

THE FAILURE OF RATIONAL DILATION ON A TRIPLY CONNECTED DOMAIN

MICHAEL A. DRITSCHEL AND SCOTT MCCULLOUGH

1. INTRODUCTION

Let R denote a domain in \mathbb{C} with boundary B . Let X denote the closure of R . An operator T on a complex Hilbert space \mathcal{H} has X as a spectral set if $\sigma(T) \subset X$ and

$$\|f(T)\| \leq \|f\|_R = \sup\{|f(z)| : z \in R\}$$

for every rational function f with poles off X . The expression $f(T)$ may be interpreted either in terms of the Riesz functional calculus, or simply by writing the rational function f as pq^{-1} for polynomials p and q , in which case $f(T) = p(T)q(T)^{-1}$.

The operator T has a normal B -dilation if there exists a Hilbert space \mathcal{K} containing \mathcal{H} and a normal operator N on \mathcal{K} so that

$$(1.1) \quad f(T) = P_{\mathcal{H}}f(N)|_{\mathcal{H}},$$

for every rational function f with poles off the closure of R . Here $P_{\mathcal{H}}$ is the orthogonal projection of \mathcal{K} onto \mathcal{H} .

It is evident that if T has a normal B -dilation, then T has X as a spectral set. The Sz.-Nagy Dilation Theorem implies that the closed unit disk is a spectral set for T if and only if T has a unitary dilation. A deep result of Agler says that if the closure of an annulus \mathbb{A} is a spectral set for T , then T has a normal boundary of \mathbb{A} -dilation [2]. Despite a great deal of effort, there has until recently been little progress in determining whether or not these examples are typical or exceptional for finitely connected spectral sets. The problem may be formulated in terms of contractive representations of the algebra of rational functions on the spectral set. Arveson showed [8] that rational dilation holds precisely when such representations are automatically completely contractive. Hence the problem bears some resemblance to the famous problem of Halmos, solved by Pisier [30], on similarity to a contraction, where the question hinged on deciding whether bounded representations of the disc algebra are completely bounded.

Agler, Harland and Raphael [5] have recently given an example of a triply connected domain R and a 4×4 matrix with X as a spectral set such that this

Received by the editors April 28, 2004.

2000 *Mathematics Subject Classification*. Primary 47A25; Secondary 30C40, 30E05, 30F10, 46E22, 47A20, 47A48.

Key words and phrases. Dilations, spectral sets, multiply connected domains, inner functions, Herglotz representations, Fay reproducing kernels, Riemann surfaces, theta functions, transfer functions, Nevanlinna-Pick interpolation.

The first author's research was supported by the EPSRC.

The second author's research was supported by the NSF.

©2005 American Mathematical Society
Reverts to public domain 28 years from publication

matrix does not have a normal B -dilation. In this paper we give a proof that for any bounded triply connected domain with analytic disjoint boundary components, there is an operator T with X as a spectral set which does not have a normal B -dilation. It can be shown that T can be taken to be a matrix [27], but the argument does not give a bound on the size. Our proof, in contrast to that of Agler, Harland and Raphael, does not use numerical computations. Nevertheless it borrows freely from and overlaps considerably with [4], [3], [5]; see also [6], [11], [12]. We implicitly employ the rank one bundle shifts over R [1], [31], the representation for the corresponding reproducing kernels in terms of theta functions due to Fay [20] and Ball and Clancey [10] (see also [15], [16], [13], [32]), and some elementary compact Riemann surface theory.

For the remainder of the paper we assume $-1 < c_1 < 0$, $0 < c_2 < 1$ are two points on the real axis, $0 < r_k < \min\{|c_k|, |1 - c_k|\}$, and that R is a region obtained by removing disks with centers c_k and radii r_k from the unit disk. By [14], any bounded triply connected region with boundary components consisting of disjoint analytic curves is conformally equivalent to an annulus with a disk removed. By scaling and rotating if necessary, we assume that the outer boundary of the annulus is the unit circle and that the center of the removed disk is on the real axis. By choosing a point between the removed disk and the central disk of the annulus and applying a Möbius transformation mapping the unit disk to itself and the chosen point to the origin, we get a region R of the sort described initially. Hence there is no loss of generality in restricting ourselves to such regions. Note that in this case, the boundary B of R consists of three components, $B_0 = \{|z| = 1\}$, $B_1 = \{|z - c_1| = r_1\}$ and $B_2 = \{|z + c_2| = r_2\}$. We set $X = R \cup B$.

As it happens, only three parameters are really needed to distinguish conformally distinct triply connected domains, so there is some redundancy in using the four parameters c_1, c_2, r_1, r_2 . However, we will later require that no minimal inner function on our domain have a zero of multiplicity three at 0 (all such functions are normalized to have 0 as a zero). We will show that in choosing the point to move to the origin in going from the annulus with the disk removed to the unit disk with disks on either side of the imaginary axis removed in the previous paragraph, there is a choice that enforces this condition on the zeros of these inner functions (Corollary 2.13).

The main theorem of the paper is the following (see also Agler, Harland and Raphael [5]).

Theorem 1.1. *For R a bounded triply connected subset of \mathbb{C} the boundary B of which consists of disjoint analytic curves, there exists a Hilbert space \mathcal{H} and a bounded operator T on \mathcal{H} so that \bar{R} is a spectral set for T , but T does not have a normal B -dilation.*

Here is the idea of the proof. We assume without loss of generality that R is the unit disk with two smaller disks removed, each centered on the real axis and on opposite sides of the imaginary axis. Let \mathcal{C} denote the cone generated by

$$\{H(z)(1 - \psi(z)\psi(w)^*)H(w)^* : \psi \in B\mathbb{H}(X), H \in M_2(\mathbb{H}(X))\},$$

where $B\mathbb{H}(X)$ is the unit ball in the supremum norm of the space of functions analytic in a neighborhood of X , $M_2(\mathbb{H}(X))$ the 2×2 matrices of functions analytic in a neighborhood of X . For $F \in M_2(\mathbb{H}(X))$, we set

$$\rho_F = \sup\{\rho > 0 : I - \rho^2 F(z)F(w)^* \in \mathcal{C}\}.$$

It happens that $\rho_F > 0$. Suppose that F is unitary-valued on B and that $\rho_F < 1$. Then there exists a Hilbert space \mathcal{H} and an operator T such that T has X as a spectral set, but T does not have a normal B -dilation (Theorem 5.2). We also show that if $F \in M_2(\mathbb{H}(X))$ is as above and $\rho_F = 1$, then under suitable assumptions regarding zeros, F has a Herglotz representation (Theorem 5.10). The kernel of the adjoint of multiplication by F in this case has its kernel spanned by a finite collection of reproducing kernels due to Fay (a variant on the Szegő kernel for harmonic measure—Lemma 5.11), and this results in F being diagonalizable (Theorem 5.13). Our counterexample is obtained by constructing a function satisfying the boundary conditions and hypotheses concerning zeros mentioned above which is not diagonalizable. For such a function, $\rho_F < 1$, and hence rational dilation fails.

The organization of the paper differs somewhat from the outline of the proof mentioned above. Minimal inner functions play a key role throughout, so we begin by detailing the results on harmonic functions and analytic functions with positive real parts which are related to these inner functions via a Cayley transform. Much of this material can be found in Grunsky's monograph [22] for more general planar domains. These inner functions are used to give a "scalar" Herglotz representation theorem for functions analytic in a neighborhood of R with unimodular boundary values (Proposition 2.14). The reader is also referred to [5], where the importance of Herglotz representations in characterizing spectral sets is cogently presented.

We then seemingly digress into some basic results on Riemann surfaces and theta functions, which are useful in constructing meromorphic functions on a compact Riemann surface with a given zero/pole structure. The compact surface Y we consider is very special: it is the double of our two-holed region R , obtained topologically by gluing a second copy of R to itself along B . Our minimal inner functions extend to meromorphic functions on the double via reflection, and indeed the same goes for matrix-valued inner functions.

Using fairly elementary tools, we are able to say quite a bit about the zero structure of the minimal inner functions on R based on the parametrization of these functions. We then extend results on scalar inner functions to some 2×2 matrix-valued inner functions, constructing a family of such functions which will ultimately give rise to our example of such a function with a particular zero structure which is not diagonalizable.

We then turn to considering Szegő kernels $K^a(\zeta, z)$ on R with respect to harmonic measure for the point $a \in R$. The truly remarkable fact discovered by Fay [20] is that these kernels extend to meromorphic functions on the double Y . Indeed, Fay gives a representation for $K^a(\zeta, z)$ in terms of theta functions and the prime form. Ball and Clancey [10] give a similar representation, which is actually a bit more explicit as it only involves theta functions for the n -torus family of Abrahamse kernels associated to a multiply connected domain of connectivity $n + 1$. The explicit nature of the theta function representation of the meromorphic kernels $K^a(\zeta, z)$ allows the determination of their zero/pole structure. Next we consider certain finite linear combinations of $K^a(\zeta, z)$ with coefficients in \mathbb{C}^2 , and show that if an analytic function is in the span of certain kernels, then it must be constant. This is used later to prove the diagonalization result mentioned above.

We next turn to representing nice 2×2 matrix-valued inner functions—these are functions which are unitary valued on the boundary with what we term a "standard" zero set—see section 3.2 for the definition. First we show the connection

between the failure of rational dilation and ρ_F being strictly less than one for some contractive analytic function F with unitary boundary values. We make a brief foray into matrix measures, and then prove that when $\rho_F = 1$ for such functions, we have a nice representation for $1 - F(z)F(w)^*$. Part of this relies on an Agler-Nevanlinna-Pick interpolation result (Proposition 5.8), proved using a transfer function representation, as well as a uniqueness result (Proposition 5.9). First we show that a Herglotz type representation holds over finite subsets of points in R for such an F . This is then extended to all of R via the interpolation theorem, and if the finite set of points was chosen in just the right way, this extension is unique. The result is what we term a “tight” representation.

In the next section we return to Fay’s kernel, and use it to prove our diagonalization result. We show that some of the matrix inner functions constructed earlier are not diagonalizable, proving our main theorem.

The last section shows why, once we know that rational dilation fails, there is a finite-dimensional example. This part is based on work of Paulsen [27].

2. SPECTRAL SETS AND SOME FUNCTION THEORY ON R

We begin with some standard material on spectral sets. Then we review harmonic functions, analytic functions of positive real part, and inner functions on R , the unit disk with two disks removed as described above.

2.1. Spectral sets. Let $\mathbb{H}(X)$ denote complex-valued functions which are analytic in a neighborhood of X , and $M_2(\mathbb{H}(X))$ the 2×2 matrices of such functions. We likewise let $\mathcal{R}(X)$ denote the rational functions with poles off of X , with $M_2(\mathcal{R}(X))$ defined in the obvious manner. For f in either $\mathbb{H}(X)$ or $M_2(\mathbb{H}(X))$ define

$$\|f\| = \|f\|_R = \sup\{\|f(z)\| : z \in R\},$$

where $\|f(z)\|$ is the modulus of the scalar $f(z)$ or the operator norm of the 2×2 matrix $f(z)$ respectively.

Throughout the rest of this section T is a bounded operator on the complex Hilbert space \mathcal{H} which has X as a spectral set.

To the operator T associate the homomorphism, $\phi_T : \mathcal{R}(X) \rightarrow \mathcal{B}(\mathcal{H})$ by $\phi_T(p/q) = p(T)q(T)^{-1}$, p and q polynomials. When T has a normal B -dilation, equation (1.1) can then be expressed as $\phi_T(f) = P_{\mathcal{H}}f(N)|_{\mathcal{H}}$ for $f \in \mathcal{R}(X)$. Using the Riesz functional calculus and Runge’s theorem, the map ϕ_T extends continuously to $\phi_T : \mathbb{H}(X) \rightarrow \mathcal{B}(\mathcal{H})$. Conversely, a contractive unital homomorphism $\pi : \mathbb{H}(X) \rightarrow \mathcal{B}(\mathcal{H})$, that is, a unital homomorphism satisfying

$$\|\pi(f)\| \leq \|f\|_R$$

for $f \in \mathbb{H}(X)$, determines an operator with X as a spectral set.

If T has a normal B -dilation as in (1.1) and if G is in $M_2(\mathcal{R}(X))$, then

$$G(T) = P_{\mathcal{H} \oplus \mathcal{H}}G(N)|_{\mathcal{H} \oplus \mathcal{H}}.$$

Since, by the maximum principle, $\|G(N)\| = \|G\|_R$, it follows that

$$(2.1) \quad \|G(T)\| \leq \|G\|_R.$$

Indeed, the same reasoning implies this for any $G \in M_n(\mathcal{R}(X))$, or equivalently, that ϕ_T is completely contractive.

If $F \in M_2(\mathbb{H}(X))$ (so that the entries $F_{j,\ell}$ of F are analytic in a neighborhood of X), but not necessarily rational, it then still makes sense to consider, on $\mathcal{H} \oplus \mathcal{H}$, the operator

$$F(T) = (F_{j,\ell}(T))$$

where the $F_{j,\ell}(T)$ are defined using the Riesz functional calculus.

Lemma 2.1. *If T has X as a spectral set and a normal B -dilation, and if $F \in \mathbb{H}(X)$ is unitary valued on B , then*

$$\|F(T)\| \leq 1.$$

Proof. Choose a compact set K so that the interior of K contains X and K is a subset of the domain of analyticity of F .

Using Runge’s Theorem (entrywise), there exists a sequence G_n of rational 2×2 matrix-valued functions with poles off K which converges uniformly to F on K . From standard results about the functional calculus, $\{G_n(T)\}$ converges to $F(T)$ in the operator norm. Since $\|F\|_R = 1$, the sequence $\{\|G_n\|_R\}$ converges to 1. An application of equation (2.1) and a limit argument completes the proof. \square

Remark 2.2. The set X is a complete spectral set for an operator $T \in \mathcal{B}(\mathcal{H})$ if $\|F(T)\| \leq \|F\|_R$ for every $n \times n$ (no bound on n) matrix-valued rational function with poles off of R (the norm $\|F\|_R$ defined in the expected way). It is a result of Arveson that if X is a complete spectral set for T , then T has a normal B -dilation [9], [8]; see also [29], [28]. (Arveson’s result is actually stated and proved for a commuting n -tuple of operators with a domain in \mathbb{C}^n .)

The following may be found in Conway [18].

Lemma 2.3. *If $\pi : \mathbb{H}(X) \rightarrow \mathcal{B}(\mathcal{H})$ is a contractive unital homomorphism, then X is a spectral set for $T = \pi(\zeta)$. Here $\zeta(z) = z$. Moreover, if $f \in \mathbb{H}(X)$, then $\pi(f) = f(T)$.*

Proof. Given $\lambda \notin X$, the function $f_\lambda(z) = (z - \lambda)^{-1} \in \mathbb{H}(X)$. Thus, as π is a unital homomorphism,

$$I = \pi(1) = \pi(f_\lambda \cdot (\zeta - \lambda)) = \pi(f_\lambda)\pi(\zeta - \lambda) = \pi(f_\lambda)(T - \lambda).$$

It follows that the spectrum of T is in X .

If $f = p/q \in \mathcal{R}(X)$, p, q polynomials, then $\pi(f) = f(T)$ since π is a homomorphism and $f(T)$ can be defined as $p(T)q(T)^{-1}$. Since $\|\pi(f)\| \leq \|f\|_R$, it follows that X is a spectral set for T .

Finally, if $f \in H(X)$, then by Runge’s Theorem, there exists a sequence $\{f_n\}$ from $\mathcal{R}(X)$ which converges uniformly to f on a compact set K containing X . By continuity, both $\{f_n(T)\}$ converges to $f(T)$ and $\{\pi(f_n)\}$ converges to $\pi(f)$ (in the operator norm). Since $f_n(T) = \pi(f_n)$, it follows that $f(T) = \pi(f)$. \square

For $F \in M_2(\mathbb{H}(X))$, define

$$\pi(F) = (\pi(F_{j,\ell})).$$

Thus, π is defined entry-wise.

Proposition 2.4. *Suppose $\pi : \mathbb{H}(X) \rightarrow \mathcal{B}(\mathcal{H})$ is a contractive unital homomorphism and let $T = \pi(\zeta)$. If T has a normal B -dilation and if F is a 2×2 matrix-valued function analytic in a neighborhood of R and unitary valued on B , then $\|\pi(F)\| \leq 1$.*

Proof. By Lemma 2.1, $\|F(T)\| \leq 1$. On the other hand, by Proposition 2.3, $F(T) = \pi(F)$. \square

2.2. Harmonic functions. Most of the results and discussion in this section come from Fisher's book [21].

For each point $z \in R$ there exists a measure ω_z on B such that if h is continuous on X and harmonic in R , then

$$h(z) = \int_B h(\zeta) d\omega_z(\zeta).$$

The measure ω_z is a Borel probability measure on B and is known as *harmonic measure* for the point z .

For $z \in R$, $d\omega_z$ and Lebesgue measure ds are mutually absolutely continuous. Thus there is a Radon-Nikodým derivative,

$$\mathbb{P}(\cdot, z) = \frac{d\omega_z}{ds}.$$

This is also the Poisson kernel for R ; that is,

$$\mathbb{P}(\cdot, z) = -\frac{1}{2\pi} \frac{\partial}{\partial n} g(\cdot, z),$$

where $g(\cdot, z)$ is the Green's function for the point z , n the outward normal.

If h is a positive harmonic function in R , then there exists a positive measure μ such that

$$(2.2) \quad h(z) = \int_B \mathbb{P}(\zeta, z) d\mu(\zeta).$$

Conversely, given a positive measure μ , the formula (2.2) defines a positive harmonic function in R . Note that

$$h(0) = \int_B d\mu = \mu(B).$$

Let h_j denote harmonic measure for B_j . Thus, h_j is the solution to the Dirichlet problem with boundary values 1 on B_j and 0 on B_ℓ , $\ell \neq j$. Alternatively,

$$h_j(z) = \int_{B_j} \mathbb{P}(\zeta, z) ds(\zeta),$$

where ds is normalized arc length measure for B (Nehari [26], section VII.3).

The symmetry in the domain R yields a simple but useful symmetry for the h_j .

Lemma 2.5. For $z \in R$, $h_j(z) = h_j(z^*)$.

Proof. This is obvious from the definition of $h_j(z)$. \square

2.3. Analytic functions with positive real part. As in the last section, some of the results can be found in the book of Fisher [21]. The material on extreme points of the set of normalized analytic functions with positive real part is a special case of that found in Grunsky [22].

If μ and h are as in the formula (2.2), then the periods $P_j(h)$ of the harmonic conjugate of h around B_j , $j = 0, 1, 2$, are given by

$$P_j(h) = \int_B Q_j d\mu,$$

where Q_j is the normal derivative of h_j . Of course h is the real part of an analytic function if and only if $P_j(h) = 0$, $j = 0, 1, 2$.

Lemma 2.6. *The functions Q_j have no zeros on B . Moreover, $Q_j > 0$ on B_j and $Q_j < 0$ on B_ℓ for $\ell \neq j$.*

Before proving Lemma 2.6 we note the following consequence.

Lemma 2.7. *If h is a nonzero positive harmonic function on R which is the real part of an analytic function and if h is represented in terms of a positive measure μ as in equation (2.2), then $\mu(B_j) > 0$ for each j .*

Proof. If $\mu(B_1) = 0$, then, as $Q_1 < 0$ on $B_0 \cup B_2$, $P_1(h) < 0$. Thus, $\mu(B_1) > 0$. A similar argument holds for $\mu(B_2)$.

Let h_0 denote harmonic measure for B_0 . Then $\sum_0^2 h_j = 1$. Consequently, $\sum_0^2 Q_j = 0$. On the other hand, from the proof of Lemma 2.6, $Q_0 > 0$. Thus, if $\mu(B_0) = 0$, then

$$\begin{aligned} \sum_{j=1}^2 \int_B Q_j d\mu &= \sum_{j,\ell=1}^2 \int_{B_\ell} Q_j d\mu \\ &= - \sum_{\ell=1}^2 \int_{B_\ell} Q_0 d\mu < 0, \end{aligned}$$

implying that not both $P_j(h) = 0$. So we must have $\mu(B_0) > 0$. □

Proof of Lemma 2.6. Let h_j denote harmonic measure for B_j . Thus, h_j is the solution to the Dirichlet problem with boundary values 1 on B_j and 0 on B_ℓ , $\ell \neq j$. Alternatively,

$$h_j(z) = \int_{B_j} \mathbb{P}(\zeta, z) ds(\zeta),$$

where ds is normalized arc length measure for B .

The functions Q_j are related to the h_j by

$$Q_j = \frac{\partial h_j}{\partial n},$$

where the derivative is with respect to the outward normal to the boundary. Note that, for $j = 0, 1, 2$, the partial derivatives $\frac{\partial h_j}{\partial n}$ are evidently nonnegative on B_j and nonpositive on B_ℓ , for $\ell \neq j$, from which it follows that $Q_j \geq 0$ on B_j and $Q_j \leq 0$ on B_ℓ for $\ell \neq j$.

Let R' denote the reflection of R about B_0 by $z \mapsto 1/z^*$. The functions h_j naturally extend to harmonic functions on $X \cup R'$ by

$$h_j(1/z^*) = -h_j(z).$$

Thus, Q_j extends to R' .

