2016

An inner-outer factorization in ℓp with applications to ARMA processes

Raymond Cheng

William T. Ross
University of Richmond, wross@richmond.edu

AN INNER-OUTER FACTORIZATION IN ℓ^p WITH APPLICATIONS TO ARMA PROCESSES

RAYMOND CHENG AND WILLIAM T. ROSS

Abstract. The following inner-outer type factorization is obtained for the sequence space ℓ^p: if the complex sequence $\mathbf{F} = (F_0, F_1, F_2, \ldots)$ decays geometrically, then for any p sufficiently close to 2 there exist $\mathbf{J}$ and $\mathbf{G}$ in ℓ^p such that $\mathbf{F} = \mathbf{J} * \mathbf{G}$; $\mathbf{J}$ is orthogonal in the Birkhoff-James sense to all of its forward shifts $S\mathbf{J}, S^2\mathbf{J}, S^3\mathbf{J}, \ldots$; $\mathbf{J}$ and $\mathbf{F}$ generate the same S-invariant subspace of ℓ^p; and $\mathbf{G}$ is a cyclic vector for S on ℓ^p.

These ideas are used to show that an ARMA equation with characteristic roots inside and outside of the unit circle has Symmetric-α-Stable solutions, in which the process and the given white noise are expressed as causal moving averages of a related i.i.d. SαS white noise. An autoregressive representation of the process is similarly obtained.

1. Introduction

In this paper we begin to examine, under certain circumstances, a possible "inner-outer" factorization for the class ℓ^p_A of analytic functions f on the open unit disk $\mathbb{D}$ whose Taylor coefficients belong to the classical sequence space ℓ^p. When $p = 2$, the sequence space ℓ^2_A is the classical Hardy space H^2, and a theorem of Beurling says that every $f \in \ell^2_A$ can be factored, uniquely up to unimodular constants, as

$$f = JG, \tag{1.1}$$

where J is an inner function and G is an outer function. More specifically, the function J can be obtained by the formula $J = f - \widehat{f}$, where $\widehat{f}$ is the orthogonal projection of f onto $S[f]$, equivalently

$$J \perp S^k J, \quad k \geq 1. \tag{1.2}$$

In the above, S is the unilateral shift operator $Sg = zg$ on ℓ^2_A and $[f]$ is the S-invariant subspace generated by f. Furthermore, the function

Date: September 1, 2016.

2010 *Mathematics Subject Classification.* 46A45, 46B25, 46E15, 60G12, 60G52.

Key words and phrases. inner-outer factorization, Hardy class, invariant subspace, time series, autoregressive, moving average, stable process.

G satisfies $[G] = \ell^2_A$, i.e., G is a cyclic vector for S, and the function J satisfies $[J] = [f]$. The classic text [13] provides a full account of all this.

When $p \neq 2$, the invariant subspaces of ℓ^p_A can be quite complicated and, especially when $p > 2$, can have many dramatic pathological properties [1]. In short, the S-invariant subspaces for $p \neq 2$ are not understood much at all. Even when $p = 1$, where ℓ^1_A is an algebra of continuous functions on the closure $\overline{\mathbb{D}}$ of $\mathbb{D}$ (known as the Wiener algebra), the S-invariant subspaces, which turn out to be the ideals of ℓ^1_A, still have delicate, but pathological, properties [18].

In this paper, we obtain a partial Beurling-type factorization as in (1.1) by replacing the standard Hilbert space orthogonality in ℓ^2_A, used in (1.2) to define the inner factor J, by the Birkhoff-James orthogonality relationship $\perp_p$ in ℓ^p_A (see (3.2) below). Our main theorem is the following.

Theorem 1.3. *If f is analytic in a neighborhood of $\overline{\mathbb{D}}$, then for any p sufficiently close to 2, there exist functions G and J, analytic in a neighborhood of $\overline{\mathbb{D}}$, such that*

$$f = JG$$

where G is cyclic vector for S on ℓ^p_A and J satisfies

$$J \perp_p S^k J \quad k = 1, 2, 3, \ldots$$

Furthermore, we have $[f] = [J]$ in ℓ^p_A. The functions J and G are unique up to multiplicative constants.

As with Beurling's theorem, notice that we have $J \perp_p S^k J$ for all $k \geq 1$, which is an equivalent to J being inner when $p = 2$, and $[G] = \ell^p_A$, which is equivalent to G being outer when $p = 2$. One might expect that "inner" in this situation depends on p. Indeed, the "inner" factor J when $p = 2$ is actually an inner function in the classical sense (i.e., a bounded analytic function on $\mathbb{D}$ with unimodular radial boundary values almost everywhere on the circle $\mathbb{T}$), up to a constant factor. When $p \neq 2$, one can have the condition $J \perp_p S^k J$ for all $k \geq 1$ but without J being inner in the classical sense (see the examples at the ends of Sections 3 and 5, respectively).

The next section sets forth the notation used in this paper and reviews the related function theory. Section 3 contains the development of the main theorem. The analytical tools derived in Section 3 are applied in Section 4 to solve a problem concerning Autoregressive Moving Average (ARMA) processes. It is shown that an infinite-variance

ARMA model has a causal stationary solution, even if its characteristic polynomials have roots both inside and outside the unit circle.

The modest ℓ^p factorization given in Theorem 1.3 is sufficient to solve the intended application to ARMA processes. However, one can see that there is much work to be done to obtain, if possible, a general "inner-outer" factorization for all $f \in \ell^p_A$, not necessarily smooth up to the boundary. We invite the reader to join in the discussion.

2. Preliminaries

For $1 \leq p < \infty$ define ℓ^p to be the set of sequences

$$\mathbf{a} = (a_0, a_1, \ldots)$$

of complex numbers for which

$$\|\mathbf{a}\|_p := \left(\sum_{k=0}^{\infty} |a_k|^p\right)^{1/p} < \infty.$$

When $1 \leq p < \infty$, the quantity $\|\mathbf{a}\|_p$ defines a norm on ℓ^p which makes ℓ^p a Banach space. In this paper, we will use the fact that when $1 < p < \infty$, the space ℓ^p is both uniformly convex and smooth [8].

For an $\mathbf{a} \in \ell^p$ we will set

$$a(z) = \sum_{k=0}^{\infty} a_k z^k \tag{2.1}$$

to be the power series whose Taylor coefficients are $\mathbf{a}$. Note the use of $\mathbf{a}$ (bold faced) to represent a sequence and a (not bold faced) to represent the corresponding power series. By Hölder's inequality we see that if p' denotes the usual conjugate index, i.e., $1/p + 1/p' = 1$, then

$$\begin{aligned} \sum_{k=0}^{\infty} |a_k||z^k| &\leq \left(\sum_{k=0}^{\infty} |a_k|^p\right)^{1/p} \left(\sum_{k=0}^{\infty} |z|^{kp'}\right)^{1/p'} \\ &= \|\mathbf{a}\|_p \left(\frac{1}{1-|z|^{p'}}\right)^{1/p'}. \end{aligned}$$

This implies that the above power series a determines an analytic function on the open unit disk $\mathbb{D}$. If we define

$$\ell^p_A = \{a : \mathbf{a} \in \ell^p\}$$

then, norming a by $\|\mathbf{a}\|_p$, ℓ^p_A becomes a Banach space of analytic functions on $\mathbb{D}$. Furthermore, for each $z \in \mathbb{D}$ and $a \in \ell^p_A$ we have

$$|a(z)| \leq \|\mathbf{a}\|_p \left(\frac{1}{1-|z|^{p'}}\right)^{1/p'} \tag{2.2}$$

and so if a sequence of functions converges in the norm of ℓ^p_A then it converges uniformly on compact subsets of $\mathbb{D}$.

For two sequences $\mathbf{a}$ and $\mathbf{b}$ the convolution $\mathbf{a} * \mathbf{b}$ is the sequence

$$\left\{ \sum_{k=0}^{n} a_k b_{n-k} \right\}_{n \geq 0}.$$

By multiplying Taylor series coefficients, notice how $\mathbf{a} * \mathbf{b}$ corresponds via (2.1) to the product $a(z)b(z)$. The space ℓ^p is convolution algebra (i.e., $\mathbf{a}*\mathbf{b} \in \ell^p$ whenever $\mathbf{a}, \mathbf{b} \in \ell^p$) only when $p = 1$, and ℓ^1 is called the Wiener algebra. From here note that the corresponding function space ℓ^1_A is a Banach algebra of analytic functions on $\mathbb{D}$ which are continuous on $\overline{\mathbb{D}}$. Also observe that $\ell^1 \subset \ell^p$ for all $p \geq 1$ and recall Young's inequality [22, p. 37]

$$\|\mathbf{a} * \mathbf{b}\|_p \leq \|\mathbf{a}\|_p \|\mathbf{b}\|_1, \quad \mathbf{a} \in \ell^p, \mathbf{b} \in \ell^1. \tag{2.3}$$

For $1 \leq p < \infty$ the classical Hardy space H^p consists of the analytic functions f on $\mathbb{D}$ for which

$$\|f\|_{H^p} = \left(\sup_{0<r<1} \int_{\mathbb{T}} |f(re^{i\theta})|^p d\lambda(e^{i\theta}) \right)^{1/p} < \infty,$$

where $d\lambda$ is normalized Lebesgue measure on the unit circle $\mathbb{T}$. Functions in H^p are known to have radial limits

$$f(e^{i\theta}) := \lim_{r \to 1^-} f(re^{i\theta}), \quad \lambda\text{-a.e.}$$

and the corresponding boundary function belongs to $L^p(\mathbb{T}, \lambda)$. When $p = 2$, Parseval's theorem shows that H^2 is a Hilbert space with the standard $L^2(\mathbb{T}, \lambda)$ inner product.

