FREE SEMIGROUP ALGEBRAS A SURVEY

KENNETH R. DAVIDSON

Free semigroup algebras are WOT-closed algebras generated by n isometries with pairwise orthogonal ranges. They were introduced in [27] as an interesting class of operator algebras in their own right. The prototype algebra, obtained from the left regular representation of the free semigroup on n letters, was introduced by Popescu [45] in connection with multi-variable non-commutative dilation theory. This algebra has a great deal of analytic structure associated to the unit ball in \mathbb{C}^n which justifies its name as the noncommutative analytic Toeplitz algebra. The general free semigroup algebras contain interesting computable information about the unitary invariants for the *n*-tuple of generators. This has allowed the classification of large classes of representations of the Cuntz algebra. Such classifications are important in various applications of C^{*}-algebras. In particular, the work of Bratteli and Jorgensen [15] uses such representations to generate wavelets, and unitary invariants for special classes of representations are central to their work. In this article, we will survey results about free semigroup algebras themselves, with passing reference to various applications.

1. Connections to C*-Algebras

Let S_1, \ldots, S_n denote an *n*-tuple of isometries with pairwise orthogonal ranges. The orthogonality relations are given algebraically by $S_i^* S_j = \delta_{ij} I$ for $1 \leq i, j \leq n$, or equivalently by

$$S_i^* S_i = I$$
 for $1 \le i \le n$ and $\sum_{i=1}^n S_i S_i^* \le I$.

The C*-algebra generated by such an *n*-tuple was introduced by Cuntz [21]. There are only two possibilities. When $\sum_{i=1}^{n} S_i S_i^* = I$, the C*-algebra is *-isomorphic to the Cuntz algebra \mathcal{O}_n , which is simple and purely infinite. When $\sum_{i=1}^{n} S_i S_i^* < I$, there is again only one *-isomorphism class of C*algebras, known as the Cuntz-Toeplitz algebra \mathcal{E}_n . This algebra contains an ideal \mathcal{K} generated by $I - \sum_{i=1}^{n} S_i S_i^*$ which is isomorphic to the compact operators \mathfrak{K} . This is the only proper ideal, and $\mathcal{E}_n/\mathcal{K}$ is isomorphic to \mathcal{O}_n .

Perhaps because there are only two possibilities for this C*-algebra, it is difficult to use the C*-algebraic structure to determine spatial invariants (up to unitary equivalence). A famous theorem of Glimm [32] states that,

¹⁹⁹¹ Mathematics Subject Classification. 47L80.

The author was partially supported by an NSERC grant.

at least considering the parameterization of (cyclic) representations via the state space and the GNS construction, it is essentially impossible to determine complete unitary invariants. More precisely, Glimm's theorem states that for any non-type I C*-algebra, there is no countable family of Borel functions on the state space which distinguish the corresponding representations up to unitary equivalence. As the only simple type I algebras are the finite matrix algebras \mathfrak{M}_n and the compact operators \mathfrak{K} , it follows that this classification is impossible for most C*-algebras.

Nevertheless, there are good reasons for wanting to do this. We mention two. Bob Powers [50] introduced the study of WOT-continuous *endomorphisms of $\mathcal{B}(\mathcal{H})$. Such a map is determined by what it does to \mathfrak{K} , which has only one irreducible representation up to unitary equivalence, the identity map. So the restriction of this endomorphism π to \mathfrak{K} is just a multiple, say n, of the identity map. This extended integer n is known as the *Powers index*. Note then that there are n isometries S_i for $1 \leq i \leq n$ so that $\pi(A) = \sum_{i=1}^n S_i A S_i^*$. These isometries are not unique, but they are unique up to an action of the $n \times n$ unitary group \mathfrak{U}_n , which amounts to the choice of a basis for the range of $\pi(E_{11})$, where E_{11} is a matrix unit of \mathfrak{K} . From this one finds that the C*-algebra generated by these S_1, \ldots, S_n is unique, and the choice of generators is determined up to a so-called *gauge automorphism*. See [37, 18, 14] for further information.

The second application is work of Bratteli, Jorgensen and others [14, 15, 17, 34, 13] using representations of the Cuntz algebra to generate wavelets. They introduced [14] a class of representations known as *finitely correlated representations* in connection with the endomorphisms above. We shall see can that they be completely classified by exploiting a connection with dilation theory in our work. Their isometries act on $L^2(\mathbb{T})$, and have the form $S_i f(z) = m_i(z) f(z^n)$ where m_i are functions of modulus 1 satisfying the orthogonality relations

$$\sum_{k=1}^{n} m_i(z\omega^k) \overline{m_j(z\omega^k)} = \delta_{ij} n$$

where $\omega = e^{2\pi i/n}$ [16]. These are used to generate wavelet bases of $L^2(\mathbb{R})$ with n-1 mother wavelets under integer translation and *n*-fold dilation. The interested reader is referred to the papers cited in the references.

We make an elementary observation about the representation theory of \mathcal{E}_n . Suppose that π is a *-representation of \mathcal{E}_n . The restriction to \mathcal{K} must be equivalent to a multiple α of the identity representation because \mathcal{K} is isomorphic to the compact operators. Standard C*-algebra theory (c.f.[23]) shows that the restriction of π to $\overline{\pi(\mathcal{K})\mathcal{H}}$ is also unitarily equivalent to $\mathrm{id}^{(\alpha)}$, and the restriction to the complement factors through $\mathcal{E}_n/\mathcal{K} \simeq \mathcal{O}_n$. Hence $\pi \simeq \mathrm{id}^{(\alpha)} \oplus \sigma q$ where q is the quotient map onto \mathcal{O}_n and σ is a representation of \mathcal{O}_n . Below we shall see that this decomposition is equivalent to the Wold decomposition.

2. Definitions and Some Examples

The free semigroup algebra generated by an *n*-tuple of isometries with orthogonal ranges $S = (S_1, \ldots, S_n)$ is the unital wor-closed algebra $\mathfrak{S} = \overline{\operatorname{Alg}\{S_1, \ldots, S_n\}}^{WOT}$.

The fundamental example is the left regular representation of the free semigroup \mathbb{F}_n^+ consisting of all words in n non-commuting letters $\{1, 2, \ldots, n\}$. The empty word \emptyset is the identity element for \mathbb{F}_n^+ . The Hilbert space $\mathcal{K}_n = \ell^2(\mathbb{F}_n^+)$ has the orthonormal basis $\{\xi_w : w \in \mathbb{F}_n^+\}$. Isometries L_v are defined for every word $v \in \mathbb{F}_n^+$ by $L_v \xi_w = \xi_{vw}$. In particular, the generators determine the n-tuple of isometries L_1, \ldots, L_n . Notice that the L_i have orthogonal ranges spanned by all words beginning with the letter ifor $1 \leq i \leq n$, and that the vector ξ_{\emptyset} spans the complement of the sum $\sum_{i=1}^n L_i \mathcal{K}_n$. The free semigroup algebra generated by $L = (L_1, \ldots, L_n)$ is denoted by \mathfrak{L}_n and is called the *non-commutative analytic Toeplitz algebra*.

The semigroup \mathbb{F}_n^+ is graded by the length function |w| of a word w, since |vw| = |v| |w|. The space \mathcal{K}_n is known as Fock space, and may be decomposed into a direct sum of subspaces $\mathcal{K}_n = \sum_{k\geq 0}^{\oplus} \mathcal{H}_k$ where $\mathcal{H}_k =$ $\operatorname{span}\{\xi_w : |w| = k\}$. Then $\dim \mathcal{H}_k = n^k$. Thus $\mathcal{H}_0 = \mathbb{C}, \ \mathcal{H}_1 \simeq \ell_n^2$ and $\mathcal{H}_k \simeq \mathcal{H}_1^{\otimes k}$, the tensor product of k copies of n-dimensional space. The isometries L_i are known as creation operators in this context, and are written by mathematical physicists as $L_i \zeta = \xi_i \otimes \zeta$ in this presentation of \mathcal{K}_n .

Since $P_0 := I - \sum_{i=1}^n L_i L_i^* = \xi_0 \xi_0^*$ is non-zero, it follows that $C^*(L) = \mathcal{E}_n$. Note also that \mathbb{F}_n^+ has a right regular representation given by $R_i \xi_w = \xi_{wi}$ for $1 \leq i \leq n$. It is routine to verify that $R_v \xi_w = \xi_{w\tilde{v}}$ where \tilde{v} is the word v in reverse order. Clearly the free semigroup algebra \mathfrak{R}_n generated by $R = (R_1, \ldots, R_n)$ commutes with \mathfrak{L}_n . Moreover this algebra is unitarily equivalent to \mathfrak{L}_n via the unitary $W \xi_w = \xi_{\tilde{w}}$.

When n = 1, this representation just yields the unilateral shift. In this case, \mathfrak{L}_1 is just the classical analytic Toeplitz algebra on H^2 isomorphic to H^{∞} of the unit disk. Also $\mathfrak{R}_1 = \mathfrak{L}_1 = \mathfrak{L}'_1$. When $n \geq 2$, there is much analogous analytic structure associated to the unit ball in \mathbb{C}^n which will be examined in detail later. At this point, we content ourselves with a simple observation. Each element $A \in \mathfrak{L}_n$ is determined by $A\xi_{\varnothing} = \sum_{v \in \mathbb{F}_n^+} a_v \xi_v$ because

$$A\xi_w = AR_{\widetilde{w}}\xi_{\varnothing} = R_{\widetilde{w}}A\xi_{\varnothing} = \sum_{v \in \mathbb{F}_n^+} a_v\xi_{vw}.$$

At least when this is a finite sum, it follows that $A = \sum_{v \in \mathbb{F}_n^+} a_v L_v$. In general, we think of this infinite sum as a Fourier series of analytic type.

In fact, if we define maps $\Phi_k(A) = \sum_{|v|=k} a_w L_w$, then Cesaro means may be defined by $\Sigma_k(A) = \sum_{j=0}^{k-1} \left(1 - \frac{j}{k}\right) \Phi_j(A)$. These are completely positive unital maps of \mathfrak{L}_n into itself which converge to the identity map in the pointwise strong operator topology [27, Lemma 1.1].

Now suppose that $S = (S_1, \ldots, S_n)$ is an arbitrary set of isometries with orthogonal ranges. We shall write S_v for the monomial $v(S) = S_{i_1} \ldots S_{i_k}$ for any word $v = i_1 \ldots i_k \in \mathbb{F}_n^+$. A wandering subspace is a subspace \mathcal{W} such that the subspaces $\{S_v \mathcal{W} : v \in \mathbb{F}_n^+\}$ are pairwise orthogonal. The span of these subspaces is the \mathfrak{S} -invariant subspace $\mathfrak{S}[\mathcal{W}]$ generated by \mathcal{W} . Clearly the restriction of the isometries S_i to $\mathfrak{S}[\mathcal{W}]$ is unitarily equivalent to a direct sum of dim \mathcal{W} copies of the left regular representation. We shall see that such subspaces are omnipresent in all known examples.

Popescu [42] established the analogue of the Wold decomposition in this context, and the proof is essentially the same as for a single isometry. Let $\mathcal{W} = (I - \sum_{i=1}^{n} S_i S_i^*) \mathcal{H}$. It is readily verified that \mathcal{W} is a wandering space and that $\mathfrak{S}[\mathcal{W}]$ is a *reducing* subspace. The restriction to $\mathfrak{S}[\mathcal{W}]$ is a multiple of the left regular representation, as noted above, and the restriction to the complement yields a representation of the Cuntz algebra because the sum of the ranges there is the whole space. The C*-algebraic view of this was mentioned in the previous section.

A representation which is a multiple of the left regular representation is called *pure*, and a representation which generates the Cuntz algebra is said to be of *Cuntz type*. There are many representations of the latter type. We provide a few to keep in mind.

Example 2.1. Let $u = i_1 \dots i_k$ be a word in \mathbb{F}_n^+ and let $\lambda \in \mathbb{T}$ be a scalar of modulus one. Define a Hilbert space $\mathcal{K}_u \simeq \mathbb{C}^k \oplus \mathcal{K}_n^{k(n-1)}$ with orthonormal basis ζ_1, \dots, ζ_k for \mathbb{C}^k and index the copies of \mathcal{K}_n by (s,j), where $1 \leq s \leq k$, $1 \leq j \leq n$ and $j \neq i_s$, with basis $\{\xi_{s,j,w} : w \in \mathbb{F}_n^+\}$. Define a representation $\sigma_{u,\lambda}$ of \mathbb{F}_n^+ and isometries $S_i = \sigma_{u,\lambda}(i)$ by

$$S_i \zeta_s = \zeta_{s-1} \quad \text{if} \quad i = i_s, \, s > 1$$

$$S_i \zeta_1 = \lambda \zeta_k \quad \text{if} \quad i = i_1$$

$$S_i \zeta_s = \xi_{s,i,\emptyset} \quad \text{if} \quad i \neq i_s$$

$$S_i \xi_{s,i,w} = \xi_{s,i,iw} \quad \text{for all} \quad i, s, j, w$$

Notice that each $\xi_{s,j,\emptyset}$ is a wandering vector generating the (s,j)th copy of \mathcal{K}_n . The sum of the ranges of the S_i 's is the whole space, so this is a Cuntz representation. The copy of \mathbb{C}^k is invariant for the S_i^* 's and each basis vector ζ_s is a cyclic vector for \mathfrak{S} . Think of this as a ring of k basis vectors being permuted around the ring by the appropriate S_{i_s} at each point. Off of each node ζ_s there are n-1 copies of Fock space, one for each $i \neq i_s$.

The reader can quickly verify that every wandering vector is orthogonal to span{ ζ_1, \ldots, ζ_k }. Thus the span of all wandering vectors is precisely span{ ζ_1, \ldots, ζ_k }^{\perp}. We shall be able to detect this in \mathfrak{S} because it turns out that the projection P onto span{ ζ_1, \ldots, ζ_k } belongs to \mathfrak{S} , and is the largest projection in \mathfrak{S} such that $P\mathfrak{S}P$ is self-adjoint. This will allow us to completely describe the algebra \mathfrak{S} , and to decompose the corresponding representation of \mathcal{O}_n into a direct sum of irreducible representations. In particular, notice that the restriction of \mathfrak{S} to $P^{\perp}\mathcal{H}$ is unitarily equivalent

to the direct sum of k(n-1) copies of \mathfrak{L}_n . Since $P \in \mathfrak{S}$, we need only understand $\mathfrak{S}P$ to have a complete picture of the algebra \mathfrak{S} . This turns out to be $\mathfrak{W}P$ where \mathfrak{W} is the von Neumann algebra generated by S. Moreover because $P\mathfrak{W}P$ is the finite dimensional C*-algebra on $P\mathcal{H}$ generated by the restrictions of the S_i^* 's, it will be straightforward to completely analyze this example.

