A PIECEWISE CONTRACTIVE DYNAMICAL SYSTEM AND ELECTION METHODS

SVANTE JANSON AND ANDERS ÖBERG

Abstract. We prove some basic results for a dynamical system given by a piecewise linear and contractive map on the unit interval that takes two possible values at a point of discontinuity. We prove that there exists a universal limit cycle in the non-exceptional cases, and that the exceptional parameter set is very tiny in terms of gauge functions. The exceptional two-dimensional parameter is shown to have Hausdorff-dimension one. We also study the invariant sets and the limit sets; these are sometimes different and there are several cases to consider. In addition, we give a thorough investigation of the dynamics; studying the cases of rational and irrational rotation numbers separately, and we show the existence of a unique invariant measure. We apply some of our results to a combinatorial problem involving an election method suggested by Phragmén and show that the proportion of elected seats for each party converges to a limit, which is a rational number except for a very small exceptional set of parameters. This is in contrast to a related election method suggested by Thiele, which we study at the end of this paper, for which the limit can be irrational also in typical cases and hence there is no typical ultimate periodicity as in the case of Phragmén's method.

Contents

Date: 19 September, 2017.
2010 *Mathematics Subject Classification.* 37E05, 91B12, 28A78.

1. Introduction

The purpose of this paper is to study the dynamical system $f_\pm : [0,1] \to [0,1]$ given by the multi-valued function $x \mapsto \{f_-(x), f_+(x)\}$, where

$$f_-(x) = \{ax + b\}, \tag{1.1}$$

where a and b are given constants with $0 < a < 1$ and $0 \leqslant b < 1$, $\{\cdot\}$ denotes the usual fractional part taking values in $[0,1)$, and where $f_+(x)$ takes the value 1 instead of 0 for x such that $ax+b$ is an integer, but otherwise equals $f_-(x)$. We write $f_+(x) = \{ax + b\}_+$.

The dynamical system given by $f_- : [0,1) \to [0,1)$ has been studied from time to time and looks deceptively simple; it is locally contractive, but it has (typically) a discontinuity which makes the behaviour non-trivial. It has been studied in a variety of contexts, see, e.g., [34], [17], [5], [6], [3], and [9]; furthermore, it is a special case of more general locally contractive dynamical systems in one or several dimensions studied in [4] and [8]. The recent works by Nogueira and Pires [24], Nogueira, Pires and Rosales [25], and, especially, that of Laurent and Nogueira [21], are close to our investigation.

We study the dynamical system given by the multi-valued function $f_\pm$ instead of just f_-, both in order to obtain complete (and symmetric) results concerning the invariant set and the limit set, and because we need $f_\pm$ for our application to an election method in Section 11. The study of the dynamics given by $f_\pm$ becomes somewhat more complicated than for f_-, for example when studying the possible orbits, but we are rewarded by clear and useful results; see for example the results in Sections 8 and 9.

Earlier studies of f_-, show that (ignoring a few complications that disappear when considering $f_\pm$) the limit set may be either a periodic orbit or a Cantor set, and that these cases correspond to rational and irrational rotation numbers. These results are easily extended to $f_\pm$; much of the extension is straight-forward, but we also add some details and special features for $f_\pm$ that make the picture more complete.

In Sections 2 and 3 we make a preliminary investigation of the invariant set $\Lambda_\pm := \bigcap_{n=0}^{\infty} f_\pm^n([0,1])$ and the limit set $\omega_{f_\pm}(x)$ of $f_\pm$ for $x \in [0,1]$. (See Section 2 for the definition of the limit set in this context.) We also show that if there exists a periodic orbit, then it is a universal limit cycle in the sense that every orbit converges to it. In particular, there is at most one periodic orbit. We further give examples when $\omega_{f_\pm}(x) \subsetneq \Lambda_\pm$ for all $x \in [0,1]$, and show that even if $f_\pm$ has a universal limit cycle, the invariant set may be different from it, in analogy with the higher dimensional case, see [8].

In Section 4 we study all possibilities for orbits of $f_\pm$, with different cases depending on whether a periodic orbit exists or not, and also on whether the periodic orbit (if it exists) contains the point of discontinuity (the point of two values) of $f_\pm$ or not.

Next, building on the work by Bugeaud [5], Bugeaud and Conze [6], and Coutinho [9], we study in Sections 5 and 6 the rotation number of $f_\pm$, with special attention to whether the rotation number is rational or irrational. Furthermore, we show in Section 5 that every orbit has a well-defined average, and that this is related to the rotation number. In Section 6 we identify the set of parameters (a, b) that gives rise to a certain rotation number.

As shown by Bugeaud [5] and Bugeaud and Conze [6], the rotation number of this dynamical system is typically rational; the exceptional set of parameters (a, b) such that the rotation number is irrational has Lebesgue measure 0, and Laurent and Nogueira [21] showed, furthermore, that the set of exceptional b for a fixed a has Hausdorff dimension 0. We improve this result on Hausdorff dimension somewhat in Section 7, in that we specify a gauge function, $h(t) = 1/|\log t|^2$, for which the Hausdorff measure of the exceptional parameter set is finite. We also give a lower bound showing that this exceptional set is not arbitrarily tiny, by showing that the Hausdorff measure is positive for the gauge function $h(t) = 1/|\log t|$. We further prove that the exceptional set of parameter pairs (a, b) (a subset of $[0,1)^2$) has Hausdorff dimension 1. We prove in Section 7 also that the Hausdorff dimension of the invariant set $\Lambda_\pm$ is zero and that its Hausdorff measure is finite for the gauge function $h(t) = 1/|\log t|$. We leave it as an open question whether this gauge function is best possible in some sense.

In Section 8 we prove that the dynamical system given by $f_\pm$ has a rational rotation number if and only if it has a universal limit cycle. In Section 9, we study the case of an irrational rotation number and classify the limit sets for f_-, f_+ and $f_\pm$; we prove in particular that the limit set $\omega_{f_\pm}(x)$ (then a Cantor set) is equal to the invariant set $\Lambda_\pm$ for all $x \in [0,1]$.

In Section 10, we show that the dynamical system $F_\pm$ has a unique invariant measure with support in the invariant set. Furthermore, the empirical measure of any orbit converges to this invariant measure.

The dynamical system we consider, or rather the one given by f_-, has been studied in several applications, of which we here only mention a couple of interesting ones: the work by Feely and Chua [14] in signal theory, which inspired [5] and [6], and the paper by Coutinho *et al.* [10] studying genetic regulatory networks.

Two election methods. We also have an application in mind, and this was our original motivation for the present work. We wanted to understand a curious behaviour recently found by Mora and Oliver [23] of an election method that was suggested in 1894 by the Swedish mathematician Edvard Phragmén [26].

As a background, consider election methods where a given number $n \geqslant 1$ of persons are to be elected from some list of candidates without any formal parties, and each voter votes for set of candidates (without ranking), where the set may be chosen arbitrarily (except that possibly its size is restricted). One such method is simple *plurality*, where the n persons with the largest number of votes are elected. (In this case, usually each voter is restricted to vote for at most n candidates; this system is also called *block vote*. The version where a voter may vote for any number of candidates is called *appoval voting*.) This method has been widely used, and it is still widely used in e.g. associations and societies without (formal or informal) parties. However, for general elections with political parties, it will typically lead to the largest party getting all seats; hence this method has for such purposes in most places been replaced by other methods that tend to give representation also to smaller parties, for example proportional methods based on parties with separate lists such as *D'Hondt's method* [11, 12] or *Sainte-Laguë's method* [31]. (Many different election methods are and have been used, or proposed; see e.g. [1], [16] and [30] for discussions of several important ones, including also practical and political aspects.)

Another way to achieve some kind of proportionality is to keep the system above, where each voter votes with a ballot containing an arbitrary set of candidates, but elect the n persons sequentially and reduce the voting power of the ballots where some candidates already have been elected. Two different such systems were proposed in 1894 and 1895 by the Swedish mathematican Edvard Phragmén (1863–1937) [26, 27] and the Danish astronomer and mathematician Thorvald Nicolai Thiele (1838–1910) [33], respectively; see also [28, 29] and [20]. Both methods can be seen as generalizations of D'Hondt's method to a situation without formal parties [27, 20]. (To be precise, in a situation with organized parties, if every voter votes for a party list, then both methods yield the same result as D'Hondt's method. This is easy to see from the descriptions in Sections 11.2 and 12.1.)

We describe Phragmén's and Thiele's methods in Sections 11 and 12; see [20] for further discussion.

A party version. Mora and Oliver [23] recently considered an extension of Phragmén's method, where the individual candidates are replaced by (disjoint) groups of candidates; these groups are called *candidatures* in [23], but we shall call them *parties*. Mathematically, the difference is that a party may get several members elected; the seats are allocated to the parties one by one as in the original method, but we allow repetitions so a party may be selected several times. We assume in this paper (unlike [23]) that the parties are sufficiently large (with potentially infinite lists of candidates) so that they do not run out of persons to fill their seats. We also consider the same extension of Thiele's method.

Remark 1.1. We have presented the party version as an extension of the original method, but it can also be considered as a special case. Consider the original method with individual candidates and assume that there are parties consisting of disjoint sets of candidates that are regarded as equivalent by all voters (and by us), so that each voter votes for either all candidates from a party or for none of them, for each party. In other words, each voter votes for the union of some set of parties. It is then easy to see, for both Phragmén's and Thiele's methods, that the result is the same for the party version and for the original version (with the party representatives chosen e.g. by lot, since all from the same party will tie each time).

We consider an election using the party version of either Phragmén's or Thiele's method, with some set of parties and some set of votes (where each vote thus is for one or several parties). We let $n \geqslant 1$ seats be distributed in the election, and let n_i be the number of seats given to a party i and $p_{in} := n_i/n$ the corresponding proportion of seats. Our main interest is in the asymptotics of these proportions as the number n of elected seats tends to infinity, for a fixed set of votes. (This makes sense for the party version, but not for the original version.) In the case when each voter votes for exactly one party, both methods reduce to D'Hondt's method, as said above, and it is well-known and easy to see that the proportion p_{in} of elected seats for a party then converges to the proportion of votes for that party. (For more precise results, see [19].)

Mora and Oliver [23, Section 7.7] studied in particular the party version of Phragmén's method in the case with only two parties, A and B, and found numerically that the proportions n_A/n and $n_B/n = 1-n_A/n$ of elected seats for each party do converge; however, the limit has an unexpected singular 'Devil's staircase' structure as a function of the proportions of votes for different ballots: it seemed that the limit is always a rational number and that each rational number in $(0, 1)$ is the limit for some range of the vote proportions. We show that this is indeed the case, with the modification that irrational limits exist but only for a null set of the parameters, by interpreting the party version of Phragmén's method as a dynamical system, which in the case of two parties can be transformed to a dynamical system of the type considered in the present paper. This leads to the following theorem, which is one of our main results. The proof is given in Section 11. Recall that in the present context each vote is either for party A, party B or the set $\{A, B\}$, which we denote by AB.

Theorem 1.2. *Consider the party version of Phragmén's election method, with two parties A and B, and let the proportions of votes on A, B and AB be α, β and $\zeta = 1 - \alpha - \beta$, respectively, and assume that $\alpha + \beta > 0$. Let n_A and n_B be the numbers of seats given to the two parties when n seats have been distributed; then the fractions n_A/n and n_B/n of seats given to the two parties converge to some limits p_A and $p_B = 1-p_A$, respectively, as $n \to \infty$. Furthermore, the following holds.*

(i) *$n_A = p_A n + O(1)$ and $n_B = p_B n + O(1)$.*

(ii) *If $\alpha \geqslant \beta > 0$, then*

$$p_B = \frac{1}{2 + b_0 + \rho}, \tag{1.2}$$

where ρ is the rotation number of the dynamical system

$$f_\pm(x) = \{\{ax + b\}, \{ax + b\}_+\} \tag{1.3}$$

and we define

$$a := \frac{\alpha\beta}{(\alpha+\zeta)(\beta+\zeta)} = \frac{\alpha\beta}{(1-\alpha)(1-\beta)} \in (0,1], \tag{1.4}$$

$$b^* := \frac{\alpha-\beta}{\beta} + \frac{\alpha(1-\alpha-\beta)}{(1-\alpha)(1-\beta)}, \tag{1.5}$$

$$b := \{b^*\}, \tag{1.6}$$

$$b_0 := \lfloor b^* \rfloor. \tag{1.7}$$

We have $a < 1 \iff \zeta > 0$.

