2018

The range and valence of a real Smirnov function

Timothy Ferguson

William T. Ross
University of Richmond, wross@richmond.edu

Recommended Citation

The range and valence of a real Smirnov function

Timothy Ferguson[1] · William T. Ross[2]

Abstract We give a complete description of the possible ranges of real Smirnov functions (quotients of two bounded analytic functions on the open unit disk where the denominator is outer and such that the radial boundary values are real almost everywhere on the unit circle). Our techniques use the theory of unbounded symmetric Toeplitz operators, some general theory of unbounded symmetric operators, classical Hardy spaces, and an application of the uniformization theorem. In addition, we completely characterize the possible valences for these real Smirnov functions when the valence is finite. To do so we construct Riemann surfaces we call disk trees by welding together copies of the unit disk and its complement in the Riemann sphere. We also make use of certain trees we call valence trees that mirror the structure of disk trees.

Keywords Hardy spaces · Real outer functions · Valence · Riemann surfaces · Conformal welding

Mathematics Subject Classification 30J05 · 30H10 · 46E22

✉ William T. Ross
    wross@richmond.edu

    Timothy Ferguson
    tjferguson1@ua.edu

[1] Department of Mathematics, University of Alabama, Tuscaloosa, AL, USA

[2] Department of Mathematics and Computer Science, University of Richmond, Richmond, VA 23173, USA

1 Introduction

This paper explores the range and valence of real Smirnov functions. Smirnov functions, a well studied class of functions [10], are analytic functions on the open unit disk $\mathbb{D}$ which can be written as the quotient of two bounded analytic functions where the denominator is an outer function. Real Smirnov functions, studied in [14, 15, 19, 20, 22], are those Smirnov functions which have real boundary values almost everywhere. In a nutshell, we will characterize all possible ranges of such functions and all possible finite valences on their range. The two main theorems of this paper (terminology, motivation, and plenty of examples to be reviewed below) are the following:

Theorem *If φ is a non-constant real Smirnov function, then $\varphi(\mathbb{D})$ is either*

$$\varphi(\mathbb{D}) = \mathbb{C}_+\backslash F \;\; or \;\; \varphi(\mathbb{D}) = \mathbb{C}_-\backslash G \;\; or \;\; \varphi(\mathbb{D}) = \mathbb{C}\backslash(F \cup G \cup E),$$

where $E \subsetneq \mathbb{R}$ and closed, $F \subseteq \mathbb{C}_+$ is relatively closed and has capacity zero, and $G \subseteq \mathbb{C}_-$ is relatively closed and has capacity zero. Moreover, given any closed $E \subsetneq \mathbb{R}$, any relatively closed $F \subseteq \mathbb{C}_+$ of capacity zero, and any relatively closed $G \subseteq \mathbb{C}_-$ of capacity zero, there are real Smirnov functions with ranges $\mathbb{C}_+\backslash F$, $\mathbb{C}_-\backslash G$, and $\mathbb{C}\backslash(E \cup F \cup G)$.

Theorem *The valence of every real Smirnov function with finite valence is given by the valence of a plane valence tree, and any valence arising from a plane valence tree is the valence of a real Smirnov function.*

The inspiration for this paper, and what informs our results, comes from the study of unbounded Toeplitz operators on the Hardy space H^2 of the open unit disk $\mathbb{D}$ (see [3, 26] and below). Here, for a general analytic function φ on $\mathbb{D}$, one can define the Toeplitz operator

$$T_\varphi : \mathcal{D}(T_\varphi) \to H^2, \quad T_\varphi f = \varphi f,$$

where $\mathcal{D}(T_\varphi)$, the domain of T_φ, is defined by

$$\mathcal{D}(T_\varphi) = \left\{ f \in H^2 : \varphi f \in H^2 \right\}.$$

Sarason [26] showed that $\mathcal{D}(T_\varphi) \neq \{0\}$ if and only if

$$\varphi = \frac{b}{a}, \tag{1.1}$$

where b and a are bounded analytic functions on $\mathbb{D}$ and a has no zeros. Such φ comprise the well-known *Nevanlinna class N* [10]. It can also be arranged so that

$$a(0) > 0 \;\; \text{and} \;\; |a|^2 + |b|^2 = 1$$

almost everywhere on the unit circle $\mathbb{T}$. With these normalizing conditions, this representation is unique. In the same paper, Sarason also showed that $\mathcal{D}(T_\varphi)$ is dense in H^2 if and only a in (1.1) is an outer function. These φ comprise the *Smirnov class* N^+. Observe from [10, Ch. 2] that

$$\bigcup_{p>0} H^p \subseteq N^+, \tag{1.2}$$

where H^p, the *Hardy classes*, are the analytic functions f on $\mathbb{D}$ for which the p-integral means

$$M_p(r, f) := \left(\int_0^{2\pi} |f(re^{it})|^p \frac{dt}{2\pi} \right)^{1/p} \tag{1.3}$$

are uniformly bounded for $r \in [0, 1)$.

Classical theorems of Fatou and Riesz [10, Ch.1, 2] say that for each $\varphi \in N^+$ the radial limit

$$\varphi(e^{it}) := \lim_{r \to 1^-} \varphi(re^{it}) \tag{1.4}$$

exists (and is non-zero) for almost every $t \in [0, 2\pi]$. We say $\varphi \in N^+$ belongs to the *real Smirnov class* $N_{\mathbb{R}}^+$ if

$$\varphi(e^{it}) \in \mathbb{R}$$

for almost every t (see some examples below). These real Smirnov functions have been studied in [14,15,19,20,22] and a full characterization of them was given by Helson [19,20] as

$$\varphi \in N_{\mathbb{R}}^+ \iff \varphi = i \frac{u+v}{u-v}, \tag{1.5}$$

where u and v are inner functions and $u - v$ is an outer function.

When $\varphi \in N_{\mathbb{R}}^+$ and

$$\langle f, g \rangle = \int_0^{2\pi} f(e^{it}) \overline{g(e^{it})} \frac{dt}{2\pi}$$

denotes the usual inner product on H^2 (considered in the usual way, via radial limit functions, as a closed subspace of L^2), one can use the fact that $\varphi(e^{it})$ is real for almost every t to see that

$$\langle T_\varphi f, g \rangle = \langle f, T_\varphi g \rangle, \quad f, g \in \mathcal{D}(T_\varphi).$$

In other words, T_φ is a densely defined symmetric operator on H^2.

Standard results from the theory of unbounded symmetric operators [2, Vol. II, Ch. VII] show that when $\varphi \in N_{\mathbb{R}}^+$ and $\lambda \notin \mathbb{R}$, the densely defined operator $T_\varphi - \lambda I$ has closed range and the deficiency numbers

$$d_\varphi(\lambda) := \dim(\text{Rng}(T_\varphi - \lambda I))^\perp, \tag{1.6}$$

where Rng denotes the range of an operator, are constant on each of the half planes

$$\mathbb{C}_+ = \{z : \Im z > 0\}, \quad \mathbb{C}_- = \{z : \Im z < 0\}.$$

Moreover, given a pair (m, n), where $m, n \in \mathbb{N}_0 \cup \{\infty\}$, there is a $\varphi \in N_{\mathbb{R}}^+$ with

$$(d_\varphi(i), d_\varphi(-i)) = (m, n).$$

It is also the case that both deficiency numbers are finite if and only if $\varphi \in N_{\mathbb{R}}^+$ is a rational function.

To get to our discussion of the range of φ, the focus of this paper, we unpack this a bit further as was done in [19]. Observe that $\mathrm{Rng}(T_\varphi - \lambda I)$ is not only a closed subspace of H^2 but it is also invariant under the shift operator $Sf = zf$ on H^2 and thus, by Beurling's theorem [10, Ch. 7],

$$\mathrm{Rng}(T_\varphi - \lambda I) = \Theta H^2$$

for some inner function Θ. Let Θ_λ denote the inner factor of $\varphi - \lambda$. All functions in ΘH^2 have Θ_λ as a divisor. Moreover, since we can write $\varphi - \lambda = b/a$ where b and a are in H^∞ and a is outer, it follows that the inner factor of $(\varphi - \lambda)$ is the inner factor of b, and thus the inner factor of $(\varphi - \lambda)a$ is precisely Θ_λ. Thus $\Theta = \Theta_\lambda$. This means that

$$(\mathrm{Rng}(T_\varphi - \lambda I))^\perp = (\Theta_\lambda H^2)^\perp$$

which is a model space [5,9,13,24], a typical invariant subspace for the backward shift operator S^*. Moreover, the model space $(\Theta_\lambda H^2)^\perp$ has finite dimension n if and only if Θ_λ is a finite Blaschke product of degree n. Since

$$\ker(T_\varphi^* - \bar{\lambda} I) = (\mathrm{Rng}(T_\varphi - \lambda I))^\perp = (\Theta_\lambda H^2)^\perp,$$

and for each $w \in \mathbb{D}$,

$$T_\varphi^* k_w = \overline{\varphi(w)} k_w, \quad k_w(z) = \frac{1}{1 - \bar{w} z},$$

we see that

$$\bigvee \{k_w : \varphi(w) = \lambda\} \subseteq \ker(T_\varphi^* - \bar{\lambda} I) = (\Theta_\lambda H^2)^\perp.$$

In the above, $\bigvee$ denotes the closed linear span. Hence, using the fact that Θ_λ is the inner factor for $\varphi - \lambda$, we see that for $\lambda \in \mathbb{C} \backslash \mathbb{R}$,

$$\lambda \in \varphi(\mathbb{D}) \iff \Theta_\lambda(w) = 0 \text{ for some } w \in \mathbb{D}.$$

Furthermore, the valence

$$v_\varphi(\lambda) := \mathrm{card}\{w \in \mathbb{D} : \varphi(w) = \lambda\} \tag{1.7}$$

will be the degree of the Blaschke factor of Θ_λ. For example, if the inner factor of Θ_λ is either a unimodular constant or a singular inner function (which will have no zeros in $\mathbb{D}$), then $\lambda \notin \varphi(\mathbb{D})$, i.e., $v_\varphi(\lambda) = 0$. On the other hand, if Θ_λ has is an infinite Blaschke factor, then $v_\varphi(\lambda) = \infty$. Note that

$$v_\varphi(\lambda) \leqslant d_\varphi(\lambda), \quad \lambda \notin \mathbb{R},$$

but equality does not always hold. For example, Θ_λ might be the product of a finite Blaschke product of degree n and a singular inner function. In this case $v_\varphi(\lambda) = n$ while $d_\varphi(\lambda) = \infty$.