We define the *forward shift* operator

$$S : \ell^p \to \ell^p, \quad S\mathbf{a} = (0, a_0, a_1, a_2, \ldots)$$

and observe that S is an isometry on ℓ^p. For $\mathbf{a} \in \ell^p$, let $[\mathbf{a}]$ denote the S-invariant subspace generated by $\mathbf{a}$, that is,

$$[\mathbf{a}] = \bigvee \{\mathbf{a}, S\mathbf{a}, S^2\mathbf{a}, \ldots\},$$

where $\bigvee$ denotes the closed linear span in ℓ^p. A vector $\mathbf{a} \in \ell^p$ is said to be *cyclic* if $[\mathbf{a}] = \ell^p$. It will be useful to write $\mathbf{I} = (1, 0, 0, 0, \ldots)$ and notice via (2.1) how this corresponds to the constant function 1.

One can view the shift S on ℓ^p as the operator

$$a(z) \mapsto za(z)$$

of multiplication by the independent variable z on the corresponding function space ℓ^p_A. From this viewpoint, note that for $a \in \ell^p_A$, $[a]$ is the ℓ^p_A-closure of the set of all pa, where p is an analytic polynomial. We will identify the shift operator on the sequence space ℓ^p with the multiplication (by z) operator on the function space ℓ^p_A, and denote both by S .

If Θ is an inner function, then ΘH^2 is a closed S-invariant subspace of H^2. Beurling showed that every S-invariant subspace of H^2 takes the form ΘH^2 for some unique (up to multiplicatitive unimodular constants) inner function Θ. Perhaps more relevant to this paper is the well-known fact that every $f \in H^2$ can be factored as

$$f = \Theta F,$$

where $[f] = [\Theta]$ and $[F] = H^2$, i.e., $F \in H^2$ is an outer function [13].

Beurling's Theorem and the inner-outer factorization carry over directly to all the Hardy spaces H^p, $1 < p < \infty$. However, for the spaces ℓ^p_A, the situation is quite unclear when $p \neq 2$. Indeed, it is known that the invariant subspaces for ℓ^p_A (especially when $p > 2$) are extremely complicated and can have many pathological properties [1].

3. Inner-Outer Factorization in ℓ^p

A key concept in our factorization and its application to ARMA processes is the notion of Birkhoff-James orthogonality. Some good sources for this material are [2, 14]. Let X and Y be vectors belonging to a normed linear space $\mathfrak{X}$. We say that X is orthogonal to Y in the Birkhoff-James sense if

$$\|X + \beta Y\|_{\mathfrak{X}} \geq \|X\|_{\mathfrak{X}} \tag{3.1}$$

for all scalars β. In this situation we write $X \perp_{\mathfrak{X}} Y$. An easy exercise shows that when $\mathfrak{X}$ is a Hilbert space with inner product $\perp$, then $X \perp Y \iff X \perp_{\mathfrak{X}} Y$. The relation $\perp_{\mathfrak{X}}$ is generally neither symmetric nor linear. In the special case $\mathfrak{X} = \ell^p$ $(1 < p < \infty)$, let us write $\perp_p$ in place of $\perp_{\ell^p}$. Of particular importance here is the following criterion for the relation $\perp_p$:

$$\mathbf{a} \perp_p \mathbf{b} \iff \sum_{k=0}^{\infty} |a_k|^{p-2} \bar{a}_k b_k = 0 \tag{3.2}$$

[14, Example 8.1], where any occurrence of "$|0|^{p-2}0$" in the above expression is interpreted as zero. Note that $\perp_p$ is therefore linear in its second argument, and it makes sense to speak of a vector being orthogonal to a subspace of ℓ^p.

Birkhoff-James orthogonality arises in a natural way in the study of p-stationary processes. These processes include α-stable processes with $1 < \alpha \leq 2$, L^p-harmonizable processes, and strictly stationary L^p

processes. The orthogonality condition is connected to the associated prediction problems, Wold-type decompositions, and moving-average representations [7, 10, 11, 16].

In light of the orthogonality criterion in (3.2), let us write

$$\mathbf{a}^{\langle p-1\rangle} = \{|a_k|^{p-2}\bar{a}_k\}_{k\geq 0}.$$

Thus, in accordance with our notational scheme in (2.1), $a^{\langle p-1\rangle}(z)$ represents the power series

$$a^{\langle p-1\rangle}(z) = \sum_{k=0}^{\infty} |a_k|^{p-2}\bar{a}_k z^k.$$

We remind the reader that for $1 < p < \infty$, p' will denote the conjugate index for p, i.e., $1/p + 1/p' = 1$.

Proposition 3.3. *For any complex number b,*

$$(b^{\langle p'-1\rangle})^{\langle p-1\rangle} = (b^{\langle p-1\rangle})^{\langle p'-1\rangle} = b. \tag{3.4}$$

If $\mathbf{b} \in \ell^p$, then $\mathbf{b}^{\langle p-1\rangle} \in \ell^{p'}$, and

$$\|\mathbf{b}\|_p^p = \|\mathbf{b}^{\langle p-1\rangle}\|_{p'}^{p'}. \tag{3.5}$$

Proof. By definition,

$$\begin{aligned}
(b^{\langle p'-1\rangle})^{\langle p-1\rangle} &= \left||b|^{p'-2}\bar{b}\right|^{p-2}|b|^{p'-2}b \\
&= |b|^{(p'-2)(p-1)-(p-2)}b \\
&= |b|^{(p'-1)(p-1)-1}b \\
&= b.
\end{aligned}$$

This verifies part of (3.4). The other part is similar. The proof of (3.5) comes from the identity

$$|b_k^{\langle p-1\rangle}|^{p'} = |b_k|^{(p-1)p'} = |b_k|^p, \quad k = 0, 1, 2, \ldots \qquad \square$$

Given $\mathbf{f} \in \ell^p$, we wish to describe the S-invariant subspace generated by $\mathbf{f}$. When $p = 2$, we set $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$, where $\widehat{\mathbf{f}}$ is the orthogonal projection of $\mathbf{f}$ onto $S[\mathbf{f}]$ (which is a closed subspace of ℓ^2 since S is an isometry). It follows that

$$\mathbf{J} \perp S^k\mathbf{J}, \quad k = 1, 2, \ldots$$

Expressing this orthogonality in terms of $L^2(\mathbb{T}, \lambda)$ functions, we get

$$\int_{\mathbb{T}} |J(e^{i\theta})|^2 e^{-ik\theta} d\lambda(e^{i\theta}) = 0, \quad k = 1, 2, \ldots.$$

Taking complex conjugates of the above expression shows that all of the Fourier coefficients of $|J|$ (except the 0-th one) are zero. This means that $|J|$ must be a constant function on $\mathbb{T}$, and hence inner, up to a constant factor. Since $\mathbf{J} \in [\mathbf{f}]$ we see that $[\mathbf{J}] \subset [\mathbf{f}]$. The reverse inclusion follows by observing that if $\mathbf{g} \in [\mathbf{f}] \ominus [\mathbf{J}]$ then $\mathbf{g} \perp [\mathbf{J}]$, which, written in integral form, says that

$$\int_{\mathbb{T}} g(e^{i\theta})\overline{J(e^{i\theta})}e^{-ik\theta}d\lambda(e^{i\theta}) = 0, \quad k = 0, 1, 2, \ldots.$$

But since $\mathbf{g} \in [\mathbf{f}]$ and $\mathbf{J} \perp S[\mathbf{f}]$ we also have, again writing this in integral form,

$$\int_{\mathbb{T}} J(e^{i\theta})e^{-ik\theta}\overline{g(e^{i\theta})}d\lambda(e^{i\theta}) = 0, \quad k = 1, 2, \ldots.$$

The above two integral identities show that the Fourier coefficients of $g\overline{J}$ vanish identically, which makes g the zero function. We may thus conclude that

$$[\mathbf{f}] = [\mathbf{J}]. \tag{3.6}$$

When $p \neq 2$, the orthogonality relationship is not given by an integral, but the infinite series expression in (3.2), and thus the powerful tools of Fourier analysis used above (and by Beurling) are not at our disposal. However, we can proceed with parts of the above proof but we will eventually have to replace the tools of Fourier analysis with something else.

Indeed if $\mathbf{f} \in \ell^p$ we can still examine the closed S-invariant subspace $S[\mathbf{f}]$ and, since ℓ^p is uniformly convex [8, Thm. 11.10], there exists a unique $\widehat{\mathbf{f}} \in \mathbf{S}[\mathbf{f}]$ which is closest to $S[\mathbf{f}]$ [8, Thm. 11.3(b)], i.e., the metric projection of $\mathbf{f}$ onto $S[\mathbf{f}]$. The definition of the Birkhoff-James orthogonality in (3.1) (and the fact that $\widehat{\mathbf{f}}$ is the vector in $S[\mathbf{f}]$ closest to $\mathbf{f}$) shows that

$$(\mathbf{f} - \widehat{\mathbf{f}}) \perp_p S[\mathbf{f}]. \tag{3.7}$$

As before, set $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$ and note that

$$\mathbf{J} \perp_p S^k\mathbf{f}, \quad k = 1, 2, \ldots \tag{3.8}$$

and, since the criterion for Birkhoff-James orthogonality $\perp_p$ from (3.2) is linear in the second argument, we also have

$$\mathbf{J} \perp_p S^k\mathbf{J}, \quad k = 1, 2, \ldots \tag{3.9}$$

These two orthogonality relations will be useful later. The following technical detail will also be useful.

