Proposition 4.1.2 is from [**182**]. The full GGT representation appeared in Constantinescu [**212**] (see also [**75**, p. 49]). He finds it as the Naimark dilation of the GGT matrix. Its use to prove Verblunsky's theorem seems to be new, although I regard it as a natural step to take, given the GGT representation.

The analysis of the case where $d\mu$ has a single point in its essential spectrum is due to Golinskii, Nevai, and van Assche [**466**] and is discussed further in the next section. L. Golinskii [**458**] discussed the case where the essential spectrum is finite.

In some of the literature, what we call the GGT representation is called "the Hessenberg matrix." We have not used this name since "Hessenberg matrix" is generic for a matrix, M, with $M_{ij} = 0$ if $i > j+1$ and does not even include the unitarity.

L. Golinskii [**456**] used the GGT representation to prove an invariance of the a.c. spectrum if $\sum_{j=0}^\infty |\alpha_j - \beta_j| < \infty$ and $\inf|\alpha_j| > 0$. Since we will prove this without the assumption $\inf|\alpha_j| > 0$ in Section 4.3, we do not provide the details.

The history of Rakhmanov's lemma is discussed in Section 4.3. While he used it to prove his theorem, it should not be confused with his theorem.

4.2. The CMV Representation

In this section we discuss a representation of Cantero, Moral, and Velázquez [**181**]. The GGT representation had the attractive property that columns were finite but, alas, rows were not. The reason was that if $\mathcal{H}^{(n)} = \mathcal{H}_{(0,n)}$ is the span of $\{\delta_0, \delta_1, \dots, \delta_n\}$, then $\mathcal{H}^{(0)} \subset \mathcal{H}^{(1)} \subset \cdots$ with each $\mathcal{H}^{(n)}$ finite-dimensional and $U[\mathcal{H}^{(n)}] \subset$ some $\mathcal{H}^{(k)}$ (indeed, $\mathcal{H}^{(n+1)}$) with $U =$ multiplication by z. For rows to be finite, we would need $U^*[\mathcal{H}^{(n)}]$ in some $\mathcal{H}^{(k)}$, and that is not necessarily true since $\langle \varphi_j, z^{-1}\varphi_k\rangle$ can be nonzero for k fixed and all $j \geq k-1$.

To hope for finite rows and columns, we need to choose $\mathcal{H}^{(n)}$ so both multiplication by z and z^{-1} take $\mathcal{H}^{(n)}$ into some $\mathcal{H}^{(k)}$. The natural choice is to take $\mathcal{H}^{(n)}$ to be the span of the first $n+1$ elements of the ordered set: $1, z, z^{-1}, z^2, z^{-2}, \dots$. Clearly, $z[\mathcal{H}^{(n)}] \subset \mathcal{H}^{(n+2)}$, $z^{-1}[\mathcal{H}^{(n)}] \subset \mathcal{H}^{(n+2)}$, which means if $\{\chi_j\}_{j=0}^\infty$ is a basis with $\chi_j \in \mathcal{H}^{(j+1)} \cap \mathcal{H}^{(j)\perp}$, then $\langle \chi_j, z\chi_k\rangle = 0$ if $|j-k| \geq 3$, that is, each row and column has at most five nonzero elements, and in this basis, $U \equiv$ multiplication by z is five-diagonal. In fact, since either $z[\mathcal{H}^{(n)}] \subset \mathcal{H}^{(n+1)}$ or $z^{-1}[\mathcal{H}^{(n)}] \subset \mathcal{H}^{(n+1)}$ (depending on whether n is even or odd), each row and column has at most four nonzero elements. Analyzing small n carefully, one sees the pattern of possible nonzero elements is

$$\begin{pmatrix} * & * & * & & & \\ * & * & * & & & \\ & * & * & * & * & \\ & * & * & * & * & \\ & & & * & * & * & * \\ & & & * & * & * & * \end{pmatrix} \tag{4.2.1}$$

so there are 2×4 blocks with the main diagonal slicing at positions (12) and (23).

The key fact is that if the natural choice is made for the χ_j (i.e., Gram-Schmidt), the χ's can be expressed in terms of the φ's and the $\langle \chi_j, U\chi_k\rangle$ in terms of the Verblunsky coefficients $\{\alpha_j\}_{j=0}^\infty$. Before doing so, we want to address the issue of whether — with a bit more cleverness — one could not arrange for a tridiagonal or four-diagonal representation. Notice that every tridiagonal matrix is Hessenberg (i.e., only one row below the diagonal) and every four-diagonal matrix M has either M or M^t Hessenberg. Thus the following result implies that an infinite unitary matrix with a cyclic vector cannot be less than five-diagonal:

PROPOSITION 4.2.1. *Let U be a unitary matrix on $\ell^2(\mathbb{Z}_+)$ so that $u_{ij} = 0$ if $i - j \notin \{-1, 0, 1, \dots, n\}$. Then U is a direct sum of blocks of size at most $n + 1$.*

PROOF. By Proposition 4.1.2, U has the form (4.1.17). We will prove that this implies at least one of $\beta_0, \beta_1, \dots, \beta_n$ is zero. If β_j is zero, (4.1.17) shows that M decomposes into a direct sum of $(j+1) \times (j+1)$ block and the remaining infinite block. An inductive argument then shows the original matrix is a sum of blocks of size at most $n+1$. Thus we need only prove one of $\beta_0, \dots, \beta_n$ is zero.

By hypothesis, $u_{0,n+j} = 0$ for $j = 1, 2, \dots,$

$$\alpha_{n+j}\beta_0 \dots \beta_n \dots \beta_{n+j-1} = 0 \tag{4.2.2}$$

Suppose all of $\beta_0, \dots, \beta_n$ are nonzero. By (4.2.2) for $j = 1$, $\alpha_{n+1} = 0$. But then $\beta_{n+1} \neq 0$ and so, by (4.2.2) for $j = 2$, implies $\alpha_{n+2} = 0$. Thus

$$\beta_0, \dots, \beta_n \text{ all nonzero} \Rightarrow \alpha_{n+1} = \alpha_{n+2} = \cdots = 0$$

Looking at (4.1.17), this implies that the first $n+2$ rows have zero matrix elements from columns $n+1$ onwards, that is,

$$u_{ij} = 0 \qquad \text{if } i = 0, 1, \dots, n+1;\ j = n+1, n+2, \dots$$

(this is clear from (4.1.18)). If $\delta_0, \delta_1, \dots$ is the canonical basis, this says $U^*\delta_0, \dots, U^*\delta_{n+1}$ lie in the $n+1$-dimensional span of $\{\delta_0, \dots, \delta_n\}$. Thus U^* is an isometry from an $n+2$-dimensional space to an $n+1$-dimensional space — and we have a contradiction!