Thus, characterizing the range of φ will involve a discussion of the $\lambda \in \mathbb{C}\backslash\mathbb{R}$ such that $\varphi - \lambda$ has a non-trivial Blaschke factor. Rudin [25], generalizing a classical theorem of Frostman [6, p. 37], showed that for nearly all $\lambda \in \mathbb{C}\backslash\mathbb{R}$, the inner factor of $\varphi - \lambda$ is a Blaschke product. Here "nearly all" means that this property holds with the possible exceptional set of logarithmic capacity (capacity for short) zero. See [11,23] for basic facts about logarithmic capacity and see [7,12,21] for more on Blaschke products.

In the above, we are allowing a unimodular constant to count as a Blaschke factor (of order zero). In this degenerate case we see that $\Theta_\lambda \equiv \xi$ for some $\xi \in \mathbb{T}$ and so

$$(\mathrm{Rng}(T_\varphi - \lambda I))^\perp = (\xi H^2)^\perp = \{0\}.$$

Thus if Θ_λ is a unimodular constant function for one $\lambda \in \mathbb{C}_+$ (or one $\lambda \in \mathbb{C}_-$) then, since d_φ is constant on each of $\mathbb{C}_+$ or $\mathbb{C}_-$, it follows that Θ_λ is a constant unimodular function for all $\lambda \in \mathbb{C}_+$ (or all $\lambda \in \mathbb{C}_-$). If Θ_λ is a constant unimodular function for one $\lambda \in \mathbb{C}_+$, then $\varphi(\mathbb{D}) \cap \mathbb{C}_+ = \varnothing$ (similarly for some $\lambda \in \mathbb{C}_-$ and hence $\varphi(\mathbb{D}) \cap \mathbb{C}_- = \varnothing$). Hence, for example, if $\varphi(\mathbb{D}) \cap \mathbb{C}_+$ omits an open disk about $\lambda \in \mathbb{C}_+$, then $\varphi - \lambda$ is an outer function, i.e., Θ_λ is a constant unimodular function. In this case the above discussion implies that $\varphi(\mathbb{D}) \cap \mathbb{C}_+ = \varnothing$.

Thus, using the above analysis, along with the fact that $\varphi(\mathbb{D})$ is an open connected subset of $\mathbb{C}$ (open mapping theorem), we have the following possibilities for the range of $\varphi \in N_{\mathbb{R}}^+$:

$$\varphi(\mathbb{D}) = \mathbb{C}_+\backslash F \;\; \mathrm{or} \;\; \varphi(\mathbb{D}) = \mathbb{C}_-\backslash G \;\; \mathrm{or} \;\; \varphi(\mathbb{D}) = \mathbb{C}\backslash(F \cup G \cup E), \tag{1.8}$$

where $F \subseteq \mathbb{C}_+$ and $G \subseteq \mathbb{C}_-$ are relatively closed subsets of capacity zero and $E \subsetneq \mathbb{R}$ is closed. The question we ask and answer in this paper is whether or not we can actually obtain all of these possibilities. In addition, we also discuss the valence on these ranges.

To give the reader a feel for where we are heading, let us consider a few simple examples of ranges of $\varphi \in N_{\mathbb{R}}^+$.

Example 1.1 If

$$\varphi_1(z) = i\frac{1+z}{1-z},$$

then $\varphi_1 \in N^+$ (since it is the quotient of two bounded analytic functions and the denominator $1 - z$ is outer) and

$$\varphi_1(e^{it}) = -\cot(t/2) \in \mathbb{R},$$

which says that $\varphi_1 \in N_{\mathbb{R}}^+$. Furthermore, $\varphi_1(\mathbb{D}) = \mathbb{C}_+$. In a similar way, we see that if

$$\varphi_2(z) = -i\frac{1+z}{1-z},$$

then $\varphi_2 \in N_{\mathbb{R}}^+$ and $\varphi_2(\mathbb{D}) = \mathbb{C}_-$.

If θ is the singular inner function

$$\theta(z) = \exp\left(\frac{1+z}{1-z}\right),$$

then $\theta(\mathbb{D}) = \mathbb{D}\backslash\{0\}$ and thus if $\psi_1 := \varphi_1 \circ \theta$, then $\varphi_1 \in N_{\mathbb{R}}^+$ and $\psi_1(\mathbb{D}) = \mathbb{C}_+\backslash\{i\}$. Observe that the singleton $\{i\}$ has capacity zero [23, p. 140].

Given a relatively closed subset $W \subseteq \mathbb{D}$ of capacity zero, there is an inner function σ such that $\sigma(\mathbb{D}) = \mathbb{D}\backslash W$ [6]. Then $\psi_2 = \varphi_1 \circ \sigma \in N_{\mathbb{R}}^+$ and $\psi_2(\mathbb{D}) = \mathbb{C}_+\backslash F$, where $F = \psi_1(W)$ has capacity zero.

Example 1.2 If

$$\varphi_3(z) = \left(\frac{1+z}{1-z}\right)^4$$

then

$$\varphi_3(e^{it}) = \cot^4(t/2) \in \mathbb{R}$$

and, since

$$z \mapsto \frac{1+z}{1-z}$$

maps $\mathbb{D}$ onto the right-half plane $\{z : \Re z > 0\}$, then $\varphi_3(\mathbb{D}) = \mathbb{C}\backslash\{0\}$.

Example 1.3 If

$$\varphi_4(z) = \frac{z}{(1-z)^2},$$

the well-known Koebe function, then

$$\varphi_4(e^{it}) = -\frac{1}{2}\frac{1}{1-\cos t} \in \mathbb{R}$$

and $\varphi_4(\mathbb{D})$ is the single slit domain $\mathbb{C}\backslash(-\infty, -\frac{1}{4}]$.

Example 1.4 If

$$\varphi_5(z) = \frac{iz}{1-z^2},$$

then

$$\varphi_5(e^{it}) = -\frac{1}{2}\csc t \in \mathbb{R}$$

and $\varphi_5(\mathbb{D})$ turns out to be the double slit domain

$$\mathbb{C}\backslash\left((-\infty, -\tfrac{1}{2}]\cup[\tfrac{1}{2}, \infty)\right).$$

Of course one can compose any of the functions φ_j from these examples with interesting inner functions, like was done in the first example, to obtain ranges taking the form $\mathbb{C}_+\backslash F$, $\mathbb{C}_-\backslash G$, and $\mathbb{C}\backslash(F \cup G \cup E)$ for relatively closed sets F and G of capacity zero and a closed set $E \subsetneq \mathbb{R}$. Can we obtain all of these possibilities as ranges for given E, F, G?

2 The main range result

Our main result about the range of $\varphi \in N_{\mathbb{R}}^+$ is the following:

Theorem 2.1 *If $\varphi \in N_{\mathbb{R}}^+$ and non-constant, then $\varphi(\mathbb{D})$ is either*

$$\varphi(\mathbb{D}) = \mathbb{C}_+\backslash F \ \text{ or } \ \varphi(\mathbb{D}) = \mathbb{C}_-\backslash G \ \text{ or } \ \varphi(\mathbb{D}) = \mathbb{C}\backslash(F \cup G \cup E),$$

where $E \subsetneq \mathbb{R}$ and closed, $F \subseteq \mathbb{C}_+$ is relatively closed and has capacity zero, and $G \subseteq \mathbb{C}_-$ is relatively closed and has capacity zero. Moreover, given any closed $E \subsetneq \mathbb{R}$, any relatively closed $F \subseteq \mathbb{C}_+$ of capacity zero, and any relatively closed $G \subseteq \mathbb{C}_-$ of capacity zero, there are functions in $N_{\mathbb{R}}^+$ with ranges $\mathbb{C}_+\backslash F$, $\mathbb{C}_-\backslash G$, and $\mathbb{C}\backslash(E \cup F \cup G)$.

Our proof needs a variation of a result from [11, p. 119].

Lemma 2.2 *Suppose f is a non-constant function belonging to H^p for some $p \in (0, \infty)$ and $\mathcal{E} \subseteq \mathbb{T}$ of positive Lebesgue measure for which*

$$\lim_{r\to 1^-} f(r\xi) =: f(\xi)$$

exists for each $\xi \in \mathcal{E}$. If $E = \{f(\xi) : \xi \in \mathcal{E}\}$, then E has positive capacity.