If Q_j were to have infinitely many zeros in B , then Q_j would be zero. By the Cauchy-Riemann equations, \tilde{h}_j , a harmonic conjugate of h_j exists on B_0 for $j = 1, 2$. Moreover, since $Q_j \leq 0$ on B_0 and has only finitely many zeros on B_0 , \tilde{h}_j is strictly decreasing (in the positive orientation on B_0). In particular, \tilde{h}_j is one-one on B_0 . Observe, if Q_j is zero at a point $\zeta \in B_0$, then, as the derivative of h_j tangential to B at ζ is also 0, the function $f = h_j + i\tilde{h}_j$ is analytic near ζ and has zero derivative at ζ . It follows that f is at least two

to one in a sufficiently small neighborhoods of ζ . But then \tilde{h}_j could not be one to one on B_0 near ζ , a contradiction. Hence $Q_j < 0$ on B_0 ($j = 1, 2$).

Similar arguments show $Q_j > 0$ on B_j and $Q_j < 0$ on B_ℓ for $\ell \neq j$. □

Let $\Pi = B_0 \times B_1 \times B_2$.

Lemma 2.8. *For each $p \in \Pi$, the kernel of*

$$M(p) = \begin{pmatrix} Q_1(p_0) & Q_1(p_1) & Q_1(p_2) \\ Q_2(p_0) & Q_2(p_1) & Q_2(p_2) \end{pmatrix}$$

is one-dimensional and spanned by a vector with all entries strictly positive. In particular, there is a continuous function $\tau : \Pi \rightarrow \mathbb{R}^3$ such that $\tau(p)$ is entry-wise positive, the sum of the entries is one, and $\tau(p)$ is in the kernel of $M(p)$.

Moreover, τ reflects the symmetry in the domain. Namely,

$$\tau(p_0, p_1, p_2) = \tau(p_0^*, p_1, p_2) = \tau(p_0, p_1^*, p_2) = \tau(p_0, p_1, p_2^*).$$

Proof. Computing the cross product of the rows of $M(p)$, we get the vector with entries

$$\begin{aligned} & Q_1(p_1)Q_2(p_2) - Q_1(p_2)Q_2(p_1), \\ & Q_1(p_2)Q_2(p_0) - Q_1(p_0)Q_2(p_2), \\ & Q_1(p_0)Q_2(p_1) - Q_1(p_1)Q_2(p_0). \end{aligned}$$

This vector is in the kernel of $M(p)$ and by considering the signs of the Q_j on the boundary components B_ℓ , one easily checks that the signs of the last two entries are positive. It is also clear from the signs of the entries that the rows of $M(p)$ are linearly independent. Hence $M(p)$ is rank two and its kernel is one dimensional. To finish the proof of the first part of the lemma, it remains to show that the sign of the first entry is positive. Since $\sum_{j=0}^2 h_j(P) = 1$ for $P \in X$, it follows that for $P \in B_1$ or $P \in B_2$, that

$$Q_1(P) + Q_2(P) = -Q_0(P) > 0.$$

Thus,

$$\begin{aligned} & Q_1(p_1)Q_2(p_2) - Q_1(p_2)Q_2(p_1) \\ &= Q_1(p_1)Q_2(p_2) + Q_1(p_1)Q_1(p_2) - Q_1(p_1)Q_1(p_2) - Q_1(p_2)Q_2(p_1) \\ &= Q_1(p_1)(Q_2(p_2) + Q_1(p_2)) - Q_1(p_2)(Q_1(p_1) + Q_2(p_1)) \\ &= -Q_1(p_1)Q_0(p_2) + Q_1(p_2)Q_0(p_1). \end{aligned}$$

Examination of the signs of the terms on the right-hand side above shows that the first term is indeed positive.

To prove the last part of the lemma, simply note that the symmetry in the domain implies each h_j is symmetric, $h_j(z^*) = h_j(z)$ and thus, $Q_j(z^*) = Q_j(z)$. □

The lemma allows the construction of canonical analytic functions of positive real part on R , since, by construction,

$$(2.3) \quad h_p = \sum \tau_j(p) \mathbb{P}(\cdot, p_j)$$

is a positive harmonic function with no periods. Indeed, h_p corresponds to the measure $\sum \tau_j(p) \delta_{p_j}$ in equation (2.2). In particular, by our normalization of τ ,

$$h_p(0) = 1.$$

Let f_p denote the analytic function with real part h_p such that $f(p) = 1$.

Let $H(R)$ be the locally convex metrizable topological space of holomorphic functions on R with the topology of uniform convergence on compact subsets of R . The space $H(R)$ has the Heine-Borel property; i.e., closed bounded subsets of $H(R)$ are compact. Let

$$\mathbb{K} = \{f \in H(R) : f(0) = 1, f + f^* > 0\}.$$

The set \mathbb{K} is easily seen to be closed.

Lemma 2.9. *The set \mathbb{K} is compact.*

What needs to be shown is that \mathbb{K} is bounded—in other words, that for each compact subset $K \subset R$ there exists an M_K so that $|f(z)| \leq M_K$ for all $f \in \mathbb{K}$ and $z \in K$.

That \mathbb{K} is bounded is straightforward in the case that R is replaced by the unit disk, and we consider this case to begin with.

Let $P_r(\theta)$ denote the Poisson kernel and $Q_r(\theta)$ the conjugate Poisson kernel,

$$P_r(\theta) = \frac{1 - r^2}{1 + r^2 - 2r \cos(\theta)}$$

$$Q_r(\theta) = \frac{2r \sin(\theta)}{1 + r^2 - 2r \cos(\theta)}.$$

If U is a positive harmonic in the unit disk \mathbb{D} with $U(0) = 1$, then there is a probability measure μ so that

$$U(r \exp(i\theta)) = \int_{-\pi}^{\pi} P_r(\theta - t) d\mu(t).$$

The harmonic conjugate V of U normalized by $V(0) = 0$ is given by

$$V(r \exp(i\theta)) = \int_{-\pi}^{\pi} Q_r(\theta - t) d\mu(t).$$

Thus, as μ is a probability measure, for each $0 < r < 1$ there is a constant M_r so that if $|z| \leq r$, then $|U(z)|, |V(z)| \leq M_r/2$. By conformal mapping, the result is seen to hold for any bounded simply connected domain.

The proof of the lemma is based upon the above result for simply connected domains, using the fact that R can be written as the union of two simply connected domains.

Proof of Lemma 2.9. It suffices to show that for each compact set K there is a constant M_K so that if $f \in \mathbb{K}$ and $z \in K$, $|f(z)| \leq M_K$. In fact, it is enough to work with the compact sets of the form

$$K_\epsilon = \{|z| \leq 1 - \epsilon\} \cap \{|z + c_1| \geq r_1 + \epsilon\} \cap \{|z + c_2| \geq r_2 + \epsilon\},$$

for $\epsilon > 0$ sufficiently small.

First, observe that if f is holomorphic with positive real part on R and $f(0) = 1$, then for each $\epsilon > 0$ there is a constant N_ϵ so that for all $z \in K_\epsilon$, $h(z) \leq N_\epsilon$, by virtue of the representation (2.2).

Given $\epsilon > 0$, the open simply connected sets

$$R_\epsilon^\pm = \left\{ z = a + bi \in K_{\frac{\epsilon}{2}} : \pm b \geq -\frac{\min(r_1, r_2)}{2} \right\}$$

contain 0 and their union is $K_{\frac{\epsilon}{2}}$. After conformal mapping from the disk taking 0 to 0, it follows that there exists M_{ϵ}^{\pm} so that if f is analytic on R with $f(0) = 1$, then for all $z \in R_{\epsilon}^{\pm}$, $|f(z)| \leq M_{\epsilon}^{\pm}$. \square

Lemma 2.10 ([22]). *The extreme points of \mathbb{K} are precisely $\{f_p : p \in \Pi\}$.*

Proof. It is evident that each f_p is an extreme point of \mathbb{K} . So consider the converse.

Let $f \in \mathbb{K}$. The real part of f is a positive harmonic function h with $h(0) = 1$. Hence there exists a probability measure μ such that

$$h(z) = \int_B \mathbb{P}(\zeta, z) d\mu(\zeta).$$

Suppose the support of μ on B_0 contains more than one point and so can be written as the union of disjoint sets $A_1, A_2 \subset B_0$ with $\mu(A_j) > 0$.

Let

$$\alpha_{j,\ell} = \int_{A_\ell} Q_j d\mu, \quad \ell = 1, 2,$$

and

$$\kappa_{j,m} = \int_{B_m} Q_j d\mu, \quad m = 1, 2.$$

Since h is the real part of an analytic function,

$$0 = \int_B Q_j d\mu.$$

Thus, $\kappa_{j,1} + \kappa_{j,2} + (\alpha_{j,1} + \alpha_{j,2}) = 0$. Since also $Q_j < 0$ on B_0 for $j = 1, 2$,

$$\kappa_{j,1} + \kappa_{j,2} = -(\alpha_{j,1} + \alpha_{j,2}) > 0.$$

This gives $\kappa_{1,1} \geq |\kappa_{1,2}| = -\kappa_{1,2}$. Hence the determinant of $\kappa = (\kappa_{i,j})$ is positive.

Since the determinant of $\kappa = (\kappa_{i,j})$ is positive the solution of

$$\begin{pmatrix} \kappa_{1,1} & \kappa_{1,2} \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix} \begin{pmatrix} \beta_{1,\ell} \\ \beta_{2,\ell} \end{pmatrix} = - \begin{pmatrix} \alpha_{1,\ell} \\ \alpha_{2,\ell} \end{pmatrix}$$

is given by

$$\frac{1}{\det k} \begin{pmatrix} \kappa_{2,2} & -\kappa_{1,2} \\ -\kappa_{2,1} & \kappa_{1,1} \end{pmatrix} \begin{pmatrix} -\alpha_{1,\ell} \\ -\alpha_{2,\ell} \end{pmatrix} = \begin{pmatrix} \beta_{1,\ell} \\ \beta_{2,\ell} \end{pmatrix}.$$

In view of the signs of the $\kappa_{j,m}$ and $\alpha_{j,\ell} < 0$, it follows that $\beta_{1,\ell}, \beta_{2,\ell} \geq 0$.

Define positive measures ν_ℓ by

$$\nu_\ell(A) = \mu(A \cap A_\ell) + \beta_{1,\ell} \mu(A \cap B_1) + \beta_{2,\ell} \mu(A \cap B_2).$$

Then

$$\int_B Q_j d\nu_\ell = \alpha_{j,\ell} + \kappa_{j,1} \beta_{1,\ell} + \kappa_{j,2} \beta_{2,\ell} = 0$$

and therefore each

$$h_\ell = \int_B \mathbb{P}(\cdot, \zeta) d\nu_\ell(\zeta)$$

is the real part of an analytic function g_ℓ with $g_\ell(0) = \nu(B)$. Since also $\nu_1 + \nu_2 = \mu$, we have $h_1 + h_2 = h$. Thus, $\frac{g_\ell}{g_\ell(0)} \in \mathbb{K}$ and

$$f = g_1(0) \left(\frac{g_1}{g_1(0)} \right) + g_2(0) \left(\frac{g_2}{g_2(0)} \right).$$

We conclude that f is not an extreme point.

Next suppose B_1 is a disjoint union of sets A_1, A_2 with $\mu(A_j) > 0$. Let

$$\alpha_{j,\ell} = \int_{A_\ell} Q_j d\mu, \quad \ell = 1, 2,$$

and

$$\kappa_{j,m} = \int_{B_m} Q_j d\mu, \quad m = 0, 1.$$

This time the signs are $\alpha_{1,\ell} \geq 0$, $\alpha_{2,\ell} \leq 0$, $\kappa_{j,0} \leq 0$, $\kappa_{1,2} < 0$, and $\kappa_{2,2} > 0$. So, with

$$\kappa = \begin{pmatrix} \kappa_{1,0} & \kappa_{1,2} \\ \kappa_{2,0} & \kappa_{2,2} \end{pmatrix}$$

the determinant of κ is negative. Thus, the solution of

$$\kappa \begin{pmatrix} \beta_{0,\ell} \\ \beta_{2,\ell} \end{pmatrix} = - \begin{pmatrix} \alpha_{1,\ell} \\ \alpha_{2,\ell} \end{pmatrix}$$

is given by

$$\frac{1}{\det \kappa} \begin{pmatrix} \kappa_{2,2} & -\kappa_{1,2} \\ -\kappa_{2,0} & \kappa_{1,0} \end{pmatrix} \begin{pmatrix} -\alpha_{1,\ell} \\ -\alpha_{2,\ell} \end{pmatrix} = \begin{pmatrix} \beta_{1,\ell} \\ \beta_{2,\ell} \end{pmatrix}.$$

In view of the signs of the entries and of $\det(\kappa)$, the $\beta_{j,\ell}$ are all nonnegative.

Define positive measures ν_ℓ by

$$\nu_\ell(A) = \mu(A \cap A_\ell) + \beta_{0,\ell} \mu(A \cap B_0) + \beta_{2,\ell} \mu(A \cap B_2).$$

Then

$$\int_B Q_j d\nu_\ell = \alpha_{j,\ell} + \kappa_{j,0} \beta_{0,\ell} + \kappa_{j,2} \beta_{2,\ell} = 0$$

and therefore each

$$h_\ell = \int_B \mathbb{P}(\cdot, \zeta) d\nu_\ell(\zeta)$$

is the real part of an analytic function g_ℓ with $g_\ell(0) = \nu(B)$. Since also $\nu_1 + \nu_2 = \mu$, we have $h_1 + h_2 = h$. The argument proceeds as before, with $\frac{g_\ell}{g_\ell(0)} \in \mathbb{K}$ and

$$f = g_1(0) \left(\frac{g_1}{g_1(0)} \right) + g_2(0) \left(\frac{g_2}{g_2(0)} \right).$$

We conclude that f is not an extreme point. □

Lemma 2.11. *The set of extreme points of \mathbb{K} is a closed set and the function taking Π to \mathbb{K} by $p \mapsto f_p$ is a homeomorphism.*

Proof. It suffices to show, if $p(n) \in \Pi$ converges to $p(0)$ in Π , then $f_{p(n)}$ converges uniformly on compact subsets of R to $f_{p(0)}$. Since \mathbb{K} is compact and $\{f_{p(n)}\}$ is a sequence in \mathbb{K} , some subsequence, still denoted $\{f_{p(n)}\}$, converges to some $f \in \mathbb{K}$. Let $h_{p(n)}$ denote the real part of $f_{p(n)}$ and let μ_n denote the measure which represents $h_{p(n)}$ so that

$$h_{p(n)}(z) = \int_B \mathbb{P}(\zeta, z) d\mu_n$$

and $\mu_n = \sum \tau_j(p(n)) \delta_{p(n)_j}$, where τ is defined in Lemma 2.8. See also equation (2.3).

The measures μ_n converge to the measure μ_0 weakly so that $h_{p(n)}(z)$ converges to $h_{p(0)}(z)$ pointwise in R . It follows that $h_{p(0)}$ is the real part of f . Thus, $\{f_{p(n)}\}$ converges, in $H(R)$, to $f_{p(0)}$. Since every subsequence of our original subsequence has a subsequence which converges to $f_{p(0)}$, the whole subsequence converges to

$f_p(0)$. This shows that the mapping is continuous. Since Π is compact, it follows that our mapping is a homeomorphism and its range is compact. \square

2.4. Scalar inner functions on R . Up to post composition by a Möbius transformation, the inner functions on R with precisely three zeros are canonically parameterized by the 2-torus, \mathbb{T}^2 . Details may be found in the book by Fay [20]. This section contains an alternate description of this family using results from the previous sections. The following is well known (see, for example, Fisher [21], Ch. 4, ex. 6,7).

Proposition 2.12. *A nonconstant inner function ψ on R has at least three zeros counting with multiplicity. Moreover, if ψ has exactly three zeros, z_0, z_1, z_2 , then for $j = 1, 2$,*

$$\sum_{\ell=0}^2 h_j(z_\ell) = 1.$$

As a first application, the last proposition allows us to show that we may assume that our region supports no inner function with a zero of multiplicity three at 0, and so consequently we will take it that R has this property throughout the remainder of the paper.

Corollary 2.13. *We may assume without loss of generality that there is no inner function on R with a zero of multiplicity three at 0.*

Proof. Suppose f is an inner function on R with exactly three zeros counting multiplicity. Then if f takes the value $f(x)$ with multiplicity three where $x \in (c_1 + r_1, c_2 - r_2)$ (i.e., $f(z) - f(x)$ has a zero of multiplicity three at x), there is a Möbius transformation m of the disk to itself fixing 1 and moving $f(x)$ to 0. Thus $m \circ f$ is an inner function with zero of multiplicity three at the origin. Hence by Proposition 2.12, $h_j(x) = 1/3$. Since h_1 is continuous and equal to 1 on B_1 , if we choose x close enough to $c_1 + r_1 \in B_1$, it will be the case that $h_1(x) > 1/3$, and so f does not take the value $f(x)$ with multiplicity three for any inner function with three zeros. Now take a Möbius transformation mapping x to the origin. This takes R to a new region which again has the property that it is a region which is the unit disk with two smaller disks removed on either side of the imaginary axis and centers on the real axis. \square

For $p \in \Pi$, let

$$\phi_p = \frac{f_p - 1}{f_p + 1},$$

where f_p is the extreme point of \mathbb{K} corresponding to the point $p = (p_0, p_1, p_2)$ in Π as in the previous subsection. The real part, h_p , of f_p is harmonic across $B \setminus \{p_0, p_1, p_2\}$ and therefore f_p is (at least locally) analytic across $B \setminus \{p_0, p_1, p_2\}$. Furthermore, near p_j , f_p has the form $\frac{g_j(z)}{z - p_j}$ for some g_j analytic in a neighborhood of s_j and nonvanishing at p_j ([21], Ch. 4). From these properties of f_p , it follows that ϕ_p is continuous onto B and $|\phi_p| = 1$ on B . By the reflection principle, ϕ_p extends to be analytic in a neighborhood of B . Since ϕ_p is inner and extends analytically across B , and $\phi_p^{-1}(\{1\}) = \{p_0, p_1, p_2\}$, it follows that the preimage of each point $z \in \mathbb{D}$ is exactly three points, counted with multiplicity. In particular, ϕ_p has precisely three zeros.

On the other hand, suppose ϕ is analytic in a neighborhood of R , has modulus one on B , and three zeros in R . As above, it follows that $\phi^{-1}(1)$ consists of three points. Moreover, the real part of

$$f = \frac{1 + \phi}{1 - \phi}$$

is a positive harmonic function which is zero on B except at those points z where $\phi(z) = 1$. By Lemma 2.7 we must have $\phi^{-1}(\{1\}) = \{p_0, p_1, p_2\}$ where $p_j \in B_j$. So if we also assume $\phi(0) = 0$, then $\phi = \phi_p$ for some $p \in \Pi$.

Proposition 2.14. *If ψ is analytic in R and if $|\psi| \leq 1$ on R , then there exists a positive measure μ on Π and a measurable function h defined on Π whose values are functions $h(\cdot, p)$ analytic in R so that*

$$1 - \psi(z)\psi(w)^* = \int h(z, p)[1 - \phi_p(z)\phi_p(w)^*]h(w, p)^* d\mu(p).$$

Proof. First suppose $\psi(0) = 0$.

Let

$$f = \frac{1 + \psi}{1 - \psi}.$$

Verify

$$\psi = \frac{f - 1}{f + 1}$$

and hence,

$$(2.4) \quad 1 - \psi(z)\psi(w)^* = 2 \frac{f(z) + f(w)^*}{(f(z) + 1)(f(w)^* + 1)}.$$

Since h , the real part of f , is positive and $f(0) = 1$, the function f is in \mathbb{K} defined in the previous subsection.

Since \mathbb{K} is a compact subset of the topological vector space $H(R)$ and the set of extreme points $\{f_p : p \in \Pi\}$ of \mathbb{K} is a compact set by Lemma 2.11, there exists a (regular Borel) probability measure ν on Π so that

$$f = \int_{\Pi} f_p d\nu(p).$$

Using the definition of ϕ_p and equation (2.4), verify

$$1 - \psi(z)\psi(w)^* = \int_{\Pi} \frac{1 - \phi_p(z)\phi_p(w)^*}{(f(z) + 1)(1 - \phi_p(z))(1 - \phi_p(w)^*)(f(w)^* + 1)} d\nu(p).$$

If $\psi(0) = a$, then one has a representation as above since

$$1 - \left(\frac{\psi(z) - a}{1 - a^*\psi(z)} \right) \left(\frac{\psi(w) - a}{1 - a^*\psi(w)} \right)^* = \frac{(1 - aa^*)(1 - \psi(z)\psi(w)^*)}{(1 - a^*\psi(z))(1 - a\psi(w)^*)}.$$

□

While we have used the three parameters in Π to parameterize the inner functions with exactly three zeros, after rotation, really only two are needed. Indeed,

$$\phi_p(1)^* \phi_p = \phi_q,$$

where $q = (1, q_1, q_2)$, $q_j \in B_j$ for $j = 1, 2$, are the unique points such that $\phi_p(q_j) = \phi_p(1)$.

3. SOME RIEMANN SURFACES AND THETA FUNCTIONS

We review some of the results from the theory of Riemann surfaces which we will need subsequently. In particular, we look at theta functions, the use of which in operator theory was pioneered by Clancey [15], [16]. The presentation here borrows heavily from Ball and Clancey [10] as well as Mumford [24], [25], and Farkas and Kra [19].