(iii) *If the rotation number ρ is rational, and furthermore $\zeta > 0$, then the sequence of awarded seats is eventually periodic.*

Furthermore, (1.2) can be combined with Theorem 6.5 or Theorems 7.1–7.2, which all imply that the rotation number, and thus p_B, is rational for almost all values of the parameters α, β, and that each rational number in $(0,1)$ is attained for some set of (α, β) with a non-empty interior, verifying the observed Devil's staircase behaviour. The reader can compare [6, Figure 1] and [23, Figura 2], which show this phenomenon from two different points of view, connected by our Theorem 1.2.

Remark 1.3. In particular, as shown by Laurent and Nogueira [21], see Theorem 8.6 below, the rotation number is rational whenever a and b are rational (or even algebraic) numbers; hence Theorem 1.2 shows that p_B is rational whenever α and β are rational (or algebraic), which explains why only rational limits were observed in [23]. See further Theorem 11.5.

Problem 1.4. Consider the party version of Phragmén's method in a case with $N \geqslant 3$ parties, and given numbers of votes. Will the proportions of seats n_i/n given to the different parties converge as $n \to \infty$? What are the limits?

In Section 12, we consider instead the party version of Thiele's method (with an arbitrary number of parties), and obtain very different results. We show that, under weak hypoteses, the proportions of seats for each party converge as $n \to \infty$ for Thiele's method too, but now each limit is a smooth function of the vote proportions; moreover, the limits can be irrational numbers also in simple cases with integer numbers of votes. We do not know whether there is a quasi-periodic behaviour in this case. In any case, we find this difference between the two election methods interesting.

Remark 1.5. Phragmén's and Thiele's methods were devised for a situation without a completely developed party system, and for small constituencies. Here, in contrast, we study the methods in the opposite situation with well-organized parties and a very large number of seats. The results are therefore not directly relevant for the original situation, and our investigation is mainly for mathematical curiosity; nevertheless, the results might give insight into some aspects of the methods.

For small numbers of seats, Thiele's method sometimes yields undesirable results, while Phragmén's method seems more robust, as discussed with many examples in the 1913 report of the Swedish Royal Commission on the Proportional Election Method [15], see also [20]. For very large numbers of seats, our result indicate the opposite, with a smoother behaviour of Theiele's method.

Historical note. Thiele's method was used in Swedish parliamentary elections 1909–1920 for the distribution of seats within parties (in combination with a special rule); it was in 1921 replaced by an ordered version of Phragmén's method. This version of Phragmén's method is still formally used but nowadays in combination with a system of personal votes and in reality the method has a very minor role. See further [20, Appendix D].

Acknowledgements. First of all, we would like to thank Mark Pollicott for helping us with this project. We are also grateful to Arnaldo Nogueira and Jean-Pierre Conze for valuable guidance, and to Anders Johansson for many valuable discussions. The first author was supported in part by the Knut and Alice Wallenberg Foundation.

2. Notation and some basic properties

We assume throughout that a and b are given constants with $0 < a < 1$ and $0 \leqslant b < 1$. (See Remark 2.2 for other parameter values.)

We let, as usual, $\lfloor x \rfloor$ and $\{x\}$ denote the integer and fractional parts of a real number x; thus $\lfloor x \rfloor \in \mathbb{Z}$ and $\{x\} := x - \lfloor x \rfloor \in [0,1)$. Furthermore, $\lceil x \rceil := -\lfloor -x \rfloor$ is the smallest integer $\geqslant x$. We further define $\{x\}_+$ as the left-continuous version of $\{x\}$; thus, when $x \in \mathbb{R} \setminus \mathbb{Z}$, then $\{x\}_+ = \{x\} \in (0,1)$, but if $x \in \mathbb{Z}$, then $\{x\} = 0$ and $\{x\}_+ = 1$. (Equivalently, $\{x\}_+ := 1 - \{-x\}$.)

For a function f defined on (a subset of) $\mathbb{R}$, let $f(x-) := \lim_{y \nearrow x} f(y)$ and $f(x+) := \lim_{y \searrow x} f(y)$, when the limits exist.

The Lebesgue measure of a set $E \subseteq \mathbb{R}$ is denoted $|E|$.

2.1. **The basic functions.** Let us first dismiss a trivial case.

Example 2.1. Suppose that $a + b < 1$. Then (1.1) is $f_-(x) = ax + b$ for all $x \in [0,1]$. This is a linear contraction, and trivially $f_-^n(x) \to p_0$ as $n \to \infty$ for every x, where $p_0 := b/(1-a) \in [0,1)$ is the (unique) fixed point of f_-.

If $b > 0$, then $f_+ = f_-$, and thus $f_\pm(x)^n \to p_0$ as $n \to \infty$, for every x. We return to the case $b = 0$ in Example 2.5 below.

In the sequel we thus focus on the case $a + b \geqslant 1$.

Let $\tau \in [0,1]$ be the point of discontinuity of $\{ax + b\}$ in [0,1], if any. Thus, if $a + b \geqslant 1$ (our main case), then $\tau = (1-b)/a$ is the solution of $ax + b = 1$; note that in this case $\tau \in (0,1]$. In the exceptional case $b = 0$, we have $\tau = 0$, and in the trivial case $a + b < 1$ with $b > 0$ (see Example 2.1), τ does not exist.

As said in the introduction, we allow an ambiguity at the discontinuity point τ, and we thus define two versions of (1.1), both for $x \in [0,1]$:

$$f_-(x) := \{ax+b\} = ax+b-\lfloor ax+b \rfloor, \tag{2.1}$$

$$f_+(x) := \{ax+b\}_+ = ax+b-(\lceil ax+b \rceil - 1). \tag{2.2}$$

Thus, explicitly, in the case $a+b \geqslant 1$, when $\tau > 0$,

$$f_-(x) = \begin{cases} ax+b, & 0 \leqslant x < \tau; \\ ax+b-1, & \tau \leqslant x \leqslant 1; \end{cases} \tag{2.3}$$

$$f_+(x) = \begin{cases} ax+b, & 0 \leqslant x \leqslant \tau; \\ ax+b-1, & \tau < x \leqslant 1. \end{cases} \tag{2.4}$$

If $\tau = b = 0$, then (2.3)–(2.4) are modified by replacing b by 1. In the trivial case when τ does not exist, $f_-(x) = f_+(x) = ax+b$ for all $x \in [0,1]$.

Note that $f_-(x) = f_+(x)$ except at the discontinuity $x = \tau$, where $f_-(\tau) = 0$ and $f_+(\tau) = 1$. Note also that f_- is right-continuous on $[0,1]$ and f_+ is left-continuous. Furthermore, $f_- : [0,1] \to [0,1)$ and $f_+ : [0,1] \to (0,1]$.

Finally, let $f_\pm(x)$ denote the multi-valued function $x \mapsto \{f_-(x), f_+(x)\}$. Formally, this is a set-valued function, but we usually regard it as a function $[0,1] \to [0,1]$ that is indeterminate at τ, where we can choose freely between $f(\tau) = 0$ and $f(\tau) = 1$; for $x \in [0,1] \setminus \{\tau\}$, $f_\pm(x)$ is a unique single value in $[0,1]$.

Note that $f_\pm$ is injective but not surjective, and that it has a continuous single-valued inverse $f_\pm^{-1} : [0, a+b-1] \cup [b,1] \to [0,1]$ (when $a+b > 1$).

Remark 2.2. We thus assume $0 < a < 1$ and $0 \leqslant b < 1$. The assumption $0 \leqslant b < 1$ is without loss of generality, since only the fractional part of b matters. However, it is also possible to consider other values of a. The main reason for our assumption $0 < a < 1$ is that we want the dynamical system to be locally contractive, which rules out $|a| \geqslant 1$.

The case $-1 < a < 0$ is locally contractive but decreasing instead of increasing; this seems to be another interesting case, and we expect results similar to the ones in the present paper, but this case will not be studied here.

Note also that the limiting cases $a = 0$ and $a = 1$ are trivial: when $a = 0$, f is constant, and when $a = 1$, $f_-(x) = \{x+b\}$ is just a translation (rotation) on the circle group $\mathbb{R}/\mathbb{Z}$.

Remark 2.3. The reflection $\sigma(x) := 1-x$ maps the dynamical system to another one of the same kind. More precisely, indicating the parameters a, b by subscripts, if we reflect the left-continuous $f_{a,b;+}$ we obtain the right-continuous

$$\begin{aligned} \sigma \circ f_{a,b,+} \circ \sigma(x) &= 1 - f_{a,b;+}(1-x) = 1 - \{a - ax + b\}_+ \\ &= \{-(a-ax+b)\} = \{ax - (a+b)\} = f_{a,\tilde{b};-}(x), \end{aligned} \tag{2.5}$$

where

$$\tilde{b} := \{-(a+b)\}. \tag{2.6}$$

Similarly, the reflection of $f_{a,b,-}$ is $f_{a,\tilde{b};+}$, and consequently the reflection of $f_{a,b,\pm}$ is $f_{a,\tilde{b};\pm}$.

If $a+b>1$ (the most interesting case), (2.6) yields $\tilde{b}=2-a-b$.

2.2. **Orbits and periodic points.** For the single-valued function f_-, the *orbit* of a point $x\in[0,1]$ is, as usual, the sequence $(f_-^n(x))_{n=0}^\infty$, and similarly for f_+. For the multi-valued $f_\pm$, we say that an orbit of $x\in[0,1]$ is any sequence $(x_n)_0^\infty$ such that $x_0=x$ and $x_{n+1}\in f_\pm(x_n)$, $n\geqslant 0$. In other words, an orbit is any possible sequence obtained by repeatedly applying $f_\pm$, making arbitrary choices each time there is a choice (i.e., when the orbit visits τ).

A *periodic orbit* is an orbit $(x_n)_0^\infty$ with $x_{n+q}=x_n$ for some $q\geqslant 1$ (the *period*) and all $n\geqslant 0$; in this case we also write the orbit as $\{x_0,\dots,x_{q-1}\}$. If furthermore $x_0,\dots,x_{q-1}$ are distinct, we say that this is a *minimal periodic orbit*. Note that, also for a multi-valued function such as $f_\pm$, a non-minimal periodic orbit always can be seen as a combination of several minimal periodic orbits (identical or not, and possibly with different initial points and inserted into each other).

A periodic orbit with period 1 is the same as a fixed point.

A *periodic point* is a point x that has a periodic orbit.

We consider a few simple examples with a periodic orbit (for example a fixed point), but where the multi-valuedness of $f_\pm$ causes complications because τ is in the periodic orbit. The general case is studied in Section 3.

Example 2.4. Suppose that $a+b=1$. Then $\tau=1$, and 1 is both a fixed point of $ax+b$ and a discontinuity point, since $f_\pm(1)=\{0,1\}$. If $0\leqslant x<1$, then x has a unique orbit $(x_n)_0^\infty=(f_\pm^n(x))_0^\infty=(f_-^n(x))_0^\infty=(f_+^n(x))_0^\infty$ with, by induction, $x_n=1-a^n(1-x)$; the orbit converges to the fixed point 1, but it never reaches 1 and thus there is never any choice.

However, if we start with $x=1$, then there is one periodic orbit 1 with period 1, but there are also infinitely many other orbits, starting with 1 repeated an arbitrary number of times followed by a jump to 0; from that point the orbit follows the unique orbit starting at 0 and thus converges to 1 as said above.

Consequently, in this example, all possible orbits converge to the fixed point 1. However, note that they do not converge uniformly, since an orbit starting at 1 may reach 0 at any given later time.

Example 2.5. Suppose that $b=0$. This is a special case of Example 2.1, and $f_-(x)=ax$ which is a contraction with fixed point 0, so all orbits of f_- converge to 0.

However, in this case (unlike the case $a+b<1$ with $b>0$), Example 2.1 does not give the full story for $f_\pm$, since $f_+(0)=1$. Hence, the fixed point 0 is also the discontinuity point τ, and 0 has infinitely many orbits, the periodic orbit 0 and orbits starting 0 repeated an arbitrary number of times followed by 1 and then converging back to 0, without ever reaching it.

