The Origins of the Subject

Prehistory. Cauchy – Fourier – Poisson – Weierstrass – Stieltjes – Fatou – Lebesgue – Hilbert – Parseval – Jensen.

History. Lebesgue – Hardy – Luzin – Privalov – Schur – the Riesz brothers – Szegő – Nevanlinna – Smirnov – Littlewood – Kolmogorov – Paley – Wiener – Zygmund.

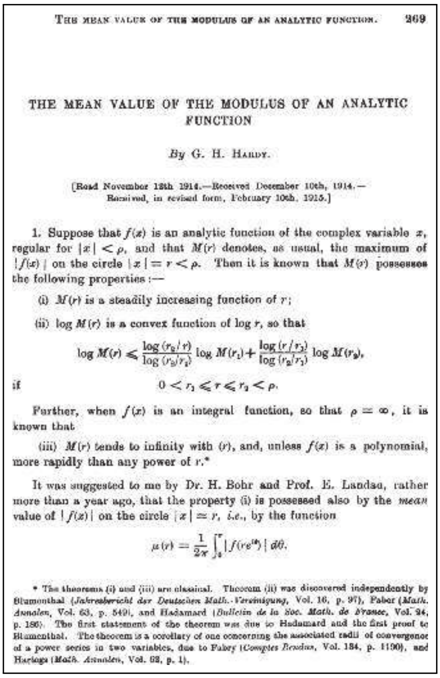
Legacy/Continuation. Stein – Fefferman – de Branges – Helson – Kahane – Garnett – Gamelin – Carleson – Sarason – Havin – Douglas – Arveson – Sz.-Nagy – Foias – Fuhrmann – Lax – Phillips – Lacey, etc.

The birth of Hardy spaces dates back to the year 1915, at Cambridge University. At the time, it went virtually unnoticed. Admittedly, the year 1915 can be considered as "unremarkable" only for their creator, the British mathematician G. H. Hardy (1877–1947). Sure enough, as usual, he had published a dozen (!) articles and research notes, but apparently no salient result emerged from his efforts that year, with one exception – if we equate a definition with a result.

Godfrey Harold (G. H.) Hardy (1877–1947) was one of the founding fathers of modern "hard" analysis, and the author of several fundamental ideas that transformed such disciplines as Diophantine analysis, Tauberian theory, the summation of divergent series, Fourier series, the distribution of prime numbers, and the theory of the Euler ζ function. David Hilbert called him "the best mathematician in England." Several theorems and mathematical creations are named after Hardy. His book *A Mathematician's Apology* (1940) is a masterpiece on the philosophy and psychology of a mathematician. His remarkable essay "Orders of infinity: The 'Infinitärcalcül' of Paul Du Bois-Reymond" (1910) inspired a chapter in Bourbaki's treatise. He was a friend of the novelist and scientist C. P. Snow and a co-author with Littlewood, Ramanujan, Titchmarsh, Ingham, Landau, and Marcel Riesz.

Trinity College, Cambridge.

Specifically, in part of a short nine-page article published in the 1915 *Proceedings of the London Mathematical Society*, Hardy defined a family of spaces ("function classes") of holomorphic functions. At the time, the event was barely noticed: either by the general public (preoccupied by the



THE MEAN VALUE OF THE MODULUS OF AN ANALYTIC FUNCTION. 269

THE MEAN VALUE OF THE MODULUS OF AN ANALYTIC FUNCTION

By G. H. Hardy.

[Read November 12th 1914.—Received December 10th, 1914.—Received, in revised form, February 10th, 1915.]

1. Suppose that $f(x)$ is an analytic function of the complex variable x, regular for $|x| < \rho$, and that $M(r)$ denotes, as usual, the maximum of $|f(x)|$ on the circle $|x| = r < \rho$. Then it is known that $M(r)$ possesses the following properties:—

(i) $M(r)$ is a steadily increasing function of r;

(ii) $\log M(r)$ is a convex function of $\log r$, so that

$$\log M(r) \leqslant \frac{\log(r_2/r)}{\log(r_2/r_1)} \log M(r_1) + \frac{\log(r/r_1)}{\log(r_2/r_1)} \log M(r_2),$$

if $$0 < r_1 \leqslant r \leqslant r_2 < \rho.$$

Further, when $f(x)$ is an integral function, so that $\rho = \infty$, it is known that