Proposition 3.10. *Let $\mathbf{a} \in \ell^p$ and $\mathbf{b} \in [\mathbf{a}]$. If $w \in \mathbb{D}$ is a zero of $a(z)$ of multiplicity m, then w is a zero of $b(z)$ with multiplicity no less than m.*

Proof. Viewing everything in ℓ^p_A, we see that

$$b \in \bigvee\{z^k a : k \geq 0\}$$

and so there is a sequence of polynomials $\{p_n\}_{n\geq 1}$ so that $p_n a \to b$ in ℓ^p_A. By (2.2) we see that $p_k b \to a$ uniformly on compact subsets of $\mathbb{D}$, and the claim follows from the Cauchy integral formula. □

Next we derive some structural information about the co-projection $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$, and the related sequence $\mathbf{J}^{\langle p-1\rangle}$, in the important special case when f is a polynomial with all of its roots inside the open unit disk $\mathbb{D}$.

Theorem 3.11. *Suppose that the polynomial*

$$f(z) = \prod_{k=1}^{d} \frac{1}{r_k}(r_k - z)$$

has roots lying only in $\mathbb{D} \setminus \{0\}$. *Let* $\mathbf{f}$ *be its coefficient sequence, and let* $\widehat{\mathbf{f}}$ *be the metric projection of* $\mathbf{f}$ *onto* $S[\mathbf{f}]$ *in the norm of* ℓ^p. *Then the co-projection* $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$ *satisfies*

$$J(e^{i\theta}) = f(e^{i\theta})U(e^{i\theta}) \tag{3.12}$$

for some function $U(z)$ *with geometrically decaying Taylor coefficients.*

Furthermore,

$$J^{\langle p-1\rangle}(e^{i\theta}) = \frac{Q_0 e^{id\theta} + Q_1 e^{i(d-1)\theta} + \cdots + Q_d}{f_0 e^{id\theta} + f_1 e^{i(d-1)\theta} + \cdots + f_d} \tag{3.13}$$

for some polynomial $Q(z) = Q_0 + Q_1 z + \cdots + Q_d z^d$ *of degree exactly* d.

Proof. As observed in (3.8) we have $\mathbf{J} \perp S^k\mathbf{f}$ for all $k \geq 1$ and therefore by (3.2),

$$\sum_{j=0}^{\infty} |J_{j+k}|^{p-2} \bar{J}_{j+k} f_j = 0, \quad k = 1, 2, 3, \ldots \tag{3.14}$$

The function $J^{\langle p-1\rangle}$ is certainly analytic in $\mathbb{D}$. Condition (3.14) tells us that for all $k \geq 1$, the meromorphic function $f(1/z)J^{\langle p-1\rangle}(z)$ satisfies

$$\oint_{\mathbf{T_r}} f(1/z) J^{\langle p-1\rangle}(z) \frac{dz}{z^{k+1}} = 0$$

where T_r is any circle of radius $r \in (0, 1)$, centered at the origin. On the other hand, the function

$$z^d f(1/z) J^{\langle p-1\rangle}(z)$$

is analytic in $\mathbb{D}$. As a result, the Laurent series of $f(1/z)J^{\langle p-1\rangle}(z)$ must take the form

$$f(1/z)J^{\langle p-1\rangle}(z) = Q_0 + Q_1(1/z) + \cdots + Q_d(1/z^d)$$

for some polynomial Q. Since $Q_d = J_0^{\langle p-1\rangle} f_d \neq 0$, the degree of Q is exactly d. We may therefore take limits as $r \uparrow 1$, and conclude that

$$\begin{aligned} J^{\langle p-1\rangle}(e^{i\theta}) &= \frac{Q_0 + Q_1 e^{-i\theta} + \cdots + Q_d e^{-id\theta}}{f_0 + f_1 e^{-i\theta} + \cdots + f_d e^{-id\theta}} \\ &= \frac{Q_0 e^{id\theta} + Q_1 e^{i(d-1)\theta} + \cdots + Q_d}{f_0 e^{id\theta} + f_1 e^{i(d-1)\theta} + \cdots + f_d} \end{aligned}$$

which verifies (3.13). Note that all of the roots of f lie inside $\mathbb{D}$, so that this can be further expressed as

$$J^{\langle p-1\rangle}(z) = (Q_0 z^d + Q_1 z^{(d-1)} + \cdots + Q_d) \prod_{k=1}^{d} \Big(\sum_{j=0}^{\infty} r_k^{j-1} z^j \Big)$$

which has radius of convergence no less than

$$\rho = \min\{1/r_1, 1/r_2, 1/r_3, \ldots, 1/r_d\} > 1.$$

Thus the elements of the sequence $\mathbf{J}^{\langle p-1\rangle}$ decay geometrically at the rate $1/\rho^k$ and thus $J^{\langle p-1\rangle}(e^{i\theta})$ is a bounded function on $\mathbb{T}$. From this and Proposition 3.3, it follows that the elements of the sequence $(\mathbf{J}^{\langle p-1\rangle})^{\langle p'-1\rangle} = \mathbf{J}$ decay as $1/(\rho^{p'-1})^k$. Hence J is analytic in a neighborhood of $\overline{\mathbb{D}}$ as well.

Condition (3.9) can be written as

$$\sum_{j=0}^{\infty} |J_{j+k}|^{p-2} \bar{J}_{j+k} J_j = 0, \quad k = 1, 2, 3 \ldots,$$

which can, in turn, be expressed in integral form as

$$\int_{\mathbb{T}} J(e^{i\theta}) J^{\langle p-1\rangle}(e^{-i\theta}) e^{ik\theta} \, d\lambda(e^{i\theta}) = 0, \quad k = 1, 2, 3, \ldots$$

Note the use of the facts that J and $J^{\langle p-1\rangle}$ are analytic in the neighborhood of $\overline{\mathbb{D}}$ and so their values on $\mathbb{T}$ are well defined. By the F. and M. Riesz theorem [13, p. 41] $J(e^{i\theta})J^{\langle p-1\rangle}(e^{-i\theta})$ is some function $K(e^{i\theta})$ in H^∞ (the bounded analytic functions on $\mathbb{D}$). Using (3.13), we find that

$$J(e^{i\theta}) = \frac{f(e^{i\theta})}{Q(e^{i\theta})} K(e^{i\theta})$$

Since $\mathbf{J} \in [\mathbf{f}]$, all the zeros of f are zeros of J (by Proposition 3.10), so all of the zeros of Q lying inside $\mathbb{D}$ must also be zeros of K, counting multiplicities.

We may therefore speak of $U = K/Q$ as an analytic function on the disk of radius $\rho^{p'-1} > 1$. This proves (3.12). □

When $p = 2$, J is simply the finite Blaschke factor carrying the roots of f. From this it is easy to see that $G = f/J$ is analytic and nonvanishing in a neighborhood of $\overline{\mathbb{D}}$, and hence $\mathbf{G}$ is cyclic in ℓ^2_A. For $p \neq 2$, it is much more difficult to obtain information on the location of the zeros of J. In the following lemma, we use the continuity of zero sets to see that for p sufficiently close to 2, the resulting J has the same zeros in $\mathbb{D}$ as f. However, our methods are unable to discern whether or how this could extend to the full range of parameter values $1 < p < \infty$.

Lemma 3.15. *Suppose that the polynomial f has exactly d roots, counting multiplicity, all lying in $\mathbb{D} \setminus \{0\}$. There exists $\delta > 0$ such that $|2 - p| \leq \delta$ implies that the associated $J = f - \widehat{f}$ has exactly d zeros, counting multiplicity, in $\mathbb{D}$.*

Proof. For the present proof only, let us adopt the notation $\widehat{f}_{(p)}$, $J_{(p)}$ and $Q_{(p)}$ in place of $\widehat{f}$, J and Q, respectively, to emphasize their dependence on p. We will see that they are well behaved as p converges to 2. Fix some $p_0 > 3$. Let us also temporarily rescale f, if necessary, so that $\|\mathbf{f}\|_p < 1$ for all p, $1 < p \leq p_0$. This is harmless since $\widehat{\mathbf{f}}$ and $\mathbf{J}$ also scale linearly and their zero sets are unaffected.

From Theorem 3.11 we know there exists $\rho > 1$ such that $J_{(p)}^{\langle p-1 \rangle}$ is analytic in $\{z : |z| < \rho\}$ for all p, $1 < p < \infty$. Thus $J_{(p)}$ is analytic in $\{z : |z| < \rho^{p'-1}\}$ for all p, $1 < p < \infty$. In particular, for $1 < p \leq p_0$, $J_{(p)}$ is analytic in the fixed disk $\{z : |z| < \rho^{p'_0-1}\}$.

Choose r satisfying $1 < r < \rho^{p'_0-1}$. Let

$$\mathcal{A} = \{\mathbf{a} : \|\mathbf{a}\|_1 \leq 1\}.$$

For $1 < p \leq p_0$, we have

$$\begin{aligned} \big||a_k|^p - |a_k|^2\big| &= |a_k| \cdot \big||a_k|^{p-1} - |a_k|\big| \\ &\leq |a_k| \cdot M_p \end{aligned}$$

where $M_p = \max\{|x^{p-1} - x| : 0 \leq x \leq 1\}$. It is easy to see that M_p tends to zero as p approaches 2. Consequently,

$$\|\mathbf{a}\|_p^p - \|\mathbf{a}\|_2^2 \leq \|\mathbf{a}\|_1 \cdot M_p \leq M_p$$

uniformly for all $\mathbf{a} \in \mathcal{A}$. By the continuity of elementary power functions, it follows that

$$\|\mathbf{a}\|_p \to \|\mathbf{a}\|_2$$

uniformly for all $\mathbf{a} \in \mathcal{A}$.

Notice that if $1 < p \leq p_0$, then

$$\|J_{(2)}\|_2 = \|f - \widehat{f}_{(2)}\|_2 \leq \|f\|_2 < 1$$
$$\|J_{(p)}\|_p = \|f - \widehat{f}_{(p)}\|_p \leq \|f\|_p < 1.$$

By Hölder's inequality, each such $\mathbf{J}_{(p)}$ belongs to $\mathcal{A}$.