The situation is as in Example 2.4, with 0 and 1 interchanged; in fact, the two examples are the mirror images of each other by the reflection discussed in Remark 2.3.

Example 2.6. Consider $a=1/2$ and $b=2/3$, i.e., $f_-(x)=\{\frac{1}{2}x+\frac{2}{3}\}$. Then $\tau=2/3$. Furthermore, $f_\pm(0)=2/3$, and thus $\{0,\frac{2}{3}\}$ is a periodic orbit with period 2. But 0 and 2/3 also have an infinite number of orbits that include

$f_+(2/3) = 1$, for example $\frac{2}{3}, 1, \frac{1}{6}, \dots$. Each such orbit continues from 1 along the unique orbit of 1, which is $1, \frac{1}{6}, \frac{3}{4}, \frac{1}{24}, \frac{11}{16}, \dots$, where $x_{2n} = (2 + 2^{-2n})/3$ and $x_{2n+1} = 2^{-2n-1}/3$; hence each such orbit converges to the periodic orbit $\{0, \frac{2}{3}\}$.

2.3. **The invariant set.** If $K \subseteq [0,1]$, then

$$f_\pm(K) = f_+(K \cap [0,\tau]) \cup f_-(K \cap [\tau,1]). \tag{2.7}$$

Since f_+ is continuous on $[0,\tau]$ and f_- on $[\tau,1]$, it follows that if $K \subseteq [0,1]$ is compact, then $f_\pm(K)$ is compact.

Consequently (by induction), $f_\pm^n([0,1])$, $n \geqslant 0$, is a decreasing sequence of non-empty compact subsets of $[0,1]$, and thus

$$\Lambda_\pm := \bigcap_{n=0}^{\infty} f_\pm^n([0,1]) \tag{2.8}$$

is a non-empty compact set.

Note that $f_\pm(\Lambda_\pm) = \Lambda_\pm$ and (since $f_\pm^{-1}$ is single-valued) $f_\pm^{-1}(\Lambda_\pm) = \Lambda_\pm$. In particular, since $f_\pm(\tau) = \{0,1\}$,

$$0 \in \Lambda_\pm \iff \tau \in \Lambda_\pm \iff 1 \in \Lambda_\pm. \tag{2.9}$$

Moreover, if $0, \tau, 1 \notin \Lambda_\pm$, then $f_\pm$ is single-valued on $\Lambda_\pm$, and thus $f_\pm : \Lambda_\pm \to \Lambda_\pm$ then is a homeomorphism. (We shall see in Sections 8 and 9 that this happens only when $\Lambda_\pm$ is finite, cf. the general [8, Theorem 3.1].)

We can also define the corresponding sets for f_- and f_+:

$$\Lambda_- := \bigcap_{n=0}^{\infty} f_-^n([0,1]), \qquad \Lambda_+ := \bigcap_{n=0}^{\infty} f_+^n([0,1]). \tag{2.10}$$

However, these may be empty, as seen by the following example (and its mirror image Example 2.5); furthermore, Λ_- and Λ_+ are not always closed sets, see Theorem 9.2. Hence $f_\pm$ and (2.8) yield a more satisfactory definition. We describe the sets $\Lambda_\pm, \Lambda_-, \Lambda_+$ completely in Theorems 8.2 and 9.2.

Example 2.7. Consider again Example 2.4 with $a+b=1$. Clearly the fixed point $1 \in \Lambda_\pm$, and thus every orbit of 1 is contained in $\Lambda_\pm$; furthermore, by applying $f_\pm^{-1}$ repeatedly, it is easily seen that no further points belong to $\Lambda_\pm$. Thus $\Lambda_\pm = \{1 - a^n : n \geqslant 0\} \cup \{1\}$. It is also easily seen that $\Lambda_- = \emptyset$ and $\Lambda_+ = \{1\}$.

Remark 2.8. The invariant set is sometimes called the *attractor*, see [8] (where our definition corresponds not to Definition 2.2 but to the version given immediately afterwards; these are not always equivalent). However, in the present context, this name seems less appropriate. For example, in Example 2.7, every orbit is attracted to 1, see Example 2.4.

2.4. **The limit set.** As in the higher-dimensional case (see [8]) the invariant set $\Lambda_\pm$ for our multivalued $f_\pm$ can be quite large, and too large for some purposes, see Example 2.7 and Remark 2.8. It is convenient to introduce the notion of a *limit set* for $f_\pm$. For single-valued functions, we define the ω-limit set as in, e.g., [24] and [8]: for a single-valued function f, we say that a point p is an ω-limit point of x if there is a strictly increasing sequence of positive integers $\{n_\ell\}$ such that $\lim_{\ell\to\infty} f^{n_\ell}(x) = p$. The collection of all such limit points is the ω-limit set of x, denoted by $\omega_f(x)$. Equivalently,

$$\omega_f(x) = \bigcap_{m\geq 0} \overline{\bigcup_{k\geq m} \{f^k(x)\}}. \tag{2.11}$$

We adjust this definition for the multi-valued function $f_\pm$ with the convention that we follow a specific orbit. More precisely, for $f_\pm$, we say that p is an ω-limit point of x if there exists an orbit $(x_n)_0^\infty$ of x and a subsequence $\{n_\ell\}_{\ell=1}^\infty$ of positive integers such that $x_{n_\ell} \to p$ as $\ell \to \infty$.

Remark 2.9. The function f_- maps into $[0, 1)$, so it may be regarded as a dynamical system either $f_i : [0, 1) \to [0, 1)$ or $f_i : [0, 1] \to [0, 1]$. (The difference is of course trivial, and usually does not matter.) For definiteness, we interpret (2.11) in $[0, 1]$, so $\omega_{f_-}(x)$ is a closed subset of $[0, 1]$, defined for all $x \in [0, 1]$. The same applies to f_+.

For a specific periodic orbit $C = \{y_0, \dots, y_{k-1}\}$, we say that an orbit $(x_n)_{n=0}^\infty$ converges to C if there exists j such that $x_n - y_{j+n \bmod k} \to 0$ as $n \to \infty$. We further say that C is a *limit cycle* of x if every orbit starting at x converges to C; in this case we also say that x *is attracted to* C. If C is a limit cycle of x, then $\omega_{f_\pm}(x) = C$. Conversely, using Lemma 3.1 below, it is easy to see that if C is a periodic orbit of $f_\pm$, and $\omega_{f_\pm}(x) = C$, then C is a limit cycle of x.

We say that C is a *universal limit cycle* if it is a limit cycle for every $x \in [0, 1]$, or, equivalently, that $\omega_{f_\pm}(x) = C$ for every x. In other words, every orbit with any initial point is attracted to C.

A related notion is that $f_\pm$ is *asymptotically periodic* if $\omega_{f_\pm}(x)$ is a periodic orbit of $f_\pm$ for every $x \in [0, 1]$. As shown in Section 3 below, $f_\pm$ has at most one periodic orbit, and thus $f_\pm$ is asymptotically periodic if and only if $f_\pm$ has a universal limit cycle. (Cf. [4] and [24], where this notion is studied in situations where several periodic orbits may occur.)

It is easy to see that $\omega_{f_\pm}(x) \subseteq \Lambda_\pm$. We note that in Example 2.4 we have $\omega_{f_\pm}(x) = \{1\}$ for every x, and thus, see Example 2.7, $\omega_{f_\pm}(x) \subsetneq \Lambda_\pm$ for every x. This is also the case in the following example, which illustrates one possible situation when there is a periodic orbit, see Section 4. See also Remarks 8.3 and 9.3 where the relation between the limit sets and the invariant sets is studied further.

Example 2.10. Consider again Example 2.6 with $a = 1/2$ and $b = 2/3$. Then the ω-limit set $\omega_{f_\pm}(x) = \{0, \frac{2}{3}\}$ for every $x \in [0, 1]$, and thus the periodic orbit $\{0, \frac{2}{3}\}$ is a universal limit cycle with period 2. But $\tau = 2/3$ is mapped to 0 or 1 and this makes it impossible to get a uniform bound on the rate of convergence to the limit cycle. This phenomenon will occur for any $f_\pm$ as soon as $\tau \in \Lambda_\pm$ and is in contrast to the uniform rates for f_- and f_+ (see [3, Theorem 2.2(2)]).

Remark 2.11. Another related notion, is the *non-wandering set* of $f_\pm$, as defined in e.g. [8]. In our case, it can be shown, e.g. using Theorems 8.2 and 9.2, that the non-wandering set is equal to the ω-limit set $\omega_{f_\pm}(x)$ for all $x \in [0, 1]$. We shall therefore not consider the non-wandering set further.

2.5. **The lifts.** We define lifts $F_-, F_+ : \mathbb{R} \to \mathbb{R}$ of f_- and f_+ by

$$F_-(x) := a\{x\} + b + \lfloor x \rfloor \ = ax + b + (1-a)\lfloor x \rfloor, \tag{2.12}$$

$$F_+(x) := F_-(x-) = ax + b - (1-a)\lfloor 1 - x \rfloor. \tag{2.13}$$

Note that $F_-(x) = F_+(x)$ unless x is an integer.

We collect some standard properties that follow immediately from the definition.

Lemma 2.12 (Cf. [9, p. 15])**.** *Let $F_-, F_+ : \mathbb{R} \to \mathbb{R}$ be the lifts defined in* (2.12)–(2.13)*. Then*

(i) $F_-(x+1) = F_-(x) + 1$, $F_+(x+1) = F_+(x) + 1$.
(ii) $\pi_- \circ F_- = f_- \circ \pi_-$, *where* $\pi_- : \mathbb{R} \to [0,1)$ *is given by* $\pi_-(x) = \{x\}$*;* $\pi_+ \circ F_+ = f_+ \circ \pi_+$, *where* $\pi_+ : \mathbb{R} \to (0,1]$ *is given by* $\pi_+(x) = \{x\}_+$.
(iii) *F_- and F_+ are strictly increasing.*
(iv) *F_- and F_+ are continuous except at the integers; F_- is right-continuous and F_+ is left-continuous.*

Proof. Obvious. □

2.6. **The rotation number.** It is well-known that the dynamical system f_- has a well-defined *rotation number*, see e.g. [5], [6], [9], This is easily extended to $f_\pm$ in the following sense. We give a proof in Section 5.

Lemma 2.13. *There exists a number $\rho = \rho(f_\pm) \in [0,1)$, called the rotation number of $f_\pm$, such that, for any $x \in \mathbb{R}$, as $n \to \infty$,*

$$F_-^n(x)/n \to \rho, \qquad F_+^n(x)/n \to \rho. \tag{2.14}$$

In fact,

$$F_-^n(x) = x + \rho n + O(1), \qquad F_+^n(x) = x + \rho n + O(1), \tag{2.15}$$

uniformly in $x \in \mathbb{R}$ and $n \geqslant 0$. We have

$$a + b - 1 \leqslant \rho \leqslant b. \tag{2.16}$$

Furthermore, $\rho = 0 \iff a + b \leqslant 1$.

We also use the notation $\rho(a,b)$.

The rotation number will be important in the sequel. In particular, we shall see (in Section 8) that there exists a periodic orbit if and only if the rotation number is rational; moreover, in this case the periodic orbit is unique and is a universal limit cycle, i.e., it attracts every orbit.

2.7. **Symbolic dynamics.** In the case $a + b \geqslant 1$ (and thus $\tau > 0$), we code an orbit $(x_i)_0^\infty$ for $f_\pm$ by a symbolic sequence $(\varepsilon_i)_0^\infty$, where $\varepsilon_i \in \{0,1\}$ is defined by

$$\varepsilon_i := \begin{cases} 0 & x_i \in [0,\tau) \text{ or } (x_i = \tau \text{ and } x_{i+1} = 1), \\ 1 & x_i \in (\tau,1] \text{ or } (x_i = \tau \text{ and } x_{i+1} = 0). \end{cases} \tag{2.17}$$

See e.g. [13], [14], [9] for equivalent versions (in the single-valued case); see also [17] for deep study of symbolic dynamics in a more general situation.