(iii) $M(r)$ tends to infinity with (r), and, unless $f(x)$ is a polynomial, more rapidly than any power of r.*

It was suggested to me by Dr. H. Bohr and Prof. E. Landau, rather more than a year ago, that the property (i) is possessed also by the *mean* value of $|f(x)|$ on the circle $|x| = r$, *i.e.*, by the function

$$\mu(r) = \frac{1}{2\pi}\int_0^{\pi} |f(re^{i\theta})|\, d\theta.$$

* The theorems (i) and (iii) are classical. Theorem (ii) was discovered independently by Blumenthal (*Jahresbericht der Deutschen Math.-Vereinigung*, Vol. 16, p. 97), Faber (*Math. Annalen*, Vol. 63, p. 549), and Hadamard (*Bulletin de la Soc. Math. de France*, Vol. 24, p. 186). The first statement of the theorem was due to Hadamard and the first proof to Blumenthal. The theorem is a corollary of one concerning the associated radii of convergence of a power series in two variables, due to Faber (*Comptes Rendus*, Vol. 184, p. 1190), and Hartogs (*Math. Annalen*, Vol. 62, p. 1).

The first page of Hardy's nine-page paper of 1915 defining "Hardy classes." Who could have prophesied that this acorn would grow into such a mighty oak?

First World War), or by the scientific world (1915 was above all the year of Einstein's General Relativity, as well as Wegener's theory of *Pangaea*), or even by mathematicians. Nevertheless, it was a turning point for a number of disciplines linked to mathematical analysis: complex analysis (then flourishing), harmonic analysis, signal processing, and in particular several theories non-existent at the time, but crucial today – the theory of operators, optimal control, diffusion theory, random processes.

Later on in his career, Hardy himself returned several times to the theory of the spaces he had defined in 1915, which, at first glance, seemed to be merely an auxiliary tool. However, for its transformation into an indispensable, extremely powerful technique of analysis and for the majority of its applications, we are highly indebted to the efforts of the "Golden Team" of analysts of that time (such as Schur, Marcel Riesz, Frigyes Riesz, Szegő, Nevanlinna, Luzin, Privalov, Smirnov, Kolmogorov, Paley, Wiener, Zygmund), and to their equally brilliant successors (such as Beurling, Stein, Fefferman, de Branges, Helson, Carleson, Kahane, Garnett, Gamelin, Sarason, Havin, Douglas, Sz.-Nagy, Foias, Fuhrmann, Lax, Phillips).

The explanation for its success can perhaps be summed up in just a few points: (1) the dynamics of the Hardy space $e^{inx}H^2$, $n \in \mathbb{Z}$, generates an orthonormal basis $e^{inx} \in e^{inx}H^2 \ominus e^{i(n+1)x}H^2$ in the Lebesgue space $L^2(-\pi, \pi)$; (2) the space H^2 is the "analytic half" of $L^2(-\pi, \pi)$; (3) in H^2, there is a property of factorization into elementary factors, similar to that of polynomials (in a sense, H^2 is a "factorial ring"). First of course come the definition and the basic properties.

A remark for the experts: the current dominant approach to Hardy spaces is via real harmonic analysis (maximal functions, Hilbert transforms, etc.); thus it is unnecessary to differentiate between H^2 and H^p, $p \neq 2$, or between the groups where the space is defined ($\mathbb{T}$, $\mathbb{T}^n$, $\mathbb{R}$, $\mathbb{R}^n$, etc., and even without any group structure). In this book, I follow a combination of the "genetic" approach based on analysis of a single complex variable, and the spectral analysis of a unitary representation of $\mathbb{Z}$. Why this choice? It is indeed the most elementary and direct route to obtain all the results of the theory needed for applications. Let us add that, so far, the true value of the powerful methods of real variables remains purely theoretical. As soon as we are faced with practical applications of Hardy spaces, we use the complex presentation and its techniques – beginning with signal processing and operator theory, and then H^∞ optimal control and diffusion theory, or even stochastic processes or the Euler ζ function. Our work is especially concerned with the spaces H^2, H^1, and H^∞.

