The space $H^2(\mathbb{T})$: An Archetypal Invariant Subspace

Topics. Lebesgue spaces $L^p(\mathbb{T}, \mu)$, Hardy spaces $H^p(\mathbb{T})$, lattice of invariant subspaces, the shift operator (reducing subspaces – Wiener’s theorem – and invariant subspaces – Helson’s theorem), uniqueness theorem, and inner and outer functions.

In this chapter we mainly work in the context of the Hilbert spaces $L^2(\mathbb{T}, \mu)$, $L^2(\mathbb{T})$, $H^2(\mathbb{T})$; the other $H^p$ appear occasionally.

1.1 Notation and Terminology of Operators

Let $H$ be a Hilbert space (always over the field of complex numbers $\mathbb{C}$) and let $T: H \to H$ be a bounded linear operator on $H$. The space (the algebra) of operators on $H$ is denoted $L(H)$. Let $E \subset H$ be a subspace of $H$ (= closed linear subspace). $E$ is said to be invariant for $T \in L(H)$ if

$$x \in E \implies Tx \in E$$

(in short, $TE \subset E$). The set $\text{Lat}(T)$ of invariant subspaces is a lattice with respect to the operations $\cap$ and span (= closed linear hull). If $T$ is a family of operators on $H$, we set $\text{Lat}(T) = \bigcap_{T \in \mathcal{T}} \text{Lat}(T)$.

In the particular case of $\mathcal{T} = \{T, T^*\}$, where $T \in L(H)$ and $T^*$ is the adjoint operator of $T$ (see Appendix E), a subspace $E \in \text{Lat}(T, T^*)$ is said to be reducing.

The goal of this section is to describe the lattice $\text{Lat}(M_z)$ where $M_z$ is the operator of multiplication by an “independent variable” in the space $L^2(\mathbb{T}, \mu)$,
with \( \mu \) a finite Borel measure on the circle \( \mathbb{T} = \{ \zeta \in \mathbb{C} : |\zeta| = 1 \} \),

\[
M_z f = zf(z), \quad z \in \mathbb{T}.
\]

The operator \( M_z \) is called the bilateral shift operator.

### 1.2 Reducing Subspaces of the Bilateral Shift \( M_z \)

In the years 1920–1930, Norbert Wiener developed the mathematical theory of stationary filters. Since the tools he needed could not be found in the Analysis of the time, he created them himself, thus profoundly enriching harmonic analysis and spectral theory.
was denied a position at Harvard because of the anti-Semitic atmosphere of the establishment (George Birkhoff is often cited as one of his principal opponents, behind the scenes). Unlike other top-level scientists, Wiener was not invited to participate in the Manhattan Project. A confirmed pacifist, he systematically refused all government financing of his research after the Second World War and never participated in military projects.

In particular, for filtering theory, Wiener needed to solve the problem of the recognition (identification) of filters (see the details below in Chapter 5). As a first step, he proved the following theorem (in the case where \( \mu = m \), the normalized Lebesgue measure on the circle \( \mathbb{T} \); 80 years later, we prove it in a somewhat more general form).

**Theorem 1.2.1** (Wiener, 1932) *Let \( \mu \) be a positive Borel measure in \( \mathbb{C} \) with compact support and \( E \) a (closed) subspace of \( L^2(\mu) \). The following assertions are equivalent.*

1. \( E \in \text{Lat}(M_z, M_z^*) \).

2. There exists a Borel set \( A \subset \mathbb{C} \) such that

\[
E = \chi_AL^2(\mu) = \{ f \in L^2(\mu) : f = 0 \ \mu\text{-a.e. \ on the complement } A' = \mathbb{C} \setminus A \}.
\]

The set \( A \) in (2) is unique modulo \( \mu \): \( \chi_AL^2(\mu) = \chi_BL^2(\mu) \) if and only if \( \chi_A = \chi_B \) \( \mu\text{-a.e.} \), i.e. if and only if \( \mu(A \triangle B) = 0 \), where \( A \triangle B = (A \setminus B) \cup (B \setminus A) \) is the symmetric difference.

**Proof** First observe that \( M_z^* = M_z \) and \( \frac{1}{2}(z + \bar{z}) = X, \frac{1}{2i}(z - \bar{z}) = Y \) imply that a subspace \( E \) is reducing for \( M_z \) if and only if, for every polynomial \( p = p(X, Y) \), we have \( p \cdot E \subset E \). Let \( \mathcal{P} \) denote the set of polynomials in \( X \) and \( Y \).

Let us show (1) \( \Rightarrow \) (2). Let \( f \in E \) and \( g \in E^\perp = \{ g \in L^2(\mu) : (h, g) = 0, \forall h \in E \} \) (orthogonal complement of \( E \)). Then

\[
0 = (pf, g) = \int pf\bar{g} \, d\mu, \quad \forall p \in \mathcal{P}.
\]
The space $H^2_\mathbb{T}$: An Archetypal Invariant Subspace

Since $\mathcal{P}$ is dense in the space $C(\text{supp}(\mu))$ of continuous functions on a compact set $\text{supp}(\mu)$ (Weierstrass’s theorem), we obtain $\int g \, d\mu = 0$ (the null measure), hence $\overline{f}g = 0$ $\mu$-a.e. Then, as $L^2(\mu)$ is separable, so is $E^\perp$. By taking a sequence $(g_n)$ dense in $E^\perp$, we set

$$A = \bigcap_n Z(g_n), \quad Z(g_n) = \{z: g_n(z) = 0\}.$$  

(More rigorously, we define $Z(g_n)$ by choosing a measurable representative in the equivalence class $g_n$ of $L^2(\mu)$; another choice of representative would lead to a set $A'$ differing from $A$ only by a negligible set, hence $\chi_A = \chi_{A'}$ in the space $L^2(\mu)$.) We obtain, for any $f \in E$ and every $n$, $f g_n = 0$ $\mu$-a.e., and thus $f = 0$ a.e. on the set $\bigcup_n Z(g_n)^c = A^c$. This means that $f \in \chi_A L^2(\mu)$, and hence $E \subset \chi_A L^2(\mu)$.

Conversely, if $f \in \chi_A L^2(\mu)$, then (clearly) $f = 0$ $\mu$-a.e. on $A^c$. Since $g_n = 0$ on $A$, we have $f g_n = 0$ $\mu$-a.e., thus $(f, g_n) = 0$, $\forall n$. By the density of $(g_n)$ in $E^\perp$, we obtain $f \perp E^\perp$, hence $f \in E$. The two inclusions give $E = \chi_A L^2(\mu)$. The implication $(2) \Rightarrow (1)$ is evident.

For the uniqueness, the equality $\chi_A L^2(\mu) = \chi_B L^2(\mu)$ implies $\chi_A \in \chi_B L^2(\mu)$, thus $\chi_A = 0$ a.e. on $B^c$, meaning that $A \subset B$ up to a $\mu$-negligible set (i.e., $\mu(A \setminus B) = 0$). Similarly, $\mu(B \setminus A) = 0$, which completes the proof. ■

1.3 Non-reducing Subspaces of the Bilateral Shift $M_z$

In order to catalog the non-reducing subspaces of $M_z$, we use two related (but not coincident) orthogonal decompositions. The first is given by Lemma 1.3.1 below and concerns an invariant subspace of an arbitrary operator. The second is the Radon–Nikodym decomposition (see Appendix A)

$$L^2(\mu) = L^2(\mu_a) \oplus L^2(\mu_s),$$

where $\mu$ is a Borel measure on the circle $\mathbb{T}$, and $\mu_a$, $\mu_s$ denote, respectively, the absolutely continuous and singular components of $\mu$ with respect to the normalized Lebesgue measure $m$, $m[e^{it}: \theta_1 \leq t \leq \theta_2] = (\theta_2 - \theta_1)/2\pi \leq 1$.

**Lemma 1.3.1** Let $T: H \to H$ be a bounded linear operator on a Hilbert space $H$ and let $E \subset H$ be a closed subspace.

1. $E \in \text{Lat}(T) \iff E^\perp \in \text{Lat}(T^*)$.
2. $E \in \text{Lat}(T, T^*) \iff E \in \text{Lat}(T), E^\perp \in \text{Lat}(T)$.
3. For every $E \in \text{Lat}(T)$, $E = E_R \oplus E_N$.

https://doi.org/10.1017/9781316882108.003 Published online by Cambridge University Press
where \( E_R \in \text{Lat}(T, T^*) \) (a reducing subspace of \( T \)) and \( E_N \in \text{Lat}(T) \) is a completely non-reducing subspace, i.e. such that \( E' \subset E_N \), \( E' \in \text{Lat}(T, T^*) \Rightarrow E' = \{0\} \). This representation is unique.

**Proof** (1) We first show the implication “\( \Rightarrow \)”. Let \( y \in E^\perp \). Then, \((T^*y, x) = (y, Tx) = 0\) for every \( x \in E \), and hence \( T^*y \in E^\perp \). It ensues that \( T^*E^\perp \subset E^\perp \).

The implication “\( \Leftarrow \)” is immediate since \( T = (T^*)^* \).

(2) It is immediate by (1) since \( T = (T^*)^* \).

(3) Clearly the “span” (closed linear hull) of a family of reducing subspaces is still in \( \text{Lat}(T, T^*) \). Set \( E_R = \text{span}(E' : E' \subset E, E' \in \text{Lat}(T, T^*)) \), \( E_N = E \ominus E_R \).

Then \( E = E_R \oplus E_N \) and \( E_R \in \text{Lat}(T, T^*) \). Moreover, \( E_N = E \cap (E_R)^\perp \) and hence, by (1), \( E_N \in \text{Lat}(T) \). If \( E' \subset E_N \) and \( E' \in \text{Lat}(T, T^*) \), then \( E' \subset E_R \) by the definition of the latter. Thus \( E' = \{0\} \). The uniqueness is also immediate. ■

**Lemma 1.3.2** Let \( \mu \) be a finite Borel measure on \( \mathbb{T} \), with \( \mu = \mu_a + \mu_s = w \cdot m + \mu_s \) its Radon–Nikodym decomposition (see Appendix A), and let \( E \subset L^2(\mu) \) be a NON-reducing invariant subspace of \( M_z : L^2(\mu) \rightarrow L^2(\mu) \). Then:

1. There exists a function \( q \in E \) such that \( |q|^2 w = 1 \) m-a.e.
2. \( E_R \subset L^2(\mu_s) \), where \( E_R \) is the reducing part of \( E \) according to Lemma 1.3.1.

**Proof** (1) Our subspace \( E \) satisfies the properties \( M_z E \subset E, M_z E \neq E \); indeed, if we had \( M_z E = E \), then \( M_z E = M_z M_z E = E \), hence \( E \in \text{Lat}(M_z, M_z^*) \) which is not the case. Moreover, \( M_z \) is an isometric (and even unitary) operator, and thus the image \( M_z E \) is closed. Let

\[
q \in E \oplus M_z E = E \cap (M_z E)^\perp, \quad ||q|| = 1.
\]

Since \( q \in E \) and \( M_z^n q \in M_z E \) for all \( n \geq 1 \), we obtain

\[
0 = (z^n q, q) = \int_{\mathbb{T}} z^n q \bar{q} d\mu = \int_{\mathbb{T}} z^n |q|^2 d\mu, \quad n \geq 1.
\]

We conclude, by complex conjugation, that all the Fourier coefficients of the measure \( |q|^2 d\mu \), except for one, are zero, and hence there exists a constant \( c \) such that \( (|q|^2 d\mu)(n) = c \hat{m}(n) \) for all \( n, n \in \mathbb{Z} \). By the theorem of uniqueness (see Appendix A), \( |q|^2 d\mu = m \) (\( c = 1 \) by the normalization
\[ \|q\| = 1. \] Thus, \(|q|^2 d\mu_a + |q|^2 d\mu_s = m, \] and by the uniqueness of the Radon–Nikodym decomposition \( m = |q|^2 d\mu_a = |q|^2 w, \] which is equivalent to \( |q|^2 w = 1 \) m-a.e.