Thus, one of $\beta_0, \dots, \beta_n$ must be zero. $\square$

We now define the CMV basis and representation explicitly. Let $\mathcal{H}_{(k,\ell)}$ be the space of Laurent polynomials spanned by $\{z^j\}_{j=k}^{\ell}$ and $P_{(k,\ell)}$ the orthogonal projection onto $\mathcal{H}_{(k,\ell)}$ in $L^2(\partial\mathbb{D}, d\mu)$. Define

$$\mathcal{H}^{(n)} = \begin{cases} \mathcal{H}_{(-k,k)} & n = 2k \\ \mathcal{H}_{(-k,k+1)} & n = 2k+1 \end{cases} \tag{4.2.3}$$

and $P^{(n)} =$ projection onto $\mathcal{H}^{(n)}$.

Define $\chi_n^{(0)}$ by

$$\chi_n^{(0)} = \begin{cases} z^{-k} & n = 2k \\ z^{k+1} & n = 2k+1 \end{cases} \tag{4.2.4}$$

and the *CMV basis* by

$$\chi_n = \frac{(1 - P^{(n-1)})\chi_n^{(0)}}{\|(1 - P^{(n-1)})\chi_n^{(0)}\|} \tag{4.2.5}$$

where we use the nontriviality of $d\mu$ to conclude that the $\chi_n^{(0)}$ are linearly dependent, so $(1 - P^{(n-1)})\chi_n^{(0)} \neq 0$.

Clearly, it is just as natural to take the ordered set $1, z^{-1}, z, z^{-2}, z^2, \dots$ in place of $1, z, z^{-1}, z^2, z^{-2}, \dots$. We define

$$\widetilde{\mathcal{H}}^{(n)} = \begin{cases} \mathcal{H}_{(-k,k)} & n = 2k \\ \mathcal{H}_{(-k-1,k)} & n = 2k+1 \end{cases} \tag{4.2.6}$$

and $\tilde{P}^{(n)}$ = projection onto $\widetilde{\mathcal{H}}^{(n)}$.

$$x_n^{(0)} = \begin{cases} z^k & n = 2k \\ z^{-k-1} & n = 2k+1 \end{cases}$$

and the *alternate CMV basis* by

$$x_n = \frac{(1 - \tilde{P}^{(n-1)})x_n^{(0)}}{\|(1 - \tilde{P}^{(n-1)})x_n^{(0)}\|}$$

It will be convenient to define

$$\sigma_n = \chi_{2n} \qquad \tau_n = \chi_{2n-1} \qquad s_n = x_{2n} \qquad t_n = x_{2n-1}$$

so σ_n, s_n are labelled by $n = 0, 1, 2, \dots$ and τ_n, t_n by $n = 1, 2, \dots$.

PROPOSITION 4.2.2. (i) *We have that*

$$\tau_n = z^{-n+1}\varphi_{2n-1} \tag{4.2.7}$$

$$\sigma_n = z^{-n}\varphi_{2n}^* \tag{4.2.8}$$

$$t_n = z^{-n}\varphi_{2n-1}^* \tag{4.2.9}$$

$$s_n = z^{-n}\varphi_{2n} \tag{4.2.10}$$

(ii)

$$x_n(z) = \overline{\chi_n(1/\bar{z})} \tag{4.2.11}$$

PROOF. (i) φ_{2n-1} is $(1 - P_{(0,2n-2)})z^{2n-1}/\|\dots\|$. Since

$$z^\ell P_{(k,m)} z^{-\ell} = P_{(k+\ell, m+\ell)}$$

we have

$$\begin{aligned} z^{-n+1}\varphi_{2n-1} &= \frac{[z^{-n+1}(1 - P_{(0,2n-2)})z^{n-1}]z^n}{\|\dots\|} \\ &= \frac{(1 - P_{(-n+1,n-1)})z^n}{\|\dots\|} \\ &= \frac{(1 - P^{(2n-2)})\chi_{2n-1}^{(0)}}{\|\dots\|} \\ &= \chi_{2n-1} = \tau_n \end{aligned}$$

proving (4.2.7). The proofs of the others are similar if we note that $\varphi_\ell^* = (1 - P_{(1,\ell)})1/\|\dots\|$.

(ii) this is immediate from (i) and (1.1.6). □

The *CMV representation*, $\mathcal{C}(d\mu)$, is

$$\mathcal{C}_{ij}(d\mu) = \langle \chi_i, z\chi_j \rangle \tag{4.2.12}$$

where $\{\chi_j\}_{j=0}^\infty$ is the CMV basis, and the *alternate CMV representation*, $\tilde{\mathcal{C}}(d\mu)$, is the matrix

$$\tilde{\mathcal{C}}_{ij}(d\mu) = \langle x_i, zx_j \rangle \tag{4.2.13}$$

where $\{x_j\}_{j=0}^\infty$ is the alternate CMV basis (we will see that $\tilde{\mathcal{C}}$ is the transpose of $\mathcal{C}$).

PROPOSITION 4.2.3. *$\mathcal{C}(d\mu)$ is given by*

$$\begin{aligned} \langle\sigma_{j-1}, z\sigma_j\rangle &= \rho_{2j-1}\rho_{2j-2} & \langle\sigma_j, z\sigma_j\rangle &= -\bar\alpha_{2j}\alpha_{2j-1} \\ \langle\tau_j, z\sigma_j\rangle &= -\alpha_{2j-2}\rho_{2j-1} & \langle\tau_{j+1}, z\sigma_j\rangle &= -\alpha_{2j-1}\rho_{2j} \\ \langle\sigma_{j-1}, z\tau_j\rangle &= \bar\alpha_{2j-1}\rho_{2j-2} & \langle\sigma_j, z\tau_j\rangle &= \bar\alpha_{2j}\rho_{2j-1} \\ \langle\tau_j, z\tau_j\rangle &= -\bar\alpha_{2j-1}\alpha_{2j-2} & \langle\tau_{j+1}, z\tau_j\rangle &= \rho_{2j}\rho_{2j-1} \end{aligned}$$

All other matrix elements are zero.

Remarks. 1. As usual, $\alpha_{-1} = -1$.

2. The terms that have a minus are precisely those with a factor of α_ℓ for some ℓ (thinking of α_ℓ and $\bar\alpha_\ell$ as independent variables).