Example 2.2. Let $x = i_1 i_2 i_3 \dots$ be an infinite word in $\{1, \dots, n\}$. Define a sequence $x_m = i_1 i_2 \dots i_m$ for $m \ge 0$. Let $\mathbb{F}_n^+ x^{-1}$ denote the collection of words in the free group on n generators of the form $v = u x_m^{-1}$ for u in \mathbb{F}_n^+ and some $m \ge 0$. Identify words which are the same after cancellation, namely $u x_m^{-1} = (u i_{m+1}) x_{m+1}^{-1}$. Let \mathcal{H}_x be the Hilbert space with orthonormal basis $\{\xi_v : v \in \mathbb{F}_n^+ x^{-1}\}$. Define a representation π_x of \mathbb{F}_n^+ and isometries $S_i = \pi_x(i)$ on \mathcal{H}_x by $\pi_x(w) \xi_v = \xi_{wv}$ for $v \in \mathbb{F}_n^+ x^{-1}$ and $w \in \mathbb{F}_n^+$.

This is also a Cuntz representation because it is readily apparent that each basis vector is the image of some other basis vector, whence the sum of the ranges of the S_i 's is the whole space. Indeed, notice that each $\xi_{x_m^{-1}}$ is a wandering vector which generates $\ell^2(\mathbb{F}_n^+x_m^{-1})$. Consequently \mathcal{H}_x is expressed as the increasing union of a nested sequence of invariant subspaces on which \mathfrak{S} acts like the left regular representation. Such Cuntz representations are in a natural sense inductive limits of (possibly multiples of) the left regular representation, and hence are said to be of *inductive type*. In particular, it is easy to convince yourself that the algebra \mathfrak{S} is canonically isometrically isomorphic to \mathfrak{L}_n .

3. DILATION THEORY

The Sz.Nagy dilation theorem [56] shows that an arbitrary contraction on a Hilbert space \mathcal{H} dilates to an isometry on a larger space. Franzo [31] and Bunce [19] observed that there is a natural analogue for a contractive *n*-tuple $A = (A_1, \ldots, A_n)$. There is no commutativity required, only the norm condition $||A|| = ||\sum A_i A_i^*||^{1/2} \leq 1$. Perhaps the simplest proof uses the Schaeffer construction. Namely consider $D_A = (I_n - A^*A)^{1/2}$ which acts on the direct sum $\mathcal{H}^{(n)}$ of *n* copies of \mathcal{H} (where I_n is the identity operator on $\mathcal{H}^{(n)}$). Then $\begin{bmatrix} A \\ D_A \end{bmatrix}$ is an isometry from $\mathcal{H}^{(n)}$ into $\mathcal{H} \oplus \mathcal{H}^{(n)}$. Thus the *n* columns $\begin{bmatrix} A_i \\ X_i \end{bmatrix}$ of this isometry, for $1 \leq i \leq n$, are isometries from \mathcal{H} into $\mathcal{H} \oplus \mathcal{H}^{(n)}$ with pairwise orthogonal ranges. Consider the Hilbert space $\mathcal{K} = \mathcal{H} \oplus (\mathcal{H}^{(n)} \otimes \mathcal{K}_n)$, and define isometries

$$S_i = \begin{bmatrix} A_i & 0\\ X_i & I_n \otimes L_i \end{bmatrix}$$

There is a slight abuse of notation where we identify X_i with its composition with the natural inclusion of $\mathcal{H}^{(n)}$ onto $\mathcal{H}^{(n)} \otimes \mathbb{C}\xi_{\emptyset}$ as a subspace of $\mathcal{H}^{(n)} \otimes \mathcal{K}_n$.

Popescu [42] showed that there is a unique minimal dilation, and extends this argument to a countable row contraction. The minimal dilation is obtained from the construction above by restricting each S_i to $\operatorname{span}\{S_v\mathcal{H}: v \in \mathbb{F}_n^+\}$, which is evidently the minimal invariant subspace for the S_i 's containing \mathcal{H} in this dilation.

Popescu pursues the analogy with the one-variable case and finds many parallels with the Sz.Nagy–Foiaş theory. In particular, he establishes the analogue of the commutant lifting theorem [44]: if X is a contraction commuting with A_1, \ldots, A_n , then X dilates to a contraction Y on the larger space which commutes with the dilation S_1, \ldots, S_n .

There is a natural analogue [45] of the von Neumann inequality. The point is the same as Sz.Nagy's proof in one variable, namely if p is a non-commuting polynomial in *n*-variables, then the 1, 1 entry of $p(S_1, \ldots, S_n)$ is just $p(A_1, \ldots, A_n)$. Hence

$$||p(A_1,\ldots,A_n)|| \le ||p(S_1,\ldots,S_n)||$$

Moreover the right hand side does not depend on the particular dilation! Indeed, this norm may be computed in $C^*(S)$, and so there are only two possibilities, namely \mathcal{E}_n and \mathcal{O}_n . Since \mathcal{O}_n is a quotient of \mathcal{E}_n , the norm is apparently at least as large in \mathcal{E}_n . Thus

$$||p(A_1,\ldots,A_n)|| \le ||p(L_1,\ldots,L_n)||.$$

However on the norm closed unital algebra \mathcal{A}_n generated by L_1, \ldots, L_n , the quotient map is completely isometric. Popescu calls this algebra \mathcal{A}_n the non-commutative disk algebra.

Evidently von Neumann's inequality allows the immediate construction of a functional calculus on \mathcal{A}_n for every row contraction A. Following Arveson's approach to non-commutative dilation theory [6, 7], it follows that there is a unique unital completely positive map from \mathcal{O}_n onto $C^*(A_1, \ldots, A_n)$ which sends S_i onto A_i for $1 \leq i \leq n$. In [49], Popescu develops an analogue of the Poisson kernel to provide an explicit formula for this map. Since \mathcal{O}_n is simple, it is the smallest C*-algebra which contains \mathcal{A}_n completely isometrically. By Hamana's Theorem [33], there is a unique minimal C*algebra containing any unital operator algebra completely isometrically, which Arveson named the C*-envelope. Hence \mathcal{O}_n is the C*-envelope of \mathcal{A}_n . This shows in particular that the algebras \mathcal{A}_n are not isomorphic for different values of n [48].

In the one variable case, it is desirable when possible to extend the disk algebra functional calculus to $H^{\infty}(\mathbb{D})$. This requires the contraction to dilate to an absolutely continuous isometry (meaning that the unitary part of the Wold decomposition has a spectral measure which is absolutely continuous with respect to Lebesgue measure on the circle). Popescu [45, 46] identifies the corresponding class of row contractions and extends the functional calculus to \mathfrak{L}_n , which Popescu calls \mathcal{F}_n^{∞} . A special case which is especially

tractable occurs when ||A|| = r < 1. In this instance, Bunce [19] basically showed that there is a unique way to extend the functional calculus which sends L_i to A_i to all of \mathfrak{L}_n by setting $\Phi(X) = \sum_{w \in \mathbb{F}_n^+} x_w w(A)$ for $X = \sum_{w \in \mathbb{F}_n^+} x_w L_w$ in \mathfrak{L}_n . The strict contraction condition may be used to show that $||\sum_{|w|=k} x_w w(A)|| \le r^k ||X||$, and thus $\Phi(X)$ is a norm convergent series.

We also mention [41, 43] as related papers of Popescu developing Sz.Nagy– Foiaş models.

4. The Spatial Structure of \mathfrak{L}_n

In this section, we develop the spatial structure theory of the non-commutative analytic Toeplitz algebra \mathfrak{L}_n . The results of this section are due to Popescu [43, 47], myself and Pitts [27] and Arias and Popescu [4]. Our goal is to stress the analogue with the analytic Toeplitz algebra $\mathcal{T}(H^{\infty})$ generated by the unilateral shift.

Suppose that A is an operator commuting with the right regular algebra \mathfrak{R}_n . Then $A\xi_{\varnothing} = \sum_{v \in \mathbb{F}_n^+} a_v \xi_v$. The Cesaro means defined in section 2 suggest that we consider the sequence $\sum_{|v| < k} \left(1 - \frac{|v|}{k}\right) a_v L_v$. These converge strongly to A, which shows that the commutant of \mathfrak{R}_n is just \mathfrak{L}_n . Conversely $\mathfrak{L}'_n = \mathfrak{R}_n$ as these algebras are unitarily equivalent.

This leads to several immediate consequences, such as the fact that \mathfrak{L}_n is inverse closed, i.e. if $A \in \mathfrak{L}_n$ is invertible in $\mathcal{B}(\mathcal{K}_n)$, then the inverse lies in \mathfrak{L}_n . Also since \mathfrak{R}_n contains isometries with orthogonal ranges, it is immediate that $||A||_e = ||A||$ for all $A \in \mathfrak{L}_n$. In particular this establishes the fact that the quotient from \mathcal{E}_n to \mathcal{O}_n is (completely) isometric on \mathcal{A}_n .

With a bit more work, we see that \mathfrak{L}_n contains no normal elements which are not scalar. To see this, notice that ξ_{\varnothing} is in the kernel of each L_i^* , and thus is an eigenvector for every A^* with $A \in \mathfrak{L}_n$. Indeed, $A^*\xi_{\varnothing} = \overline{a_{\varnothing}}\xi_{\varnothing}$. Suppose that A were normal. Then ξ_{\varnothing} would also be an eigenvector for Aand so $A\xi_{\varnothing} = a_{\varnothing}\xi_{\varnothing}$. From the Fourier series, it follows that $A = a_{\varnothing}I$.

A more careful study shows [27, Theorem 1.7] that every non-zero element of \mathfrak{L}_n is injective and has non-zero connected essential spectrum $\sigma_e(A) = \sigma(A)$. In particular, \mathfrak{L}_n contains no proper projections or nonzero quasinilpotents. Hence \mathfrak{L}_n is a semisimple algebra.

The first deep connection to function theory is an analogue of the Beurling Theorem for \mathfrak{L}_n [43] (c.f. [27]). The Beurling Theorem [10] says that every invariant subspace of the unilateral shift represented on H^2 as the Toeplitz operator T_z has the form wH^2 where w is an inner function ($w \in H^{\infty}(\mathbb{D})$) and $|w(e^{i\theta})| = 1$ a.e.). These subspaces are always cyclic (w is the cyclic vector) and they are the range of the isometry T_w . Now the only isometries in the Toeplitz algebra $\mathcal{T}(H^{\infty})$ are of the form T_w for w inner, and $\mathcal{T}(H^{\infty})$ is maximal abelian. So Beurling's theorem may be restated as saying that the invariant subspaces of the unilateral shift are the ranges of isometries in the commutant. This is the form which generalizes.

Theorem 4.1. Every invariant subspace of \mathfrak{L}_n is the direct sum of cyclic subspaces. The cyclic invariant subspaces of \mathfrak{L}_n are precisely the ranges of isometries in \mathfrak{R}_n ; and the choice of isometry is unique up to a scalar.

The proof starts with an invariant subspace \mathcal{M} and forms the wandering space $\mathcal{W} = \mathcal{M} \ominus \left(\sum_{i} L_{i}\mathcal{M}\right)$. The grading on Fock space makes it straightforward to see that there can be no Cuntz part, and thus \mathcal{M} splits into a direct sum of cyclic subspaces obtained by choosing an orthonormal basis for \mathcal{W} . For each such vector ζ , one constructs an isometry R_{ζ} in \mathfrak{R}_{n} by setting $R_{\zeta}\xi_{w} = L_{w}\zeta$ and extending by linearity. This is the desired isometry onto the cyclic subspace $\mathfrak{L}_{n}[\zeta]$.

This naturally leads to an inner-outer factorization as follows. Take any $A \in \mathfrak{L}_n$. The subspace $\overline{\operatorname{Ran}(A)}$ is invariant for \mathfrak{R}_n with cyclic vector $A\xi_{\varnothing}$. Thus it is the range of an isometry $L \in \mathfrak{L}_n$. Therefore A factors as A = LB where $B = L^*A$. One verifies that B commutes with each R_i , and thus lies in \mathfrak{L}_n . Moreover B evidently has dense range, and by analogy is called an *outer operator*.

Like $\mathcal{T}(\mathcal{H}^{\infty})$, the algebra \mathfrak{L}_n has a rich collection of invariant subspaces of co-dimension one. These correspond to eigenvectors for the adjoint algebra \mathfrak{L}_n^* . When n = 1, each point $\lambda \in \mathbb{D}$ determines a kernel function $k_{\lambda} = (1 - |\lambda|^2)^{1/2}(1 - \overline{\lambda}z)^{-1}$ which are the eigenvectors for the backward shift $T_z^*k_{\lambda} = \overline{\lambda}k_{\lambda}$. This vector yields a weak-* continuous point evaluation $h(\lambda) = \langle T_h k_{\lambda}, k_{\lambda} \rangle$ for $h \in H^{\infty}$. Exactly the same thing occurs for $n \geq 2$ corresponding to each point λ in the unit ball \mathbb{B}_n of \mathbb{C}^n [27, Theorem 2.6].

Theorem 4.2. The eigenvectors for \mathfrak{L}_n^* are the vectors

$$\nu_{\lambda} = (1 - \|\lambda\|^2)^{1/2} \sum_{w \in \mathbb{F}_n^+} \overline{w(\lambda)} \xi_w = (1 - \|\lambda\|^2)^{1/2} \left(I - \sum_{i=1}^n \overline{\lambda_i} L_i\right)^{-1} \xi_{\emptyset}$$

for λ in the unit ball \mathbb{B}_n . They satisfy $L_i^*\nu_\lambda = \overline{\lambda_i}\nu_\lambda$ for $1 \leq i \leq n$.

Moreover $\langle p(L)\nu_{\lambda},\nu_{\lambda}\rangle = p(\lambda)$ for every polynomial $p = \sum_{w} a_{w}w$ in the semigroup algebra \mathbb{CF}_{n}^{+} . This extends to a WOT-continuous multiplicative linear functional on \mathfrak{L}_{n} given by $\varphi_{\lambda}(A) = \langle A\nu_{\lambda},\nu_{\lambda}\rangle$.

The vector ν_{λ} is cyclic for \mathfrak{L}_n . The subspace $\{\nu_{\lambda}\}^{\perp}$ is \mathfrak{L}_n -invariant, and its wandering subspace is n-dimensional.

The general issue of the structure of weak-* continuous functionals on operator algebras and the connection to reflexivity has its roots in the Scott Brown technique c.f.[11]. The fact that $\mathcal{L}'_n = \mathfrak{R}_n$ contains isometries with orthogonal ranges makes the following result straightforward [29]. A somewhat more delicate analysis [12] shows that \mathcal{L}_n has the related property $X_{0,1}$ which we will not define.

Theorem 4.3. Let φ be a weak-* continuous linear functional on the spatial tensor product $\mathcal{B}(\mathcal{H}) \otimes \mathfrak{L}_n$ with $\|\varphi\| < 1$. Then there are unit vectors η and ζ so that $\varphi(A) = \langle A\eta, \zeta \rangle$ for all $A \in \mathcal{B}(\mathcal{H}) \otimes \mathfrak{L}_n$.

In particular, every weak-* continuous functional on \mathfrak{L}_n is given by a vector functional. So one obtains the important consequence that the weak-* and WOT topologies coincide on \mathfrak{L}_n .

