L_i^* H^* \\ R_{r,i+1} &= R_{r,i} - L_i^* H^* R_{e,i}^{-1} H L_i \end{aligned}$$

*The recursion is initialized as follows:  $K_0 = F \Pi_0 H^* + G S$  and  $R_{e,0} = R + H \Pi_0 H^*$ . Then factor get  $L_0$  and  $R_{r,0}$*

$$\delta P_0 := F \Pi_0 F^* + G Q G^* - K_0 R_{e,0}^{-1} K_0^* - \Pi_0 =: -L_0 R_{r,0}^{-1} L_0^*$$

*where  $L_0$  is  $m \times \alpha$  and  $R_{r,0}$  is  $\alpha \times \alpha$ , nonsingular and Hermitian.*

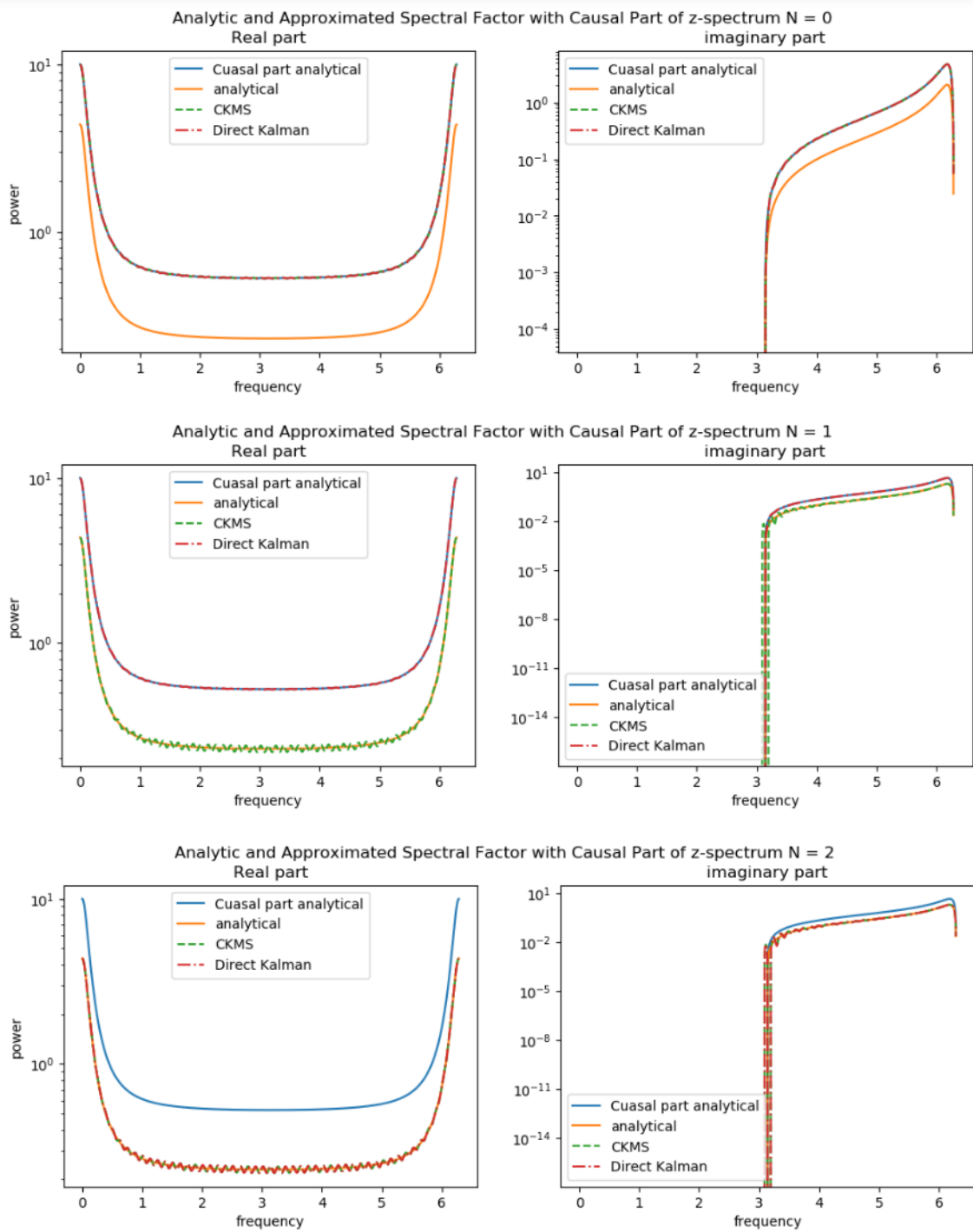
An implementation of this method is summarized in Algorithm 2.

## APPENDIX B. PROOF OF LEMMA 1

Lemma 1 has been restated for convenience.



