The parabola theorem on continued fractions

Journal Item

CMFT-MS 15094

The parabola theorem on continued fractions

Ian Short

Received: 24 November 2015 / Accepted: to be determined

Abstract Using geometric methods borrowed from the theory of Kleinian groups, we interpret the parabola theorem on continued fractions in terms of sequences of Möbius transformations. This geometric approach allows us to relate the Stern–Stolz series, which features in the parabola theorem, to the dynamics of certain sequences of Möbius transformations acting on three-dimensional hyperbolic space. We also obtain a version of the parabola theorem in several dimensions.

Keywords Conical limit points · continued fractions · hyperbolic geometry · Möbius transformations · parabola theorem · Stern–Stolz series

Mathematics Subject Classification (2000) Primary: 40A15; Secondary: 30B70, 30F45.

1 Introduction

The object of this paper is to discuss the parabola theorem on continued fractions using the geometry of Möbius transformations. The parabola theorem is about *infinite complex continued fractions* (henceforth described more briefly as *continued fractions*) which are quantities

$$\mathbf{K}(a_n|\,b_n) = \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cdots}}},$$

where the coefficients a_i and b_j are complex numbers, and none of the coefficients a_i are 0. Let $t_n(z) = a_n/(b_n+z)$ and $T_n = t_1 \circ \cdots \circ t_n$ (in future we just write $T_n = t_1 \cdots t_n$). The continued fraction $\mathbf{K}(a_n|\,b_n)$ is said to *converge* if the sequence $T_n(0)$ converges within the

Ian Short
Department of Mathematics and Statistics
The Open University
Milton Keynes MK7 6AA
United Kingdom
E-mail: ian.short@open.ac.uk

extended complex plane. Excepting Section 4, we will assume that $b_j = 1$ for $j = 1, 2, \ldots$, in which case $t_n(z) = a_n/(1+z)$ and the continued fraction will be denoted by $\mathbf{K}(a_n|\,1)$.

The first version of the parabola theorem was proven by Scott and Wall in [19]. Shortly after, more general versions were found by Leighton and Thron [13] and Paydon and Wall [17]. The theorem was extended by some of these authors in a number of subsequent papers including [14,21], and the statement of the theorem from [21] is recast in the books by Jones and Thron [11, Thm. 4.42] and Lorentzen and Waadeland [15, Thm. 3.43]. It is this version of the parabola theorem that we work with. Let P_α be the region given by

$$P_\alpha = \left\{ z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant \tfrac{1}{2}\cos^2\alpha \right\},$$

where $-\pi/2 < \alpha < \pi/2$, shown in Fig. 1.1, which is bounded by a parabola.

The parabola theorem *Suppose that $a_n \in P_\alpha$ and $a_n \neq 0$ for $n = 1, 2, \ldots$. Then the odd and even sequences $T_{2n-1}(0)$ and $T_{2n}(0)$ of the continued fraction $\mathbf{K}(a_n|\,1)$ both converge. They converge to the same limit, and hence $\mathbf{K}(a_n|\,1)$ converges, if and only if the series*

$$\left|\frac{1}{a_1}\right| + \left|\frac{a_1}{a_2}\right| + \left|\frac{a_2}{a_1a_3}\right| + \left|\frac{a_1a_3}{a_2a_4}\right| + \left|\frac{a_2a_4}{a_1a_3a_5}\right| + \cdots$$

diverges.

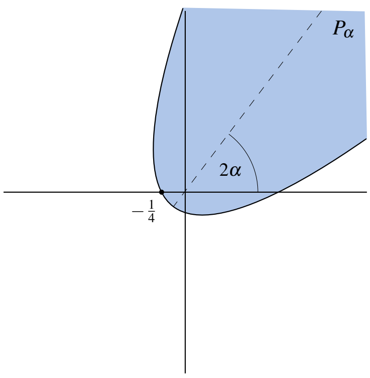

Fig. 1.1 The parabolic region $P_\alpha = \left\{ z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant \frac{1}{2}\cos^2\alpha \right\}$

Inspired by the philosophy set out by Beardon in [3], our aim is to understand in detail what the conditions of the parabola theorem say about the sequence of Möbius transformations T_n by using properties of Möbius transformations that are often employed in the theory of Kleinian groups. In particular, we wish to determine the behaviour of the sequence T_n not only at distinguished points such as 0 or ∞, but on the whole of the extended complex plane. To do this, we use the topological group structure of the group of Möbius transformations and the isometric action of this group on three-dimensional hyperbolic space. This approach gives insight into the parabola theorem and allows us to make precise statements about convergence. Another strength of the geometric approach is that it generalises easily to several dimensions, although for simplicity we state our results in the complex plane. Only at the very end of the paper do we discuss the parabola theorem in higher dimensions.

Our strategy is to split the parabola theorem in two, and deal with the condition $a_n \in P_\alpha$, and convergence of the series

$$\left|\frac{1}{a_1}\right| + \left|\frac{a_1}{a_2}\right| + \left|\frac{a_2}{a_1a_3}\right| + \left|\frac{a_1a_3}{a_2a_4}\right| + \left|\frac{a_2a_4}{a_1a_3a_5}\right| + \cdots, \tag{1.1}$$

separately. To understand the significance of the condition $a_n \in P_\alpha$, we use a recent version of the Hillam–Thron theorem, a theorem that was originally proven in [7]. The type of argument we use for this part of the parabola theorem is well known to continued fraction theorists, so we describe it only briefly, in Section 3.

The main result of this paper (Theorem 1.1, below) is about the significance of the series (1.1), which is known as the *Stern–Stolz series*. It converges only if the sequence $a_1, a_2, \dots$ grows sufficiently quickly. When a_n is large the map $t_n(z) = a_n/(1+z)$ is close to the map $s_n(z) = a_n/z$, in a sense that will later be made precise. Thus it will be shown that if the Stern–Stolz series converges, then we can understand the behaviour of the maps T_n using the simpler maps $S_n = s_1 \cdots s_n$.

Theorem 1.1 (to follow) gives a host of equivalent conditions involving the maps S_n and T_n. Let us summarise the terminology used in that theorem. We denote the upper half-space model of three-dimensional hyperbolic space by $\mathbb{H}^3$ (this is the upper half of $\mathbb{R}^3$) and we denote the hyperbolic metric on $\mathbb{H}^3$ by ρ. We identify the ideal boundary of $\mathbb{H}^3$ with the extended complex plane $\mathbb{C}_\infty$ in the usual way. The closure $\overline{\mathbb{H}^3}$ of $\mathbb{H}^3$ in $\mathbb{R}^3 \cup \{\infty\}$ consists of $\mathbb{H}^3$ together with its ideal boundary $\mathbb{C}_\infty$. The chordal metric χ is a complete metric on $\overline{\mathbb{H}^3}$; it is the metric inherited from the Euclidean metric on the unit ball by stereographic projection. The group of Möbius transformations acts on $\mathbb{C}_\infty$, and it can also act on $\mathbb{H}^3$. In fact, it is the full group of conformal isometries of $\mathbb{H}^3$. We can measure the distance between two Möbius transformations using the supremum metric χ_0, which is given by

$$\chi_0(f,g) = \sup_{z\in\mathbb{C}_\infty} \chi(f(z),g(z)).$$

Let $j = (0,0,1)$. A point p in $\mathbb{C}_\infty$ is said to be a *backward limit point* of a sequence $F_1, F_2, \dots$ of Möbius transformations if there is a subsequence of $F_1^{-1}(j), F_2^{-1}(j), \dots$ that converges to p in the chordal metric. A backward limit point p is a *conical limit point* of $F_1, F_2, \dots$ if there is a geodesic γ in $\mathbb{H}^3$ with one end-point at p and a subsequence of $F_1^{-1}(j), F_2^{-1}(j), \dots$ that lies within a bounded hyperbolic distance of γ and converges to p in the chordal metric. The sequence $F_1, F_2, \dots$ is said to be a *rapid escape sequence* if $\sum_n \exp[-\rho(j, F_n(j))]$ converges. These are familiar concepts from Kleinian group theory. They will be explained in more detail later on.

Theorem 1.1 *The following are equivalent:*

(i) $\left|\frac{1}{a_1}\right| + \left|\frac{a_1}{a_2}\right| + \left|\frac{a_2}{a_1a_3}\right| + \left|\frac{a_1a_3}{a_2a_4}\right| + \left|\frac{a_2a_4}{a_1a_3a_5}\right| + \cdots$ *converges;*

(ii) $\sum_n \chi_0(T_nS_n^{-1}, T_{n+1}S_{n+1}^{-1})$ *converges;*

(iii) $\sum_n \chi(T_n(0), T_{n+2}(0))$ *converges, and the sequences* $T_{2n-1}(0)$ *and* $T_{2n}(0)$ *converge to two distinct values;*

(iv) $\sum_n \rho\left(T_nS_n^{-1}(j), T_{n+1}S_{n+1}^{-1}(j)\right)$ *converges;*

(v) S_n *is a rapid escape sequence, and* ∞ *is its only backward limit point;*

(vi) *T_n is a rapid escape sequence, and ∞ is its only backward limit point and its only conical limit point.*

Further, if (i)–(vi) *hold, then there are distinct points p and q such that $T_{2n-1} \to p$ and $T_{2n} \to q$ locally uniformly on $\mathbb{C}$, and $T_{2n-1}(\infty) \to q$ and $T_{2n}(\infty) \to p$.*

The equivalence of (i) and (iii) has been proven already by Lane and Wall [12] in a direct algebraic fashion. Our shorter proof uses geometric properties of Möbius transformations.