(2) Let \( f \in E_R. \) Given that \( E_R \) is reducing and \( M^\ast R = M_T = M_T^{-1}, \) we have \( \tau^n f \in E_R \subset E \) for all \( n \in \mathbb{Z}. \) Then \( \tau^n f = \tau(\tau^{n-1} f) \in \tau E, \) and by the definition of \( q \) we obtain

\[ 0 = (\tau^n f, q) = \int_T \tau^n f \overline{q} \, d\mu, \quad \forall n \in \mathbb{Z}. \]

Therefore, \( f \overline{q} = 0 \) \( \mu \)-a.e., hence \( \mu_a, \mu_s \)-a.e., and thus (given that \( m = |q|^2 d\mu_a \) \( f \overline{q} = 0 \) \( m \)-a.e. However \( q \neq 0 \) \( m \)-a.e., hence \( f = 0 \) \( m \)-a.e., which means \( f \in L^2(\mu_s). \) We thus obtain \( E_R \subset L^2(\mu_s). \)

**Corollary 1.3.3** Every invariant subspace of \( L^2(\mu) \) contained in \( L^2(\mu_s) \) is reducing and can be written \( E = \chi_A L^2(\mu_s) \) with \( A \) a Borel set.

Indeed, if \( E \) were not reducing, it would contain a function \( q \) satisfying \( |q|^2 \neq 0 \) \( m \)-a.e., which is impossible. \( \blacksquare \)

**Definition 1.3.4** (the space \( H^2(\mathbb{T}) \), the generic non-reducing subspace) Let \( L^2(\mathbb{T}) = L^2(\mathbb{T}, m) \) (normalized Lebesgue measure). The Hardy space \( H^2(\mathbb{T}) \) is defined as the following subspace of \( L^2(\mathbb{T}) \):

\[ H^2(\mathbb{T}) = \{ f \in L^2(\mathbb{T}) : \hat{f}(n) = 0 \text{ for all integers } n < 0 \}. \]

*Reminder* The exponentials \((\tau^n)_{n \in \mathbb{Z}} = (e^{int})_{n \in \mathbb{Z}}\) form an orthonormal basis of the space \( L^2(\mathbb{T}) \), and hence every function \( f \in L^2(\mathbb{T}) \) is the sum of its Fourier series

\[ f = \sum_{n \in \mathbb{Z}} \hat{f}(n) \tau^n, \]

norm-\(L^2(\mathbb{T})\) convergent for the symmetric partial sums \( \sum_{n=-N}^N \hat{f}(n) \tau^n \) (for \( N \to \infty \)), or even for “disordered” sums \( \sum_{n \in \sigma(N)} \hat{f}(n) \tau^n \) where \( \sigma(N) \subset \mathbb{Z}, \sigma(N) \not\subset \mathbb{Z} \) for \( N \to \infty \):

\[ \lim_{N} \left\| f - \sum_{n \in \sigma(N)} \hat{f}(n) \tau^n \right\|_{L^2(\mathbb{T})} = 0. \]

With this reminder, we can say

\[ H^2(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) : f = \sum_{n \geq 0} \hat{f}(n) \tau^n \right\} = \left\{ \sum_{n \geq 0} a_n \tau^n : \sum_{n \geq 0} |a_n|^2 < \infty \right\}. \]

Moreover, the use of properties of orthogonal bases leads to

\[ H^2(\mathbb{T}) = \text{span}_{L^2(\mathbb{T})} (\tau^n : n = 0, 1, \ldots). \]
1.3 Non-reducing Subspaces of the Bilateral Shift $M_z$

By the above, clearly

$$M_zH^2(\mathbb{T}) \subset H^2(\mathbb{T})$$

and $M_zH^2(\mathbb{T}) \neq H^2(\mathbb{T})$ (hence, $H^2(\mathbb{T})$ is a non-reducing invariant subspace). Moreover, $H^2(\mathbb{T})$ is completely non-reducing since for every $f \in H^2(\mathbb{T}), f \neq 0$, there exists a positive integer $n$ such that $M_z^n f = \overline{f} \neq H^2(\mathbb{T})$. The following theorem shows that this is a generic example: any other completely non-reducing subspace coincides with $H^2(\mathbb{T})$ up to a factor of correction. This was proved in the 1960s by Henry Helson, professor at the University of California (Berkeley).

**Henry Helson** (1927–2010), one of the primary experts of his generation in harmonic analysis, was a professor at the University of California (Berkeley), 1955–2010. His work on Hardy spaces and “abstract Hardy spaces” (1960–1965, in collaboration with David Lowdenslager, a mathematician from Yale), as well as his perfectly written research monographs (such as *Lectures on Invariant Subspaces* (Helson, 1964)) profoundly influenced the development of the subject. A rich personality with an extraordinary range of talent (in particular, he was a violinist and cellist at a professional level), he had a truly singular career: as a dedicated Quaker, he turned down a position in California in 1948, because he refused to take a “loyalty oath” (mandatory in California in the era of McCarthyism), and left for Europe where he continued his studies in Poland (with Szpilrajn), then in Sweden (with Beurling) and in France, at Nancy (with Schwartz, Dieudonné, Godement, and Grothendieck).

**Theorem 1.3.5** (Helson, 1964) Let $\mu$ be a finite Borel measure on $\mathbb{T}$, $\mu = \mu_a + \mu_s = w\mu + \mu_s$, and let $E \subset L^2(\mu)$ be an invariant subspace of $M_z$. Then:

1. either $M_zE = E$, and then $E = \chi_A L^2(\mu)$ with $A$ a Borel set,
2. or $M_zE \neq E$, and then $E = \chi_A L^2(\mu_s) \oplus qH^2(\mathbb{T})$, where $|q|^2w = 1$ m.a.e. and $A$ is a Borel set; such a function $q$ is unique, and so is $A$ (meaning $q = \lambda q'$ with $\lambda \in \mathbb{T}$ and $\chi_A = \chi_A'$ $\mu_s$-a.e.).

Conversely, each $A$ and $q$ satisfying this equation generate a reducing subspace by the formula $E = \chi_A L^2(\mu)$, and a non-reducing subspace by $E = \chi_A L^2(\mu_s) \oplus qH^2(\mathbb{T})$. The latter subspace is completely non-reducing if and only if $\chi_A L^2(\mu_s) = \{0\}$ ($\iff \chi_A = 0 \mu_s$-a.e.).

**Proof** (1) This is Wiener’s Theorem 1.2.1.
The space $H^2(\mathbb{T})$: An Archetypal Invariant Subspace

(2) Let $E \in \text{Lat}(M_z)$, $E \neq M_zE$. By Lemma 1.3.2, there exists a function $q \in E \cap L^2(\mu_a)$, $q \perp M_zE$ such that $|q|^2w = 1$ $m$-a.e. The sequence $(z^nq)_{n \in \mathbb{Z}}$ is orthonormal:

$$(z^nq, z^m q) = \int_{\mathbb{T}} z^{n-m}|q|^2 \, dm = \int_{\mathbb{T}} z^{n-m} |q|^2 \, w \, dm = \int_{\mathbb{T}} z^{n-m} \, dm = \delta_{m,n},$$

where $\delta_{m,n}$ is the Kronecker delta ($= 0$ if $m \neq n$, and = 1 if $m = n$). Consequently,

$$\text{span}_{L^2(\mu)} (z^nq: n \geq 0) = \left\{ \sum_{n \geq 0} a_n z^n q : \sum_{n \geq 0} |a_n|^2 < \infty \right\}.$$ 

Moreover, clearly the mapping $U: f \mapsto qf$ is a linear isometry of $L^2(\mathbb{T}) = L^2(\mathbb{T}, m) \rightarrow L^2(\mathbb{T}, \mu_a)$:

$$\|Uf\|_{L^2(\mu_a)}^2 = \int_{\mathbb{T}} |f|^2 |q|^2 \, w \, dm = \int_{\mathbb{T}} |f|^2 \, dm = \|f\|_{L^2(\mathbb{T})}^2.$$ 

Hence $\text{span}_{L^2(\mu)} (z^nq: n \geq 0) = U(\mathcal{H}^2(\mathbb{T})) = q\mathcal{H}^2(\mathbb{T}) \subset E$. Let $E = E' \oplus q\mathcal{H}^2(\mathbb{T})$ where $E' = E \cap (q\mathcal{H}^2(\mathbb{T}))^\perp$ (orthogonal complement in $L^2(\mu_a)$). For an arbitrary function $f \in E' \subset E$, we have

$$\int_{\mathbb{T}} z^n f \overline{q} \, dm = (q, z^n f) = 0 \text{ for } n \geq 1, \text{ and } \int_{\mathbb{T}} f \overline{z^n q} \, dm = (f, qz^n) = 0 \text{ for } n \geq 0,$$

so $f \overline{q} \, dm = 0$, and hence $f \overline{q} = 0$ $\mu$-a.e. However $q \neq 0$ $\mu_a$-a.e., thus $f = 0$ $\mu_a$-a.e., and then $f \in L^2(\mu_s)$. We have shown that $E' \subset L^2(\mu_s)$, and – because the converse $E \cap L^2(\mu_s) \subset E'$ is clear – $E' = E \cap L^2(\mu_s)$. As both $E$ and $L^2(\mu_s)$ are $M_z$-invariant, then Corollary 1.3.3 leads to $E' = \chi_A L^2(\mu_s)$.

For the uniqueness, let $\chi_A L^2(\mu_s) \oplus q\mathcal{H}^2(\mathbb{T}) = \chi_A L^2(\mu_s) \oplus q'\mathcal{H}^2(\mathbb{T})$ where $|q'|^2w = 1$ $m$-a.e. Then clearly $\chi_A L^2(\mu_s) = \chi_A L^2(\mu_s)$ and $q\mathcal{H}^2(\mathbb{T}) = q'\mathcal{H}^2(\mathbb{T})$, hence $\chi_A = \chi_A' \mu$-a.e. The second equation implies $q/q' \in \mathcal{H}^2(\mathbb{T})$ and $q'/q \in \mathcal{H}^2(\mathbb{T})$, and since $|q| = |q'|$ $m$-a.e., all the Fourier coefficients of $q/q'$, with the exception of $(q/q')(0)$, are zero. Hence $q/q' = \text{constant} = \lambda$; clearly $|\lambda| = 1$.

The rest of the statement is also evident.

**Corollary 1.3.6** The space $L^2(\mathbb{T}, \mu)$ contains a non-reducing invariant subspace $E$ (i.e. $M_zE \subset E$, $M_zE \neq E$) if and only if $m \ll \mu$ (i.e. $w > 0$ $m$-a.e. on $\mathbb{T}$).