We now have in hand ample information to study reflexivity. An operator algebra \mathfrak{A} is *reflexive* if the algebra can be recovered from its invariant subspace lattice $\mathcal{L} = \text{Lat}(\mathfrak{A})$ as the set $\text{Alg}(\mathcal{L})$ of all operators leaving each subspace invariant. The lattice \mathcal{L} determines a seminorm on $\mathcal{B}(\mathcal{H})$ by

$$\beta_{\mathcal{L}}(T) := \sup_{L \in \mathcal{L}} \|L^{\perp}TL\|.$$

Clearly, $\beta_{\mathcal{L}}(T) = 0$ precisely when T belongs to Alg(\mathcal{L}). Moreover, it is elementary to show that

$$\beta_{\mathcal{L}}(T) \leq \operatorname{dist}(T, \operatorname{Alg}(\mathcal{L})) \text{ for all } T \in \mathcal{B}(\mathcal{H}).$$

The algebra is said to be *hyper-reflexive* if these norms are comparable. In this case, the hyper-reflexivity constant is the smallest number C such that

$$\operatorname{dist}(T, \operatorname{Alg}(\mathcal{L})) \leq C\beta_{\mathcal{L}}(T) \quad \text{for all} \quad T \in \mathcal{B}(\mathcal{H}).$$

When an operator \mathfrak{A} is the commutant of another algebra, there is another natural measure of the distance to \mathfrak{A} . Define the derivation δ_T by $\delta_T(X) = XT - TX$. Notice that the restriction of δ_T to \mathfrak{A}' is zero if and only if Tbelongs to $\mathfrak{A}'' = \mathfrak{A}$. Suppose that dist $(T, \mathfrak{A}) = d$ and choose $A \in \mathfrak{A}$ such that ||T - A|| = d. This is possible because \mathfrak{A} is wor-closed. Then

$$\|\delta_T|_{\mathfrak{A}'}\| = \|\delta_{T-A}|_{\mathfrak{A}'}\| \le 2\|T-A\| = 2d.$$

On the other hand, it is not automatic that the distance to \mathfrak{A} is bounded by a constant times $\|\delta_T\|_{\mathfrak{A}}\|$.

The list of algebras known to be hyper-reflexive is rather short. It includes nest algebras which have constant 1 due to the Arveson Distance Formula [8]. Christensen [20] showed that injective von Neumann algebras have constant at most 4. Von Neumann algebras with abelian commutant have constant at most 2, as do abelian von Neumann algebras [52]. Von Neumann algebras are also commutants. Since von Neumann algebras are spanned by their projections, it is easy to relate the quantities $\beta_{\text{Lat }\mathfrak{A}}(T)$ and $\|\delta_T\|_{\mathfrak{A}'}\|$ in this situation. Lastly, the case most closely related to our study is the analytic Toeplitz algebra $\mathcal{T}(H^{\infty})$, which has distance constant at most 19 [22].

The non-commutative analytic Toeplitz algebras \mathcal{L}_n are reflexive [4]. Indeed they are hyper-reflexive [27]. The constant we found was 51. However Bercovici [12] obtained a beautiful general result that yields distance constant 3 for any algebra with property $X_{0,1}$. We have:

Theorem 4.4. The algebras \mathfrak{L}_n are hyper-reflexive. Moreover, for all T in $\mathcal{B}(\mathcal{K}_n)$,

$$\frac{1}{3}\operatorname{dist}(T,\mathfrak{L}_n) \leq \sup_{L \in \operatorname{Lat}(\mathfrak{L}_n)} \|L^{\perp}TL\| \leq \|\delta_T|_{\mathfrak{R}_n}\| \leq 2\operatorname{dist}(T,\mathfrak{L}_n).$$

5. The Algebraic Structure of \mathfrak{L}_n

This section deals with [28] which develops the algebraic structure of \mathfrak{L}_n , culminating in a description of the automorphism group, and ties the algebra even more strongly to analytic function theory on the complex *n*-ball.

5.1. **Ideals.** First consider WOT-closed ideals. Let $\mathrm{Id}_r(\mathfrak{L}_n)$, $\mathrm{Id}_l(\mathfrak{L}_n)$ and $\mathrm{Id}(\mathfrak{L}_n)$ denote the sets of all WOT-closed right, left and two-sided ideals respectively. Suppose that \mathfrak{J} belongs to $\mathrm{Id}_r(\mathfrak{L}_n)$. Observe that the subspace $\overline{\mathfrak{J}\xi_{\varnothing}}$ belongs to $\mathrm{Lat}\,\mathfrak{R}_n$. To see this, note that

$$\mathfrak{R}_n\overline{\mathfrak{J}\xi_arnothing}=\overline{\mathfrak{J}\mathfrak{R}_n\xi_arnothing}=\overline{\mathfrak{J}\mathcal{K}_n}=\overline{\mathfrak{J}\mathfrak{L}_n\xi_arnothing}=\overline{\mathfrak{J}\xi_arnothing}.$$

Thus $\overline{\mathfrak{J}\xi_{\varnothing}} = \overline{\mathfrak{J}\mathcal{K}_n}$ is the range of \mathfrak{J} and is \mathfrak{R}_n invariant. Similarly when \mathfrak{J} belongs to $\mathrm{Id}_{\ell}(\mathfrak{L}_n)$, we have $\mathfrak{L}_n\mathfrak{J}\xi_{\varnothing} = \mathfrak{J}\xi_{\varnothing}$; so $\overline{\mathfrak{J}\xi_{\varnothing}}$ is \mathfrak{L}_n invariant. Hence when \mathfrak{J} is a two-sided ideal, $\overline{\mathfrak{J}\xi_{\varnothing}}$ belongs to $\mathrm{Lat}(\mathfrak{L}_n) \cap \mathrm{Lat}(\mathfrak{R}_n)$. We define the map $\mu(\mathfrak{J}) = \overline{\mathfrak{J}\xi_{\varnothing}}$.

For the moment, we consider right ideals. Left ideals are *not* handled in the same way. Suppose that \mathcal{M} is an invariant subspace for \mathfrak{R}_n . Define $\iota(\mathcal{M}) = \{J \in \mathfrak{L}_n : J\xi_{\varnothing} \in \mathcal{M}\}$. Clearly this is a WOT-closed subspace of \mathfrak{L}_n . For any $J \in \iota(\mathcal{M})$ and $A \in \mathfrak{L}_n$,

$$JA\xi_{\varnothing} \in \overline{J\mathcal{K}_n} = \overline{J\mathfrak{R}_n\xi_{\varnothing}} = \overline{\mathfrak{R}_nJ\xi_{\varnothing}} \subset \mathcal{M}.$$

Whence $\iota(\mathcal{M})$ is a right ideal. Likewise, if $\mathcal{M} \in \operatorname{Lat}(\mathfrak{L}_n)$, then $\iota(\mathcal{M})$ is a wort-closed left ideal.

It is easy to see that μ respects sums: $\mu(\mathfrak{J}_1 + \mathfrak{J}_2) = \mu(\mathfrak{J}_1) \lor \mu(\mathfrak{J}_2)$. However it also respects intersections, which relies on the fact that μ turns out to be a bijection. This fact relies on the Beurling Theorem 4.1 and Theorem 4.3 on weak-* continuous functionals.

Theorem 5.1. Let $\mu : \operatorname{Id}_r(\mathfrak{L}_n) \to \operatorname{Lat}(\mathfrak{R}_n)$ be given by $\mu(\mathfrak{J}) = \overline{\mathfrak{J}\xi_{\varnothing}}$. Then μ a complete lattice isomorphism. The restriction of μ to the set $\operatorname{Id}(\mathfrak{L}_n)$ is a complete lattice isomorphism onto $\operatorname{Lat} \mathfrak{L}_n \cap \operatorname{Lat} \mathfrak{R}_n$. The inverse map is ι .

It turns out that there are significant differences between right and left ideals. The tight correspondence between right ideals and invariant subspaces, which also works well for two sided ideals, has no good corresponding result for left ideals. The reason is that there are good factorization results for elements of \mathfrak{L}_n with isometries on the left, but not on the right. Indeed, notice that if L is an isometry, then ||LA|| = ||A||, but ||AL|| may be much smaller. Kribs [**35**] found significant pathology in the structure of left ideals which arose because of strange factorization results. For example, the isometry L_1 cannot be properly factored as $L_1 = AB$ if ||A|| = ||B|| = 1. However it can be properly factored in \mathfrak{L}_n if the norm condition is dropped. Results of this kind are used to show that the map μ is neither one-to-one nor surjective from $\mathrm{Id}_l(\mathfrak{L}_n)$ to $\mathrm{Lat}(\mathfrak{L}_n)$.

A weak-* continuous character on \mathfrak{L}_n has a kernel which is a weak-* closed two-sided ideal of co-dimension one. As the weak-* and WOT topologies coincide, it is WOT-closed. Its range is an invariant subspace of co-dimension one, and thus it determines an eigenvector for \mathfrak{L}_n^* . This leads to the conclusion that the only weak-* continuous characters are the maps φ_{λ} for $\lambda \in \mathbb{B}_n$ which we found in the previous section.

One important ideal which plays a central role is the WOT-closed commutator ideal \mathfrak{C} . Let \mathcal{K}_n^s denote symmetric Fock space, which consists of all vectors of the form $\zeta = \sum_{w \in \mathbb{F}_n^+} a_w \xi_w$ such that $a_w = a_v$ if w(z) = v(z), where evaluation is taken at a commuting *n*-tuple $z = (z_1, \ldots, z_n)$ of indeterminates. In other words, if |w| = k and $\sigma \in S_k$ is a permutation of ksymbols and if $\sigma(w)$ denotes the permutation of the k letters in w via σ , then $a_{\sigma(w)} = a_w$. In particular, the eigenvectors ν_{λ} are evidently symmetric, and in fact they span \mathcal{K}_n^s . The main result about \mathfrak{C} is that

$$\mathfrak{C} = \bigcap_{\lambda \in \mathbb{B}_n} \ker \varphi_\lambda \quad \text{ and } \quad \mu(\mathfrak{C}) = \mathcal{K}_n^{s \perp}.$$

Another set of important ideals are denoted $\mathfrak{L}_n^{0,k}$. The ideal $\mathfrak{L}_n^0 = \ker \varphi_0$ consists of those elements whose leading Fourier coefficient vanishes. The ideals $\mathfrak{L}_n^{0,k}$ are the *k*th powers of this ideal, and consist of elements whose Fourier coefficients vanish up to order k - 1. It is clear that the range $\mu(\mathfrak{L}_n^{0,k}) = \operatorname{span}\{\xi_w : |w| \ge k\}$. This subspace has wandering dimension n^k . The ideal structure can be used to show that every element of $\mathfrak{L}_n^{0,k}$ factors as a sum $A = \sum_{|w|=k} L_w A_w = LB$ where *L* is the row isometry with n^k entries L_w for |w| = k, and *B*, the column operator with the n^k entries A_w , has norm ||A||. This norm control is crucial in later developments.

5.2. Representations. In the category of unital operator algebras, we take the viewpoint that the natural representations are completely contractive and unital. Given an operator algebra \mathfrak{A} , for each $1 \leq k < \infty$, we let $\operatorname{Rep}_k(\mathfrak{A})$ denote the set of completely contractive representations of \mathfrak{A} into $\mathcal{B}(\mathcal{H}_k)$, where \mathcal{H}_k is a fixed Hilbert space of dimension k. Put the topology of pointwise convergence on this space. Even though \mathfrak{L}_n has a natural weak-* topology, there are good reasons to consider norm continuous representations which are not weak-* continuous at this juncture. This space is a reasonable topological object because k is finite, and we can obtain a lot of information from these finite dimensional representations.

There are many irreducible representations of \mathfrak{L}_n (even WOT-continuous ones) on spaces of every dimension. Indeed, take any *n*-tuple $A = (A_1, \ldots, A_n)$ of $k \times k$ matrices such that ||A|| = r < 1 and so that the A_i 's generate the full matrix algebra \mathfrak{M}_k . Then the functional calculus of Bunce and Popescu which comes out of the dilation theory yields a WOT-continuous completely contractive representation sending L_i to A_i .

Conversely, suppose that Φ is a completely contractive representation of \mathcal{L}_n on $\mathcal{B}(\mathcal{H})$ with dim $\mathcal{H} = k < \infty$. Then

$$\Phi(L) = \begin{bmatrix} \Phi(L_1) & \dots & \Phi(L_n) \end{bmatrix} =: \begin{bmatrix} A_1 & \dots & A_n \end{bmatrix}$$

must be a contraction. This determines a natural continuous projection $\pi_{n,k}$ of $\operatorname{Rep}_k(\mathfrak{L}_n)$ into the set $\overline{\mathbb{B}}_{n,k}$ of all contractive *n*-tuples of $k \times k$ matrices.

Both $\operatorname{Rep}_k(\mathfrak{L}_n)$ and $\overline{\mathbb{B}}_{n,k}$ are compact Hausdorff spaces and $\pi_{n,k}$ is continuous. Moreover we observed above that the open ball $\mathbb{B}_{n,k}$ is in the range of $\pi_{n,k}$. Hence this map is surjective. The Bunce argument may be modified to show uniqueness of the functional calculus when ||A|| = r < 1, and it follows that the restriction of $\pi_{n,k}^{-1}$ to $\mathbb{B}_{n,k}$ is a homeomorphism. However, over the boundary points, the projection is generally not injective, and the fibre $\pi_{n,k}^{-1}(A)$ may be very large.

When n = 1, $\pi_{1,1}$ is the natural projection from the maximal ideal space of H^{∞} onto the closed $\overline{\mathbb{D}}$. Over each point λ in the open disk \mathbb{D} , there is the unique functional of point evaluation at λ . However over each point on the boundary circle, there is a huge fibre.

When n > 1 and k = 1, the points $\lambda \in \mathbb{B}_n$ correspond to the point evaluations φ_{λ} ; so the set of weak-* continuous characters on \mathfrak{L}_n is homeomorphic to \mathbb{B}_n . Thus each element $A \in \mathfrak{L}_n$ determines a function $\hat{A}(\lambda) = \varphi_{\lambda}(A)$ on \mathbb{B}_n via the Gelfand transform. Moreover it is clear that $\hat{L}_i = z_i$ is the *i*th coordinate function. Moreover wor-convergent sequences in \mathfrak{L}_n are sent to sequences of functions which converge uniformly on compact subsets of the ball. Hence it follows that \hat{A} is analytic, and that the map from A to \hat{A} is a contractive map of \mathfrak{L}_n into $H^{\infty}(\mathbb{B}_n)$. This map is not surjective, and the exact description of the range will be dealt with in section 8.

In the k = 1 case, all the boundary points are essentially the same, so the fibre $\pi_{n,1}^{-1}(1, 0, \ldots, 0)$ is typical. There is a natural WOT-continuous map of \mathfrak{L}_n onto $H^{\infty}(\mathbb{D})$ with kernel generated by $\{L_i : i \geq 2\}$. Composing this map with any multiplicative linear function of H^{∞} sending z to 1 yields an element of the fibre $\pi_{n,1}^{-1}(1, 0, \ldots, 0)$. In particular, this fibre is huge! It is an open question whether there are any other multiplicative linear functionals in the fibre which do not factor in this manner.