By (2.3)–(2.4), we have

$$\varepsilon_i = ax_i + b - x_{i+1}. \tag{2.18}$$

For completeness, we define ε_i by (2.18) also when $a+b<1$, although this case is not very interesting: if $a+b<1$ and $b>0$, then $\varepsilon_i=0$ for all i, and if $b=0$, then $\varepsilon_i=0$ except possibly for one i, where we have $\varepsilon_i=-1$.

The proportion of 1's in the symbolic sequence converges for any orbit, and the limit equals the rotation number. This was shown for f_- by Coutinho [9]; we extend this to $f_\pm$ in the next theorem; the proof is given in Section 5.2.

Theorem 2.14. *For any orbit $(x_i)_0^\infty$ for $f_\pm$, the corresponding symbolic sequence $(\varepsilon_i)_0^\infty$ satisfies*

$$\sum_{i=0}^{n}\varepsilon_i=\rho n+O(1), \tag{2.19}$$

where ρ is the rotation number of $f_\pm$. In particular, $\sum_{i=0}^{n-1}\varepsilon_i/n\to\rho$ as $n\to\infty$.

3. Periodic points

Recall the definition of periodic points in Section 2.2.

Lemma 3.1. 0 *and* 1 *cannot both be periodic points of $f_\pm$.*

Proof. Suppose that 0 is a periodic point, and consider a minimal periodic orbit $x_0,\dots,x_{k-1}$ with $x_0=0$. Recall that $f_\pm^{-1}$ is single-valued, and $f_\pm^{-1}(0)=\tau$. Thus $x_{k-1}=\tau$. Furthermore, if $x_i=1$ for some $i\leqslant k-1$, then $i>0$ and $x_{i-1}=f_\pm^{-1}(1)=\tau=x_{k-1}$, which is impossible since this periodic orbit is minimal. Consequently, the backwards orbit $Q:=\{f_\pm^{-n}(\tau):n\geqslant 0\}=\{x_j:0\leqslant j<k\}$ contains 0 but not 1.

Similarly, if 1 is a periodic point, then Q contains 1 but not 0.

Thus these two events exclude each other. □

Note that the proof is valid also when $\tau\in\{0,1\}$, which occurs precisely in the simple cases in Examples 2.4 and 2.5, and when τ does not exist (then 0 and 1 are not in the image of $f_\pm$, and thus certainly not periodic points).

Lemma 3.2. *Suppose that $p\in[0,1]$ is a periodic point of $f_\pm$. Then p is a periodic point of f_- or f_+ (or both).*

Proof. By assumption, there exists $k\geqslant 1$ and a periodic orbit $C=\{p_0,\dots,p_{k-1}\}$ with $p_0=p$. By Lemma 3.1, 0 and 1 cannot both appear in C. If $0\notin C$, then C is a periodic orbit of f_+, and if $1\notin C$, then C is a periodic orbit of f_-. □

Theorem 3.3. *Suppose that $f_\pm$ has a periodic orbit C. Then $f_\pm$ is asymptotically periodic and C is the universal limit cycle for $f_\pm$.*

Proof. By assumption, there exists a periodic orbit $C=\{p_0,\dots,p_{q-1}\}$ of $f_\pm$.

Suppose first that 1 is not a periodic point of $f_\pm$. Then $p_i<1$ for every i, and it follows, as in the proof of Lemma 3.2, that C is a periodic orbit of f_-. We may assume that the orbit is minimal, so $p_0,\dots,p_{q-1}$ are distinct. We consider first only the action of f_-.

Let $\xi_0, \dots, \xi_{q-1}$ be $p_0, \dots, p_{q-1}$ arranged in increasing order; thus $0 \leqslant \xi_0 < \dots < \xi_{q-1} < 1$. Extend this to a doubly infinite increasing sequence $\Xi = \{\xi_n\}_{-\infty}^{\infty}$ by

$$\xi_{mq+i} := \xi_i + m, \qquad 0 \leqslant i < q, \ m \in \mathbb{Z}. \tag{3.1}$$

It follows, using Lemma 2.12, that F_- maps the set Ξ into itself. Moreover, if $0 \leqslant i < q$, then $\pi_- \circ F_-^q(p_i) = f_-^q \circ \pi_-(p_i) = f_-^q(p_i) = p_i$ and thus $F_-^q(p_i) = p_i + r_i$ for some $r_i \in \mathbb{Z}$. It follows, using Lemma 2.12 again, that $F_-^q(\Xi) = \Xi$, and thus $F_- : \Xi \to \Xi$ is onto. Since F_- is strictly increasing, it follows that there exists an integer r such that

$$F_-(\xi_n) = \xi_{n+r}, \qquad n \in \mathbb{Z}. \tag{3.2}$$

In particular, this implies that, recalling (3.1),

$$F_-^q(\xi_n) = \xi_{n+qr} = \xi_n + r, \qquad n \in \mathbb{Z}. \tag{3.3}$$

Let $I_i := (\xi_i, \xi_{i+1}]$ and $\bar{I}_i := [\xi_i, \xi_{i+1}]$, for $i \in \mathbb{Z}$. Since F_- is strictly increasing, (3.2) implies that $F_-(\bar{I}_i) \subseteq \bar{I}_{i+r}$. Moreover, if $I_i \cap \mathbb{Z} = \emptyset$, then F_- is linear (and thus continuous) on $\bar{I}_i$, and $F_-(\bar{I}_i) = \bar{I}_{i+r}$; since F_- has contraction factor a, this implies $|\bar{I}_{i+r}| = a|\bar{I}_i|$.

Suppose that none of the q intervals $I_i, I_{i+r}, \dots, I_{i+(q-1)r}$ contains an integer. Then F_- is a linear contraction $\bar{I}_{i+jr} \to \bar{I}_{i+(j+1)r}$ for each j, and in particular $|\bar{I}_{i+(j+1)r}| = a|\bar{I}_{i+jr}|$. Hence, $|\bar{I}_{i+qr}| = a^q|\bar{I}_i|$, which is a contradiction, since $\bar{I}_{i+qr} = \bar{I}_i + r$ by (3.1).

Consequently, for each i, at least one of the q intervals $I_i, I_{i+r}, \dots, I_{i+(q-1)r}$ contains an integer. Taking $i = i_0, \dots, i_0 + r - 1$ for some i_0, we see that the rq disjoint intervals I_j, $i_0 \leqslant j < i_0 + rq$, contain at least r integers. On the other hand, the union of these intervals is $(\xi_{i_0}, \xi_{i_0+rq}] = (\xi_{i_0}, \xi_{i_0} + r]$, which contains exactly r integers. It follows that for every $i \in \mathbb{Z}$, exactly one of the q intervals $I_i, I_{i+r}, \dots, I_{i+(q-1)r}$ contains an integer. (Also, no I_i contains two integers.)

Suppose that $j \in \mathbb{Z}$ is such that I_j contains an integer ℓ_j. Then F_- is linear on $I'_j := [\xi_j, \ell_j)$ and on $I''_j := [\ell_j, \xi_{j+1}]$, and maps both intervals into $\bar{I}_{j+r}$. Since there is no integer in any of $I_{j+r}, \dots, I_{j+(q-1)r}$ by the argument above, we can apply F_- repeatedly and see that F_-^m is linear on I'_j and I''_j for $1 \leqslant m \leqslant q$. In particular, F_-^q is linear on I'_j and I''_j. Since $F_-^q(\xi_j) = \xi_j + r$ and $F_-^q(\xi_{j+1}) = \xi_{j+1} + r$ by (3.3), and F_-^q has contraction factor $a^q < 1$, it follows that $F_-^q : I'_j \to I'_j + r$ and $F_-^q : I''_j \to I''_j + r$, and we can thus iterate further. Consequently, if $x \in I'_j$ then, for every $n \geqslant 0$,

$$F_-^n(x) - F_-^n(\xi_j) = a^n(x - \xi_j). \tag{3.4}$$

It follows also, for example by (2.12) and (3.4) for n and $n+1$, that $\lfloor F_-^n(x) \rfloor = \lfloor F_-^n(\xi_j) \rfloor$, and thus, using (3.4) again and Lemma 2.12(ii),

$$f_-^n(\{x\}) - f_-^n(\{\xi_j\}) = \{F_-^n(x)\} - \{F_-^n(\xi_j)\} = a^n(x - \xi_j). \tag{3.5}$$

Hence, $f_-^n(\{x\}) - f_-^n(\{\xi_j\}) \to 0$ as $n \to \infty$, and since $\{\xi_j\} = \xi_{j \bmod q} \in C$, $\{x\}$ is attracted to the periodic orbit C by f_-. Similarly, if $x \in I''_j$, then $f_-^n(\{\xi_{j+1}\}) - f_-^n(\{x\}) \to 0$ as $n \to \infty$, and again $\{x\}$ is attracted to C. We have shown that if $x \in \bar{I}_j$ and $I_j \cap \mathbb{Z} \neq \emptyset$, then $\{x\} \in [0,1)$ is attraced to C by f_-.

Now let $x \in \bar{I}_j$ with j arbitrary. Then there exists m with $0 \leqslant m < q$ such that $I_{j+mr} \cap \mathbb{Z} \neq \emptyset$. Furthermore, $F_-^m(x) \in \bar{I}_{j+mr}$, and thus the argument above applies to $F_-^m(x)$, and shows that $\{F_-^m(x)\} = f_-^m(\{x\})$ is attracted to C by f_-; consequently also $\{x\}$ is attracted to C.

This shows that every $x \in [0, 1)$ is attracted to the periodic orbit C by f_-. Moreover, $f_-(1) \in [0, 1)$, and thus it follows that 1 too is attracted to C by f_-.

It remains to show that every point is attracted to C also by $f_\pm$, i.e., even when we allow $\tau \to f_+(\tau) = 1$ instead of $\tau \to f_-(\tau) = 0$. If $\{x_n\}$ is an orbit that makes the transition $\tau \mapsto 1$ only once, then the development after this is by f_-, and thus the sequence is attracted to C. The only possible problem is thus when we make the transition $\tau \mapsto 1$ at least twice, but then 1 appears at least twice in the orbit $\{x_n\}$, and thus there is a periodic orbit containing 1, contradicting our assumption.

This completes the proof that if 1 is not a periodic point, then every orbit is attracted to C.

If 0 is not a periodic point, the same conclusion holds by mirror symmetry, see Remark 2.3, or by repeating the proof above with F_+ instead of F_-, mutatis mutandis.

Since either 0 or 1 is not a periodic point by Lemma 3.1, this completes the proof. □

Corollary 3.4. *The dynamical system $f_\pm$ has at most one periodic orbit.* □

It follows from (3.3) in the proof above that if $f_\pm$ has a periodic orbit, then the rotation number is rational (r/q in the notation above). In fact, the converse holds too; we return to this in Theorem 8.1.

4. A classification of orbits

We now clarify what the possibilities are for orbits of $f_\pm$.

If $x \in [0, 1]$ has an orbit for $f_\pm$ that does not contain τ, then there is never any choice, and this orbit is simultaneously the orbit of x for both f_- and f_+, and the unique orbit for $f_\pm$. Hence, our consideration of the multi-valued $f_\pm$ lead to complications only when x has an orbit containing τ, i.e., when x is in the countable (or finite) set $A^- := \{f_\pm^{-n}(\tau) : n \geqslant 0\}$.

Consider first the case when τ does not belong to any periodic orbit. Then no orbit can contain τ more than once; hence if x has an orbit containing τ, then τ will not appear again, which means that there are no further choices. Consequently, if $x \in A^-$, then x has exactly two orbits for $f_\pm$, one is its orbit for f_- and the other is its orbit for f_+; furthermore, both orbits agree until they reach τ, and then they follow the unique orbits of 0 and 1 (for f_-, f_+ or $f_\pm$). Hence, for the asymptotical behaviour of the orbits, it does not matter whether we consider f_-, f_+ or $f_\pm$.

On the other hand, if τ belongs to a periodic orbit C, and $x \in A^-$, then x has an infinite number of orbits for $f_\pm$: the orbit is unique until we reach τ, but then we can either continue along the periodic orbit C repeatedly for ever, or we can go around C N times, where $N = 0, 1, 2, \ldots$, and then make the other choice at τ; this brings us to either 0 or $1 \notin C$, and then we cannot come back to τ, by Lemma 3.1, so the orbit continues with the unique orbit of 0 or 1.