3. The Θ-factorization below will provide another proof of this that is conceptually simpler than the brute force calculation.

PROOF. That these are the only nonzero matrix elements follows from the block structure (4.2.1). By Proposition 4.2.2, the eight matrix elements here correspond precisely to the eight matrix elements in Proposition 1.5.9(i); for example,

$$\begin{aligned} \langle\sigma_{j-1}, z\sigma_j\rangle &= \langle z^{-j+1}\varphi^*_{2j-2}, zz^{-j}\varphi^*_{2j}\rangle \\ &= \langle\varphi^*_{2j-2}, \varphi^*_{2j}\rangle \end{aligned}$$

is given by (1.5.72) and

$$\begin{aligned} \langle\sigma_j, z\tau_j\rangle &= \langle z^{-j}\varphi^*_{2j}, z^2z^{-j}\varphi_{2j-1}\rangle \\ &= \langle\varphi^*_{2j}, z^2\varphi_{2j-1}\rangle \end{aligned}$$

is given by (1.5.70). $\square$

Thus

$$\mathcal{C} = \begin{pmatrix} \bar\alpha_0 & \bar\alpha_1\rho_0 & \rho_1\rho_0 & 0 & 0 & \dots \\ \rho_0 & -\bar\alpha_1\alpha_0 & -\rho_1\alpha_0 & 0 & 0 & \dots \\ 0 & \bar\alpha_2\rho_1 & -\bar\alpha_2\alpha_1 & \bar\alpha_3\rho_2 & \rho_3\rho_2 & \dots \\ 0 & \rho_2\rho_1 & -\rho_2\alpha_1 & -\bar\alpha_3\alpha_2 & -\rho_3\alpha_2 & \dots \\ 0 & 0 & 0 & \bar\alpha_4\rho_3 & -\bar\alpha_4\alpha_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \tag{4.2.14}$$

A similar calculation or Corollary 4.2.6 below shows that

$$\tilde{\mathcal{C}} = \begin{pmatrix} \bar\alpha_0 & \rho_0 & 0 & 0 & 0 & \dots \\ \bar\alpha_1\rho_0 & -\bar\alpha_1\alpha_0 & \bar\alpha_2\rho_1 & \rho_2\rho_1 & 0 & \dots \\ \rho_1\rho_0 & -\rho_1\alpha_0 & -\bar\alpha_2\alpha_1 & -\rho_2\alpha_1 & 0 & \dots \\ 0 & 0 & \bar\alpha_3\rho_2 & -\bar\alpha_3\alpha_2 & \bar\alpha_4\rho_3 & \dots \\ 0 & 0 & \rho_3\rho_2 & -\rho_3\alpha_2 & -\bar\alpha_4\alpha_3 & \dots \end{pmatrix} \tag{4.2.15}$$

There is a way of writing $\mathcal{C}$ as a product that is illuminating and useful for computations. It involves the pair of bases $\{\chi_j\}_{j=0}^\infty$ and $\{x_j\}_{j=0}^\infty$. Define

$$\mathcal{M}_{ij}(d\mu) = \langle x_i, \chi_j\rangle \tag{4.2.16}$$

$$\mathcal{L}_{ij}(d\mu) = \langle \chi_i, zx_j\rangle \tag{4.2.17}$$

PROPOSITION 4.2.4. (i)

$$\mathcal{C} = \mathcal{L}\mathcal{M} \tag{4.2.18}$$

(ii)
$$\tilde{\mathcal{C}} = \mathcal{M}\mathcal{L} \tag{4.2.19}$$

(iii) *$\mathcal{M}$ is tridiagonal with a single 1×1 block followed by 2×2 blocks.*
(iv) *$\mathcal{L}$ is tridiagonal with 2×2 blocks.*

PROOF. (i)
$$\begin{aligned}
\mathcal{C}_{ij} &= \langle \chi_i, z\chi_j \rangle \\
&= \langle z^{-1}\chi_i, \chi_j \rangle \\
&= \sum_{\ell=0}^{\infty} \langle z^{-1}\chi_i, x_\ell \rangle \langle x_\ell, \chi_j \rangle \\
&= \sum_{\ell=0}^{\infty} \langle \chi_i, zx_\ell \rangle \langle x_\ell, \chi_j \rangle \\
&= \sum_{\ell=0}^{\infty} \mathcal{L}_{i\ell}\mathcal{M}_{\ell j} = (\mathcal{L}\mathcal{M})_{ij}
\end{aligned}$$

(ii) Similar to (i).

(iii) $\{\chi_j\}_{j=0}^{2n}$ and $\{x_j\}_{j=0}^{2n}$ both span the space $\mathcal{H}_{(-n,n)}$ generated by $\{z^k\}_{k=-n}^{n}$. Thus $\{\tau_n, \sigma_n\}$ and $\{t_n, s_n\}$ are both orthonormal bases for the two-dimensional space $\mathcal{H}_{(-n,n)} \cap \mathcal{H}^{\perp}_{(-n+1,n-1)}$. It follows that $\mathcal{M}$, the change of basis unitary, has a 2×2 block structure except for $x_0 = \chi_0 = 1$ at the start.

(iv) $\{\chi_j\}_{j=0}^{2n+1}$ and $\{zx_j\}_{j=0}^{2n+1}$ both span the space $\mathcal{H}_{(-n,n+1)}$ generated by $\{z^k\}_{k=-n}^{n+1}$. Thus $\{\sigma_n, \tau_{n+1}\}$ and $\{zs_n, zt_{n+1}\}$ are orthonormal bases for the two-dimensional space $\mathcal{H}_{(-n,n+1)} \cap \mathcal{H}^{\perp}_{(-n+1,n)}$, so $\mathcal{L}$ has the stated 2×2 block structure. $\square$

THEOREM 4.2.5. *Let*
$$\Theta_j = \begin{pmatrix} \bar{\alpha}_j & \rho_j \\ \rho_j & -\alpha_j \end{pmatrix} \tag{4.2.20}$$
Then
$$\mathcal{M} = \begin{pmatrix} 1 & & & \\ & \Theta_1 & & \\ & & \Theta_3 & \\ & & & \ddots \end{pmatrix} \qquad \mathcal{L} = \begin{pmatrix} \Theta_0 & & & \\ & \Theta_2 & & \\ & & \Theta_4 & \\ & & & \ddots \end{pmatrix} \tag{4.2.21}$$

Remarks. 1. For this reason, we call (4.2.18), (4.2.19) the Θ-factorization.