Theorem 1.1 shows that if the Stern–Stolz series converges, then the sequences S_n and T_n have extremely strong convergence properties, even though the sequence $T_n(0)$ itself does not converge.

As a consequence of Theorem 1.1, we obtain the following version of the parabola theorem.

Theorem 1.2 *Let H be a half-plane that contains 0, but does not contain -1, even in its closure. Suppose that $t_n(H) \subseteq H$ for $n = 1,2,\dots$. If the Stern–Stolz series*

$$\left|\frac{1}{a_1}\right| + \left|\frac{a_1}{a_2}\right| + \left|\frac{a_2}{a_1a_3}\right| + \left|\frac{a_1a_3}{a_2a_4}\right| + \left|\frac{a_2a_4}{a_1a_3a_5}\right| + \cdots$$

converges, then there are distinct points p and q such that $T_{2n-1} \to p$ and $T_{2n} \to q$ locally uniformly on $\mathbb{C}$, and $T_{2n-1}(\infty) \to q$ and $T_{2n}(\infty) \to p$. If the Stern–Stolz series diverges then T_n converges locally uniformly on H, and almost everywhere on $\mathbb{C}_\infty$, to a single point.

It is well known (and will be shown in the next section) that the condition $t_n(H) \subseteq H$ in this theorem is equivalent to the condition that the coefficients a_n lie within a parabolic region such as that shown in Fig. 1.1. The more substantial difference between Theorem 1.2 and other statements of the parabola theorem, including the one given earlier, is that Theorem 1.2 contains detailed information about the behaviour of the sequence T_n. Some of the new features of this theorem could be extracted from existing accounts of the parabola theorem; however, other accounts tend to focus on the continued fraction $\mathbf{K}(a_n|\,1)$ rather than describing in detail the dynamics of the sequence T_n.

In Section 6 we provide an example to show that when the Stern–Stolz series diverges, the sequence T_n may diverge on an uncountable, dense subset of $\mathbb{C}_\infty \setminus H$.

2 A geometric explanation of the parabolic region

Let H_α be the half-plane given by

$$H_\alpha = \left\{-\tfrac{1}{2} + e^{i\alpha}z : \mathrm{Re}[z] > 0\right\},$$

and recall that P_α denotes the parabolic region

$$\left\{z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant \tfrac{1}{2}\cos^2\alpha\right\}.$$

The following theorem, illustrated by Fig. 2.1, is well known (see, for example, [15, Thm. 3.43]).

Theorem 2.1 *The Möbius transformation $t_n(z) = a_n/(1+z)$ satisfies $t_n(H_\alpha) \subseteq H_\alpha$ if and only if $a_n \in P_\alpha$.*

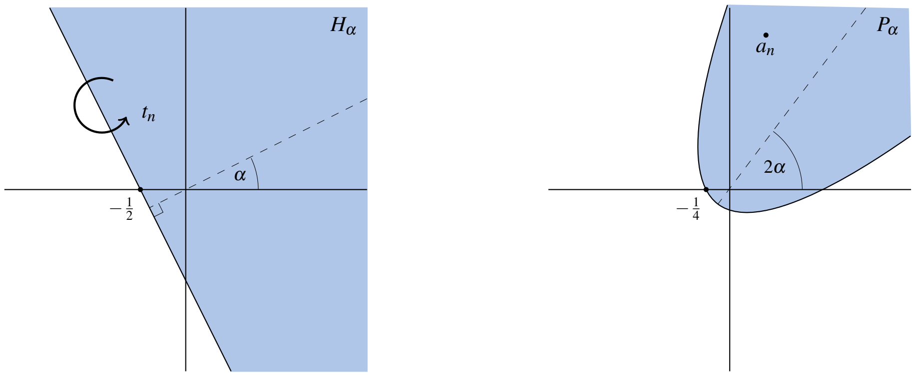

Fig. 2.1 $t_n(H_\alpha) \subseteq H_\alpha$ if and only if $a_n \in P_\alpha$

Theorem 2.1 allows us to apply the Hillam–Thron theorem to help prove the parabola theorem in the usual manner, as we shall see in the next section. Here we give a new geometric proof of Theorem 2.1, which illuminates some of the features of the parabola theorem needed later, and, using only a few basic properties of Möbius transformations, it gets us off to an undemanding start.

Let H be an open Euclidean half-plane that contains 0, but does not contain -1, even in its closure. The boundary line ∂H must cut the real axis somewhere between -1 and 0. If it cuts at $-1/2$ then H is one of the half-planes H_α, but there is no need for us to assume that. Let $t(z) = a/(1+z)$, where $a \neq 0$. Denote by u the inverse point of -1 in the boundary line ∂H. This is the image of -1 under reflection in ∂H. Inverse points of lines and circles are preserved by Möbius transformations, which implies that $t(u)$ is the inverse point of $t(-1)$ in $t(\partial H)$. Now, we know that $t(-1) = \infty$, which implies that $t(\partial H)$ is a Euclidean circle with centre $t(u)$. We also know that $t(\partial H)$ contains the point $t(\infty) = 0$.

The half-plane H and its image disc $t(H)$ are shown in Fig. 2.2.

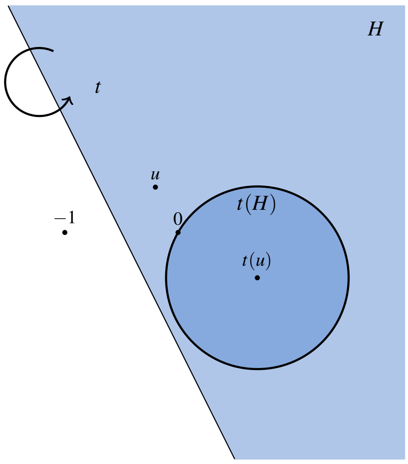

Fig. 2.2 The disc $t(H)$ is contained within the half-plane H

The radius of $t(H)$ is $|t(u)-0| = |t(u)|$. Since the centre of $t(H)$ is $t(u)$ it follows that $t(H) \subseteq H$ if and only if $t(u)$ lies in the set

$$\{z \in \mathbb{C} : |z| \leqslant d(z, \partial H)\},$$

where d denotes the Euclidean metric. This is a region bounded by a parabola with focus 0 and directrix ∂H. Since $t(u) = a/(1+u)$, we have proved the following theorem.

Theorem 2.2 *The Möbius map $t(z) = a/(1+z)$ satisfies $t(H) \subseteq H$ if and only if the coefficient a lies in the parabolic region*

$$P = (1+u)\{z \in \mathbb{C} : |z| \leqslant d(z, \partial H)\}.$$

The parabolic region P has focus 0 and directrix $(1+u)\partial H$.

Let us now show that this theorem implies Theorem 2.1. We can write H in the form $s + e^{i\alpha}\mathbb{K}$, where $\mathbb{K}$ is the right half-plane, and $-1 < s < 0$ and $-\pi/2 < \alpha < \pi/2$. Let τ denote the reflection in ∂H, which is given by

$$\tau(z) = s + e^{2i\alpha}(s - \bar{z}).$$

Then $u = \tau(-1)$, and so

$$1 + u = 1 + \tau(-1) = 2(s+1)e^{i\alpha}\cos\alpha.$$

Also, for $z \in H$,

$$d(z, \partial H) = \tfrac{1}{2}|z - \tau(z)| = \mathrm{Re}[(z-s)e^{-i\alpha}] = \mathrm{Re}[ze^{-i\alpha}] - s\cos\alpha.$$

Therefore

$$\begin{aligned} P &= 2(s+1)e^{i\alpha}\cos\alpha\left\{z \in \mathbb{C} : |z| \leqslant \mathrm{Re}[ze^{-i\alpha}] - s\cos\alpha\right\} \\ &= \left\{z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant -2s(s+1)\cos^2\alpha\right\}. \end{aligned}$$

When $s = -1/2$, so that H is the half-plane $H_\alpha = -\frac{1}{2} + e^{i\alpha}\mathbb{K}$, we find that

$$P = \left\{z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant \tfrac{1}{2}\cos^2\alpha\right\},$$

which is the parabolic region P_α. Therefore Theorem 2.2 does indeed imply Theorem 2.1.

Notice that the expression $-2s(s+1)$ takes its maximum value, namely $1/2$, when $s = -1/2$. This shows that, for a given angle α, P_α contains the parabolic region

$$\left\{z \in \mathbb{C} : |z| - \mathrm{Re}[ze^{-2i\alpha}] \leqslant -2s(s+1)\cos^2\alpha\right\},$$

no matter the value of s. This explains why the original statement of the parabola theorem given near the beginning of the introduction cannot be improved by allowing s to take values other than $-1/2$.

Let us finish this section with some remarks on related geometric constructions that arise in the theory of continued fractions. We have seen that a Möbius transformation $t(z) = a/(1+z)$ satisfies $t(H) \subseteq H$ if and only if a belongs to a parabolic region P. Suppose now that H is a Euclidean disc instead of a half-plane. It is only possible for t to map H within itself if $-1 \notin \overline{H}$. Given this condition, it can be shown that $t(H) \subseteq H$ if and only if a belongs to a region bounded by a Cartesian oval, and there is a corresponding theorem in continued fraction theory called the oval theorem, which was first proven by Lorentzen (formerly known as Jacobsen) and Thron in [9].