This leads to the following possibilities for the orbits of an arbitrary $x \in [0,1]$.

Case 1. There exists a periodic orbit C. By Theorem 3.3 (and Corollary 3.4), C is the only periodic orbit, and every orbit is asymptotic to C. We distinguish two subcases.

Case 1a. $\tau \notin C$. Then τ does not belong to any periodic orbit, and thus no orbit can contain τ more than once. Hence, starting at an arbitrary $x \in [0,1]$, either there is a unique orbit for $f_\pm$ ($x \notin A^-$), or there are two orbits ($x \in A^-$), one (the orbit for f_-) containing 0 and one (the orbit for f_+) containing 1. All orbits are asymptotic to C. Hence, $\omega_{f_\pm}(x) = C$ for every $x \in [0,1]$. Furthermore, it follows from the proof of Theorem 3.3 (see (3.5)) that the orbits converge uniformly to C, and thus $\Lambda_\pm = \Lambda_- = \Lambda_+ = C$.

Case 1b. $\tau \in C$. Then either $0 \in C$ or $1 \in C$, but not both (Lemma 3.1). Suppose that $0 \in C$. (The case $1 \in C$ is symmetric, with 0 and 1 and the indices $+$ and $-$ interchanged below.)

If $x \notin A^-$, then x has a unique orbit, which by Theorem 3.3 is asymptotic to C. If $x \in A^-$, then x has an infinite number of orbits, as described above; one follows eventually C for ever (this is the orbit for f_-), while all others eventually follow the unique orbit of 1. Each orbit is asymptotic to C, and $\omega_{f_\pm}(x) = C$ for every $x \in [0,1]$. However, for $x \in A^-$, the orbits do not converge to C uniformly. It follows easily that if O_1 is the (unique) orbit of 1, then $\Lambda_\pm = C \cup O_1$, $\Lambda_- = C$ and $\Lambda_+ = \emptyset$.

Case 2. There is no periodic orbit of $f_\pm$. As in Case 1a, any $x \in [0,1]$ has either one or two orbits. $\Lambda_\pm$ is infinite, and we shall see in Section 9 that $\omega_{f_\pm}(x) = \Lambda_\pm$ for every $x \in [0,1]$. Furthermore, the orbits converge to $\Lambda_\pm$ uniformly.

5. The rotation number

For completeness, we supply a simple proof of the existence of a rotation number in our context (Lemma 2.13), based on earlier proofs for f_-, see e.g. [6], [9]. We also, again for completeness, prove the simple consequence Theorem 2.14, that the proportion of 1's in the symbolic sequence converges to the rotation number; see again [9] for f_-. Finally, we use this to show that every orbit has an asymptotic average, which is independent of the orbit.

5.1. **Existence of the rotation number.**

Proof of Lemma 2.13. We first observe that for any $x \in \mathbb{R}$ and any $n \geqslant 0$,

$$\left|F_-^n(x) - x - F_-^n(0)\right| < 1. \tag{5.1}$$

In fact, $F_-^n(x) - x$ has period 1, so it suffices to consider $x \in [0,1)$, and then

$$F_-^n(x) - x \leqslant F_-^n(x) < F_-^n(1) = F_-^n(0) + 1 \tag{5.2}$$

and

$$F_-^n(x) - x > F_-^n(x) - 1 \geqslant F_-^n(0) - 1, \tag{5.3}$$

which verifies (5.1).

Taking $x = F_-^m(0)$ in (5.1) we obtain, for $m, n \geqslant 0$,

$$\left|F_-^{m+n}(0) - F_-^m(0) - F_-^n(0)\right| < 1. \tag{5.4}$$

Consequently,

$$F_-^{m+n}(0) + 1 \leqslant \bigl(F_-^m(0) + 1\bigr) + \bigl(F_-^n(0) + 1\bigr), \tag{5.5}$$

i.e., the sequence $F_-^n(0) + 1$ is subadditive. As is well-known, this implies the existence of the limit

$$\rho = \lim_{n\to\infty} \frac{F_-^n(0) + 1}{n} = \inf_{n\geqslant 1} \frac{F_-^n(0) + 1}{n} \geqslant -\infty. \tag{5.6}$$

We thus have $F_-^n(0)/n \to \rho$ as $n \to \infty$, and it follows from (5.1) that $F_-^n(x)/n \to \rho$ for any $x \in \mathbb{R}$.

The corresponding result for F_+ then holds too since, if $x' < x < x''$, then, by Lemma 2.12(iii) and $F_+(x) = F_-(x-)$, $F_-(x') < F_+(x) < F_-(x'')$, and thus by induction $F_-^n(x') < F_+^n(x) < F_-^n(x'')$. Hence, (2.14) holds.

Furthermore, (5.4) implies similarly that sequence $F_-^n(0) - 1$ is superadditive, and thus also

$$\rho = \lim_{n\to\infty} \frac{F_-^n(0) - 1}{n} = \sup_{n\geqslant 1} \frac{F_-^n(0) - 1}{n}. \tag{5.7}$$

By (5.6) and (5.7), $n\rho \leqslant F_-^n(0) + 1$ and $n\rho \geqslant F_-^n(0) - 1$. Consequently,

$$\rho n - 1 \leqslant F_-^n(0) \leqslant \rho n + 1, \qquad n \geqslant 0. \tag{5.8}$$

It follows from (5.8) and (5.1) that for any real x,

$$\rho n + x - 2 < F_-^n(x) < \rho n + x + 2, \qquad n \geqslant 0, \tag{5.9}$$

showing (2.15).

If $x \geqslant 0$, then by (2.12), $F_-(x) \geqslant F_-(0) = b \geqslant 0$, and thus by induction $F_-^n(0) \geqslant 0$ for all $n \geqslant 1$; hence $\rho \geqslant 0$. Similarly, (2.12) implies $F_-(x) - x = b-(1-a)\{x\} \in [b+a-1, b]$, and hence by induction $n(a+b-1) \leqslant F_-^n(0) \leqslant nb$. Consequently, $a + b - 1 \leqslant \rho \leqslant b < 1$, showing both (2.16) and $\rho \in [0, 1)$.

Finally, if $a + b \leqslant 1$, then $x \in [0, 1)$ implies by (2.12) $F_-(x) = ax + b < a + b \leqslant 1$ and thus $F_-(x) \in [0, 1)$; hence $F_-^n(0) \in [0, 1)$, and $\rho = \lim_{n\to\infty} F^n(0)/n = 0$. The converse follows by (2.16). □

5.2. **Proof of Theorem 2.14.** Suppose first that the orbit does not contain 1; then $x_{n+1} = f_-(x_n) = \{ax_n + b\}$ for $n \geqslant 0$, and it follows from (2.12) and (2.18) by induction that

$$F_-^n(x_0) = x_n + \sum_{i=0}^{n-1} \varepsilon_i. \tag{5.10}$$

Hence, $\sum_{i=0}^{n-1} \varepsilon_i = F_-^n(x_0) + O(1) = n\rho + O(1)$ by (2.15), and the result follows.

If the orbit contains only a finite number of 1's, then the result follows by considering the part of the orbit after the last 1.

Similarly, if the orbit does not contain 0, then

$$F_+^n(x_0) = x_n + \sum_{i=0}^{n-1} \varepsilon_i, \tag{5.11}$$

and the conclusion follows by (2.15). Again, this extends to any orbit with a finite number of 0's.

The only remaining case is thus an orbit that contains an infinite number of 0's and an infinite number of 1's. However, no such orbit can exist; in fact, if there were an orbit with both 0 and 1 ocurring more than once, then both 0 and 1 would be periodic points, but that is impossible by Lemma 3.1. □

5.3. **The average of an orbit.** The following theorem shows that every orbit has an average, in the sense of the limit of the average of the n first points; furthermore, this limit is independent of the orbit, and we provide an explicit formula.

Theorem 5.1. *Let $(x_n)_0^\infty$ be any orbit of $f_\pm$, with any initial point $x_0 \in [0,1]$. Then, as $n \to \infty$,*

$$\frac{1}{n}\sum_{i=0}^{n-1} x_i \to \chi := \frac{b-\rho}{1-a}. \tag{5.12}$$

Proof. Let $S_n := \sum_{i=0}^{n-1} x_i$. Then, using (2.18) and Theorem 2.14,

$$\begin{aligned} aS_n + nb &= \sum_{i=0}^{n-1}(ax_i+b) = \sum_{i=0}^{n-1}(x_{i+1}+\varepsilon_i) = \sum_{i=1}^{n} x_i + \sum_{i=0}^{n-1}\varepsilon_i \\ &= S_n + x_n - x_0 + \rho n + O(1) = S_n + n\rho + O(1). \end{aligned} \tag{5.13}$$

Consequently,

$$S_n = n\frac{b-\rho}{1-a} + O(1). \tag{5.14}$$

This implies (5.12). □

In particular, if there exists a periodic orbit $(x_n)_0^{k-1}$, then the average of the points in the orbit is χ. For an example, see Example 2.6, where $\rho = 1/2$ and $\chi = 1/3$.

For a more trivial example, suppose that there is a fixed point p_0. Then $\rho = 0$, and (5.12) implies that $p_0 = \chi = b/(1-a)$, as is immediately seen directly.

6. Location of the rotation number

The dependency of the rotation number $\rho(a,b)$ on a and b was investigated by Ding and Hemmer [13], Bugeaud [5], Bugeaud and Conze [6] and Coutinho [9]. We use and combine some of their ideas and develop them further. There are large overlaps with the results of the references just mentioned; we nevertheless give full proofs.

In this section, ρ denotes an arbitrary real number. We do not assume that ρ equals the rotation number $\rho(a,b) = \rho(f_\pm)$ unless explicitly said so; on the contrary, our aim is to let ρ vary freely in order to eventually derive conditions for the equality $\rho = \rho(f_\pm)$.

We define, following Coutinho [9], for $\rho \in \mathbb{R}$ and $x \in \mathbb{R}$,

$$\phi_\rho(x) = \phi_{\rho,a,b}(x) := \frac{b}{1-a} + (1-a)\sum_{j=0}^{\infty} a^j \lfloor x-(j+1)\rho \rfloor. \tag{6.1}$$

The sum obviously converges absolutely, so each ϕ_ρ is a function $\mathbb{R} \to \mathbb{R}$.

It follows from (6.1) that

$$\phi_\rho(x+1) = \phi_\rho(x) + 1, \qquad x \in \mathbb{R}. \tag{6.2}$$

We state some further simple properties of the function ϕ_ρ.

Lemma 6.1. *For any $\rho \in \mathbb{R}$, $\phi_\rho : \mathbb{R} \to \mathbb{R}$ has the following properties.*

(i) *ϕ_ρ is weakly increasing: if $x \leqslant y$, then $\phi_\rho(x) \leqslant \phi_\rho(y)$.*

(ii) *If ρ is irrational, then ϕ_ρ is strictly increasing, while if ρ is rational, with denominator q, then ϕ_ρ is constant on each interval $[\frac{k}{q}, \frac{k+1}{q})$.*

(iii) *The set of discontinuity points of ϕ_ρ is*

$$D_\rho := \{n + m\rho : m \in \mathbb{Z}_{>0}, n \in \mathbb{Z}\}, \tag{6.3}$$

and ϕ_ρ has a jump discontinuity at each $x \in D_\rho$. In particular, if ρ is irrational, then the set of discontinuity points is dense in $\mathbb{R}$.

(iv) *$\phi_\rho(x)$ is right-continuous.*

Proof. (i): This is clear from (6.1), because each $\lfloor x - (j+1)\rho \rfloor$ is weakly increasing and the coefficients in (6.1) are positive.

(iii): First, note that each discontinuity is a jump discontinuity by (i).

Let $y \in D_\rho$, so $y = n + m\rho$ with $m \geqslant 1$. Then, in the sum in (6.1), the term with $j = m-1$ has a positive jump at $x = y$. The sum of all other terms is a weakly increasing function of x, since each term is; hence, the sum in (6.1) has a positive jump at y, and $\phi_\rho(y) > \phi_\rho(y-)$.

Conversely, if $y \notin D_\rho$, then every term in the sum in (6.1) is continuous at $x = y$. Since the sum converges uniformly on bounded sets, it follows that $\phi_\rho(x)$ is continuous at y.