We have shown that for any $\epsilon > 0$, there exists a $\delta > 0$ so that

$$|2 - p| \leq \delta \implies \Big| \|\mathbf{a}\|_p - \|\mathbf{a}\|_2 \Big| < \epsilon \quad \forall \mathbf{a} \in \mathcal{A}.$$

For p satisfying $2 - \delta < p < \min\{2 + \delta, p_0\}$, we have

$$\begin{aligned} \|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_p &\leq \|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_p \\ &= \|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_2 + (\|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_p - \|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_2) \\ &\leq \|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_2 + \epsilon \end{aligned}$$

and

$$\begin{aligned} \|\mathbf{f} - \widehat{\mathbf{f}}_{(2)}\|_2 &\leq \|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_2 \\ &= \|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_p + (\|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_2 - \|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_p) \\ &\leq \|\mathbf{f} - \widehat{\mathbf{f}}_{(p)}\|_p + \epsilon. \end{aligned}$$

Due to the uniqueness of nearest points in ℓ^2, it follows that $\mathbf{J}_{(p)}$ converges to $\mathbf{J}_{(2)}$ in ℓ^2. Their respective kth coefficients must also converge.

But from Theorem 3.11 we have

$$J_{(p)}^{\langle p-1 \rangle}(z) = \frac{z^d Q_{(p)}(1/z)}{z^d f(1/z)}.$$

Each $Q_{(p)}(z)$ is a polynomial of degree d, and so the convergence of $z^d Q_{(p)}(1/z)$ to $z^d Q_{(2)}(1/z)$, as p approaches 2, is uniform on any compact set in the complex plane. Dividing by $z^d f(1/z)$ preserves uniform convergence on all compact subsets of the disk $\{z : |z| < \rho\}$. This is because division by $z^d f(1/z)$ amounts to multiplying by a factor $\sum_{k=0}^{\infty} w^k z^k$ for each root w of f. Indeed, the kth Taylor coefficients of $J_{(p)}^{\langle p-1 \rangle}$ are uniformly bounded by a constant times $1/\rho^k$. Hence $J_{(p)}^{\langle p-1 \rangle}$ converges to $J_{(2)}$ uniformly on any compact subset of $\{z : |z| < \rho\}$.

Similarly, for $1 < p < p_0$, the kth Taylor coefficients of $J_{(p)}$ are uniformly bounded by a constant times $1/\rho^{(\langle p_0' - 1 \rangle)k}$. From this we also see that $J_{(p)}$ converges uniformly to $J_{(2)}$ on some closed disk $\{z : |z| \leq$

$s\}$. The derivatives $J'_{(p)}$ converge uniformly to $J'_{(2)}$ on the $\{z : |z| \leq s\}$ as well.

Finally, consider the contour integral

$$\oint_\Gamma \frac{J'_{(p)}(z)}{J_{(p)}(z)} \frac{dz}{2\pi i}$$

where Γ is a circle of radius between 1 and s, centered at the origin, traversed once counterclockwise. This is a continuous function of p at $p = 2$. For p close to 2, $J_{(p)}(z)$ has no poles or zeros on Γ. It follows from the Argument Principle, that for such p, $J_{(p)}$ and $J_{(2)}$ have the same number of zeros, counting multiplicity, in $\mathbb{D}$. Since $J_{(2)}$ is just the finite Blaschke product carrying the roots of f, it has exactly d zeros in $\mathbb{D}$. $\square$

This next result establishes (3.6) when $p \neq 2$. The proof has the same flavor as the $p = 2$ case (it uses orthogonality), but the techniques need to avoid Fourier analysis.

Proposition 3.16. *Suppose that a polynomial f has roots lying only in $\mathbb{D} \setminus \{0\}$, and let $\widehat{\mathbf{f}}$ be the metric projection of $\mathbf{f}$ onto $S[\mathbf{f}]$ in ℓ^p. If p is sufficiently close to 2, then the co-projection $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$ has the property $[\mathbf{f}] = [\mathbf{J}]$.*

Proof. Lemma 3.15 furnishes a $\delta > 0$ such that $|2 - p| < \delta$ implies that J has exactly d zeros in $\mathbb{D}$. Assume p is in this range.

The inclusion $[\mathbf{J}] \subseteq [\mathbf{f}]$ is obviously true, and so it remains to establish the reverse inclusion. Suppose that $\mathbf{g}$ belongs to $[\mathbf{f}]$ and $\mathbf{g} \perp_p [\mathbf{J}]$. This can be expressed as

$$\sum_{k=0}^{\infty} g_{k+n}^{\langle p-1 \rangle} J_k = 0, \quad n = 0, 1, 2, \ldots \tag{3.17}$$

Let us further interpret this in terms of functions. Recall that J is analytic in an open disk with radius larger than 1. We may then speak of the function $J(1/z)$, analytic in an annulus $A = \{z : \tau < z < 1\}$ for some $\tau < 1$ (actually, from the proof of Theorem 3.11, we see that $\tau \leq 1/\rho^{p'-1} < 1$).

The condition (3.17) can now be expressed as

$$\int_{\mathbb{T}} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} g_j^{\langle p-1 \rangle} r^j e^{ij\theta} \cdot J_k r^{-k} e^{-in\theta} \cdot r^{-n} e^{-ik\theta} \, d\lambda(e^{i\theta}) = 0$$

$$\frac{1}{2\pi i} \oint_C g^{\langle p-1 \rangle}(z) J(1/z)(1/z)^{n+1} \, dz = 0$$

for all $n \geq 0$, where C is any circle centered at the origin, with radius r between τ and 1, traversed once counterclockwise. (Indeed, since these functions have geometrically converging coefficients, there is no problem passing the integral and sums through each other.) In other words, the function

$$R(z) = g^{\langle p-1\rangle}(z)J(1/z)$$

has Laurent series about the origin with only negative indices. In particular, $R_0 = 0$.

Next, we know that $J = fU$ where U is analytic in a disk with radius greater than 1. Note that because J has exactly d roots inside $\mathbb{D}$, they must coincide with the roots of f, and consequently U is nonvanishing in $\mathbb{D}$.

Suppose that

$$U(z) = V(z)\prod_{k=1}^{N}(v_k - z)$$

where $v_1,\ v_2,\ldots,v_N$ are the zeros of U of unit modulus, repeated in accordance with multiplicity. Then V is analytic and nonvanishing in some neighborhood of $\overline{\mathbb{D}}$, hence we may now write

$$g^{\langle p-1\rangle}(z)J(1/z) = R(z)$$

$$g^{\langle p-1\rangle}(z)f(1/z)U(1/z) = R(z)$$

$$g^{\langle p-1\rangle}(z)f(1/z)z^d\prod_{k=1}^{N}(v_k z - 1) = z^{d+N}R(z)/V(1/z) \tag{3.18}$$

where d is the degree of the polynomial f. The left hand side is analytic in a neighborhood of $\overline{\mathbb{D}}$, while $R(z)/V(1/z)$ is analytic in an annulus $\{z : \sigma < |z| < 1\}$, with $\tau \leq \sigma < 1$, and its Laurent series about the origin is nonvanishing only for negative indices. This can only happen if $R(z)/V(1/z)$ is a finite sum of the form

$$R(z)/V(1/z) = c_d(1/z)^{d+N} + c_{d-1}(1/z)^{d+N-1} + \cdots + c_1(1/z).$$

Let P be the polynomial

$$P(z) = z^{d+N}[c_d(1/z)^{d+N} + c_{d-1}(1/z)^{d+N-1} + \cdots + c_1(1/z)]$$

and observe that it has degree strictly less than $d + N$. Comparing both sides of (3.18), we see that $1/v_1,\ 1/v_2,\ldots,1/v_N$ must be roots of P. Therefore

$$P_0(z) = P(z)/\prod_{k=1}^{N}(v_k z - 1)$$

is a polynomial of degree less than d.

Consequently, it must be that

$$\begin{aligned}\int_{\mathbb{T}} g^{\langle p-1\rangle}(e^{i\theta}) f(e^{-i\theta}) e^{-ik\theta}\, d\lambda(e^{i\theta}) &= \int_{\mathbb{T}} \frac{P_0(e^{i\theta})}{e^{id\theta} f(e^{-i\theta})} f(e^{-i\theta}) e^{-ik\theta}\, d\lambda(e^{i\theta}) \\ &= \int_{\mathbb{T}} P_0(e^{i\theta}) e^{-i(k+d)\theta}\, d\lambda(e^{i\theta}) \\ &= 0\end{aligned}$$

for all $k \geq 0$. This is another way of saying that $\mathbf{g} \perp_p [\mathbf{f}]$. But $\mathbf{g} \in [\mathbf{f}]$ and so $\mathbf{g} \perp_p \mathbf{g}$. From the criterion for $\perp_p$ in (3.2) we have

$$\mathbf{g} \perp_p \mathbf{g} \iff \sum_{k=0}^{\infty} |g_k|^p = 0 \iff \mathbf{g} = \mathbf{0}.$$

This proves that $[\mathbf{f}] = [\mathbf{J}]$. □

The next three propositions identify classes of cyclic vectors in ℓ^p, which will be needed in the main theorem.