2. Taking the matrix product and using (4.2.18) provides a new proof of Proposition 4.2.3.

3. In a sense, Theorem 4.2.5 makes Proposition 4.2.4(iii) and (iv) unnecessary. The direct calculation below shows L and M have the claimed block structure. We have included Proposition 4.2.4(iii) and (iv) because we feel that its proof explains why the block structure occurs in a way that a mere calculation does not.

PROOF. By (1.5.43),
$$\sigma_n = -\alpha_{2n-1}s_n + \rho_{2n-1}t_n \tag{4.2.22}$$

and by (1.5.25)

$$\tau_n = \rho_{2n-1} s_n + \bar{\alpha}_{2n-1} t_n \tag{4.2.23}$$

This implies the formula for $\mathcal{M}$, for example,

$$\langle s_n, \sigma_n \rangle = \langle s_n, -\alpha_{2n-1} s_n + \rho_{2n-1} t_n \rangle = -\alpha_{2n-1}$$

yielding the jj matrix elements of $\mathcal{M}$ for $j = 3, 5, 7, \dots$.

By (1.5.25),

$$z s_n = -\bar{\alpha}_{2n} \sigma_n + \rho_{2n} \tau_{n+1} \tag{4.2.24}$$

$$z t_{n+1} = \rho_{2n} \sigma_n - \alpha_{2n} \tau_{n+1} \tag{4.2.25}$$

which yields the formula for $\mathcal{L}$. □

COROLLARY 4.2.6. $\tilde{\mathcal{C}} = \mathcal{C}^t$

Remarks. 1. ^{t} is transpose, that is, adjoint without the complex conjugative.

2. This can also be proven using (4.2.11) and the fact that $\mathcal{C}^*$ is multiplication by $z^{-1} = \bar{z}$.

EXAMPLE 4.2.7. In the free case, $d\mu = \frac{d\theta}{2\pi}$, $\alpha_j \equiv 0$ and

$$\mathcal{L} = \begin{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & & \\ & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \\ & & \ddots \end{pmatrix}$$

$$\mathcal{M} = \begin{pmatrix} 1 & & & \\ & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & & \\ & & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} & \\ & & & \ddots \end{pmatrix}$$

$$\mathcal{C} = \begin{pmatrix} 0 & 0 & 1 & & & & \\ 1 & 0 & 0 & & & & \\ & 0 & 0 & 0 & 1 & & \\ & 1 & 0 & 0 & 0 & & \\ & & & 0 & 0 & 0 & 1 \\ & & & 1 & 0 & 0 & 0 \end{pmatrix}$$

with $\left(\begin{smallmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{smallmatrix}\right)$ blocks. □

PROOF. Clearly, $\Theta_j^t = \Theta_j$ so $\mathcal{M}^t = \mathcal{M}$ and $\mathcal{L}^t = \mathcal{L}$. Thus

$$\tilde{\mathcal{C}} = \mathcal{M}\mathcal{L} = \mathcal{M}^t \mathcal{L}^t = (\mathcal{L}\mathcal{M})^t = \mathcal{C}^t$$ □

The first benefit of the CMV representation is our final proof of Verblunsky's theorem.

THEOREM 4.2.8 (Fourth Proof of Verblunsky's Theorem). *Let* $\{\alpha_j\}_{j=0}^\infty \in \times_{n=0}^\infty \mathbb{D}$. *Define* Θ *by* (4.2.20), $\mathcal{M}, \mathcal{L}$ *by* (4.2.21), *and* $\mathcal{C} = \mathcal{L}\mathcal{M}$. *Then* $\mathcal{C}$ *is unitary. Let* $d\mu$ *be the spectral measure for* $\mathcal{C}$ *and vector,* δ_0. *Then*

$$\alpha_j(d\mu) = \alpha_j$$

In particular, $\{\alpha_j\}_{j=0}^\infty$ *are the Verblunsky coefficients of a measure.*

Remark. To prove Verblunsky's theorem using the GGT representation required some extensive computation to show $\mathcal{G}(\{\alpha_j\}_{j=0}^\infty)$ was unitary. In the CMV representation, the Θ-factorization makes it immediately evident that $\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)$ is unitary.

PROOF. Since Θ is unitary (whenever $|\alpha| \leq 1$), $\mathcal{M}$, $\mathcal{L}$, and $\mathcal{C}$ are unitary. Let $\delta_0, \delta_1, \dots$ be the canonical basis for $\ell^2(\mathbb{Z}_+)$. From (4.2.14) and Proposition 4.2.3 (which come from multiplication of L and M), we note the following critical matrix elements in $\mathcal{C}$:

$$\langle \delta_0, \mathcal{C}\delta_0 \rangle = \bar{\alpha}_0 \tag{4.2.26}$$

$$\langle \delta_{2n+1}, \mathcal{C}\delta_{2n-1} \rangle = \rho_{2n}\rho_{2n-1} \qquad n = 1, 2, 3, \dots \tag{4.2.27}$$

$$\langle \delta_{2n}, \mathcal{C}\delta_{2n+2} \rangle = \rho_{2n+1}\rho_{2n} \qquad n = 0, 1, 2, \dots \tag{4.2.28}$$

$$\langle \delta_{2n}, \mathcal{C}\delta_{2n-1} \rangle = \bar{\alpha}_{2n}\rho_{2n-1} \qquad n = 1, 2, 3, \dots \tag{4.2.29}$$

$$\langle \delta_{2n-2}, \mathcal{C}\delta_{2n-1} \rangle = \bar{\alpha}_{2n-1}\rho_{2n-2} \qquad n = 1, 2, 3, \dots \tag{4.2.30}$$

From (4.2.27), the limiting value $\langle \delta_1, \mathcal{C}\delta_0 \rangle = \rho_0$, and the shape (4.2.1), we see that

$$\mathcal{C}^{n+1}\delta_0 = \rho_0\rho_1 \dots \rho_{2n}\delta_{2n+1} + \text{l.c.}\{\delta_0, \dots, \delta_{2n}\} \tag{4.2.31}$$

where l.c.$\{\dots\}$ means a linear combination of the vectors in $\{\dots\}$. From $\mathcal{C}^{-1} = \mathcal{C}^*$, (4.2.28), and the shape (4.2.1), we see

$$\mathcal{C}^{-n}\delta_0 = \rho_0 \dots \rho_{2n-1}\delta_{2n} + \text{l.c.}\{\delta_0, \dots, \delta_{2n-1}\} \tag{4.2.32}$$