Let Y denote the double of the bordered Riemann surface $X = R \cup B$. Recall Y is obtained topologically by gluing a second copy R' of R along B . The complex atlas is then found by anti-holomorphically reflecting the complex structure from X to R' . Since B_0 is the unit circle, the anti-copy R' of R may be thought of as the reflection of R in B_0 given by $z \mapsto z^{*-1}$. In particular, there is an anti-holomorphic involution $J : Y \rightarrow Y$ which fixes B . For $\zeta \in R$, $J\zeta$ is its twin in R' .

Note that Y is a compact Riemann surface. It is important since we can sometimes extend analytic functions defined on R to meromorphic functions on Y . In particular this is true for inner functions, which extend by reflection.

We get a lot of mileage out of a few basic tools for analyzing meromorphic functions on compact Riemann surfaces. To begin with, if such a function is nonconstant and takes a given value n times, it takes all of its values n times (counting multiplicity). Furthermore, and as a consequence, such a function has equal numbers of zeros and poles. Finally, a meromorphic function on a compact Riemann surface without zeros or poles is constant.

The material in subsections 3.3–3.5 is solely for describing the zeros of the Fay kernel in Section 4. The reader who is willing to accept the statement of Theorem 4.2 may skip these parts.

3.1. Minimal meromorphic functions on R . Our Riemann surface Y also comes naturally equipped with a conformal involution which fixes 6 points, namely

$$\iota(\zeta) = J(\zeta^*).$$

Geometrically, ι is rotation by π in the axis through ± 1 . The fixed points of the involution are precisely the Weierstrass points of Y , namely $\pm 1, c_1 \pm r_1, c_2 \pm r_2$ ([19], Cor. 1, p. 108).

If f is a meromorphic function on a compact Riemann surface, then there is a number n , the degree of f , so that f takes each value n times, counting multiplicity. In particular, if f has just one pole, then f is one to one and the Riemann surface is conformally equivalent to the Riemann sphere. Thus, a nonconstant meromorphic function on Y must have at least two poles (and zeros).

Up to post composition by a Möbius transformation, there is a unique meromorphic function ϕ on Y with precisely two poles. Moreover, ϕ is ramified with branching one at the Weierstrass points.

The construction of a meromorphic function with two poles and zeros found in Farkas and Kra ([19], Theorem III.7.3) is the following. The Riemann surface \tilde{Y} , obtained as the quotient of Y by the map $\zeta \mapsto \iota^{-1}(\zeta)$, is a Riemann surface of genus zero by the Riemann-Hurwitz formula. Thus, \tilde{Y} is conformally equivalent to the Riemann sphere, and the canonical quotient map

$$Y \mapsto \tilde{Y}$$

followed by a conformal map from \tilde{Y} to the Riemann sphere generates the desired meromorphic function. From this construction it is evident that if ϕ is any meromorphic function with two poles and two zeros, then $\phi\iota = \phi$. Another way to see this is that $\phi - \phi\iota$ has zeros at the six Weierstrass points, while it has at most four poles, and so must be constantly zero.

Proposition 3.1. *If f is a meromorphic function on Y with precisely two poles Q_1 and Q_2 , then $Q_2 = \iota(Q_1)$. In particular, f cannot have one pole in R and the other in $X = R \cup B$. The same statement holds if we consider zeros rather than poles.*

Proof. From the discussion above, $f \circ \iota = f$. The conclusion for the poles of f follows. By considering $1/f$ the conclusion for the zeros of f is also seen to hold. □

Using the proposition it is possible to show that a unimodular function on R which takes the value 1 at exactly one point on each boundary component B_j and is 0 at 0 is uniquely determined. Of course, this is also evident by construction and has already been used.

Corollary 3.2. *Suppose $p_j \in B_j$, for $j = 0, 1, 2$. If $f, g : X \rightarrow \mathbb{C}$ are analytic in a neighborhood of X , unimodular on B , $f^{-1}(\{1\}) \cap X = g^{-1}(\{1\}) \cap X = \{p_0, p_1, p_2\}$, and $f(0) = g(0) = 0$, then $f = g$.*

Proof. By the maximum modulus principle, f and g are strictly less than one in modulus in R . The functions f and g reflect to meromorphic functions on Y , which can take the value 1 only on the boundary of B . Thus, the functions are globally three to one and so have three zeros in R and three poles, at the reflected points, in R' . In particular, f and g share the zero 0 and pole $J0$.

Consider the function

$$\psi = \frac{1 - f}{1 - g}.$$

The zeros of $1 - f$ and those of $1 - g$ agree, so that ψ can only have poles at the poles of f . Since one of these poles matches a pole of g , ψ has only two poles (and two zeros). However, these poles are poles of f which both are in R' , contradicting the proposition. Hence, it must be that ψ is constant so that $1 - f = c(1 - g)$. Since $f(0) = g(0) = 0$, $c = 1$. □

We can also describe minimal inner functions which agree at two of their zeros.

Lemma 3.3. *If ϕ_p and $\phi_{\tilde{p}}$ have two zeros in common (0 and one other) and $\phi_p(1) = \tilde{\phi}_p(1) = 1$, then either they are equal or the remaining zeros form a conjugate pair. In the case that $p = (1, p_1, p_2)$ and $\tilde{p} = (1, p_1, \tilde{p}_2)$, then $\tilde{p}_2 = p_2$ and so $\phi_{\tilde{p}} = \phi_p$.*

Proof. If ϕ_p and $\phi_{\tilde{p}}$ have all three zeros in common, then they are equal, and there is nothing to prove. So assume instead that they have two common zeros (0 and z_1) and unequal zeros z_2, \tilde{z}_2 for ϕ_p and $\phi_{\tilde{p}}$, respectively. Then $\phi_p/\phi_{\tilde{p}}$ has two simple zeros and poles, $z_2, J\tilde{z}_2$ and \tilde{z}_2, Jz_2 , respectively. By Proposition 3.1, $Jz_2 = \iota\tilde{z}_2 = J(\tilde{z}_2^*)$, or $\tilde{z}_2 = z_2^*$.

Now suppose that ϕ_p and $\phi_{\tilde{p}}$ have common zeros 0 and z_1 , and that $p = (1, p_1, p_2)$ and $\tilde{p} = (1, p_1, \tilde{p}_2)$. Then unless $\phi_p/\phi_{\tilde{p}}$ is constant, $\phi_p/\phi_{\tilde{p}} - 1$ has the same two poles as $\phi_p/\phi_{\tilde{p}}$. It would then have exactly two zeros, namely 1 and p_1 . Now apply Proposition 3.1 and note that $\iota(1) = 1 \neq p_1$. The result then follows. □

Lemma 3.4. *Let $p = (1, p_1, p_2)$. If the zeros of ϕ_p are all real, then $p_1, p_2 \in \mathbb{R}$.*

Conversely, for $p = (1, c_1 - r_1, c_2 - r_2)$, the zeros of ϕ_p are all real. In fact, ϕ_p has a zero at zero and one zero in each of the intervals $(-1, c_1 - r_1)$ and $(c_2 + r_2, 1)$ on the real axis in R .

It turns out that for $p = (1, c_1 + r_1, c_2 - r_2)$ or $p = (1, c_1 - r_1, c_2 + r_2)$, the zeros of ϕ_p are also real.

Proof. Suppose that the zeros of ϕ_p are real, and define $\tilde{\phi}_p = \phi_p(\zeta^*)^*$. Then $\tilde{\phi}_p$ has the same zeros as ϕ_p . Hence the ratio of these two functions has no zeros and so must be constant. Since both equal 1 at 1, it follows that they are equal everywhere. In particular,

$$\phi_p(c_j \pm r_j)^* = \phi_p((c_j \pm r_j)^*)^* = \phi_p(c_j \pm r_j)$$

so that $\phi_p(c_j \pm r_j)$ is real (in each of the four cases). Since ϕ_p maps each B_j one-one onto the unit circle, it follows that ϕ_p takes the value 1 at one of the points $c_j \pm r_j$ on B_j , $j = 1, 2$ and this gives the conclusion of the lemma.

For the converse, define $\tilde{\phi}_p$ as above. Since the entries of p are real, Corollary 3.2 implies $\phi_p(\zeta) = \tilde{\phi}_p(\zeta^*)^*$ so that ϕ_p is real on the real axis. Since also ϕ_p takes the value 1 at $c_j - r_j$ and by assumption there is only one point on B_j where ϕ_p equals 1, it takes the value -1 at the points $c_j + r_j$. The intermediate value theorem now implies that ϕ_p has at least one zero in each of the intervals, $(-1, c_1 - r_1)$, $(c_1 + r_1, c_2 - r_2)$, and $(c_2 + r_2, 1)$. Since ϕ_p has exactly three zeros the result follows. □

For $t = (t_1, t_2)$, $t_1, t_2 \in [0, 2\pi)$, let $p_1(t) = c_1 - r_1 \exp(it_1)$ and $p_2 = c_2 - r_2 \exp(it_2)$. These points are on the two inner components of the boundary. On the outer boundary we just use 1. Accordingly, let $p(t) = (1, p_1(t), p_2(t)) \in \Pi$ and, for notational ease, let $\psi_t = \phi_{p(t)}$. Then $\psi_{-t} = \phi_{p(-t)}$.

Lemma 3.5. *There exist t_1, t_2 nonzero such that ψ_t has three distinct zeros $0, z_1, z_2$ with z_1, z_2 both nonreal.*

Proof. Write 0 for $(0, 0)$. From the second part of Lemma 3.4, ψ_0 has distinct real zeros.

Now suppose that $t = (t_1, 0)$. If $\pi > t_1 > 0$ is small enough, the zeros $0, z_1, z_2$ of ψ_t are still distinct. By the first part of Lemma 3.4, if these zeros are all real, then t_1 is a multiple of π . On the other hand suppose that one of the nonzero zeros of ψ_t , say z_1 , is real. Then ψ_t and ψ_{-t} share 0 and z_1 as zeros, as well as both being equal to 1 at 1 and $c_2 - r_2$. Hence by Lemma 3.3 they are equal, implying that $p_1(t)$ is real, or equivalently, that t_1 is a multiple of π . Since we have chosen $0 < t_1 < \pi$, this cannot happen. Thus we must have that both zeros of ψ_t are nonreal for small enough t_1 .

Finally, if we choose t_1 such that $\psi_{(t_1, 0)}$ has nonreal roots, then by continuity, for small enough t_2 , ψ_t will still have nonreal roots with $t = (t_1, t_2)$. Again, for t small enough, these zeros will still be distinct. □

3.2. Some matrix inner functions. We now construct a family of 2×2 matrix-valued analytic functions which are unimodular on the boundary B and have precisely 6 zeros in R starting from certain positive 2×2 matrix-valued harmonic functions on R . This will later provide the basis for our counterexample.

For our purposes, a matrix-valued $F : R \rightarrow M_2(\mathbb{C})$ has *simple zeros* in R if $Z = \{a_1, \dots, a_n\}$, the zeros of $\det(F)$, have multiplicity either one or two, and if the zero $z \in Z$ has multiplicity two, then $F(z) = 0$.

Definition 3.6. We say that F has a *standard zero set* if F has distinct simple zeros $0, a_1, \dots, a_4 \in \mathbb{R}$, with the first having multiplicity two and the others multiplicity one. If furthermore, $\delta_j \neq 0$ are such that $F(a_j)^* \delta_j = 0$, $j = 1, \dots, 4$, then no three of the δ_j 's are collinear.

While it is not required immediately, we also ask that $Ja_j \neq P_k$, $k = 1, 2$, where P_1, P_2 are poles of the Fay kernel $K^0(\cdot, z)$ (defined in Section 4). This last assumption plays an important role in the diagonalization arguments in Section 5.8. In this regard it is also useful to know that the points P_1, P_2 are real and distinct (see Lemma 4.1). Also not required immediately but ultimately useful is the assumption that each δ_j is sufficiently close to some e_k , a standard basis element of \mathbb{C}^2 —how close being determined by Theorem 4.3.

We take t and $p(t)$ as at the end of the last section.

Recall $M(p)$ and $\tau(p)$ from Lemma 2.8. For notational purposes, write $p(t) = (1, q_1, q_2)$. Let τ denote the common value of $\tau(1, q_1, q_2)$, $\tau(1, q_1^*, q_2)$, $\tau(1, q_1, q_2^*)$, and $\tau(1, q_1^*, q_2^*)$.

For $0 \leq \eta \leq 1$, let P_+^η denote the projection onto the span of $\eta e_1 + (1 - \eta^2)^{\frac{1}{2}} e_2$ and P_-^η denote the projection onto $(1 - \eta^2)^{\frac{1}{2}} e_1 - \eta e_2$ (so $P_+^\eta + P_-^\eta = I$). Define $H_{\eta,t}$ by

$$H_{\eta,t} = \tau_0 \mathbb{P}(\cdot, 1)I + \tau_1 \mathbb{P}(\cdot, c_1 - r_1 \exp(it_1))P_+^1 + \tau_1 \mathbb{P}(\cdot, c_1 - r_1 \exp(-it_1))P_-^1 \\ + \tau_2 \mathbb{P}(\cdot, -c_2 - r_2 \exp(it_2))P_+^\eta + \tau_2 \mathbb{P}(\cdot, -c_2 - r_2 \exp(it_2))P_-^\eta.$$

That is, given $x \in \mathbb{C}^2$ a unit vector, $\langle H_{\eta,t}(\zeta)x, x \rangle$ is the positive harmonic function on B corresponding to the measure

$$\mu_{x,x} = \tau_0 \delta_1 + \tau_1 (\delta_{q_1} |x_1|^2 + \delta_{q_1^*} |x_2|^2) \\ + \tau_2 \left(\delta_{q_2} |\eta x_1 + (1 - \eta^2)^{\frac{1}{2}} x_2|^2 + \delta_{q_2^*} |(1 - \eta^2)^{\frac{1}{2}} x_1 - \eta x_2|^2 \right).$$

Then

$$\int_B Q_j d\mu_{x,x} = \tau_0 Q_j(1) + \tau_1 (Q_j(q_1) |x_1|^2 + Q_j(q_1^*) |x_2|^2) \\ + \tau_2 \left(Q_j(q_2) |\eta x_1 + (1 - \eta^2)^{\frac{1}{2}} x_2|^2 + Q_j(q_2^*) |(1 - \eta^2)^{\frac{1}{2}} x_1 - \eta x_2|^2 \right).$$

Using $Q_j(q_k^*) = Q_j(q_k)$ and this choice of τ gives

$$\int_B Q_j d\mu_{x,x} = \tau_0 Q_j(1) + \tau_1 Q_j(q_1) (|x_1|^2 + |x_2|^2) \\ + \tau_2 Q_j(q_2) \left(|\eta x_1 + (1 - \eta^2)^{\frac{1}{2}} x_2|^2 + |(1 - \eta^2)^{\frac{1}{2}} x_1 - \eta x_2|^2 \right) \\ = \tau_0 Q_j(1) + \tau_1 Q_j(q_1) + \tau_2 Q_j(q_2) \\ = 0.$$

Hence $\langle H_{\eta,t}(\zeta)x, x \rangle$ is the real part of an analytic function. It follows that $H_{\eta,t}$ is the real part of an analytic 2×2 matrix-valued function $G_{\eta,t}(\zeta)$ normalized by $G_{\eta,t}(0) = I$.

Our desired functions are

$$(3.1) \quad \Psi_{\eta,t} = (G_{\eta,t} - I)(G_{\eta,t} + I)^{-1}.$$

Lemma 3.7. *For each η ,*

- (1) $\Psi_{\eta,t}$ is analytic in a neighborhood of X and unitary-valued on B ;
- (2) $\Psi_{\eta,t}(0) = 0$;
- (3) $\Psi_{\eta,t}(1) = I$;
- (4) $\Psi_{\eta,t}(q_1)e_1 = e_1$ and $\Psi_{\eta,t}(q_1^*)e_2 = e_2$;
- (5) $\Psi_{\eta,t}(q_2)(\eta e_1 + (1 - \eta^2)^{\frac{1}{2}}e_2) = (\eta e_1 + (1 - \eta^2)^{\frac{1}{2}}e_2)$ and
 $\Psi_{\eta,t}(q_2^*)((1 - \eta^2)^{\frac{1}{2}}e_1 - \eta e_2) = (1 - \eta^2)^{\frac{1}{2}}e_1 - \eta e_2$;
- (6) $\Psi_{0,t} = \begin{pmatrix} \psi_t & 0 \\ 0 & \psi_{-t} \end{pmatrix}$.

Proof. In the neighborhood of a point $p \in B$, the Poisson kernel $\mathbb{P}(\zeta, p)$ is the real part of a function of the form $g_p(\zeta)(\zeta - p)^{-1}$ where g_p is analytic in a neighborhood of p and does not vanish at p . Near any other point $q \in B$, $\mathbb{P}(\zeta, p)$ extends to a harmonic function in a neighborhood of q (see Fisher [21], chapter 4, proposition 6.4). Thus $\mathbb{P}(\zeta, p)$ is, near q , the real part of an analytic function, the real part of which is 0 at q .

From the above discussion, if $p \in B$ is different from 1, q_1, q_1^*, q_2, q_2^* , then $G_{\eta,t}$ is, at least locally, analytic in a neighborhood of p . Furthermore, $G_{\eta,t} + I$ is invertible near p , since $G_{\eta,t}(\zeta) = H_{\eta,t}(\zeta) + iA(\zeta)$, for some self-adjoint matrix-valued function $A(\zeta)$ and $H_{\eta,t}(p) = 0$ for such a p . Thus, $G_{\eta,t} + I$ is invertible at p and by continuity, also invertible near p . The relation

$$I - \Psi_{\eta,t}(\zeta)\Psi_{\eta,t}(\zeta)^* = 2(G_{\eta,t}(\zeta) + I)^{-1}(G_{\eta,t}(\zeta) + G_{\eta,t}(\zeta)^*)(G_{\eta,t}(\zeta) + I)^{* - 1}$$

now shows that $\Psi_{\eta,t}$ is unitary at p .

From the definition of $G_{\eta,t}$, in a neighborhood of 1, there exist analytic functions g_1, g_2, h_1 , and h_2 so that the real parts of h_j are 0 at 1, each g_j is different from 0 at 1, and

$$G_{\eta,t}(\zeta) = \begin{pmatrix} \frac{g_1}{\zeta - 1} & h_1 \\ h_2 & \frac{g_2}{\zeta - 1} \end{pmatrix}.$$

Thus,

$$\begin{aligned} & (G_{\eta,t}(\zeta) + I)^{-1} \\ &= \frac{1}{\frac{g_1 + \zeta - 1}{\zeta - 1} \frac{g_2 + \zeta - 1}{\zeta - 1} - h_1 h_2} \begin{pmatrix} \frac{g_2 + \zeta - 1}{\zeta - 1} & -h_1 \\ -h_2 & \frac{g_1 + \zeta - 1}{\zeta - 1} \end{pmatrix} \\ &= \frac{\begin{pmatrix} (g_2 + \zeta - 1)(\zeta - 1) & -h_1(\zeta - 1)^2 \\ -h_2(\zeta - 1)^2 & (g_1 + \zeta - 1)(\zeta - 1) \end{pmatrix}}{(g_1 + \zeta - 1)(g_2 + \zeta - 1) - h_1 h_2 (\zeta - 1)^2}. \end{aligned}$$

Note that the determinant in the denominator is indeed different from 0 near 1, and as $\zeta \rightarrow 1$, it goes to $g_1(1)g_2(1) \neq 0$. Hence $G_{\eta,t} + I$ is in fact invertible. Next, by directly computing $(G_{\eta,t} - I)(G_{\eta,t} + I)^{-1} = \Psi_{\eta,t}$ we see that $\Psi_{\eta,t}$ is analytic in a neighborhood of 1 and $\Psi_{\eta,t}(1) = I$.

Now move on to q_1 . Near this point there exist analytic functions g, h_2, h_3 , and h_4 so that the real parts of the h_j are 0 at q_1 , the function g does not vanish at q_1

and

$$G_{\eta,t} = \begin{pmatrix} \frac{g}{\zeta - q_1} & h_2 \\ h_3 & h_4 \end{pmatrix}.$$

Since $h_4 + 1$ has real part 1 at q_1 , whereas $g(\zeta - q_1)^{-1}$ has a pole and h_2, h_3 are analytic at q_1 , we see that $G_{\eta,t} + I$ is invertible near q_1 . Furthermore, by direct computation of $\Psi_{\eta,t} = (G_{\eta,t} - I)(G_{\eta,t} + I)^{-1}$ we see that $\Psi_{\eta,t}$ is analytic in a neighborhood of q_1 and

$$\Psi_{\eta,t}(q_1) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{h_4(q_1)-1}{h_4(q_1)+1} \end{pmatrix}.$$

The analogous result holds for the point q_1^* . For the points q_2, q_2^* , the same argument prevails by writing all matrices with respect to the orthonormal basis $\{\eta e_1 + (1 - \eta^2)^{\frac{1}{2}} e_2, \eta e_2 + (1 - \eta^2)^{\frac{1}{2}} e_1\}$ of \mathbb{C}^2 .