Proposition 3.19. *If a is analytic and nonvanishing in a neighborhood of $\overline{\mathbb{D}}$, then its coefficient sequence $\mathbf{a}$ is cyclic in ℓ^p.*

Proof. In any case, $\mathbf{a}$ belongs to ℓ^p, since the coefficients converge geometrically. The function $1/a$ is analytic in a (possibly smaller) neighborhood $\overline{\mathbb{D}}$, and so it has a geometrically convergent coefficient sequence $\mathbf{b} = (b_k)_{k=0}^{\infty}$. With $\mathbf{I} = (1, 0, 0, 0, \ldots)$, we have

$$\Big\| \mathbf{I} - \sum_{k=0}^{\infty} b_k S^k \mathbf{a} \Big\|_p = \|\mathbf{I} - \mathbf{a} * \mathbf{b}\|_p = \Big\| 1 - \frac{1}{a} \cdot a \Big\|_p = 0.$$ □

Proposition 3.20. *If $|w| = 1$, then the vector $\mathbf{a} = (w, -1, 0, 0, 0, \ldots)$ is cyclic in ℓ^p.*

Proof. It suffices to look at the case $w = 1$. Then we have

$$\begin{aligned}&\Big\| \mathbf{I} - \sum_{k=0}^{\infty} r^k S^k \mathbf{a} \Big\|_p^p \\ &= \Big\| \mathbf{I} - \Big(\mathbf{I} + \sum_{k=1}^{\infty} (1-r) r^{k-1} S^k \mathbf{I} \Big) \Big\|_p^p \\ &= (1-r)^p (1 + r^p + r^{2p} + r^{3p} + \cdots) \\ &= \frac{(1-r)^p}{1-r^p}\end{aligned}$$

for any r with $0 < r < 1$. One application of L'Hôpital's Rule confirms that this expression tends to zero as r increases to 1. □

Remark 3.21. When $p = 1$ the above fact is no longer true since $a(z) = w - z$ and the zero of a lies on $\mathbb{T}$. Since ℓ^1_A is an algebra of analytic functions which are continuous up to $\overline{\mathbb{D}}$ then a can not possibly be a cyclic vector for S.

Proposition 3.22. *If* $\mathbf{a}$ *and* $\mathbf{b}$ *are cyclic vectors in* ℓ^p *and* $\mathbf{b} \in \ell^1$*, then* $\mathbf{a} * \mathbf{b}$ *is cyclic in* ℓ^p*.*

Proof. First note that since $\mathbf{a}, \mathbf{b} \in \ell^1$ and ℓ^1 is a convolution algebra then

$$\mathbf{a} * \mathbf{b} \in \ell^1 \subset \ell^p$$

(see (2.3)). Since $\mathbf{a}$ is cyclic, there exists a sequence of polynomials $\{q_1, q_2, q_3, \ldots\}$ such that $\mathbf{a} * \mathbf{q_k}$ converges to $\mathbf{I}$ in ℓ^p. Then

$$\begin{aligned}\|\mathbf{b} - \mathbf{a} * \mathbf{b} * \mathbf{q_k}\|_{\ell^p} &= \left\| \sum_{j=0}^{\infty} b_j S^j (\mathbf{I} - \mathbf{a} * \mathbf{q_k}) \right\|_{\ell^p} \\ &\leq \left(\sum_{j=0}^{\infty} |b_j| \right) \cdot \|\mathbf{I} - \mathbf{a} * \mathbf{q_k}\|_{\ell^p}\end{aligned}$$

which tends to zero as k increases to infinity. This shows that the subspace $[\mathbf{a} * \mathbf{b}]$ contains the vector $\mathbf{b}$. By assumption, this vector is cyclic, and therefore $[\mathbf{a} * \mathbf{b}]$ must be all of ℓ^p. ☐

We are now able to show that the coefficient vector for J/f is cyclic in ℓ^p.

Lemma 3.23. *Suppose that the polynomial* f *has roots lying only in* $\mathbb{D} \setminus \{0\}$*, and let* $\mathbf{J} = \mathbf{f} - \widehat{\mathbf{f}}$ *be the co-projection of* $\mathbf{f}$ *onto* $S[\mathbf{f}]$ *in* ℓ^p*. If* p *is sufficiently close to 2, then* $U = J/f$ *has the property that* $\mathbf{U}$ *is cyclic in* ℓ^p*.*

Proof. We already know from Theorem 3.11 that U is analytic in a neighborhood of $\overline{\mathbb{D}}$. Proposition 3.16 assures that $[\mathbf{f}] = [\mathbf{J}]$ when p is sufficiently close to 2. There exist polynomials $q_1, q_2, q_3, \ldots$ such that

$$\|\mathbf{f} + \mathbf{q_k} * \mathbf{J}\|_{\ell^p} \longrightarrow 0.$$

Since U is analytic in a neighborhood of $\overline{\mathbb{D}}$, it has at most a finite number of zeros $v_1, v_2, \ldots, v_N$, repeated according to multiplicity, on the circle $\mathbb{T}$. Once again let

$$V(z) = \frac{U(z)}{\prod_{k=1}^{N} (v_k - z)}$$

Then V is analytic and nonvanishing in some neighborhood of $\overline{\mathbb{D}}$, and its reciprocal is analytic is a (possibly smaller) neighborhood of $\overline{\mathbb{D}}$. Accordingly, the Taylor coefficients of V are absolutely summable, i.e., $V \in \ell^1_A$. By the Wiener-Lévy Theorem [22, p. 245], the Taylor coefficients of $1/V$ are also absolutely summable. Thus, writing

$$\frac{1}{V(e^{i\theta})} = \sum_{k=0}^{\infty} b_k e^{ik\theta},$$

we have

$$\mathbf{I} = \sum_{k=0}^{\infty} b_k S^k \mathbf{V}$$

The series converges in the norm of ℓ^p, since

$$\Big\| \sum_{k=n_1}^{n_2} b_k S^k \mathbf{V} \Big\|_p \leq \|V\|_p \sum_{k=n_1}^{n_2} |b_k|$$

for all indices $n_1 \leq n_2$, and the coefficients $(b_k)_{k=0}^{\infty}$ are absolutely summable. Thus $\mathbf{V}$ is cyclic. The cyclicity of $\mathbf{U}$ now follows from Propositions 3.20 and 3.22. ☐

Finally, here is the our main theorem with its proof.

Theorem 3.24. *If F is analytic in a neighborhood of $\overline{\mathbb{D}}$, then for any p sufficiently close to 2, there exist functions G and J, analytic in a neighborhood of $\overline{\mathbb{D}}$, such that*

$$F = JG$$

where $\mathbf{G}$ is cyclic vector in ℓ^p and $\mathbf{J}$ satisfies

$$\mathbf{J} \perp_p S^k \mathbf{J}, \quad k = 1, 2, 3, \ldots.$$

Furthermore, we have $[\mathbf{F}] = [\mathbf{J}]$ in ℓ^p. The functions J and G are unique up to multiplicative constants.

Proof. There is no harm in supposing that $F(0) = 1$, as any factor az^k of $F(z)$ can be absorbed into J. Let all of the roots of F inside $\mathbb{D}$ be removed by the polynomial f. Then the coefficient vector $\mathbf{a}$ of the function $a = F/f$ is cyclic. To see this, write

$$a(z) = H(z) \prod_{j=1}^{q} (z - \zeta_j),$$

where $\zeta_1, \ldots, \zeta_q$ are the (possible) zeros of a on $\mathbb{T}$ and H is analytic and zero free in a neighborhood of $\overline{\mathbb{D}}$. Now apply Propositions 3.19 and 3.20.

Next apply Theorem 3.11 to obtain $J = fU$, where $\mathbf{J}$ has the claimed orthogonality property, and the terms of $\mathbf{U}$ decay geometrically. Lemma 3.15 assures that U is zero free in an open neighborhood of $\overline{\mathbb{D}}$ for p in some interval $[2-\delta, 2+\delta]$. Consequently, $1/U$ is is analytic in a neighborhood of $\overline{\mathbb{D}}$ (and, of course, zero free). Thus by Proposition 3.19, the coefficient vector for $1/U$ is cyclic in ℓ^p for $p \in [2-\delta, 2+\delta]$.

Define $G = a/U$ and apply Proposition 3.22 to see that $\mathbf{G}$ is cyclic in ℓ^p. Now observe that

$$F = af, \quad F = \frac{a}{U}Uf, \quad F = GJ.$$

The equality $[\mathbf{f}] = [\mathbf{J}]$ was established in Proposition 3.16. The coefficient vector for $a = F/f$ cyclic in ℓ^p. Consequently, Young's Inequality (see (2.3)) assures that the expressions

$$\begin{aligned} \|\mathbf{f} - \mathbf{F} * \mathbf{h}\|_p &= \|\mathbf{f} * (\mathbf{I} - \mathbf{a} * \mathbf{h})\|_p \\ &\leq \|\mathbf{f}\|_1 \cdot \|\mathbf{I} - \mathbf{a} * \mathbf{h})\|_p \end{aligned}$$

can be made arbitrarily small by judicious selection of the polynomial h. This gives $[\mathbf{f}] \subseteq [\mathbf{F}]$ and hence $[\mathbf{J}] \subseteq [\mathbf{F}]$. To see that $[\mathbf{F}] \subseteq [\mathbf{f}]$ (and hence $[\mathbf{J}] = [\mathbf{F}]$) observe that $F = af$ and so

$$\begin{aligned} \|\mathbf{F} - \mathbf{h} * \mathbf{f}\|_p &= \|\mathbf{f} * (\mathbf{a} - \mathbf{h})\|_p \\ &\leq \|\mathbf{f}\|_1 \|\mathbf{a} - \mathbf{h}\|_p \end{aligned}$$

which, as before, can be made arbitrarily small by a judicious selection of the polynomial h (indeed, the N-th Taylor polynomial of a).