Proof We follow the original proof from [11, p. 119] with a slight variation. First, note that since the almost everywhere defined boundary $\xi \mapsto f(\xi)$ function on $\mathbb{T}$ is a limit of continuous functions (the dilations $f_r(\xi) := f(r\xi)$) on $\mathbb{T}$, by Egorov's theorem there is a set of positive measure that is a subset of $\mathcal{E}$ on which the boundary function f is continuous. By the inner regularity of Lebesgue measure, this set has a compact subset of positive measure. Without loss of generality we may take $\mathcal{E}$ to be this set. Then $f(\mathcal{E})$ is compact, and $|f| \leqslant K$ on $\mathcal{E}$ for some $K > 0$. Suppose towards a contradiction that E has zero logarithmic capacity. Then by Evan's theorem [11, p. 33], for some probability measure σ on $\mathbb{T}$, the logarithmic potential

$$p(z) = -\int_E \log|z - \zeta| d\sigma(\zeta)$$

satisfies

$$\lim_{z \to \xi} p(z) = +\infty, \quad \xi \in E.$$

The function $u = p(f)$ is harmonic on $\mathbb{D}$ and satisfies the condition

$$\lim_{r \to 1^-} u(r\xi) = +\infty, \quad \xi \in \mathcal{E}.$$

Let v be the harmonic conjugate of u on $\mathbb{D}$ and define $F = e^{-u-iv}$. Assuming that $F \in H^p$ (a fact we will prove momentarily), we see that

$$\lim_{r \to 1^-} |F(r\xi)| = 0, \quad \xi \in \mathcal{E}.$$

But since $\mathcal{E}$ has positive Lebesgue measure, this would mean that $f \equiv 0$ [10, p. 17], a contradiction.

To finish, we now show that $F \in H^p$. Observe that a use of Jensen's inequality [23, p. 43] and the integral means definition of H^p from (1.3) shows that

$$\begin{aligned}
\int_0^{2\pi} |F(re^{it})|^p \frac{dt}{2\pi} &= \int_0^{2\pi} \exp\left(\int_E \log|f(re^{it}) - \zeta|^p d\sigma(\zeta)\right) \frac{dt}{2\pi} \\
&\leqslant \int_0^{2\pi} \left(\int_E |f(re^{it}) - \zeta|^p d\sigma(\zeta)\right) \frac{dt}{2\pi} \\
&= \int_E \left(\int_0^{2\pi} |f(re^{it}) - \zeta|^p \frac{dt}{2\pi}\right) d\sigma(\zeta) \\
&\leqslant 2^p \left(\sup_{0 < r < 1} M(r, f)^p + K^p\right).
\end{aligned}$$

Since the last quantity above is independent of $r \in (0, 1)$, this shows that $F \in H^p$ and thus completes the proof. $\qquad\square$

Proof of Theorem 2.1 The first part of the theorem follows from the discussion preceding (1.8). For the second part (obtaining all possible types of ranges), we proceed as follows. First consider the case where the range contains points in both the upper and lower half planes. Since E is a proper closed subset of $\mathbb{R}$, it follows that $\mathbb{R}\setminus E$ has at least one component. If it has *exactly one* component, then E must be one of the following four options:

$$\varnothing, \quad (-\infty, c], \quad [c, \infty), \quad (-\infty, a] \cup [b, \infty), \quad (a < b). \tag{2.1}$$

We will first deal with the case where $\mathbb{R}\setminus E$ has *at least two* components. By means of a real translation of our function at the end, we can assume $0 \in E$ and, for some $a < 0$ and $0 < b < c$, that

$$E \subseteq (-\infty, a] \cup [0, b] \cup [c, \infty). \tag{2.2}$$

Define

$$E_+ = E \cap [0, \infty), \quad E_- = E \cap (-\infty, 0)$$

and the open set

$$\Omega = \{\Re z > 0\}\setminus(\widetilde{E}_1 \cup \widetilde{E}_2 \cup \widetilde{E}_3 \cup \widetilde{F} \cup \widetilde{G}), \tag{2.3}$$

where

$$\widetilde{E}_1 = \{x^{\frac{1}{4}} : x \in E_+\},$$

$$\widetilde{E}_2 = e^{i\frac{\pi}{4}}\{(-x)^{\frac{1}{4}} : x \in E_-\},$$

$$\widetilde{E}_3 = e^{-i\frac{\pi}{4}}\{(-x)^{\frac{1}{4}} : x \in E_-\},$$

and $\widetilde{F}$ and $\widetilde{G}$ are the intersections of the right half plane with the images of F and G respectively under the multivalued map $z \mapsto z^{1/4}$. See Fig. 1 for an illustration of Ω when

$$E = (-\infty, a] \cup [0, b] \cup [c, \infty).$$

Since Ω is contained in

$$\{\Re z > 0\}\setminus \left([0, b^{\frac{1}{4}}] \cup [c^{\frac{1}{4}}, \infty) \cup e^{i\pi/4}[|a|^{\frac{1}{4}}, \infty) \cup e^{-i\pi/4}[|a|^{\frac{1}{4}}, \infty)\right)$$

and this last set is connected, one concludes, also using the containment in (2.2) along with the fact that extracting a set of capacity zero does not disconnect a domain [23, p. 68], that Ω is connected. By [8, p. 125] there exists an analytic covering map for

Fig. 1 The region Ω when
$E = (-\infty, a] \cup [0, b] \cup [c, \infty)$
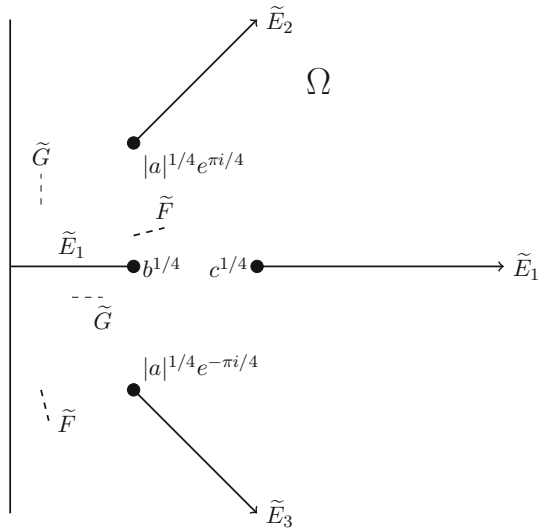
Ω, such that $\psi(\mathbb{D}) = \Omega$. Furthermore, since $\psi(\mathbb{D})$ is contained in a half-plane, then ψ belongs to H^p for all $p \in (0, 1)$ [16, p. 109]. Thus, by (1.2), $\psi \in N^+$.

The map ψ is a covering map from $\mathbb{D}$ to Ω, which means that each point of Ω is contained in an open neighborhood U such that $\psi^{-1}(U)$ consists of disjoint open sets each of which is homeomorphic to U under ψ.

We now claim that if the radial limit

$$\lim_{r \to 1^-} \psi(re^{it}) \tag{2.4}$$

exists, which it will for almost every t [see (1.4)], then this value must belong to $\partial\Omega$. Indeed, if this were not the case, then, for some particular t, the limit would be equal to some $w \in \Omega$. Now choose an open neighborhood U of w such that $\psi^{-1}(U)$ consists of disjoint sets each of which is homeomorphic to U under ψ. Let W be an open neighborhood of w contained in U and such that $\overline{W} \cap U$ is compact. Then

$$\psi^{-1}(W) = \bigcup_{a \in A} V_a,$$

where A is some index set and the V_a are pairwise disjoint open sets that are each homeomorphic to W under ψ. Thus each V_a has compact closure in $\psi^{-1}(U)$. For some $b \in [0, 1)$, the curve

$$r \to \psi([re^{it}, e^{it})), \quad r \in [b, 1),$$

must lie entirely in W since ψ has radial limit of w at e^{it}. But since the V_a are disjoint open sets, this means that the ray

$$[re^{it}, e^{it}), \quad r \in [b, 1),$$

must lie entirely in one of the V_a. But this is impossible because each of the V_a has compact closure in $\psi^{-1}(U)$ but $e^{it} \notin \psi^{-1}(U)$.

The set $\widetilde{F}$ is the union of two images of F under maps that are Lipschitz in any annulus centered at the origin. Since F has capacity 0, each of the images has capacity 0 [23, p. 137], and their union $\widetilde{F}$ also has capacity zero [23, p. 57]. The same applies to $\widetilde{G}$, and thus $\widetilde{F} \cup \widetilde{G}$ has capacity zero.

By Lemma 2.2, the values in those sets cannot be boundary values of ψ, except possibly on a set of measure 0.

Setting $\varphi = \psi^4$ we see that $\psi \in H^p$ for all $p \in (0, \frac{1}{4})$ and thus, again from (1.2), $\psi \in N^+$. Moreover, since

$$\lim_{r \to 1^-} \psi(re^{it}) \in \partial\Omega \backslash (\widetilde{F} \cup \widetilde{G})$$

for almost every θ, we see that

$$\lim_{r \to 1^-} \varphi(re^{it}) \in \mathbb{R}$$

for almost every t. Thus $\varphi \in N_{\mathbb{R}}^+$. The construction of Ω from (2.3), and the fact that $0 \in E$, will show that

$$\varphi(\mathbb{D}) = \{z^4 : z \in \Omega\} = \mathbb{C} \backslash (E \cup F \cup G).$$

For the cases where the desired range is $\mathbb{C}_+ \backslash F$ or $\mathbb{C}_- \backslash G$, and for the case where $E = (-\infty, c]$ or $E = [c, \infty)$ or $E = (-\infty, a] \cup [b, \infty)$ and the desired range is $\mathbb{C} \backslash (E \cup F \cup G)$, we let φ be the covering map from $\mathbb{D}$ onto the desired range. The proof that this map has the required properties is similar to the proof of the first case above. We use the fact that any mapping from $\mathbb{D}$ into a domain with at least one slit that goes to ∞ is in H^p for all $p \in (0, \frac{1}{2})$ [16, p. 110].