When $n \geq 2$ and $k \geq 2$, a boundary point of $\overline{\mathbb{B}}_{n,k}$ may have a unique pre-image or it may be huge, depending on the point.

5.3. Automorphisms. It turns out that the automorphism group of \mathfrak{L}_n exhibits significant analytic structure.

A natural first question is about continuity. Every automorphism Θ of \mathfrak{L}_n is automatically norm and WOT-continuous. Norm continuity is easy, and follows from a *gliding hump* argument. That is, under the assumption that the map is unbounded, one constructs a single element A which encodes an appropriate sequence on which the norm blows up, allowing one to show that $\|\Theta(A)\|$ is infinite, reaching a contradiction.

The WOT-continuity is more subtle, and relies in part on the equivalence of the WOT and weak-* topologies. Indeed, for a dual topology, one can apply the Krein–Smulian Theorem which says that a subspace is weak-* closed if its intersection with the unit ball is weak-* closed. This is used to show that $\Theta(\mathfrak{L}_n^0)$ is weak-* closed. This leads to the conclusion that $\varphi = \varphi_0 \Theta^{-1}$ is a weak-* continuous character, and thus equals some φ_{λ} . Similarly, for each $\lambda \in \mathbb{B}_n$, there is a $\mu \in \mathbb{B}_n$ so that $\varphi_{\lambda} \Theta^{-1} = \varphi_{\mu}$. This determines a map from the ball into itself given by $\tau_{\Theta}(\lambda) = \mu$. A computation shows that

$$\tau_{\Theta}(\lambda) = \varphi_{\lambda} \Theta^{-1}(L) := \left(\varphi_{\lambda} \Theta^{-1}(L_1), \dots, \varphi_{\lambda} \Theta^{-1}(L_n)\right) = \hat{T}(\lambda)$$

where $T = \Theta^{-1}(L) = \begin{bmatrix} \Theta^{-1}(L_1) & \dots & \Theta^{-1}(L_n) \end{bmatrix}$ is a $1 \times n$ row operator. Thus this map is analytic. Another computation shows that $\tau_{\Theta_1}\tau_{\Theta_2} = \tau_{\Theta_1\Theta_2}$. In particular, $\tau_{\Theta^{-1}} = \tau_{\Theta}^{-1}$. So each τ_{Θ} is a biholomorphic homeomorphism of \mathbb{B}_n . Thus we have constructed a homomorphism τ from $\operatorname{Aut}(\mathfrak{L}_n)$ into $\operatorname{Aut}(\mathbb{B}_n)$, the group of conformal automorphisms of the ball.

The object now is to compute the kernel of this map, and to show that τ is surjective. The kernel is straightforward because of the weak-* continuity of Θ . Clearly $\tau_{\Theta} =$ id precisely when $\Theta(L_i) - L_i$ belongs to $\bigcap \ker(\varphi_{\lambda})$, which we saw earlier is the commutator ideal \mathfrak{C} . However this easily implies that $\Theta(A) - A \in \mathfrak{C}$ for every polynomial in the generators. The weak-* continuity allows us to extend this to all of \mathfrak{L}_n . Hence Θ is trivial modulo the commutator ideal. Such automorphisms are called *quasi-inner*, and the subgroup of quasi-inner automorphisms is denoted q-Inn(\mathfrak{L}_n).

Now consider the question of surjectivity. The conformal maps of \mathbb{B}_n which fix the origin are just the unitary maps on \mathbb{C}^n . A class of automorphisms known as gauge automorphisms of the Cuntz-Toeplitz algebra are well-known from quantum mechanics. For any unitary U in \mathfrak{U}_n , construct a unitary on Fock space by $\widetilde{U} = 1 \oplus \sum_{k\geq 1} \oplus U^{\otimes k}$ where 1 acts on \mathbb{C} , U acts on the *n*-dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^n$, and $U^{\otimes k} = U \otimes \cdots \otimes U$ acts on $\mathcal{H}^{\otimes k} = \mathcal{H} \otimes \cdots \otimes \mathcal{H}$. It is an easy calculation to show that the left creation operator L_{ζ} which tensors on the left by a vector $\zeta \in \mathcal{H}$ is conjugated by \widetilde{U} as $\widetilde{U}L_{\zeta}\widetilde{U}^* = L_{U\zeta}$. Thus Ad \widetilde{U} satisfies Ad $\widetilde{U}(L_i) = L_{\zeta_i}$ where ζ_1, \ldots, ζ_n is an orthonormal basis for \mathcal{H} . Thus it follows that span $\{L_1, \ldots, L_n\}$ is fixed by Ad \widetilde{U} . Hence Ad \widetilde{U} implements an automorphism Θ_U of \mathfrak{L}_n .

An easy calculation shows that $\Theta_U \Theta_V = \Theta_{UV}$; so this is a homomorphism of the unitary group \mathcal{U}_n into the automorphism group $\operatorname{Aut}(\mathfrak{L}_n)$. However a bit of a complication appears since $\tau_{\Theta_U} = \overline{U}$, the coordinate-wise conjugate of U, rather than U itself.

Now it is easy to see that any transitive subgroup of $\operatorname{Aut}(\mathbb{B}_n)$ containing \mathfrak{U}_n is the whole group [53]. So it suffices to construct automorphisms with $\tau_{\Theta}(0) = \lambda$ for every $\lambda \in \mathbb{B}_n$. One way to accomplish this by a unitarily implemented automorphism is to use the fact that $\{\nu_{\lambda}\}^{\perp}$ has an *n*-dimensional wandering space \mathcal{W}_{λ} . An orthonormal basis ζ_1, \ldots, ζ_n for this subspace yields a set of isometries L_{ζ_i} which generate \mathfrak{L}_n . The automorphism sending L_i to L_{ζ_i} does the job, although the argument is a bit long. We describe another approach used by Voiculescu [57] to construct unitarily implemented automorphisms of the Cuntz-Toeplitz algebra.

Consider the Lie group U(1, n) consisting of those $(n+1) \times (n+1)$ matrices X such that $X^*JX = J$, where $J = \begin{bmatrix} -1 & 0 \\ 0 & I_n \end{bmatrix}$. These matrices have the form $X = \begin{bmatrix} x_0 & \eta_1^* \\ \eta_2 & X_1 \end{bmatrix}$ where the coefficients satisfy the relations:

- (i)
- (ii)
- $\begin{aligned} \|\eta_1\|^2 &= \|\eta_2\|^2 = |x_0|^2 1\\ X_1\eta_1 &= \overline{x_0}\eta_2 \quad \text{and} \quad X_1^*\eta_2 = x_0\eta_1\\ X_1^*X_1 &= I_n + \eta_1\eta_1^* \quad \text{and} \quad X_1X_1^* = I_n + \eta_2\eta_2^*. \end{aligned}$ (iii)

The group $\operatorname{Aut}(\mathbb{B}_n)$ is a quotient of U(1,n), which is obtained by representing the conformal maps as fractional linear transformations. Indeed the map $X \to \theta_X$ given by

$$\theta_X(\lambda) = \frac{X_1 \lambda + \eta_2}{x_0 + \langle \lambda, \eta_1 \rangle} \quad \text{for} \quad \lambda \in \mathbb{B}^n.$$

is the desired homomorphism, and the kernel is the subgroup \mathbb{T} of scalars of modulus one.

Voiculescu observed that there is a related automorphism of \mathcal{E}_n determined by its action on generators:

$$\Theta_X(L_{\zeta}) = (x_0 I - L_{\eta_2})^{-1} (L_{X_1 \zeta} - \langle \zeta, \eta_1 \rangle I).$$

Moreover he constructs a unitary operator U_X by

$$U_X(A\xi_{\varnothing}) = \Theta_X(A)(x_0I - L_{\eta_2})^{-1}\xi_{\varnothing} \quad \text{for all} \quad A \in \mathcal{A}_n$$

so that $\Theta_X(A) = U_X A U_X^*$ for all A in \mathcal{A}_n . Unitarily implemented automorphisms are WOT-continuous, and thus this maps \mathfrak{L}_n into itself. The beauty of this approach is that once one finds this lovely formula, it is routine to verify that the map is a homomorphism, and hence each Θ_X is invertible and so is an automorphism.

Again the computation of τ has a twist. It turns out that $\tau(\Theta_X) = \theta_{\overline{X}}$ where \overline{X} is the pointwise conjugate of X. This is evidently enough to show that τ is surjective. As a bonus, we have constructed a section from Aut(\mathbb{B}_n) into the subgroup $\operatorname{Aut}_u(\mathfrak{L}_n)$ of all unitarily implemented automorphisms of \mathfrak{L}_n . In fact, this is the full subgroup. To see this, note that if Θ is just isometric and $\tau_{\Theta} = id$, then $1 = \Theta(L_i) = L_i + C_i$ where $C_i \in \mathfrak{C}$. However

$$1 \ge \|(L_i + C_i)\xi_{\varnothing}\|^2 = \|\xi_i + C_i\xi_{\varnothing}\|^2 = 1 + \|C_i\xi_{\varnothing}\|^2.$$

It follows that $C_i \xi_{\emptyset} = 0$ and so $C_i = 0$. Hence $\Theta = \text{Id}$.

The complete result is

Theorem 5.2. There is a natural short exact sequence

$$0 \longrightarrow \operatorname{q-Inn}(\mathfrak{L}_n) \longrightarrow \operatorname{Aut}(\mathfrak{L}_n) \xrightarrow{\tau} \operatorname{Aut}(\mathbb{B}_n) \longrightarrow 0.$$

The map τ takes Θ to $\tau_{\Theta}(\lambda) = \varphi_{\lambda} \Theta^{-1}(L)$ for $\lambda \in \mathbb{B}_n$. Moreover, τ has a continuous section which carries $\operatorname{Aut}(\mathbb{B}_n)$ onto the subgroup $\operatorname{Aut}_u(\mathfrak{L}_n)$ of unitarily implemented automorphisms. So $\operatorname{Aut}(\mathfrak{L}_n)$ is a semidirect product.

FREE SEMIGROUP ALGEBRAS

We now turn to the analysis of more general free semigroup algebras. In this section, we will be concerned with the classification of certain special classes of algebras. The atomic representations are classified in [27] and the finitely correlated representations are classified in [25]. In the next section, we shall see how much of this apparently special structure actually extends to all free semigroup algebras.

6.1. Atomic Representations. Say that an *n*-tuple of isometries $S = (S_1, \ldots, S_n)$ with orthogonal ranges is *atomic* if there is an orthonormal basis $\{\xi_k\}$ for \mathcal{H} which is permuted up to scalars by each S_i . This requires that there be endomorphisms $\pi_i : \mathbb{N} \to \mathbb{N}$ for $1 \leq i \leq n$ and scalars $\lambda_{i,k} \in \mathbb{T}$ such that $S_i \xi_k = \lambda_{i,k} \xi_{\pi_i(k)}$. Equivalently, this says that there is an atomic masa containing all the range projections $P_w = w(S)w(S)^*$ for $w \in \mathbb{F}_n^+$. We call the corresponding representation of \mathcal{E}_n and the free semigroup algebra \mathfrak{S} atomic as well.

There is a connection between these representations and the permutation representations of \mathcal{O}_n recently introduced and studied by Bratteli and Jorgensen in [15]. Permutation representations are the subclass of atomic representations for which all scalars $\lambda_{i,k} = 1$ and $\sum_{i=1}^{n} S_i S_i^* = I$. Bratteli and Jorgensen were interested in decomposing permutation representations into direct sums of irreducible representations. However, the condition that $\lambda_{i,k} = 1$ forced them to make certain restrictive assumptions. In general, to obtain a decomposition into irreducible representations, arbitrary scalars are needed as we shall see.

The basic idea of this analysis is easy. Start with any basis vector ξ , and see where it is sent by the action of the semigroup \mathbb{F}_n^+ . Either it is a wandering vector, or there is some word which maps ξ back to a scalar multiple of itself. When it returns, one shows that one obtains a reducing subspace on which the representation looks like Example 2.1. When ξ is wandering, one checks to see if it is in the range of one of the S_i 's. One begins pulling back until one reaches a basis vector orthogonal to the range of all of the S_i 's, one enters a recurring loop as above, or the chain continues indefinitely. In the first case, this produces a wandering vector which generates a reducing subspace on which the algebra is unitarily equivalent to \mathfrak{L}_n . And the latter case produces a summand of inductive type unitarily equivalent to Example 2.2.

This shows that every atomic representation decomposes into a direct sum of the basic building blocks which we have already identified. The discussion now turns to deciding when these representations are irreducible, what the free semigroup looks like, and how this decomposition may be accomplished.

Consider Example 2.1 first. An easy observation is that when $u = v^p$ for $p \ge 2$ is a power of a smaller word v, the ring structure has p-fold symmetry. This allows the representation $\sigma_{u,\lambda}$ to be split into p summands σ_{v,μ_j} where μ_j are the p pth roots of λ . So even if $\lambda = 1$, other roots of unity are needed in this decomposition.

On the other hand, one needs to establish that $\sigma_{u,\lambda}$ is irreducible when u is *primitive*, meaning that it is not a power of a smaller word. It is here that the free semigroup algebra $\mathfrak{S}_{u,\lambda}$ that it generates is useful. Consider the polynomial $(\overline{\lambda}u(S))^k$. This is an isometry in $\mathfrak{S}_{u,\lambda}$ which maps the vector ζ_k onto itself. Consider what it does to the other vectors in this basis. If it is another basis vector in the ring, it cannot be mapped back to itself, for that would imply symmetry. So eventually it is mapped out of the ring onto a wandering vector, and then it is mapped off into the far reaches of one of the wandering spaces. The same happens to each basis vector which is not in the ring. Hence $(\overline{\lambda}u(S))^k$ converges in the WOT topology to the rank one projection $P = \zeta_k \zeta_k^*$ onto $\mathbb{C}\zeta_k$. So P belongs to $\mathfrak{S}_{u,\lambda}$.

Suppose that this representation splits into a sum $\mathcal{M} \oplus \mathcal{M}^{\perp}$ of reducing subspaces. Either \mathcal{M} or \mathcal{M}^{\perp} is not orthogonal to ζ_k , say \mathcal{M} . Since \mathcal{M} is invariant for P, it must contain ζ_k , which is a cyclic vector. Hence $\mathcal{M} = \mathcal{H}$, and therefore this representation is irreducible.