FIGURE 1. The causal part of the  $z$ -spectrum together with the analytic and approximated spectral factor



**Lemma.** Given the inner product defined under (2) and the norm induced by it, under (3), let  $x \in C^d$  and  $y_i \in \mathbb{C}^m$  for  $i = 0, \dots, N$  be random variables. If  $\hat{x} \in \text{span}_d\{y_i, i = 0, 1, \dots, N\}$

**Algorithm 2** Spectral Factorization by CKMS

---

1: **procedure** KALMANSPECFAC( $C_{\text{win}}[0], C_{\text{win}}[1], \dots, C_{\text{win}}[m]$ )

**Initialize variables**

2:  $F \leftarrow \begin{pmatrix} 0 & & & \\ I & 0 & & \\ & \ddots & \ddots & \\ & & I & 0 \end{pmatrix}$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)  
3:  $H \leftarrow (0 \quad \dots \quad 0 \quad I)$   $\triangleright \in \mathbb{C}^{dm \times dm}$  (use sparse array)  
4:  $K \leftarrow \text{col}(C_{\text{win}}[m-i], i = 0, 1, \dots, m-1)$   $\triangleright \in \mathbb{C}^{dm \times d}$   
5:  $L \leftarrow \text{col}(C_{\text{win}}[m-i], i = 0, 1, \dots, m-1)$   $\triangleright \in \mathbb{C}^{dm \times d}$   
6:  $R_e \leftarrow C_{\text{win}}[0]$   $\triangleright \in \mathbb{C}^{d \times d}$   
7:  $R_r \leftarrow C_{\text{win}}[0]$   $\triangleright \in \mathbb{C}^{d \times d}$ 
**Run CKMS recursion**

8: Assign  $N > 0$   
9: **for**  $i = 1$  to  $N$  **do**  
10:  $K_{\text{new}} \leftarrow K - FLR_r^{-1}L^*H^*$   
11:  $L_{\text{new}} \leftarrow FL - KR_e^{-1}hL$   
12:  $R_{e,\text{new}} \leftarrow R_e - HLR_r^{-1}L^*H^*$   
13:  $R_{r,\text{new}} \leftarrow R_r - L^*H^*R_e^{-1}HL$   
14:  $K \leftarrow K_{\text{new}}$   
15:  $L \leftarrow L_{\text{new}}$   
16:  $R_e \leftarrow R_{e,\text{new}}$   
17:  $R_r \leftarrow R_{r,\text{new}}$ 
**Extract spectral factor coefficients**

18:  $K \leftarrow K/R_e$   
19:  $R_e \leftarrow \sqrt{R_e}$   
20:  $\ell \leftarrow \text{zeros}(d, d, m+1)$   
21: **for**  $i = 0$  to  $m$  **do**  
22:  $\ell[:, :, m-i+1] \leftarrow K[di+1 : d(i+1), :]R_e$   
**return**  $\ell$ 


---

is chosen such that

$$(31) \quad (x - \hat{x}) \perp y_i \quad \text{for} \quad i = 0, 1, \dots, N,$$

then  $\hat{x}$  is the linear LMSE for  $x$  given  $\{y_i, i = 0, \dots, N\}$ , meaning

$$\mathbb{E}(x - \hat{x})(x - \hat{x})^* = \|x - \hat{x}\|^2 = \inf_{x' \in \text{span}_d\{y_i, \forall i\}} \|x - x'\|^2.$$

*Proof.* Let  $x' \in \text{span}_d\{y_i, \forall i\}$  and observe that since  $\hat{x} - x' \in \text{span}_d\{y_i, \forall i\}$ . The orthogonality assumption amounts to  $x - \hat{x}$  being orthogonal to every element in  $\text{span}_d\{y_i, \forall i\}$  and

so  $(x - \hat{x}) \perp (\hat{x} - x')$ . Then we can write

$$\|x - x'\|^2 = \|x - \hat{x} + \hat{x} - x'\|^2 = \|x - \hat{x}\|^2 + \|\hat{x} - x'\|^2 \geq \|x - \hat{x}\|^2$$

which completes the proof.  $\square$

## APPENDIX C. ILLUSTRATION

Consider this iterative algorithm for finding the square root of a positive number  $x$ .

---

### Algorithm 3 Square Root (Analogue of CKMS)

---

```

1: procedure MYSQRT1( $x$ )
2:    $x_0 \leftarrow x$ 
3:   Assign  $N > 0$ 
4:   for  $i = 1$  to  $N$  do
5:      $x \leftarrow \frac{1}{2} \left( x + \frac{x_0}{x} \right)$ 
   return  $x$ 

```

---



---

### Algorithm 4 Square Root (Analogue of Direct Kalman)

---

```

1: procedure MYSQRT2( $x$ )
2:    $x_0 \leftarrow x$ 
3:    $x \leftarrow 1$ 
4:   Assign  $N > 0$ 
5:   for  $i = 1$  to  $N$  do
6:      $x \leftarrow \frac{1}{2} \left( x + \frac{x_0}{x} \right)$ 
   return  $x$ 

```

---

## REFERENCES

- [1] Ali H Sayed and Thomas Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.
- [2] Thomas Kailath, Ali H Sayed, and Babak Hassibi. *Linear estimation*. Number BOOK. Prentice Hall, 2000.