Another possibility is that H is the complement of a Euclidean disc, which contains 0 in its interior. In this case, $t(H) \subseteq H$ if and only if a belongs to another region bounded by a Cartesian oval. This case has received little if any attention.

3 The Hillam–Thron theorem

In this section we explore the ramifications of the condition $a_n \in P_\alpha$ from the parabola theorem further. Theorem 2.1 tells us that $a_n \in P_\alpha$ if and only if $t_n(z) = a_n/(1+z)$ satisfies $t_n(H_\alpha) \subseteq H_\alpha$. We also know that t_n maps -1 to ∞ and ∞ to 0. The point -1 lies outside the closure $\overline{H_\alpha}$ of H_α, ∞ lies on the boundary ∂H_α of H_α, and 0 lies inside H_α.

More generally, consider a sequence f_n of Möbius transformations that satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$ for $n = 1, 2, \ldots$, where D is an open disc in $\mathbb{C}_\infty$, and a, b, and c are three points with $a \in D$, $b \in \partial D$, and $c \notin \overline{D}$. These properties are illustrated in Fig. 3.1. The sequence t_n is of this type, as we can see by choosing $a = 0$, $b = \infty$, $c = -1$, and $D = H_\alpha$.

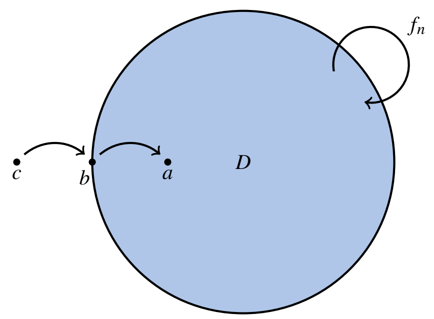

Fig. 3.1 The map f_n satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$

The next theorem shows that, given such a collection of maps f_n, the sequence $F_n = f_1 \cdots f_n$ has strong convergence properties. This theorem uses the chordal metric χ on $\mathbb{C}_\infty$, which is the metric inherited from the Euclidean metric on the unit sphere by stereographic projection. We could instead use the spherical metric on the unit sphere, which is additive along geodesics, but the chordal metric is simpler algebraically. It is given by

$$\chi(z,w) = \frac{2|z-w|}{\sqrt{1+|z|^2}\sqrt{1+|w|^2}}\,, \quad \chi(z,\infty) = \frac{2|z|}{\sqrt{1+|z|^2}}\,,$$

where $z, w \neq \infty$.

Theorem 3.1 *Suppose that D is an open disc and a, b, and c are three points with $a \in D$, $b \in \partial D$, and $c \notin \overline{D}$. Suppose also that $f_1, f_2, \ldots$ is a sequence of Möbius transformations that satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$ for $n = 1, 2, \ldots$. Let $F_n = f_1 \cdots f_n$. Then $\sum_n \chi(F_n(a), F_{n+2}(a))$ converges. Furthermore, F_{2n-1} converges locally uniformly on D, and almost everywhere on $\mathbb{C}_\infty$, to a point p, and F_{2n} converges locally uniformly on D, and almost everywhere on $\mathbb{C}_\infty$, to a point q.*

The first assertion of the parabola theorem, that the odd and even sequences $T_{2n-1}(0)$ and $T_{2n}(0)$ both converge, follows immediately from Theorem 3.1.

Theorem 3.1 is a corollary of the following version of the Hillam–Thron theorem, which is [20, Thm. 4.1] paraphrased. This procedure of deducing a result such as Theorem 3.1 from a version of the Hillam–Thron theorem is standard within the continued fractions literature; the only original part of Theorem 3.1, which we need later, is the statement about almost everywhere convergence.

Theorem 3.2 *Suppose that D is an open disc, u is a point in D, v is a point that lies outside $\overline{D}$, and $g_1, g_2, \dots$ is a sequence of Möbius transformations with $g_n(D) \subseteq D$ and $g_n(v) = u$ for $n = 1, 2, \dots$. Then the sequence $G_n = g_1 \cdots g_n$ satisfies*

(i) $\sum_n \chi(G_n(u), G_{n+1}(u))$ *converges;*
(ii) G_n *converges locally uniformly on D to a point p;*
(iii) G_n *converges everywhere but on a set of Hausdorff dimension at most 1 to p.*

We can now prove Theorem 3.1.

Proof (of Theorem 3.1) Recall that $f_1, f_2, \dots$ is a sequence of Möbius transformations that satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$ for $n = 1, 2, \dots$. Let $g_n = f_{2n-1} f_{2n}$. Then $g_n(D) \subseteq f_{2n-1}(D) \subseteq D$ and $g_n(c) = f_{2n-1}(b) = a$. It follows from Theorem 3.2 that the series $\sum_n \chi(F_{2n}(a), F_{2n+2}(a))$ converges and F_{2n} converges locally uniformly on D, and everywhere on $\mathbb{C}_\infty$ but on a set of Hausdorff dimension at most one, to a point p. A similar argument can be applied to F_{2n-1}. The result follows, because sets of Hausdorff dimension one have Lebesgue measure 0. □

As the proof indicates, each assertion about almost everywhere convergence in Theorem 3.1 can be replaced by a stronger assertion about convergence on a set whose complement has Hausdorff dimension at most one. The same can be said of Theorem 1.2. There is an example in Section 6 that shows that F_n may diverge on an uncountable, dense subset of $\mathbb{C}_\infty \setminus D$.

We conclude this section by discussing other theorems of a similar nature to Theorem 3.1. Suppose that we continue to assume that the sequence f_n satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$, but adjust the configuration of a, b, c, and D. A few possibilities are shown in Fig. 3.2.

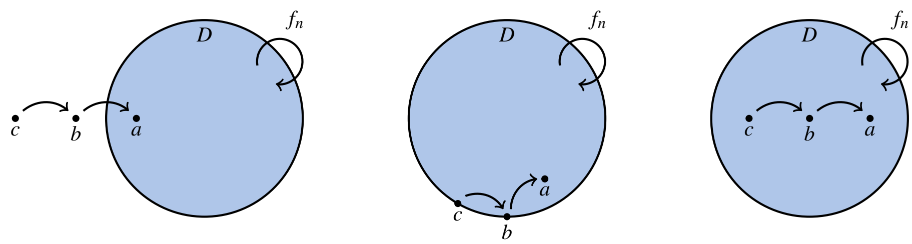

Fig. 3.2 Each map f_n satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$

The left-hand configuration arises in the oval theorem, which was referred to near the end of the previous section. There are other theorems corresponding to the other configurations: some trivial, and some no less significant, from this perspective, than the parabola theorem.

4 The Stern–Stolz series

For this section only we consider more general continued fractions of the form $\mathbf{K}(a_n | b_n)$, where the coefficients b_n need not necessarily equal 1. Our goal is to understand the geometric significance of the *Stern–Stolz series*

$$\left|b_1\frac{1}{a_1}\right|+\left|b_2\frac{a_1}{a_2}\right|+\left|b_3\frac{a_2}{a_1a_3}\right|+\left|b_4\frac{a_1a_3}{a_2a_4}\right|+\left|b_5\frac{a_2a_4}{a_1a_3a_5}\right|+\cdots, \tag{4.1}$$

which features heavily in the continued fractions literature. Roughly speaking, we will show that the series measures how close $T_nS_n^{-1}$ is to a Möbius transformation. Here, as usual, $t_n(z)=a_n/(b_n+z)$ and $T_n=t_1\cdots t_n$, and $s_n(z)=a_n/z$ and $S_n=s_1\cdots s_n$.

Let us first go over some of the theory of Möbius transformations, which can be found in [2,18]. Using the Poincaré extension, the group $\mathscr{M}$ of Möbius transformations acts on the upper half-space model $\mathbb{H}^3$ of three-dimensional hyperbolic space. Each Möbius transformation is a conformal isometry of $\mathbb{H}^3$, and every conformal isometry of $\mathbb{H}^3$ arises in this fashion. Let $j=(0,0,1)$. A useful formula for the hyperbolic metric ρ on $\mathbb{H}^3$ is

$$\sinh^2\tfrac{1}{2}\rho(u,v)=\frac{|u-v|^2}{4(u\cdot j)(v\cdot j)}, \tag{4.2}$$

where $u\cdot j$ is the scalar product of u and j (that is, if $u=(u_1,u_2,u_3)$ then $u\cdot j=u_3$).

We also make use of the metric of uniform convergence χ_0 on $\mathscr{M}$, given by

$$\chi_0(f,g)=\sup_{z\in\mathbb{C}_\infty}\chi(f(z),g(z)).$$

The metric space $(\mathscr{M},\chi_0)$ is complete, and it is a topological group. The metric χ_0 is right-invariant. It is not left-invariant; however, given a Möbius transformation h, we can define $M(h)=\exp[\rho(j,h(j))]$, and then

$$\frac{1}{M(h)}\chi_0(f,g)\leqslant\chi_0(hf,hg)\leqslant M(h)\chi_0(f,g) \tag{4.3}$$

for all Möbius maps f and g.