It follows by induction that $\delta_0, \delta_1, \delta_2, \dots$ is obtained by applying the Gram-Schmidt process to $\delta_0, \mathcal{C}\delta_0, \mathcal{C}^{-1}\delta_0, \mathcal{C}^2\delta_0, \mathcal{C}^{-2}\delta_0, \dots$. Thus, if $V : \ell^2 \to L^2(\partial\mathbb{D}, d\mu)$ so $V\mathcal{C}V^{-1}$ = multiplication by z and $V\delta_0 = 1$, then

$$V\delta_n = \chi_n(z, d\mu)$$

This means that $\mathcal{C}$ is the CMV matrix of $d\mu$, so by (4.2.27)–(4.2.31),

$$\bar{\alpha}_0(d\mu) = \bar{\alpha}_0 \tag{4.2.33}$$

$$\rho_{2n}(d\mu)\rho_{2n-1}(d\mu) = \rho_{2n}\rho_{2n-1} \tag{4.2.34}$$

$$\rho_{2n+1}(d\mu)\rho_{2n}(d\mu) = \rho_{2n+1}\rho_{2n} \tag{4.2.35}$$

$$\bar{\alpha}_{2n}(d\mu)\rho_{2n-1}(d\mu) = \bar{\alpha}_{2n}\rho_{2n-1} \tag{4.2.36}$$

$$\bar{\alpha}_{2n-1}(d\mu)\rho_{2n-2}(d\mu) = \bar{\alpha}_{2n-1}\rho_{2n-2} \tag{4.2.37}$$

Because all $\rho_j \neq 0$, the first three equations imply $\rho_j(d\mu) = \rho_j$ for all j, and then the last two that $\alpha_j(d\mu) = \alpha_j$. $\square$

We first turn to the understanding of the change $\alpha_j \to \lambda\alpha_j$ from the CMV point of view. A glance at (4.2.14) shows that $\mathcal{C}(\{\lambda\alpha_j\}_{j=0}^\infty)$ and $\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)$ differ by an infinite rank operator (unless only finitely many α's are nonzero) unlike the case for $\mathcal{G}$. This puzzle is resolved by a unitary equivalence. Given $\lambda \in \partial\mathbb{D}$, let

$\gamma^2 = \lambda$. We will pick the square root with $\operatorname{Im}\gamma \geq 0$ for definiteness. Let $U(\lambda)$ be the diagonal unitary matrix

$$U(\lambda) = \begin{pmatrix} \bar\gamma & & & & \\ & \gamma & & 0 & \\ & & \bar\gamma & & \\ & 0 & & \gamma & \\ & & & & \ddots \end{pmatrix} \tag{4.2.38}$$

THEOREM 4.2.9. *Let $\lambda \in \partial\mathbb{D}$. Then*

(i) *$U(\lambda)\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)U(\lambda)^{-1}$ and $\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)$ have the same spectral measure associated with the vector δ_0.*

(ii) *$\mathcal{C}(\{\lambda\alpha_j\}_{j=0}^\infty) - U(\lambda)\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)U(\lambda)^{-1}$ is a rank one operator.*

Remarks. 1. As the proof shows, the difference is only in column 1, where $\mathcal{C}(\{\lambda\alpha_j\}_{j=0}^\infty)$ has elements $(\bar\lambda\bar\alpha_0, \rho_0, 0, \dots)$ and $U(\lambda)\mathcal{C}(\{a_j\}_{j=0}^\infty)U(\lambda)^{-1}$ has elements $(\bar\alpha_0, \lambda\rho_0, 0, \dots) = \lambda(\bar\lambda\bar\alpha_0, \rho_0, 0, \dots)$.

2. This result and the theory of rank one perturbations in Subsection 1.4.16 provide a CMV proof of Theorem 3.2.14.

3. This result and (1.4.25) provide a new proof of the eigenvalue interlacing in (c) of Theorem 3.2.16.

PROOF. First, since $U(\lambda)^{-1}\delta_0 = \gamma\delta_0$, if A_1 and A_2 are the two operators in (i), then $\langle \delta_0, A_1^k\delta_0\rangle = \langle\delta_0, A_2^k\delta_0\rangle$ for all k, so the spectral measures are the same. Let

$$u(\lambda) = \begin{pmatrix} \bar\gamma & 0 \\ 0 & \gamma \end{pmatrix}$$

Then (note that it is $u(\lambda)$, not $u(\lambda)^{-1}$, to the right of Θ)

$$u(\lambda)\Theta(\alpha)u(\lambda) = \Theta(\lambda\alpha)$$

so

$$U(\lambda)\mathcal{L}(\{\alpha_j\}_{j=0}^\infty)U(\lambda) = \mathcal{L}(\{\lambda\alpha_j\}_{j=0}^\infty) \tag{4.2.39}$$

and

$$U(\lambda)^{-1}\mathcal{M}(\{\alpha_j\}_{j=0}^\infty)U(\lambda)^{-1} = \mathcal{M}(\{\lambda\alpha_j\}_{j=0}^\infty) + (\lambda - 1)P_0 \tag{4.2.40}$$

where P_0 is the projection onto δ_0. (4.2.40) follows because all $\mathcal{M}$'s have 1 as their 11 matrix element, while $U(\lambda)^{-1}\mathcal{M}U(\lambda)^{-1}$ has λ as its matrix element.

Multiplying (4.2.39) by (4.2.40),

$$U(\lambda)\mathcal{C}(\{\alpha_j\}_{j=0}^\infty)U(\lambda)^{-1} = \mathcal{C}(\{\lambda\alpha_j\}_{j=0}^\infty) + (\lambda-1)\mathcal{L}(\{\lambda\alpha_j\}_{j=0}^\infty)P_0 \tag{4.2.41}$$

$\square$

On the other hand, internal variations are more transparent in the CMV representation than in the GGT representation:

THEOREM 4.2.10. *Let $\{\alpha_j\}_{j=0}^\infty$ and $\{\tilde\alpha_j\}_{j=0}^\infty$ be two sets of Verblunsky coefficients so that*

$$\tilde\alpha_j = \alpha_j \qquad j \neq k, k+1, \dots, k+\ell-1$$

for some $k \in \{0, 1, 2, \dots\}$ and $\ell \geq 1$. Then $\mathcal{C}(\{\tilde\alpha_j\}_{j=0}^\infty) - \mathcal{C}(\{\alpha_j\}_{j=0}^\infty)$ is rank at most $\ell + 1$ if ℓ is odd or $k = 0$ and at most $\ell + 2$ if ℓ is even and $k \geq 1$.