If the desired range is $\mathbb{C} \backslash (F \cup G)$, we let $\widetilde{F}$ and $\widetilde{G}$ be the images of F and G respectively under the multivalued map $z \mapsto z^{1/2}$—which are of capacity zero [23, p. 137]. Let ψ be the covering map from $\mathbb{D}$ onto $\mathbb{C} \backslash (\widetilde{F} \cup \widetilde{G} \cup [1, \infty))$, and let $\varphi = \psi^2$. The proof that mapping has the required properties is similar to the proofs above. □

Remark 2.3 1. The mapping φ constructed above is actually outer, as long as 0 is not in the range. To see this, observe that, in the first part of the proof, $\psi(\mathbb{D}) \subseteq \{\Re z > 0\}$. Such functions are outer [16, p. 109]. Since the product of outer functions is another outer function, this means that $\varphi = \psi^4$ is outer. In the other cases in which 0 is not in the range, we have that φ is a map onto a domain with a linear slit containing 0 and going to ∞. But any map onto such a domain is outer, since we can define the square root of such a mapping, and that mapping will be onto a half plane omitting 0 and thus outer (and belong to H^p for all $p \in (0, 1)$ [16, p. 109]).

2. The proof of Theorem 2.1 also shows that we can find a φ with the desired properties that is in H^p for each $p \in (0, \frac{1}{4})$.

3 Controlling the valence

A key step in the construction in Theorem 2.1 was the uniformization theorem [8, p. 125]. However, it is not clear from our construction how one can control the valence of φ.

In this regard, one can ask the following question: Suppose we are given an closed set $E \subsetneq \mathbb{R}$ and a pair (m, n), $m, n \in \mathbb{N}_0 \cup \{\infty\}$. Can we find a $\varphi \in N_{\mathbb{R}}^+$ such that the valence of φ is equal to m on $\mathbb{C}_+$, n on $\mathbb{C}_-$, and such that $\varphi(\mathbb{D}) = \mathbb{C}\backslash E$? Can we say anything about the valence of φ on $\mathbb{R}\backslash E$?

Let us start with a few observations. Extending a standard proof of the open mapping theorem for analytic functions, one can prove the following. Recall that v_φ is the valence function from (1.7) and d_φ is the deficiency index from (1.6).

Proposition 3.1 *For $\varphi \in N_{\mathbb{R}}^+$ and $N = 1, 2, \ldots$, the set*

$$\{w \in \mathbb{C} : v_\varphi(w) \geqslant N\}$$

is an open subset of $\mathbb{C}$.

Note that the above result does not hold when $N = \infty$.

Example 3.2 Theorem 2.1 says that we can find a $\varphi \in N_{\mathbb{R}}^+$ such that

$$\varphi(\mathbb{D}) = \mathbb{C}\backslash[0, \infty).$$

However, the previous proposition says that we can't find a $\varphi \in N_{\mathbb{R}}^+$ with the same range and that also satisfies

$$v_\varphi|_{\mathbb{C}_+} = 1, \quad v_\varphi|_{\mathbb{C}_-} = 2, \quad v_\varphi|_{\{x<0\}} = 3$$

since $v_\varphi|_{\mathbb{C}_+}$ and $v_\varphi|_{\mathbb{C}_-}$ must be at least 3.

Next we explore when the valence is finite.

Proposition 3.3 *For $\varphi \in N_{\mathbb{R}}^+$ the following are equivalent.*

(i) *φ is a rational function;*
(ii) *$v_\varphi|_{\mathbb{C}_+}$ and $v_\varphi|_{\mathbb{C}_-}$ are finite;*
(iii) *There are two relatively prime finite Blaschke products B_1, B_2 such that $B_1 - B_2$ has no zeros on $\mathbb{D}$ and such that*

$$\varphi = i\frac{B_1 + B_2}{B_1 - B_2}. \tag{3.1}$$

(iv) *$d_\varphi(i)$ and $d_\varphi(-i)$ are finite.*

Furthermore,

(a) *if any of the above equivalent conditions hold, we have*

$$v_\varphi|_{\mathbb{C}_+} = \deg(B_2), \quad v_\varphi|_{\mathbb{C}_-} = \deg(B_1).$$

(b) *if any of the above conditions do not hold then either $v_\varphi|_{\mathbb{C}_+}$ or $v_\varphi|_{\mathbb{C}_-}$ is infinite nearly everywhere.*

Proof of Theorem 2.1 $(i) \iff (iv)$ is from [19].

$(iii) \implies (i)$: One can show directly that a φ given by (3.1) is a rational function in $N_\mathbb{R}^+$.

$(i) \implies (iii)$: Set $g = (\varphi - i)/(\varphi + i)$ and observe that g is meromorphic function on $\mathbb{D}$ that is continuous with unimodular boundary values on $\mathbb{T}$. A classical theorem of Fatou [12] says that

$$g = \frac{B_1}{B_2},$$

where B_1 and B_2 are relatively prime Blaschke products. The result now follows with a simple computation.

$(iii) \implies (ii)$: We follow an argument from [12]. The Möbius transformation

$$\psi(z) = \frac{z - i}{z + i}$$

is injective and maps $\mathbb{C}_+$ onto $\mathbb{D}$ and $\mathbb{C}_-$ onto $\widehat{\mathbb{C}} \backslash \mathbb{D}^-$. If $w \in \mathbb{C}_+$, the number of solutions to $\varphi(z) = w$ is the same as the number of solutions to $\psi \circ \varphi(z) = \psi(w) = \eta \in \mathbb{D}$. Writing this out, this is same as the number of solutions to

$$\eta = \frac{\varphi(z) - i}{\varphi(z) + i} = \frac{B_2(z)}{B_1(z)}.$$

To examine the number of zeros in $\mathbb{D}$ of $B_2 - \eta B_1$, observe that on $\mathbb{T}$ we have

$$|\eta B_1| = |\eta| < 1 = |B_2|$$

and so, by Rouche's Theorem, the number of zeros in $\mathbb{D}$ of B_2 and $B_2 - \eta B_1$ are the same. This proves that $v_\varphi(w) = \deg(B_2)$ whenever $w \in \mathbb{C}_+$. The corresponding valence on $\mathbb{C}_-$ follows in a similar way. This also verifies (a).

$(ii) \implies (iv)$: Suppose $v_\varphi|_{\mathbb{C}_+}$ is finite but $d_\varphi(i) = \infty$. Since $d_\varphi|_{\mathbb{C}_+}$ is constant, we see that $d_\varphi(\lambda) = \infty$ for all $\lambda \in \mathbb{C}_+$. By [25], the inner factor Θ_λ of $\varphi - \lambda$ is a Blaschke product for all $\lambda \notin \mathbb{R}$ except possibly for a set of capacity zero. As discussed in (1.7),

$$v_\varphi(\lambda) = \deg(\Theta_\lambda) = \infty$$

for nearly all $\lambda \in \mathbb{C}_+$, a contradiction. An analogous argument holds for $v_\varphi|_{\mathbb{C}_-}$. This also proves (*b*). □

The techniques in the above proof also show the following.

Corollary 3.4 *If* $\varphi \in N_{\mathbb{R}}^+$ *and* $v_\varphi|_{\mathbb{C}_+}$ *is non-constant, then* $v_\varphi(\lambda) = \infty$ *for nearly all* $\lambda \in \mathbb{C}_+$. *An analogous result holds for* $v_\varphi|_{\mathbb{C}_-}$

Remark 3.5 We point out a paper [27] which gives some information about how one can, under mild technical conditions, define an inner function whose valence can be prescribed on various closed subsets of the $\mathbb{D}$ of zero capacity. Since real Smirnov functions take the form $i(u+v)/(u-v)$ (recall Helson's characterization from (1.5)), where u and v are inner, this creates examples of real Smirnov functions with wild valence behavior.

We now characterize the possible valences of real Smirnov functions of finite valence. As seen in Proposition 3.3, these are the rational real Smirnov functions.

A *disk tree* is a type of Riemann surface made by welding together copies of $\mathbb{D}$ and

$$\mathbb{D}^* := \left\{ z \in \widehat{\mathbb{C}} : |z| > 1 \text{ or } z = \infty \right\}.$$

To each copy of $\mathbb{D}$ and $\mathbb{D}^*$ in the disk tree is associated a positive integer m which we call the valence. A *admissible* arc on a disk tree is an arc on the boundary of a copy of $\mathbb{D}$ or $\mathbb{D}^*$ of the form $\{e^{i\theta} : a < \theta < b\}$ where (a, b) contains no multiple of $2\pi/m$. The *image arc* for a given admissible arc is the arc $\{e^{i\theta} : am < \theta < bm\}$. This is the image of the admissible arc under the function z^m. An admissible arc is called a *free arc* if it is part of the boundary of the disk tree, in other words, if it has *not* been welded to another arc in the disk tree. An admissible arc is a *welded arc* if it has been welded to another arc in the disk tree.

We will now formally define disk trees inductively. A copy of $\mathbb{D}$ or $\mathbb{D}^*$ with a valence is a disk tree. Let X be a disk tree. Let Z be a new copy of $\mathbb{D}$ or $\mathbb{D}^*$ with valence m. Let Y be a copy of $\mathbb{D}$ or $\mathbb{D}^*$ in X with valence n, where Y is a copy of $\mathbb{D}$ if Z is a copy of $\mathbb{D}^*$, and vice versa. Suppose we are given a free arc on Y and a free arc on Z and that both of them have the same image arc. We may weld Y and Z together on their free arcs by the map $Y \ni z \mapsto z^{n/m} \in Z$. See [1, II.3C] for more on welding Riemann surfaces. We will also give an explicit example of welding in Example 3.6. We say the resulting surface of X welded to Z is a disk tree if it still has a free arc remaining.

Any disk tree is simply connected (van Kampen's theorem [17, Ch. 1]). It is also conformally equivalent to the disk. Indeed, by the uniformization theorem, it is equivalent to the disk, the plane, or the Riemann sphere. But we may weld another copy of the disk (or complement of the disk) to any disk tree and still obtain a simply connected Riemann surface, since any disk tree has a free arc. Thus we can obtain the original disk tree from the new Riemann surface by removing a set with infinitely many points. But if we take a set with infinitely many points away from either the Riemann sphere, the plane, or the disk, and we are left with a simply connected set, that set must be conformally equivalent to the disk.