The uniqueness of the factorization, up to multiplicative constants, follows from the uniqueness of J as the co-projection of f onto $S[f]$. $\square$

Note that Theorem 3.24 could equivalently be expressed in terms of sequences. That is, if $\mathbf{F}$ is a geometrically decaying sequence, then for p sufficiently close to 2 there exist $\mathbf{G}$ and $\mathbf{J}$ in ℓ^p such that $\mathbf{F} = \mathbf{J} * \mathbf{G}$; $\mathbf{G}$ is cyclic vector in ℓ^p; and $\mathbf{J}$ satisfies $\mathbf{J} \perp_p S^k\mathbf{J}$ for all $k \geq 1$.

Example 3.25. For a simple example, consider the polynomial $f(z) = 1 - z/w$, where $w \in \mathbb{D} \setminus \{0\}$. Then

$$J(z) = \frac{1 - z/w}{1 - w^{\langle p'-1 \rangle} z}$$

and

$$G(z) = 1 - w^{\langle p'-1 \rangle} z.$$

To see this, check that $J \in [f]$, and then straightforward calculation shows that $\mathbf{J} \perp_p S^k\mathbf{f}$ for all $k \geq 1$. Note that J really is just the Blaschke factor for f when $p = 2$.

The main theorem presents a sort of inner-outer factorization related to ℓ^p. We would like to be able to extend its validity to the full parameter range $1 < p < \infty$, or show that this is impossible. This effort would entail investigating more thoroughly the zero set behavior of J, and understanding the analytical behavior of the mapping $\mathbf{a} \mapsto \mathbf{a}^{\langle p-1\rangle}$. These matters are the subject of future projects.

4. A comment about the Hardy space

As mentioned in the introduction, the classical Hardy space $H^p, p \in (1, \infty)$, is the Banach space of analytic functions f on $\mathbb{D}$ for which the norm

$$\|f\|_{H^p} = \sup_{0<r<1} \left(\int_{\mathbb{T}} |f(re^{i\theta})|^p d\lambda(e^{i\theta}) \right)^{\frac{1}{p}}$$

is finite. Standard theory [13] says that for every $f \in H^p$

$$f(e^{i\theta}) := \lim_{r \to 1^-} f(re^{i\theta})$$

exists for λ-almost every $e^{i\theta}$ and moreover,

$$\|f\|_{H^p} = \left(\int_{\mathbb{T}} |f(e^{i\theta})|^p d\lambda(e^{i\theta}) \right)^{\frac{1}{p}}.$$

In other words, via radial limits, H^p is a closed subspace of L^p. In fact, H^p can be characterized by the "vanishing negative Fourier coefficients" criterion:

$$H^p = \left\{ f \in L^p : \int_{\mathbb{T}} f(e^{i\theta}) e^{ik\theta} d\lambda(e^{i\theta}) = 0 \; \forall k > 1 \right\}.$$

Similar to the ℓ^p case, the Birkhoff-James orthogonality in L^p is

$$f \perp_{L^p} g \iff \int_{\mathbb{T}} |f(e^{i\theta})|^{p-2} \overline{f(e^{i\theta})} g(e^{i\theta}) d\lambda(e^{i\theta}) = 0.$$

It is well known that every $f \in H^p$ can be factored (uniquely up to multiplicative unimodular constants) as

$$f = jg,$$

where j is an inner function, i.e., $|j(e^{i\theta})| = 1$ for almost every $e^{i\theta}$, and g is outer, i.e.,

$$[g] = \bigvee \{S^k g : k = 0, 1, \ldots\} = H^p,$$

where $(Sg)(z) = zg(z)$ is the unilateral shift on H^p.

If $\mathcal{P}_0$ denotes the analytic polynomials which vanish at $z = 0$, then for $f = jg$ as above,

$$\begin{aligned}\inf\{\|f - p(S)f\|_{H^p}^p : p \in \mathcal{P}_0\} &= \inf\{\|j(g - p(S)g)\|_{H^p}^p : p \in \mathcal{P}_0\} \\ &= \inf\{\|g - p(S)g\|_{H^p}^p : p \in \mathcal{P}_0\} \\ &= |g(0)|^p \\ &= \|g(0)j\|_{H^p}^p.\end{aligned}$$

The above says that the co-projection of f onto $S[f]$ is a constant multiple of j, which is consistent with what we obtained for ℓ^p_A.

Furthermore, that j is orthogonal in the Birkhoff-James sense to all of its forward shifts, i.e.,

$$j \perp_{H^p} S^k j, \quad k = 1, 2, \ldots,$$

follows from

$$\begin{aligned}\int_{\mathbb{T}} j(e^{i\theta})^{\langle p-1\rangle} e^{ik\theta} j(e^{i\theta})\, d\lambda(e^{i\theta}) &= \int_{\mathbb{T}} |j(e^{i\theta})|^{p-2}\overline{j(e^{i\theta})} e^{ik\theta} j(e^{i\theta})\, d\lambda(e^{i\theta}) \\ &= \int_{\mathbb{T}} |j(e^{i\theta})|^p e^{ik\theta}\, d\lambda(e^{i\theta}) \\ &= \int_{\mathbb{T}} e^{ik\theta}\, d\lambda(e^{i\theta}) \\ &= 0.\end{aligned}$$

In summary, applying the Birkhoff-James orthogonality to obtain a factorization in H^p, as was done for ℓ^p_A, yields the classical inner-outer factorization in H^p.

5. ARMA Processes

In this section we apply the analytical methods of the previous sections to solve a problem associated with certain stochastic processes. We say that the real valued random process $\{X_k\}_{k=-\infty}^{\infty}$ is an Autoregressive Moving Average (ARMA) process if it is (weakly or strictly) stationary and satisfies

$$X_k - \phi_1 X_{k-1} - \cdots - \phi_n X_{k-n} = Z_k + \theta_1 Z_{k-1} + \cdots + \theta_q Z_{k-q} \tag{5.1}$$

for all k, where $\{Z_k\}_{k=-\infty}^{\infty}$ is a white noise. We speak of

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_n z^n \tag{5.2}$$

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{5.3}$$

as the *characteristic polynomials*. ARMA models are important and widely used in statistical time series analysis. Throughout this section let us assume that ϕ and θ have real coefficients, they have no common roots, and they have no roots on the unit circle $\mathbb{T}$.

It is well known that the ARMA model (5.1) has a weakly stationary solution of the form

$$X_k = \sum_{j=0}^{\infty} a_j Z_{k-j} \tag{5.4}$$

if and only if ϕ has no roots in the closed unit disk $\overline{\mathbb{D}}$ (see, for example, Brockwell and Davis [5, Theorem 3.1.1]). In this case, the coefficients $\{a_k\}_{k=0}^{\infty}$ are determined by the condition

$$\sum_{k=0}^{\infty} a_k z^k = \theta(z)/\phi(z)$$

for all $z \in \mathbb{D}$. The series in (5.4) converges absolutely with probability one, since the coefficients $\{a_k\}_{k=0}^{\infty}$ are absolutely summable.

An important feature of the solution (5.4) is that it is *causal*; that is, the value of X_k depends only on the present and past values of the noise, $\{\ldots, Z_{k-2}, Z_{k-1}, Z_k\}$. If ϕ does have some roots inside $\mathbb{D}$, then a solution of the form

$$X_k = \sum_{j=-\infty}^{\infty} a_j Z_{k-j} \tag{5.5}$$

exists, in which values of X_k may also depend on future values of the noise. For that reason, however, this solution is unsuitable for many practical uses.

It is also well known that there is an inverse or autoregressive representation

$$X_k = Z_k - \sum_{j=1}^{\infty} \pi_j X_{k-j}$$

with absolutely summable coefficients $\{\pi_k\}_{k=0}^{\infty}$ if and only if θ has no roots in the closed disk $\overline{\mathbb{D}}$ [5, Theorem 3.1.2]. In this case, the coefficients can be determined from

$$\sum_{k=0}^{\infty} \pi_k z^k = \phi(z)/\theta(z).$$

Both the solution and inversion carry over to infinite variance stable processes with essentially the same proofs; see Samorodnitsky and Taqqu [20], p. 376ff.

It could happen in practice, however, that the polynomials ϕ and θ are determined by fitting to a data set, and there is at least one resulting characteristic root inside $\mathbb{D}$. In this situation, the extant methods described above do not apply. Instead, using analytical tools from the previous section, we shall derive a Symmetric-α-Stable (SαS) solution to (5.1), involving a second white noise process $\{\epsilon_k\}$ which is an i.i.d. SαS process. In this solution, both the process $\{X_k\}$ and the given noise $\{Z_k\}$ are expressed as causal moving averages of $\{\epsilon_k\}$. A similar approach is taken to obtain an autoregressive representation for $\{X_k\}$. Simple examples then conclude this paper.

The books [5, 3, 19] cover weakly stationary ARMA models and their extensions. The analogous infinite variance case is handled in [20], and their asymptotic dependence structure is studied in [15]. For more general prediction problems associated with infinite variance processes, see [7, 12, 17]. The rates of convergence of SαS moving averages and autoregressive series are explored in [9]. In [6], strictly stationary, possibly non-causal solutions are obtained for ARMA equations with characteristic roots inside and outside of $\mathbf{T}$. This was extended to the multivariate case in [4], and the infinite-dimensional case in [21]. All of the solutions presented here satisfy the causality requirement.

Let us present some background on SαS processes. We defer the elaborate definition to [20], since ultimately all we need is a Banach space isomorphism between a certain linear space of random variables onto ℓ^α. Suffice to say that SαS distributions are a family of distributions that are stable under independent sums and scalar multiplication; when $\alpha = 2$ we have the Gaussian distribution; when $0 < \alpha < 2$ the density function is bell-shaped but has heavy tails.