Finally, it is well known that if ρ is irrational, then the sequence $(\{m\rho\})_{m\geqslant 1}$ is dense in $[0,1)$, and thus D_ρ is dense in $\mathbb{R}$.

(ii): If $\rho = p/q$, and $x \in [\frac{k}{q}, \frac{k+1}{q})$, then $\lfloor x - (j+1)\rho \rfloor = \lfloor \frac{k}{q} - (j+1)\rho \rfloor$ for every j and thus $\phi_\rho(x) = \phi_\rho(\frac{k}{q})$.

On the other hand, if ρ is irrational and $x < y$, then there exists by (iii) a discontinuity point $z \in (x,y)$. Hence $\phi_\rho(x) \leqslant \phi_\rho(z-) < \phi_\rho(z+) \leqslant \phi_\rho(y)$.

(iv): This follows because each $\lfloor x - (j+1)\rho \rfloor$ is right-continuous, and the sum in (6.1) converges uniformly on compact intervals. ☐

In particular, it follows from (6.3) that $0 \in D_\rho$ if and only if ρ is rational, and hence

$$\begin{cases} \phi_\rho(0) > \phi_\rho(0-), & \text{if } \rho \in \mathbb{Q}, \\ \phi_\rho(0) = \phi_\rho(0-), & \text{if } \rho \notin \mathbb{Q}. \end{cases} \tag{6.4}$$

Lemma 6.2. *Suppose that*

$$\phi_\rho(0-) \leqslant 0 \leqslant \phi_\rho(0). \tag{6.5}$$

Then

(i) *If ρ is irrational, or $\phi_\rho(0-) < 0$, then, for all $x \in \mathbb{R}$,*

$$\lfloor \phi_\rho(x) \rfloor = \lfloor x \rfloor, \tag{6.6}$$

$$\{\phi_\rho(x)\} = \phi_\rho(\{x\}), \tag{6.7}$$

and

$$F_-(\phi_\rho(x)) = \phi_\rho(x+\rho), \tag{6.8}$$

$$f_-(\{\phi_\rho(x)\}) = \{\phi_\rho(x+\rho)\} = \phi_\rho(\{x+\rho\}). \tag{6.9}$$

(ii) *If ρ is irrational, or $\phi_\rho(0) > 0$, then, for all $x \in \mathbb{R}$,*

$$\lceil \phi_\rho(x-) \rceil = \lceil x \rceil, \tag{6.10}$$

and

$$F_+(\phi_\rho(x-)) = \phi_\rho((x+\rho)-), \tag{6.11}$$

$$f_+(\{\phi_\rho(x-)\}_+) = \{\phi_\rho((x+\rho)-)\}_+. \tag{6.12}$$

Note that (6.4) shows that if (6.5) holds, then at least one of (i) and (ii) applies. Furthermore, if ρ is irrational, then (6.5) holds if and only if $\phi_\rho(0) = 0$.

Proof. (i): By monotonicity and (6.5), if $x \geqslant 0$, then $\phi_\rho(x) \geqslant \phi_\rho(0) \geqslant 0$. Similarly, if $x < 0$ and $\phi_\rho(0-) < 0$, then $\phi_\rho(x) \leqslant \phi_\rho(0-) < 0$. Furthermore, if $x < 0$ and $\rho \notin \mathbb{Q}$, then $\phi_\rho(x) < \phi_\rho(x/2) \leqslant \phi_\rho(0-) \leqslant 0$ by Lemma 6.1(ii) and (6.5). Hence, in both cases, $x < 0 \implies \phi_\rho(x) < 0$, and it follows from (6.2) that $x < 1 \implies \phi_\rho(x) < 1$. Consequently, $x \in [0,1) \implies \phi_\rho(x) \in [0,1)$, which yields (6.6) and (6.7) by (6.2).

Moreover, by (2.12), (6.6) and (6.1),

$$\begin{aligned} F_-(\phi_\rho(x)) &= a\phi_\rho(x) + b + (1-a)\lfloor \phi_\rho(x) \rfloor = a\phi_\rho(x) + b + (1-a)\lfloor x \rfloor \\ &= \frac{ab}{1-a} + (1-a)\sum_{j=0}^{\infty} a^{j+1}\lfloor x-(j+1)\rho \rfloor + b + (1-a)\lfloor x \rfloor \\ &= \frac{b}{1-a} + (1-a)\sum_{k=0}^{\infty} a^k \lfloor x - k\rho \rfloor = \phi_\rho(x+\rho). \end{aligned} \tag{6.13}$$

Finally, (6.9) follows from (6.13) by Lemma 2.12 and (6.7).

(ii): In this case we similarly see that $x \in (0,1] \implies \phi_\rho(x-) \in (0,1]$, and (6.10) follows by (6.2). Then, (6.11) follows as in (6.13). (By (6.2), it suffices to consider $x \in (0,1]$.) Finally, Lemma 2.12 yields (6.12). ∎

Let, for $\rho \in \mathbb{R}$,

$$\begin{aligned} \psi(\rho) := \phi_\rho(0) &= \frac{b}{1-a} + (1-a)\sum_{j=0}^{\infty} a^j \lfloor -(j+1)\rho \rfloor \\ &= \frac{b}{1-a} - (1-a)\sum_{j=0}^{\infty} a^j \lceil (j+1)\rho \rceil. \end{aligned} \tag{6.14}$$

Lemma 6.3. (i) *$\psi(\rho)$ is left-continuous and strictly decreasing.*

(ii) *$\psi(\rho)$ is continuous at every irrational ρ and has a jump at every rational ρ.*

(iii) *The right limits are given by*

$$\psi(\rho+) = \phi_\rho(0-) = \frac{b}{1-a} - 1 - (1-a)\sum_{j=0}^{\infty} a^j \lfloor (j+1)\rho \rfloor. \tag{6.15}$$

(iv) $\psi(0) \geqslant 0$ *and* $\psi(1) < 0$*. Furthermore,* $\psi(0+) > 0 \iff a + b > 1$.

Proof. (i): The left-continuity follows from (6.14), since each $\lceil (j+1)\rho \rceil$ is left-continuous, and the sum converges uniformly on bounded domains.

That $\psi(\rho)$ is weakly decreasing follows also from (6.14). Furthermore, if $\rho_1 < \rho_2$, then there exist j such that $(j+1)(\rho_2 - \rho_1) > 1$ and then $\lceil (j+1)\rho_1 \rceil < \lceil (j+1)\rho_2 \rceil$; hence $\psi(\rho_1) > \psi(\rho_2)$. Thus ψ is strictly decreasing.

(ii): If ρ is irrational, then every $\lceil (j+1)\rho \rceil$ is continuous at ρ, and thus (6.14) implies that ψ is continuous at ρ, again using the fact that the sum converges uniformly on bounded domains.

Conversely, if ρ is rational, then $(j+1)\rho \in \mathbb{Z}$ for some j, and then $\lceil (j+1)\rho \rceil$ has a jump at ρ. (There will be infinitely many such j, but all jumps are in the same direction, so there is no cancellation.)

(iii): For any $x, \rho \in \mathbb{R}$ and $j \geqslant 0$,

$$\lim_{\rho' \searrow \rho} \lfloor x - (j+1)\rho' \rfloor = \lim_{x' \nearrow x} \lfloor x' - (j+1)\rho \rfloor. \tag{6.16}$$

Hence, (6.1) yields, using local uniform convergence of the sums again,

$$\phi_{\rho+}(x) := \lim_{\rho' \searrow \rho} \phi_{\rho'}(x) = \lim_{x' \nearrow x} \phi_\rho(x') = \phi_\rho(x-). \tag{6.17}$$

Now take $x = 0$ to obtain $\psi(\rho+) = \phi_{\rho+}(0) = \phi_\rho(0-)$. Finally, use (6.14) and $\lceil y+ \rceil = \lfloor y \rfloor + 1$.

(iv): Simple calculations using (6.14) and (6.15) yield

$$\psi(0) = \frac{b}{1-a}, \tag{6.18}$$

$$\psi(1) = \frac{b}{1-a} - \frac{1}{1-a} = -\frac{1-b}{1-a}, \tag{6.19}$$

$$\psi(0+) = \frac{b}{1-a} - 1 = \frac{a+b-1}{1-a}. \tag{6.20}$$

□

By (6.14) and (6.15), (6.5) is equivalent to

$$\psi(\rho+) \leqslant 0 \leqslant \psi(\rho). \tag{6.21}$$

Lemma 6.4. *Let* $\rho \in \mathbb{R}$*. Then* ρ *equals the rotation number* $\rho(f_\pm) = \rho(a, b)$ *of* $f_\pm$ *if and only if* (6.21) *holds (or, equivalently,* (6.5) *holds).*

Proof. Suppose first that (6.21) holds, and thus also (6.5). As noted above, then Lemma 6.2(i) or (ii) applies. If Lemma 6.2(i) applies, then (6.6) implies $|\phi_\rho(x) - x| < 1$, and thus by iterating (6.8),

$$F_-^n(\phi_\rho(0)) = \phi_\rho(n\rho) = n\rho + O(1), \qquad n \geqslant 0; \tag{6.22}$$

hence $F_-^n(\phi_\rho(0))/n \to \rho$ as $n \to \infty$, and thus the rotation number $\rho(f_\pm) = \rho$.

A similar argument works if Lemma 6.2(ii) applies.

For the converse, let

$$\bar\rho := \sup\{\rho : \psi(\rho) \geqslant 0\}. \tag{6.23}$$

Lemma 6.3 implies that $\bar\rho$ is well-defined, with $0 \leqslant \bar\rho \leqslant 1$; furthermore, the left-continuity of ψ implies $\psi(\bar\rho) \geqslant 0$, so the supremum in (6.23) is attained (and is thus a maximum). Furthermore, by (6.23), $\psi(\rho) < 0$ for $\rho > \bar\rho$, and thus $\psi(\bar\rho+) \leqslant 0$.

Hence, $\psi(\bar\rho+) \leqslant 0 \leqslant \psi(\bar\rho)$, i.e. (6.21) holds for $\rho = \bar\rho$; as shown above this implies that $\bar\rho$ equals the rotation number $\rho(f_\pm)$. Consequently, (6.21) holds when $\rho = \rho(f_\pm)$. ☐

The rotation number $\rho(f_\pm) = \rho(a, b)$ depends on a and b in a rather complicated way. Similarly, the function $\psi(\rho)$ depends on a and ρ in rather complicated ways, but its dependency on b is simple.

We define

$$b_-(a,\rho) = (1-a)^2 \sum_{j=0}^{\infty} a^j \lceil (j+1)\rho \rceil, \tag{6.24}$$

$$b_+(a,\rho) = 1 - a + (1-a)^2 \sum_{j=0}^{\infty} a^j \lfloor (j+1)\rho \rfloor. \tag{6.25}$$

Then, by (6.14) and (6.15),

$$(1-a)\psi(\rho) = b - b_-(a,\rho), \tag{6.26}$$

$$(1-a)\psi(\rho+) = b - b_+(a,\rho). \tag{6.27}$$

Note that $b_-(a,\rho) \leqslant b_+(a,\rho)$, with equality if and only if ρ is irrational, as is easily seen directly from (6.24)–(6.25), or by (6.26)–(6.27) and Lemma 6.3(ii). Furthermore, $b_-(a,\rho)$ and $b_+(a,\rho)$ are strictly increasing functions of ρ, and $b_+(a,\rho) = b_-(a,\rho+)$.

By (6.26) and (6.27),

$$\psi(\rho) \geqslant 0 \iff b \geqslant b_-(a,\rho), \tag{6.28}$$

$$\psi(\rho+) \leqslant 0 \iff b \leqslant b_+(a,\rho), \tag{6.29}$$

We can now rephrase and expand Lemma 6.4, regarding a and ρ as given and b as varying. This yields the following theorem, essentially due to Bugeaud [5] (in a different form, see Remark 6.6 below), see also Bugeaud and Conze [6] and Ding and Hemmer [13].