For a disk tree X, define

$$f_X : X \mapsto \widehat{\mathbb{C}}\backslash\{1\}$$

by $f_X(z) = z^m$ for z in a copy of $\mathbb{D}$ or $\mathbb{D}^*$ with valence m, or for z in a welded arc that belongs to a copy of $\mathbb{D}$ with valence m. The function f_X is clearly meromorphic in each copy of $\mathbb{D}$ or $\mathbb{D}^*$. By construction, it is continuous in a neighborhood of each welded arc. To see this, suppose that some admissible arc I in a copy of $\mathbb{D}$ or $\mathbb{D}^*$ with valence n (call the copy Y) is welded to an admissible arc J in a copy of $\mathbb{D}$ or $\mathbb{D}^*$ of valence m (call the copy Z). Let Y' and Z' be sufficiently small neighborhoods of I and J in $Y \cup I$ and $Z \cup J$, respectively. Note that by definition of the welding the map

$$\varphi(z) = \begin{cases} z^{n/m} & \text{if } z \in Y' \\ z & \text{if } z \in Z' \end{cases}$$

is well defined, continuous and even conformal. But $f_X(z) = \varphi(z)^m$ in $Y' \cup Z'$. Thus by Morera's theorem, f_X is analytic in a neighborhood of each welded arc and thus meromorphic on X. When restricted to a copy of $\mathbb{D}$ or $\mathbb{D}^*$ that has valence m the function f_X has valence m at each point of $\mathbb{D}$ or $\mathbb{D}^*$ respectively, and valence 0 elsewhere. Also, when restricted to a welded arc, the function f_X has valence 1 on the image arc and 0 elsewhere. Thus, f_X has valence on $\mathbb{D}$ equal to the sum of the valences of the copies of $\mathbb{D}$ in the disk tree, and similarly for $\mathbb{D}^*$. Its valence on a point in $\partial\mathbb{D}\backslash\{1\}$ is equal to the number of image arcs in which it appears, where if two arcs are welded together in the disk tree we count their image arc (which is the same for both the welded arcs) as appearing only once. Let $\varphi : \mathbb{D} \mapsto X$ be a conformal map from $\mathbb{D}$ to X. Then $f_X \circ \varphi$ is a (meromorphic) map from $\mathbb{D}$ to $\widehat{\mathbb{C}}\backslash\{1\}$.

Example 3.6 We give an explicit construction of a disk tree. This disk tree will have valence 1 on $\mathbb{D}$ and 2 on $\mathbb{D}^*$. See Fig. 2 for an illustration of some aspects of this example. Let X consist of one copy each of $\mathbb{D}$ and $\mathbb{D}^*$, together with the boundary of $\partial\mathbb{D}\backslash\{1\}$. For $0 < \theta < 2\pi$, identify the point $e^{i\theta}$ on $\partial\mathbb{D}\backslash\{1\}$ with the point $e^{i\theta/2}$ on $\partial\mathbb{D}^*\backslash\{1\}$. We will weld along these identified boundary points. To do this explicitly, let

$$\overline{\mathbb{D}} \supset U_1 = \left\{ re^{i\theta} : 1/2 < r \leqslant 1 \text{ and } 0 < \theta < 2\pi \right\}$$

and

$$\overline{\mathbb{D}^*} \supset U_2 = \left\{ re^{i\theta} : 1 \leqslant r < 2 \text{ and } 0 < \theta < \pi \right\}$$

and let $U = U_1 \cup U_2$. Take coordinate charts $\varphi_1 : \mathbb{D} \to \mathbb{C}$ and $\varphi_2 : \mathbb{D}^* \to \mathbb{C}$ and $\varphi_3 : U \to \mathbb{C}$, where $\varphi_1(z) = z$ and $\varphi_2(z) = 1/z$ and

$$\varphi_3(z) = \begin{cases} z^{1/2} & \text{if } z \in U_1, \\ z & \text{if } z \in U_2 \end{cases}$$

Fig. 2 An illustration of φ_3 and f_X on U, from Example 3.6
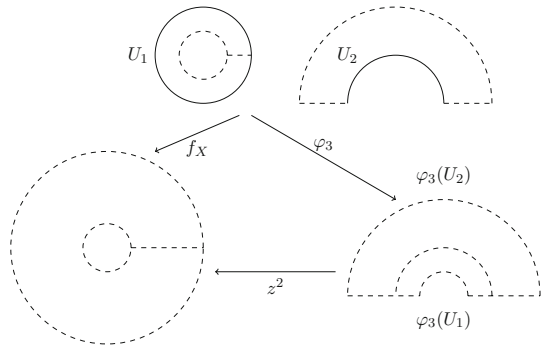
where we take the branch of $z^{1/2}$ with $(-1)^{1/2} = i$ and branch cut along the positive real axis. We can take as a basis for the open sets in X sets that are open in $\mathbb{D}$ or in $\mathbb{D}^*$ or sets that are the inverse images of open sets under φ_3. Thus U is an open set. (Note that any set that is the inverse image of an open set under φ_3 and lies entirely in $\mathbb{D}$ is open in $\mathbb{D}$; the same may be said for $\mathbb{D}^*$.) Notice that φ_3 is continuous, and that $\varphi_1(\varphi_3^{-1}(z)) = z^2$ and $\varphi_2(\varphi_3^{-1}(z)) = z^{-1}$. Both of these maps are analytic in their domains. Thus we have made X into a Riemann surface with the given charts.

Define

$$f_X(z) = \begin{cases} z & \text{for} \quad z \in \mathbb{D} \quad \text{or} \quad z \in \partial\mathbb{D} \\ z^2 & \text{for} \quad z \in \mathbb{D}^* \quad \text{or} \quad z \in \partial\mathbb{D}^*. \end{cases}$$

Then f_X is analytic from X into $\widehat{\mathbb{C}}\backslash\{1\}$. To see this, note that

$$f_X(\varphi_1^{-1}(z)) = z, \quad f_X(\varphi_2^{-1}(z)) = 1/z^2, \quad f_X(\varphi_3^{-1}(z)) = z^2.$$

Also, f has valence 1 on $\mathbb{D}$, valence 2 on $\mathbb{D}^*$, and valence 1 on $\partial\mathbb{D}\backslash\{1\}$.

We now define some types of graphs which we need to state the main result. A *plane valence tree* is a graph that is a tree. To each node is associated a label of either $\mathbb{C}_+$ or $\mathbb{C}_-$ and positive integer m, called the valence. A node with label $\mathbb{C}_+$ may only be adjacent to nodes labeled $\mathbb{C}_-$, and nodes labeled $\mathbb{C}_-$ may only be adjacent to nodes labeled $\mathbb{C}_+$. To each edge is associated an open interval in $\mathbb{R}$. The interval may be all of $\mathbb{R}$ but may not be empty. We make the requirement that a disjoint union of all the intervals on edges coming from a node is a subset of a disjoint union of m copies of $\mathbb{R}$, where m is the valence of the node. We require that some node has the property that a disjoint union of m copies of $\mathbb{R}$, where m is the valence of the node, contains a disjoint union of all the intervals on edges coming from the node, as well as another open interval in $\mathbb{R}$. We say that such a node has a free interval. For a plane valence tree, the valence of a point in $\mathbb{C}_+$ is the sum of the valences of the $\mathbb{C}_+$ nodes, and similarly for $\mathbb{C}_-$. For a point in $\mathbb{R}$, it is the number of times it appears in an edge of the valence tree.

Fig. 3 An example plane valence tree
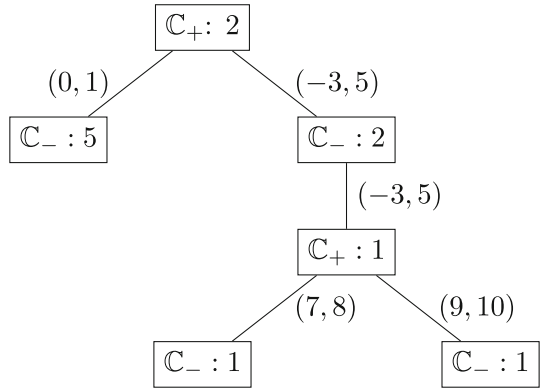
Define

$$\psi = -i\frac{z+1}{z-1}.$$

Then ψ maps $\mathbb{D}$ to $\mathbb{C}_+$ and $\mathbb{D}^*$ to $\mathbb{C}_-$ and $\mathbb{R}$ to $\partial\mathbb{D}\backslash\{1\}$.

We may form a *disk valence tree* by mapping $\mathbb{C}_+$ and $\mathbb{C}_-$ and $\mathbb{R}$ in the labeling of the valence tree to $\mathbb{D}$ and $\mathbb{D}^*$ and $\partial\mathbb{D}\backslash\{1\}$ under ψ^{-1}.

Example 3.7 Figure 3 is an example of a plane valence tree. By Theorem 3.8 (see below), there is a real Smirnov function with valence 3 on $\mathbb{C}_+$, valence 9 on $\mathbb{C}_-$, valence 3 on $(0, 1)$, valence 2 on $(-3, 0] \cup [1, 5)$, valence 1 on $(7, 8)$ and $(9, 10)$, and valence 0 elsewhere.

Theorem 3.8 *The valence of every real Smirnov function with finite valence is given by the valence of a plane valence tree, and any valence arising from a plane valence tree is the valence of a real Smirnov function.*

The proof of Theorem 3.8 needs the following valence result. It is an exercise in [18] (Example 3.1 of Section 3.3), but for the sake of completeness we give the proof.