Indeed, according to Theorem 1.3.5(2), it is a question of the existence of a function $q$ such that $|q|^2w = 1$ $m$-a.e., which is equivalent to the condition of the corollary.
1.3 Non-reducing Subspaces of the Bilateral Shift $M_z$

1.3.1 $H^p(\mathbb{T})$ Spaces

Let $1 \leq p \leq \infty$. The Hardy space $H^p(\mathbb{T})$ is defined similarly to the space $H^2(\mathbb{T})$

$$H^p(\mathbb{T}) = \{ f \in L^p(\mathbb{T}) : \hat{f}(n) = 0 \text{ for any integer } n < 0 \}.$$

The $H^p(\mathbb{T})$ spaces share many of the properties of the space $H^2(\mathbb{T})$, but of course not all of them, and always after some modifications. For example, the exponentials $(z^n)_{n \in \mathbb{Z}}$ no longer form an unconditional basis in $L^p(\mathbb{T})$, $p \neq 2$, but form a Schauder basis for $1 < p < \infty$, as will be seen in Chapter 2, Exercise 2.8.4(f).

Here, we limit ourselves to a short list of initial properties. For more information, see Chapter 2, in particular Exercise 2.8.1 (concerning an analog in $L^p(\mathbb{T})$ of the theorems of Beurling and Helson).

1. $H^p(\mathbb{T})$ is a closed vector subspace of $L^p(\mathbb{T})$.

2. If $f, \overline{f} \in H^p(\mathbb{T})$, then $f = \text{constant}$.

Indeed, all the Fourier coefficients of $f$ are zero, except perhaps $\hat{f}(0)$. ■

3. If $f \in H^p(\mathbb{T})$ and $f = 0$ on a set $A \subset \mathbb{T}$, with $m(A) > 0$, then $f = 0$.

For a proof, see Corollary 2.3.3 below.

4. Let

$$H^p_0(\mathbb{T}) = \{ f \in L^p(\mathbb{T}) : \hat{f}(n) = 0 \text{ for every integer } n \geq 0 \}.$$

Then, $H^p(\mathbb{T}) \cap H^p_0(\mathbb{T}) = \{0\}$ and, if $p < \infty$, $\text{clos}_{L^p}(H^p(\mathbb{T}) + H^p_0(\mathbb{T})) = L^p(\mathbb{T})$.

(For $p = \infty$ see Exercise 2.8.4(i). In fact, for $1 < p < \infty$, the sum is already closed, by Marcel Riesz’s Theorem 2.8.4(e)).

Indeed, the first equation holds for the reason used for (2), the second because $H^p(\mathbb{T}) + H^p_0(\mathbb{T}) \supset \mathcal{P}$. ■

5. The invariant subspaces of $L^p(\mathbb{T})$ were described by Srinivasan (1963): let $E$ be a subspace invariant under the shift operator $M_z : L^p(\mathbb{T}) \to L^p(\mathbb{T})$. Then:

(a) either $zE = E$, and then $E = \chi_A L^p(\mathbb{T})$ for some Borel set $A$,

(b) or $zE \neq E$, and then there exists $q$, measurable on $\mathbb{T}$, with $|q| = 1$ m-a.e., such that $E = qH^p(\mathbb{T})$.

The parameters $A$ and $q$ are uniquely defined by $E$ in the same sense as in Theorem 1.3.5.
For the proof, see Exercise 2.8.1 where this analog of Theorem 1.3.5 is a corollary of a more general proposition.

### 1.4 Beurling “Inner Functions”

The special case $\mu = m$ is particularly important.

**Corollary 1.4.1** (invariant subspaces of $L^2(\mathbb{T})$) Let $E$ be a subspace invariant under the shift operator $M_z: L^2(\mathbb{T}) \to L^2(\mathbb{T})$. Then:

1. either $zE = E$, and then $E = \chi_A L^2(\mathbb{T})$ for some Borel set $A$,
2. or $zE \neq E$, and then there exists a function $q \in L^\infty(\mathbb{T})$, with $|q| = 1$ m-a.e., such that $E = qH^2(\mathbb{T})$. The parameters $A$ and $q$ are uniquely defined by $E$ in the same sense as in Theorem 1.3.5.

Indeed, this is Theorem 1.3.5 with $\mu_s = 0$ and $w = 1$. ■

The following even more specialized case, called “Beurling’s theorem,” is important for not only its consequences, but also for its role in the development of the theory of Hardy spaces. Even though the proof given below is totally different from the original proof, we still need the following definition introduced by Beurling: that of a special class of “inner functions” in $H^2(\mathbb{T})$ which, today, plays a fundamental role in the entire theory. See also the historical remarks in the biographical sketch below, and in §§ 1.9, 2.9, 3.6.

---

**Arne Beurling** (1905–1986), was a Swedish mathematician, the author of numerous remarkable works in mathematical analysis and cryptography, and a professor (1937–1954) at the University of Uppsala (a university founded in 1477, where Carl Linnaeus and Anders Celsius worked), and later at the Institute for Advanced Study at Princeton, USA. Simultaneously with Gelfand, he discovered the fundamental principles of Banach algebras and introduced an important class of weighted algebras (“Beurling algebras”), described the invariant subspaces of the isometric shift operator, and (with Malliavin) resolved the problem of the completeness radius of families of exponentials by proving the “multiplier theorem,” important for the uncertainty principle in harmonic analysis. His doctoral students include Carleson, Domar, Esseen, Hall, and Nyman.
Beurling is also famous for having single-handedly (in 1940) deciphered the German Nazi secret code known as *Geheimfernschreiber* (“secret teleprinter”), based on a machine that could create $10^{18}$ different combinations (many more than the Enigma machine, famous for its role in Operation Overlord!). This feat allowed the Swedish secret service to systematically decipher coded messages that were passing through Sweden via a cable linking Norway with Nazi Germany. The invasion plan Barbarossa and the date at which it was to start (June 22, 1941) were intercepted and communicated to the Soviets, but they did not believe the information as its source was not revealed.

**Definition 1.4.2** (Beurling inner functions) A function on the circle $\mathbb{T}$ is said to be inner (in the sense of Beurling) if

$$\varphi \in H^2(\mathbb{T}) \text{ and } |\varphi| = 1 \text{ m-a.e.}$$

**Corollary 1.4.3** (Beurling’s Theorem, 1949) Let $E \subset H^2(\mathbb{T}) \subset L^2(\mathbb{T})$ be a subspace of $H^2(\mathbb{T})$, $E \neq \{0\}$. Then, $E$ is $M_z$-invariant if and only if there exists an inner function $q$ such that

$$E = qH^2(\mathbb{T}).$$

There is a bijective correspondence between $\text{Lat}(M_z|H^2(\mathbb{T}))$ and the set of inner functions $q$ whose first non-zero Fourier coefficient is positive.

Indeed, by applying Corollary 1.4.1, we see that case (1) is impossible: if $zE = E$, then we would have $\bar{z}^n f \in E \subset H^2(\mathbb{T})$ for all $n \geq 0$ and every function $f \in E$; however if $f \neq 0$, $f = \sum_{k \geq n} a_k z^k$ with $a_n \neq 0$, we obtain $\bar{z}^{n+1} f = a_n \bar{z} + \sum_{k > n} a_k z^{k-n-1}$ and hence $\bar{z}^{n+1} f \not\in H^2(\mathbb{T})$. Consequently, $zE \neq E$. In case (2) of Corollary 1.4.1, we have $E = qH^2(\mathbb{T}) \subset H^2(\mathbb{T})$, thus $q \in H^2(\mathbb{T})$, and the result follows.

**Corollary 1.4.4** (boundary uniqueness theorem) If $f \in H^2(\mathbb{T})$ and $f = 0$ on a set $A \subset \mathbb{T}$ such that $m(A) > 0$, then $f = 0$.

Indeed, let

$$E_f := \text{span}_{L^2(\mathbb{T})}(\bar{z}^n f: n \geq 0) = \text{clos}_{L^2(\mathbb{T})}(f\mathcal{P}_a),$$

the smallest $M_z$-invariant subspace containing $f$, where $\mathcal{P}_a$ is the space of analytic polynomials,

$$\mathcal{P}_a := \mathcal{P} \cap H^2(\mathbb{T}).$$
If we suppose $f \neq 0$, then by Corollary 1.4.3, $E_f = qH^2(\mathbb{T})$ with an inner function $q$ (hence, $|q| = 1 \text{ m-a.e.}$). In particular, $q \in E_f$, which is impossible, since for any polynomial $p$,

$$\|q - pf\|^2 \geq \int_A |q - pf|^2 \, dm = \int_A |q|^2 \, dm = m(A) > 0,$$

thus a contradiction. 

Furthermore, with regard to the uniqueness theorem (proved by Frigyes and Marcel Riesz in 1916: see the biographical sketch in § 1.5), we can add that in Chapter 3 a more complete (even definitive) description of the subject will be presented. Finally, note that numerous examples of inner functions are known (see Exercises § 1.8.3), and better still, that there exists an intelligible description of all inner functions. This was well known long before Beurling’s theorem (see §§ 1.9, 2.9 for details).

### 1.5 $H^2(\mu)$ Spaces and the Riesz Brothers’ Theorem

We begin with the definition of the space $H^2$ associated with an arbitrary Borel measure $\mu$ on the circle $\mathbb{T}$ (in place of the Lebesgue measure $m$) and the Radon–Nikodym decomposition lemma of invariant subspaces.

**Definition 1.5.1** Let $\mu$ be a finite measure on $\mathbb{T}$. The Hardy space associated with $\mu$ is defined by

$$H^2(\mu) := \text{span}_{L^2(\mu)}(\{z^n : n \geq 0\}) = \text{clos}_{L^2(\mu)} \mathcal{P}_a,$$

where $\mathcal{P}_a = \text{Lin}(\{z^n : n \geq 0\})$ is again the space of analytic polynomials. Clearly, $H^2(m) = H^2(\mathbb{T})$.

**Lemma 1.5.2** Let $\mu$ be a (finite) Borel measure on $\mathbb{T}$. Then:

1. For every $E \in \text{Lat}(M_z)$, with $M_z : L^2(\mu) \to L^2(\mu) = L^2(\mu_a) \oplus L^2(\mu_s)$, we have

$$E = E_a \oplus E_s$$

where $E_a = E \cap L^2(\mu_a)$, $E_s = E \cap L^2(\mu_s)$.

2. $H^2(\mu) = H^2(\mu_a) \oplus L^2(\mu_s)$.

**Proof** (1) By Helson’s Theorem 1.3.5, either $E = \chi_A L^2(\mu) = \chi_A L^2(\mu_a) \oplus \chi_A L^2(\mu_s)$, or $E = \chi_A L^2(\mu_s) \oplus qH^2(\mathbb{T})$ and $qH^2(\mathbb{T}) \subset L^2(\mu_a)$, and the result follows.
(2) Since $H^2(\mu) \in \text{Lat}(M_\mu)$, by (1) we have $E := H^2(\mu) = E_\alpha \oplus E_\sigma$. Then by Corollary 1.3.3, $E_\sigma = \chi_A L^2(\mu_\sigma)$ with a Borel set $A$. As $1 \in H^2(\mu)$, we have $\chi_A = 1 \mu_\sigma$-a.e., and hence $E_\sigma = L^2(\mu_\sigma)$.