Finally, (6) is easily seen from the definition of $H_{0,t}$ and the functions $G_{0,t}, \Psi_{0,t}$. □

We are able to obtain some information on the zeros of $\Psi_{\eta,t}$, at least for certain η .

Lemma 3.8. *There exists an $\epsilon > 0$ such that for all η nonzero but less than ϵ , $\Psi_{\eta,t}$ has a standard zero set.*

Proof. Recall from our choice of t , the zeros of ψ_t are distinct and, aside from 0, are not real. Of course the zeros of ψ_{-t} are just the complex conjugates of those of ψ_t . Thus

$$\Psi_0 = \begin{pmatrix} \psi_t & 0 \\ 0 & \psi_{-t} \end{pmatrix}$$

has a standard zero set.

By choosing a sequence η_n converging to 0 and considering the uniformly bounded sequence $\Psi_n = \Psi_{\eta_n,t}$, there is some subsequence, still written as Ψ_n , which converges uniformly on compact subsets of R to some Ψ . Consequently,

$$G_n = (I + \Psi_n)(I - \Psi_n)^{-1}$$

converges uniformly on compact subsets of R to

$$G = (I + \Psi)(I - \Psi)^{-1}.$$

Now H_n , the real part of G_n , is harmonic and

$$H_n - H_0 = \mathbb{P}(\cdot, q_2)(\tau_2(\eta_n)P_+^{\eta_n} - \tau_2(0)P_+^0) + \mathbb{P}(\cdot, q_2^*)(\tau_2(\eta_n)P_-^{\eta_n} - \tau_2(0)P_-^0).$$

Thus, as both $(P_{\pm}^{\eta_n} - P_{\pm}^0)$ and $(\tau_2(\eta_n) - \tau_2(0))$ converge to 0 with n (τ is continuous by Lemma 2.8), the H_n converge to H_0 uniformly on compact sets. Since also $G(0) = I = G_0(0)$, it follows that the G_n converge to G_0 and hence the Ψ_n converge uniformly to Ψ_0 on compact sets.

Let $d_n(\zeta) = \det(\Psi_n(\zeta))$. This is analytic and unimodular on the boundary and d_n converges uniformly to d_0 on compact sets. From Rouché's theorem (or else Hurwitz's theorem—see Conway [17]), the number of zeros of d_n is constant, equaling 6. Indeed, these zeros vary continuously with n , and d_n has a double zero at 0. If we let $0, w_1, w_2$ denote the zeros of ψ_t , then the zeros of ψ_{-t} are $0, w_1^*, w_2^*$. Hence the zeros of Ψ_0 are exactly given by $\Psi_0(0) = 0, \Psi_0(w_j)e_1 = 0$, and

$\Psi_0(w_j^*)e_2 = 0$, $\ell = 1, 2$, and, listed as $0, w_1, w_2, w_1^*, w_2^*$, they are distinct. Thus, for large n , the zeros $0, a_1^n, a_2^n, a_3^n, a_4^n$ of Ψ_n satisfy this assumption too.

Finally, if $\Psi_n(a_1^n)^* \delta_1^n = 0$ and a_1 is close to w_1 , then

$$\Psi_0(w_1)^* \delta_1^n = (\Psi_0(w_1) - \Psi_0(a_1^n))\delta_1^n + (\Psi_0(a_1^n) - \Psi_n(a_1^n))\delta_1^n.$$

For n large both terms on the right-hand side are small and thus for n large δ is close to (a scalar multiple of) e_1 . Thus, for n large enough the δ_j^n satisfy the assumption that no three are collinear. Finally, since the zeros vary continuously with η and for $\eta = 0$ the zeros, aside from 0, are not real, whereas P_1, P_2 are real (and not zero—see Lemma 4.1), for small enough η , neither P_1 nor P_2 will be a zero of $\Psi_{\eta,t}$ and the nonzero zeros of $\Psi_{\eta,t}$ will be distinct.

This completes the proof. □

3.3. The period matrix and Abel-Jacobi map. Since the harmonic measures h_j vanish on B_0 , for $j = 1, 2$, they reflect across B_0 by

$$\begin{aligned} h_j(1/z^*) &= -h_j(z), \\ h_j(Jz) &= -h_j(z) \end{aligned}$$

to a harmonic function on $Y \setminus B_j$. Note, when R' is viewed as the reflection of R across B_0 , h_j is -1 on the reflection of B_j .

While a harmonic conjugate \tilde{h}_j of h_j is only locally defined, the differential

$$\frac{d(h_j + \tilde{h}_j)}{dz} dz$$

is uniquely determined and globally defined. If ζ is a point of Y , then

$$\chi(\zeta) = \frac{1}{2} \left(\int_{-1}^{\zeta} \frac{d(h_1 + \tilde{h}_1)}{dz} dz \right) - \frac{1}{2} \left(\int_{-1}^{\zeta} \frac{d(h_2 + \tilde{h}_2)}{dz} dz \right)$$

depends upon the path of integration from -1 to ζ , but only up to the number of windings of the path around the boundary components B_1, B_2 and the number of crossings of B_1, B_2 (in passing from R to R'). The choice of -1 for the base point is fairly canonical as it is a Weierstrass point for Y (see subsection 3.1).

The multiple-valued nature of $\chi : Y \rightarrow \mathbb{C}^2$ is concisely described by the period matrix and period lattice for Y . Let

$$\mathbf{P}_{j,\ell} = \frac{1}{i} \int_{B_\ell} \frac{d(h_j + \tilde{h}_j)}{dz} dz = \int_{B_\ell} \frac{\partial h_j}{\partial n} ds.$$

Thus, $\mathbf{P}_{j,\ell}$ is the period of the harmonic conjugate of h_j around B_ℓ . The 2×2 matrix \mathbf{P} has positive definite real part and is called the period matrix for R . (Our period matrix differs by a factor of $1/i$ from what most call the period matrix.)

Let L denote the lattice $\mathbb{Z}^2 + i\mathbf{P}\mathbb{Z}^2$. The Jacobi variety of Y is the quotient $\text{Jac}(Y) = \mathbb{C}^2/L$. Let $[z]$ denote the class of $z \in \mathbb{C}^2$ in $\text{Jac}(Y)$. The mapping $\chi_0 : Y \rightarrow \text{Jac}(Y)$,

$$\chi_0(\zeta) = [\chi(\zeta)]$$

is well defined and known as the Abel-Jacobi map. However, it will often be convenient to work with χ , the multiple-valued version of χ_0 .

Let e_1, e_2 be the usual basis for \mathbb{C}^2 .

Proposition 3.9 ([19], p. 92). *The Abel-Jacobi map has the following properties:*

- (1) χ_0 is a one-to-one conformal map of Y onto its image in $\text{Jac}(Y)$;

- (2) $\chi_0(J\zeta) = -\chi_0(\zeta)^*$;
- (3) If $\zeta \in B$, then $-\chi(\zeta)^* = \chi(\zeta) + n$ for some $n \in \mathbb{Z}^2$.

Property (2) depends upon the choice of basepoint $-1 \in B_0$. Property (3) follows from the observation that for $\zeta \in B$, $h_j(\zeta)$ is either 0 or 1.

3.4. The theta function and its zeros. Details for most of what follows in this subsection can be found in Ch. IV of [19] and Ch. 2 of [24].

Let L be the lattice defined in the last subsection. The Riemann theta function associated to L is the entire function on \mathbb{C}^2 defined by

$$\theta(z) = \sum_{n \in \mathbb{Z}^2} \exp(-\pi \langle \mathbf{P}n, n \rangle + 2\pi i \langle z, n \rangle),$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product on \mathbb{C}^2 .

Straightforward manipulations show that $\theta(z^*) = \theta(z)^*$ and $\theta(-z) = \theta(z)$.

The quasi-periodic behavior of θ with respect to L is given by

$$(3.2) \quad \begin{aligned} \theta(z + \ell) &= \theta(z), \\ \theta(z + i\mathbf{P}m) &= \exp(\pi \langle \mathbf{P}m, m \rangle - 2\pi i \langle z, m \rangle) \theta(z), \end{aligned}$$

where $\ell, m \in \mathbb{Z}^2$.

Given $e \in \mathbb{C}^2$ write $e = u + i\mathbf{P}v$, and define the theta function with characteristic $[e]$ by

$$\theta[e](z) = \theta \begin{bmatrix} u \\ v \end{bmatrix} (z) = \exp(\pi i (\langle \mathbf{P}v, v \rangle + 2\langle u - z, v \rangle)) \theta(z - e).$$

The function $\theta[e] : \mathbb{C}^2 \rightarrow \mathbb{C}$ obeys the period laws

$$(3.3) \quad \begin{aligned} \theta[e](z + \ell) &= \exp(-2\pi \langle \ell, v \rangle) \theta[e](z), \\ \theta[e](z + i\mathbf{P}m) &= \exp(2\pi \langle m, v \rangle) \exp(\pi \langle \mathbf{P}m, m \rangle - 2\pi i \langle z, m \rangle) \theta[e](z). \end{aligned}$$

It turns out that almost all meromorphic functions and differentials on Y can be represented in terms of translates $\theta[e](\chi(\zeta)) : Y \rightarrow \mathbb{C}^2$, which despite being multiple valued has a well-defined zero set, the description of which is due to Riemann.

Theorem 3.10. *There exists a constant vector Δ (depending upon the choice of basepoint) so that for each $e \in \mathbb{C}^2$, either $\theta[e](\chi(\zeta))$ is identically zero, or $\theta[e](\chi(\zeta))$ has exactly 2 zeros P_1, P_2 , and*

$$\chi(P_1) + \chi(P_2) = e - \Delta \quad \text{modulo } L.$$

Here Δ is known as the vector of Riemann constants.

The set

$$\mathcal{N} = \{e \in \text{Jac}(Y) : \theta(\chi(\zeta) - e) \text{ is identically zero}\}$$

is a proper closed subset of $\text{Jac}(Y)$.

3.5. The prime form. The following fact plays an important role in construction of the prime form and thus multiple-valued meromorphic functions on Y with prescribed poles and zeros. Given $e \in \mathbb{C}^2$ such that $\theta(e) = 0$, define

$$\mathcal{E}_e(\zeta, \xi) = \theta(\chi(\zeta) - \chi(\xi) - e).$$

Theorem 3.11 ([24], Ch. 2, Lemma 3.4). *If $e \in \mathbb{C}^2$, $\theta(e) = 0$, and \mathcal{E}_e is not identically zero, then there exists $P \in Y$ so that for each $\xi \in Y$, $\xi \neq P$, the zeros of $\theta[e + \chi(\xi)](\chi(\zeta))$, which coincide with the zeros of $\mathcal{E}_e(\zeta, \xi)$, are precisely ξ and P .*

The following can be found in Mumford [25] (Lemma 1, p. 3.208), but see also [20] and [10].

Theorem 3.12. *There exists $e_* = \frac{1}{2}(u + i\mathbf{P}v) \in \mathbb{C}^2$ such that $2e_* = 0$ modulo L , $\langle u, v \rangle$ is odd (equal to 1 modulo $2\mathbb{Z}$), and \mathcal{E}_{e_*} is not identically zero.*

An e_* as in this theorem is called a nonsingular odd half period and for the remainder we take $e_* = \frac{1}{2}(u_* + i\mathbf{P}v_*)$ as fixed. Note that $\theta(e_*) = 0$, as the fact that $\langle u_*, v_* \rangle$ is an odd integer implies $\theta(e_*) = -\theta(-e_*)$. Indeed, $e_* + e_*^* = 2u_* \in \mathbb{Z}^2$ so that, using the periodicity of θ , for $z \in \mathbb{C}^2$,

$$\theta(z + e_*^*) = \theta(z + e_*^* + 2u_*) = \theta(z - e_*).$$

Hence $\theta(e_*) = \theta(-e_*)$, and so $\theta(e_*) = 0$.

Lemma 3.13. *There exists a $P \in B$ so that for each $\xi \in R$, the multiple-valued function $\mathcal{E}_{e_*}(\zeta, \xi) : Y \rightarrow \mathbb{C}$ given by $\mathcal{E}_{e_*}(\zeta, \xi) = \theta(\chi(\zeta) - \chi(\xi) - e_*)$ is not identically zero and has zeros at precisely P and ξ .*

Proof. From Theorem 3.11, there is a P so that either $\mathcal{E}_{e_*}(\zeta, \xi)$ is identically zero, or has zeros P and ξ . Accordingly, consider the multiple-valued function $g : Y \rightarrow \mathbb{C}$ defined by $g(\zeta) = \theta(\chi(\zeta) - \chi(-1) - e_*)$. Since $[\chi(-1)] = 0$, we may assume $\chi(-1) = 0$. Thus $g(\zeta) = \theta(\chi(\zeta) - e_*)$ and the fact that e_* is nonsingular means g is not identically zero. Hence g has zeros P and -1 .

As $e_* + e_*^* \in \mathbb{Z}^2$, there exists an $n \in \mathbb{Z}^2$ such that $e_*^* = -e_* + n$. Similarly, as $\chi(JP) = -\chi(P)^*$ modulo L , there exists $a, b \in \mathbb{Z}^2$ so that $\chi(JP) = -\chi(P)^* + a + i\mathbf{P}b$. Hence,

$$\begin{aligned} g(JP)^* &= \theta(\chi(JP) - e_*)^* \\ &= \theta(-\chi(P)^* + a + i\mathbf{P}b - e_*)^* \\ &= (\kappa\theta(-\chi(P)^* - e_*))^* \\ &= \kappa^*\theta(-\chi(P) - e_*^*) \\ &= \kappa^*\theta(-\chi(P) + e_* - n) \\ &= \kappa^*\theta(-\chi(P) + e_*) \\ &= \kappa^*\theta(\chi(P) - e_*) \\ &= \kappa^*g(P), \end{aligned}$$

where $\kappa = \exp[\pi(\mathbf{P}a, a) - 2\pi i\langle -\chi(P)^* - e_*, n \rangle]$ is nonzero. Thus, $g(JP) = 0$. It follows that $JP = P$ or $JP = -1$, in which case $P = -1$. Thus, $JP = P$ and so P is in B .

Since P is in B , $\mathcal{E}_{e_*}(\cdot, \xi)$ is not identically zero and the lemma follows from Theorem 3.11. □

Let

$$\theta_*(z) = \theta[e_*(z)] = \exp[\frac{1}{2}\pi i(\langle \mathbf{P}v_*, v_* \rangle + \langle u_* - z, v_* \rangle)]\theta(z - e_*).$$

If $z, w \in Y$ are different from the P of Lemma 3.13, then

$$(3.4) \quad \frac{\theta_*(\chi(\zeta) - \chi(z))}{\theta_*(\chi(\zeta) - \chi(w))} = e^{\pi i[\chi(z) - \chi(w)]} \frac{\theta(\chi(\zeta) - \chi(z) - e_*)}{\theta(\chi(\zeta) - \chi(w) - e_*)}$$

is multiple valued, but its zero/pole structure is well defined: it has a zero at z and a pole at w . As we shall see, this will play an important role in defining reproducing kernels on R with respect to harmonic measure.

4. THE FAY KERNEL FUNCTIONS OF R

We now introduce the reproducing kernel K^a as found in Fay [20]. The description of these kernels involves the critical points for the Green's function $g(\cdot, a)$ for R at the point $a \in R$. A point w is a critical point if the gradient of $g(\zeta, a)$ is 0 at $\zeta = w$. It is well known that in a region of connectivity $n + 1$ there are n of these critical points ([26], p. 133). Thus, in R there are two. In the sequel we will have use of the following fact about the location of these critical points for the choice $a = 0$.

Lemma 4.1. *The critical points of the Green's function $g(\zeta, 0)$ are on the real axis, one in each of the intervals, $(-1, c_1 - r_1)$ and $(c_2 + r_2, 1)$.*

Proof. For notational ease, let $g(\zeta) = g(\zeta, 0)$. From the symmetry of the domain, $\frac{\partial g}{\partial y} = 0$ on the x -axis in R . Since g is 0 at the points -1 and $c_1 - r_1$, Rolle's Theorem implies there is a point $-1 < w_1 < c_1 - r_1$ so that $\frac{\partial g}{\partial x}(w_1) = 0$. Thus, the gradient of g is zero at w_1 . Similarly, there is a point $c_2 + r_2 < w_2 < 1$ such that the gradient of g at w_2 is also zero. \square

For the remainder of the paper, we let $P_1 = Jw_1$ and $P_2 = Jw_2$, where w_1, w_2 are the critical points for the Green's function for R at 0, $g(\cdot, 0)$. Thus, $P_1, P_2 \in R'$. These are the points P_1, P_2 which appear in the definition of a standard zero set, Definition 3.6.

Theorem 4.2. *There is a reproducing kernel K^a for the Hardy space $\mathbb{H}^2(R, \omega_a)$ of functions analytic in R with boundary values in $L^2(\omega_a)$, where ω_a is harmonic measure for the point a . If $z = a$, then $K^a(\zeta, z) = 1$. Otherwise, $K^a(\zeta, z)$ has precisely the poles $P_1(a), P_2(a), Jz$ (where $JP_1(a)$ and $JP_2(a)$ are the critical points for the Green's function for R at a), and three zeros in Y , one of which is at Ja .*

Incidentally, the Hardy space in the theorem corresponds to just one of the two torus parameter families of rank one bundle shifts over R [1]. Indeed, Ball and Clancey [10] give a theta function representation for all of the corresponding reproducing kernels, avoiding the use of the Klein prime form which Fay uses in his formula for the kernel.

The reader who has skipped sections 3.3-3.5 may wish to skip the proof and proceed directly to subsection 4.1.

Proof of Theorem 4.2. There is an $e \in \text{Jac}(Y)$ so that

$$(4.1) \quad K^a(\zeta, z) = \frac{\theta(\chi(\zeta) + \chi(z)^* + e) \theta(\chi(a) + \chi(a)^* + e) \theta_*(\chi(a) + \chi(z)^*) \theta_*(\chi(\zeta) + \chi(a)^*)}{\theta(\chi(a) + \chi(z)^* + e) \theta(\chi(\zeta) + \chi(a)^* + e) \theta_*(\chi(\zeta) + \chi(z)^*) \theta_*(\chi(a) + \chi(a)^*)}.$$

(See Fay [20], Proposition 6.15 and Ball and Clancey, [10].) Straightforward computation using the periodic nature of θ checks that the right-hand side is invariant under $\chi(\zeta) \mapsto \chi(\zeta) + n + i\mathbf{P}m$ for $m, n \in \mathbb{C}^2$ so that $K^a(\zeta, z)$ is in fact single valued and meromorphic in Y . Furthermore, according to Fay, for $z \in X$, $K^a(\zeta, z)$ is analytic as a function of $\zeta \in R$.

One readily verifies from (4.1) that $K^a(\zeta, a) = 1$, so assume $z \neq a$. Now examine the zero/pole structure of the portion of the right side of (4.1) depending on ζ :

$$(4.2) \quad \frac{\theta(\chi(\zeta) + \chi(z)^* + e) \theta_*(\chi(\zeta) + \chi(a)^*)}{\theta(\chi(\zeta) + \chi(a)^* + e) \theta_*(\chi(\zeta) + \chi(z)^*)}.$$

From the comments following (3.4), $\theta_*(\chi(\zeta) + \chi(a)^*)/\theta_*(\chi(\zeta) + \chi(z)^*)$ has a zero at Ja and a pole at Jz . Each of the remaining theta functions has two zeros, which we label $Z_1(z), Z_2(z)$ for the top term and $P_1(a), P_2(a)$ for the bottom term. Hence (4.2) (and so (4.1)) has zeros at $Z_1(z), Z_2(z), Ja$ and poles at $P_1(a), P_2(a), Jz$, the latter of which are all in $R' \cup B$, since K^a is analytic.

If any of the poles and zeros were to cancel, either $K^a(\zeta, z)$ is constant or has two poles and two zeros. Since these kernels are linearly independent and $K^a(\zeta, a) = 1$, the first possibility cannot occur. In the second case we would be left with two poles in $R' \cup B$, which by Proposition 3.1, is also impossible unless Jz cancels with Ja and the other two poles are in B . But then $z = a$ and $K^a(\zeta, a) = 1$, which we have already ruled out. Hence $K^a(\zeta, a)$ has order three, with zeros and poles as claimed. □

4.1. Application of the theta function representation of K^a . We assume throughout that for $j = 1, \dots, 4$, $a_j \in B$ are distinct and $0 \neq \delta_j \in \mathbb{C}^2$ have the property that no three are collinear (the properties of a standard zero set). As usual e_1, e_2 are the standard basis for \mathbb{C}^2 . We write P_1, P_2 for $P_1(0), P_2(0)$, the poles of $K^0(\zeta, z)$.

Let \mathcal{M}_δ^0 denote the span of

$$\begin{aligned} & \{K^0(\zeta, 0)e_1, K^0(\zeta, 0)e_2, K^0(\zeta, a_1)\delta_1, \dots, K(\zeta, a_4)\delta_4\} \\ & = \{e_1, e_2, K^0(\zeta, a_1)\delta_1, \dots, K(\zeta, a_4)\delta_4\}. \end{aligned}$$

Recall that given $p \in \Pi$ there exists an $s \in B_1 \times B_2 = \mathbb{T}^2$ so that, identifying s with $(1, s)$, $\phi_s = \phi_p(1)^* \phi_p$. In particular, the zeros of ϕ_p , $z_0^s = 0, z_1^s, z_2^s$, depend only upon s .