Fix $1 < \alpha \le 2$, and let $\mathcal{X}$ be a linear space of (real) SαS random variables. A norm is defined on $\mathcal{X}$ in the following manner. If $Y_1, Y_2, \ldots, Y_d$ belong to $\mathcal{X}$, then there exists a symmetric finite measure $\Gamma_{Y_1,Y_2,\ldots,Y_d}$ on the unit sphere S_d in $\mathbf{R}^d$ such that

$$E\exp\{i(\theta_1 Y_1 + \cdots \theta_d Y_d)\}$$

is equal to

$$\exp\Big\{-\int_{S_d} |\theta_1 t_1 + \cdots + \theta_d t_d|^\alpha \, d\Gamma_{Y_1,Y_2,\ldots,Y_d}(t_1,\ldots,t_d)\Big\}$$

The measure $\Gamma_{Y_1,Y_2,\ldots,Y_d}$ is the *spectral measure* of the random vector $(Y_1, Y_2, \ldots, Y_d)$. Given any pair of random variables X and Y in $\mathcal{X}$, the *covariation* of X and Y is then defined to be

$$[X,Y]_\alpha = \int_{S_2} t_1 |t_2|^{\alpha-1} \operatorname{sign}(t_2)\, d\Gamma_{X,Y}(t_1,t_2)$$

This generalizes the notion of covariance to infinite variance stable variables. The *covariation norm* of X in $\mathcal{X}$ is now given by

$$\|X\|_\alpha = [X, X]_\alpha$$

This is indeed a norm on $\mathcal{X}$, and it turns out that for every X in $\mathcal{X}$ we have $\|X\|_\alpha = \sigma_X$, the scale parameter of X [20, Sections 2.3, 2.4, 2.7 and 2.8]. (We exclude the parameter range $0 < \alpha \leq 1$, for then, the covariation is not defined). The closure of $\mathcal{X}$ under the covariation norm is a Banach space.

If $\mathcal{X}$ is spanned by an i.i.d. sequence $\{\epsilon_k\}_{k=-\infty}^{\infty}$ of SαS random variables with unit scale parameter, then this space is isometrically isomorphic to the space ℓ^α of real sequences over the index set:

$$\Big\| \sum_{k=-\infty}^{\infty} a_k \epsilon_k \Big\|_\alpha = \Big(\sum_{k=-\infty}^{\infty} |a_k|^\alpha \Big)^{1/\alpha} \tag{5.6}$$

for all sequences of real constants $\{a_k\}_{k=-\infty}^{\infty}$. This isomorphism enables us to bring in the results for ℓ^p spaces established above. Indeed, the notions of metric projection and Birkhoff-James orthogonality carry over in a straightforward way from ℓ^α to $\mathcal{X}$. We are primarily interested in real one-sided sequences. Given $\mathbf{a} = (a_0, a_1, a_2, \ldots)$ and $\mathbf{b} = (b_0, b_1, b_2, \ldots)$ in ℓ^α, we have

$$\sum_{k=0}^{\infty} a_k \epsilon_{-k} \quad \perp_\alpha \quad \sum_{k=0}^{\infty} b_k \epsilon_{-k}$$

in $\mathcal{X}$ if and only if $\mathbf{a} \perp_\alpha \mathbf{b}$ in ℓ^α. This, in turn, occurs precisely when

$$\sum_{k=0}^{\infty} a_k^{\langle \alpha - 1 \rangle} b_k = 0$$

It should be noted that restricting the scalars from the complex field to the real field presents no difficulties when utilizing the ℓ^p results of previous sections. This is because if f has real coefficients in Theorem 3.11, then so do the associated functions J, $J^{\langle p-1 \rangle}$, K and Q.

Let us say that a sequence $\{Y_k\}$ in $\mathcal{X}$ is an *orthogonal sequence* if $Y_j \perp Y_k$ whenever $j > k$; that is, each Y_j is orthogonal to its past. We emphasize that orthogonality here is very much a one-sided affair: such Y_j need not be orthogonal to its future. In this section the term *white noise* refers to a SαS orthogonal sequence with a common nonzero scale parameter σ. It may or may not be i.i.d. This is not standard terminology; however, when $\alpha = 2$ a white noise is orthogonal in the usual (Hilbert space) sense.

Now with the ARMA equation (5.1) given, we will construct a causal solution based on an i.i.d. SαS white noise $\{\epsilon_k\}$. We write B for the backward shift operator

$$B\epsilon_k = \epsilon_{k-1}$$

defined on the closed linear span of $\{\epsilon_k\}$.

Let $\mathbf{\Phi}$ be the infinite sequence

$$(1, -\phi_1, -\phi_2, \ldots, -\phi_n, 0, 0, \ldots)$$

in ℓ^α (the index set is $\{0, 1, 2, \ldots\}$). Let the roots of ϕ in $\mathbb{D}$ be removed by the polynomial f_1, with $f_1(0) = 1$. Apply Theorem 3.11 to obtain functions J and U, analytic in a neighborhood of $\overline{\mathbb{D}}$, such that $J = fU$ and $\mathbf{J}$ is orthogonal in ℓ^α to all of its forward shifts.

With these ingredients, here is our generalized solution to (5.1) for infinite variance processes.

Theorem 5.7. *Let $1 < \alpha \leq 2$, and let $\{\epsilon_k\}$ be an i.i.d. SαS white noise. The ARMA equation (5.1) has a stationary SαS solution of the form*

$$X_k = \sum_{j=0}^{\infty} a_j \epsilon_{k-j} \tag{5.8}$$

$$Z_k = J(B)\epsilon_k \tag{5.9}$$

where the coefficients $\{a_k\}_{k=0}^\infty$ are determined by

$$\sum_{k=0}^{\infty} a_k z^k = \theta(z)U(z)\frac{f(z)}{\phi(z)}. \tag{5.10}$$

The representations (5.8) and (5.9) converge in covariation norm, and the process $\{Z_k\}$ is an SαS white noise.

Proof. The function ϕ/f is analytic and nonvanishing in a neighborhood of $\overline{\mathbb{D}}$; therefore its reciprocal is an analytic function in the same domain and has geometrically decaying Taylor coefficients. It follows that the series in (5.10) is absolutely summable. Hence the moving average (5.8) converges in covariation norm.

The coefficients of J are absolutely summable, and hence $J(B)\epsilon_k$ converges in covariation norm. The fact that $\mathbf{J} \perp S[\mathbf{J}]$ in ℓ^α implies that $\{Z_k\}$ is an SαS white noise.

Finally, define

$$X_k = \left(\frac{f}{\phi}\right)(B)\theta(B)U(B)\epsilon_k.$$

These functions of B are summable series, and hence they converge in operator norm. Now observe that

$$\begin{aligned} X_k &= \left(\frac{f}{\phi}\right)(B)\theta(B)U(B)\epsilon_k \\ \left(\frac{\phi}{f}\right)(B)X_k &= \theta(B)U(B)\epsilon_k \\ \left(\frac{\phi}{f}\right)(B)f(B)X_k &= \theta(B)f(B)U(B)\epsilon_k \\ \phi(B)X_k &= \theta(B)J(B)\epsilon_k \\ \phi(B)X_k &= \theta(B)Z_k. \end{aligned}$$ □

An autoregressive representation of X_k can be achieved in a similar way. Let the zeros of θ in $\mathbb{D}$ be removed by the polynomial $F(z)$, with $F(0) = 1$. Invoke Theorem 3.11 once again to obtain functions $\Psi(z)$ and $W(z)$, analytic in a neighborhood of $\overline{\mathbb{D}}$, such that

$$\Psi(z) = F(z)W(z)$$

and $\mathbf{\Psi} \perp S[\mathbf{\Psi}]$ in ℓ^α.

Theorem 5.11. *Let $1 < \alpha \leq 2$ be fixed, and let $\{\epsilon_k\}$ be an i.i.d. $S\alpha S$ white noise with unit scale parameter. The solution from Theorem 5.7 of (5.1) admits the autoregressive representation*

$$X_k = \eta_k - \sum_{j=1}^{\infty} \pi_j X_{k-j} \tag{5.12}$$

where

$$\eta_k = \Psi(B)Z_k = \Psi(B)J(B)\epsilon_k \tag{5.13}$$

the coefficients $\{\pi_j\}$ derive from

$$1 + \sum_{j=1}^{\infty} \pi_j z^j = W(z)\phi(z)\frac{F(z)}{\theta(z)}. \tag{5.14}$$

The series in (5.12) converges absolutely with probability one, and the representation (5.13) converges in covariation norm.

Proof. Since θ/F is analytic and nonvanishing in a neighborhood of $\overline{\mathbb{D}}$, its reciprocal is also analytic on that domain and has geometrically decaying Taylor coefficients. It follows that the sequence $\{\pi_k\}$ in equation (5.14) is absolutely summable, and thus the representation (5.12) converges absolutely with probability one. We now have

$$\begin{aligned}
\theta(B)Z_k &= \phi(B)X_k \\
\Big(\frac{\theta}{F}\Big)(B)F(B)\theta(B)Z_k &= \phi(B)X_k \\
F(B)\theta(B)Z_k &= \Big(\frac{F}{\theta}\Big)(B)\phi(B)X_k \\
W(B)F(B)\theta(B)Z_k &= W(B)\Big(\frac{F}{\theta}\Big)(B)\phi(B)X_k \\
\theta(B)\Psi(B)Z_k &= W(B)\Big(\frac{F}{\theta}\Big)(B)\phi(B)X_k \\
\theta(B)\eta_k &= W(B)\Big(\frac{F}{\theta}\Big)(B)\phi(B)X_k
\end{aligned}$$

which confirms (5.12). Both representations in (5.13) converge in covariation norm since the coefficients of $\mathbf{\Psi}$ and $\mathbf{J}$ converge geometrically. □

We have obtained in this section causal moving average and autoregressive representations for ARMA processes, in terms of related white noise processes, when an infinite variance phenomenon is proposed or in evidence. Because the characteristic roots may lie in $\mathbb{D}$, this extends existing methods. All of the results of this section hold true when $\alpha = 2$, in which case the processes are Gaussian; the conclusions are then consistent with the results of Theorem [5, Theorem 3.5.2].