Moreover, by writing $1 = 1_\alpha \oplus 1_\sigma \in E_\alpha \oplus E_\sigma$, we obtain $1_\alpha \in E_\alpha$ ($1_\alpha = 1 \mu_\alpha$-a.e.), and hence $H^2(\mu_\alpha) \subset E_\alpha$. However the reverse inclusion is evident, since for every $f \in E_\alpha$ and every sequence of polynomials $p_n \in P_\alpha$ converging to $f$, we have

$$ \|f - p_n\|_{L^2(\mu_\alpha)}^2 \leq \|f - p_n\|_{L^2(\mu_\alpha)}^2 + \|p_n\|_{L^2(\mu_\sigma)}^2 = \|f - p_n\|_{L^2(\mu)}^2 \to 0, $$

hence $f \in H^2(\mu_\alpha)$.

**Remark 1.5.3** What equality (2) in the lemma means is that there is a simultaneous polynomial approximation: $\forall f \in H^2(w \cdot m)$, $\forall g \in L^2(\mu_\sigma)$, there exists a sequence of polynomials $(p_n) \subset P_\alpha$ such that, simultaneously, $p_n \to f$ (in $L^2(\mu_\alpha)$) and $p_n \to g$ (in $L^2(\mu_\sigma)$).

The following theorem (usually called the Riesz Brothers Theorem) is a cornerstone in the construction of Hardy spaces and of harmonic analysis on the circle $\mathbb{T}$ (moreover, there exist analogs of this statement for other groups such as $\mathbb{R}$, $\mathbb{T}^n$, $\mathbb{R}^n$; see § 1.9). *A priori*, this is somewhat unexpected: certain restrictions on the Fourier spectrum $\sigma_F(\mu)$ of a complex measure $\mu$, where

$$ \sigma_F(\mu) = \text{supp}(\hat{\mu}) = \{ n \in \mathbb{Z} : \hat{\mu}(n) \neq 0, \hat{\mu}(n) = \int_\mathbb{T} z^n d\mu, $$

imply consequences on the size of the (Borel) support of $\mu$.

**Theorem 1.5.4** (Riesz and Riesz, 1916) *Let $\mu$ be a complex Borel measure on $\mathbb{T}$, assumed “analytic,” i.e. its Fourier coefficients of negative index are zero:*

$$ \hat{\mu}(-n) := \int_\mathbb{T} z^n d\mu = 0, \quad n \geq 0. $$

*Then, $\mu \ll m$ (\(m\) is absolutely continuous with respect to \(m\)) and $\mu = h m$ with $h \in H^1_0$, where

$$ H^1_0 := \{ f \in L^1(\mathbb{T}) : \hat{f}(k) = 0 \text{ for } k \leq 0 \}. $$

**Proof** Let $|\mu|$ be the variation of the measure $\mu$ (see Appendix A). Clearly $\mu \ll |\mu|$; let $\epsilon$ be the corresponding Radon–Nikodym derivative: $\mu = \epsilon|\mu|$. It is well known that $|\epsilon| = 1 |\mu|$-a.e. (see Appendix A). Since $\int_\mathbb{T} z^n d\mu = \int_\mathbb{T} z^n \epsilon d|\mu|$, the hypothesis on $\mu$ means that $\epsilon \perp H^2(|\mu|)$ in the space $L^2(|\mu|)$. However, by Lemma 1.5.2, $H^2(|\mu|) = H^2(|\mu_\alpha|) \oplus L^2(|\mu_\sigma|)$, which implies $\epsilon \perp L^2(|\mu_\sigma|)$, and hence $\epsilon = 0 |\mu_\sigma|$-a.e. Since, at the same time, $|\epsilon| = 1 |\mu_\sigma|$-a.e., we obtain...
The measure $\mu = \epsilon|\mu| = \epsilon|\mu|_{a}$ is absolutely continuous with respect to $m$, and hence by Radon–Nikodym there exists $h \in L^1(m)$ such that $\mu = hm$. Clearly the hypothesis on $\mu$ can be translated to $h \in H^1_0$. ■

The brothers **Frigyes (Frédéric)** and **Marcell (Marcel) Riesz** were two pillars of analysis in the twentieth century. They founded various domains of analysis, thus offering a rare example of familial scientific endeavor at such a high level. The elder, Frigyes Riesz (1880–1956), laid the foundations of functional analysis and operator theory as separate disciplines (1910); he was strongly influenced by the ideas of Fréchet, Lebesgue, and Hilbert. The representation theorem of linear functionals, as well as the *Riesz–Fischer theorem*, bear his name. He also founded the János Bolyai Mathematical Institute and the journal *Acta Scientiarum Mathematicarum* (Szeged), and with his student Béla Sz.-Nagy co-authored an influential text, *Leçons d’analyse fonctionelle*.

Marcel Riesz (1886–1969) spent (almost) all of his career at the University of Lund (Sweden). His contribution to analysis was enormous: his discoveries include *Riesz transformation*, the *Riesz potential*, the *Riesz (–Bochner) mean*, and the *Riesz–Thorin theorem*. Curiously, in his search for a permanent position, he was classed in second place twice in a row (for different positions), each time behind Torsten Carleman. His doctoral students included Thorin, Cramér, Hille, Frostman, and Hörmander.

Frigyes and Marcel Riesz wrote only one article together (Riesz and Riesz, 1916): it contains the *Riesz brothers’ theorem*, which subsequently became so important for harmonic analysis and its applications.

Furthermore, with regard to Theorem 1.5.4, it can be mentioned that the original proof was much more complicated than that above; however, our proof depends on invariant subspaces and on some of the already developed theory, hence it is indirect. An alternative proof is presented below, which is completely elementary and depends only on the definition of absolutely continuous measures.

### 1.5.1 Elementary Proof of Theorem 1.5.4 (Øksendal, 1971)

First note that the hypothesis on $\mu$ implies $\int_{\mathbb{R}} p \, d\mu = 0$ for every $p \in \mathcal{P}_a$, and then $\int_{\mathbb{R}} f \, d\mu = 0$ for every function $f$ defined and holomorphic in a
disk \((1 + \epsilon)D = D(0, 1 + \epsilon), \epsilon > 0\): indeed such a function \(f\) is the sum of a power series normally convergent on \(T\), \(f(z) = \sum_{k \geq 0} \hat{f}(k)z^k\) (with radius of convergence \(\geq 1 + \epsilon\)), hence the series can be integrated term by term.

In particular, for every rational function \(f = p/q (p, q \in \mathcal{P}_a)\) having poles (zeros of the denominator \(q\)) in \(\mathbb{C} \setminus \overline{D}\), we have

\[
\int_T f \, d\mu = 0.
\]

By the definition of \(\mu \ll m\), we must show that, for every Borel set \(A \subset \mathbb{T}\), \(m(A) = 0 \Rightarrow \mu(A) = 0\). In fact, it is sufficient to do this only for a closed \(A = F\), because of the regularity of the variation \(|\mu|\) (see Appendix A).

So, let \(F \subset \mathbb{T}\), with \(F = \overline{F}, m(F) = 0\). We are going to construct a sequence of rational functions \((h_n)\) such that

1. \(|h_n(z)| \leq 2\) on \(\mathbb{T}\),
2. \(\lim_n h_n(z) = \chi_F(z)\) for all \(z \in \mathbb{T}\).

Then, by the dominated convergence theorem,

\[
0 = \int_T h_n \, d\mu \to \int_T \chi_F \, d\mu = \mu(F) \quad \text{when} \ n \to \infty,
\]

thus \(\mu(F) = 0\), which will complete the proof.

Construction of a sequence \((h_n)\): since \(m(F) = 0\), for every \(n \geq 1\), there exist disks \(D(z_i, r_i), i = 1, \ldots, N\), such that

\[
z_i \in F \subset \mathbb{T}, \quad F \subset \bigcup_i D(z_i, r_i), \quad \text{and} \quad \sum_{i=1}^N r_i < \frac{1}{n^2}.
\]

Set

\[
f_n = \prod_{i=1}^N \frac{z - z_i}{z - z_i - nr_i z_i}.
\]

The \(f_n\) satisfy the following properties.

1. The functions \(f_n\) are rational, with poles \(z = (1 + nr_i)z_i\) in \(\mathbb{C} \setminus \overline{D}\), hence

\[
\int_T f_n \, d\mu = 0.
\]
2. For \(z \in \mathbb{T}\), by elementary geometry we have \(|z - z_i| < |z - z_i(1 + nr_i)|\), thus \(|f_n(z)| < 1\).
The space \( H^2(\mathbb{T}) \): An Archetypal Invariant Subspace

(3) For \( z \in \mathbb{T} \cap D(z_i, r_i) \), we obtain \( |z - z_i| < r_i \) and
\[
|z - z_i(1 + nr_i)| > |z_i|nr_i - |z - z_i| > nr_i - r_i = (n - 1)r_i, \text{ hence}
\]
\[
\frac{|z - z_i|}{|z - z_i - z_i nr_i|} < \frac{r_i}{(n - 1)r_i} = \frac{1}{n - 1}.
\]
However, the other factors of \( f_n \) are bounded above by 1 (see (2)): hence
\[
|f_n(z)| < 1/(n - 1) \text{ for every point of } F.
\]
(4) For \( z \in \mathbb{T} \setminus F \), let \( d = \text{dist}(z, F) \); we have \( |z - z_i| \geq d > 0 \) for every \( i \).
Writing
\[
\frac{1}{f(z)} = \prod_{i=1}^{N} \left(1 - \frac{nr_iz_i}{z - z_i}\right) = \exp \left\{ \sum_{i=1}^{N} \log \left(1 - \frac{nr_iz_i}{z - z_i}\right) \right\},
\]
we observe that for \( n > 2/d \),
\[
\left|\frac{nr_iz_i}{z - z_i}\right| \leq \frac{nr_i}{d} \leq \frac{1}{nd} < \frac{1}{2},
\]
and by using \( |\log(1 - w)| \leq 2|w| \) for \( |w| \leq 1/2 \), we obtain
\[
\left| \sum_{i=1}^{N} \log \left(1 - \frac{nr_iz_i}{z - z_i}\right) \right| \leq 2 \sum_{i=1}^{N} \frac{nr_i}{|z - z_i|} \leq \frac{2n}{d} \sum_{i=1}^{N} r_i < \frac{2}{dn},
\]
and hence \( \lim_n(1/f_n(z)) = 1 \).

In conclusion, the functions \( h_n = 1 - f_n \) satisfy all the required properties, which completes the proof.

1.6 The Past and the Future: The Prediction Problem

The problems of prediction, prognosis, and extrapolation of stochastic (random) processes have played an extraordinary role in the history of Hardy spaces.

**Definition 1.6.1** A discrete time stationary process (also known as a stationary sequence) is a sequence \( (x_n)_{n \in \mathbb{Z}} \) in a Hilbert space \( H \) such that the elements of its correlation matrix \( \{ (x_n, x_k)_H \} \) depend only on the difference \( n - k \), i.e.
\[
(x_n, x_k) = (x_{n+j}, x_{k+j}), \quad \forall n, k, j \in \mathbb{Z},
\]
and \( H = \text{span}_H(x_n : n \in \mathbb{Z}) \).

A subspace \( E_- = \text{span}_H(x_n : n < 0) \) is said to be the past of the process, and \( E_+ = \text{span}_H(x_n : n \geq 0) \) the future of the process. The process is said to...
be singular (or deterministic) if \(E_\sim = H\), and regular (or non-deterministic) if \(E_\sim \neq H\).