Theorem 4.3. *Let a_1^0, \dots, a_4^0 be points in R so that $P_1, P_2, J0, Ja_1^0, Ja_2^0, Ja_3^0, Ja_4^0$ are all distinct. Let $\{e_1, e_2\}$ denote the standard basis for \mathbb{C}^2 and let $\delta_1^0 = \delta_2^0 = e_1$ and $\delta_3^0 = \delta_4^0 = e_2$. There exists an $\epsilon > 0$ so that if $|a_j^0 - a_j|, \|\delta_j^0 - \delta_j\| < \epsilon$, and if*

$$(4.3) \quad h(\zeta) = \sum c_j K^0(\zeta, a_j) \delta_j + v$$

does not have poles at P_1, P_2 , then h is constant; i.e., each $c_j = 0$.

Furthermore, if $h \neq 0$ has a representation as in equation (4.3), and if there exist $z_1, z_2 \in B$ (not assumed to be distinct but not both zero) such that

$$h(\zeta)K^0(\zeta, z_k) = \sum c_j^k K^0(\zeta, a_j) \delta_j + v_k,$$

then h is constant, $z_1 = a_{j_1}$, and $z_2 = a_{j_2}$, the corresponding δ_{j_k} 's may be taken to be equal to h , and all the other terms are zero.

Note that the theorem is really a statement about the meromorphic functions $K^0(\zeta, a_j)$ on the double Y and so we view ζ as a local coordinate on Y . Indeed, by restricting ζ to be near either P_1 or P_2 it may be assumed that all the points $\zeta, P_1, P_2, Ja_1, \dots, Ja_4$ are in a single chart $U \subset R'$ (U is an open simply connected subset of Y).

With fixed $a \in R$ distinct from $0, JP_1, JP_2$, where P_1, P_2 , are the poles of the kernel $K^0(\zeta, a)$, the residue of the pole of $K^0(\zeta, a)$ at P_j is given by the value of the analytic function of ζ ,

$$(\zeta - P_j)K^0(\zeta, a),$$

at the point $\zeta = P_j$. Let $R_j(a)$ denote this residue.

Lemma 4.4. *The residue $R_j(a)$ varies continuously with a .*

Proof. Consider the theta function representation for $K^0(\zeta, a)$ from Theorem 4.2. The function

$$f(\zeta) = \theta(\chi(\zeta) + \chi(0)^* + e)$$

is analytic and single valued in U . Furthermore, $f(\zeta)$ vanishes to order one at P_j and thus can be written as

$$f(\zeta) = (\zeta - P_j)f_j(\zeta),$$

where f_j is analytic in U and $f_j(P_j) \neq 0$. Given a set $W \subset U$, let $W^* = \{z^* : z \in W\}$. Choose neighborhoods V_j and W of U so that $F : V_j \times W^* \rightarrow \mathbb{C}$ by

$$\begin{aligned} F(\zeta, a^*) &= f(\zeta)K^0(\zeta, a) \\ &= \frac{\theta(\chi(\zeta) + \chi(z)^* + e) \theta(\chi(0) + \chi(0)^* + e) \theta_*(\chi(0) + \chi(z)^*) \theta_*(\chi(\zeta) + \chi(0)^*)}{\theta(\chi(0) + \chi(z)^* + e) \theta_*(\chi(\zeta) + \chi(z)^*) \theta_*(\chi(0) + \chi(0)^*)} \end{aligned}$$

is analytic. Rewriting gives

$$(\zeta - P_j)K^0(\zeta, a) = \frac{F(\zeta, a)}{f_j(\zeta)}.$$

The lemma now follows from the fact that the right-hand side is analytic in $(\zeta, a^*) \in V_j \times W$. \square

Proof of Theorem 4.3. Without loss of generality, we can assume $\epsilon > 0$ is small enough that the points $P_1, P_2, Ja_1, \dots, Ja_4$ are all distinct. Define

$$R(a_1, a_2) = \begin{pmatrix} R_1(a_1) & R_1(a_2) \\ R_2(a_1) & R_2(a_2) \end{pmatrix}$$

and

$$R(a_3, a_4) = \begin{pmatrix} R_1(a_3) & R_1(a_4) \\ R_2(a_3) & R_2(a_4) \end{pmatrix}$$

where $R_j(a)$ is the residue of $K^0(\zeta, a)$ at P_j as in Lemma 4.4.

To prove that $R(a_1, a_2)$ is invertible, let

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

and $f_c = c_1K^0(\zeta, a_1) + c_2K^0(\zeta, a_2)$. Note that $R(a_1, a_2)c = 0$ if and only if f_c does not have poles at either P_1 or P_2 . In this case, if f_c is not constant, then the poles of f_c are precisely the points Ja_1 and Ja_2 , which gives the usual contradiction, since both of these points are in R' . Thus, f_c is constant. The kernel $K^0(\zeta, 0) = 1$, so we can express this as

$$0 = c_0K^0(\zeta, 0) + c_1K^0(\zeta, a_1) + c_2K^0(\zeta, a_2).$$

Since $0, a_1, a_2$ are distinct, the functions $K^0(\zeta, 0), K^0(\zeta, a_1), K^0(\zeta, a_2)$ are linearly independent and hence $c_1 = c_2 = 0$. Summarizing, if $R(a_1, a_2)c = 0$, then $c = 0$.

It follows that $R(a_1, a_2)$ is invertible and by an identical argument, $R(a_3, a_4)$ is invertible.

Consider the function F defined for δ_j near δ_j^0 by

$$F = \begin{pmatrix} R_1(a_1)\delta_1 & R_1(a_2)\delta_2 & R_1(a_3)\delta_3 & R_1(a_4)\delta_4 \\ R_2(a_1)\delta_1 & R_2(a_2)\delta_2 & R_2(a_3)\delta_3 & R_2(a_4)\delta_4 \end{pmatrix}.$$

Thus, F takes values in M_4 , the 4×4 matrices, viewed as 2×4 matrices with entries from \mathbb{C}^2 . Clearly F is continuous in δ_j . By Lemma 4.4 it is also continuous in a_1, \dots, a_4 . Indeed, from the form of F it is jointly continuous in a_j, δ_j . Since F is invertible at a_j^0, δ_j^0 , it follows that there is an $\epsilon > 0$ so that if $|a_j^0 - a_j|, \|\delta_j^0 - \delta_j\| < \epsilon$, then F is invertible.

If a_j and δ_j are chosen such that F is invertible and

$$h(\zeta) = \sum c_j K^0(\zeta, a_j)\delta_j + v$$

does not have poles at P_j , then

$$\begin{aligned} 0 &= \begin{pmatrix} \sum c_j R_1(a_j)\delta_j \\ \sum c_j R_2(a_j)\delta_j \end{pmatrix} \\ &= F \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}. \end{aligned}$$

Hence $c = 0$ and h is constant.

Now suppose $h \neq 0$ and there exist $z_1, z_2 \in B$ (not assumed distinct but not both zero) such that

$$(4.4) \quad h(\zeta)K^0(\zeta, z_k) = \sum c_j^k K^0(\zeta, a_j)\delta_j + v_k, \quad k = 0, 1, 2,$$

where $z_0 = 0$ (and so $K^0(\zeta, z_0) = 1$). Using $k = 1$ we see that P_1, P_2 are not poles of h , since by the assumptions on the distinctness of the P_k 's and a_j 's, the right side has a pole of order at most one at each P_k , while the left side has a pole of order at least one at these points. Using $k = 0$ we see that h satisfies the hypothesis of the part of the theorem which has already been proved. Thus h is constant.

The rest of the result now easily follows using the linear independence of the kernels. □

5. REPRESENTING NICE MATRIX-VALUED INNER FUNCTIONS

5.1. Hahn-Banach separation. Recall that $\mathbb{H}(X)$ denotes the set of functions analytic in a neighborhood of X and $\mathcal{R}(X)$ the rational functions with poles off X . Of particular interest is the set $B\mathbb{H}(X)$ consisting of those $f \in \mathbb{H}(X)$ with $\|f\|_R \leq 1$.

Let $M_2(\mathbb{H}(X))$ denote the 2×2 matrices with entries from $\mathbb{H}(X)$ and similarly define $M_2(\mathcal{R}(X))$. For $f \in \mathbb{H}(X)$, f^* denotes its pointwise complex conjugate while for $F \in M_2(\mathbb{H}(X))$, F^* is the pointwise adjoint.

For $f, g \in \mathbb{H}(X)$ or $M_2(\mathbb{H}(X))$ and $h(z, w) = \sum f(z)g(w)^*$ (with only finitely many terms) we use the convention $\sum f(z)g(w)^*(T) = \sum f(T)g(T)^*$ to define $h(T) = h(T, T^*)$.

Let \mathcal{C} be the cone generated by

$$\{H(z)(1 - \psi(z)\psi(w)^*)H(w)^* : \psi \in B\mathbb{H}(X), H \in M_2(\mathbb{H}(X))\}.$$

Obviously we would get the same set if we were to instead assume that H is a \mathbb{C}^2 -valued function.

Lemma 5.1. *If $F \in M_2(\mathbb{H}(X))$, then there exists $\rho > 0$ such that $I - \rho^2 F(z)F(w)^* \in \mathcal{C}$.*

Proof. First, suppose

$$F = \begin{pmatrix} f \\ g \end{pmatrix}.$$

Choose $0 < \tau$ so large that $\frac{f}{\tau}$ and $\frac{g}{\tau}$ are in $B\mathbb{H}(X)$. Then,

$$\begin{aligned} 2\tau^2 I - F(z)F(w)^* &= 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} (\tau^2 - f(z)f(w)^*) \begin{pmatrix} 1 & 0 \end{pmatrix} \\ &+ 2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} (\tau_g^2 - g(z)g(w)^*) \begin{pmatrix} 0 & 1 \end{pmatrix} + \begin{pmatrix} f(z) \\ -g(z) \end{pmatrix} (1 - 0) \begin{pmatrix} f(w)^* & g(w)^* \end{pmatrix}. \end{aligned}$$

Thus $\rho = \frac{1}{\sqrt{2\tau}}$ satisfies the conclusion of the lemma.

For general F write $F(z)F(w)^* = G(z)G(w)^* + H(z)H(w)^*$, where G and H are the first and second columns of F respectively. There exist ρ_G and ρ_H so that both $I - \rho_G^2 G(z)G(w)^*$ and $I - \rho_H^2 H(z)H(w)^*$ are in \mathcal{C} . With $\rho^2 = \frac{1}{2} \min\{\rho_G^2, \rho_H^2\}$,

$$\begin{aligned} I - \rho^2 F(z)F(w)^* &= I - \rho^2 G(z)G(w)^* - \rho^2 H(z)H(w)^* \\ &= \frac{1}{2}(I - \rho_G^2 G(z)G(w)^*) + \frac{1}{2}(I - \rho_H^2 H(z)H(w)^*) \\ &+ \left(\frac{1}{2}\rho_G^2 + \frac{1}{2}\rho_H^2 - \rho^2\right). \end{aligned}$$

Each term on the right-hand side is evidently in \mathcal{C} . This completes the proof. \square

Henceforth, for $F \in M_2(\mathbb{H}(X))$, we set

$$\rho_F = \sup\{\rho > 0 : I - \rho^2 F(z)F(w)^* \in \mathcal{C}\}.$$

The following theorem is an application of the Hahn-Banach theorem. It is central to our construction.

Theorem 5.2. *If there exists a function $F : R \rightarrow M_2(\mathbb{C})$ which is analytic in a neighborhood of X and unitary-valued on B such that $\rho_F < 1$, then there exists a Hilbert space \mathcal{H} and an operator $T \in \mathcal{B}(\mathcal{H})$ having X as a spectral set, but not having a normal B -dilation.*

Proof. The proof features a familiar Hahn-Banach separation argument and GNS construction.

From the hypotheses, there exists a $\rho < 1$ so that $I - \rho^2 F(z)F(w)^* \notin \mathcal{C}$.

Let \mathcal{P} be the vector space of finite sums

$$\sum h_j(z)g_j(w)^*$$

where h_j, g_j are \mathbb{C}^2 -valued functions analytic in a neighborhood of X . Note, $h(z)g(w)^*$ is pointwise a 2×2 matrix.

The cone \mathcal{C} is a convex subset of \mathcal{P} not containing $I - \rho^2 F(z)F(w)^*$ and by Lemma 5.1, I is an internal point of \mathcal{C} . Hence, there exists a nonconstant linear functional $\lambda : \mathcal{P} \rightarrow \mathbb{C}$ so that $\lambda \geq 0$ on \mathcal{C} and $\lambda(I - \rho^2 F(z)F(w)^*) \leq 0$. We have $\lambda(I) > 0$, as otherwise, from Lemma 5.1, $\lambda(H(z)H(w)^*) = 0$ for all H and hence $\lambda = 0$ (see, for example, Holmes [23], § 11.E).

Let $\mathbb{H}_2(X)$ denote the \mathbb{C}^2 -valued functions analytic in a neighborhood of X . For $h, g \in \mathbb{H}_2(X)$, define

$$[h, g] = \lambda(h(z)g(w)^*).$$

Since the cone \mathcal{C} contains $h(z)h(w)^*$, the form $[\cdot, \cdot]$ is positive semidefinite on $\mathbb{H}_2(X)$.

Given f analytic in a neighborhood of X , consider the mapping $M_f : \mathbb{H}_2(X) \rightarrow \mathbb{H}_2(X)$ defined by multiplication by f so that $M_f g = fg$. Let C_f be the infimum over all positive numbers such that $\frac{f}{C_f}$ takes values in the closed unit disk. Obviously, $C_f = \|f\|_R$. For any $g \in \mathbb{C}^2$,

$$(C_f^2 g(z)) \left(1 - \frac{f(z)}{C_f} \frac{f(w)^*}{C_f}\right) (C_f g(w))^* = g(z)(C_f^2 - f(z)f(w)^*)g(w)^* \in \mathcal{C}.$$

Thus,

$$\begin{aligned} C_f^2 [g, g] - [M_f g, M_f g] &= C_f^2 \lambda(g(z)g(w)^*) - \lambda(f(z)g(z)g(w)^* f(w)) \\ &= \lambda(g(z) (C_f^2 - f(z)f(w)^*) g(w)^*) \geq 0, \end{aligned}$$

as λ is nonnegative on \mathcal{C} . It follows that each M_f defines a bounded operator, still denoted by M_f , on the Hilbert space \mathcal{H} obtained from $\mathbb{H}_2(X)$ by modding out $[\cdot, \cdot]$ -null vectors and completing. Furthermore, $\|M_f\| = C_f$. In particular, with $T = M_\zeta$, where $\zeta(z) = z$ is the identity function, the set X is a spectral set for T . Here we are using $f(T) = M_f$.

To see that T does not have a dilation to a normal operator with spectrum in X , it suffices to show that $F(T)$ is not a contraction for the F in the statement of the theorem. To this end, write

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}$$

so that

$$F^t(T) = \begin{pmatrix} M_{F_{11}} & M_{F_{21}} \\ M_{F_{12}} & M_{F_{22}} \end{pmatrix},$$

where F^t denotes the pointwise transpose of F . Let e_1, e_2 denote the (class of) constant functions, $e_j(z) = e_j$, as elements of \mathcal{H} . Compute

$$\begin{aligned} \left\langle F^t(T) \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, F^t(T) \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \right\rangle &= \left\langle \begin{pmatrix} F_{11} \\ F_{21} \\ F_{12} \\ F_{22} \end{pmatrix}, \begin{pmatrix} F_{11} \\ F_{21} \\ F_{12} \\ F_{22} \end{pmatrix} \right\rangle \\ &= \lambda \left(\begin{pmatrix} F_{11}(z) \\ F_{21}(z) \end{pmatrix} (F_{11}(w)^* \quad F_{21}(w)^*) + \begin{pmatrix} F_{12}(z) \\ F_{22}(z) \end{pmatrix} (F_{12}(w)^* \quad F_{22}(w)^*) \right) \\ &= \lambda(F(z)F(w)^*). \end{aligned}$$

On the other hand,

$$\left\langle \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \right\rangle = \lambda(e_1 e_1^*) + \lambda(e_2 e_2^*) = \lambda(I).$$

Combining the last two equalities gives

$$\begin{aligned} \left\langle (I - F^t(T)^* F^t(T)) \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}, \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \right\rangle &= \lambda(I - F(z)F(w)^*) \\ &= \lambda \left(\frac{1}{\rho^2} - F(z)F(w)^* \right) - \left(\frac{1}{\rho^2} - 1 \right) \lambda(I) < 0. \end{aligned}$$

Therefore, $\|F^t(T)\| > 1$. On the other hand, $\|F^t\|_\infty = \sup\{\|F^t(z)\| : z \in X\}$ is the same as $\|F\|_\infty$, since the norms of a matrix and its transpose are the same. In fact, as F is unitary valued on B , so is F^t and thus $\|F^t\|_\infty = 1$. It now follows from Lemma 2.1 that T does not dilate to a normal operator with spectrum in X . \square

5.2. Matrix measures. This subsection is a brief digression from the main line of development to collect some needed facts about matrix-valued measures.

Given a compact Hausdorff space X , an $m \times m$ matrix-valued measure

$$\mu = (\mu_{j,\ell})_{j,\ell=1}^m$$

on X is an $m \times m$ matrix whose entries $\mu_{j,\ell}$ are complex-valued regular Borel measures on X . The measure μ is positive, written $\mu \geq 0$, if, for each continuous function $f : X \rightarrow \mathbb{C}^m$,

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$$

we have

$$0 \leq \int_X f^* d\mu f = \sum_{j,\ell} \int_X f_j^* f_\ell d\mu_{j,\ell}.$$

The positive measure μ is bounded by $C > 0$ if

$$CI_m - (\mu_{j,\ell}(X)) \geq 0$$

is positive semidefinite, where I_m is the identity $m \times m$ matrix.

Lemma 5.3. *The $m \times m$ matrix-valued measure μ is positive if and only if for each Borel set ω the $m \times m$ matrix*

$$(\mu_{j,\ell}(\omega))$$

is positive semi-definite.

Furthermore, if there is a κ so that each diagonal entry $\mu_{j,j}(X) \leq \kappa$, then each entry $\mu_{j,\ell}$ of μ has total variation at most κ . In particular, if μ is bounded by C , then each entry has variation at most C .

Proof. First suppose μ is positive. Let $C(X)$ denote the continuous complex-valued functions on X . Fix a vector $c \in \mathbb{C}^m$. Given $f \in C(X)$, the function cf is a continuous \mathbb{C}^m -valued function. Thus, $\Phi_c : C(X) \rightarrow \mathbb{C}$ given by

$$\Phi_c(f) = \sum \int_X c_j^* c_\ell f d\mu_{j,\ell}$$

is a positive linear functional. Hence $\sum c_j^* c_\ell \mu_{j,\ell}$ is a positive measure on X . If ω is a Borel set, then

$$\langle (\mu_{j,\ell}(\omega)) c, c \rangle = \sum c_j^* c_\ell \mu_{j,\ell}(\omega) > 0.$$

Since c was arbitrary, the matrix in the lemma is positive semidefinite for each ω .

Conversely, suppose the matrix in the lemma is positive semidefinite for each Borel set ω . If $f : X \rightarrow \mathbb{C}^m$ is a measurable simple function, $f = \sum v_j \chi_{\omega_j}$, then

$$\int_X f^* d\mu f = \sum_j \langle \mu(\omega_j) v_j, v_j \rangle \geq 0.$$

For more general $f : X \rightarrow \mathbb{C}^m$, choose a sequence of measurable simple functions converging to f pointwise. Then,

$$0 \leq \int_X f_n^* d\mu f_n \rightarrow \int_X f^* d\mu f.$$

Now suppose that each diagonal entry $\mu_{j,j}(X) \leq \kappa$. For $g \in C(X)$, let $\mu_{j,\ell}(g)$ denote the integral of g with respect to $d\mu_{j,\ell}$, and for $j \neq \ell$ let

$$\lambda = -\frac{\mu_{j,\ell}(g)^*}{|\mu_{j,\ell}(g)^*|} \|g\|_\infty.$$

Observe

$$\begin{aligned} 0 &\leq \left\langle \begin{pmatrix} \mu_{j,j} & \mu_{j,\ell} \\ \mu_{\ell,j} & \mu_{\ell,\ell} \end{pmatrix} \begin{pmatrix} \lambda \\ g \end{pmatrix}, \begin{pmatrix} \lambda \\ g \end{pmatrix} \right\rangle \\ &= \|g\|_\infty^2 \mu_{j,j}(X) + \mu_{\ell,\ell}(|g|^2) - 2|\mu_{j,\ell}(g)| \|g\|_\infty \\ &\leq 2(\|g\|_\infty^2 \kappa - |\mu_{j,\ell}(g)| \|g\|_\infty). \end{aligned}$$

It follows that $\mu_{j,\ell}$ is a continuous linear functional on $C(X)$ with norm at most κ . In particular, the variation of $\mu_{j,\ell}$ is at most κ .