The rest of this section is devoted to proving the following theorem, which itself will later be used to prove Theorem 1.1.

Theorem 4.1 *The following are equivalent:*

(i) $\left|b_1\frac{1}{a_1}\right|+\left|b_2\frac{a_1}{a_2}\right|+\left|b_3\frac{a_2}{a_1a_3}\right|+\left|b_4\frac{a_1a_3}{a_2a_4}\right|+\left|b_5\frac{a_2a_4}{a_1a_3a_5}\right|+\cdots$ *converges;*

(ii) $\sum_n\chi_0(T_nS_n^{-1},T_{n+1}S_{n+1}^{-1})$ *converges;*

(iii) $\sum_n\chi\left(T_n(0),T_{n+2}(0)\right)$ *converges, and the sequences* $T_{2n-1}(0)$ *and* $T_{2n}(0)$ *converge to two distinct values;*

(iv) $\sum_n\rho\left(T_nS_n^{-1}(j),T_{n+1}S_{n+1}^{-1}(j)\right)$ *converges.*

Further, if (i)–(iv) *hold, then there is a Möbius transformation* g *such that* $\chi_0(T_n,gS_n)\to 0$ *as* $n\to\infty$.

For continued fractions of the form $\mathbf{K}(a_n|1)$, this theorem gives part of Theorem 1.1. For continued fractions of the form $\mathbf{K}(1|b_n)$, this theorem gives [5, Thm. 1.7].

Define, for each positive integer n,

$$\lambda_{2n-1}=\frac{a_2a_4\cdots a_{2n-2}}{a_1a_3\cdots a_{2n-1}},\quad \lambda_{2n}=\frac{a_1a_3\cdots a_{2n-1}}{a_2a_4\cdots a_{2n}}. \tag{4.4}$$

Then $\sum_n |b_n\lambda_n|$ is the Stern–Stolz series. Observe that

$$S_{2n-1}(z)=\frac{1}{\lambda_{2n-1}z},\quad S_{2n}(z)=\lambda_{2n}z,\quad S_{2n-1}^{-1}(z)=\frac{1}{\lambda_{2n-1}z},\quad S_{2n}^{-1}(z)=\frac{z}{\lambda_{2n}}.$$

Let $\tau_n(z)=z+b_n\lambda_n$ for $n=1,2,\dots$, and let $\sigma(z)=1/z$.

Lemma 4.1 *We have*

$$S_{n-1}t_nS_n^{-1}=\begin{cases}\sigma\tau_n\sigma & \text{if } n \text{ is odd},\\ \tau_n & \text{if } n \text{ is even}.\end{cases}$$

Proof Let $\beta_n(z)=z+b_n$. Then $t_n=s_n\beta_n$. Therefore $S_{n-1}t_nS_n^{-1}=S_n\beta_nS_n^{-1}$. It is now straightforward to check the odd and even cases separately. □

We now give a pair of lemmas that will be used to handle statement (ii) of Theorem 4.1, which involves the metric χ_0. We denote the identity Möbius transformation by I.

Lemma 4.2 *Let $\tau(z)=z+\mu$. Then*

$$\chi_0(\tau,I)=\begin{cases}2 & \text{if } |\mu|\geqslant 2,\\ \frac{8|\mu|}{4+|\mu|^2} & \text{if } |\mu|\leqslant 2.\end{cases}$$

Proof Observe that

$$\chi_0(\tau,I)=\sup_{z\in\mathbb{C}}\frac{2|\mu|}{\sqrt{1+|z|^2}\sqrt{1+|z+\mu|^2}}.$$

If $|\mu|\geqslant 2$ then this supremum attains the value 2 (the largest possible value of χ_0) at $z=-\frac{1}{2}\mu(1+\sqrt{1-4/|\mu|^2})$. If $|\mu|\leqslant 2$ then the supremum can be obtained by finding the minimum of $(1+|z|^2)(1+|z+\mu|^2)$ over $\mathbb{C}$. The minimum occurs at $z=-\mu/2$, and we omit the details. □

Recall that $\tau_n(z)=z+b_n\lambda_n$ and $\sigma(z)=1/z$. The map σ is a chordal isometry because, acting on the unit sphere, it is a rotation by π that interchanges the north and south poles. It follows that $\chi_0(\sigma f,\sigma g)=\chi_0(f,g)$ for any Möbius maps f and g.

Lemma 4.3 *We have $\chi_0(S_nT_n^{-1},S_{n-1}T_{n-1}^{-1})=\chi_0(\tau_n,I)$.*

Proof Using right-invariance we obtain

$$\chi_0(S_nT_n^{-1},S_{n-1}T_{n-1}^{-1})=\chi_0(S_n,S_{n-1}t_n)=\chi_0(I,S_{n-1}t_nS_n^{-1}).$$

When n is even the result follows immediately from Lemma 4.1. When n is odd, Lemma 4.1 tells us that $\chi_0(I,S_{n-1}t_nS_n^{-1})=\chi_0(I,\sigma\tau_n\sigma)$. Since χ_0 is right-invariant, and σ is a chordal isometry, it again follows that $\chi_0(I,S_{n-1}t_nS_n^{-1})=\chi_0(I,\tau_n)$. □

Next we give a pair of lemmas that will be used to handle statement (iv) of Theorem 4.1, which involves the hyperbolic metric ρ. Recall that $j=(0,0,1)$.

Lemma 4.4 *Let $\tau(z)=z+\mu$. Then $2\sinh\frac{1}{2}\rho(\tau(j),j)=|\mu|$.*

Proof This follows immediately from the hyperbolic metric formula (4.2). □

On $\mathbb{H}^3$, the map $\sigma(z)=1/z$ acts as an inversion in the unit sphere followed by a reflection in the plane $x_2=0$. In particular, σ fixes the point j.

Lemma 4.5 *We have* $\rho(T_nS_n^{-1}(j), T_{n+1}S_{n+1}^{-1}(j)) = \rho(\tau_n(j), j)$.

Proof Using left-invariance we obtain

$$\rho(T_nS_n^{-1}(j), T_{n-1}S_{n-1}^{-1}(j)) = \rho(t_nS_n^{-1}(j), S_{n-1}^{-1}(j)) = \rho(S_{n-1}t_nS_n^{-1}(j), j).$$

When n is even the result follows immediately from Lemma 4.1. When n is odd, Lemma 4.1 tells us that $\rho(S_{n-1}t_nS_n^{-1}(j), j) = \rho(\sigma\tau_n\sigma(j), j)$. Since ρ is left-invariant, and $\sigma(j) = j$, it again follows that $\rho(S_{n-1}t_nS_n^{-1}(j), j) = \rho(\tau_n(j), j)$. □

Next, we develop some preparatory results for dealing with statement (iii) of Theorem 4.1. We need the *cross ratio* $[a,b,c,d]$ of four points a, b, c, and d in $\mathbb{C}_\infty$. It is given by

$$[a,b,c,d] = \frac{(a-b)(c-d)}{(a-c)(b-d)},$$

where the usual conventions are adopted if one of the points a, b, c, or d is ∞. It is well known that if g is a Möbius transformation then

$$[g(a), g(b), g(c), g(d)] = [a,b,c,d].$$

Remember that $t_n(z) = a_n/(b_n + z)$ and $T_n = t_1 \cdots t_n$.

Lemma 4.6 *We have*

$$\frac{T_n(0) - T_{n-2}(0)}{T_n(0) - T_{n-1}(0)} = \frac{b_n}{a_n}\, T_{n-1}^{-1}(\infty).$$

Proof The result follows by expanding the identity

$$[t_n(0), \infty, 0, T_{n-1}^{-1}(\infty)] = [T_n(0), T_{n-1}(\infty), T_{n-1}(0), \infty]$$

and observing that $T_{n-1}(\infty) = T_{n-2}(0)$. □

Recall the sequence λ_n defined by (4.4).

Corollary 4.1 *We have*

$$\frac{T_n(0) - T_{n-2}(0)}{T_n(0) - T_{n-1}(0)} = -\lambda_{n-1}\lambda_n T_{n-1}^{-1}(\infty)T_n^{-1}(\infty) + 1.$$

Proof Notice that

$$T_n^{-1}(\infty) = t_n^{-1}(T_{n-1}^{-1}(\infty)) = -b_n + \frac{a_n}{T_{n-1}^{-1}(\infty)}.$$

Therefore

$$\frac{b_n}{a_n}\, T_{n-1}^{-1}(\infty) = -\frac{1}{a_n}\, T_{n-1}^{-1}(\infty)T_n^{-1}(\infty) + 1.$$

The result follows, because $\lambda_{n-1}\lambda_n = 1/a_n$. □

We need one final lemma, on the convergence of series.

Lemma 4.7 *Suppose that z_n is a sequence of complex numbers such that $\sum_n |z_nz_{n+1} - 1|$ converges. Then z_{2n-1} converges to a non-zero limit z, and z_{2n} converges to $1/z$.*

Proof As the sum $\sum_n |z_{2n}z_{2n+1}-1|$ converges, we see that the product $z_2z_3\cdots z_{2n+1}$ converges to a non-zero value, and as the sum $\sum_n |z_{2n-1}z_{2n}-1|$ converges, we see that the product $z_1z_2\cdots z_{2n}$ converges to a non-zero value. Thus z_{2n+1} converges to a non-zero number z, and $z_{2n}\to 1/z$ because $z_nz_{n+1}\to 1$. $\square$

Finally we can prove Theorem 4.1.