Proof. Suppose first that k is even and ℓ is odd. Write $\mathcal{C}(\{\tilde{\alpha}\}_{j=0}^{\infty}) - \mathcal{C}(\{\alpha_j\}_{j=0}^{\infty}) = \delta\mathcal{C}$ and similarly for $\delta\mathcal{M}$ and $\delta\mathcal{L}$. Then

$$\delta\mathcal{C} = \delta\mathcal{L}\mathcal{M}(\{\alpha\}_{j=0}^{\infty}) + \mathcal{L}(\{\tilde{\alpha}\}_{j=0}^{\infty})\delta\mathcal{M}$$

Clearly, we have that $\operatorname{ran}(\delta\mathcal{L}) \subset \operatorname{span}[\delta_k, \delta_{k+1}, \dots, \delta_{k+\ell}]$. Moreover, $\operatorname{ran}(\delta\mathcal{M}) \subset \operatorname{span}[\delta_{k+1}, \dots, \delta_{k+\ell-1}]$ and $\mathcal{L}$ takes that into $\operatorname{span}[\delta_k, \dots, \delta_{k+\ell}]$, so $\operatorname{ran}(\delta\mathcal{C}) \subset \operatorname{span}[\delta_k, \dots, \delta_{k+\ell}]$ has dimension $\ell + 1$.

If k is even and ℓ is even, $\operatorname{ran}(\delta\mathcal{L}) \subset \operatorname{span}[\delta_k, \delta_{k+1}, \dots, \delta_{k+\ell-1}]$, $\operatorname{ran}(\delta\mathcal{M}) \subset \operatorname{span}[\delta_{k+1}, \dots, \delta_{k+\ell}]$ and $\mathcal{L}$ takes that into $\operatorname{span}[\delta_k, \dots, \delta_{k+\ell+1}]$, so $\operatorname{ran}(\delta C) \subset \operatorname{span}[\delta_k, \dots, \delta_{k+\ell+1}]$, which has dimension $\ell + 2$.

If k is odd or $k = 0$, ℓ even, we look at $\operatorname{ran}(\delta\mathcal{C}^t)$. $\square$

Next, we turn to the case of small essential spectra.

Theorem 4.2.11. *Let μ be a nontrivial probability measure on $\partial\mathbb{D}$ and let $\tau \in \partial\mathbb{D}$. The following are equivalent:*

(i) *The essential support of $d\mu$ is $\{\tau\}$.*

(ii)
$$\lim_{n\to\infty} \bar{\alpha}_{n+1}\alpha_n = -\tau \tag{4.2.42}$$

(iii) *For any continuous function f on $\partial\mathbb{D}$,*
$$\lim_{n\to\infty} \int f(e^{i\theta})|\varphi_n(e^{i\theta})|^2\, d\mu(\theta) = f(\tau)$$

(iv) *For any continuous function f on $\partial\mathbb{D}$ and any $k \in \mathbb{Z}$,*
$$\lim_{n\to\infty} \int f(e^{i\theta})\varphi_n(e^{i\theta})\,\overline{\varphi_{n+k}(e^{i\theta})}\, d\mu(\theta) = f(\tau)\delta_{k0}$$

(v) *We have*
$$\lim_{n\to\infty} \int e^{i\theta}|\varphi_n(e^{i\theta})|^2\, d\mu(\theta) = \tau \tag{4.2.43}$$

(vi) *We have for any $k \in \mathbb{Z}$,*
$$\lim_{n\to\infty} \int e^{i\theta}\varphi_n(e^{i\theta})\,\overline{\varphi_{n+k}(e^{i\theta})}\, d\mu(\theta) = \tau\delta_{k0} \tag{4.2.44}$$

Proof. We will show that

$$\begin{array}{ccccccccccc} & & & & \text{(iii)} & & & & & & \\ & & & \nearrow & & \searrow & & & & & \\ \text{(i)} & \Longrightarrow & \text{(iv)} & & & & \text{(v)} & \Longrightarrow & \text{(ii)} & \Longrightarrow & \text{(i)} \\ & & & \searrow & & \nearrow & & & & & \\ & & & & \text{(vi)} & & & & & & \end{array}$$

<u>(i) $\Rightarrow$ (iv)</u>. (i) means that $\mathcal{C} - \tau\mathbf{1}$ is compact, which implies $f(\mathcal{C}) - f(\tau)\mathbf{1}$ is compact, first for any polynomial (by factorization) and then for any continuous function (by polynomial approximation). Then $\|(f(\mathcal{C}) - f(\tau))\varphi_n\| \to 0$, which implies (iv).

<u>(iv) $\Rightarrow$ (iii)</u> and <u>(iv) $\Rightarrow$ (vi)</u> as special cases.

<u>(iii) $\Rightarrow$ (v)</u> and <u>(vi) $\Rightarrow$ (v)</u> as special cases.

<u>(v) $\Rightarrow$ (ii)</u>. By (4.1.5), the left side of (4.2.43) is $-\bar{\alpha}_n\alpha_{n-1}$.

(ii) $\Rightarrow$ (i). (4.2.42) implies $|\alpha_n| \to 1$ so $\rho_n \to 0$. A glance at (4.2.14) or Proposition 4.2.3 shows that all matrix elements of $\mathcal{C} - \tau\mathbf{1}$ go to zero as the indices go to infinity. Since $\mathcal{C}$ is a five-diagonal matrix, this implies that $\mathcal{C} - \tau\mathbf{1}$ is compact, so τ is the only point in its essential spectrum. $\square$

THEOREM 4.2.12. *If $d\mu$ has only N points in its essential spectrum, then*

$$\lim_{n\to\infty} \prod_{j=1}^{N} \rho_{n+j} = 0 \tag{4.2.45}$$

In particular,

$$\limsup_{n\to\infty} |\alpha_n| = 1 \tag{4.2.46}$$

PROOF. This result is easier to prove using the GGT representation, so we will use that. Let $\tau_1, \dots, \tau_N$ be the essential spectrum of $\mathcal{G}$. Then $\prod_{j=1}^{N}(\mathcal{G} - \tau_j\mathbf{1})$ is compact. By (4.1.6),

$$\left\langle \delta_{n+N+1}, \prod_{j=1}^{N}(\mathcal{G} - \tau_j\mathbf{1})\delta_{n+1} \right\rangle = \prod_{j=1}^{N} \rho_{n+j}$$

Since $\delta_{n+1} \to 0$ weakly, compactness implies (4.2.45).