Lemma 3.9 *Let f be an analytic function in $\mathbb{D}$ of valence at most m. Then $f \in H^p$ for every $p \in (0, \frac{1}{2m})$.*

Proof of Theorem 2.1 Recall the definition of the p-integral means $M_p(r, f)$ from (1.3) and define

$$M_\infty(r, f) := \sup_{|z|=r} |f(z)|.$$

By [4, Thm. 1] (see also [18, Sec. 2.3]) we have

$$M_\infty(r, f) = O\left(\frac{1}{(1-r)^{2m}}\right). \tag{3.2}$$

From [18, Theorem 3.2] we see that if f is m-valent and $0 < r_0 < r < 1$ then

$$M_p(r, f) \leqslant M_\infty(r_0, f)^p + m \max\left(m, \frac{m^2}{2}\right) \int_{r_0}^{r} \frac{M_\infty(t, f)^p}{t} \, dt.$$

Applying the estimate in (3.2) for $M_\infty(r, f)$ shows that the function

$$r \mapsto M_p(r, f)$$

is bounded when $p \in (0, \frac{1}{2m})$, i.e., $f \in H^p$ for all $p \in (0, \frac{1}{2m})$. $\qquad\square$

Proof of Theorem 3.8 Given a plane valence tree, form the associated (disk) valence tree under the mapping ψ^{-1}. Construct a disk tree X where nodes labeled $\mathbb{D}$ with valence m correspond to disks of valence m, and similarly for $\mathbb{D}^*$. If Y is a copy of $\mathbb{D}$ or $\mathbb{D}^*$ with valence m, consider the arcs labeling edges connected to the corresponding node. Call these arcs $I_1, \ldots, I_k$. The disjoint union of these arcs are a subset of a disjoint union of m copies of $\partial\mathbb{D}\backslash\{1\}$. This means that we can find a disjoint set of admissible arcs on the boundary of Y whose image arcs are precisely the arcs $I_1, \ldots, I_k$.

Given copies of $\mathbb{D}$ and $\mathbb{D}^*$ with corresponding nodes connected by an edge, weld them together together on arcs with image arcs equal to the arc labeling the edge between them. This is possible by the above remarks. Since the valence tree has the free arc property and is a tree, X will have a free arc on some copy of $\mathbb{D}$ or $\mathbb{D}^*$ contained in it and will be a disk tree.

Let φ be a conformal map from $\mathbb{D}$ onto X. The map $g = \psi \circ f_X \circ \varphi$ has valence equal to the valence of the plane valence tree. Since g has finite valence, it is in the Smirnov class by Lemma 3.9. Every point in $\mathbb{C}\backslash(\mathbb{R} \cup \{i, -i\})$ has a neighborhood U such that $g^{-1}(U)$ consists of disjoint sets that are each homeomorphic to U under f. The reason that we exclude i and $-i$ is that $i = \psi(0)$ and $-i = \psi(\infty)$, and the points 0 and ∞ may be the image under f_X of points where f_X has zero derivative. The fact that $\{i, -i\}$ has zero capacity together with the same argument used to prove (2.4) shows that g has real boundary values almost everywhere. We mention in passing that, in fact, since g has finite valence and thus is rational (by the result from [19]), it has real boundary values everywhere, except for a finite number of points where it has ∞ as a boundary value.

For the other direction, suppose that f is real Smirnov with finite valence. Then f is rational and thus continuous on $\mathbb{D}$, when viewed as a map into $\widehat{\mathbb{C}}$.

Consider $f^{-1}(-\infty, \infty)$. This is a set of branched analytic arcs in $\mathbb{D}$, with endpoints only at $\mathbb{D}$ and branch points only at points where f' is zero. Each branch point has an even number of analytic arcs coming from it. Also, $f^{-1}(\mathbb{C}_+)$ is a finite disjoint union of open sets we will call upper regions (similarly for $\mathbb{C}_-$ and lower regions). By the maximum principle for harmonic functions applied to the real part of f, no upper region can have more than two of the analytic arcs going into a branch point as boundary arcs. Also by the maximum principle, no two upper regions can share a boundary curve, and each upper region is simply connected. Given an upper region, note that its boundary curve is mapped into $\mathbb{R}$, and so by the argument principle each point in $\mathbb{C}_+$ has the same valence m under f restricted to the upper region. This means

Fig. 4 $I_1 = (-1, \infty)$, $I_2 = (0, 3)$, $I_3 = (-\infty, -2)$. Valence is 3 on $\mathbb{C}_+$, 2 on $\mathbb{C}_-$, 2 on $(0, 3)$, 1 on $(-1, 0]$ and $[3, \infty)$, and 1 on $(-\infty, -2)$. Note that as we proceed counterclockwise around the boundary of an upper region and clockwise around the boundary of a lower region, the image under f increases on the real line
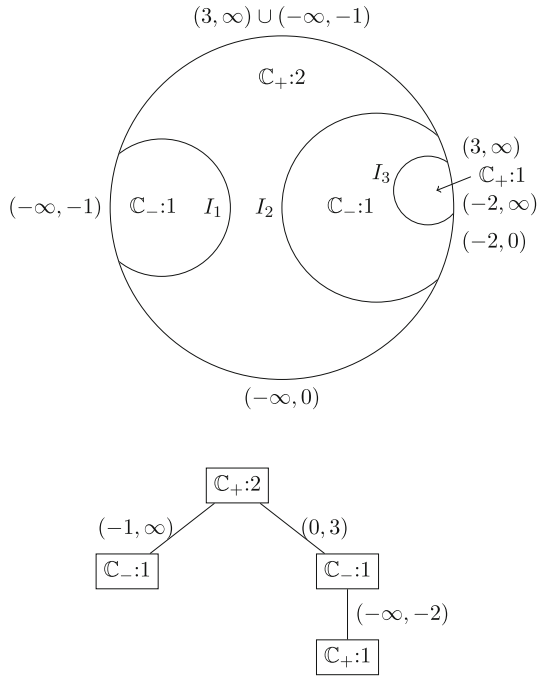

the boundary of the upper region must contain some point that maps to ∞. The same applies to lower regions. By the orientation preserving properties of analytic functions, each upper region maps its boundary (considered as positively oriented) to the real line in a way so that moving along the boundary increases the value on the real line (except at ∞). A lower region has the opposite property. Thus, by the argument principle, the boundary of an upper or lower region must map onto $\mathbb{R} \cup \{\infty\}$ exactly m times. We give Figs. 4 and 5 for illustration.

We will now give a method that, given a finite valence real Smirnov function, constructs a plane valence tree such that the valence of the tree corresponds to the valence of the function. We proceed by induction on the number of regions. If there is only one region, and this region has valence m, construct a plane valence tree with only one node of valence m.

Suppose there is more than one region. We may replace upper regions by lower regions in the following argument if needed. Take an upper region, and consider all upper regions sharing common boundary points *inside the unit disk* with the given region. Take the union, including the boundary points and boundary arcs. Repeat the process for all the new upper regions added, and continue until it is no longer possible to do so. This forms a finite union of upper regions (and their boundaries) - call it X, and call it an upper collection. By the maximum principle for harmonic functions, X is simply connected. X is not the whole disk by our assumption.

The boundary of X must intersect the boundary of the disk since some point in the boundary of each upper region maps to ∞. Thus, the complement of X consists of a union of simply connected sets. For the jth set, let φ_j be the conformal map from the

Fig. 5 The disk with upper and lower regions, and the corresponding valence tree. Upper regions are labeled $\mathbb{C}_+$ and lower regions are labeled $\mathbb{C}_-$. Edges and valences are not labeled. Note that, if we start with upper region A, we obtain the upper collection with A, B, C, and D as the first node
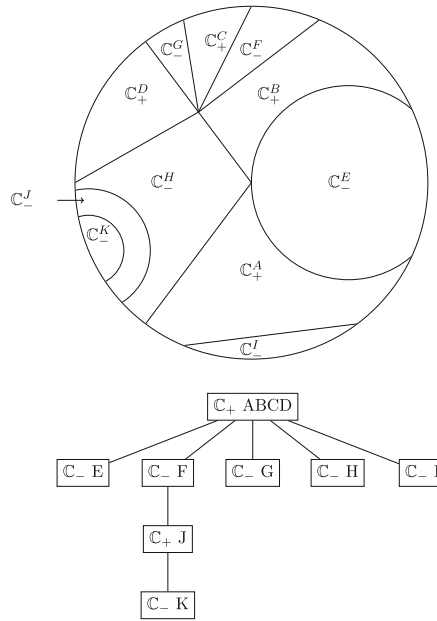
unit disk onto this set. By the induction hypothesis, we may form a valence tree for the jth set, using the function $f \circ \varphi_j$ instead of the function f, and by starting with a lower region instead of an upper region. (We could also apply the above reasoning directly to the jth set without using the conformal map φ_j.) Let T_j denote the valence tree for the jth set. Now suppose the total valence of all components of X is M. Draw a $\mathbb{C}_+$ node with valence M. Draw edges from the node for X to the nodes in the trees T_j that correspond to lower collections sharing boundary arcs with X. There is at least one edge to each T_j because every T_j shares a boundary arc with X. We will later see that there is exactly one edge to each T_j. Label each arc's edges with the intervals corresponding to the values of f on the edges.

We will show the graph formed is a tree. Consider an analytic arc in $f(-\infty, \infty)$ that approaches the boundary of the circle and is part of the boundary of the upper collection. If we start from a point in the arc that is on the unit circle and follow the arc, it either terminates at another point of the circle, or at a branch point. If this is the case, some other arc going from the branch point must be the boundary of the same lower region as the original arc; follow the new arc. We may continue until we hit the boundary of the unit circle, which we must since the arc can never approach the same branch point again, and there are finitely many branch points. Call the combination of arcs γ. The combination of arcs γ will be part of the boundary of some lower region, call it L. The lower region L can only have a common boundary with the upper collection X along γ, since the upper collection X is connected and γ intersects the circle at its two ends. The component of the complement of the upper collection X that contains L can have common boundary with X only on γ, for the same reason. This shows that the graph we form is a tree.