Proof (of Theorem 4.1) We begin by proving the final statement of Theorem 4.1. That statement follows from (ii), because (ii) implies that the sequence $T_nS_n^{-1}$ is a Cauchy sequence in $(\mathscr{M},\chi_0)$, and hence there is a Möbius transformation g such that $\chi_0(T_nS_n^{-1},g)\to 0$. Then right-invariance of χ_0 implies that $\chi_0(T_n,gS_n)\to 0$.

Next we show that the series $\sum_n \chi_0(T_nS_n^{-1},T_{n+1}S_{n+1}^{-1})$ converges if and only if the series $\sum_n \chi_0(S_nT_n^{-1},S_{n+1}T_{n+1}^{-1})$ converges. Suppose that $\sum_n \chi_0(T_nS_n^{-1},T_{n+1}S_{n+1}^{-1})$ converges. Using (4.3) we have

$$\begin{aligned}\chi_0(S_nT_n^{-1},S_{n+1}T_{n+1}^{-1}) &= \chi_0(S_nT_n^{-1}T_{n+1}S_{n+1}^{-1},S_nT_n^{-1}T_nS_n^{-1})\\ &\leqslant M(S_nT_n^{-1})\ \chi_0(T_{n+1}S_{n+1}^{-1},T_nS_n^{-1}).\end{aligned}$$

We have seen that $T_nS_n^{-1}$ converges to a Möbius map g. Since $(\mathscr{M},\chi_0)$ is a topological group it follows that $S_nT_n^{-1}$ converges to g^{-1}. Therefore the sequence $M(S_nT_n^{-1})$ is bounded above. Hence $\sum_n \chi_0(S_nT_n^{-1},S_{n+1}T_{n+1}^{-1})$ converges. This argument can be run in reverse, so $\sum_n \chi_0(T_nS_n^{-1},T_{n+1}S_{n+1}^{-1})$ converges if and only if $\sum_n \chi_0(S_nT_n^{-1},S_{n+1}T_{n+1}^{-1})$ converges.

Lemmas 4.2 and 4.3 tell us that $\sum_n \chi_0(S_nT_n^{-1},S_{n+1}T_{n+1}^{-1})$ converges if and only if the Stern–Stolz series $\sum_n |b_n\lambda_n|$ converges. The equivalence of (i) and (ii) follows immediately. Lemmas 4.4 and 4.5 tell us that the series $\sum_n \rho(T_nS_n^{-1}(j),T_{n+1}S_{n+1}^{-1}(j))$ converges if and only if $\sum_n |b_n\lambda_n|$ converges, which implies that (i) and (iv) are equivalent.

Next we show that (ii) implies (iii). Observe first that

$$\chi(T_n(0),T_{n+2}(0)) = \chi(T_{n+1}(\infty),T_{n+2}(0)).$$

If n is odd then $S_{n+1}^{-1}(\infty)=\infty$ and $S_{n+2}^{-1}(\infty)=0$, and if n is even then $S_{n+1}^{-1}(0)=\infty$ and $S_{n+2}^{-1}(0)=0$. Therefore $\chi(T_{n+1}(\infty),T_{n+2}(0))$ is less than

$$\chi\left(T_{n+1}S_{n+1}^{-1}(0),T_{n+2}S_{n+2}^{-1}(0)\right)+\chi\left(T_{n+1}S_{n+1}^{-1}(\infty),T_{n+2}S_{n+2}^{-1}(\infty)\right).$$

Since both terms in this expression do not exceed $\chi_0\left(T_{n+1}S_{n+1}^{-1},T_{n+2}S_{n+2}^{-1}\right)$ we see that

$$\chi(T_n(0),T_{n+2}(0)) \leqslant 2\chi_0\left(T_{n+1}S_{n+1}^{-1},T_{n+2}S_{n+2}^{-1}\right).$$

Therefore $\sum_n \chi(T_n(0),T_{n+2}(0))$ converges. Recall from the start of this proof that $T_nS_n^{-1}$ converges to a Möbius map g. It follows that $T_{2n-1}(0)\to g(\infty)$ and $T_{2n}(0)\to g(0)$, and these two limits are distinct.

It remains to show that (iii) implies (i). We can assume, by adjusting b_1 if necessary, that the two distinct limits p and q of the sequences $T_{2n-1}(0)$ and $T_{2n}(0)$ are finite. Since Euclidean and chordal metrics are locally equivalent, and $\sum_n \chi(T_n(0),T_{n-2}(0))$ converges, we deduce that $\sum_n |T_n(0)-T_{n-2}(0)|$ converges. Furthermore, because $|T_n(0)-T_{n-1}(0)|\to |p-q|$ we see that

$$\sum_n \left|\frac{T_n(0)-T_{n-2}(0)}{T_n(0)-T_{n-1}(0)}\right|$$

converges. Let $z_n = \lambda_n T_n^{-1}(\infty)$. Then, from Corollary 4.1, we find that $\sum_n |z_n z_{n+1} - 1|$ converges. Hence, by Lemma 4.7, the sequence $|z_n|$ is bounded below by a positive constant M. Next, from Lemma 4.6 we have that $\sum_n \left|\frac{b_n}{a_n} T_{n-1}^{-1}(\infty)\right|$ converges. Now

$$|b_n\lambda_n| = \left|\frac{b_n}{\lambda_{n-1}a_n}\right| = \left|\frac{1}{z_{n-1}}\right|\left|\frac{b_n}{a_n}\,T_{n-1}^{-1}(\infty)\right| \leqslant \frac{1}{M}\left|\frac{b_n}{a_n}\,T_{n-1}^{-1}(\infty)\right|.$$

Therefore the Stern–Stolz series $\sum_n |b_n\lambda_n|$ converges. □

5 Proof of Theorem 1.1

Before we prove Theorem 1.1 let us revise a few concepts from the theory of Kleinian groups that apply to continued fractions. These ideas, within the context of Kleinian groups, can be found in [16].

Recall that $j = (0,0,1)$. A sequence $F_1, F_2, \dots$ of Möbius transformations is *restrained* if the sequence $F_n(j)$ accumulates only on the ideal boundary $\mathbb{C}_\infty$ of $\mathbb{H}^3$. This terminology was introduced by Lorentzen (formerly Jacobsen) and Thron in [10]. Our definition is taken from [1] and [3, Sec. 7], and it differs from (but is equivalent to) Lorentzen's original definition. The sequence F_n is restrained if and only if $\rho(F_n(j), j) \to \infty$ as $n \to \infty$. Since $\rho(F_n(j), j) = \rho(j, F_n^{-1}(j))$ it follows that F_n^{-1} is also restrained. We are often interested in sequences F_n for which $F_n(j)$ converges to a point on the ideal boundary $\mathbb{C}_\infty$ (in the chordal metric). Such sequences are said to be *generally convergent*; they were first studied in the context of continued fractions by Lorentzen in [8].

A *backward limit point* of a restrained sequence F_n is an accumulation point of the backward orbit $F_n^{-1}(j)$. That is, a point p in $\mathbb{C}_\infty$ is a backward limit point of F_n if there is a subsequence of $F_1^{-1}(j), F_2^{-1}(j), \dots$ that converges to p in the chordal metric. The *backward limit set* $\Lambda(F_n)$ of F_n is the collection of all backward limit points. A *forward limit point* of F_n is an accumulation point of the forward orbit $F_1(j), F_2(j), \dots$, although we have little use for this definition. If $F_1, F_2, \dots$ are all the elements of a Kleinian group then, because groups are closed under taking inverses, backward limit points and forward limit points are the same. This is why there is no distinction between backward and forward limit points in the theory of Kleinian groups.

The point j has no special significance in the definitions so far, and we can replace it with any other point in $\mathbb{H}^3$ without consequence. In particular, the set of accumulation points of $F_1^{-1}(w), F_2^{-1}(w), \dots$ is $\Lambda(F_n)$ for any point w in $\mathbb{H}^3$. In fact, providing a point p in $\mathbb{C}_\infty$ is not a forward limit point of F_n, the set of accumulation points of $F_1^{-1}(p), F_2^{-1}(p), \dots$ is also $\Lambda(F_n)$ (see, for example, [1, Thm. 3.5]).

A point p in $\mathbb{C}_\infty$ is a *conical limit point* of the restrained sequence F_n if there is a hyperbolic geodesic γ with one end-point at p and a subsequence of $F_1^{-1}(j), F_2^{-1}(j), \dots$ that lies within a bounded hyperbolic distance of γ and converges to p in the chordal metric. Elementary hyperbolic geometry can be used to show that this definition is independent of the choice of geodesic γ, and j can be replaced by any other point in $\mathbb{H}^3$. The *conical limit set* $\Lambda_c(F_n)$ of F_n is the set of all conical limit points of F_n. This set is contained in the backward limit set of F_n. The conical limit set is important in continued fraction theory because of the following theorem due to Aebischer [1, Thm. 5.2] .

Theorem 5.1 *Let F_n be a restrained sequence of Möbius transformations and let $p \in \mathbb{C}_\infty$. Then $\chi(F_n(j), F_n(p)) \to 0$ if and only if $p \notin \Lambda_c(F_n)$.*

We also record a corollary of this theorem [1, Prop. 5.3]. Remember that a sequence F_n is generally convergent if $F_n(j)$ converges in the chordal metric to a point on $\mathbb{C}_\infty$.