Consider what the free semigroup algebra looks like in this case. Since it contains P, it also contains $v_1(S)Pv_2(S)$. If v_1 and v_2 are chosen properly, we can build the set of matrix units $\zeta_i \zeta_j^*$ for $1 \leq i, j \leq k$. Let $Q = \sum_{i=1}^k \zeta_i \zeta_i^*$. Then $Q\mathfrak{S}_{u,\lambda}Q$ is isomorphic to \mathfrak{M}_k , the algebra of $k \times k$ matrices. Because each ζ_i is cyclic, every non-zero vector in the range of Q is cyclic. So $\mathfrak{S}_{u,\lambda}Q = \mathcal{B}(\mathcal{H})Q$. Finally, $Q^{\perp}\mathcal{H}$ is the direct sum of k(n-1) copies of Fock space on which $\mathfrak{S}_{u,\lambda}$ looks like k(n-1) copies of \mathfrak{L}_n . Hence $\mathfrak{S}_{u,\lambda}$ has the form

$$\mathfrak{B}_{n,k} = \begin{bmatrix} \mathfrak{M}_k & 0\\ Q^{\perp} \mathcal{B}(\mathcal{H}) Q & \mathfrak{L}_n^{(k(n-1))} \end{bmatrix}$$

with respect to the decomposition of $\mathcal{H} = Q\mathcal{H} \oplus Q^{\perp}\mathcal{H}$. Notice that the algebra (rather than the specific representation) depends only on n and |u| = k.

Now consider how to identify a summand of type (u, λ) in an arbitrary atomic representation. Again study the action of the polynomial $(\overline{\lambda}u(S))^k$ on a basis vector. If it lives in the *k*th position of a ring corresponding to the word primitive *u*, then it is mapped back to a multiple of itself. In a representation $\sigma_{v,\mu}$ where *v* is not a cyclic permutation of *u*, every standard basis vector is mapped off to infinity by these polynomials. Likewise in any pure or inductive type representation, it is equally evident that each basis vector is sent off to infinity. However in a representation $\sigma_{u,\mu}$ for $\mu \neq \lambda$, ζ_k is sent to $(\overline{\lambda}\mu)^k \zeta_k$. On any other basis vector ζ , the sequence $(\overline{\lambda}u(S))^k \zeta$ heads off to infinity as before. One deals with this added complication by setting $p_m(x) = \frac{1}{m!} \sum_{j=1}^{m!} x^{m!+j}$, and considering $p_m(\overline{\lambda}u(S))$. This sequence converges in the strong operator topology to the projection $P_{u,\lambda}$ onto the span of the vectors corresponding to ζ_k in each summand of type $\sigma_{u,\lambda}$.

Thus the multiplicity of $\sigma_{u,\lambda}$ is computed as the rank of $P_{u,\lambda}$. In particular we now see that $\sigma_{u,\lambda}$ and $\sigma_{v,\mu}$ are unitarily equivalent only if v is a cyclical permutation of u and $\mu = \lambda$.

Summing $P_{u,\lambda}$ over all pairs (u,λ) yields a projection Q in the free semigroup algebra \mathfrak{S} corresponding to the projection onto all of the rings. Observe that $Q\mathfrak{S}Q$ is a (type I) von Neumann algebra. Every standard basis vector in $Q^{\perp}\mathcal{H}$ is a wandering vector. Indeed, $Q^{\perp}\mathfrak{S}Q^{\perp}$ is the direct sum of copies of \mathfrak{L}_n and of inductive algebras, which were observed in Example 2.2 to be (completely) isometrically isomorphic to \mathfrak{L}_n . Thus it follows that $Q^{\perp}\mathfrak{S}Q^{\perp}$ is completely isometrically isomorphic to \mathfrak{L}_n . As $Q^{\perp}\mathcal{H}$ is invariant for \mathfrak{S} , the algebra has a lower triangular form. The 2,1 entry is specified by the structure of the pieces coming from each ring algebra. In fact, $\mathfrak{S}Q = \mathfrak{M}Q$ where \mathfrak{M} is the von Neumann algebra generated by \mathfrak{S} , as we see in the next section.

Lastly consider the representations of inductive type. It is clear that if one can delete a finite number of terms from the beginning of two words in order to make them equal, then the two representations are unitarily equivalent. In this case, the two sequences are called *shift tail equivalent*. Otherwise they are inequivalent. They are irreducible except when the word is equivalent to a periodic word. The representation π_x where $x = uuuuu \dots$ is equivalent to a direct integral of the representation $\sigma_{u,\lambda}$ over the unit circle with respect to Lebesgue measure.

To complete the picture, one may compute the multiplicity of the pure part as rank $(I - \sum_{i=1}^{n} S_i S_i^*)$ and of each inductive representation π_x . Recall from Example 2.2 the sequence x_m . Then note that $x_m(S)s_m(S)^*$ is a projection onto the range of $x_m(S)$ spanned by a subset of the standard basis. As *m* increases, one obtains precisely those basis vectors which can be pulled back indefinitely along this sequence. When it is not periodic, there is one vector in each summand of π_x and none in any other. Whence the multiplicity is the rank of SOT-lim $x_m(S)s_m(S)^*$. In the periodic case, one obtains instead a subspace on which u(S) is a bilateral shift, and one computes its multiplicity.

6.2. Finitely Correlated Representations. Our approach here ties dilation theory strongly to the classification of an important class of representations of \mathcal{E}_n . Bratteli and Jorgensen [14] introduced this class to study endomorphism of $\mathcal{B}(\mathcal{H})$ and then showed them to be of central importance in generating wavelets [16, 17]. Their class of representations is obtained by the GNS construction from a state ψ which has the property that the cyclic vector ξ_{ψ} generates a *finite dimensional* subspace $\mathcal{V} = \operatorname{span}\{w(S)^*\xi_{\psi} : w \in \mathbb{F}_n^+\}$ for the adjoints S_i^* of the generating isometries. Our slightly weaker definition does not insist that the representation have a cyclic vector, only that \mathcal{V} be cyclic for \mathfrak{S} . That is, a free semigroup \mathfrak{S} is *finitely correlated* if there is a finite dimensional subspace \mathcal{V} which is invariant for \mathfrak{S}^* and is cyclic for \mathfrak{S} . In particular this class contains all of the atomic ring representations mentioned above.

Our viewpoint about where such representations arise comes from dilation theory. Let $A_i = (S_i^*|_{\mathcal{V}})^*$ be the compressions of each S_i to \mathcal{V} . A simple calculation shows that

$$\sum_{i=1}^{n} A_i A_i^* = P_{\mathcal{V}} \sum_{i=1}^{n} S_i S_i^* \big|_{\mathcal{V}} \le I_{\mathcal{V}}$$

Thus $A = (A_1, \ldots, A_n)$ is a row contraction; and it is a row isometry if (and only if) S is a Cuntz representation. Notice that S is an isometric dilation of the row contraction A. Moreover since \mathcal{V} is cyclic, this is a minimal dilation. By the uniqueness of the minimal isometric dilation, we see that Sis completely determined by A. Our goal is to start with A and find complete unitary invariants for the set of isometries S. In particular, we will be able to decide if A and B determine unitarily equivalent representations.

For convenience we will consider only the Cuntz case $AA^* = I$. For the general case, we mention only that the multiplicity of the pure part may be computed directly from A as rank $(I - AA^*)$. The additional complication in decomposing the representation can be found in [25].

It is easy to see that \mathcal{V}^{\perp} is of pure type, so that the restriction of \mathfrak{S} to this subspace is a multiple $\mathfrak{L}_n^{(\alpha)}$ of \mathfrak{L}_n . However \mathcal{V}^{\perp} is not normally a maximal subspace of this type. It does place \mathfrak{S} inside the algebra $\mathcal{B}(\mathcal{H})P_{\mathcal{V}} + P_{\mathcal{V}}^{\perp}\mathfrak{S}P_{\mathcal{V}}^{\perp}$, which is unitarily equivalent to $\mathfrak{B}_{n,d}$ where $d = \dim \mathcal{V}$. In particular, we see immediately that the weak-* and WOT topologies coincide.

We search for a (finite rank) projection in \mathfrak{S} . We focus on the algebra $\mathfrak{A} = \operatorname{Alg}\{A_1, \ldots, A_n\}$ of matrices generated by A. The key technical result is that for every non-zero vector $x \in \mathcal{H}$, the subspace $\mathfrak{S}^*[x]$ always intersects \mathcal{V} . In the special case in which $\mathfrak{A} = \mathcal{B}(\mathcal{V})$, one can now show that $P_{\mathcal{V}}$ belongs to \mathfrak{S} . From this we deduce, as in the atomic ring case, that the representation is irreducible. The free semigroup algebra in this case is exactly the algebra $\mathfrak{B}_{n,d}$ above.

More generally, one finds that whenever \mathcal{M} is a minimal invariant subspace for \mathfrak{A}^* , that $\mathfrak{S}[\mathcal{M}]$ is a reducing subspace. The restriction of \mathfrak{S} to this subspace is irreducible, and it contains the projection $P_{\mathcal{M}}$. The restriction of \mathfrak{A}^* to \mathcal{M} is an algebra of matrices with no proper invariant subspace. So by Burnside's Theorem, $\mathfrak{A}^*|_{\mathcal{M}} = \mathcal{B}(\mathcal{M})$. We deduce that $\mathfrak{S}_{\mathfrak{S}[\mathcal{M}]} \simeq \mathfrak{B}_{n,d}$ where $d = \dim \mathcal{M}$.

We then form a maximal family $\{\mathcal{M}_i\}$ of pairwise orthogonal minimal \mathfrak{A}^* -invariant subspaces. This yields an orthogonal decomposition of the free semigroup algebra into a direct sum of irreducible algebras on $\mathfrak{S}[\mathcal{M}_i]$. This procedure does not depend on how this maximal family is obtained. In particular the span of the \mathcal{M}_i 's is the subspace $\widetilde{\mathcal{V}}$ spanned by all minimal \mathfrak{A}^* -invariant subspaces.

At this stage, you might suspect that because of the structure $\mathfrak{A}^*|_{\mathcal{M}_i} = \mathcal{B}(\mathcal{M}_i)$ that $\mathfrak{A}^*|_{\widetilde{\mathcal{V}}}$ is a C*-algebra. Remarkably this is true, and once this is established, one can use the Wedderburn Theorem to decompose it into a sum of full matrix algebras. This yields a corresponding orthogonal decomposition of \mathfrak{S} .

Thus the decomposition problem now turns on deciding when two such irreducible representations are unitarily equivalent. This suggests studying intertwining maps between two such representations. The tool that makes this possible is the completely positive map Φ on $\mathcal{B}(\mathcal{V})$ given by

$$\Phi(X) = \sum_{i=1}^{n} A_i X A_i^*.$$

Suppose that there is an invertible map T between two minimal \mathfrak{A}^{*-} invariant subspaces \mathcal{V}_1 and \mathcal{V}_2 which commutes with \mathfrak{A} . Write $B_i := P_{\mathcal{V}_1} A_i|_{\mathcal{V}_1}$ and $C_i := P_{\mathcal{V}_2} A_i|_{\mathcal{V}_2}$. Since $TB_i T^{-1} = C_i$ and

$$\sum_{i=1}^{n} B_i B_i^* = I_{\mathcal{V}_1} \quad \text{and} \quad \sum_{i=1}^{n} C_i C_i^* = I_{\mathcal{V}_2},$$

we compute that

$$I_{\mathcal{V}_2} = \sum_{i=1}^n (TB_i T^{-1}) (TB_i T^{-1})^* = T\Phi(T^{-1} T^{*-1}) T^*.$$

Therefore

$$\Phi(T^{-1}T^{*-1}) = T^{-1}T^{*-1}.$$

This leads to the discovery of two interesting facts about completely positive maps on \mathfrak{M}_k . The first is that a non-scalar operator X such that $\Phi(X) = X$ implies that $\mathfrak{A}^* = \operatorname{Alg}\{A_1^*, \ldots, A_n^*\}$ has two pairwise orthogonal minimal invariant subspaces. This had gone unnoticed because no-one had ever started with a completely positive map Φ on \mathfrak{M}_k , found a Stinespring decomposition which leads to the form as above and then looked at the algebra generated by the A_i^* 's. It is not at all obvious that the uniqueness in the Stinespring condition would make this an interesting object; but it is!

The second consequence is even more surprising. When \mathcal{V} is the orthogonal direct sum of minimal \mathfrak{A}^* -invariant subspaces, then \mathfrak{A} is a C*-algebra and the fixed point set of Φ coincides with the commutant of \mathfrak{A} . The following example puts this result in perspective.

Example 6.1. Let

$$A_1 = \begin{bmatrix} 1/\sqrt{2} & 0 & 0\\ 1/2\sqrt{2} & 1/2 & 1/2\sqrt{2}\\ 0 & 0 & 1/\sqrt{2} \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 1/\sqrt{2} & 0 & 0\\ -1/2\sqrt{2} & 1/2 & -1/2\sqrt{2}\\ 0 & 0 & 1/\sqrt{2} \end{bmatrix}.$$

Note that $A_1A_1^* + A_2A_2^* = I$. A calculation shows that the fixed point set of Φ is the set of matrices $X = [x_{ij}]$ such that $x_{12} = x_{21} = x_{23} = x_{32} = 0$ and $x_{11} + x_{13} + x_{31} + x_{33} = 2x_{22}$. In particular, this is not an algebra. The algebra \mathfrak{A}^* has two minimal invariant subspaces, $\mathbb{C}e_1$ and $\mathbb{C}e_3$. Note that the compression of \mathfrak{A} to span $\{e_1, e_3\}$ consists of scalar matrices, and the fixed point set of the restricted completely positive map is the full 2×2 matrix algebra.

Putting all of this information together yields the complete picture:

Theorem 6.2. Let A_1, \ldots, A_n be operators on a finite dimensional space \mathcal{V} such that $\sum_{i=1}^n A_i A_i^* = I$, and let S_1, \ldots, S_n be their joint isometric dilation. Let $\widetilde{\mathcal{V}}$ be the subspace of \mathcal{V} spanned by all minimal \mathfrak{A}^* -invariant subspaces. Then the compression $\widetilde{\mathfrak{A}}$ of \mathfrak{A} to $\widetilde{\mathcal{V}}$ is a C^* -algebra. Let $\widetilde{\mathfrak{A}}$ be decomposed as $\sum_{g \in G}^{\oplus} \mathfrak{M}_{dg} \otimes \mathbb{C}^{m_g}$ with respect to a decomposition $\widetilde{\mathcal{V}} = \sum_{g \in G}^{\oplus} \mathcal{V}_g^{(m_g)}$, where \mathcal{V}_g has dimension d_g and multiplicity m_g . Then the dilation acts on the space $\mathcal{H} = \sum_{g \in G} \mathcal{H}_g^{(m_g)}$ where $\mathcal{H}_g = \mathcal{V}_g \oplus \mathcal{K}_n^{(d_g(n-1))}$. The algebra \mathfrak{S} decomposes as $\mathfrak{S} \simeq \sum_{g \in G}^{\oplus} \mathcal{B}_{n,d_g}^{(m_g)}$.

In particular, the algebra \mathfrak{S} contains the projection $P_{\widetilde{V}}$, and $P_{\widetilde{V}}\mathfrak{S}P_{\widetilde{V}}$ is the C*-algebra $\widetilde{\mathfrak{A}}$. This is what makes the decomposition possible. This description looks a bit complicated, but it yields some useful and easily implemented consequences.

Corollary 6.3. Let \mathfrak{S} be the algebra determined by the joint isometric dilation of a contractive n-tuple A on a finite dimensional space \mathcal{V} such that $AA^* = I$. Then \mathfrak{S} is irreducible if and only if the fixed point set of $\Phi(X) = \sum_{i=1}^{n} A_i X A_i^*$ is trivial, i.e. $\{X : \Phi(X) = X\} = \mathbb{C}I$.