The arc γ (not counting points on the unit circle) maps to some subset of $(-\infty, \infty)$ once. This follows since as a point travels along γ, the image of the point always increases on the real line, or always decreases, since γ is part of the boundary of a lower region. This shows each edge is labeled with an interval in $\mathbb{R}$. Since the boundary of the upper collection maps onto $\mathbb{R} \cup \{\infty\}$ exactly M times, the disjoint union of all the intervals labeling edges coming from the upper collection is contained in a disjoint union of M copies of $\mathbb{R}$.

Note that there is some node (say of valence m) such that m disjoint copies of $\mathbb{R}$ minus the disjoint union of its edges contains an (open) interval. This follows from the fact that some upper or lower region must have a boundary arc in common with the unit disk. Thus, the graph we construct is a plane valence tree. The valence of the tree is the same as that of f by construction. □

Note that if a holomorphic function has bounded finite valence on $\mathbb{C}_+$ and $\mathbb{C}_-$, it must have finite valence on $\mathbb{R}$ by the open mapping theorem.

We now give some examples.

Example 3.10 Let $n \geqslant 1$ and

$$(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$$

be a finite set of open intervals such that none of the intervals is the entire real line and (a_j, b_j) is disjoint from (a_{j+1}, b_{j+1}) for each j. Then there is a function from $N_{\mathbb{R}}^+$ whose range is

$$\bigcup_{j=1}^{n} (a_j, b_j) \cup \mathbb{C}_+ \cup \mathbb{C}_-.$$

Moreover, the valence of each point of $\mathbb{C}_+$ is $\lfloor n/2 \rfloor + 1$ and the valence of each point of $\mathbb{C}_-$ is $\lceil n/2 \rceil$. The valence of each point in $\mathbb{R}$ is equal to the number of the intervals (a_j, b_j) in which it lies.

Clearly we can interchange the roles of $\mathbb{C}_+$ and $\mathbb{C}_-$ in the above example.

We could deduce this from our previous theorem, but we will first give an independent proof that is simpler than the proof of the previous theorem.

Proof of Theorem 3.8 Construct a Riemann surface as follows. Weld a copy of $\mathbb{C}_+$ to $\mathbb{C}_-$ along the interval (a_1, b_1). Now weld the copy of $\mathbb{C}_-$ to a different copy of $\mathbb{C}_+$ along the interval (a_2, b_2). Now weld this copy of $\mathbb{C}_+$ to a different copy of $\mathbb{C}_-$ along the interval (a_3, b_3). Proceed in this manner until all of the intervals are exhausted. Call this Riemann surface X. Let θ be the projection map from X to $\mathbb{C}$ that takes a given point in the Riemann surface to the corresponding point in either $\mathbb{C}_+$, $\mathbb{C}_-$, or $\mathbb{R}$.

We now claim that the Riemann surface is conformally equivalent to $\mathbb{D}$. If it were not, then, since it is simply connected, it would be equivalent to either the Riemann sphere or the complex plane (uniformization theorem). It is not equivalent to the sphere since it is not compact.

It is not equivalent to the complex plane since we could weld another half plane onto the last half plane welded onto X and still have a simply connected surface, and if we remove infinitely many points from any simply connected Riemann surface and are left with a simply connected surface, the new surface must be equivalent to the disk by the uniformization theorem and the Riemann mapping theorem.
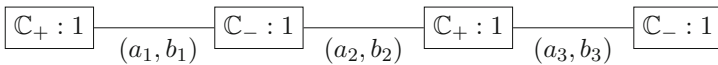
Let φ be the conformal map from $\mathbb{D}$ to the Riemann surface and

$$f = \theta \circ \varphi.$$

Then f maps from $\mathbb{D}$ into $\mathbb{C}$. Since f has valence at most $\lfloor n/2 \rfloor + 1$ at each point, it belongs to some Hardy space (Lemma 3.9) and thus belongs to N^+.

We will now show that f has real (radial) boundary values almost everywhere. To see this, note that every point in $\mathbb{C} \backslash \mathbb{R}$ has a neighborhood U such that $\theta^{-1}(U)$ consists of disjoint sets that are each homeomorphic to U under θ. Thus every point in $\mathbb{C} \backslash \mathbb{R}$ has a neighborhood U such that $f^{-1}(U)$ consists of disjoint sets that are each homeomorphic to U under f. The same argument as used to prove (2.4) shows that no (radial) boundary values of f lie in $\mathbb{C} \backslash \mathbb{R}$. □

For example, for $n = 3$, the valence tree for this example is as shown below. The nodes are shown with their valences.

$$\boxed{\mathbb{C}_+ : 1} \underset{(a_1, b_1)}{\rule{3em}{0.4pt}} \boxed{\mathbb{C}_- : 1} \underset{(a_2, b_2)}{\rule{3em}{0.4pt}} \boxed{\mathbb{C}_+ : 1} \underset{(a_3, b_3)}{\rule{3em}{0.4pt}} \boxed{\mathbb{C}_- : 1}$$

Example 3.11 The only possible valence trees for a real Smirnov function with valence 1 on $\mathbb{C}_+$ and valence 1 on $\mathbb{C}_-$ have a $\mathbb{C}_+$ node of valence 1 connected to a $\mathbb{C}_-$ node of valence 1. The interval labeling the edge connecting them can be any nonempty open interval, except $\mathbb{R}$, since if the edge was labeled $\mathbb{R}$ neither node would have a free interval. Thus the valence of such a real Smirnov function is 1 on a nonempty open proper subinterval of $\mathbb{R}$, and 0 elsewhere on $\mathbb{R}$.

Example 3.12 Let us find all possible valences for real Smirnov functions with valence 2 on $\mathbb{C}_+$ and valence 1 on $\mathbb{C}_-$. The valence tree for such a function has either one $\mathbb{C}_+$ node of valence 2, or two $\mathbb{C}_+$ nodes of valence 1. These nodes cannot be adjacent. Figure 6 shows all possible valence trees. In tree I, I_1 and I_2 must be disjoint; both the top and bottom node automatically have free intervals. In tree II, I_1 can be an arbitrary non-empty open interval. Because the $\mathbb{C}_+$ node has valence 2 and only one edge, it has a free arc automatically. So the valence at every point in $\mathbb{R}$ is either 1 or 0. The range on $\mathbb{R}$ is either the union of two open intervals, or is one interval.

Example 3.13 Let us find all possible valences for real Smirnov functions with valence 2 on $\mathbb{C}_+$ and valence 3 on $\mathbb{C}_-$. The valence tree for such a function has either one $\mathbb{C}_+$ node of valence 2, or two $\mathbb{C}_+$ nodes of valence 1. These nodes cannot be adjacent. Figure 7 shows all possible valence trees. Note that if all the other conditions for being a real valence tree are satisfied, some node in a tree must have a free interval if there is some node of valence one connected to at least two nodes, one of which connects

Fig. 6 All possible valence tress with valence 2 on $\mathbb{C}_+$ and 1 on $\mathbb{C}_-$
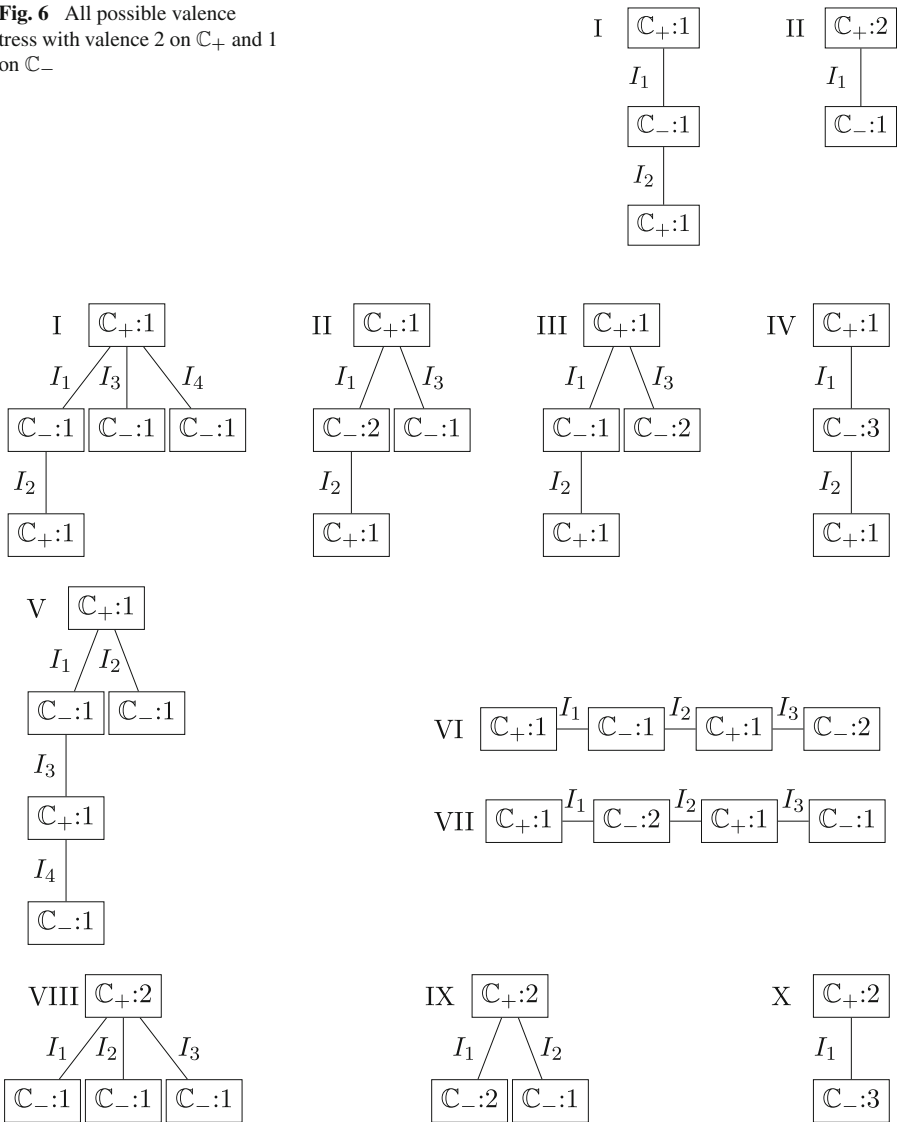
I $\boxed{\mathbb{C}_+\text{:}1}$
I_1
$\boxed{\mathbb{C}_-\text{:}1}$
I_2
$\boxed{\mathbb{C}_+\text{:}1}$