Theorem 6.5. *Fix $a \in (0,1)$ and $\rho \in [0,1)$. Then $0 \leqslant b_-(a,\rho) \leqslant b_+(a,\rho) < 1$. Moreover, the rotation number $\rho(a,b)$ of $f_\pm$ equals ρ if and only if*

$$b_-(a,\rho) \leqslant b \leqslant b_+(a,\rho). \tag{6.30}$$

Furthermore,

(i) *If $\rho \notin \mathbb{Q}$, then $b_-(a,\rho) = b_+(a,\rho)$. Hence there is a unique value of b such that the rotation number $\rho(a,b)$ equals ρ.*

(ii) *If $\rho \in \mathbb{Q}$, then $b_-(a,\rho) < b_+(a,\rho)$. Hence, there is an interval $I_{a,\rho} := [b_-(a,\rho), b_+(a,\rho)]$ of b that give the same rotation number ρ of $f_\pm$. If ρ has denominator q (in lowest terms), then $I_{a,\rho}$ has length*

$$|I_{a,\rho}| = b_+(a,\rho) - b_-(a,\rho) = a^{q-1}(1-a)^2/(1-a^q). \tag{6.31}$$

Proof. First, by (6.24), $b_-(a,0) = 0$ and $b_-(a,1) = 1$. Hence, $0 \leqslant \rho < 1$ implies $b_-(a,\rho) \geqslant 0$ and $b_+(a,\rho) = b_-(a,\rho+) < 1$.

By Lemma 6.4, $\rho = \rho(f_\pm)$ if and only if (6.21) holds, which by (6.28)–(6.29) is equivalent to (6.30).

We have already remarked that $b_-(a,\rho) = b_+(a,\rho)$ if and only if $\rho \notin \mathbb{Q}$. Hence it only remains to calculate $|I_{a,\rho}|$. We have, by (6.24)–(6.25),

$$b_+(a,\rho) - b_-(a,\rho) = (1-a)^2 \sum_{j=0}^{\infty} a^j \Bigl(1 + \lfloor (j+1)\rho \rfloor - \lceil (j+1)\rho \rceil\Bigr). \tag{6.32}$$

The big bracket in this sum is 0 or 1, and 1 if and only if $(j+1)\rho \in \mathbb{Z}$. If $\rho = p/q$, this happens when $j = kq - 1$ with $k \geqslant 1$; hence

$$b_+(a,\rho) - b_-(a,\rho) = (1-a)^2 \sum_{k=1}^{\infty} a^{kq-1} = (1-a)^2 \frac{a^{q-1}}{1-a^q}. \tag{6.33}$$

□

As remarked by Ding and Hemmer [13] and Bugeaud and Conze [6], it follows from [18, Theorem 309] that for any $a \in (0,1)$, the sum of the lengths $|I_{a,\rho}|$ for all rational $\rho \in [0,1)$ is, considering only p/q in lowest terms and letting φ be the Euler totient function,

$$\begin{aligned}\Bigl|\bigcup_{\rho\in\mathbb{Q}\cap[0,1)} I_{a,\rho}\Bigr| &= \sum_{\rho\in\mathbb{Q}\cap[0,1)} |I_{a,\rho}| = (1-a)^2 \sum_{p/q\in\mathbb{Q}\cap[0,1)} \frac{a^{q-1}}{1-a^q} \\ &= (1-a)^2 \sum_{q=1}^{\infty} \varphi(q) \frac{a^{q-1}}{1-a^q} = 1 \end{aligned} \tag{6.34}$$

and hence for any fixed a, the rotation number is rational for almost every $b \in [0,1)$. Furthermore, the exceptional set of b has Hausdorff dimension 0, see [21] and Theorem 7.1 below.

Remark 6.6. As simple consequences of (6.24)–(6.25), we also have

$$b_-(a,\rho) = (1-a) \sum_{j=0}^{\infty} a^j \bigl(\lceil (j+1)\rho \rceil - \lceil j\rho \rceil\bigr) \tag{6.35}$$

$$b_+(a,\rho) = (1-a)\Bigl(1 + \sum_{j=0}^{\infty} a^j \bigl(\lfloor (j+1)\rho \rfloor - \lfloor j\rho \rfloor\bigr)\Bigr). \tag{6.36}$$

This shows that $b_-(a,\rho)$ and $b_+(a,\rho)$ coincide with the functions defined (for the same purpose) by Bugeaud [5] and Bugeaud and Conze [6, 7]. In their notation, our $b_-(a,\rho)$ is written $\tau_a(\rho)$ when ρ is irrational, and $P_q^p(a)/(1 + a + \cdots + a^{q-1})$ when $\rho = p/q$ is rational; $P_q^p(a)$ is a polynomial, and these polynomials are studied further in [5, 6, 7].

Example 6.7. For $\rho = 1/2$, (6.24) yields

$$b_-(a,\tfrac12) = (1-a)^2 \sum_{k=0}^{\infty} \bigl(a^{2k} + a^{2k+1}\bigr)\bigl(k+1\bigr) = (1-a)^2 \frac{1+a}{(1-a^2)^2} = \frac{1}{1+a} \tag{6.37}$$

and then (6.31) yields

$$b_+(a,\tfrac12) = b_-(a,\tfrac12) + \frac{a(1-a)^2}{1-a^2} = \frac{1+a-a^2}{1+a}. \tag{6.38}$$

Consequently,

$$\rho(f_\pm) = \frac{1}{2} \iff \frac{1}{1+a} \leqslant b \leqslant \frac{1+a-a^2}{1+a}. \tag{6.39}$$

7. Hausdorff dimension

We use the results above to prove three theorems about the Hausdorff dimension of important sets. The first two concern the exceptional set of parameters for which the rotation number is irrational, and thus the invariant set of $f_\pm$ is a Cantor set; in the third theorem we study the invariant set itself.

As said after Theorem 6.5, Bugeaud and Conze [6] showed that for any fixed a, the exceptional set of b that yield an irrational rotation number $\rho(a,b)$ has Lebesgue measure 0; moreover, Laurent and Nogueira [21, Theorem 2] show the sharper result that this exceptional set has Hausdorff dimension 0. See also [13]. We supply gauge functions that provide even finer information, including both an upper and a lower bound on the 'size' of the exceptional set. Furthermore, we consider in Theorem 7.2 the Hausdorff dimension of the two-dimensional parameter set (a,b) that yield irrational rotation numbers.

Let $\mathcal{E}$ be the exceptional set of all $(a,b) \in (0,1)\times[0,1)$ such that $f_{\pm,a,b}$ has irrational rotation number; furthermore, for $a\in(0,1)$, let $\mathcal{E}_a$ be the set of $b\in[0,1)$ such that $(a,b)\in\mathcal{E}$.

Theorem 7.1. *For every $a \in (0,1)$, the Hausdorff dimension of $\mathcal{E}_a$ is 0. Moreover, the Hausdorff measure $\mathcal{H}_h(\mathcal{E}_a) < \infty$ for the gauge function $h(t) = 1/|\log t|^2$, but $\mathcal{H}_h(\mathcal{E}_a) > 0$ for the gauge function $h(t) = 1/|\log t|$,*

Proof. Fix $N > 1$. There are less that N^2 intervals $I_{a,p/q}$ with $q \leqslant N$. (Here and throughout the proof we consider only $I_{a,p/q}$ with $p/q \in [0,1)$ and p/q in lowest terms.) Hence, their complement $A_N := (0,1)\setminus\bigcup_{q\leqslant N} I_{a,p/q}$ is a union of at most N^2 (open) intervals. Each of these intervals has length at most, recalling (6.34),

$$\begin{aligned} |A_N| = 1 - \sum_{q\leqslant N}\sum_p \bigl|I_{a,p/q}\bigr| &= \sum_{q>N}\sum_p \bigl|I_{a,p/q}\bigr| \leqslant \sum_{q>N} q(1-a)^2\frac{a^{q-1}}{1-a^n} \\ &\leqslant (1-a)\sum_{q>N} qa^{q-1} = \bigl(N+(1-a)^{-1}\bigr)a^N. \end{aligned} \tag{7.1}$$

Since $\mathcal{E}_a \subset A_N$, it follows that, for any gauge function h

$$\mathcal{H}_h(\mathcal{E}_a) \leqslant \liminf_{N\to\infty}\bigl(N^2 h(2Na^N)\bigr). \tag{7.2}$$

Taking $h(t) = t^\alpha$, we find $\mathcal{H}_\alpha(\mathcal{E}_a) = 0$ for every $\alpha > 0$, and thus the Hausdorff dimension is 0.

Furthermore, taking $h(t) = 1/|\log t|^2$ in (7.15) we obtain $\mathcal{H}_h(\Lambda_\pm) < \infty$.

For the lower bound for the gauge function $h(t) = 1/|\log t|$, suppose that we have a covering

$$\mathcal{E}_a \subseteq \bigcup_{k=1}^\infty I_k, \tag{7.3}$$

where $I_k = [b_k', b_k''] \subseteq [0,1]$.

Let $J_k := [\rho(a,b'_k), \rho(a,b''_k)]$. Then, every irrational $\rho \in (0,1)$ equals $\rho(a,b)$ for some $b \in \mathcal{E}_a$; thus $b \in I_k$ for some k and then $\rho \in J_k$. Consequently, $\bigcup_k J_k \supseteq (0,1) \setminus \mathbb{Q}$, and taking the Lebesgue measure we obtain

$$\sum_k |J_k| \geqslant 1. \tag{7.4}$$

We shrink each J_k to $[\rho'_k, \rho''_k] \subseteq J_k$ with ρ'_k, ρ''_k irrational and $\rho''_k - \rho'_k \geqslant \frac{1}{2}|J_k|$. (Ignore J_k with $|J_k| = 0$, if any.) Then $b_-(a,\rho'_k), b_-(a,\rho''_k) \in I_k$.

Let $j_k := \lfloor (\rho''_k - \rho'_k)^{-1} \rfloor \leqslant 2|J_k|^{-1}$. Then $(j_k+1)\rho''_k \geqslant (j_k+1)\rho'_k + 1$, and thus $\lceil (j_k+1)\rho''_k \rceil \geqslant \lceil (j_k+1)\rho'_k \rceil + 1$. Hence, (6.24) implies

$$|I_k| \geqslant b_-(a,\rho''_k) - b_-(a,\rho'_k) \geqslant (1-a)^2 a^{j_k}. \tag{7.5}$$

If $|I_k| \geqslant (1-a)^4$, then (7.5) implies $a^{j_k} \leqslant (1-a)^2$, and thus by (7.5) again, $|I_k| \geqslant a^{2j_k}$ and

$$\frac{1}{\log(1/|I_k|)} \geqslant \frac{1}{2j_k \log(1/a)} \geqslant \frac{|J_k|}{4\log(1/a)}. \tag{7.6}$$

Hence, for any covering (7.3) with $\sup|I_k| \leqslant (1-a)^4$, using (7.4),

$$\sum_k \frac{1}{\log(1/|I_k|)} \geqslant \sum_k \frac{|J_k|}{4\log(1/a)} \geqslant \frac{1}{4\log(1/a)}. \tag{7.7}$$

Consequently, with the gauge function $h(t) = 1/|\log t|$ we have

$$\mathcal{H}_h(\mathcal{E}_a) \geqslant 1/(4\log(1/a)). \tag{7.8}$$

□

For each fixed $\rho \in [0,1]$, the functions $b_-(a,\rho)$ and $b_+(a,\rho)$ defined in (6.24)–(6.25) are analytic functions of $a \in (0,1)$, and by Theorem 6.5, for every irrational $\rho \in (0,1)$, the set $(a,b) \in (0,1) \times [0,1)$ such that $f_{\pm,a,b}$ has rotation number ρ is the smooth curve $\Gamma_\rho := \{(a, b_-(a,\rho)) : a \in (0,1)\}$. Hence $\mathcal{E} = \bigcup_{\rho \in (0,1)\setminus\mathbb{Q}} \Gamma_\rho$ is an uncountable union of these smooth curves. Each curve Γ_ρ obviously has Hausdorff dimension 1. We show that the same holds for their union $\mathcal{E}$.

Theorem 7.2. *The Hausdorff dimension of $\mathcal{E}$ is* 1.

Proof. We develop the argument in the proof of Theorem 7.1 further, taking into account the dependence on a.

Let $a_* \in (0,1)$ and consider only $a \in (0,a_*]$; let $\mathcal{E}_{\leqslant a_*} := E \cap \big((0,a_*] \times [0,1)\big)$. We let C denote unspecified constants that may depend on a_* (but not on N below).