Corollary 5.1 *If F_n is a generally convergent sequence of Möbius transformations then, providing it has more than one backward limit point, F_n diverges everywhere on its conical limit set.*

A sequence $F_1, F_2, \dots$ of Möbius transformations is said to be a *rapid escape sequence* if, for a point w in $\mathbb{H}^3$, the sum $\sum_n \exp[-\rho(w, F_n(w))]$ converges. Once again, this definition is independent of the particular point w chosen. We use the phrase "rapid escape" because the forward orbit of a rapid escape sequence at a point w approaches the ideal boundary particularly quickly. Clearly, rapid escape sequences are restrained. The Hausdorff dimension of the conical limit set of a rapid escape sequence does not exceed one (see [16, Cor. 9.3.2]). That is the reason why we obtain convergence everywhere but on a set of Hausdorff dimension at most one in the proof of Theorem 3.1. This issue is covered in more detail in [20]. Rapid escape sequences, and the related concept of the *critical exponent*, play an important role in Kleinian group theory (see [16]).

Lemma 5.1 *Let F_n and G_n be two sequences of Möbius transformations such that F_n is restrained and $G_nF_n^{-1}$ converges uniformly to another Möbius transformation g. Then G_n is also restrained and*

(i) *p is a backward limit point of F_n if and only if p is a backward limit point of G_n;*
(ii) *p is a conical limit point of F_n if and only if p is a conical limit point of G_n;*
(iii) *F_n is a rapid escape sequence if and only if G_n is a rapid escape sequence.*

Proof Since $G_nF_n^{-1}(j) \to g(j)$ it follows that the sequence $\rho(G_nF_n^{-1}(j), j)$ is bounded above. But $\rho(F_n^{-1}(j), G_n^{-1}(j)) = \rho(G_nF_n^{-1}(j), j)$, so the sequence $\rho(F_n^{-1}(j), G_n^{-1}(j))$ is bounded above. It is now immediate that G_n is restrained and (i), (ii), and (iii) hold. □

Recall that $s_n(z) = a_n/z$ and $S_n = s_1 \cdots s_n$. Remember also the definition of λ_n given in (4.4), and the formulas for S_n^{-1} that follow that definition.

Lemma 5.2 *If $S_n^{-1}(j)$ has modulus greater than 1 then*

$$\exp\left[-\rho(S_n^{-1}(j), j)\right] = |\lambda_n|.$$

Proof If μ is a complex number with modulus greater than 1 then $\rho(\mu j, j) = \log|\mu|$. Thus $\exp[-\rho(\mu j, j)] = 1/|\mu|$. Since $S_n^{-1}(j) = j/|\lambda_n|$, we see that $\exp\left[-\rho(S_n^{-1}(j), j)\right] = |\lambda_n|$. □

The sequence $S_n^{-1}(j)$ is confined to the vertical geodesic from 0 to ∞. It follows that the only possible backward limit points of S_n are 0 and ∞, and each of these points is a backward limit point if and only if it is a conical limit point.

We need one final lemma on hyperbolic geometry (see [2, Thm. 7.9.1]).

Lemma 5.3 *Let γ be a geodesic in $\mathbb{H}^3$ that lands at points a and b in $\mathbb{C}_\infty$. Then*

$$\cosh \rho(j, \gamma) = \frac{2}{\chi(a,b)}.$$

In particular,

$$\exp[-\rho(j, \gamma)] \geqslant \frac{\chi(a,b)}{4}.$$

We can now prove Theorem 1.1.

Proof (of Theorem 1.1) The equivalence of (i)–(iv) follows from Theorem 4.1.

Let us show that (i) implies (v). If the Stern–Stolz series $\sum_n |\lambda_n|$ converges then $\lambda_n \to 0$, so eventually $|\lambda_n| < 1$. Since $|S_n^{-1}(j)| = 1/|\lambda_n|$ it follows that eventually $|S_n^{-1}(j)| > 1$, and then Lemma 5.2 tells us that the sum $\sum_n \exp\left[-\rho(S_n^{-1}(j), j)\right]$ converges. Also, because $S_n^{-1}(j) \to \infty$, we see that ∞ is the only backward limit point of S_n.

Now we show that (v) implies (i). Because S_n is a rapid escape sequence it follows that $\rho(S_n^{-1}(j), j) \to \infty$ (that is, S_n is restrained). Since ∞ is the only backward limit point of S_n we deduce that $S_n^{-1}(j) \to \infty$ as $n \to \infty$. Therefore $|S_n^{-1}(j)| > 1$ for sufficiently large n, and again we appeal to Lemma 5.2, this time to see that convergence of $\sum_n \exp\left[-\rho(S_n^{-1}(j), j)\right]$ implies convergence of the Stern–Stolz series $\sum_n |\lambda_n|$.

Next we show that (ii) and (v) imply (vi). We know that S_n is restrained, by (v), because it is a rapid escape sequence. Furthermore, because $S_n^{-1}(j)$ is constrained to lie on the vertical geodesic between 0 and ∞, S_n has only a single backward limit point, and only a single conical limit point, ∞. From Theorem 4.1, and using (ii), we know that $T_nS_n^{-1}$ converges to another Möbius transformation g. We can now apply Lemma 5.1 to deduce that T_n, like S_n, is a rapid escape sequence with only a single backward limit point, and only a single conical limit point, ∞.

Last we show that (vi) implies (iii). Let γ denote the hyperbolic geodesic between -1 and 0. Let w_n denote the point on γ such that $\rho(\gamma, T_n^{-1}(j)) = \rho(w_n, T_n^{-1}(j))$. Since T_n is restrained, and it has only a single backward limit point ∞, we see that $T_n^{-1}(j) \to \infty$ as $n \to \infty$. It follows that w_n converges to the highest point on γ, namely $(-1 + j)/2$. In particular, there is a positive constant K such that $\rho(w_n, j) < K$ for all positive integers n. Applying the triangle inequality gives

$$\rho(j, T_n^{-1}(j)) \leqslant \rho(j, w_n) + \rho(w_n, T_n^{-1}(j)) < K + \rho(\gamma, T_n^{-1}(j)).$$

Therefore, as the series $\sum_n \exp[-\rho(j, T_n^{-1}(j))]$ converges, the series $\sum_n \exp[-\rho(\gamma, T_n^{-1}(j))]$ also converges. Hence $\sum_n \exp[-\rho(T_n(\gamma), j)]$ converges. It follows from Lemma 5.3 that the series $\sum_n \chi(T_n(-1), T_n(0))$ converges. Since $T_n(-1) = T_{n-2}(0)$ we see that the series $\sum_n \chi(T_{n-2}(0), T_n(0))$ converges too. This implies that the sequences $T_{2n-1}(0)$ and $T_{2n}(0)$ each converge. Suppose they converge to the same limit p. Then $T_n(0) \to p$ and $T_n(\infty) \to p$. Moreover, because j lies on the geodesic between 0 and ∞ it follows that $T_n(j) \to p$. However, we know from Theorem 5.1 that $\chi(T_n(j), T_n(\infty)) \nrightarrow 0$ as $n \to \infty$ because ∞ is a conical limit point. This contradiction show that the limits of $T_{2n-1}(0)$ and $T_{2n}(0)$ are distinct.

It remains to prove the final assertion in Theorem 1.1. Recall that $S_{2n-1}(z) = 1/(\lambda_{2n-1}z)$ and $S_{2n}(z) = \lambda_{2n}z$, where $|\lambda_n|$ is the nth term in the Stern–Stolz series. When $\sum_n |\lambda_n|$ converges (statement (i)), and hence $\lambda_n \to 0$, it follows that $S_{2n-1} \to \infty$ and $S_{2n} \to 0$ locally uniformly on $\mathbb{C}$, and $S_{2n-1}(\infty) = 0$ and $S_{2n}(\infty) = \infty$. Now, we know from the final assertion in Theorem 4.1 that there is a Möbius map g such that $\chi_0(T_n, gS_n) \to 0$ as $n \to \infty$. Therefore $T_{2n-1} \to g(\infty)$ and $T_{2n} \to g(0)$ locally uniformly on $\mathbb{C}$, and $T_{2n-1}(\infty) \to g(0)$ and $T_{2n}(\infty) \to g(\infty)$. □

6 Proof of Theorem 1.2

Theorem 1.2 can easily be deduced from Theorems 1.1 and 3.1.

Proof (of Theorem 1.2) Suppose first that the Stern–Stolz series converges. Then the final assertion of Theorem 1.1 tells us that there are distinct points p and q such that $T_{2n-1} \to p$ and $T_{2n} \to q$ locally uniformly on $\mathbb{C}$, and $T_{2n-1}(\infty) \to q$ and $T_{2n}(\infty) \to p$.