One limitation of Theorem 5.11 is that in contrast to the finite variance case, $\{\Psi(B)J(B)\epsilon_k\}$ is generally not an orthogonal process under the covariation norm with index α. However, both $\{\Psi(B)\epsilon_k\}$ and $\{J(B)\epsilon_k\}$ are orthogonal white noise processes. Thus, if either θ or ϕ has no roots inside (D), then the left side of the autoregressive representation (5.12) is an SαS white noise. Another limitation is that the above methods depend heavily on the isometry (5.6) between a linear space of SαS random variables and the sequence space ℓ^α. Other classes of infinite variance processes would require a different approach.

We conclude with two simple examples. Let r be a real number with $0 < |r| < 1$. First, the complete solution is provided for $\phi(z) = 1 - z/r$ and $\theta(z) = 1$. Then the autoregressive representation is obtained when $\phi(z) = 1$ and $\theta(z) = 1 - z/r$.

Proposition 5.15. *Fix a real number r, with $0 < |r| < 1$. Suppose that $\{\epsilon_k\}$ is an i.i.d. SαS white noise. Then*

$$Z_k = X_k - \frac{1}{r}X_{k-1}$$

has a solution

$$\begin{aligned} X_k &= \sum_{j=0}^{\infty} r^{\langle\alpha'-1\rangle j}\epsilon_{k-j} \\ Z_k &= \epsilon_0 - \Big(\frac{1}{r}-1\Big)\sum_{j=1}^{\infty} r^{\langle\alpha'-1\rangle j}\epsilon_{k-j} \end{aligned}$$

where $\{Z_k\}$ *is an* $S\alpha S$ *white noise. The finite variance case is obtained by taking* $\alpha = 2$.

Proof. Following the definitions and calculations leading to Theorem 5.7, we find that

$$\begin{aligned} \phi(z) &= f(z) = 1 - \frac{z}{r} \\ J(z) &= \frac{1}{r}\frac{r-z}{1-r^{\langle\alpha'-1\rangle}z} \\ J^{\langle\alpha-1\rangle}(z) &= 1 - \Big(\frac{1-|r|^{\alpha'}}{|r|^{\alpha'}}\Big)^{\alpha/\alpha'}\frac{rz}{1-rz} \\ U(z) &= \frac{1-rz}{1-r^{\langle\alpha'-1\rangle}z} \end{aligned}$$

Indeed, it is clear that the proposed $\mathbf{J}$ lies in the span of ϕ. Note that $J(0) = 1$. To see that $\mathbf{J} \perp S^k\mathbf{f}$ for all $k = 1, 2, 3, \ldots$, we need only check

$$\begin{aligned} &J_k^{\langle\alpha-1\rangle} f_0 - J_{k+1}^{\langle\alpha-1\rangle} f_1 \\ &= -\Big(\frac{1-|r|^{\alpha'}}{|r|^{\alpha'}}\Big)^{\alpha/\alpha'} r\cdot r^k \cdot 1 + \Big(\frac{1-|r|^{\alpha'}}{|r|^{\alpha'}}\Big)^{\alpha/\alpha'} r\cdot r^{k+1}\cdot\frac{1}{r} \\ &= 0 \end{aligned}$$

for all $k = 1, 2, 3, \ldots$. This affirms that $\mathbf{J} = \mathbf{f} - \hat{\mathbf{f}}$. The rest is routine algebra. In the finite variance ($\alpha = 2$) case, this simplifies to

$$\begin{aligned} &J(z) = J^{\langle\alpha-1\rangle}(z) = \frac{1}{r}\frac{r-z}{1-rz} \\ &U(z) = 1 \end{aligned}$$

□

Proposition 5.16. *Let* $\{Z_k\}$ *be an i.i.d.* $S\alpha S$ *white noise. Fix a real number* r*, with* $0 < |r| < 1$*. Then the process* $\{X_k\}$ *satisfying*

$$Z_k - \frac{1}{r}Z_{k-1} = X_k$$

has an autoregressive representation

$$X_k = \eta_k - \sum_{j=1}^{\infty} r^{\langle \alpha'-1\rangle j} X_{k-j}$$

where

$$\eta_k = \Big(\frac{1}{r} - 1\Big) \sum_{j=1}^{\infty} r^{\langle \alpha'-1\rangle j} Z_{k-j}$$

is an $S\alpha S$ white noise. The finite variance case is obtained by taking $\alpha = 2$.

Proof. Straightforward calculation yields the identifications

$$\begin{aligned}
\phi(z) &= 1\\
\theta(z) &= 1 - z/r\\
F(z) &= 1 - \frac{1}{r}z\\
\Psi(z) &= \frac{1}{r}\frac{r-z}{1 - r^{\langle \alpha'-1\rangle} z}\\
\Psi^{\langle \alpha-1\rangle}(z) &= 1 - \Big(\frac{1-|r|^{\alpha'}}{|r|^{\alpha'}}\Big)^{\alpha/\alpha'} \frac{rz}{1-rz}\\
Q(z) &= (-1/r)z + [1 + (1-|r|^{\alpha'})^{\alpha/\alpha'}|r|^{-\alpha}]\\
W(z) &= \frac{1}{1 - r^{\langle \alpha'-1\rangle} z}.
\end{aligned}$$

Indeed, we have already established that $\boldsymbol{\Psi} = \mathbf{f} - \hat{\mathbf{f}}$. In the finite variance case this simplifies to

$$\begin{aligned}
\Psi(z) = \Psi^{\langle \alpha-1\rangle}(z) &= \frac{1}{r}\frac{r-z}{1-rz}\\
Q(z) &= \frac{1}{r^2}(1-rz)\\
W(z) &= \frac{1}{1-rz}.
\end{aligned}$$ □

The ARMA(1,1) case can be handled using the above results. However, if the characteristic polynomials have two or more roots in $\mathbb{D}$, then the calculation of $J(z)$ can be very challenging.

References

[1] Evgeny Abakumov and Alexander Borichev. Shift invariant subspaces with arbitrary indices in l^p spaces. *J. Funct. Anal.*, 188(1):1–26, 2002.

[2] Javier Alonso, Horst Martini, and Senlin Wu. On Birkhoff orthogonality and isosceles orthogonality in normed linear spaces. *Aequationes Math.*, 83(1-2):153–189, 2012.

[3] George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2008. Forecasting and control.

[4] Peter Brockwell, Alexander Lindner, and Bernd Vollenbröker. Strictly stationary solutions of multivariate ARMA equations with i.i.d. noise. *Ann. Inst. Statist. Math.*, 64(6):1089–1119, 2012.

[5] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer, New York, 2006. Reprint of the second (1991) edition.

[6] Peter J. Brockwell and Alexander Lindner. Strictly stationary solutions of autoregressive moving average equations. *Biometrika*, 97(3):765–772, 2010.

[7] Stamatis Cambanis, Clyde D. Hardin, Jr., and Aleksander Weron. Innovations and Wold decompositions of stable sequences. *Probab. Theory Related Fields*, 79(1):1–27, 1988.

[8] N. L. Carothers. *A Short Course on Banach Space Theory*, volume 64 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2005.

[9] R. Cheng and C. Harris. Mixed norm spaces and prediction of SαS moving averages. *J. Time Ser. Anal.*, 36:853–875, 2015.

[10] R. Cheng, A. G. Miamee, and M. Pourahmadi. Regularity and minimality of infinite variance processes. *J. Theoret. Probab.*, 13(4):1115–1122, 2000.

[11] R. Cheng and W. T. Ross. Weak parallelogram laws on Banach spaces and applications to prediction. *Period. Math. Hungar.*, 71(1):45–58, 2015.

[12] Daren B. H. Cline and Peter J. Brockwell. Linear prediction of ARMA processes with infinite variance. *Stochastic Process. Appl.*, 19(2):281–296, 1985.

[13] Peter L. Duren. *Theory of H^p Spaces*. Pure and Applied Mathematics, Vol. 38. Academic Press, New York-London, 1970.

[14] Robert C. James. Orthogonality and linear functionals in normed linear spaces. *Trans. Amer. Math. Soc.*, 61:265–292, 1947.

[15] Piotr S. Kokoszka and Murad S. Taqqu. Infinite variance stable ARMA processes. *J. Time Ser. Anal.*, 15(2):203–220, 1994.

[16] A. G. Miamee and M. Pourahmadi. Wold decomposition, prediction and parameterization of stationary processes with infinite variance. *Probab. Theory Related Fields*, 79(1):145–164, 1988.

[17] Mohammad Mohammadi and Adel Mohammadpour. Best linear prediction for α-stable random processes. *Statist. Probab. Lett.*, 79(21):2266–2272, 2009.

[18] Raymond Mortini and Michael von Renteln. Ideals in the Wiener algebra W^+. *J. Austral. Math. Soc. Ser. A*, 46(2):220–228, 1989.

[19] Mohsen Pourahmadi. *Foundations of Time Series Analysis and Prediction Theory*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York, 2001.

[20] Gennady Samorodnitsky and Murad S. Taqqu. *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. Stochastic models with infinite variance.

[21] Felix Spangenberg. Strictly stationary solutions of ARMA equations in Banach spaces. *J. Multivariate Anal.*, 121:127–138, 2013.

[22] A. Zygmund. *Trigonometric Series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. With a foreword by Robert A. Fefferman.

Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA 23529

E-mail address: rcheng@odu.edu

Department of Mathematics and Computer Science, University of Richmond, Richmond, VA 23173, USA

E-mail address: wross@richmond.edu