And the general question of unitarily equivalent representations of this type reduces to a (possibly difficult) finite dimensional problem.

Theorem 6.4. Let $A = (A_1, \ldots, A_n)$ and $B = (B_1, \ldots, B_n)$ be contractive n-tuples on finite dimensional spaces \mathcal{V}_A and \mathcal{V}_B respectively such that $AA^* = I_{\mathcal{V}_A}$ and $BB^* = I_{\mathcal{V}_B}$. Let $S = (S_1, \ldots, S_n)$ and $T = (T_1, \ldots, T_n)$ be their joint minimal isometric dilations on Hilbert spaces \mathcal{H}_A and \mathcal{H}_B ; and let σ_A and σ_B be the induced representations of \mathcal{O}_n . Let $\widetilde{\mathcal{V}}_A$ be the subspace spanned by all minimal \mathfrak{A}^* -invariant subspaces; and similarly define $\widetilde{\mathcal{V}}_B$. Then σ_A and σ_B are unitarily equivalent if and only if $A^*|_{\widetilde{\mathcal{V}}_A}$ is unitarily equivalent to $B^*|_{\widetilde{\mathcal{V}}_p}$.

A result of Kribs [36] that shows that one can detect the \mathfrak{A}^* -invariant subspaces from the completely positive map Φ without computing the A_i 's explicitly. He shows that \mathcal{M} is \mathfrak{A}^* -invariant if and only if $\Phi(P_{\mathcal{M}}) \leq P_{\mathcal{M}}$.

The structure of these algebras is so tight that one can explicitly describe many of their invariant subspaces. In particular, the column $\mathfrak{S}P_{\mathcal{V}}$ is a 'slice' of a type I von Neumann algebra, and thus is hyper-reflexive with distance constant 4 by Christensen's result [20]. The restriction to $\widetilde{\mathcal{V}}^{\perp}$ is a multiple of \mathfrak{L}_n , and thus has hyper-reflexivity constant 3 by Bercovici [12]. Combining these estimates yields:

Corollary 6.5. The algebra \mathfrak{S} determined by the joint isometric dilation of a contractive n-tuple on a finite dimensional space is hyper-reflexive with distance constant at most 5.

7. Structure Theory

Our attention earlier has been on special classes which have some useful extra property. However it is now possible to extend some of the structural picture to all free semigroup algebras. The focus in on several features which have emerged as important in these special classes. The first feature is the central role played by wandering vectors, because the restriction of S to a subspace generated by a wandering vector is unitarily equivalent to the left regular representation L. Secondly there is the role of projections which occur in the algebra \mathfrak{S} . Since \mathfrak{L}_n has no proper projections itself, projections in \mathfrak{S} have made it possible to decompose the algebra into pieces.

We shall say that an algebra \mathfrak{S} is type L if it is canonically isomorphic to \mathfrak{L}_n , meaning that the map sending S_i to L_i may be extended to an algebra isomorphism. In addition to multiples of \mathfrak{L}_n , we have also seen that atomic representations of inductive type are type L. More generally any inductive representation (meaning that \mathcal{H} is the increasing union of invariant subspaces on which \mathfrak{S} is pure) must be type L. In fact these algebras are completely isometrically isomorphic and weak-* homeomorphic by the canonical WOT-continuous map connecting them. This turns out to be true for all type L algebras.

An important tool in this general analysis is the ideal \mathfrak{S}_0 of \mathfrak{S} generated by $\{S_1, \ldots, S_n\}$. This ideal is either co-dimension 1 if it does not contain I or it is all of \mathfrak{S} . If \mathfrak{S} has a wandering vector ξ , then $\varphi(A) = \langle A\xi, \xi \rangle$ is a unital wort-continuous functional with kernel \mathfrak{S}_0 . Whence $\mathfrak{S}_0 \neq \mathfrak{S}$. Conversely, if $\mathfrak{S}_0 \neq \mathfrak{S}$, then there is a WOT-continuous functional φ with ker $\varphi = \mathfrak{S}_0$. This functional may be represented as $\varphi(A) = \sum_{i=1}^p \langle A\zeta_i, \eta_i \rangle$. Hence on the p-fold ampliation $\mathfrak{S}^{(p)}$ acting on $\mathcal{H}^{(p)}$, one has $\varphi(A) = \langle A^{(p)}\zeta, \eta \rangle$ where ζ has coefficients ζ_i and η has coefficients η_i . Consequently ζ is orthogonal to $\overline{\mathfrak{S}_0^{(p)}\zeta}$. The projection of ζ onto the orthogonal complement will be a wandering vector for the subspace $\mathfrak{S}^{(p)}[\zeta]$.

On the other hand, if $\mathfrak{S}_0 = \mathfrak{S}$, then *I* is the WOT-limit of a net A_α in \mathfrak{S}_0 . It follows easily that S_i^*A is a net in \mathfrak{S} which converges WOT to S_i^* . Thus \mathfrak{S} contains each S_i^* , and so is the von Neumann algebra \mathfrak{W} generated by *S*.

There are two important open questions related to this circle of ideas:

Question 7.1. Can \mathfrak{S} be a von Neumann algebra?

Question 7.2. Does every (type L) representation have a wandering vector?

We have some partial information which pushes the ampliation argument significantly further, and adds some information to the second question.

Theorem 7.3. Suppose that \mathfrak{S} is type *L*. Then for *p* sufficiently large, the space $\mathcal{H}^{(p)}$ is spanned by wandering vectors of $\mathfrak{S}^{(p)}$.

It follows from our argument above that whenever $\mathfrak{S}_0 \neq \mathfrak{S}$, there is a wort-continuous map from \mathfrak{S} into \mathfrak{L}_n sending each generator S_i to L_i , just by sending A to $A^{(p)}|_{\mathfrak{S}^{(p)}[\mathcal{L}]}$. This map is always surjective. Indeed,

Theorem 7.4. Suppose that $\Phi : \mathfrak{S} \to \mathfrak{L}_n$ is a WOT-continuous homomorphism such that $\Phi(S_i) = L_i$ for $1 \leq i \leq n$. Then Φ is surjective, and $\mathfrak{S}/\ker(\Phi)$ is completely isometrically isomorphic and weak-* homeomorphic to \mathfrak{L}_n .

Clearly the map Φ is unique when it exists. We need to identify its kernel. It is not difficult to convince yourself that ker $\Phi = \bigcap_{k\geq 1} \mathfrak{S}_0^k =: \mathfrak{J}$. The important observation is that \mathfrak{J} is invariant under left multiplication by S_i^* as well as by S_i , and thus it is a WOT-closed left ideal in \mathfrak{W} . Consequently \mathfrak{J} contains a self-adjoint projection P so that $\mathfrak{J} = \mathfrak{W}P$. This is a crucial step in the main structure theorem that we are seeking.

Theorem 7.5. Let \mathfrak{S} be a free semigroup algebra, and let \mathfrak{W} be its enveloping von Neumann algebra. Then there is a largest projection P in \mathfrak{S} such that $P\mathfrak{S}P$ is self-adjoint. It has the following properties:

- (i) $\mathfrak{W}P = \bigcap_{k \ge 1} \mathfrak{S}_0^k$,
- (ii) $P^{\perp}\mathcal{H}$ is invariant for \mathfrak{S} ,
- (iii) if $P \neq I$, then $\mathfrak{S}P^{\perp}$ is completely isometrically isomorphic and weak-* homeomorphic to \mathfrak{L}_n via the canonical WOT-continuous homomorphism Φ with $\Phi(S_i) = L_i$ for $1 \leq i \leq n$, and

(iv)
$$\mathfrak{S} = \mathfrak{W}P + P^{\perp}\mathfrak{S}P^{\perp}$$
.

Note that the description of P is given only in terms of \mathfrak{S} , and hence P is an invariant of the algebra and is not dependent on a choice of generators. The theorem yields a canonical decomposition of \mathfrak{S} into a lower triangular form where the first column is a slice of a von Neumann algebra and the (2, 2) entry is type L.

Since type L algebras contain no proper projections, we may conclude that any free semigroup algebra which is merely algebraically isomorphic to a subalgebra of \mathfrak{L}_n is automatically type L, and thus is completely isometrically isomorphic and weak-* homeomorphic to \mathfrak{L}_n via the canonical map.

As an immediate corollary, we can characterize the radical.

Corollary 7.6. With $n \ge 2$ and notation as above, the radical of \mathfrak{S} is $P^{\perp}\mathfrak{S}P$. Thus the following are equivalent:

- (i) \mathfrak{S} is semisimple and is not self-adjoint
- (ii) \mathfrak{S} is type L
- (iii) S has no non-scalar idempotents
- (iv) \mathfrak{S} has no non-zero quasinilpotent elements.

One interesting and non-trivial consequence of the structure theorem is information about the geometry of the unit ball. The Russo–Dye Theorem

[54] states that in any C*-algebra, the convex hull of all unitary elements is the whole unit ball. Also the algebra H^{∞} is the convex hull of the inner functions [38]. In a free semigroup algebra which may contain no non-scalar unitaries at all (when it is type L), we instead consider the rich collection of isometries. One useful consequence of the information about wandering vectors in ampliations is that whenever V is an isometry in \mathfrak{L}_n and Φ is the canonical map of a type L algebra \mathfrak{S} onto \mathfrak{L}_n , then $\Phi^{-1}(V)$ is also an isometry. So there are many isometries in any free semigroup algebra. We obtain an analogue of the Russo-Dye Theorem in our context.

Theorem 7.7. The convex hull of $\text{Isom}(\mathfrak{S})$ contains the open unit ball of \mathfrak{S} . Moreover, if $||A|| < 1 - \frac{1}{k}$ for k > 0 an even integer, then A is the average of 6k isometries.

Examples 7.8. We conclude this section by considering a few examples.

Consider the finitely correlated representations studied in the previous section. The projection P is the projection onto $\tilde{\mathcal{V}}$ and $P\mathfrak{S}P = \mathfrak{A}$ is a C*-algebra. The type L portion is in fact pure.

Now consider the atomic representations. The projection P has range equal to the direct sum of all of the rings. The type L portion consists of the span of all wandering vectors, which includes all of the inductive representations.

Here is an example which yields a large class of inductive type L representations. Let $S = (S_1, \ldots, S_n)$ be any *n*-tuple of isometries with orthogonal ranges acting on \mathcal{H} . Let U be the unitary of multiplication by zon $L^2(\mathbb{T})$. Define a new *n*-tuple of isometries on $\mathcal{H} \otimes L^2(\mathbb{T})$ by $S \otimes U = (S_1 \otimes U, \ldots, S_n \otimes U)$. Evidently

$$\sum_{i=1}^{n} (S_i \otimes U)(S_i \otimes U)^* = \sum_{i=1}^{n} S_i S_i^* \otimes I.$$

So these isometries have orthogonal ranges. In addition, if S is of Cuntz type, then so is the tensored n-tuple.

However, this new representation has a spanning set of wandering vectors of the form $\xi \otimes z^k$ for any $\xi \in \mathcal{H}$ and $k \in \mathbb{Z}$, as a simple calculation shows. Thus this representation has type L. In fact it is inductive, since the restriction to $\mathcal{H} \otimes \overline{z^k} H^2(\mathbb{T})$ is pure with wandering space $\mathcal{H} \otimes \mathbb{C}\overline{z^k}$.

Consider the representation $\pi_{1^{\infty}}$. This is an atomic representation of inductive type, and hence is of type L. Because 1^{∞} is periodic, it is also a direct integral $\pi_{1^{\infty}} \simeq \int_{\mathbb{T}} \sigma_{1,\lambda} d\lambda$. Indeed, let $\mathcal{K} = \mathbb{C} \oplus \mathcal{K}_n$. The representation $\sigma_{1,\lambda}$ is determined by generators

$$S_1^{\lambda} = \begin{bmatrix} \lambda & 0\\ 0 & L_1 \end{bmatrix}$$
 and $S_2^{\lambda} = \begin{bmatrix} 0 & 0\\ \xi_{\varnothing} & L_2 \end{bmatrix}$.

Thus the representation $\pi_{1^{\infty}}$ may be represented on $\mathcal{H}_T := L^2(\mathbb{T}) \otimes \mathcal{K}$ by

$$S_1 = \begin{bmatrix} U & 0 \\ 0 & I \otimes L_1 \end{bmatrix} \quad \text{and} \quad S_2 = \begin{bmatrix} 0 & 0 \\ I \otimes \xi_{\varnothing} & I \otimes L_2 \end{bmatrix}$$

where U is multiplication by z on $L^2(\mathbb{T})$.

Let *E* be a measurable subset of \mathbb{T} with positive measure. Let *V* denote $U|_{L^2(E)}$ and $J = I_{L^2(E)}$. Now consider the representation ρ_E on $\mathcal{H}_E = L^2(E) \otimes \mathcal{K}$ by

$$T_1 = \begin{bmatrix} V & 0\\ 0 & J \otimes L_i \end{bmatrix} \quad \text{and} \quad T_2 = \begin{bmatrix} 0 & 0\\ J \otimes \xi_{\varnothing} & J \otimes L_2 \end{bmatrix}.$$

It is evident that any vector of the form $0 \oplus (f \otimes \xi_{\emptyset})$ is a wandering vector. It can be shown that the restriction of a type L representation to an invariant subspace containing a wandering vector remains type L. In fact in this example, \mathcal{H}_E is spanned by wandering vectors.

There are a number of open questions raised by this example about subinductive representations. Is \mathcal{H}_E of inductive type? Is $\mathcal{H}_E^{(\infty)}$ of inductive type? Is $\mathcal{H}_T \oplus \mathcal{H}_E$ of inductive type? Also more generally if the restriction of any type L representation to an invariant subspace still type L?

Consider the case n = 1. We are given an isometry S, which decomposes using the Wold decomposition and the spectral theory of unitary operators as $S \simeq U_{+}^{(\alpha)} \oplus U_a \oplus U_s$ where U_{+} is the unilateral shift, U_a is a unitary with spectral measure absolutely continuous with respect to Lebesgue measure m, and U_s is a singular unitary. Let m_a and m_s denote scalar measures equivalent to the spectral measures of U_a and U_s respectively. If $\alpha > 0$ or if $m_a = m$, then by [58]

$$\mathfrak{S} = W(S) \simeq H^{\infty}(U_+^{(\alpha)} \oplus U_a) \oplus L^{\infty}(m_s)(U_s).$$

The von Neumann algebra it generates is

$$\mathfrak{W} = W^*(S) = \mathcal{B}(\mathcal{H})^{(\alpha)} \oplus L^{\infty}(U_a) \oplus L^{\infty}(m_s)(U_s).$$

The projection P of the Structure Theorem is just the projection onto the singular part. In this case, it is always a direct summand.

If $\alpha = 0$ and the essential support of U_a is a proper measurable subset of the circle, then $\mathfrak{S} = \mathfrak{W}$ is self-adjoint. Note that this may occur even though the spectrum of S is the whole circle.