The problem of optimal (quadratic, one step ahead) prediction is to calculate

\[
\text{dist}_H(x_n, H_n) = \inf_{x \in H_n} \| x_n - x \|
\]

where \(H_n = \text{span}_H(x_k: k < n)\) is the past of \(x_n\).

The main problem concerning random processes is to study the "dependence of the future of a process on its past," and in particular, to measure the best prediction of its state one or several step(s) ahead. The following theorem introduces the central concept of the theory.

**Theorem 1.6.2** (Kolmogorov, 1939) Let \((x_n)_{n \in \mathbb{N}}\) be a stationary random process, \(H = \text{span}_H(x_n: n \in \mathbb{Z})\). Then, there exist a unique Borel measure \(\mu\) on \(T\) and a unitary operator \(U: H \to L^2(\mu)\) such that

\[
U x_n = z^n, \quad n \in \mathbb{Z}.
\]

Conversely, for any \(\mu\) and every unitary operator \(U: H \to L^2(\mu)\), the sequence \((U^{-1}z^n)_{n \in \mathbb{Z}}\) is a stationary process.

The measure \(\mu\) is called the **spectral measure** of the process.

**Proof** First, observe that for every linear combination of \(x_n\), we have

\[
\left\| \sum a_n x_n \right\|^2 = \sum_{n,k} a_n \overline{a}_k(x_n, x_k) = \sum_{n,k} a_n \overline{a}_k(x_{n+1}, x_{k+1}) = \left\| \sum a_n x_{n+1} \right\|^2.
\]

This means that the mapping defined by

\[
V x_n = x_{n+1}, \quad n \in \mathbb{Z},
\]

can be extended by linearity \(V(\sum a_n x_n) = \sum a_n x_{n+1}\) to an isometric mapping \(H \to H\) such that \(VH\) is dense in \(H\); hence \(VH = H\) and \(V\) is unitary. Moreover, \(x_n = V^n x_0\) for every \(n \in \mathbb{Z}\). By the spectral theorem (Appendix E), there exists a unique Borel measure \(\mu\) on \(T\) and a unitary mapping \(U: H \to L^2(\mu)\) such that \(U x_0 = 1\) and \(V = U^{-1} M_z U\), where \(M_z\) is the shift operator on \(L^2(\mu)\). The rest of the statement is immediate. \(\blacksquare\)

**Corollary 1.6.3** A stationary process \((x_n)_{n \in \mathbb{Z}}\) is singular if and only if \(H_\sim^2(\mu) = L^2(\mu)\), with \(H_\sim^2(\mu) = \text{span}_{L^2(\mu)}(z^n: n < 0)\) and \(\mu\) the spectral measure of \((x_n)_{n \in \mathbb{Z}}\).

The corollary is immediate by the definitions and the theorem. \(\blacksquare\)
Andrey Nikolaevich Kolmogorov (1903–1987) was a Russian mathematician, one of the greatest geniuses in mathematics of the twentieth century, creator of the modern mathematical theory of probability, the KAM (Kolmogorov–Arnold–Moser) theory, the Kolmogorov complexity theory, turbulence theory, etc. Dozens of concepts of mathematics and their applications bear Kolmogorov’s name: the Kolmogorov A-integral, Kolmogorov’s inequality, the Kolmogorov–Smirnov test in statistics, the Kolmogorov 0–1 law, the Chapman–Kolmogorov equation, the entropy of a dynamic system, etc. Originally a member of Nikolai Luzin’s famous group of students (at the University of Moscow), throughout his career he founded a number of scientific schools in different domains, eventually training a total of 69 doctoral students, of which 18 became members of the Academies of Science of various countries.

Among other achievements, Kolmogorov is famous for his solution (with Vladimir Arnold) of Hilbert’s 13th problem (1957). He published more than 300 articles, as well as several books that became classics. He was awarded the Chebyshev Prize (1950), the Balzan Prize (1962), the Wolf Prize (1980), and the Lobachevsky Prize (1986), and was a member of dozens of scientific academies and societies.

Luzin wrote to him: Вам дан высокий дух, и я хочу, чтобы Вы его силы берегли для вещей, которые под силу очень немногим . . . (“You were given a great spirit, and I want you to save its strength to achieve exploits accessible by only a very few”). A caveat from Kolmogorov: “Beware of those said to be ‘good mathematicians’ by engineers and ‘good engineers’ by mathematicians.”
The study of the problem of prediction starts with a few lemmas, somewhat technical but very useful. This study will be continued in Chapters 2 and 3.

In the lemmas, \( \mu \) always stands for the spectral measure of a stationary process \( (x_n)_{n \in \mathbb{Z}} \).

**Lemma 1.6.4** For every \( n \in \mathbb{Z} \), we have

\[
\text{dist}_H(x_n, H_n) = \text{dist}_H(x_0, H_0) = \text{dist}_{L^2(\mu)}(1, H_0^2(\mu)) = \text{dist}_{L^2(\mu)}(1, H_0^2(\mu)) := d,
\]

where \( H_0^2(\mu) =: \text{span}_{L^2(\mu)}(z^n : n > 0) = \text{clos}_{L^2(\mu)}(zP_a) \).

**Proof** We first use the isometric nature of the operators \( U \) and \( V \) in the proof of Theorem 1.6.2; then, because \( p \in zP_a \Leftrightarrow \overline{p} \in \text{Lin}(z^n : n < 0) \), for every polynomial \( p \in P \), we have

\[
\|1 - p\|_{L^2(\mu)} = \|1 - \overline{p}\|_{L^2(\mu)}.
\]

\[\blacksquare\]
Lemma 1.6.5 Let \( \mu \) be a finite Borel measure on \( \mathbb{T} \). The following assertions are equivalent.

1. \( d = 0 \) (\( d \) is defined in Lemma 1.6.4).
2. \( 1 \in H_0^2(\mu) := \text{span}_{L^2(\mu)}(z^n : n > 0) \).
3. \( 1 \in H_0^2(\mu) := \text{span}_{L^2(\mu)}(z^n : n < 0) \).
4. \( \tilde{z} \in H^2(\mu) := \text{span}_{L^2(\mu)}(z^n : n \geq 0) \).
5. \( H^2(\mu) = L^2(\mu) \).
6. \( \tilde{z}H^2(\mu) = zH^2(\mu) \).
7. \( zH^2_0(\mu) = H^2_0(\mu) \).

Proof (1) \(\Leftrightarrow\) (2) since in a metric space \( X \), \( x_0 \in \text{clos}(A) \Leftrightarrow \text{dist}_X(x_0, A) = 0 \).

(2) \(\Leftrightarrow\) (4) by Lemma 1.6.4.

(2) \(\Leftrightarrow\) (6) since the mapping \( f \mapsto \tilde{z}f \) is unitary on \( L^2(\mu) \).

(4) implies \( \lim_n \|\tilde{z} - p_n\|_{L^2(\mu)} = 0 \) for a sequence \( p_n \in \mathcal{P}_a \), and hence \( \lim_n \|\tilde{z} - p_nq\|_{L^2(\mu)} = 0 \) for every \( q \in \mathcal{P}_a \), thus \( \mathcal{ZP}_a \subset H^2(\mu) \). Since \( \lim_n \|\tilde{z} - z^{k-1}p_n\|_{L^2(\mu)} = 0 \) for every \( k \geq 1 \), then by induction \( \tilde{z}\mathcal{P}_a \subset H^2(\mu) \), for every \( n \geq 0 \). Therefore, \( \mathcal{P} \subset H^2(\mu) \), and we obtain (5):

\[ H^2(\mu) = L^2(\mu). \]

(5) \(\Rightarrow\) (4) is evident.

(3) \(\Leftrightarrow\) (6) for the same reason that (2) \(\Leftrightarrow\) (5).

Finally, clearly, (5) \(\Rightarrow\) (7); and (7) \(\Rightarrow\) (2), since \( z \in zH^2(\mu) \) implies \( 1 \in H^2_0(\mu) \) (and the same manipulation with \( H^2(\mu) \)).

\[ \square \]

Lemma 1.6.6 Let \( \mu = \text{wm} + \text{us} \) be the Radon–Nikodym decomposition of a finite Borel measure on \( \mathbb{T} \). Then,

\[ d^2 := \text{dist}_{L^2(\mu)}^2(1, H^2_0(\mu)) = \text{dist}_{L^2(\text{wm})}^2(1, H^2_0(\text{wm})) = \inf_{p \in \mathcal{P}_a} \int_{\mathbb{T}} |1 - p|^2 w \, dm. \]

Proof By Lemma 1.5.2, \( H^2(\mu) = H^2(\mu_a) \oplus L^2(\mu_s) \); hence

\[ H^2_0(\mu) = zH^2(\mu) = H^2_0(\mu_a) \oplus zL^2(\mu_s) = H^2_0(\mu_a) \oplus L^2(\mu_s). \]

Writing \( 1 = 1_a \oplus 1_s \) (according to the decomposition \( L^2(\mu) = L^2(\mu_a) \oplus L^2(\mu_s) \)), we have
\[ d^2 = \text{dist}^2_{L^2(\mu)}(1_a \otimes 1_s, H^2_0(\mu_a) \oplus L^2(\mu_s)) \]
\[ = \text{dist}^2_{L^2(\mu_a)}(1_a, H^2_0(\mu_a)) + \text{dist}^2_{L^2(\mu_s)}(1_s, L^2(\mu_s)) = \text{dist}^2_{L^2(\mu_a)}(1_a, H^2_0(\mu_a)). \]

The spaces \( H^2(\mu) \) are the principal tools used in \( \S \) 1.7, but the final conclusion on the subject will be made in Chapter 2, \( \S \) 2.7.2.

### 1.7 Inner–Outer Factorization and Szegő’s Infimum

Recall that in Definition 1.4.2 we defined the inner functions, in the sense of Beurling. We now complete this terminology as follows.

**Definition 1.7.1** Let \( f \in H^2(\mathbb{T}) \). It is said to be outer if \( E_f = H^2(\mathbb{T}) \), where

\[ E_f := \text{span}_{H^2}(z^n f : n \geq 0), \]

i.e. \( E_f \) is the smallest (closed) invariant subspace of \( M_z \) containing \( f \).

**Theorem 1.7.2** (Smirnov, 1928a,b) Every function \( f \in H^2(\mathbb{T}), f \neq 0 \), can be factorized as

\[ f = f_{in} f_{out}, \]

where \( f_{in} \) is an inner function and \( f_{out} \) is outer. This factorization is unique up to a constant factor: if \( f = f_{in}' f_{out}' \) is another inner–outer factorization, then \( f_{in} = \lambda f_{in}', f_{out} = \overline{\lambda} f_{out}' \) with some \( \lambda \in \mathbb{T} \).

**Proof** By Corollary 1.4.3, there exists an inner function \( q \) such that \( E_f = qH^2(\mathbb{T}) \). In particular, \( f = qg \) where \( g \in H^2(\mathbb{T}) \). Let us show that \( g \) is outer. Indeed, for every function \( h \in H^2(\mathbb{T}) \) there exist polynomials \( p_n \in \mathcal{P}_d \) such that \( \lim_n \| p_n f - qh \| = 0 \). However, \( \| p_n f - qh \|^2 = \| q p_n g - qh \|^2 = \int_{\mathbb{T}} |q(p_n g - h)|^2 \, dm = \| p_n g - h \|^2 \), which shows that \( h \in E_g \), and hence \( E_g = H^2(\mathbb{T}) \) (i.e. \( g \) is outer). By setting \( f_{in} = q, f_{out} = g \), we obtain the desired factorization.