The memorable events of 1915

- Einstein's theory of General Relativity.
- Wegener's theory of *Pangaea*.
- The use of chemical weapons by Germany on a massive scale (Second Battle of Ypres).
- The Mexican Revolution.
- The birth of Paul Tibbets (future pilot in the US Air Force, to be assigned the task of dropping the first atomic bomb on Hiroshima on August 6, 1945).
- The thesis of Nikolai Luzin (future founder of the Moscow school of analysis), written in Paris and defended in Moscow.
- G. H. Hardy's definition of H^p spaces.

1

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

$$x \in E \Rightarrow Tx \in E$$

(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 $\mathcal{T}$ is a family of operators on H, we set $\text{Lat}(\mathcal{T}) = \bigcap_{T \in \mathcal{T}} \text{Lat}(T)$.

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*.

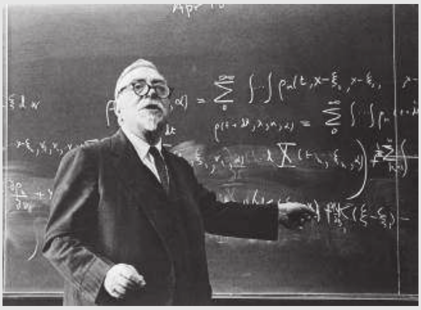
The goal of this section is to describe the lattice $\text{Lat}(M_z)$ where M_z is the operator of multiplication by an "independent variable" in the space $L^2(\mathbb{T}, \mu)$, with μ a finite Borel measure on the circle $\mathbb{T} = \{\zeta \in \mathbb{C}: |\zeta| = 1\}$,

$$M_z f = z f(z), z \in \mathbb{T}.$$

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.



Norbert Wiener (1894–1964) was an American mathematician (MIT: Massachusetts Institute of Technology), creator of cybernetics (1948) and communication theory (co-founded with Kotelnikov and Shannon). He also created the theories of stochastic processes and generalized harmonic analysis (1930, the Wiener measure and Brownian motion), Tauberian theory, and also, independently of Stefan Banach, invented *Banach spaces* (1923). He authored innovative works in mathematical physics, in potential theory and the optimal prediction of random processes (with applications to the automatic correction of the firing of anti-aircraft guns, shared with Kolmogorov). An admirer of Leibniz, Lebesgue, and Hadamard, Norbert Wiener was one of the geniuses of the twentieth century, who revolutionized mathematics and science. The reader can find a remarkable overview of Wiener's scientific impact (as well as a biographical article by Norman Levinson) in vol. 72, issue 1-II (1966) of the *Bulletin of the American Mathematical Society*. Having received his Bachelor's degree at the age of 14, Wiener followed a Master's program in zoology at Harvard, in philosophy at Cornell, and then in mathematics at Harvard. After submitting his thesis in 1912 (at the age of 17), he came to Europe for post-doctoral studies. Upon his return to the USA, Wiener 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.

Massachusetts Institute of Technology.

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 μ 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) $E \in \mathrm{Lat}(M_z, M_z^*)$.
(2) *There exists a Borel set $A \subset \mathbb{C}$ such that*

$$E = \chi_A L^2(\mu) = \{f \in L^2(\mu)\colon\ f = 0\ \mu\text{-a.e. on the complement } A' = \mathbb{C} \setminus A\}.$$

The set A in (2) is unique modulo μ: $\chi_A L^2(\mu) = \chi_B L^2(\mu)$ if and only if $\chi_A = \chi_B$ μ-a.e., i.e. if and only if $\mu(A \bigtriangleup B) = 0$, where $A \bigtriangleup B = (A \setminus B) \cup (B \setminus A)$ is the symmetric difference.

Proof First observe that $M_z^* = M_{\bar z}$ 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 $g \in E^\perp = \{g \in L^2(\mu)\colon (h, g) = 0,\ \forall h \in E\}$ (orthogonal complement of E). Then

$$0 = (pf, g) = \int pf\bar g\, d\mu, \quad \forall p \in \mathcal{P}.$$

Since $\mathcal{P}$ is dense in the space $C(\mathrm{supp}(\mu))$ of continuous functions on a compact set $\mathrm{supp}(\mu)$ (Weierstrass's theorem), we obtain $f\bar{g}\, d\mu = 0$ (the null measure), hence $f\bar{g} = 0$ μ-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

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

(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, $f\bar{g}_n = 0$ μ-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) $f = 0$ μ-a.e. on A^c. Since $g_n = 0$ on A, we have $f\bar{g}_n = 0$ μ-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 μ-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 μ is a Borel measure on the circle $\mathbb{T}$, and μ_a, μ_s denote, respectively, the absolutely continuous and singular components of μ with respect to the normalized Lebesgue measure m, $m\{e^{it}\colon \theta_1 \le t \le \theta_2\} = (\theta_2 - \theta_1)/2\pi \le 1$.