Suppose now that the Stern–Stolz series diverges. Theorem 3.1 tells us that T_{2n-1} and T_{2n} converge locally uniformly on H, and almost everywhere on $\mathbb{C}_\infty$, to points p and q. That theorem also says that $\sum_n \chi(T_n(0), T_{n+2}(0))$ converges. By comparing statements (i) and (iii) of Theorem 1.1 we see that $p = q$. Therefore T_n converges locally uniformly on H, and almost everywhere on $\mathbb{C}_\infty$, to a single point. □

It is interesting that many well-known concepts from continued fraction theory can be interpreted using geometric properties of Möbius transformations. The geometric approach often provides insight into continued fractions that is difficult to obtain from algebra alone. We provide another example of this here, by constructing a convergent continued fraction $\mathbf{K}(a_n|\,1)$ such that the sequence $t_1, t_2, \dots$ satisfies $t_n(H) \subseteq H$ for $n = 1, 2, \dots$ and T_n diverges on an uncountable, dense subset of the complement of H. This shows that the assertion about almost everywhere convergence in Theorem 1.2 cannot be strengthened significantly.

For simplicity, let us choose H to be the half-plane given by $\mathrm{Re}[z] > -1/2$, although a similar construction works for other half-planes. Suppose that $\mathbf{K}(a_n|\,1)$ is a convergent continued fraction such that $t_n(H) \subseteq H$ for $n = 1, 2, \dots$. The value of $\mathbf{K}(a_n|\,1)$ – that is, the limit of the sequence $T_n(0)$ – is necessarily contained in $\overline{H}$ (and it cannot be 0 or ∞). In fact, it is known that every element in $\overline{H} \setminus \{0, \infty\}$ is the value of some such continued fraction (see, for example, [15, Thm. 3.47]). This observation, which is implicit in our arguments below, is key to our construction. We shortcut the usual proofs of the observation by considering the dynamics of anticonformal Möbius transformations.

Choose any point q of H, and define t to be the anticonformal Möbius transformation $t(z) = a/(1+\bar{z})$, where $a = q + |q|^2$. This has exactly two fixed points, namely q and $\tau(q)$, where τ is the reflection in ∂H, given by $\tau(z) = -1 - \bar{z}$. Each anticonformal Möbius map with exactly two fixed points is conjugate to a map of the form $z \mapsto \lambda \bar{z}$, where $\lambda > 1$. One can check that q is the attracting fixed point of t and $\tau(q)$ is the repelling fixed point. Therefore the sequence t^n converges to q locally uniformly on the complement of $\tau(q)$. It follows, in particular, that $t(H) \subseteq H$.

Lemma 6.1 *Given points p and q in H, and $\varepsilon > 0$, there are Möbius maps $t_i(z) = a_i/(1+z)$, $i = 1, \dots, n$, that satisfy $t_i(H) \subseteq H$ and are such that $\chi(t_1 \cdots t_n(p), q) < \varepsilon$.*

Proof Choose n to be suitably large that $\chi(t^n(\bar{p}), q) < \varepsilon$ and $\chi(t^n(p), q) < \varepsilon$, where $t(z) = a/(1+\bar{z})$ and $a = q + |q|^2$. Let a_i equal a if i is odd, and $\bar{a}$ if i is even. Then $t_1 \cdots t_n(p)$ is equal to either $t^n(\bar{p})$ if n is odd, or $t^n(p)$ if n is even, so the result follows. □

Using Lemma 6.1 we can choose an infinite sequence $t_1, t_2, \dots$ of Möbius maps, where $t_n(z) = a_n/(1+z)$ and $t_n(H) \subseteq H$, such that the orbit $t_n \cdots t_1(0)$, $n = 1, 2, \dots$, is dense in H. Notice that we work with the sequence $t_n \cdots t_1(0)$ rather than the usual sequence $t_1 \cdots t_n(0)$. Let α be the involution given by $\alpha(z) = -1 - z$, and let L denote the half-plane given by $\mathrm{Re}[z] < -1/2$. The map α interchanges L and H. Since $\alpha(-1) = 0$, it follows that the orbit $\alpha t_n \cdots t_1 \alpha(-1)$ is dense in L. Observe that $\alpha t_i \alpha = t_i^{-1}$ for each positive integer i, and so

$$\alpha t_n \cdots t_1 \alpha = t_n^{-1} \cdots t_1^{-1} = T_n^{-1}.$$

Therefore we have shown that the orbit $T_n^{-1}(-1)$ is dense in L. Now, -1 is not a forward limit point of T_n, or in other words the sequence $T_n(j)$ does not accumulate at -1, because this sequence is confined to the hyperbolic half-space with ideal boundary H. It follows that the backward limit set $\Lambda(T_n)$ of T_n, which is the set of accumulation points of

$T_1^{-1}(j), T_2^{-1}(j), \ldots$, is equal to the set of accumulation points of $T_1^{-1}(-1), T_2^{-1}(-1), \ldots$. Therefore $\Lambda(T_n) = \overline{L}$.

Since the backward limit set of T_n contains points other than ∞, we see from the equivalence of (i) and (vi) in Theorem 1.1 that the Stern–Stolz series diverges. Therefore $T_n(0)$ and $T_n(\infty)$ converge to a point p (in the chordal metric). Since $T_n(j)$ lies on the geodesic joining $T_n(0)$ and $T_n(\infty)$, we see that $T_n(j)$ converges to p also. That is, T_n is generally convergent, with limit p.

It is known that if the backward limit set of a restrained sequence of Möbius transformations contains an open disc E, then the conical limit set is uncountable and its closure contains E (see [6, Lem. 5.12]). As $\Lambda(T_n) = \overline{L}$, it follows that $\Lambda_c(T_n)$ is uncountable and $\overline{\Lambda_c(T_n)} \supseteq L$. Therefore $\Lambda_c(T_n)$ is dense in L. This completes our construction because, by Corollary 5.1, T_n diverges everywhere on $\Lambda_c(T_n)$.

7 The geometry of the parabola theorem

As we saw in Section 2, the significance of the parabolic region in the parabola theorem is that it gives rise to the inclusions $t_n(H) \subseteq H$, where H is a half-plane. This leads to a nested sequence of discs

$$H \supseteq T_1(H) \supseteq T_2(H) \supseteq T_3(H) \supseteq \cdots.$$

Associated to this sequence of discs is a sequence of points $T_1(\infty), T_2(\infty), \ldots$. Since $\infty \in \partial H$, it follows that $T_n(\infty) \in T_n(\partial H)$ for each integer n. Our geometric approach to the parabola theorem allows us to describe all possible sequences of discs and points that arise in that theorem. Beardon carried out a similar programme for the Śleszyśki–Pringsheim theorem in [4].

Given an open disc D in $\mathbb{C}_\infty$, a point a in D, a point b in ∂D, and a point c that lies outside $\overline{D}$, we define a *parabola sequence* to be a sequence of Möbius transformations $f_1, f_2, \ldots$ that satisfies $f_n(D) \subseteq D$, $f_n(c) = b$, and $f_n(b) = a$ for each integer n. When $a = 0$, $b = \infty$, $c = -1$, and $D = H_\alpha$ we recover the usual sequence of maps t_n that arise in the classical parabola theorem. For simplicity, we focus only on *symmetric* parabola sequences, which are parabola sequences such that a and c are inverse points with respect to ∂D. Let $F_n = f_1 \cdots f_n$, and let F_0 be the identity map.

Theorem 7.1 *Let $f_1, f_2, \ldots$ be a symmetric parabola sequence with associated disc D and points a, b, and c. Let $D_n = F_n(D)$ and $z_n = F_n(b)$. Then*

(i) $D_0 \supseteq D_1 \supseteq D_2 \supseteq \cdots$,
(ii) $z_n \in \partial D_n$,
(iii) $z_{n-1} \in \mathbb{C}_\infty \setminus \overline{D_n}$ *and* $z_{n+1} \in D_n$, *and these are inverse points with respect to* ∂D_n.

Conversely, any sequence of discs D_n and points z_n satisfying (i), (ii), *and* (iii) *arise as the F_n-images of D and b for some symmetric parabola sequence f_n.*

There is a similar theorem for general parabola sequences, which has a more elaborate version of statement (iii) involving hyperbolic distance.

Theorem 7.1 is illustrated in Fig. 7.1.

Proof (of Theorem 7.1) Suppose that f_n is a symmetric parabola sequence. Property (i) follows from the inclusion $F_n(D) \subseteq F_{n-1}(D)$. Property (ii) holds because $b \in \partial D$. Property (iii) follows by preservation of inverse points under Möbius transformations, as $z_{n-1} = F_n(c)$ and $z_{n+1} = F_n(a)$.

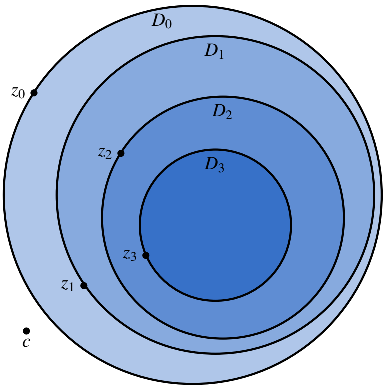

Fig. 7.1 Orbits of discs and points under a symmetric parabola sequence

Now suppose that D_n and z_n have been given to satisfy (i), (ii), and (iii). Define $H = D_0$, $b = z_0$, $a = z_1$, and let c be the inverse point of a in ∂D. For each positive integer n choose a Möbius transformation F_n that satisfies $F_n(D) = D_n$, $F_n(b) = z_n$, and $F_n(c) = z_{n-1}$. From property (iii), and because Möbius transformations preserve inverse points, we see that $F_n(a) = z_{n+1}$. Therefore the sequence $f_1, f_2, \dots$ defined by $f_n = F_{n-1}^{-1} F_n$ is a symmetric parabola sequence with $D_n = F_n(D)$ and $z_n = F_n(b)$. □

There is a subtle geometric fact that is not immediately apparent from Theorem 7.1: if the sequence z_n converges then the intersection of the nested closed discs $\overline{D_n}$ is a single point. A proof of this can be extracted from [17]. In general, the intersection of a nested sequence of closed discs is either a single point or a closed disc, and it is usual in continued fraction theory to refer to these two alternatives as the *limit-point case* and the *limit-disc case*, respectively.