Finally, by choosing $c = e_j$, where e_j is the j -th standard basis vector for \mathbb{C}^m , it follows that each $\mu_{j,j}$ is a positive measure. If we now suppose μ is bounded by C , then $\mu_{j,j}(X) \leq C$. □

Lemma 5.4. *If μ^n is a sequence of positive $m \times m$ matrix-valued measures on X which are all bounded above by C , then there is a positive $m \times m$ matrix-valued measure μ on X also bounded above by C . Hence there is a subsequence μ^{n_k} of μ^n converging to μ weak-*, i.e., for each pair of continuous functions $f, g : X \rightarrow \mathbb{C}^m$,*

$$\sum_{j,\ell} \int_X f_\ell g_j^* d\mu_{j,\ell}^n \rightarrow \sum_{j,\ell} \int_X f_\ell g_j^* d\mu_{j,\ell}.$$

Proof. By the previous lemma, for each j, ℓ the measures $\mu_{j,\ell}^n$ are bounded in variation by C . Hence, we can find a subsequence, which for convenience, we will still denote by μ^n so that each $\mu_{j,\ell}^n$ converges weak-* to some $\mu_{j,\ell}$ with variation at most C .

Suppose $f : X \rightarrow \mathbb{C}^m$ is continuous. We have

$$0 \leq \int_X f^* d\mu^n f = \sum_{j,\ell} \int_X f_j^* f_\ell d\mu_{j,\ell}^n \rightarrow \sum_{j,\ell} \int_X f_j^* f_\ell d\mu_{j,\ell} = \int_X f^* d\mu f.$$

Hence μ is a positive measure. Furthermore, for a vector $c \in \mathbb{C}^m$ thought of as a constant function,

$$0 \leq C\|c\|^2 - \langle (\mu_{j,\ell}^n(X)) c, c \rangle \rightarrow C\|c\|^2 - \langle (\mu_{j,\ell}(X)) c, c \rangle.$$

Thus, μ is bounded above by C . □

Lemma 5.5. *If μ is a positive $m \times m$ matrix-valued measure on X , then the diagonal entries, $\mu_{j,j}$ are positive measures. Furthermore, with $\nu = \sum_{j=1}^m \mu_{j,j}$, there exists an $m \times m$ matrix-valued function $\Delta : X \rightarrow M_m(\mathbb{C})$ so that $\Delta(x)$ is positive semidefinite for each $x \in X$ and $d\mu = \Delta d\nu$; i.e., for each pair of continuous functions $f, g : X \rightarrow \mathbb{C}^m$,*

$$\sum_{j,\ell} \int_X g_j^* f_\ell d\mu_{j,\ell} = \sum_{j,\ell} \int_X g_j^* \Delta_{j,\ell} f_\ell d\nu.$$

Proof. From Lemma 5.3, if ω is a Borel set and $\mu_{j,j}(\omega) = 0$, then $\mu_{j,\ell}(\omega) = 0$ for each j . Thus, for each j, ℓ the measure $\mu_{j,\ell}$ is absolutely continuous with respect to ν . By the Radon-Nikodým Theorem, there exists ν integrable functions $\Delta_{j,\ell}$ so that $d\mu_{j,\ell} = \Delta_{j,\ell} d\nu$.

Once again fix a vector $c \in \mathbb{C}^m$. By Lemma 5.3, for each Borel set ω ,

$$0 \leq \langle (\mu_{j,\ell}(\omega)) c, c \rangle = \left\langle \left(\int_{\omega} \Delta_{j,\ell} d\nu \right) c, c \right\rangle = \int_{\omega} \sum_{j,\ell} c_j^* c_{\ell} \Delta_{j,\ell} d\nu.$$

Thus, $\sum_{j,\ell} c_j^* c_{\ell} \Delta_{j,\ell} \geq 0$ almost everywhere with respect to ν .

Choose a countable dense subset $\{c^n\}$ of \mathbb{C}^m . For each n there is a set E_n of ν -measure zero such that off of E_n the function $\sum_{j,\ell} (c_j^n)^* c_{\ell}^n \Delta_{j,\ell}$ is nonnegative. We have $\sum_{j,\ell} (c_j^n)^* c_{\ell}^n \Delta_{j,\ell}(x) \geq 0$ for each n and for $x \in X \setminus (\bigcup E_n)$. By continuity of the inner product in \mathbb{C}^m , it follows that $\sum_{j,\ell} c_j^* c_{\ell} \Delta_{j,\ell}(x) \geq 0$ for all $c \in \mathbb{C}^m$ for almost all x ; that is, the matrices $\Delta(x) = (\Delta_{j,\ell}(x))$, for $x \in X \setminus (\bigcup E_n)$, are positive semidefinite. \square

5.3. Representations in terms of the ϕ_p . The following is a companion to Proposition 5.2. Given a subset $S \subset R$, a function $\Gamma : S \times S \times \Pi \rightarrow \mathbb{C}$ is a positive kernel if for each $p \in \Pi$, the matrix

$$(5.1) \quad (\Gamma(z, w; p))_{z,w \in S'},$$

S' a finite subset of S . If $S = R$, then Γ is analytic if $\Gamma(z, w; p)$ is analytic in z and conjugate analytic in w for all p .

The point of the next proposition is that we do not know *a priori* that $1 - FF^* \in \mathcal{C}$ even if $\rho_F = 1$.

Proposition 5.6. *Suppose F is a 2×2 matrix-valued function analytic in a neighborhood of R , F is unitary-valued on B , and $F(0) = 0$. If $\rho_F = 1$ and if $S \subset R$ is a finite set, then there exists a probability measure μ on Π and a positive kernel $\Gamma : S \times S \times \Pi \rightarrow \mathbb{C}$ so that*

$$1 - F(z)F(w)^* = \int_{\Pi} (1 - \phi_p(z)\phi_p(w)^*)\Gamma(z, w; p) d\mu(p)$$

for all $z, w \in S$.

Proof. Choose a sequence $0 < \rho_n < 1$ such that ρ_n converges to 1. For each n , there exist vector functions $H_{n,j}$ and functions $\psi_{n,j}$ analytic in a neighborhood of R such that $\psi_{n,j} \in B\mathbb{H}(X)$ and

$$1 - \rho_n^2 F(z)F(w)^* = \sum_{j=1}^{N_n} H_{n,j}(z)(1 - \psi_{n,j}(z)\psi_{n,j}(w)^*)H_{n,j}(w)^*.$$

By post composition with a Möbius transformation if necessary, it may be assumed without loss of generality that $\psi_{n,j}(0) = 0$ for all n, j .

For each n, j , there exist $h_{n,j}(z, p)$, analytic as a function of z in a neighborhood of R , and a probability measure $\nu_{n,j}$ on Π so that by Proposition 2.4,

$$1 - \psi_{n,j}(z)\psi_{n,j}(w)^* = \int_{\Pi} h_{n,j}(z, p)(1 - \phi_p(z)\phi_p(w)^*)h_{n,j}(w, p)^* d\nu_{n,j}(p).$$

Observe, as all the ψ and ϕ vanish at 0,

$$1 = \int h_{n,j}(0, p)h_{n,j}(0, p)^* d\nu_{n,j}(p).$$

Let

$$\nu_n = \sum_{j=1}^{N_n} \nu_{n,j}.$$

By the Radon-Nikodým theorem, there exists a nonnegative function $u_{n,j}(p)$ such that

$$d\nu_{j,n} = u_{n,j}(p)^2 d\nu_n.$$

Thus

$$1 - \psi_{n,j}(z)\psi_{n,j}(w)^* = \int_{\Pi} u_{n,j}(p)h_{n,j}(z,p)[1 - \phi_p(z)\phi_p(w)^*]h_{n,j}(w,p)^*u_{n,j}(p) d\nu_n(p).$$

Let

$$\Gamma_n(z, w; p) = \sum_j H_{j,n}(z) u_{n,j}(p) h_{n,j}(z, p) h_{n,j}(w, p)^* u_{n,j}(p)^* H_{n,j}(w)^*.$$

By construction, Γ_n is analytic in z and conjugate analytic in w in a neighborhood of R , is positive semidefinite as a kernel, and

$$(5.2) \quad I - \rho_n^2 F(z)F(w)^* = \int_{\Pi} [1 - \phi_p(z)\phi_p(w)^*]\Gamma_n(z, w; p) d\nu_n(p).$$

For fixed z , $I - \rho_n^2 F(z)F(z)^* \leq I$ and there exists an ϵ_z such that $1 - |\phi_p(z)|^2 \geq \epsilon_z$. Thus,

$$I \geq \epsilon_z \int_{\Pi} \Gamma_n(z, z; p) d\nu_n(p).$$

Let C denote the maximum of the set $\{\epsilon_z^{-1} : z \in S\}$ and m the cardinality of S . The sequence of $m \times m$ measures with 2×2 entries

$$d\mu_n = (\Gamma_n(z, w; p) d\nu_n(p))_{z,w \in S}$$

are positive and the diagonal entries are bounded by C . A positive $k \times k$ matrix whose diagonal entries are at most C is bounded above by kCI_k . It then follows from the results of the previous section that there exists a positive measure ν and a pointwise positive definite matrix-valued function

$$\Gamma(p) = (\Gamma(z, w; p))_{z,w \in S}$$

so that some subsequence of μ_n converges in the weak-* topology to $\Gamma d\nu$, where ν can be taken to be a probability measure by scaling Γ if necessary. For notational ease, we continue to denote the subsequence by μ_n .

For $z, w \in S$ fixed, the expression $(1 - \phi_p(z)\phi_p(w)^*)$ is continuous in p by Lemma 2.11. Thus, letting n tend to infinity in equation (5.2) gives

$$I - F(z)F(w)^* = \int_{\Pi} [(1 - \phi_p(z)\phi_p(w)^*)\Gamma(z, w; p) d\nu(p).$$

□

5.4. Transfer function representations. For present purposes, a unitary colligation $\Sigma = (U, K, \mu)$ consists of a probability measure μ on Π , a Hilbert space K , and a unitary operator U on the direct sum $(L^2(\mu) \otimes K) \oplus \mathbb{C}^2$, written as

$$(5.3) \quad U = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

with respect to the direct sum decomposition. Here $L^2(\mu) \otimes K$ signifies K -valued $L^2(\mu)$.

Define $\Phi : R \rightarrow B(L^2(\mu) \otimes K)$ by $(\Phi(z)f)(p) = \phi_p(z)f(p)$. Of course Φ depends upon μ and K , but this dependence is suppressed. The transfer function associated to Σ is

$$(5.4) \quad W = W_\Sigma(z) = D + C\Phi(z)(I - A\Phi(z))^{-1}B.$$

Note that as A must be a contraction and $\Phi(z)$ is a strict contraction, the inverse in (5.4) exists for $z \in R$. Moreover, since $\Phi(z)(I - A\Phi(z))^{-1} = (I - \Phi(z)A)^{-1}\Phi(z)$, the transfer function of Σ may also be expressed as

$$(5.5) \quad W = D + C(I - \Phi(z)A)^{-1}\Phi(z)B.$$

Proposition 5.7. *The transfer function is contraction-valued; that is, $\|W_\Sigma(z)\| \leq 1$ for all $z \in R$. Indeed, for $z, w \in R$,*

$$I - W_\Sigma(z)W_\Sigma(w)^* = C(I - \Phi(z)A)^{-1}(I - \Phi(z)\Phi(w)^*)(I - \Phi(w)A)^{-1}C^*.$$

By now the proof is entirely standard. Simply use the equation (5.4), and $DB^* = -CA^*$, $BB^* = I - AA^*$, and $DD^* = I - CC^*$ to verify:

$$\begin{aligned} I - W(z)W(w)^* &= 1 - [D + C\Phi(z)(I - A\Phi(z))^{-1}B][D^* + B^*(I - \Phi^*(w)A)^{-1}\Phi^*(w)C] \\ &= 1 - DD^* - [C\Phi(z)(I - A\Phi(z))^{-1}B][B^*(I - \Phi^*(w)A^*)^{-1}\Phi^*(w)C] \\ &\quad - DB^*(I - \Phi^*(w)A^*)^{-1}\Phi^*(w)C^* - C\Phi(z)(I - A\Phi(z))^{-1}BD^* \\ &= C[1 - \Phi(z)(I - A\Phi(z))^{-1}(1 - AA^*)(I - \Phi^*(w)A^*)^{-1}\Phi^*(w) \\ &\quad + A^*(I - \Phi^*(w)A^*)^{-1}\Phi^*(w) + \Phi(z)(I - A\Phi(z))^{-1}A]C^* \\ &= C[1 - (I - \Phi(z)A)^{-1}\Phi(z)(1 - AA^*)\Phi^*(w)(I - A^*\Phi^*(w))^{-1} \\ &\quad + A^*\Phi^*(w)(I - A^*\Phi^*(w))^{-1} + (I - \Phi(z)A)^{-1}\Phi(z)A]C^* \\ &= C(I - \Phi(z)A)^{-1}[(I - \Phi(z)A)(I - A^*\Phi^*(w)) - \Phi(z)(1 - AA^*)\Phi^*(w) \\ &\quad + (I - \Phi(z)A)A^*\Phi^*(w) + \Phi(z)A(I - A^*\Phi^*(w))] (I - A^*\Phi^*(w))^{-1}C^* \\ &= C(I - \Phi(z)A)^{-1}[1 - \Phi(z)\Phi^*(w)](I - A^*\Phi^*(w))^{-1}C^*. \end{aligned}$$

Note that pointwise on R , we can define $H(w) = (I - A^*\Phi(w)^*)^{-1}C^* : \mathbb{C}^2 \rightarrow L^2(\mu) \otimes K$. Thus, for w fixed, $H(w)^*$ is a function on Π , which we emphasize by writing as $H_p(w)^*$. Interpreting the representation in Proposition 5.7 in terms of the space $L^2(\mu) \otimes K$ gives

$$I - W(z)W(w)^* = \int (1 - \phi_p(z)\phi_p(w)^*)H_p(z)H_p(w)^* d\mu(p).$$

5.5. Nevanlinna-Pick interpolation. The following proposition is an Agler-Pick type interpolation theorem for some matrix-valued functions on R . The proof proceeds via a transfer function realization for a solution. The now standard methods are purloined from those pioneered by Agler, and followed by Ball and others. We eventually show that, roughly speaking, this theorem only applies to 2×2 matrix functions which are, up to a fixed unitary, the direct sum of scalar contractive functions.

Proposition 5.8. *If $S \subset R$ is a finite set, $W : S \rightarrow M_2(\mathbb{C})$, and if there is a positive kernel $\Gamma : S \times S \times \Pi \rightarrow M_2(\mathbb{C})$ such that*

$$I - W(z)W(w)^* = \int_{\Pi} (1 - \phi_p(z)\phi_p(w)^*)\Gamma(z, w; p) d\mu(p)$$

for all $z, w \in S$, then there exists $G : R \rightarrow M_2(\mathbb{C})$ such that $\|G(z)\| \leq 1$ and $G(z) = W(z)$ for $z \in S$. Indeed, there exists a finite-dimensional Hilbert space K (dimension at most twice the cardinality of S) and a unitary colligation $\Sigma = (U, K, \mu)$ so that

$$G = W_{\Sigma},$$

and hence there exists $\Delta : R \times R \times \Pi \rightarrow M_2(\mathbb{C})$ a positive analytic kernel such that

$$I - G(z)G(w)^* = \int_{\Pi} [1 - \phi_p(z)\phi_p(w)^*]\Delta(z, w; p) d\mu(p)$$

for all $z, w \in R$.

Proof. The proof employs a conventional lurking isometry argument. For $p \in \Pi$, the rank of the block matrix with 2×2 matrix entries

$$(\Gamma(z, w; p))_{z, w \in S}$$

is at most $2N$, where N is the cardinality of S . Thus, by Kolmogorov’s theorem (see, for example, [7], Theorem 2.53, especially the second proof), there exists a Hilbert space K of dimension $2N$ and a function $H : S \rightarrow L^2(\mu) \otimes B(\mathbb{C}^2, K)$, denoted $H_p(z)$, such that $\Gamma(z, w; p) = H_p(z)H_p(w)^*$ (μ almost everywhere).

Let \mathcal{E} and \mathcal{F} denote the subspaces of $(L^2(\mu) \otimes K) \oplus \mathbb{C}^2$ spanned by

$$\left\{ \begin{pmatrix} H_s(w)^*x \\ W(w)^*x \end{pmatrix} : x \in \mathbb{C}^2, w \in S \right\}$$

and

$$\left\{ \begin{pmatrix} \phi_s(w)^*H_s(w)^*x \\ x \end{pmatrix} : x \in \mathbb{C}^2, w \in S \right\}$$

respectively. The mapping V from \mathcal{E} to \mathcal{F} determined by

$$V \begin{pmatrix} H_s(w)^*x \\ W(w)^*x \end{pmatrix} = \begin{pmatrix} \phi_s(w)^*H_s(w)^*x \\ x \end{pmatrix}$$

is an isometry since

$$\left\langle \begin{pmatrix} H_s(w)^*x \\ W(w)^*x \end{pmatrix}, \begin{pmatrix} H_s(z)^*y \\ W(z)^*y \end{pmatrix} \right\rangle = \left\langle \int H_s(z)H_s(w)^* d\mu(s) x, y \right\rangle + \langle W(z)W(w)^*x, y \rangle$$

and

$$\begin{aligned} & \left\langle \left(\begin{array}{c} \phi_s(w)^* H_s(w)^* x \\ x \end{array} \right), \left(\begin{array}{c} \phi_s(z)^* H_s(z)^* y \\ y \end{array} \right) \right\rangle \\ &= \left\langle \int \phi_s(z) \phi_s(w)^* H_s(z) H_s(w)^* d\mu(s) x, y \right\rangle + \langle x, y \rangle. \end{aligned}$$

Both \mathcal{E} and \mathcal{F} are finite dimensional, so there exists a unitary U on $(L^2(\mu) \otimes K) \oplus \mathbb{C}^2$ such that U^* extends V . Let $\Sigma = (U, K, \mu)$ denote the resulting unitary colligation.

Write U as in equation (5.3). Since U^* restricted to \mathcal{E} is V ,

$$\begin{pmatrix} A^* & C^* \\ B^* & D^* \end{pmatrix} \begin{pmatrix} \phi_s(w)^* H_s(w)^* x \\ x \end{pmatrix} = \begin{pmatrix} H_s(w)^* x \\ W(w)^* x \end{pmatrix}.$$

Expressing this as a system of equations we have

$$\begin{aligned} A^* \phi_s(w)^* H_s(w)^* x + C^* x &= H_s(w)^* x, \\ B^* \phi_s(w)^* H_s(w)^* x + D^* x &= W(w)^* x. \end{aligned}$$

Solving the first equation for $H_s(w)^* x$ gives

$$H_s(w)^* x = (I - A^* \Phi(w)^*)^{-1} C^* x.$$

Substituting into this the second equation now gives

$$B^* \Phi(w)^* (I - A^* \Phi(w)^*)^{-1} C^* x = W(w)^* x.$$

It follows that for each $z \in S$,

$$W_\Sigma(z) = W(z).$$

□

Next is a uniqueness result for Nevanlinna-Pick interpolation on R .

Proposition 5.9. *Suppose $F : R \rightarrow M_2(\mathbb{C})$ is analytic in a neighborhood of X , unitary on B , and with a standard zero set. Then there exists $S \subset R$ a set with seven elements such that, if $Z : R \rightarrow M_2(\mathbb{C})$ is contraction-valued and $Z(z) = F(z)$ for $z \in S$, then $Z = F$.*

Proof. Let K^0 denote the Fay kernel for R defined in Section 4. That is, K^0 is the reproducing kernel for the Hilbert space $\mathbb{H}^2(R)$ of functions analytic in R with L^2 boundary values with respect to harmonic measure on B relative to the point 0. Let $\mathbb{H}_2^2(R)$ denote \mathbb{C}^2 -valued $\mathbb{H}^2(R)$. Since F is unitary-valued on B , the mapping V on $\mathbb{H}_2^2(R)$ defined by $VG(z) = F(z)G(z)$ is an isometry. As is shown below, the kernel of V^* is the span of $\{K^0(\cdot, a_j)\gamma_j : j = 1, 2, \dots, 6\}$ where $F(a_j)^*\gamma_j = 0$ and, of course, $\gamma_j \neq 0$; that is, the pair (a_j, γ_j) is a zero of F^* .

Before proceeding, we note that if φ is scalar-valued and analytic in a neighborhood of R , has no zeros on B , and has distinct zeros $w_1, \dots, w_n \in R$ of multiplicity one, and if $f \in \mathbb{H}^2(R)$ with $f(w_j) = 0$, then $f = \varphi g$ for a $g \in \mathbb{H}^2(R)$.

Now suppose $\psi \in \mathbb{H}^2(R)$ and for all $h \in \mathbb{H}^2(R)$, we have $\langle \psi, \varphi h \rangle = 0$. Then there is a linear combination $f = \psi - \sum_1^n c_j K^0(\cdot, w_j)$ so that $f(w_j) = 0$ as the set $\{K^0(\cdot, w_j) : 1 \leq j \leq n\}$ is linearly independent, and thus by the above remark, $f = \varphi g$ for some $g \in \mathbb{H}^2(R)$. Since $\langle K^0(\cdot, w_j), \varphi h \rangle = \varphi(w_j)^* h(w_j)^* = 0$ for each j and h , it follows that $\langle f, \varphi h \rangle = 0$ for all h . Choosing $h = g$ gives $\langle \varphi g, \varphi g \rangle = 0$ from which it follows that $g = 0$. Hence ψ is in the span of $\{K^0(\cdot, w_j) : 1 \leq j \leq n\}$.