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

(1) $E \in \mathrm{Lat}(T) \Leftrightarrow E^\perp \in \mathrm{Lat}(T^*)$.
(2) $E \in \mathrm{Lat}(T, T^*) \Leftrightarrow E \in \mathrm{Lat}(T), E^\perp \in \mathrm{Lat}(T)$.
(3) *For every* $E \in \mathrm{Lat}(T)$,

$$E = E_R \oplus E_N,$$

where $E_R \in \mathrm{Lat}(T, T^*)$ *(a reducing subspace of* T*) and* $E_N \in \mathrm{Lat}(T)$ *is a* completely non-reducing subspace*, i.e. such that* $E' \subset E_N$*,* $E' \in \mathrm{Lat}(T, T^*) \Rightarrow E' = \{0\}$*. This representation is unique.*

Proof (1) We first show the implication "⇒". 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 "⇐" 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 $\mathrm{Lat}(T, T^*)$. Set

$$E_R = \mathrm{span}\,(E' : E' \subset E, E' \in \mathrm{Lat}(T, T^*)), \quad E_N = E \ominus E_R.$$

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

Lemma 1.3.2 *Let* μ *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* $M_z : L^2(\mu) \to L^2(\mu)$*. 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 $M_z^* E = M_{\bar z} M_z E = E$, hence $E \in \mathrm{Lat}(M_z, M_z^*)$ 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

$$q \in E \ominus M_z E = E \cap (M_z E)^\perp, \quad \|q\| = 1.$$

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 $(\widehat{|q|^2\, d\mu})(n) = c\hat{m}(n)$ 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 $m = |q|^2\, d\mu_a = |q|^2 wm$, which is equivalent to $|q|^2 w = 1$ m-a.e.

(2) Let $f \in E_R$. Given that E_R is reducing and $M_z^* = M_{\bar z} = M_z^{-1}$, we have $z^n f \in E_R \subset E$ for all $n \in \mathbb{Z}$. Then $z^n f = z(z^{n-1} f) \in M_z E$, and by the definition of q we obtain

$$0 = (z^n f, q) = \int_{\mathbb{T}} z^n f \bar q \, d\mu, \quad \forall n \in \mathbb{Z}.$$

Therefore, $f\bar q = 0$ μ-a.e., hence μ_a-a.e., and thus (given that $m = |q|^2\, d\mu_a$) $f\bar q = 0$ 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. ■

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 $(z^n)_{n\in\mathbb{Z}} = (e^{int})_{n\in\mathbb{Z}}$ 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

$$f = \sum_{n\in\mathbb{Z}} \hat f(n) z^n,$$

norm-$L^2(\mathbb{T})$ convergent for the symmetric partial sums $\sum_{n=-N}^{N} \hat f(n) z^n$ (for $N \to \infty$), or even for "disordered" sums $\sum_{n\in\sigma(N)} \hat f(n) z^n$ where $\sigma(N) \subset \mathbb{Z}$, $\sigma(N) \nearrow \mathbb{Z}$ for $N \to \infty$:

$$\lim_N \Big\| f - \sum_{n\in\sigma(N)} \hat f(n) z^n \Big\|_{L^2(\mathbb{T})} = 0.$$

With this reminder, we can say

$$H^2(\mathbb{T}) = \Big\{ f \in L^2(\mathbb{T}) : f = \sum_{n\geq 0} \hat f(n) z^n \Big\} = \Big\{ \sum_{n\geq 0} a_n z^n : \sum_{n\geq 0} |a_n|^2 < \infty \Big\}.$$

Moreover, the use of properties of orthogonal bases leads to

$$H^2(\mathbb{T}) = \operatorname{span}_{L^2(\mathbb{T})} (z^n : n = 0, 1, \dots).$$