For more general parabola sequences, it is no longer true that convergence of z_n can only arise with the limit-point case. To see why this is so, consider the map

$$t(z) = \frac{-3/16}{1+z},$$

which is a loxodromic Möbius transformation with attracting fixed point $-1/4$ and repelling fixed point $-3/4$. Let $t_n = t$ for each positive integer n, so that $T_n = t^n$. If H is the half-plane given by $\mathrm{Re}[z] > -1/2$ then $t_1, t_2, \dots$ and H together form a symmetric parabola sequence, because -1 and 0 are inverse points with respect to ∂H. Because the repelling fixed point of t lies outside $\overline{H}$, the limit-point case occurs for this parabola sequence. The limit point is $-1/4$. If instead H is the half-plane $\mathrm{Re}[z] > -3/4$ then t_n and H again form a parabola sequence, but this time the parabola sequence is not symmetric, because -1 and 0 are not inverse points with respect to ∂H. The repelling fixed point of t belongs to ∂H for this parabola sequence, and it follows that the limit-disc case occurs: the limit disc is given by $|z+1/2| \leqslant 1/4$.

8 Higher dimensions

All our methods generalise to higher dimensions, and the results and their proofs go through virtually unchanged. We give just one example of this, namely a version of Theorem 1.2 in many dimensions. See [2,18] for the theory of Möbius transformations in several dimensions.

So far we have only considered Möbius transformations acting on $\mathbb{C}_\infty$, and in particular we have studied continued fractions using sequences of maps t_n given by $t_n(z) = a_n/(b_n+z)$. Each of these maps takes ∞ to 0. Now we would like to consider Möbius transformations that act on $\mathbb{R}^N \cup \{\infty\}$. Consider a sequence of Möbius maps $t_1, t_2, \dots$ acting on $\mathbb{R}^N \cup \{\infty\}$ that satisfies $t_n(\infty) = 0$ for each integer n. Let σ be inversion in the unit sphere. Then t_n can be expressed in the form

$$t_n(x) = a_n \sigma(b_n + x),$$

where $b_n \in \mathbb{R}^N$, and a_n denotes an orthogonal map of $\mathbb{R}^N$ followed by a dilation. That is, there is a positive scalar λ_n and an orthogonal map A_n such that $a_n(x) = \lambda_n A_n(x)$. In the two dimensional case a_n is a complex number, $\lambda_n = |a_n|$, and A_n is a rotation by $a_n/|a_n|$. Let us denote λ_n by $|a_n|$ even in higher dimensions. Because σ is anticonformal, we must also declare that A_n is anticonformal in order for t_n to be conformal. This is not strictly necessary, as we have not needed conformality so far, but it tallies with the two-dimensional case in which all the maps t_n are conformal. The Stern–Stolz series in higher dimensions is the series

$$|b_1|\left(\frac{1}{|a_1|}\right) + |b_2|\left(\frac{|a_1|}{|a_2|}\right) + |b_3|\left(\frac{|a_2|}{|a_1||a_3|}\right) + |b_4|\left(\frac{|a_1||a_3|}{|a_2||a_4|}\right) + \cdots.$$

Now, for the parabola theorem we need all the coefficients b_n to be equal, so let $b_n = (1,0,\dots,0)$ for each integer n, and we write this more simply as $b_n = 1$. Our maps t_n now have the form $t_n(x) = a_n\sigma(1+x)$, and as usual we let $T_n = t_1 \cdots t_n$. Let -1 denote the point $(-1,0,\dots,0)$ and let H be a Euclidean half-space that contains 0, but does not contain -1, even its closure. As before, we let u be the point inverse to -1 in ∂H, and let us also define $v = \sigma(1+u)$. Using the argument of Section 2 we see that $t_n(H) \subseteq H$ if and only if $t_n(u)$ lies in the region $\{x \in \mathbb{R}^N : |x| \leqslant d(x, \partial H)\}$ bounded by a paraboloid. Notice that $t_n(u) = a_n v$. We now have a strong version of the parabola theorem, valid in several dimensions.

The paraboloid theorem *Let H be a Euclidean half-space in $\mathbb{R}^N$ that contains 0, but does not contain -1, even in its closure, and suppose that $a_n v \in \{z \in \mathbb{C} : |z| \leqslant d(z, \partial H)\}$ for each positive integer n. If the Stern–Stolz series*

$$\left(\frac{1}{|a_1|}\right) + \left(\frac{|a_1|}{|a_2|}\right) + \left(\frac{|a_2|}{|a_1||a_3|}\right) + \left(\frac{|a_1||a_3|}{|a_2||a_4|}\right) + \cdots.$$

converges, then there are distinct points p and q such that $T_{2n-1} \to p$ and $T_{2n} \to q$ locally uniformly on $\mathbb{R}^N$, and $T_{2n-1}(\infty) \to q$ and $T_{2n}(\infty) \to p$. If the Stern–Stolz series diverges then T_n converges locally uniformly on H, and almost everywhere on $\mathbb{R}^N_\infty$, to a single point.

This framework for describing continued fractions in several dimensions can be used to generalise many other results on continued fractions. For instance, the Śleszyśki–Pringsheim theorem says that $\mathbf{K}(a_n|b_n)$ converges if $|b_n| \geqslant 1 + |a_n|$ for each integer n. In higher dimensions, using our notation $t_n(x) = a_n\sigma(b_n + x)$ and $T_n = t_1 \cdots t_n$, the same sequence of inequalities $|b_n| \geqslant 1 + |a_n|$ guarantees convergence of $T_1(0), T_2(0), \dots$.

References

1. Aebischer, B.: The limiting behaviour of sequences of Möbius transformations. Math. Zeit. 205, 49–59 (1990)
2. Beardon, A. F.: The Geometry of Discrete Groups. Graduate Texts in Mathematics, 91, Springer-Verlag, New York (1983)
3. Beardon, A. F.: Continued Fractions, Discrete Groups and Complex Dynamics. Comput. Methods and Funct. Theory 1, 535–594 (2001)
4. Beardon, A. F.: The geometry of Pringsheim's continued fractions. Geom. Dedicata 84, 125–134 (2001)
5. Beardon, A. F., Short, I.: The Seidel, Stern, Stolz and Van Vleck theorems on continued fractions. Bull. Lond. Math. Soc. 42, 457–466 (2010)
6. Crane, E., Short, I.: Conical limit sets and continued fractions. Conform. Geom. Dyn. 11, 224–249 (2007)
7. Hillam, K. L., Thron, W. J.: A general convergence criterion for continued fractions $K(a_n/b_n)$. Proc. Amer. Math. Soc. 16, 1256–1262 (1965)
8. Jacobsen, L.: General convergence of continued fractions. Trans. Amer. Math. Soc. 294, 477–485 (1986)
9. Jacobsen, L., Thron, W. J.: Oval convergence regions and circular limit regions for continued fractions $K(a_n/1)$. Analytic theory of continued fractions II, 90–126, Lecture Notes in Math., 1199, Springer, Berlin (1986)
10. Jacobsen, L., Thron, W. J.: Limiting structures for sequences of linear fractional transformations. Proc. Amer. Math. Soc. 99, 141–146 (1987)
11. Jones, W. B., Thron, W. J.: Continued fractions. Analytic theory and applications. Encyclopedia of Mathematics and its Applications, 11, Addison-Wesley Publishing Co., Reading, Mass. (1980)
12. Lane, R. E., Wall, H. S.: Continued fractions with absolutely convergent even and odd parts. Trans. Amer. Math. Soc 67, 368–380 (1949)
13. Leighton, W., Thron, W. J.: Continued fractions with complex elements. Duke Math. J. 9, 763–772 (1942)
14. Leighton, W., Thron, W. J.: On value regions of continued fractions. Bull. Amer. Math. Soc. 48, 917–920 (1942)
15. Lorentzen, L., Waadeland, H.: Continued fractions. Vol. 1. Second edition, Atlantis Studies in Mathematics for Engineering and Science, 1, Atlantis Press, Paris (2008)
16. Nicholls, P. J.: The ergodic theory of discrete groups. London Mathematical Society Lecture Note Series, 143, Cambridge University Press (1989)
17. Paydon, J. F., Wall, H. S.: The continued fraction as a sequence of linear transformations. Duke Math. J. 9, 360–372 (1942)
18. Ratcliffe, J. G.: Foundations of hyperbolic manifolds. Graduate Texts in Mathematics, 149, Springer-Verlag (1994)
19. Scott, W. T., Wall, H. S.: A convergence theorem for continued fractions. Trans. Amer. Math. Soc 47, 155–172 (1940)
20. Short, I.: Hausdorff dimension of sets of divergence arising from continued fractions. Proc. Amer. Math. Soc. 140, 1371–1385 (2012)
21. Thron, W. J.: Two families of twin convergence regions for continued fractions. Duke Math. J. 10, 677–685 (1943)