This shows that $\{K^0(\cdot, w_j) : 1 \leq j \leq n\}$ is a basis for the orthogonal complement of $\{\varphi h : h \in \mathbb{H}^2(R)\}$.

We next determine the kernel of V^* . Write $a_5 = a_6 = 0$. Since $F(0) = 0$, there is a function H analytic in a neighborhood of X so that $F = zH$. The function $\varphi = z \det(H)$ satisfies the hypothesis of the preceding paragraph.

Let

$$G = \begin{pmatrix} h_{22} & -h_{12} \\ -h_{21} & h_{11} \end{pmatrix},$$

where $H = (h_{j,\ell})$. Verify $FG = zHG = z \det(H)I$, where I is the 2×2 identity matrix.

Now suppose $x \in \mathbb{H}_2^2(R)$ and $V^*x = 0$. Let x_1, x_2 denote the coordinates of x . For each $g \in \mathbb{H}_2^2(R)$,

$$0 = \langle Gg, V^*x \rangle = \langle VGg, x \rangle = \langle z \det(H)g, x \rangle = \langle z \det(H)g_1, x_1 \rangle + \langle z \det(H)g_2, x_2 \rangle.$$

It follows from the discussion above that each x_j is in the span of $\{K^0(\zeta, a_j) : 1 \leq j \leq 6\}$ and therefore x is in the span of $\{K^0(\zeta, a_j)v : 1 \leq j \leq 6, v \in \mathbb{C}^2\}$. In particular, there exist vectors $v_j \in \mathbb{C}^2$ such that

$$x = \sum_1^6 v_j K^0(\cdot, a_j).$$

Since, as is readily verified, $V^*vK^0(\cdot, a) = F(a)^*vK^0(\cdot, a)$ and $F(0)^* = 0$,

$$0 = V^*x = \sum_1^4 F(a_j)^*v_j K^0(\cdot, a_j).$$

But $K^0(\cdot, a_j)$, $j = 1, \dots, 4$ are linearly independent, and so $F(a_j)^*v_j = 0$ for each j . Conversely, if $F(a_j)^*v_j = 0$, then $V^*v_j K^0(\cdot, a_j) = 0$ so that $\{K^0(\zeta, a_j)v_j : 1 \leq j \leq 6\}$ is a basis for the kernel of V^* .

The projection onto the kernel of V^* is $I - VV^*$ and since the kernel of V^* has dimension six, $I - VV^*$ has rank six. Thus, for any finite set $A \subset R$, the block matrix with 2×2 matrix entries

$$\begin{aligned} M_A &= \left(\left(\langle (I - VV^*)K^0(\cdot, w)e_j, K^0(\cdot, z)e_\ell \rangle \right)_{j,\ell=1,2} \right)_{z,w \in A} \\ &= \left((I - F(z)F(w)^*)K^0(z, w) \right)_{z,w \in A} \end{aligned}$$

has rank at most six. (Here $\{e_1, e_2\}$ is the usual basis for \mathbb{C}^2 .) In particular, if $A = \{a_1, \dots, a_6\}$ where the a_j are the zeros of $\det(F)$ (with $a_5 = a_6 = 0$), then M_A has exactly rank six. Choose points a_7, a_8 distinct from a_1, \dots, a_6 so that the set $S = \{a_1, \dots, a_6, a_7, a_8\}$ has exactly seven distinct points. Since S contains A , the rank of M_S is at least six. On the other hand, by what is proved above, it has rank at most six. Thus the rank of M_S is six.

Since the matrix M_S is 14×14 (viewed as a 7×7 block matrix with 2×2 matrix entries), and M_S has rank six, the kernel of M_S has dimension eight. Furthermore,

as the dimension of the subspace

$$\mathcal{L}_1 = \left\{ \left(\begin{pmatrix} \gamma_1 \\ 0 \\ \gamma_2 \\ 0 \\ \vdots \\ \gamma_7 \\ 0 \end{pmatrix} : \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_7 \end{pmatrix} \in \mathbb{C}^7 \right\}$$

of $\mathbb{C}^7 \otimes \mathbb{C}^2$ has dimension seven, it follows that there exists a nonzero $x_1 = y_1 \otimes e_1$ in \mathcal{L} and in the kernel of M_S . Similarly, by considering $\mathcal{L}_2 = \{\gamma \otimes e_2 : \gamma \in \mathbb{C}^7\}$, there exists a nonzero x_2 in the kernel of M_S of the form $x_2 = y_2 \otimes e_2$.

Let $X = (x_1 \ x_2)$; that is, X is a 14×2 matrix,

$$X = \begin{pmatrix} \begin{pmatrix} (y_1)_1 & 0 \\ 0 & (y_2)_1 \end{pmatrix} \\ \begin{pmatrix} (y_1)_2 & 0 \\ 0 & (y_2)_2 \end{pmatrix} \\ \vdots \\ \begin{pmatrix} (y_1)_7 & 0 \\ 0 & (y_2)_7 \end{pmatrix} \end{pmatrix}.$$

It is convenient to use S to index itself, so that we have

$$X(w) = (x_1(w) \ x_2(w)) = \begin{pmatrix} y_1(w) & 0 \\ 0 & y_2(w) \end{pmatrix}$$

for the $w \in S$ coordinate of X . The identity $M_S X = 0$ becomes

$$\sum_{w \in S} K^0(z, w) X(w) = F(z) \sum_{w \in S} K^0(z, w) F(w)^* X(w)$$

for each $z \in S$.

Now suppose $Z : R \rightarrow M_2(\mathbb{C})$ is analytic, contraction-valued, and $Z(z) = F(z)$ for $z \in S$. The operator W of multiplication by Z on $\mathbb{H}_2^2(R)$ is a contraction and

$$W^* K^0(\cdot, w) x = Z(w)^* x K^0(\cdot, w).$$

Given $\zeta \in R$, $\zeta \notin S$, let $S' = S \cup \{\zeta\}$ and consider the decomposition of

$$N_\zeta = \left(\left(\langle (I - Z(z)) Z(w)^* K^0(\cdot, w) e_j, K^0(\cdot, z) e_\ell \rangle \right)_{j, \ell} \right)_{z, w \in S'}$$

into blocks labelled by S and $\{\zeta\}$. Thus, N_ζ is an 8×8 matrix with 2×2 block entries. The upper left 7×7 block matrix, the block determined by S , is M_S since $Z(z) = F(z)$ for $z \in S$. Let

$$Y = \begin{pmatrix} X \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix}.$$

Since N_ζ is positive semi-definite and $M_S X = 0$, it follows that $N_\zeta Y = 0$. An examination of the last two entries (the last 2×2 block) of the product $N_\zeta Y = 0$

gives

$$(5.6) \quad \sum_{w \in S} K^0(\zeta, w)X(w) = Z(\zeta) \sum_{w \in S} Z(w)^* K^0(\zeta, w)X(w).$$

From the form of X and since the kernel functions $\{K^0(\cdot, w) : w \in S\}$ form a linearly independent set in $\mathbb{H}^2(R, \omega_0)$, it follows that the right-hand side in (5.6) has rank two for all but at most countably many ζ . Hence,

$$\sum_{w \in S} K^0(z, w)F(w)^*X(w)$$

also has rank two and Z is completely determined by the identity (5.6). Since F is also determined by this identity, $Z = F$. □

The fact that we used K^0 in the last proof is not significant: any reproducing kernel would have worked. However, certain facts that came to light in the proof will play a role in the proof of Lemma 5.11, where Fay kernels are needed.

5.6. A tight representation. The results in the previous sections now combine to produce a tight representation for $F : R \rightarrow M_2(\mathbb{C})$ which is analytic across the boundary, unitary on the boundary, has at most six simple zeros in R , and has $\rho_F = 1$.

Theorem 5.10. *Suppose F is a 2×2 matrix-valued function analytic in a neighborhood of R , which is unitary-valued on B and with a standard zero set. If $\rho_F = 1$, then there exists a unitary colligation $\Sigma = (U, K, \mu)$ such that $F = W_\Sigma$ and so that the dimension of K is at most 14. In particular, μ is a probability measure on Π , and there is an analytic function $H : R \rightarrow L^2(\mu) \otimes M_{14,2}(\mathbb{C})$, denoted $H_p(z)$ so that*

$$1 - F(z)F(w)^* = \int_{\Pi} (1 - \phi_p(z)\phi_p(w)^*)H_p(z)H_p(w)^* d\mu(p)$$

for all $z, w \in R$.

Proof. Choose, using Proposition 5.9, a finite set S of R with the property that if $G : R \rightarrow M_2(\mathbb{C})$ is analytic, contraction-valued, and $G(z) = F(z)$ for $z \in S$, then $G = F$.

Using this S , Proposition 5.6 produces a probability measure μ and positive kernel $\Gamma : S \times S \times \Pi \rightarrow M_2(\mathbb{C})$ such that

$$1 - F(z)F(w)^* = \int_{\Pi} (1 - \phi_p(z)\phi_p(w)^*)\Gamma(z, w; p) d\mu(p)$$

for all $z, w \in S$.

By Proposition 5.8 there exists a unitary colligation $\Sigma = (U, K, \mu)$ so that $W_\Sigma(z) = F(z)$ for $z \in S$. From our choice of S , we see that $F = W_\Sigma$. The integral representation follows. □

5.7. Fay kernels reprise. Suppose F satisfies the hypothesis of Theorem 5.10 and let μ denote the measure which appears in the conclusion.

Recall Fay's variant of the Szegő kernel for R , $K^0(\zeta, z)$. Also recall, for each $p \in \Pi$ there is an $s \in \{1\} \times B_1 \times B_2$ so that $\phi_s(z)\phi_s(w)^* = \phi_p(z)\phi_p(w)^*$. In this case we label the zeros of ϕ_p (which of course are the same as those of ϕ_s) $z_0^s (= 0), z_1^s, z_2^s$.

Lemma 5.11. *Suppose F satisfies the hypothesis of Theorem 5.10: F is a 2×2 matrix-valued function analytic in a neighborhood of R which is unitary-valued on B , with a standard zero set, and $\rho_F = 1$. Assume furthermore that it is represented as in the conclusion of Theorem 5.10. Let $a_5 = a_6 = 0$ and $\delta_5 = e_1$ and $\delta_6 = e_2$. Then there exists a set E of μ measure zero such that for $p \notin E$, for each $v \in \mathbb{C}^{14}$ and for $\ell = 0, 1, 2$, the vector function $H_p(\zeta)vK^0(\zeta, z_\ell)$ is in the span of $\{K^0(\zeta, a_j)\delta_j\}$, where $z_0 (= 0), z_1, z_2$ are the zeros of ϕ_p . As a consequence, H_p is analytic on R and extends to a meromorphic function on Y .*

Proof. Given a finite set $Q \subset R$,

$$M_Q = \left((I - F(z)F(w)^*)K^0(z, w) \right)_{z, w \in Q}$$

has rank at most six. Moreover, as interpreted in the proof of Proposition 5.9, the range of M_Q lies in the span of $\{(K^0(z, a_j)\delta_j)_{z \in Q} : j = 1, \dots, 6\}$. Here $(K^0(z, a_j)\delta_j)_{z \in Q}$ is a column vector indexed by Q .

From the representation for F of Theorem 5.10,

$$M_Q = \left(\int H_p(z)(1 - \phi_p(z)\phi_p(w)^*)K^0(z, w)H_p(w)^* d\mu(p) \right)_{z, w \in Q}.$$

For each p , multiplication by ϕ_p , denoted M_p , is isometric on $\mathbb{H}^2(R, \omega_0)$, the Hilbert space of analytic functions on R . Hence $1 - M_pM_p^* \geq 0$, as is $(I - M_pM_p^*) \otimes E$, where E is the $n \times n$ matrix consisting of all ones. From the reproducing property of $K^0(\cdot, z)$, $M_p^*K^0(\cdot, z) = \phi_p(z)^*K^0(\cdot, z)$. Thus, if Q is a set of n points in R and x is the vector $(K^0(\cdot, w))_{w \in Q}$, then the matrix

$$P_Q(p) = \langle (I - M_pM_p^*) \otimes E x, x \rangle = \langle (1 - \phi_p(z)\phi_p(w)^*)K^0(z, w) \rangle_{z, w \in Q} \geq 0.$$

If we set $\tilde{Q} = Q \cup \{z_j(s)\}$, for any $j = 0, 1$, or 2 , then $P_{\tilde{Q}}(p) \geq 0$ as well. Furthermore, the upper left $n \times n$ block equals $P_Q(p)$ and the right $n \times 1$ column is $(K_0(z, z_j(s)))_{z \in Q}$. Hence as a vector $(K_0(z, z_j(s)))_{z \in Q} \in \text{ran } P_Q(p)^{1/2} = \text{ran } P_Q(p)$ for $j = 0, 1, 2$.

Since $P_Q(p) \geq 0$,

$$N_Q(p) = (H_p(z)(1 - \phi_p(z)\phi_p(w)^*)K^0(z, w)H_p(w)^*)_{z, w \in Q}$$

is also positive semidefinite for each p . If $M_Qx = 0$, then

$$0 = \int \langle N_Q(p)x, x \rangle d\mu(p),$$

so that $\langle N_Q(p)x, x \rangle = 0$ for almost every p . It follows that $N_Q(p)x = 0$ almost everywhere. Choosing a basis for the kernel of M_Q , there is a set E_Q of measure zero so that for $p \notin E_Q$, the kernel of M_Q is a subspace of the kernel of $N_Q(p)$. Thus, for such p , the range of $N_Q(p)$ is a subspace of the range of M_Q . In particular, the rank of $N_Q(p)$ is at most 6.

Moreover, if we let $D_Q(p)$ denote the (2×14) block diagonal matrix with (z, z) entry $H_p(z)$ ($z \in Q$), then $N_Q(p) = D_Q(p)P_Q(p)D_Q(p)^*$. Since $P_Q(p)$ is positive semidefinite, we conclude that the range of $D_Q(p)P_Q(p)$ is a subspace of M_Q . Thus, as $(K^0(z, z_j(s)))_{z \in Q}$ is in the range of $P_Q(p)$, $(H_p(z)vK^0(z, z_j(s)))_{z \in Q}$ is in the range of M_Q for every $v \in \mathbb{C}^{14}$ and $j = 0, 1, 2$.

Now suppose $Q_n \subset R$ is a finite set with $Q_n \subset Q_{n+1}$, $Q_0 = \{a_1, \dots, a_4, a_5 (= 0)\}$, and $\mathcal{D} = \bigcup_n Q_n$ a determining set. Then since

$$(5.7) \quad (H_p(z)vK^0(z, z_j(s)))_{z \in Q_n} \in \text{ran } M_{Q_n} \subseteq \bigvee_j (K^0(z, a_j)\delta_j)_{z \in Q},$$

there are constants $c_j^n(p)$ such that

$$(5.8) \quad H_p(z)vK^0(z, z_j(s)) = \sum_{j=1}^5 c_j^n(p)K^0(z, a_j)\delta_j, \quad z \in Q_n.$$

By linear independence of the $K^0(\cdot, a_j)$'s, the $c_j^n(p)$'s are uniquely determined when $n = 0, 1, \dots$ by (5.8). Since $Q_{n+1} \supset Q_n$, we must in this case have $c_j^{n+1}(p) = c_j^n(p)$ for all n , and thus there are unique constants $c_j(p)$ such that

$$(5.9) \quad H_p(z)vK^0(z, z_j(s)) = \sum_{j=1}^5 c_j(p)K^0(z, a_j)\delta_j, \quad z \in \mathcal{D}, \quad p = 0, 1, 2.$$

By considering this equation with $j = 5$, and using the fact that $K^0(z, 0) = 1$, we see that H_p agrees with an analytic function on the determining set \mathcal{D} . It follows that we can assume that $H_p(z)$ is analytic for each $p \notin E$ and that the relation of equation (5.8) holds throughout R . Furthermore, since the $K^0(\cdot, a_j)$'s extend to meromorphic functions on Y , H_p does as well. □

5.8. Diagonalization. In this subsection we show that when $\rho_F = 1$, a contractive matrix-valued function with unitary boundary values and a standard zero set is diagonalizable.

Lemma 5.12. *Suppose F is a 2×2 matrix-valued function on R whose determinant is not identically zero. If there exists a 2×2 unitary matrix U and scalar-valued functions $\psi_1, \psi_2 : R \rightarrow \mathbb{C}$ such that $F(z)F(w)^* = UD(z)D(w)^*U^*$, where*

$$D = \begin{pmatrix} \psi_1 & 0 \\ 0 & \psi_2 \end{pmatrix},$$

then there exists a unitary matrix V so that $F = UDV$.

Proof. The hypotheses imply $D(z)^{-1}U^*F(z) = D(w)^*U^*F(w)^{-1}$ whenever $F(z)$ and $F(w)$ are invertible. Hence, $D(z)^{-1}U^*F(z) = V$ is constant. One readily verifies that $V^*V = I$. □

Theorem 5.13. *Suppose F is a 2×2 matrix-valued function which is analytic in a neighborhood of R , unitary-valued on B and has a standard zero set, $\delta_j, a_j, j = 1, \dots, 4$. Assume further that the δ_j, a_j have the property that if h satisfies*

$$h(\zeta) = \sum_1^4 c_j K^0(\zeta, a_j)\delta_j + v$$

for some $c_1, \dots, c_4 \in \mathbb{C}$ and $v \in \mathbb{C}^2$ and if h does not have pole at either P_1 or P_2 , then h is constant.

Under these conditions, if $\rho_F = 1$, then F is diagonalizable; i.e., there exist 2×2 unitary matrices U and V and analytic functions $\varphi_j : R \rightarrow \mathbb{C}$ such that, with

$$D = \begin{pmatrix} \varphi_1 & 0 \\ 0 & \varphi_2 \end{pmatrix},$$

$$F = VDU.$$

Proof. By Lemma 5.11, we may assume that, except perhaps on a set Δ_0 of measure 0, if h is a column of some H_p , then $h(\zeta)K^0(\zeta, z_\ell(s))$ is in the span of $\{K^0(\zeta, a_j)\delta_j : 1 \leq j \leq 6\}$ for $\ell = 0, 1, 2$. Here $z_0(p) = 0, z_1(p), z_2(p)$ are the zeros of ϕ_p . By hypothesis, h (and so H_p) is constant as a function of ζ . From our normalization of the domain (see Corollary 2.13), one of the zeros of ϕ_p , say $z_1(p)$, is not zero. Thus, Theorem 4.3 (or rather the argument there) also implies, that if h is not zero, then there is a $1 \leq j_1(p) \leq 4$ so that $z_1(p) = a_{j_1(p)}$ and h is a multiple of $\delta_{j_1(p)}$. Thus, every column of H_p is a multiple of $\delta_{j_1(p)}$.

Theorem 5.10 gives us the representation

$$(5.10) \quad 1 - F(z)F(w)^* = \int_{\Pi} (1 - \phi_p(z)\phi_p(w)^*)H_pH_p^* d\mu(p).$$

Substituting $w = 0$ we find

$$(5.11) \quad I = \int H_pH_p^* d\mu(p).$$

In view of equation (5.11), rearranging equation (5.10) gives

$$(5.12) \quad F(z)F(w)^* = \int_{\Pi} \phi_p(z)\phi_p(w)^*H_pH_p^* d\mu(p).$$

Since the columns of H_p are all multiples of the single vector $\delta_{j_1(p)}$, $H_pH_p^*$ is rank one and thus may be written as $G(p)G(p)^*$, for a vector $G(p) \in \mathbb{C}^2$. (Indeed, $G(p)$ is the square root of the sum of the squares of the norms of the columns of H_p times $\delta_{j_1(p)}$.) Consequently,

$$(5.13) \quad F(z)F(w)^* = \int_{\Pi} \phi_p(z)\phi_p(w)^*G(p)G(p)^* d\mu(p).$$

Since $F(a_j)^*\delta_j = 0$ for all j , equation (5.13) gives

$$0 = \delta_j^*F(a_j)F(a_j^*)\delta_j = \int |\phi_p(a_j)|^2\|G(p)^*\delta_j\|^2 d\mu(p).$$

Consequently, for each j , $\phi_p(a_j)^*G(p)^*\delta_j = 0$ for almost every p , and so off of a set $Z_0 \subset \Pi$ of measure zero, $\phi_p(a_j)^*G(p)^*\delta_j = 0$ for all p and each j . Thus, by defining $G(p) = 0$ on Z_0 , we may assume that equation (5.13) holds and

$$\phi_p(a_j)^*G(p)^*\delta_j = 0$$

for all p, j .

Let $\Delta_0 = \{p \in \Pi : G(p) = 0\}$. If $p \notin \Delta_0$, then for each j , either $\phi_p(a_j) = 0$ or $G_p^*\delta_j = 0$. Since G_p is a multiple of $\delta_{j_1(p)}$ and no three of the δ_j are collinear, it follows that ϕ_p has zeros at two of the a_j , say $a_{j_1(p)}, a_{j_2(p)}$, and the $\delta_{j_3(p)}$ and $\delta_{j_4(p)}$ are collinear (and orthogonal to $\delta_{j_1(p)}$), where $\{a_1, \dots, a_4\} = \{a_{j_1(p)}, \dots, a_{j_4(p)}\}$. We can now turn to Theorem 4.3 to conclude that $z_1(p) = a_{j_1(p)}, z_2(p) = a_{j_2(p)}$. In particular, ϕ_p has distinct zeros and $\delta_{j_1(p)}$ and $\delta_{j_2(p)}$ are collinear (and orthogonal to $\delta_{j_3(p)}$ and $\delta_{j_4(p)}$). Let $J_1 = \{a_{j_1(p)}, a_{j_2(p)}\}$ and $J_2 = \{a_{j_3(p)}, a_{j_4(p)}\}$. In addition,

let Δ_1 denote the one-dimensional subspace of \mathbb{C}^2 spanned by $\delta_{j_1(p)}$ and Δ_2 the one-dimensional subspace of \mathbb{C}^2 spanned by $\delta_{j_3(p)}$.