II $\boxed{\mathbb{C}_+\text{:}2}$
I_1
$\boxed{\mathbb{C}_-\text{:}1}$

I $\boxed{\mathbb{C}_+\text{:}1}$
$I_1 \quad I_3 \quad I_4$
$\boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}1}$
I_2
$\boxed{\mathbb{C}_+\text{:}1}$

II $\boxed{\mathbb{C}_+\text{:}1}$
$I_1 \quad I_3$
$\boxed{\mathbb{C}_-\text{:}2} \ \boxed{\mathbb{C}_-\text{:}1}$
I_2
$\boxed{\mathbb{C}_+\text{:}1}$

III $\boxed{\mathbb{C}_+\text{:}1}$
$I_1 \quad I_3$
$\boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}2}$
I_2
$\boxed{\mathbb{C}_+\text{:}1}$

IV $\boxed{\mathbb{C}_+\text{:}1}$
I_1
$\boxed{\mathbb{C}_-\text{:}3}$
I_2
$\boxed{\mathbb{C}_+\text{:}1}$

V $\boxed{\mathbb{C}_+\text{:}1}$
$I_1 \quad I_2$
$\boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}1}$
I_3
$\boxed{\mathbb{C}_+\text{:}1}$
I_4
$\boxed{\mathbb{C}_-\text{:}1}$

VI $\boxed{\mathbb{C}_+\text{:}1} \overset{I_1}{-} \boxed{\mathbb{C}_-\text{:}1} \overset{I_2}{-} \boxed{\mathbb{C}_+\text{:}1} \overset{I_3}{-} \boxed{\mathbb{C}_-\text{:}2}$

VII $\boxed{\mathbb{C}_+\text{:}1} \overset{I_1}{-} \boxed{\mathbb{C}_-\text{:}2} \overset{I_2}{-} \boxed{\mathbb{C}_+\text{:}1} \overset{I_3}{-} \boxed{\mathbb{C}_-\text{:}1}$

VIII $\boxed{\mathbb{C}_+\text{:}2}$
$I_1 \quad I_2 \quad I_3$
$\boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}1} \ \boxed{\mathbb{C}_-\text{:}1}$

IX $\boxed{\mathbb{C}_+\text{:}2}$
$I_1 \quad I_2$
$\boxed{\mathbb{C}_-\text{:}2} \ \boxed{\mathbb{C}_-\text{:}1}$

X $\boxed{\mathbb{C}_+\text{:}2}$
I_1
$\boxed{\mathbb{C}_-\text{:}3}$

Fig. 7 All possible valence tress with valence 2 on $\mathbb{C}_+$ and 3 on $\mathbb{C}_-$

to no other nodes. If a node of valence m has less than m edges connected to it, the free interval condition is also automatically satisfied. These remarks apply to all the trees in Fig. 7 except for tree VIII, which also automatically has a free interval since two of the intervals labeling its edges must be disjoint.

In tree I, we require I_1, I_3, and I_4 to be pairwise disjoint, and I_1 and I_2 to be disjoint. In tree II, we require I_1 and I_3 to be disjoint. The intervals I_1 and I_2 do not have to be disjoint, since the node they have in common has valence 2, and any two intervals

can fit disjointly into two copies of $\mathbb{R}$. In tree III, we require I_1 and I_3 to be disjoint, and I_1 and I_2 to be disjoint. In tree IV, the intervals I_1 and I_2 can be arbitrary.

In tree V, intervals I_1 and I_2 must be disjoint, intervals I_1 and I_3 are disjoint, and intervals I_3 and I_4 are disjoint. In tree VI, I_1 must be disjoint from I_2 and I_2 must be disjoint from I_3. In tree VII, I_2 must be disjoint from I_3.

In tree VIII, the intervals I_1, I_2, and I_3 must be able to fit into two disjoint copies of $\mathbb{R}$ without intersecting. This is equivalent to requiring that two of them be disjoint. In trees IX and X, there are no requirements on the intervals. Note that for all cases, the conditions above imply that some node must have a free interval, and the above intervals are allowed to be $\mathbb{R}$ if this does not conflict with any of the above conditions.

Considering all possible cases shows that the range of the function on $\mathbb{R}$ may be (counting multiplicity) the union of four open intervals I_1, I_2, I_3 and I_4 where I_1 and I_2 are disjoint, I_1 and I_3 are disjoint, and I_3 and I_4 are disjoint. The range of the function on $\mathbb{R}$ (counting multiplicity) may also be the union of three open intervals, at least two of which are disjoint, or it may be the union of two open intervals; or it will be one open interval. These are the only cases. Some of these intervals may be $\mathbb{R}$ if the conditions are satisfied (although they cannot be empty). To take a concrete example, from the first case we see that the range could be (counting multiplicities)

$$(0, 1) \cup (2, \infty) \cup (2, 3) \cup (4, 5).$$

In other words, the valence would be one on $(0, 1)$, two on $(2, 3)$ and $(4, 5)$, one on $[3, 4] \cup [5, \infty)$, and zero on the rest of $\mathbb{R}$.

References

1. Ahlfors, L.V., Sario, L.: Riemann Surfaces. Princeton Mathematical Series, No. 26. Princeton University Press, Princeton (1960)
2. Akhiezer, N.I., Glazman, I.M.: Theory of Linear Operators in Hilbert Space, vol. II. Translated from the Russian by Merlynd Nestell. Frederick Ungar Publishing Co., New York (1963)
3. Aleman, A., Martin, R.T.W., Ross, W.T.: On a theorem of livsic. J. Funct. Anal. **264**(4), 999–1048 (2013)
4. Cartwright, M.L.: Some inequalities in the theory of functions. Math. Ann. **111**(1), 98–118 (1935)
5. Cima, J.A., Ross, W.T.: The Backward Shift on the Hardy Space. Mathematical Surveys and Monographs, vol. 79. American Mathematical Society, Providence (2000)
6. Collingwood, E.F., Lohwater, A.J.: The Theory of Cluster Sets. Cambridge University Press, Cambridge (1966)
7. Colwell, P.: Blaschke Products. University of Michigan Press, Ann Arbor (1985) (Bounded analytic functions)
8. Conway, J,B.: Functions of One Complex Variable. II, volume 159 of Graduate Texts in Mathematics. Springer, New York (1995)
9. Douglas, R.G., Shapiro, H.S., Shields, A.L.: Cyclic vectors and invariant subspaces for the backward shift operator. Ann. Inst. Fourier (Grenoble) **20**(fasc.1), 37–76 (1970)
10. Duren, P.L.: Theory of H^p Spaces. Academic Press, New York (1970)
11. Fisher, S.D.: Function theory on planar domains. In: Pure and Applied Mathematics (New York). Wiley, New York (1983). A second course in complex analysis, A Wiley-Interscience Publication
12. Garcia, S., Mashreghi, J., Ross, W.: Finite Blaschke Products and Their Connections. Monographs in Mathematics. Springer, New York (2019)
13. Garcia, S.R., Mashreghi, J., Ross, W.T.: Introduction to Model Spaces and Their Operators, volume 148 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2016)

14. Garcia, S.R., Mashreghi, J., Ross, W.T.: Real complex functions. In: Recent Progress on Operator Theory and Approximation in Spaces of Analytic Functions, volume 679 of Contemp. Math., pp. 91–128. Am. Math. Soc., Providence, RI (2016)
15. Garcia, S.R., Sarason, D.: Real outer functions. Indiana Univ. Math. J. **52**(6), 1397–1412 (2003)
16. Garnett, J.B.: Bounded Analytic Functions, volume 236 of Graduate Texts in Mathematics (first edition). Springer, New York (2007)
17. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2002)
18. Hayman, W.K.: Multivalent Functions, volume 110 of Cambridge Tracts in Mathematics (second edition). Cambridge University Press, Cambridge (1994)
19. Helson, H.: Large analytic functions. In: Linear Operators in Function Spaces (Timişoara, 1988), volume 43 of *Oper. Theory Adv. Appl.*, pp. 209–216. Birkhäuser, Basel (1990)
20. Helson, H.: Large analytic functions. II. In: Analysis and Partial Differential Equations, volume 122 of Lecture Notes in Pure and Appl. Math., pp. 217–220. Dekker, New York (1990)
21. Mashreghi, J: Derivatives of Inner Functions, volume 31 of Fields Institute Monographs. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON (2013)
22. Poltoratski, A.G.: Properties of exposed points in the unit ball of H^1. Indiana Univ. Math. J. **50**(4), 1789–1806 (2001)
23. Ransford, T.: Potential Theory in the Complex Plane. London Mathematical Society Student Texts, vol. 28. Cambridge University Press, Cambridge (1995)
24. Ross, W.T., Shapiro, H.S.: Generalized Analytic Continuation. University Lecture Series, vol. 25. American Mathematical Society, Providence (2002)
25. Rudin, Walter: A generalization of a theorem of Frostman. Math. Scand. **21**(136–143), 1967 (1968)
26. Sarason, D.: Unbounded Toeplitz operators. Integral Equ. Oper. Theory **61**(2), 281–298 (2008)
27. Stephenson, K.: The finite range of an inner function. Bull. Lond. Math. Soc. **11**(3), 300–303 (1979)