The n = 1 case exhibits two phenomena which we cannot seem to replicate in the non-commutative case, and remain important open questions.

The first is the situation just noted that \mathfrak{S} can be self-adjoint. The second is that there are isometries S_1 and S_2 such that $W(S_1 \oplus S_2) \simeq H^{\infty}$, the type L case, yet neither $W(S_i)$ are type L, and in fact are von Neumann algebras. One simply takes S_i to be multiplication by z on the upper and lower half circles respectively.

FREE SEMIGROUP ALGEBRAS

8. Interpolation and the Commutative Theory

In this section, we will examine some natural interpolation questions [29, 5] about the algebra \mathfrak{L}_n including describing the image of the Gelfand map from \mathfrak{L}_n into $H^{\infty}(\mathbb{B}_n)$. The complete picture requires making a connection to the theory of commuting row contractions. We will connect the dilation theory for this class to that of the non-commuting class, and will use it to describe the image algebra precisely. Finally we will show that we obtain a reproducing kernel Hilbert space, and obtain yet another view of this interpolation.

8.1. Interpolation. Consider the Gelfand map which takes $A \in \mathfrak{L}_n$ to the function $\hat{A}(\lambda) = \varphi_{\lambda}(A) = \langle A\nu_{\lambda}, \nu_{\lambda} \rangle$ on the unit ball \mathbb{B}_n . Fix finite subsets $\{\alpha_1, \ldots, \alpha_k\}$ and $\{c_1, \ldots, c_n\}$ of \mathbb{B}_n . Consider the Nevanlinna–Pick question of whether there is an element A in the unit ball of \mathfrak{L}_n such that $\hat{A}(\alpha_j) = c_j$ for $1 \leq j \leq k$. If the norm condition is dropped, it is elementary to verify that there are solutions in \mathfrak{L}_n for any choice of constants. Following Sarason [55], we can formulate interpolation questions for \mathfrak{L}_n as operator theoretic questions. The set of all solutions is a coset of $\mathfrak{J} = \bigcap_{1 \leq j \leq k} \varphi_{\alpha_j}$, which is a WOT-closed ideal of \mathfrak{L}_n . Thus the smallest possible norm of a solution is dist (A, \mathfrak{J}) where A is any particular solution. Moreover compactness in the WOT-topology shows that this norm is attained. The issue then is to decide if this quotient has norm at most 1.

We consider the general problem of computing the distance to a WOTclosed right ideal \mathfrak{J} . Recall that the range $\mathcal{M} = \overline{\mathfrak{J}\xi_{\varnothing}} = \overline{\mathfrak{J}\mathcal{K}_n}$ is invariant for \mathfrak{R}_n . Clearly for any $J \in \mathfrak{J}$,

$$||A + J|| \ge ||P_{\mathcal{M}}^{\perp}(A + J)|| = ||P_{\mathcal{M}}^{\perp}A||.$$

The converse is also completely correct [29].

Theorem 8.1. Let \mathfrak{J} be a WOT-closed right ideal in \mathfrak{L}_n ; and let \mathcal{M} denote its range. Then $\mathfrak{L}_n/\mathfrak{J}$ is completely isometric to $P_{\mathcal{M}}^{\perp}\mathfrak{L}_n$. That is, for every matrix $A = [A_{ij}]$ in $\mathfrak{M}_p(\mathfrak{L}_n), p \geq 1$,

$$\operatorname{dist}(A,\mathfrak{M}_p(\mathfrak{J})) = \|(P_{\mathcal{M}}^{\perp} \otimes I_p)A\|.$$

The basic idea is to use the Hahn–Banach Theorem and the fact that weak-* continuous functionals on $\mathfrak{L}_n \otimes \mathcal{B}(\mathcal{H})$ are given by rank-one vector functionals. These vectors corresponding to functionals which annihilate \mathfrak{J} are generally not supported on \mathcal{M}^{\perp} . However, using the Beurling Theorem, one is able to construct a new representation of this functional which is no longer rank one, but is supported on \mathcal{M}^{\perp} .

The desired conclusion for two-sided ideals is immediate because in this case \mathcal{M} is also invariant for \mathfrak{L}_n . So $P_{\mathcal{M}}^{\perp}A = P_{\mathcal{M}}^{\perp}AP_{\mathcal{M}}^{\perp}$, and the compression to \mathcal{M}^{\perp} is a (completely contractive, unital, wor-continuous) homomorphism.

Corollary 8.2. Let \mathfrak{J} be a WOT-closed (two-sided) ideal in \mathfrak{L}_n ; and let $\mathcal{M} = \mu(\mathfrak{J}) = \overline{\mathfrak{J}\xi_{\varnothing}}$ denote its range. Then $\mathfrak{L}_n/\mathfrak{J}$ is completely isometrically isomorphic and weak-* homeomorphic to the compression $P_{\mathcal{M}}^{\perp}\mathfrak{L}_n|_{P_{\mathcal{M}}^{\perp}}$ of \mathfrak{L}_n to \mathcal{M}^{\perp} .

Now apply this to the Nevanlinna–Pick problem. The range of the ideal $\mathfrak{J} = \bigcap_{1 \leq j \leq k} \varphi_{\alpha_j}$ has complement $\mathcal{M}^{\perp} = \operatorname{span}\{\nu_{\alpha_j} : 1 \leq j \leq k\}$. Thus there is a contractive solution to $\hat{A}(\alpha_j) = c_j$ if and only of $\|P_{\mathcal{M}}^{\perp}AP_{\mathcal{M}}^{\perp}\| \leq 1$ for any solution A. This is equivalent to the condition $P_{\mathcal{M}}^{\perp}(I - AA^*)P_{\mathcal{M}}^{\perp} \geq 0$. Since we have a basis for \mathcal{M}^{\perp} , it is a standard result that this is equivalent to the positivity of the $k \times k$ matrix with entries $\langle (I - AA^*)\nu_{\alpha_i}, \nu_{\alpha_j} \rangle$. A routine calculation factoring out a few things yields the classic Pick matrix condition

$$\left[\frac{1-c_i\overline{c}_j}{1-\langle\alpha_i,\alpha_j\rangle}\right] \ge 0.$$

Taking a limit, one obtains a characterization of the image in $H^{\infty}(\mathbb{B}_n)$: an analytic function h on \mathbb{B}_n is in the image of the unit ball of \mathfrak{L}_n if and only if

$$\left[\frac{1-h(\alpha_i)\overline{h(\alpha_j)}}{1-\langle \alpha_i, \alpha_j \rangle}\right] \ge 0$$

for all finite subsets of \mathbb{B}_n . In this case, the ideal is $\bigcap_{\alpha \in \mathbb{B}_n} \varphi_\alpha$, which is the commutator ideal \mathfrak{C} . The quotient $\mathfrak{L}_n/\mathfrak{C}$ is abelian, but it is not all of $H^{\infty}(\mathbb{B}_n)$. Indeed one may take $A_k = \sum_{\{w \in \mathbb{F}_2: w(z_1, z_2) = z_1^k z_2^k\}} L_w$ and explicitly compute $\|\hat{A}_k\|_{\infty}/\|A + \mathfrak{C}\|$. This tends to 0, and thus the injective map from $\mathfrak{L}_n/\mathfrak{C}$ into $H^{\infty}(\mathbb{B}_n)$ is not surjective.

One obtains Carathéodory type conditions as well. For example, suppose that coefficients a_w are specified for $|w| \leq k$. Is there an element A in the unit ball of \mathfrak{L}_n whose initial Fourier coefficients are a_w ? This amounts to computing dist $(\sum_{|w|\leq k} a_w L_w, \mathfrak{L}_0^{k+1})$. The distance estimate reduces this to a single norm condition $||P_k \sum_{|w|\leq k} a_w L_w|| \leq 1$ where P_k projects onto span $\{\xi_w : |w| \leq k\}$.

Sarason's original approach was somewhat different, in that he established his estimates using (a prototype of) the commutant lifting theorem. Arias and Popescu [5] independently established these interpolation results, and give a proof based on Popescu's commutant lifting theorem [44].

8.2. Commuting Dilations. To better understand the abelian algebra $\mathfrak{L}_n/\mathfrak{C}$, we turn to commutative dilation theory. The appropriate models live on symmetric Fock space \mathcal{K}_n^s . Consider all *n*-tuples $k = (k_1, \ldots, k_n) \in \mathbb{N}_0^n$, and write z^k for $z_1^{k_1} \ldots z_n^{k_n}$ where z_i are commuting variables. Let $\mathcal{P}_k := \{w \in \mathbb{F}_n^+ : w(z) = z^k\}$. Consider an orthogonal but unnormalized basis for

 \mathcal{K}_n^s given by

$$\zeta^k = \frac{1}{|\mathcal{P}_k|} \sum_{w \in \mathcal{P}_k} \xi_w.$$

Indeed $|\mathcal{P}_k| = {|k|! \choose k_1! \, k_2! \cdots k_n!}$ and $||\zeta^k|| = |\mathcal{P}_k|^{-1/2}$.

Consider the operators $M_i \zeta^k = \zeta^{k+\delta_i}$, where δ_i has a 1 in the *i*th entry and 0's elsewhere. If one normalizes the basis ζ^k , one sees that the M_i 's are certain weighted shifts.

We observe that M_i is just the compression of the left creation operator L_i to symmetric Fock space. Indeed,

$$\begin{split} \langle L_i \zeta^k, \zeta^l \rangle &= \langle \zeta^k, L_i^* \zeta^l \rangle = \frac{1}{|\mathcal{P}_k| |\mathcal{P}_l|} \langle \sum_{w \in \mathcal{P}_k} \xi_w, L_i^* \sum_{v \in \mathcal{P}_l} \xi_v \rangle \\ &= \frac{1}{|\mathcal{P}_k| |\mathcal{P}_l|} \langle \sum_{w \in \mathcal{P}_k} \xi_w, \sum_{u \in \mathcal{P}_{l-\delta_i}} \xi_u \rangle \\ &= \begin{cases} 0 & \text{if } l \neq k + \delta_i \\ \frac{1}{|\mathcal{P}_k| |\mathcal{P}_l|} |\mathcal{P}_k| = \langle \zeta^l, \zeta^l \rangle & \text{if } l = k + \delta_i \\ = \langle \zeta^{k+\delta_i}, \zeta^l \rangle \end{split}$$

Thus $M_i = P_s L_i|_{\mathcal{K}_n^s}$ where P_s is the projection onto \mathcal{K}_n^s . It follows immediately that

$$MM^* = \sum_{i=1}^n M_i M_i^* = P_s \sum_{i=1}^n L_i L_i^* \big|_{\mathcal{K}_n^s} = I - \xi_{\varnothing} \xi_{\varnothing}^* \le I.$$

The main dilation result in this context is due to Drury [30]. It says that every strict commuting contraction dilates to a multiple of $M = [M_1, \ldots, M_n]$.

Theorem 8.3. Let $A = [A_1, \ldots, A_n]$ be a commuting n-tuple of operators on a Hilbert space \mathcal{H} with ||A|| < 1. Then there is an isometry V of \mathcal{H} into $\mathcal{K}_n^s \otimes \mathcal{H}$ such that $A_i V^* = V^* M_i$ for $1 \le i \le n$.

We sketch a proof based on the Frahzo–Bunce–Popescu dilation, which yields an isometry V into $\mathcal{K}_n \otimes \mathcal{H}$ such that $A_i V^* = V^*(L_i \otimes I)$ for $1 \leq i \leq n$. It suffices to show that the range of V is contained in $\mathcal{K}_n^s \otimes \mathcal{H}$. However the intertwining relation implies that the range $\mathcal{V} = V\mathcal{H}$ is invariant for each $L_i^* \otimes I$, and the commutativity of the A_i 's implies that $L_u^* L_v^* \xi = L_v^* L_u^* \xi$ for all $u, v \in \mathbb{F}_n^+$ and all $\xi \in \mathcal{V}$. So if we write $\xi = \sum_{w \in \mathbb{F}_n^+} \xi_w \otimes x_w$ where $x_w \in \mathcal{H}$, then we obtain $x_{uvw} = x_{vuw}$ for all $u, v, w \in \mathbb{F}_n^+$. At bit of thought shows that this forces $x_u = x_v$ if u(z) = v(z), and thus ξ is symmetric as claimed.

Now Arveson's dilation theory yields a canonical completely positive map φ from $C^*(M)$ into $C^*(A)$ such that $\varphi(M_i) = A_i$. In fact, $\varphi(M_u M_v^*) = A_u A_v^*$ for all $u, v \in \mathbb{F}_n^+$. This readily extends to all commuting contractions by considering $rA = (rA_1, \ldots, rA_n)$ for r < 1 and letting r increase to 1. In particular, one obtains Drury's von Neumann inequality:

Corollary 8.4. Let $A = [A_1, \ldots, A_n]$ be a commuting n-tuple of operators on a Hilbert space \mathcal{H} with $||A|| \leq 1$. Then $||p(A)|| \leq ||p(M)||$ for all polynomials p in n commuting variables.

Arveson [9] goes a step further and shows that $C^*(M)$ is the C*-envelope of M. Indeed, an easy calculation shows that $C^*(M)$ is irreducible and contains the compact operators. The commutators of M_i and M_j^* are easily computed, and are compact. Thus the quotient $C^*(M)/\mathfrak{K}$ is an abelian C^* algebra and the images $Z_i = M_i + \mathfrak{K}$ satisfy $ZZ^* = \sum_{i=1}^n |Z_i|^2 = I$. Thus the maximal ideal space is contained in the unit sphere $\partial \mathbb{B}_n$. One can show that the spectrum is invariant under the action of the unitary group \mathfrak{U}_n . Indeed, each $U \in \mathfrak{U}_n$ determines the unitary operator \widetilde{U} on Fock space, and this preserves \mathcal{K}_n^s . The restriction $\widetilde{U}|_{\mathcal{K}_n^s}$ implements an automorphism of $C^*(M)$. Thus $C^*(M)/\mathfrak{K} \simeq C(\partial \mathbb{B}_n)$.

8.3. **Reproducing kernel Hilbert spaces.** To make the connection to interpolation, we observe that \mathcal{K}_n^s is a reproducing kernel Hilbert space. The eigenvectors ν_{λ} for $\lambda \in \mathbb{B}_n$ span all of \mathcal{K}_n^s , and

$$\nu_{\lambda} = (1 - \|\lambda\|^2)^{1/2} \sum_{w \in \mathbb{F}_n^+} \overline{w(\lambda)} \xi_w = (1 - \|\lambda\|^2)^{1/2} \sum_{k \in \mathbb{N}_0^n} \overline{\lambda}^k |\mathcal{P}_k| \zeta^k.$$

We renormalize for convenience and set

$$u_{\lambda} = (1 - \|\lambda\|^2)^{-1/2} \nu_{\lambda} = \sum_{k \in \mathbb{N}_0^n} \overline{\lambda}^k |\mathcal{P}_k| \zeta^k.$$

Each vector $\zeta \in \mathcal{K}_n^s$ determines an analytic function

$$\hat{\zeta}(\lambda) = \langle \zeta, u_\lambda \rangle = \sum_{k \in \mathbb{N}_0^n} c_k \lambda^k$$

Moreover

$$|\hat{\zeta}(\lambda)| \le \|\zeta\| \|u_{\lambda}\| = \|\zeta\| (1 - \|\lambda\|^2)^{-1/2}$$

Thus $\hat{\mathcal{K}}_n^s$ becomes a Hilbert space of analytic functions in which the point evaluations are continuous. To emphasize that this is an L^2 norm on these functions, we will write $||f||_2$ for the norm of an element f in $\hat{\mathcal{K}}_n^s$.