For the uniqueness, suppose there is another factorization: then \( f_{in} f_{out} = f_{in}' f_{out}' \), hence \( f_{in} f_{in}' f_{out} = f_{out}' \). Let \( (p_n) \) be a sequence of polynomials such that \( \lim_n \| p_n f_{out} - 1 \| = 0 \). Since \( f_{in} f_{in}' \) is a unimodular function, we obtain \( \| p_n f_{out} - 1 \| = \| p_n f_{out} f_{in} f_{in}' - f_{in} f_{in}' \| \to 0 \) as \( n \to \infty \). However, \( p_n f_{out} f_{in} f_{in}' = p_n f_{out}' \in H^2(\mathbb{T}) \), and consequently \( f_{in} f_{in}' \in H^2(\mathbb{T}) \). Similarly, \( f_{in}' f_{in}' \in H^2(\mathbb{T}) \), which gives \( f_{in} f_{in}' = \text{constant} \) (compare with the proof of Theorem 1.3.5), and the result follows.
Vladimir Ivanovich Smirnov (1887–1974) was a Russian mathematician, a representative of the Saint Petersburg school (founded by Chebyshev), and a founder of modern complex analysis at Saint Petersburg. He obtained numerous important results on Hardy spaces (canonical factorization, the Smirnov “class D,” Hardy spaces on Smirnov domains, etc.), as well as in ordinary differential equations and mathematical physics. He is also known for his five-volume Course of Higher Mathematics, which for years dominated the teaching of mathematics at university level in Russia/USSR. He co-authored works with Friedman, Tamarkin, Lebedev, and others.

His notable students include Goluzin, Havin, Kantorovich (Nobel Prize in Economics, 1975), Lozinsky, Sobolev, and Yakubovich.

Moreover, Smirnov was renowned for his exceptional personality; he was irreproachable for his nobility, kindness, and generosity, even under the unforgiving circumstances of Russian/Soviet reality in the twentieth century.
Corollary 1.7.3 (Beurling, 1949) Let \( f \in H^2(\mathbb{T}) \), \( f \neq 0 \). Then, \( E_f = f_{in}H^2(\mathbb{T}) \).

Indeed:

\[
E_f = \text{clos}_{H^2}(f_{in}f_{out}P_a) = f_{in}\text{clos}_{H^2}(f_{out}P_a) = f_{in}H^2(\mathbb{T})
\]

(since \( f_{in} \) is unimodular)

(\( f_{out} \) is outer).

Theorem 1.7.2 also leads to a crucial development in the problem of \( L^2 \) optimal prediction (see Theorem 1.7.6 below), i.e. in the expression of the quantity \( d \) in Lemma 1.6.6 as a function of the measure \( \mu = wm + \mu_s \) (more precisely: of the Radon–Nikodym derivative \( w = d\mu/dm \)),

\[
d^2 = \text{dist}^2_{L^2(\mu)}(1, H^2_0(\mu))^2 = \inf_{p \in P_a} \int |1 - p|^2 d\mu.
\]

This last extremal problem appeared in the research of Gábor Szegő in the 1920s, and bears his name: the Szegő infimum.

However, we first need a property of outer functions of the type “maximum principle” (for details see Chapter 3 below).

Theorem 1.7.4 (Smirnov, 1932) Let \( f \in H^2(\mathbb{T}) \), \( f \neq 0 \). Then the following assertions are equivalent.

1. \( f \) is an outer function.
2. \( \forall g \in H^2(\mathbb{T}), g/f \in L^2(\mathbb{T}) \implies g/f \in H^2(\mathbb{T}) \).

Proof (1) \( \Rightarrow \) (2) Let \( p_n \in P_a \) such that \( \lim_n \|p_nf - 1\|_2 = 0 \), and suppose that \( g \in H^2(\mathbb{T}) \) such that \( g/f \in L^2 \), i.e. \( g = fh \) where \( h \in L^2 \). Then

\[
\int_{\mathbb{T}} |p_n g - h| dm = \int_{\mathbb{T}} |p_n fh - h| dm \leq \|h\|_2 \|p_n f - 1\|_2 \to 0 \quad \text{for} \quad n \to \infty.
\]

The convergence in \( L^1(\mathbb{T}) \) implies the convergence of the Fourier coefficients: \( \forall k \in \mathbb{Z} \) we have \( \hat{h}(k) = \lim_n (p_n g)\gamma(k) \). However \( p_n g \in H^2(\mathbb{T}) \) (because \( H^2(\mathbb{T}) \) is \( M_z \)-invariant), and hence \( \hat{h}(k) = 0 \) for every \( k < 0 \). Thus \( h \in H^2(\mathbb{T}) \).

(2) \( \Rightarrow \) (1) Let \( f = f_{in}f_{out} \) be the inner–outer factorization of \( f \). Then, \( f_{out} \in H^2(\mathbb{T}) \) and \( f_{out}/f = \overline{f_{in}} \in L^2(\mathbb{T}) \); hence, by (2), \( \overline{f_{in}} \in H^2(\mathbb{T}) \) and of course \( f_{in} \in H^2(\mathbb{T}) \). As seen several times earlier (for example, in Theorem 1.3.5), this implies \( f_{in} = \text{constant} \): hence \( f \) is an outer function.

Corollary 1.7.5

1. If \( f \in H^2(\mathbb{T}) \) is simultaneously inner and outer, then \( f = \text{constant} \).
(2) If \(f, g \in H^2(\mathbb{T})\) are outer and \(|f| = |g|\) a.e. on \(\mathbb{T}\), then \(f = \lambda g\) for some unimodular constant \(\lambda\).

Indeed, for (1), we apply Theorem 1.7.4 to 1 and \(f\), and obtain \(1/f = \overline{f} \in H^2(\mathbb{T})\), which, with \(f \in H^2(\mathbb{T})\), again implies \(f = \text{constant}\).

For (2), by setting \(h = f/g\) and applying the theorem, we obtain \(h \in H^2(\mathbb{T})\), and, by switching the roles of \(f\) and \(g\), \(\overline{h} \in H^2(\mathbb{T})\). Hence, \(h = \text{constant}\) (clearly unimodular). 

Gábor Szegő (1895–1985), a Hungarian–German–American mathematician, is known for his work in classical analysis, such as orthogonal polynomials and Toeplitz operators. After obtaining his doctorate in Budapest under the supervision of Fejér, he went to Berlin and Königsberg, but then, pressured by the Nazis, he emigrated to the USA. His famous collection of solved problems, with Pólya, Aufgaben und Lehrsätze aus der Analysis (1925), served for years as an essential source of training for generations of analysts. He is the author of several other reference monographs. His experiences in Budapest included tutoring the young child prodigy Johannes von Neumann. According to witnesses, Szegő was moved to tears by his first meeting with the young Johannes, so rapid and profoundly complete were the responses of his new student.

**Theorem 1.7.6** (Szegő, 1920; Verblunsky, 1936; Kolmogorov, 1941) Let \(\mu = \omega m + \mu_s\) be the Radon–Nikodym decomposition of a finite Borel measure on \(\mathbb{T}\). Then:

(1) either there does not exist any \(f \in H^2(\mathbb{T})\) such that \(|f|^2 = w\), and then

\[
d = \text{dist}_{L^2(\mu)}(1, H^2_0(\mu)) = 0,
\]
or there exists a (unique) outer function $F \in H^2(\mathbb{T})$ such that $|F|^2 = w$, and then

$$d = \text{dist}_{L^2(\mu)}(1, H^2_0(\mu)) = |\hat{F}(0)| > 0.$$ 

**Proof** Suppose $d > 0$. Then, $\text{dist}_{L^2(wm)}(1, H^2_0(wm)) > 0$ (Lemma 1.6.6), and hence $zH^2_0(wm) \neq H^2_0(wm)$ (Lemma 1.6.5), which implies that $H^2_0(wm)$ is an invariant non-reducing subspace of $L^2(wm)$. Helson’s theorem (Theorem 1.3.5) provides a function $q$ such that $H^2_0(\mu) = qH^2(\mathbb{T})$ and $|q|^2 w = 1$ a.e. on $\mathbb{T}$. In particular, $z = qf$ where $f \in H^2(\mathbb{T})$, which implies $|f|^2 = |z/q|^2 = w$. Setting $F = f_{out}$, we obtain $|F|^2 = w$ and

$$d^2 = \inf_{p \in P_2} \int_{\mathbb{T}} |1 - p|^2 |F|^2 \, dm = \inf_{p \in P_2} \int_{\mathbb{T}} |F - pF|^2 \, dm$$

$$= \text{dist}^2_{H^2}(F, zH^2(\mathbb{T})) = \|P_{H^2 \ominus zH^2} F\|^2 = |\hat{F}(0)|^2.$$ 

Conversely, if $w = |F|^2$ with an outer function $F \in H^2(\mathbb{T})$, then the last formula shows again that $d = |\hat{F}(0)|$. It remains to remark that $\hat{F}(0) \neq 0$ for every outer function $F$. Indeed, if we suppose $\hat{F}(0) = 0$, we would have $F = \sum_{n \geq 1} \hat{F}(n)z^n = z(\sum_{k \geq 0} \hat{F}(k + 1)z^k) \in zH^2(\mathbb{T})$, which implies $E_F \subset zH^2(\mathbb{T}) = H^2_0(\mathbb{T}) \neq H^2(\mathbb{T})$. Thus we obtain a contradiction. 

In fact, the last theorem does not resolve the prediction problem: expressing an error $d(\mu)$ of the best quadratic prediction of a process as a function of the spectral measure $\mu$. In order to obtain the famous Szegő–Verblunsky–Kolmogorov formula

$$d(\mu) = \exp\left(\int_{\mathbb{T}} \log \left| \frac{d\mu}{d\lambda} \right| \, d\lambda \right),$$

we need to develop a theory of “canonical factorization” of functions $H^2(\mathbb{T})$. This is the goal of Chapter 2.

### 1.8 Exercises

#### 1.8.1 The Wold–Kolmogorov Decomposition

Let $T : H \to H$ be a linear isometry in a Hilbert space $H$, $E \in \text{Lat}(T)$ and $W = E \ominus TE$. Prove the following.

(a) $T^nW \perp T^mW$ for every $n \neq m$ ($n, m \geq 0$) ($W$ is said to be a “wandering subspace”).
(b) The subspace $E_\infty = \bigcap_{n \geq 0} T^n E$ reduces $T$, and the restriction $T|E_\infty$ is unitary.

Solution: Let $x \in E_\infty$. For every $n \geq 0$, there exists $x_n \in E$ such that $x = T^n x_n$, which implies $Tx = T^{n+1} x_n$, and hence, $Tx \in E_\infty$. Moreover, $T^n x_n = T^{n+k} x_{n+k} \Rightarrow x_n = T^k x_{n+k}$, which in turn implies $x_n \in E_\infty$, and in particular, $x \in TE_\infty$. Hence, $TE_\infty = E_\infty$ and $T|E_\infty$ is a unitary mapping of $E_\infty$ onto itself. For the reduction property, we have $T^* x = T^* T^{n+1} x_{n+1} = T^n x_{n+1}$, hence $T^* x \in E_\infty$.

(c) The subspace $E_0 = \sum_{n \geq 0} \oplus T^n(W)$ is $T$-invariant and $T|E_0$ is completely non-unitary (i.e. if $E' \subset E_0$, $TE' \subset E'$ and $T|E'$ is unitary, then $E' = \{0\}$).