(4.2.45) implies $\liminf \rho_j = 0$ so (4.2.46) holds. $\square$

EXAMPLE 4.2.13. Suppose $\alpha_{2n} \to 0$ and $\alpha_{2n+1} \to i$. Then

$$\Theta_{2n} \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \Theta_{2n+1} \to \begin{pmatrix} -i & 0 \\ 0 & -i \end{pmatrix}$$

Thus $\mathcal{M} + i\mathbf{1}$ is compact. It follows that

$$\mathcal{L}\mathcal{M} = -i \begin{pmatrix} \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) & & \\ & \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) & \\ & & \ddots \end{pmatrix} + \text{compact}$$

so the essential spectrum is $\pm i$ (since $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ has eigenvalues ± 1). This shows $d\mu$ can have finite essential spectrum without $\liminf |\alpha_n| = 1$. A similar analysis works if $\alpha_{2n} \to \zeta_1$ and $\alpha_{2n+1} \to \zeta_2$ with either $|\zeta_1| = 1$ or $|\zeta_2| = 1$. $\square$

Finally, we note that, by Theorem 1.7.18, if $\mathcal{C}^{(n)}$ is the principal $n \times n$ block of $\mathcal{C}$, then

$$\Phi_n(z) = \det(z\mathbf{1} - \mathcal{C}^{(n)}) \tag{4.2.47}$$

This implies that

$$\begin{aligned} \Phi_n^*(z) &= z^n\, \overline{\det(\tfrac{1}{\bar z} - \mathcal{C}^{(n)})} \\ &= \overline{\det(1 - \bar z \mathcal{C}^{(n)})} \\ &= \det(1 - z\overline{\mathcal{C}^{(n)}}\,) \end{aligned} \tag{4.2.48}$$

Thus, since $\log\det(A) = \operatorname{Tr}(\log A)$ (if $\|1 - A\| < 1$)

$$\log \Phi_n^*(z) = \operatorname{Tr}(\log(1 - z\,\overline{\mathcal{C}^{(n)}}\,))$$

and so, if $\sum_{n=0}^{\infty}|\alpha_n|^2 < \infty$, the Szegő function obeys

$$\log D(z) = \lim_n [-\log \varphi_n^*(z)]$$

$$= \tfrac{1}{2}\, w_0 + \sum_{n=1}^\infty z^n\, w_n \tag{4.2.49}$$

where

$$\tfrac{1}{2}\, w_0 = \log\biggl(\prod_{n=0}^\infty (1-|\alpha_n|^2)^{1/2}\biggr) \tag{4.2.50}$$

and

$$w_n = \lim_{m\to\infty} \frac{\overline{\operatorname{Tr}((\mathcal{C}^{(m)})^n)}}{n} \tag{4.2.51}$$

It is, of course, tempting to write (4.2.51) as $w_n = \overline{\operatorname{Tr}((\mathcal{C})^n/n)}$. The problem with that is that $\mathcal{C}^n$ is not trace class; indeed, it is unitary. But $\mathcal{C}_0$ is a shift operator in the standard basis renumbered, so formally, $\operatorname{Tr}((\mathcal{C}_0)^n)=0$ and at least for $n\geq 2$, $\mathcal{C}^n - \mathcal{C}_0^n$ is trace class. Moreover, while $\mathcal{C}-\mathcal{C}_0$ is not trace class if α is only in ℓ^2 and not in ℓ^1, the sum of its diagonal elements does converge.

This suggests there is a way to take the limit. In fact,

Theorem 4.2.14. *Suppose $\{\alpha_n(d\mu)\}_{n=1}^\infty$ obeys the Szegő condition*

$$\sum_{n=0}^\infty |\alpha_n|^2 < \infty \tag{4.2.52}$$

Then the Szegő function, D, obeys for $z\in\mathbb{D}$,

$$D(0)D(z)^{-1} = \det{}_2\biggl(\frac{(1-z\bar{\mathcal{C}})}{(1-z\bar{\mathcal{C}}_0)}\biggr) e^{+zw_1} \tag{4.2.53}$$

where

$$w_1 = \alpha_0 - \sum_{n=1}^\infty \alpha_n \bar\alpha_{n-1} \tag{4.2.54}$$

If

$$\sum_{n=0}^\infty |\alpha_n| < \infty \tag{4.2.55}$$

then

$$D(0)D(z)^{-1} = \det\biggl(\frac{(1-z\bar{\mathcal{C}})}{(1-z\bar{\mathcal{C}}_0)}\biggr) \tag{4.2.56}$$

The coefficients w_n of (4.2.49) *are given by*

$$w_n = \frac{\overline{\operatorname{Tr}(\mathcal{C}^n - \mathcal{C}_0^n)}}{n} \tag{4.2.57}$$

for all $n\geq 1$ if (4.2.55) *holds and for $n\geq 2$ if* (4.2.52) *holds. In all cases, one has*

$$w_n = \sum_{j=0}^\infty \frac{(\mathcal{C}^n)_{jj}}{n} \tag{4.2.58}$$

Remark. $\det_2$ is the renormalized determinant; see Subsection 1.4.14.

Proof. Let $\alpha^{(N)}$ be the sequence $(\alpha_0,\dots,\alpha_N,0,0,\dots)$ and D_N, $d_j^{(N)}$ the associated Szegő functions and terms in (4.2.49). Since $D_N(z)\to D(z)$ (i.e., $(\varphi_N^*)^{-1}\to D$) uniformly on compact subsets of $\mathbb{D}$ and all are nonvanishing in $\mathbb{D}$, the left-hand sides of (4.2.53), (4.2.54), (4.2.56), (4.2.57), and (4.2.58) converge as $N\to\infty$. The right sides also converge by the continuity properties of det, $\det_2$, and the trace ideal results for $\mathcal{C}-\mathcal{C}_0$ in Theorem 4.3.2 below. Thus it suffices to prove the results for $\alpha^{(N)}$. In that case, $\mathcal{C}-\mathcal{C}_0$ is trace class, so (4.2.53) is equivalent to (4.2.56). (4.2.56) follows from (4.2.48),

$$\Phi_n^*(z) = \frac{\Phi_n^*(z)}{\Phi_n^*(z;\alpha_n \equiv 0)} = \det\biggl(\frac{(1-z\bar{\mathcal{C}}^{(n)})}{(1-z\bar{\mathcal{C}}_0^{(n)})}\biggr)$$

and taking $n \to \infty$ using $\mathcal{C}-\mathcal{C}_0$ trace class in the $\alpha^{(N)}$ case. Since $\mathrm{Tr}((\mathcal{C}_0^{(m)})^n) = 0$ for $n > 1$, (4.2.57) holds for α^N from (4.2.51) and taking $m \to \infty$. That yields (4.2.58) in general for $n \geq 2$. The $n = 1$ case is just (4.2.54). $\square$

The quantities d_n are used in the higher-order sum rules; see Section 2.8 and [**275**].