If $p' \notin \Pi_0$, then by arguing as above, either $G(p') \in \Delta_1$ or $G(p') \in \Delta_2$. In the former case, the zeros of $\phi_{p'}$ are in J_2 and in the latter in J_1 . Hence, for each p , either

- (0) $G(p) = 0$; or
- (1) $G(p) \in \Delta_1$ and the nonzero zeros $z_1(p), z_2(p)$ are in J_2 ; or
- (2) $G(p) \in \Delta_2$ and the nonzero zeros $z_1(p), z_2(p)$ are in J_1 .

Let

$$\begin{aligned} \Pi_0 &= \{p \in \Pi : (0) \text{ holds} \}, \\ \Pi_1 &= \{p \in \Pi : (1) \text{ holds} \}, \\ \Pi_2 &= \{p \in \Pi : (2) \text{ holds} \}. \end{aligned}$$

If $p, q \in \Pi_1$, then ϕ_p and ϕ_q have the same zeros, and are therefore equal up to a rotation. Hence, for $p, q \in \Pi_1$, $\phi_p(z)\phi_p(w)^* = \phi_q(z)\phi_q(w)^*$. Choose $p^1 \in \Pi_1$ and let $\psi_1 = \phi_{p^1}$ denote a representative. If Π_2 is not empty, choose $p^2 \in \Pi_2$ and let $\psi_2 = \phi_{p^2}$. Otherwise, let $\psi_2 = 0$. Substituting into equation (5.12) and writing the integral as the sum of the integrals over Π_1 and Π_2 gives

$$F(z)F(w)^* = h_1\psi_1(z)\psi_1(w)^*h_1 + h_2\psi_2(z)\psi_2(w)^*h_2^*$$

for some $h_j \in \Delta_j$. Substituting $z = w = 1$ and using the fact that $F(1)F(1)^* = I$ shows that $\{h_1, h_2\}$ is an orthonormal basis for \mathbb{C}^2 (and $\psi_2 \neq 0$). Thus, we can apply the previous lemma and conclude that F is diagonalizable. □

6. THE OBSTRUCTION

We now demonstrate a 2×2 matrix function Ψ which is unitary on the boundary, analytic across the boundary, has a standard zero set, and which cannot be diagonalized. Since by Theorem 5.13, any function satisfying these conditions which has $\rho_\Psi = 1$ is diagonalizable, ρ_Ψ must be less than 1 for this Ψ and so by Theorem 5.2, rational dilation does not hold.

Recall the matrix inner functions $\Psi_{\eta,t}$ introduced in subsection 3.2. The t was fixed at the outset of that section. By Lemma 3.8 for small η , $\Psi = \Psi_{\eta,t}$ has a standard zero set. Let a_j^0, δ_j^0 denote the standard zero set for $\Psi_{0,t}$. (In particular, we can assume $\delta_1^0 = \delta_2^0 = e_1$ and $\delta_3^0 = \delta_4^0 = e_2$.) Let $\epsilon^0 > 0$ denote the $\epsilon > 0$ in Theorem 4.3 corresponding to this zero set. Thus, for a small $\eta > 0$ the function $\Psi = \Psi_{\eta,t}$ has a standard zero set and this zero set satisfies the conditions in Theorem 4.3. This Ψ is the promised function.

Lemma 6.1. *Ψ has the following properties:*

- (1) Ψ is unitary-valued on B ;
- (2) Ψ has a standard zero set;
- (3) the zero set $a_1, \dots, a_4, \delta_1, \dots, \delta_4$ has the property: if h satisfies

$$h(\zeta) = \sum_1^4 c_j K^0(\zeta, a_j) \delta_j + v$$

for some $c_1, \dots, c_4 \in \mathbb{C}$ and $v \in \mathbb{C}^2$ and if h does not have a pole at either P_1 or P_2 , then h is constant;

- (4) Ψ is not diagonalizable; i.e., there do not exist fixed unitaries U and V so that $U\Psi V^*$ is pointwise diagonal.

Proof. The only thing that remains to be proved is that Ψ is not diagonalizable. We argue by contradiction. Suppose there is a diagonal function D and fixed unitaries U, V so that $D = U\Psi V^*$. Of course D must be unitary-valued on B . In particular, $D(1)$ is unitary and so by multiplying on the left (or right) by $D(1)^*$, it may be assumed that $D(1) = I$. Since $\Psi(1) = I$, $V = U$.

Let φ_1, φ_2 denote the diagonal entries of D . Since D is unitary-valued on B , both φ_1 and φ_2 are unimodular on B . Further, as $\det(\Psi)$ has 6 zeros and a function unimodular on B has at least three zeros or is constant, we conclude that either both φ_1 and φ_2 have three zeros and thus take each value in $\{|z| \leq 1\}$ exactly three times in X , or one has six zeros and the other is a unimodular constant γ . This latter case cannot occur, since then

$$0 = \Psi(0) = U^* \begin{pmatrix} \gamma & 0 \\ 0 & 0 \end{pmatrix} U.$$

From Lemma 3.7, $\Psi(q_1)e_1 = e_1$. Thus Ue_1 is an eigenvector for $D(q_1)$ corresponding to eigenvalue 1, whence at least one of $\varphi_j(q_1)$ is 1. Similarly, Ue_2 is an eigenvector for $D(q_1^*)$ with eigenvalue 1, and at least one of $\varphi_j(q_1^*)$ is 1. Now, $D(q_1)$ cannot be a multiple of the identity, as otherwise both $\varphi_1(q_1)$ and $\varphi_2(q_1)$ equal 1, in which case at least one of these two functions takes the value 1 at both q_1 and q_1^* . Therefore, without loss of generality, $Ue_1 = \lambda_1 e_1$ and $Ue_2 = \lambda_2 e_2$ for unimodular λ_1 and λ_2 . Since D is also diagonal, we can assume $\lambda_j = 1$. Hence, $\Psi = D$.

From Lemma 3.7, $\Psi(q_2)(\eta e_1 + (1 - \eta^2)^{\frac{1}{2}} e_2) = \eta e_1 + (1 - \eta^2)^{\frac{1}{2}} e_2$ so that

$$D(q_2)(\eta e_1 + (1 - \eta^2)^{\frac{1}{2}} e_2) = \eta \varphi_1(q_2) e_1 + (1 - \eta^2)^{\frac{1}{2}} \varphi_2(q_2) e_2.$$

Since $\eta \neq 0$, it follows that $\varphi_1(q_2) = 1$. Similarly, $\varphi_2(q_2^*) = 1$. As $q_2 \neq q_2^*$, the function φ_j takes the value 1 twice on B_2 , a contradiction. \square

To prove our main theorem, simply note that the first three conditions of Lemma 6.1 imply, in view of Theorem 5.13, that if $\rho_\Psi = 1$, then Ψ is diagonal. Hence, it follows that $\rho_\Psi < 1$ and this completes the proof.

7. EXISTENCE OF A FINITE-DIMENSIONAL EXAMPLE

Let \mathcal{A} denote the closure of $\mathcal{R}(X)$, the rational functions with poles off of X as a subspace of $C(X)$. That is, $\mathcal{A} = \overline{\mathcal{R}(X)} \subset C(X)$. The algebra \mathcal{A} is an (abstract) operator algebra with the family of matrix norms,

$$\|F\| = \sup\{\|F(z)\| : z \in X\},$$

for $F = (f_{j,\ell}) \in M_n(\mathcal{A})$ and where $\|F(z)\|$ is the matrix norm of the $n \times n$ matrix $F(z)$.

Given a finite subset $\Lambda \subset R$, let I_Λ denote the ideal $\{f \in \mathcal{A} : f(\lambda) = 0 \text{ for all } \lambda \in \Lambda\}$ of \mathcal{A} . The quotient

$$\mathcal{A}_\Lambda = \mathcal{A}/I_\Lambda$$

inherits, in a canonical way, an operator algebra structure from \mathcal{A} . Namely,

$$\|\pi_\Lambda(F)\| = \inf\{\|G\| : G \in M_n(\mathcal{A}) \text{ and } \pi_\Lambda(G) = \pi_\Lambda(F)\}.$$

Here π_Λ denotes the quotient map.

The following is a special case of a version of a theorem of Vern Paulsen [27]. By a representation ν of \mathcal{A} is meant a unital homomorphism $\nu : \mathcal{A} \rightarrow B(H)$, where H is a Hilbert space. The representation ν is contractive if $\|\nu(f)\| \leq \|f\|$ for all $f \in \mathcal{A}$ and is completely contractive if for each n and each $F = (f_{j,\ell}) \in M_n(\mathcal{A})$, $\|\nu(F)\| = \|(\nu(f_{j,\ell}))\| \leq \|F\|$. Make similar definitions for representations, contractive representations, and completely contractive representations of \mathcal{A}_E .

Theorem 7.1 (Paulsen). *Every contractive representation of \mathcal{A} is completely contractive if and only if every contractive representation of every \mathcal{A}_Λ is completely contractive.*

The following observation of Paulsen is also useful. We give the proof.

Proposition 7.2. *If $\nu : \mathcal{A}_\Lambda \rightarrow B(H)$ is contractive, but not completely contractive, then $\nu \circ \pi_\Lambda : \mathcal{A} \rightarrow B(H)$ is contractive, but not completely contractive.*

Proof. Since ν and π_Λ are contractive, the composition $\nu \circ \pi_\Lambda$ is contractive. On the other hand, if $\nu \circ \pi_\Lambda$ is completely contractive, then it follows from the definition of the norm on $M_n(\mathcal{A}_E)$ as the quotient norm, that ν is completely contractive. To see this, given $\pi_\Lambda(F) = (\pi_\Lambda(f_{j,\ell}))$ in $M_n(\mathcal{A}_E)$ we can assume, by the definition of the quotient norm, that F can be chosen so that $\|F\|$ is only a little larger than $\|\pi_\Lambda(F)\|$. Since $\nu \circ \pi_\Lambda$ is completely contractive, the norm of $\nu \circ \pi_\Lambda(F)$ is no larger than $\|F\|$ and the result follows. \square

We use Paulsen's results, together with our own, to argue that there is a matrix X with R as a spectral set, but which does not dilate to a normal operator with spectrum in B . It is enough to prove that there is a finite subset $\Lambda \subset R$, a finite-dimensional Hilbert space H , and a representation $\nu : \mathcal{A}_E \rightarrow B(H)$ which is contractive, but not completely contractive. We begin with a lemma.

Lemma 7.3. *Fix Λ a finite subset of R . If $\gamma : \mathcal{A}_\Lambda \rightarrow B(K)$ is a representation, then there exist subspaces \mathcal{E}_λ , $\lambda \in \Lambda$, of K so that*

- (1) each \mathcal{E}_λ is closed;
- (2) K is the algebraic direct sum $\hat{\bigoplus} \mathcal{E}_\lambda$; and
- (3) if $f \in \mathcal{A}$ and $k \in \mathcal{E}_\lambda$, then

$$\gamma(\pi_\Lambda(f))k = \lambda k.$$

Proof. For each $\lambda \in \Lambda$, choose a function $e_\lambda \in \mathcal{A}$ such that $e_\lambda(\mu)$ is 0 if $\mu \in \Lambda$, but $\mu \neq \lambda$ and such that $e_\lambda(\lambda) = 1$. It follows that the operators $E_\lambda = \gamma(\pi_\Lambda(e_\lambda))$ are idempotents such that $E_\lambda E_\mu = 0$ if $\mu \neq \lambda$ and $\sum E_\lambda$ is the identity on K . Let $\mathcal{E}_\lambda = E_\lambda K$. Properties (1) and (2) are readily verified. As for (3),

$$\begin{aligned} (\gamma(\pi_\Lambda(f)) - f(\lambda))E_\lambda &= \gamma(\pi_\Lambda(f - f(\lambda)))e_\lambda \\ &= \gamma(\pi_\Lambda(0)) = 0. \end{aligned}$$

\square

From our result, there is a representation of \mathcal{A} which is contractive, but not completely contractive. From Paulsen's theorem, there exists a Λ and Hilbert space K (possibly infinite dimensional) and a representation $\gamma : \mathcal{A}_\Lambda \rightarrow B(K)$ which is contractive, but not completely contractive. Since this representation is not completely contractive, there exists an n and $F \in M_n(\mathcal{A})$ such that

$$1 = \|\pi_\Lambda(F)\| < \|\gamma(\pi_\Lambda(F))\|.$$

Thus, there exists a vector

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \bigoplus_1^n K = \mathbb{C}^n \otimes K$$

such that $\|x\| = 1$ and

$$\|\gamma(\pi_\Lambda(F))x\| > 1.$$

Write, $x_j = \sum x_j(\lambda)$ with respect to the algebraic direct sum $\hat{\bigoplus} \mathcal{E}_\lambda$ and let \mathcal{H} denote the span of $\{x_j(\lambda) : 1 \leq j \leq n, \lambda \in \Lambda\}$. Thus, \mathcal{H} is finite dimensional (the dimension is cardinality of Λ times n). From condition (3) of our lemma, \mathcal{H} is invariant for γ . Define $\nu : \mathcal{A}_E \rightarrow B(\mathcal{H})$ by restriction: $\nu(f) = \gamma(f)|_{\mathcal{H}}$. Since also $x \in \mathbb{C}^n \otimes \mathcal{H}$ we have $\gamma(\pi_\Lambda(F))x = \nu(\pi_\Lambda(F))x$ and thus

$$\|\nu(\pi_\Lambda(F))x\| > 1 = \|\pi_\Lambda(F)\|.$$

REFERENCES

1. M. B. Abrahamse and R. G. Douglas, *A class of subnormal operators related to multiply-connected domains*, *Advances in Math.* **19** (1976), no. 1, 106–148. MR0397468 (53:1327)
2. Jim Agler, *Rational dilation on an annulus*, *Ann. of Math. (2)* **121** (1985), no. 3, 537–563. MR0794373 (87a:47007)
3. ———, *On the representation of certain holomorphic functions defined on a polydisc*, *Topics in operator theory: Ernst D. Hellinger memorial volume*, *Oper. Theory Adv. Appl.*, vol. 48, Birkhäuser, Basel, 1990, pp. 47–66. MR1207393 (93m:47013)
4. Jim Agler and John Harland, *A monograph on function theory and Herglotz formulas for multiply connected domains*.
5. Jim Agler, John Harland, and Benjamin Raphael, *Classical function theory, operator dilation theory, and machine computations on multiply-connected domains*.
6. Jim Agler and John E. McCarthy, *Nevanlinna-Pick interpolation on the bidisk*, *J. Reine Angew. Math.* **506** (1999), 191–204. MR1665697 (2000a:47034)
7. ———, *Pick interpolation and Hilbert function spaces*, *Graduate Studies in Mathematics*, vol. 44, American Mathematical Society, Providence, RI, 2002. MR1882259 (2003b:47001)
8. William Arveson, *Subalgebras of C^* -algebras. II*, *Acta Math.* **128** (1972), no. 3-4, 271–308. MR0394232 (52:15035)
9. William B. Arveson, *Subalgebras of C^* -algebras*, *Acta Math.* **123** (1969), 141–224. MR0253059 (40:6274)
10. Joseph A. Ball and Kevin F. Clancey, *Reproducing kernels for Hardy spaces on multiply connected domains*, *Integral Equations Operator Theory* **25** (1996), no. 1, 35–57. MR1386327 (97f:46042)
11. Joseph A. Ball and Tavan T. Trent, *Unitary colligations, reproducing kernel Hilbert spaces, and Nevanlinna-Pick interpolation in several variables*, *J. Funct. Anal.* **157** (1998), no. 1, 1–61. MR1637941 (2000b:47028)
12. Joseph A. Ball, Tavan T. Trent, and Victor Vinnikov, *Interpolation and commutant lifting for multipliers on reproducing kernel Hilbert spaces*, *Operator theory and analysis* (Amsterdam, 1997), *Oper. Theory Adv. Appl.*, vol. 122, Birkhäuser, Basel, 2001, pp. 89–138. MR1846055 (2002f:47028)
13. Joseph A. Ball and Victor Vinnikov, *Hardy spaces on a finite bordered Riemann surface, multivariable operator model theory and Fourier analysis along a unimodular curve*, *Systems, approximation, singular integral operators, and related topics* (Bordeaux, 2000), *Oper. Theory Adv. Appl.*, vol. 129, Birkhäuser, Basel, 2001, pp. 37–56. MR1882690 (2003f:47016)
14. Stefan Bergman and Bruce Chalmers, *A procedure for conformal mapping of triply-connected domains*, *Math. Comp.* **21** (1967), 527–542. MR0228663 (37:4243a)
15. Kevin F. Clancey, *Toeplitz operators on multiply connected domains and theta functions*, *Contributions to operator theory and its applications* (Mesa, AZ, 1987), *Oper. Theory Adv. Appl.*, vol. 35, Birkhäuser, Basel, 1988, pp. 311–355. MR1017675 (91f:47038)

16. ———, *Representing measures on multiply connected planar domains*, Illinois J. Math. **35** (1991), no. 2, 286–311. MR1091446 (92e:46110)
17. John B. Conway, *Functions of one complex variable*, second ed., Graduate Texts in Mathematics, vol. 11, Springer-Verlag, New York, 1978. MR0503901 (80c:30003)
18. ———, *The theory of subnormal operators*, Mathematical Surveys and Monographs, vol. 36, American Mathematical Society, Providence, RI, 1991. MR1112128 (92h:47026)
19. H. M. Farkas and I. Kra, *Riemann surfaces*, second ed., Graduate Texts in Mathematics, vol. 71, Springer-Verlag, New York, 1992. MR1139765 (93a:30047)
20. John D. Fay, *Theta functions on Riemann surfaces*, Springer-Verlag, Berlin, 1973, Lecture Notes in Mathematics, Vol. 352. MR0335789 (49:569)
21. Stephen D. Fisher, *Function theory on planar domains*, Pure and Applied Mathematics, John Wiley & Sons Inc., New York, 1983, A second course in complex analysis, A Wiley-Interscience Publication. MR0694693 (85d:30001)
22. Helmut Grunsky, *Lectures on theory of functions in multiply connected domains*, Vandenhoeck & Ruprecht, Göttingen, 1978, Studia Mathematica, Skript 4. MR0463413 (573365)
23. Richard B. Holmes, *Geometric functional analysis and its applications*, Springer-Verlag, New York, 1975, Graduate Texts in Mathematics, No. 24. MR0410335 (53:14085)
24. David Mumford, *Tata lectures on theta. I*, Progress in Mathematics, vol. 28, Birkhäuser Boston Inc., Boston, MA, 1983, With the assistance of C. Musili, M. Nori, E. Previato and M. Stillman. MR0688651 (85h:14026)
25. ———, *Tata lectures on theta. II*, Progress in Mathematics, vol. 43, Birkhäuser Boston Inc., Boston, MA, 1984, Jacobian theta functions and differential equations, With the collaboration of C. Musili, M. Nori, E. Previato, M. Stillman and H. Umemura. MR0742776 (86b:14017)
26. Zeev Nehari, *Conformal mapping*, Dover Publications Inc., New York, 1975, Reprinting of the 1952 edition. MR0377031 (51:13206)
27. Vern Paulsen, *Private communication*.
28. ———, *Completely bounded maps and operator algebras*, Cambridge Studies in Advanced Mathematics, vol. 78, Cambridge University Press, Cambridge, 2002. MR1976867 (2004c:46118)
29. Vern I. Paulsen, *Completely bounded maps and dilations*, Pitman Research Notes in Mathematics Series, vol. 146, Longman Scientific & Technical, Harlow, 1986. MR0868472 (88h:46111)
30. Gilles Pisier, *A polynomially bounded operator on Hilbert space which is not similar to a contraction*, J. Amer. Math. Soc. **10** (1997), no. 2, 351–369. MR1415321 (97f:47002)
31. Donald Sarason, *The H^p spaces of an annulus*, Mem. Amer. Math. Soc. No. **56** (1965), 78. MR0188824 (32:6256)
32. V. L. Vinnikov and S. I. Fedorov, *On the Nevanlinna-Pick interpolation in multiply connected domains*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **254** (1998), no. Anal. Teor. Chisel i Teor. Funkts. 15, 5–27, 244. MR1691394 (2000d:47031)

SCHOOL OF MATHEMATICS AND STATISTICS, MERZ COURT, UNIVERSITY OF NEWCASTLE UPON TYNE, NEWCASTLE UPON TYNE, NE1 7RU, UNITED KINGDOM

E-mail address: `m.a.dritschel@ncl.ac.uk`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF FLORIDA, BOX 118105, GAINESVILLE, FLORIDA 32611-8105

E-mail address: `sam@math.ufl.edu`