Solution: $E_0 = \{x : x = \sum_{n \geq 0} T^n w_n : w_n \in W, \sum_{n \geq 0} ||T^n w_n||^2 = \sum_{n \geq 0} ||w_n||^2 < \infty \}$ (convergence in norm, unique representation, see Appendix C). This implies $TE_0 \subset E_0$ and $\bigcap_{n \geq 0} T^n E_0 = \{0\}$. If $E' \subset E_0$, $TE' \subset E'$ and $T|E'$ is unitary, then $E' = \bigcap_{n \geq 0} T^n E' \subset \bigcap_{n \geq 0} T^n E_0 = \{0\}$.

(d) The Wold–Kolmogorov decomposition (1939): $E = E_0 \oplus E_\infty$.

Solution: Let $x \in E$; then, $x \in E \ominus E_0 \Leftrightarrow x \in E, x \perp T^n E \ominus T^{n+1} E$ (for every $n \geq 0$) \Leftrightarrow (consecutively, with $n = 0, 1, \ldots$) $x \in E, x \in TE, x \in T^2 E, \ldots \Leftrightarrow x \in E_\infty$.

1.8.2 The Shift Operator $M_z$ on $L^2(\mathbb{T}, \mu)$

Let $\mu$ be a finite Borel measure on $\mathbb{T}$ and $M_z : L^2(\mathbb{T}, \mu) \to L^2(\mathbb{T}, \mu)$ the shift operator (translation), $M_z f = zf$.

(a) Let $E \in \text{Lat}(M_z)$. Describe its Wold–Kolmogorov decomposition (using Helson’s Theorem 1.3.5).

Solution: If $E$ is reducing, $M_z E = E$, then $E_\infty = E$, $W = \{0\}$. Otherwise, by Theorem 1.3.5, $E = \chi_A L^2(\mu_s) \oplus q H^2(\mathbb{T})$ where $|q|^2 w = 1$ m.a.e. $(\mu = \mu_s + w m)$ is the Radon–Nikodym decomposition of $\mu$. Then clearly $M_z(\chi_A L^2(\mu_s)) = \chi_A L^2(\mu_s)$ and $W = E \ominus M_z E = q H^2(\mathbb{T}) \ominus qz H^2(\mathbb{T}) = q \mathbb{C}$ (a subspace of dim = 1 containing $q$). Consequently, $E_\infty = \chi_A L^2(\mu_s)$ and the completely non-unitary portion of $M_z$ is $M_z|q H^2(\mathbb{T})$.

(b) Let $\mu_i (i = 1, 2)$ be finite Borel measures on $\mathbb{T}$. Find a necessary and sufficient condition on $\mu_i$ so that the shift operators $S_i := M_z : L^2(\mu_i) \to L^2(\mu_i)$
(i = 1, 2) are unitarily equivalent (i.e. there exists a unitary \( U : L^2(\mu_1) \to L^2(\mu_2) \) such that \( US_1 = S_2U \)).

**Solution:** Suppose \( S_1 \) and \( S_2 \) are equivalent and \( U \) is a unitary operator such that \( US_1 = S_2U \). Then, \( US_i^k = S_i^kU \) for every \( k \in \mathbb{Z} \), and hence, for any polynomial \( p \in \mathcal{P} \), we have \( UP = p \cdot U1 \). By a passage to the limit in the last equation (for the norm \( L^2 \) on the left, and for the norm \( L^1 \) on the right) we obtain \( Uf = fU1 \) for any \( f \in L^2(\mu_1) \). Then \( U \) is unitary, and therefore

\[
\int |f|^2 |U1|^2 \, d\mu_2 = \int |f|^2 \, d\mu_1, \quad \forall f \in L^2(\mu_1),
\]

which implies that \( \mu_1 = |U1|^2 \mu_2 \), hence \( \mu_1 \ll \mu_2 \). By swapping the roles of \( S_1 \) and \( S_2 \), we obtain \( \mu_2 \ll \mu_1 \) (thus, the measures are equivalent: \( \mu_1 \sim \mu_2 \)). Conversely, if \( \mu_1 \sim \mu_2 \), then \( \mu_1 = h\mu_2 \) where \( h \in L^1(\mu_2) \) and \( 1/h \in L^1(\mu_1) \) (\( \Leftrightarrow h \neq 0 \mu_2\text{-a.e.} \)), and the mapping \( Uf = f\sqrt{h} \) is unitary: \( U : L^2(\mu_1) \to L^2(\mu_2) \) and satisfies \( US_1 = S_2U \).

**Solution:** The operators \( S_i|H^2(\mu_i) \) are isometric: they are simultaneously unitary or not, and this is the case if and only if \( H^2(\mu_i) = L^2(\mu_i) \). If this last equality holds, the question is already answered in (b); if not, we extend the operator \( U : H^2(\mu_1) \to H^2(\mu_2) \) such that \( US_1 = S_2U \) to a mapping \( U : L^2(\mu_1) \to L^2(\mu_2) \) with the same relation of commutation by the equation \( U(\overline{z}^i f) := \overline{z}^i Uf, f \in H^2(\mu_1) \). The final answer is: \( S_i|H^2(\mu_i), i = 1, 2, \) are unitarily equivalent if and only if \( \mu_1 \sim \mu_2 \) and \( H^2(\mu_i) \) simultaneously coincide (or not) with \( L^2(\mu_i) \) for \( i = 1, 2 \).

**Solution:** By Theorem 1.3.5, there exists a non-reducing invariant subspace if and only if there exists a measurable function \( q \) such that \( |q|^2w = 1 \) \( m\text{-a.e.} \) on \( \mathbb{T} \), with \( w = dm/dm \). Clearly the last property is equivalent to \( w > 0 \) \( m\text{-a.e.} \) on \( \mathbb{T} \). The answer to (d) is: it is necessary and sufficient that \( w = 0 \) on a set \( \sigma \subset \mathbb{T} \) having \( m(\sigma) > 0 \) (which is equivalent to \( m \not\ll \mu \)).

### 1.8.3 Inner and Outer Functions

A few “bare-hands” examples, without using the theory of Chapter 2, but nonetheless using knowledge of the multipliers of \( H^2(\mathbb{T}) \) (part (a) below).
(a) **Multipliers, algebra** $H^\infty$. Here $L^2 = L^2(\mathbb{T})$ and $H^\infty(\mathbb{T}) := H^2(\mathbb{T}) \cap L^\infty(\mathbb{T})$. The multiplier spaces are defined by

\[
\text{Mult}(L^2) = \{ h : f \in L^2 \Rightarrow hf \in L^2 \},
\]
\[
\text{Mult}(H^2(\mathbb{T})) = \{ h : f \in H^2(\mathbb{T}) \Rightarrow hf \in H^2(\mathbb{T}) \}.
\]

(i) **Show that** $\text{Mult}(L^2) = L^\infty(\mathbb{T})$, $\text{Mult}(H^2(\mathbb{T})) = H^\infty(\mathbb{T})$.

**Solution**: The inclusion $L^\infty(\mathbb{T}) \subset \text{Mult}(L^2)$ is evident. For the converse, let $h \in \text{Mult}(L^2)$, then $\forall f \in L^2$, $\int |f|^2 |h|^2 \, dm < \infty$. However, $g = |f|^2$ is an arbitrary positive function of $L^1(\mathbb{T})$, hence $h \in L^\infty(\mathbb{T})$ (Appendix A). For the case of $\text{Mult}(H^2(\mathbb{T}))$, clearly $\text{Mult}(H^2(\mathbb{T})) \subset H^2(\mathbb{T})$ and thus, for any $h \in \text{Mult}(H^2(\mathbb{T}))$, the multiplication operator $M_hf = hf$ is continuous $H^2(\mathbb{T}) \to L^1(\mathbb{T})$, hence it is closed for $H^2(\mathbb{T}) \to H^2(\mathbb{T})$, thus bounded (by the closed graph theorem). Consequently, the formula $M_hf = \mathbb{Z}^n M_h(z^n f)$ extends $M_h$ on the subspace $\mathbb{Z}^n H^2(\mathbb{T}) \subset L^2$ (with the same norm), and by approximation, on the whole space $L^2$. Thus $\text{Mult}(H^2(\mathbb{T})) \subset L^\infty(\mathbb{T})$, which leads to $\text{Mult}(H^2(\mathbb{T})) \subset H^\infty(\mathbb{T})$. For the converse, note that $h \in H^\infty(\mathbb{T})$, $p \in \mathcal{P}_a \Rightarrow hp \in H^2(\mathbb{T})$, and again by approximation (letting $\|p_n - f\|_2 \to 0$), we obtain $hf \in H^2(\mathbb{T})$ for any $f \in H^2(\mathbb{T})$, which shows that $H^\infty(\mathbb{T}) \subset \text{Mult}(H^2(\mathbb{T}))$.  

(ii) $H^\infty(\mathbb{T})$ is a Banach algebra for standard multiplication on $\mathbb{T}$. Moreover, for every function $f \in L^2$, we have $f \cdot H^\infty(\mathbb{T}) \subset E_f$.

**Solution**: The space of multipliers Mult$(X) = \{ h : f \in X \Rightarrow hf \in X \}$ of a function space $X$ is clearly an algebra. Moreover, the inequality $\|fg\|_\infty \leq \|f\|_\infty \|g\|_\infty$ for $f, g \in H^\infty$ is also evident, and the result follows for the algebra $H^\infty(\mathbb{T})$. For the rest, clearly $fp_a \subset E_f$. It only remains to show that $(fp_a)^{-1} \subset (fH^\infty)^{-1}$ (orthogonal complement in $L^2$). Let $g \in (fp_a)^{-1}$, i.e. $\int g \bar{f}p \, dm = 0$ for any polynomial $p \in \mathcal{P}_a$. Thus for any $h \in H^\infty$, $\int \bar{g}f h \, dm = 0$ because $\bar{g}f \in L^1$ and $h$ is a weak limit $\sigma(L^\infty, L^1)$ of its Fejé polynomials (see Appendix A).

(b) **Examples of inner functions**. Show that the following functions are inner:

(i) $b_\lambda = (\lambda - z)/(1 - \bar{\lambda}z)$ where $\lambda \in \mathbb{D} = \{ z \in \mathbb{C} : |z| < 1 \}$.

**Solution**: $b_\lambda = (\lambda - z) \sum_{n \geq 0} \bar{\lambda}^n \bar{z}^n$ ($|z| = 1$), and clearly $\hat{b}_\lambda(k) = 0$ for $k < 0$, and $\sum_{k \geq 0} |\hat{b}_\lambda(k)|^2 < \infty$; hence $b \in H^2(\mathbb{T})$. Moreover, for $|z| = 1$, we have $|\lambda - z| = |\bar{\lambda} - \bar{z}| = |1 - \bar{\lambda}z|$, thus $|\hat{b}_\lambda(z)| = 1$.  

(ii) $f = \prod_{k=1}^{N} b_{\lambda_k}$ where $\lambda_k \in \mathbb{D}$.

**Solution**: As $H^\infty(\mathbb{T}) \cdot H^\infty(\mathbb{T}) \subset H^\infty(\mathbb{T})$ (by part (ii) of (a)), a product of inner functions is inner.
(iii) $s_{\zeta, a} = \exp(-a((\zeta + z)/(\zeta - z)))$ where $a > 0$, $\zeta \in \mathbb{T}$.