Remarks and Historical Notes. The CMV representation is due to Cantero, Moral, and Velázquez [**181**], who also discuss $\tilde{\mathcal{C}}$ and the Θ-factorization. Their formulae look different from ours since what they call the Schur parameters are the a_n of (1.5.15), so their a_n and our α_n are related by $a_n = -\bar{\alpha}_{n-1}$. Moreover, they often write their matrices as the transpose of the ones we write.

The use of the CMV representation to prove Verblunsky's theorem (Theorem 4.2.8) and the discussion in Theorems 4.2.9 and 4.2.10 are new.

My initial exposure to OPUC was when [**464**] was submitted to me as an editor of Communications in Mathematical Physics. In looking over the introduction to that paper, I was puzzled by the use of polynomials, which I naively assumed would not usually be a basis. My immediate thought was that it seemed more natural to get a basis by orthogonalizing the set $\{1, z, z^{-1}, z^2, z^{-2}, \dots\}$. I did not pursue this idea, either then or later. I tell this story to illustrate that the CMV basis is exceedingly natural, and it is surprising that it took over eighty years from the earliest paper on the subject until CMV had the courage to follow through to the realization that the basis could be expressed in terms of the φ_n and the CMV matrix in terms of the α_n.

Prior to CMV, matrices in the block structure (4.2.1) or their analog on $\ell^2(\mathbb{Z})$ occurred in two places: in the study of the strong moment problem on the real line [**575, 588**] and in certain models in solid state physics [**128, 151**]. In particular, the $n = 1$ (tridiagonal) case of Proposition 4.2.1 is from [**151**]. In a preliminary version of this book, I presented their proof and conjectured that any four-diagonal matrix was a direct sum of blocks of size up to 3×3. Motivated by this, CMV proved Proposition 4.2.1 in [**182**]. In this paper, they also have an analog of Proposition 4.1.2 for matrices of the block form (4.2.1).

While not motivated by OPUC, Bourget, Howland, and Joye [**151**] considered products much like the ones that arise in the CMV factorization, but Θ is replaced by the more general

$$S = e^{i\eta}\begin{pmatrix} \bar{\alpha} & \rho \\ \bar{\rho} & -\alpha \end{pmatrix} \tag{4.2.59}$$

where $|\alpha|^2 + |\rho|^2 = 1$. Thus ρ is no longer real and there is the extra phase in front. S has three phases while Θ has only one. One phase in S can be removed by using the gauge transformations $\delta_n = e^{i\gamma_n}\delta_n$, but the classes studied by [**151**] have twice as many parameters and include CMV as a subset. [**151**] study mainly the generalization of the extended CMV operator, $\mathcal{E}$, but also generalizations of $\mathcal{C}$.

Theorem 4.2.11 is from Golinskii, Nevai, and Van Assche [**466**]. Theorem 4.2.12 is due to L. Golinskii [**458**]. This last paper has a complete analysis of the case where the essential spectrum has two points and, in particular, has Example 4.2.13 with a different (but not unrelated) analysis. See [**182**] for a further analysis when there are more than two points in σ_{ess} but σ_{ess} is finite.

Theorem 4.2.14 is new.

4.3. Spectral Consequences of the CMV Representation

This section is joint, previously unpublished, work with Leonid Golinskii [**467**]. We will need trace norm and operator norm estimates, but since it is easy to handle $\mathcal{I}_p$ norms, we will state the basic lemma (Theorem 4.3.2) in that context. For the properties of trace ideals, see Subsections 1.4.10–1.4.15 and [**440, 962**].

LEMMA 4.3.1. (i) *If A is a tridiagonal matrix and $1 \le p \le 2$, then*

$$\|A\|_p \le \left(\sum_{i,j} |a_{ij}|^p\right)^{1/p} \tag{4.3.1}$$

(ii) *If A is a tridiagonal matrix and $2 \le p \le \infty$, then*

$$\|A\|_p \le 3^{1-2/p}\left(\sum_{i,j} |a_{ij}|^p\right)^{1/p} \tag{4.3.2}$$

where the value of the sum, when $p = \infty$, is $\sup_{i,j}|a_{ij}|$.

(iii) *If A, B, C, D are any unitary operators, then for $1 \le p \le \infty$,*

$$\|AB - CD\|_p \le 2^{1-1/p}(\|A - C\|_p^p + \|B - D\|_p^p)^{1/p} \tag{4.3.3}$$

where, if $p = \infty$, the right side of (4.3.3) means $2\sup(\|A - C\|_\infty, \|B - D\|_\infty)$.

PROOF. (i) When $p = 1$, the result follows from the facts that $\|\cdot\|_1$ is a norm and a matrix with a single nonzero entry $a_{i_0 j_0}$ has trace norm $\|a_{i_0 j_0}\|$. For $p = 2$, the result is well-known. For general p in $(1, 2)$, we get the result by interpolation.

(ii) For $p = \infty$, it is obvious that

$$\|A\|_\infty \le \sup|a_{i,i-1}| + \sup|a_{i,i}| + \sup|a_{i,i+1}| \le 3\sup_{i,j}|a_{ij}|$$

since a diagonal operator has norm equal to the sup of its matrix elements. This proves (4.3.2) for $p = \infty$. $p = 2$ is well-known and general p is obtained by interpolation.

(iii)

$$\begin{aligned}\|AB - CD\|_p^p &\le (\|A - C\|_p\|D\|_\infty + \|A\|_\infty\|B - D\|_p)^p \\ &\le 2^{p-1}(\|A - C\|_p^p + \|B - D\|_p^p)\end{aligned}$$

by Hölder's inequality on $\mathbb{R}^2$. □

THEOREM 4.3.2. *Let $\{\alpha_j\}_{j=0}^\infty$ and $\{\beta_j\}_{j=0}^\infty$ be two sets of Verblunsky coefficients. Let $\mathcal{C}_\alpha$ and $\mathcal{C}_\beta$ be their CMV representations. Then for $1 \le p \le 2$,*

$$\|\mathcal{C}_\alpha - \mathcal{C}_\beta\|_p \le 2\left(\sum_{j=0}^\infty |\alpha_j - \beta_j|^p + |\rho_j - \sigma_j|^p\right)^{1/p} \tag{4.3.4}$$