Let \mathfrak{M} denote the WOT-closed algebra generated by the M_i . This is the WOT-closed algebra generated by the compressions of the L_i 's to \mathcal{K}_n^s which is the orthogonal complement of $\overline{\mathfrak{C}\mathcal{K}_n}$. Hence by Corollary 8.2, this is completely isometrically isomorphic to $\mathfrak{L}_n/\mathfrak{C}$. So we have a concrete representation of the quotient, and moreover every element of the unit ball of \mathfrak{M} is in the image of the unit ball of \mathfrak{L}_n .

The important fact is that operators in \mathfrak{M} have a nice analytic form in this functional representation. Indeed, if A is any operator in \mathfrak{L}_n and $M = P_s A|_{\mathcal{K}_n^s}$ is the compression to \mathcal{K}_n^s , then because $A^*\nu_{\lambda} = \overline{\hat{A}(\lambda)}\nu_{\lambda}$ we obtain

$$\widehat{M}\widehat{\zeta}(\lambda) = \langle A\zeta, u_{\lambda} \rangle = \langle \zeta, A^*u_{\lambda} \rangle = \widehat{A}(\lambda) \langle \zeta, u_{\lambda} \rangle = \widehat{A}(\lambda) \widehat{\zeta}(\lambda).$$

In particular, the operators M_i become multiplication operators by the coordinate functions z_i . The operators in \mathfrak{M} are analytic multipliers on $\hat{\mathcal{K}}_n^s$, and we may write $M = M_h$ where $M = A + \mathfrak{C}$ and $h = \hat{A}$. Thus the operator norm equals the multiplier norm:

$$||M|| = ||M_h|| = \sup\{||hf||_2 : ||f||_2 \le 1\}.$$

Conversely, suppose that h is a bounded multiplier; so that M_h is a bounded operator. Then the Cesaro means $h_n = C_n(f)$ are polynomials and M_{h_n} converges to M_h in the strong operator topology. Since each M_{h_n} is a polynomial in the M_i 's, it follows that M_h belongs to \mathfrak{M} . Thus \mathfrak{M} is precisely the algebra of multipliers. So the image of the Gelfand map is realized as the algebra of multipliers for this special reproducing kernel Hilbert space.

More generally, a reproducing kernel Hilbert space (see [1]) on a set X is determined by a positive definite function k(x, y) on $X \times X$, meaning that $[k(x_i, x_j)]$ is a positive definite matrix for every *finite* subset $\{x_1, \ldots, x_k\}$ of X. The functions $k_x(y) = k(x, y)$ span a space of functions on X on which we define an inner product $\langle \sum a_i k_{x_i}, \sum b_j k_{y_j} \rangle = \sum a_i \overline{b_j} k(x_i, y_j)$. The completion \mathcal{H}_k is called a reproducing kernel Hilbert space, and each vector ξ determines a function $\hat{\xi}(x) = \langle \xi, k_x \rangle$.

A multiplier is a (bounded) function h on X such that the map $M_h \xi(x) = h(x)\hat{\xi}(x)$ is continuous. This is equivalent to the boundedness of the adjoint map which has the nicer formulation $M_h^* k_x = \overline{h(x)}k_x$. The norm on h is just the operator norm $||M_h||$. One defines matrix multipliers as matrices with multipliers as coefficients. This kernel is said to have the *complete* Nevanlinna–Pick property if for every finite subset x_1, \ldots, x_s of X and finite set C_1, \ldots, C_k of $p \times p$ matrices, there is a matrix multiplier h of norm at most 1 with $\hat{h}(x_i) = C_i$ for $1 \leq i \leq s$ exactly when the Pick condition holds:

$$\left[(I_p - C_i C_j^*) \langle k_{x_j}, k_{x_i} \rangle \right]_{s \times s} \ge 0.$$

In our case, one has $\langle \nu_{\lambda}, \nu_{\mu} \rangle = \frac{1}{1 - \langle \lambda, \mu \rangle}$ which leads to the precise condition obtained above. In other words, symmetric Fock space is a complete Nevanlinna–Pick kernel.

Agler [2] reformulated the Nevanlinna–Pick problem in this way. Exactly which reproducing kernels have this property was solved by Quiggin [51] and McCullough [39]. An easy argument reduces to the irreducible case in which $k(x, y) \neq 0$ for all $x, y \in X$.

Theorem 8.5. A necessary and sufficient for an irreducible positive definite kernel k(x, y) on $X \times X$ to have the complete Nevanlinna–Pick property is that for every finite subset x_1, \ldots, x_s of X, the $s \times s$ matrix

$$\left[\frac{1}{k(x_i, x_j)}\right]$$

has exactly one positive eigenvalue counting multiplicity.

In our case, this is the matrix

$$\left[1 - \langle \lambda_i, \lambda_j \rangle\right] = \mathbf{1} \cdot \mathbf{1}^* - \Lambda^* \Lambda$$

where **1** is the column vector consisting of s 1's, and $\Lambda = (\lambda_1, \ldots, \lambda_s)$ is an $n \times s$ matrix with columns equal to λ_i . Now $\Lambda^*\Lambda$ is positive and norm less than s; while $\mathbf{1} \cdot \mathbf{1}^*$ is rank one and norm exactly s. It follows that $\mathbf{1} \cdot \mathbf{1}^* - \Lambda^*\Lambda$ has exactly one positive eigenvalue. So this is a complete NP kernel, which leads to a rather different proof than ours.

Agler and McCarthy [3] use this to show that (irreducible) complete Nevanlinna–Pick kernels have the rather specific form

$$k(x,y) = \frac{d(x)d(y)}{1 - f(x,y)}$$

where d is a non-vanishing function on X and f is a positive semidefinite function on $X \times X$ taking values in the open unit disk. If n is the rank of the Hermitian form f, they then show that k is just the restriction of the kernel for symmetric Fock space \mathcal{K}_n^s to some subset of the ball \mathbb{B}_n . In other words, the spaces \mathcal{K}_n^s are in a certain sense the universal complete Nevanlinna–Pick kernels.

References

- G. Adams, P. McGuire and V. Paulsen, Analytic reproducing kernels and multiplication operators, Illinois J. Math. 36 (1992), 404–419.
- [2] J. Agler, Nevanlinna-Pick interpolation on Sobolev space, Proc. Amer. Math. Soc. 108 (1990), 341–351.
- [3] J. Agler and J.E. McCarthy, Complete Nevanlinna-Pick kernels, preprint, 2000.
- [4] A. Arias and G. Popescu, Factorization and reflexivity on Fock spaces, Integral Eqtns. & Operator Theory 23 (1995), 268–286.
- [5] A. Arias and G. Popescu, Noncommutative interpolation and Poisson transforms, Israel J. Math. 115 (2000), 205–234.
- [6] W.B. Arveson, Subalgebras of C*-algebras, Acta Math. 123 (1969), 141-224.
- [7] W.B. Arveson, Subalgebras of C*-algebras II, Acta Math. 128 (1972), 271-308.
- [8] W.B. Arveson, Interpolation problems in nest algebras, J. Func. Anal. 20 (1975), 208–233.
- W.B. Arveson, Subalgebras of C*-algebras III: multivariable operator theory, Acta Math. 181 (1998), 159–228.
- [10] A. Beurling, On two problems concerning linear transformations in Hilbert space, Acta Math. 81 (1949), 239–255.
- [11] H. Bercovici, C. Foiaş and C. Pearcy, Dual algebras with applications to invariant subspaces and dilation theory, CBMS Notes 56, Amer. Math. Soc., Providence, 1985.
- [12] H. Bercovici, Hyper-reflexivity and the factorization of linear functionals, J. Func. Anal. 158 (1998), 242–252.
- [13] O. Bratteli, D. Evans and P. Jorgensen, Compactly supported wavelets and representations of the Cuntz relations, Appl. Comput. Harmon. Anal. 8 (2000), 166–196.
- [14] O. Bratteli and P. Jorgensen, Endomorphisms of B(H) II, J. Func. Anal. 145 (1997), 323–373.
- [15] O. Bratteli and P. Jorgensen, Iterated function systems and permutation representations of the Cuntz algebra, Mem. Amer. Math. Soc. 139 (1999), no. 663.
- [16] O. Bratteli and P. Jorgensen, Isometries, shifts, Cuntz algebras and miltiresolution wavelet analysis of scale N, Integral Eqtns. & Operator Theory 28 (1997), 382–443.

- [17] O. Bratteli and P. Jorgensen, Wavelet filters and infinite dimensional unitary groups, preprint, 2000.
- [18] O. Bratteli, P. Jorgensen, and G.L. Price, *Endomorphisms of B(H)*, in "Quantization of nonlinear partial differential equations" (W. Arveson et al, Eds.), Proc. Symp. Pure Math. **59** pp. 93–138, Amer. Math. Soc., Providence, 1996.
- [19] J. Bunce, Models for n-tuples of non-commuting operators, J. Func. Anal. 57 (1984), 21–30.
- [20] E. Christensen, Extensions of derivations II, Math. Scand. 50 (1982), 111–112.
- [21] J. Cuntz, Simple C*-algebras generated by isometries, Comm. Math. Phys. 57 (1977), 173–185.
- [22] K.R. Davidson, The distance to the analytic Toeplitz operators, Illinois J. Math. 31 (1987), 265–273.
- [23] K.R. Davidson, C*-Algebras by Example, Fields Institute Monograph Series 6, Amer. Math. Soc., Providence, RI, 1996.
- [24] K.R. Davidson, E. Katsoulis and D.R. Pitts, The Structure of Free Semigroup Algebras, J. reine angew. Math., to appear.
- [25] K.R. Davidson, D.W. Kribs and M.E. Shpigel, *Isometric Dilations of non-commuting finite rank n-tuples*, Canad. J. Math., to appear.
- [26] K.R. Davidson and M. Ordower, Some Exact Distance Constants, Lin. Alg. Appl. 208/209 (1994), 37–55.
- [27] K.R. Davidson and D.R. Pitts, Invariant subspaces and hyper-reflexivity for free semigroup algebras, Proc. London Math. Soc. 78 (1999), 401–430.
- [28] K.R. Davidson and D.R. Pitts, The algebraic structure of non-commutative analytic Toeplitz algebras, Math. Ann. 311 (1998), 275–303.
- [29] K.R. Davidson and D.R. Pitts, Nevanlinna-Pick Interpolation for non-commutative analytic Toeplitz algebras, Integral Eqtns. & Operator Theory 31 (1998), 321–337.
- [30] S.W. Drury, A generalization of von Neumann's inequality to the complex ball, Proc. Amer. Math. Soc. 68 (1978), 300–304.
- [31] A. Frahzo, Models for non-commuting operators, J. Func. Anal. 48 (1982), 1–11.
- [32] J. Glimm, Type I C*-algebras, Ann. Math. 73 (1961), 572–612.
- [33] M. Hamana, Injective envelopes of operator systems, Publ. Res. Inst. Math Sci. 15 (1979), 773–785.
- [34] P. Jorgensen. Minimality of the data in wavelet filters, preprint, 2000.
- [35] D.W. Kribs, *Factoring in non-commutative analytic Toeplitz algebras*, J. Operator Theory, to appear.
- [36] D.W. Kribs, Completely Positive Maps in Dilation and Wavelet Decompositions, preprint 2000.
- [37] M. Laca, Endomorphisms of B(H) and Cuntz algebras, J. Operator Theory 30 (1993), 381–396.
- [38] D.E. Marshall, Blaschke products generate H^{∞} , Bull. Amer. Math. Soc. 82 (1976), 494–496.
- [39] S. McCullough, The local de Branges-Rovnyak construction and complete Nevanlinna-Pick kernels, Algebraic methods in operator theory, 15–24, Birkhauser Boston, Boston, MA, 1994.
- [40] V.I. Paulsen, Completely bounded maps and dilations, Pitman Res. Notes Math. 146, Longman Sci. Tech. Harlow, 1986.
- [41] G. Popescu, Models for infinite sequences of noncommuting operators, Acta. Sci. Math. (Szeged) 53 (1989), 355–368.
- [42] G. Popescu, Isometric dilations for infinite sequences of noncommuting operators, Trans. Amer. Math. Soc. 316 (1989), 523–536.
- [43] G. Popescu, Characteristic functions for infinite sequences of noncommuting operators, J. Operator Theory 22 (1989), 51–71.

- [44] G. Popescu, Multi-analytic operators and some factorization theorems, Indiana Univ. Math. J. 38 (1989), 693-710.
- [45] G. Popescu, Von Neumann Inequality for (B(H)ⁿ)₁, Math. Scand. 68 (1991), 292– 304.
- [46] G. Popescu, Functional calculus for noncommuting operators, Mich. J. Math. 42 (1995), 345–356.
- [47] G. Popescu, Multi-analytic operators on Fock spaces, Math. Ann. 303 (1995), 31-46.
- [48] Popescu, G., Noncommuting disc algebras and their representations, Proc. Amer. Math. Soc. 124 (1996), 2137–2148.
- [49] G. Popescu, Poisson transforms on some C*-algebras generated by isometries, J. Func. Anal. 161 (1999), 27–61.
- [50] R. Powers, An index theory for semigroups of *-endomorphisms of B(H) and type II factors, Canad. J. Math. 40 (1988), 86–114.
- [51] P. Quiggin, For which reproducing kernel Hilbert spaces is Pick's theorem true? Integral Eqtns. & Operator Theory 16 (1993), 244–266.
- [52] S. Rosenoer, Distance estimates for von Neumann algebras, Proc. Amer. Math. Soc. 86 (1982), 248–252.
- [53] W. Rudin, Function Theory in the Unit Ball of \mathbb{C}^n , Grund. math. Wiss. 241, Springer-Verlag, New York, 1980.
- [54] B. Russo and H. Dye, A note on unitary operators in C*-algebras, Duke J. Math. 33 (1966), 413–416.
- [55] D. Sarason, Generalized interpolation in H[∞], Trans. Amer. Math. Soc. 127 (1967), 179–203.
- [56] B. Sz. Nagy and C. Foiaş, Harmonic analysis of operators on Hilbert space, North Holland Pub. Co., London, 1970.
- [57] D. Voiculescu, Symmetries of some reduced free product C*-algebras, Lect. Notes Math. 1132, 556–588, Springer Verlag, New York, 1985.
- [58] J. Wermer, On invariant subspaces of normal operators, Proc. Amer. Math. Soc. 3 (1952), 270–277.

PURE MATH. DEPT., U. WATERLOO, WATERLOO, ON N2L 3G1, CANADA *E-mail address:* krdavids@math.uwaterloo.ca