**Solution:** As

$$\text{Re} \left( \frac{\zeta + z}{\zeta - z} \right) = \frac{1 - |z|^2}{|\zeta - z|^2} \geq 0$$

for any $\zeta \in \mathbb{T}$, $|z| \leq 1$, $z \neq \zeta$, we obtain $|s_{\zeta, a}| = 1$ on $\mathbb{T}$. Moreover, for every $n > 0$, we have

$$\hat{s}_{\zeta, a}(-n) = \int_{\mathbb{T}} \zeta^n s_{\zeta, a}(z) \, dm = \lim_{n \to 1} \int_{\mathbb{T}} f_r(z) \, dm = 0$$

where $f(z) = \zeta^n s_{\zeta, a}(z)$ and $f_r(z) = f(rz)$, $0 \leq r < 1$ ($f_r(0) = 0$ since $f_r$ is analytic in $|z| < 1/r$ and $f_r(0) = 0$).

(iv) $f = \prod_{k=1}^{N} s_{\zeta_k, a_k}$ where $a_k > 0$, $\zeta_k \in \mathbb{T}$.

**Solution:** See the solution of (ii) above.

(c) **Examples of outer functions.** Show that the following functions are outer:

(i) $f \in H^2(\mathbb{T})$ such that $1/f \in H^\infty(\mathbb{T})$.

**Solution:** By (a,ii), clearly $1 = f \cdot 1/f \in E_f$, hence $E_f = H^2(\mathbb{T})$.

(ii) $f \in H^\infty$ such that $\text{Re}(f) \geq 0$.

**Solution:** For any $\epsilon > 0$ there exist (a large) $r > 0$ and (a small) $\delta > 0$ such that $|f + \epsilon - r| \leq (1 - \delta)r$ a.e. on $\mathbb{T}$ (to verify this, sketch the region in $\mathbb{C}$ where the values of $f(z) + \epsilon$, $|z| = 1$ are found), or $|(f + \epsilon)/r - 1| \leq (1 - \delta)$; this implies the normal convergence of the series

$$\frac{r}{f + \epsilon} = \sum_{k \geq 0} \left( 1 - \frac{f + \epsilon}{r} \right)^k.$$

However, part (ii) of (a) implies $(1 - (f + \epsilon)/r)^k \in H^\infty$, hence $r/(f + \epsilon) \in H^\infty$. By (ii) we have $f/(f + \epsilon) \in E_f$ and even

$$\lim_{\epsilon \to 0} \int_{\mathbb{T}} \left| \frac{f}{f + \epsilon} - 1 \right|^2 \, dm = 0$$

(by the dominated convergence theorem). Thus $1 \in E_f$, hence $f$ is outer.

(iii) $f = 1 + g$, $g \in H^\infty$, $\|g\|_\infty \leq 1$.

**Solution:** This is a special case of (ii).

(iv) $f \in H^2(\mathbb{T})$ such that $\text{Re}(f) \geq 0$.

**Solution:** The solution of (ii) above shows that it suffices to prove the inclusion $1/(f + \epsilon) \in H^\infty$, or the inclusion $1/(f + \epsilon) \in H^2(\mathbb{T})$ (since $1/(f + \epsilon) \in L^\infty$ is evident). To this end, fix $0 < r < 1$ and consider $f_r(z) = \sum_{k \geq 0} f(k)r^k$, $z \in \mathbb{T}$. Then, $f_r \in C(\mathbb{T})$.

https://doi.org/10.1017/9781316882108.003 Published online by Cambridge University Press
The space $H^2(\mathbb{T})$: An Archetypal Invariant Subspace

(hence bounded) and $\text{Re}(f_r) \geq 0$ (because $f_r = f \ast P_r$, which is a convolution with the positive function

$$P_r(z) = \frac{1 - r^2}{|z - r|^2} = \sum_{k \in \mathbb{Z}} r^{|k|} z^k,$$

$z \in \mathbb{T}$, see Appendix A). By the solution (ii) above, $1/(f_r + \epsilon) \in H^\infty \subset H^2(\mathbb{T})$ and by the dominated convergence theorem,

$$\lim_{r \to 1} \left\| \frac{1}{f_r + \epsilon} - \frac{1}{f + \epsilon} \right\|_2 = 0.$$

Thus $1/(f + \epsilon) \in H^2(\mathbb{T})$.

(d) An extremal problem. First, we justify Cauchy’s formula for Fourier coefficients:

(i) Let $f, g \in L^2(\mathbb{T})$ (thus $fg \in L^1(\mathbb{T})$). Show that, for every $n \in \mathbb{Z}$, $\hat{fg}(n) = \sum_{k \in \mathbb{Z}} \hat{g}(k) \hat{f}(n - k)$: the series converges absolutely.

**Solution:** By Cauchy’s inequality $\|f(g - g')\|_1 \leq \|f\|_2 \|g - g'\|_2$, the multiplication $M_g f = fg$ is continuous $L^2(\mathbb{T}) \to L^1(\mathbb{T})$. Moreover, the Fourier series $g = \sum_{k \in \mathbb{Z}} \hat{g}(k) z^k$ converges for the norm of $L^2(\mathbb{T})$. Hence, $fg = \sum_{k \in \mathbb{Z}} \hat{g}(k) z^k f$ converges in $L^1(\mathbb{T})$, which implies $\hat{fg}(n) = \sum_{k \in \mathbb{Z}} \hat{g}(k) \hat{f}(n - k)$. The calculation $\hat{(z^k f)}(n) = \hat{f}(n - k)$ is elementary.

(ii) Let $f = f_{\text{in}} f_{\text{out}} \in H^2(\mathbb{T})$. Show that

$$\sup\{ |\hat{g}(0)| : g \in H^2(\mathbb{T}), \ |g| \leq |f| \text{ a.e. on } \mathbb{T} \} = |\hat{f}_{\text{out}}(0)|.$$

**Solution:** By (i), clearly $\hat{\varphi \psi}(0) = \hat{\varphi}(0) \hat{\psi}(0)$ for all functions $\varphi, \psi \in H^2(\mathbb{T})$. Moreover, for every inner function $h$, we have $|\hat{h}(0)| \leq ||h||_1 = 1$. Given $g \in H^2(\mathbb{T})$, $|g| \leq |f|$, which implies $|\hat{g}(0)| = |\hat{g}_{\text{in}}(0) \hat{g}_{\text{out}}(0)| \leq |\hat{g}_{\text{out}}(0)|$. Then by Theorem 1.7.6,

$$|\hat{g}(0)|^2 \leq |\hat{g}_{\text{out}}(0)|^2 \leq \inf_{p \in \mathcal{P}_a} \int_\mathbb{T} |1 - p|^2 |g|^2 \, dm \leq \inf_{p \in \mathcal{P}_a} \int_\mathbb{T} |1 - p|^2 |f|^2 \, dm = |\hat{f}_{\text{out}}(0)|^2.$$

1.9 Notes and Remarks

As already mentioned, Hardy spaces $H^p$ were defined in 1915 (Hardy, 1915), and by 1930 the essentials of the theory had been constructed. At the time, it was a novel mix of fundamental ideas: complex analysis, the Lebesgue integral, and functional vector spaces. Very rapidly, Hardy spaces became one of the mainsprings of the development of analysis in the twentieth century. However, the theory had to wait another 30 years, until the 1960s, for the true
magnitude of its potential to be revealed, via the discovery of the main source of its force: the invariant subspaces of the group of translations \( (M^n_\mathbb{Z})_{n \in \mathbb{Z}} \) and its semigroup \( (M^n_\mathbb{Z})_{n \in \mathbb{Z}_+} \).

Arne Beurling (1949) formulated the correspondence between the invariant subspaces and the inner–outer factorization (in fact, the latter had been known by Smirnov for more than 20 years (Smirnov, 1928a,b)). Beurling’s work led to the discovery of the hidden heart of the theory (Helson and Lowdenslager, 1961; Helson, 1964)): the fact that analyticity is a consequence of the causality of the semigroup under consideration, and that the main feature of the subject is that this semigroup is linearly ordered (it is not very important whether we take \( \mathbb{Z} \) or \( \mathbb{R} \), as made clear in the years 1920–1940).

The presentation of this book is based on the novel version of the theory proposed by Helson (1964) (see also Nikolski (1980, 1986, 2002)): in his work, the point of view described above is accepted from the start as the cornerstone of the whole construction. This is a spectacular difference from the classical and/or post-modern presentations, i.e. Privalov (1941), Duren (1970), Garnett (1981), Stein (1993), Koosis (1980), and Pavlović (2004).

Formal references: for Theorem 1.2.1 see Wiener (1933), for Theorem 1.3.5 see Helson (1964), and for Corollary 1.4.3 see Beurling (1949). Historically, the astounding success of the approach by invariant subspaces led to the creation of an “abstract complex analysis” where analyticity is defined and studied with the aid of invariant subspaces with respect to a “semigroup” satisfying certain conditions. This theory is well-developed and is highly efficient for the study of functions of several variables, of almost periodic functions, etc.: see Gamelin (1969) and Barbey and König (1977).

The uniqueness theorem Corollary 1.4.4, as well as Theorem 1.5.4, is due to Riesz and Riesz (1916) (for the proof of § 1.5.1 see Øksendal (1971)). Theorem 1.5.4 plays an important role in several applications, in particular for different forms of the uncertainty principle in harmonic analysis. Numerous generalizations and improvements of this theorem are known; for all these subjects, see Havin and Jöricke (1994).

The contents of § 1.6 are taken from Kolmogorov (1941). The inner–outer factorization of § 1.7 was discovered by Smirnov (1928a,b) and published in a minor Russian journal (but in French! See the Russian translation in Smirnov (1988)). There, Smirnov (following Szegő (1921)) speaks of “maximal functions” instead of “outer functions” (he does not introduce a name for the “inner functions”), which finds a strong justification in several forms of the “maximum principle” (Theorem 1.7.4, found in Smirnov (1932), is one of them; for others, see § 3.3–3.4 below). Because of the isolation of Russia after the Bolshevik revolution, followed by Stalin’s Iron Curtain, these
results remained almost unknown until the 1960s. The other principal result of Beurling (1949) met with the same destiny: Corollary 1.7.3 is an almost immediate consequence of another article by Smirnov (1932).

We can also mention that the inner–outer factorization was rediscovered (practically independently of Smirnov or Beurling) by Wiener and Masani in the framework of the theory of linear prediction (by the generalized Wold–Kolmogorov decomposition), under the name of optimal-residual factorization: see Masani (1966).

Theorem 1.7.6 (with the formula mentioned at the end of this section) was proved by Szegő (1920) in the case $\mu = \mu_a$, and by Verblunsky (1936) and Kolmogorov (1941) in the general case. The role and significance of Verblunsky’s 1936 paper was overlooked by the community for many decades and was restored by a thorough historical analysis in Barry Simon’s book *Orthogonal Polynomials on the Unit Circle, Part 1: Classical Theory* (Simon, 2005; see especially pp. 141, 221).

Generalizations for continuous-time processes are also due to Kolmogorov, and for vector-valued processes to Kolmogorov, Matveev, and Rozanov (see Rozanov, 1963), as well as Wiener and Masani (1957, 1958). The Wold–Kolmogorov decomposition (Wold, 1938; Kolmogorov, 1941) plays an important role in the analysis of time series (in the prediction of random processes).