Let $N > 1$, and let $Q_N := \{\frac{p}{q} \in \mathbb{Q} \cap [0,1] : 1 \leqslant q \leqslant N\}$. Order the elements of Q_N as $0 = r_1 < \dots < r_M = 1$, where $M := |Q_N| \leqslant N^2$. (This is the well-known Farey series [18].)

By Theorem 6.5, if $b_-(a,r_j) \leqslant b \leqslant b_+(a,r_j)$, then $\rho(a,b) = r_j \in \mathbb{Q}$. Hence, recalling that $b_-(a,0) = 0$ and $b_-(a,1) = 1$,

$$\mathcal{E} \subset \bigcup_{j=1}^{M-1} \big\{(a,b) \in (0,1) \times [0,1) : b_+(a,r_j) < b < b_-(a,r_{j+1})\big\}. \tag{7.9}$$

For any $a \leqslant a_*$, and any $i < M$, (7.1) shows that

$$0 < b_-(a, r_{j+1}) - b_+(a, r_j) \leqslant \bigl(N + (1-a)^{-1}\bigr)a^N \leqslant (N + C)a_*^N. \tag{7.10}$$

Let $\delta_N := N a_*^N$, $M' := \lceil a_*/\delta_N \rceil$, and $a_i := i a_*/M'$, $i = 0, \dots, M'$; thus $a_i - a_{i-1} = a_*/M' \leqslant \delta_N$. Let

$$E_{i,j} := \bigl\{(a, b) \in (a_{i-1}, a_i] \times [0, 1) : b_+(a, r_j) < b < b_-(a, r_{j+1})\bigr\}. \tag{7.11}$$

Then, by (7.9),

$$\mathcal{E}_{\leqslant a_*} \subseteq \bigcup_{\substack{1 \leqslant i \leqslant M' \\ 1 \leqslant j < M}} E_{i,j}. \tag{7.12}$$

It follows from (6.24)–(6.25) that

$$\left|\frac{\partial}{\partial a} b_-(a, \rho)\right|, \left|\frac{\partial}{\partial a} b_+(a, \rho)\right| \leqslant C, \tag{7.13}$$

uniformly for all $a \in [0, a_*]$ and $\rho \in [0, 1]$. Consequently, if $a \in (a_{i-1}, a_i]$, then $|b_-(a, \rho) - b_-(a_i, \rho)| \leqslant C\delta_N$ and $|b_+(a, \rho) - b_+(a_i, \rho)| \leqslant C\delta_N$ for every $\rho \in [0, 1]$, and it follows from (7.11) and (7.10) that every set $E_{i,j}$ has diameter at most $(N + C)a_*^N + C\delta_N \leqslant CNa_*^N$. By (7.12), $\mathcal{E}_{\leqslant a_*}$ is covered by less than $MM' \leqslant CN^2/\delta_N = CNa_*^{-N}$ such sets. Consequently, for any $\alpha > 1$,

$$\mathcal{H}_\alpha(\mathcal{E}_{\leqslant a_*}) \leqslant \liminf_{N \to \infty} CNa_*^{-N}\bigl(CNa_*^N\bigr)^\alpha = 0. \tag{7.14}$$

Finally, $\mathcal{E} = \bigcup_n \mathcal{E}_{\leqslant 1-1/n}$, and thus $\mathcal{H}_\alpha(\mathcal{E}) = 0$ for every $\alpha > 1$. □

Our final theorem on Hausdorff dimension concerns the invariant set $\Lambda_\pm$ (or, equivalently, the ω-limit set $\omega_{f_\pm}(x)$ for any $x \in [0, 1]$, see Theorem 9.2). In the case of a rational rotation number, this set is finite or countably infinite, see Theorem 8.2 below, so it has trivially Hausdorff dimension 0. We prove that the same holds also in the irrational case, and prove a sharper result using the gauge function $h(t) = 1/|\log t|$.

Theorem 7.3. *The set $\Lambda_\pm$ has Hausdorff dimension 0. Moreover, the Hausdorff measure $\mathcal{H}_h(\Lambda_\pm)$ is finite for the gauge function $h(t) = 1/|\log t|$.*

Proof. We claim that for each $n \geqslant 0$, $f_\pm^n([0, 1])$ is the union of at most $n+1$ disjoint closed intervals (possibly of length 0) of total length a^n. In fact, this is true for $n = 0$. Suppose that it holds for some n, with $f_\pm^n([0, 1]) = \bigcup_{j=1}^{n+1} I_j$, where some of the intervals I_j may be empty. Then τ belongs to at most one interval $I_k = [x_k, y_k]$, and then $f_\pm^{n+1}(I_k) = f_+([x_k, \tau]) \cup f_-([\tau, y_k])$ is the union of two disjoint closed intervals; all other intervals are mapped to single intervals. Since furthermore, $f_\pm$ is injective, and contracts measures by a, the claim follows by induction.

Hence, $\Lambda_\pm$ can for each n be covered by $n+1$ intervals of lengths a^n, and thus, since $a^n \to 0$, for any gauge function h

$$\mathcal{H}_h(\Lambda_\pm) \leqslant \liminf_{n \to \infty}\bigl((n + 1)h(a^n)\bigr). \tag{7.15}$$

Taking $h(t) = t^\alpha$, we find $\mathcal{H}_\alpha(\Lambda_\pm) = 0$ for every $\alpha > 0$, and thus the Hausdorff dimension is 0.

Furthermore, taking $h(t) = 1/|\log t|$ in (7.15) we obtain $\mathcal{H}_h(\Lambda_\pm) \leqslant 1/|\log(a)| < \infty$. □

Alternatively, we can argue as in the proof Theorem 7.1, using (9.4) below.

Unlike in Theorem 7.1, we do not know any lower bound in Theorem 7.3, in the sense of a certain Hausdorff measure being positive. We state this as an open problem.

Problem 7.4. Find a gauge function $h(t)$ such that $\mathcal{H}_h(\Lambda_\pm) > 0$, at least for some (a, b).

In particular, we do not know whether the gauge function $1/|\log t|$ is best possible in Theorem 7.3. We suspect that the answer might depend on the parameters; it seems possible that $1/|\log t|$ is best possible in Theorem 7.3 if, for example, $\rho = 1/\sqrt{2}$ or $(\sqrt{5}-1)/2$, but not if ρ is a Liouville number.

Similarly, we do not know whether the gauge functions in Theorem 7.1 are best possible.

Problem 7.5. Improve, if possible, one or both of the gauge functions $1/|\log t|^2$ and $1/|\log t|$ in Theorem 7.1.

Again, it seems possible that the answer depends on a.

8. Rational rotation number

We return to the study of orbits. We first use the results of Section 6 to show that $f_\pm$ has a periodic orbit if and only if the rotation number is rational, as claimed at the end of Section 3.

Theorem 8.1. (i) *Suppose that the rotation number $\rho = \rho(f_\pm)$ of $f_\pm$ is rational, say $\rho = p/q$ (in lowest terms). Then $f_\pm$ has a periodic orbit C of length exactly q. Furthermore, $C = \{\phi_\rho(k/q) : k = 0, \dots, q-1\}$. In particular,*

$$\min C = \phi_\rho(0) = \psi(\rho), \tag{8.1}$$

$$\max C = \phi_\rho((q-1)/q) = \phi_\rho(1-) = 1 + \psi(\rho+). \tag{8.2}$$

(ii) *Conversely, if $f_\pm$ has a periodic orbit, then the rotation number is rational. Moreover, if the periodic orbit is minimal and has length q, then $\rho(f_\pm)$ has denominator q in lowest terms.*

Proof. (i): By Lemma 6.4 and (6.21), $\psi(\rho+) \leqslant 0 \leqslant \psi(\rho)$. Define $x_k := \phi_\rho(k/q)$, $k \in \mathbb{Z}$, and note that, by (6.2),

$$x_{k+q} = \phi_\rho(k/q + 1) = x_k + 1. \tag{8.3}$$

By Lemma 6.1(iii), $x_k < x_{k+1}$. Furthermore,

$$x_0 = \phi_\rho(0) = \psi(\rho) \geqslant 0, \tag{8.4}$$

and, recalling Lemma 6.1(ii),

$$x_{q-1} = \phi_\rho((q-1)/q) = \phi_\rho(1-) = 1 + \phi_\rho(0-) = 1 + \psi(\rho+) \leqslant 1. \tag{8.5}$$

Suppose first that $\psi(\rho+) < 0$. Then, recalling (6.15), Lemma 6.2(i) applies, and (6.8) holds. Consequently, for any $k \in \mathbb{Z}$,

$$F_-(x_k) = F_-\Bigl(\phi_\rho\Bigl(\frac{k}{q}\Bigr)\Bigr) = \phi_\rho\Bigl(\frac{k}{q} + \rho\Bigr) = \phi_\rho\Bigl(\frac{k}{q} + \frac{p}{q}\Bigr) = x_{k+p}. \tag{8.6}$$

This implies, by Lemma 2.12(ii), $f_-(\{x_k\}) = \{F_-(x_k)\} = \{x_{k+p}\}$, and thus by iteration $f_-^n(\{x_k\}) = \{x_{k+np}\}$ for any $n \geqslant 0$. Taking $n = q$ we find, using

(8.3), $f_-^q(\{x_k\}) = \{x_k + p\} = \{x_k\}$, so $\{x_k\}$ lies in a periodic orbit C of f_-. Moreover, it is easy to see that

$$C = \bigl\{\{x_k\}\bigr\}_{k\in\mathbb{Z}} = \bigl\{\{x_k\}\bigr\}_{k=0}^{q-1} = \{x_k\}_{k=0}^{q-1}, \tag{8.7}$$

using the fact that $x_k \in [0,1)$ for $0 \leqslant k \leqslant q-1$ by (8.4)–(8.5). We thus have $\min C = x_0$ and $\max C = x_{q-1}$; hence (8.4)–(8.5) yield (8.1)–(8.2).

If $\psi(\rho+) = 0$, then necessarily $\psi(\rho) = \phi_\rho(0) > 0$, see (6.4). In this case, Lemma 6.2(ii) applies, and (6.11) holds. By Lemma 6.1, ϕ_ρ is constant on the interval $[\frac{k}{q}, \frac{k+1}{q})$, and thus, using (6.11),

$$\begin{aligned} F_+(x_k) &= F_+\Bigl(\phi_\rho\Bigl(\frac{k}{q}\Bigr)\Bigr) = F_+\Bigl(\phi_\rho\Bigl(\frac{k+1}{q}-\Bigr)\Bigr) = \phi_\rho\Bigl(\Bigl(\frac{k+1}{q}+\rho\Bigr)-\Bigr) \\ &= \phi_\rho\Bigl(\frac{k+1+p}{q}-\Bigr) = \phi_\rho\Bigl(\frac{k+p}{q}\Bigr) = x_{k+p}. \end{aligned} \tag{8.8}$$

We can now repeat the arguments above, using f_+, F_+ and $\{\cdot\}_+$ instead of f_-, F_- and $\{\cdot\}$; this shows that $C = \{x_k\}_{k=0}^{q-1}$ now is a periodic orbit for f_+. Note that in the present case, $C \subset (0,1]$.

(ii): Suppose that $f_\pm$ has a periodic orbit. By Lemma 3.2, either f_- or f_+ has a periodic orbit; let us assume that f_- has one. Then, for some $x \in [0,1)$ and some $q \geqslant 1$, $f_-^q(x) = x$, which by Lemma 2.12 implies $F_-^q(x) = x+p$ for some integer p. Consequently, $F_-^{nq}(x) = x + np$ for every $n \geqslant 0$, and thus $F_-^{nq}(x)/n \to p/q$; hence the rotation number is p/q.

If q is minimal, then p and q are coprime, as a consequence of (i) and Corollary 3.4 (or by a simple direct argument which we omit). □

By Theorem 3.3, $f_\pm$ has a universal limit cycle. Combining these results, we obtain the following.

Theorem 8.2. *Suppose that $a \in (0,1)$ and $\rho \in [0,1)$ with ρ rational. Then $f_\pm$ has rotation number $\rho(f_\pm) = \rho$ if and only if one of the following three cases holds.*

(i) $b = b_-(a,\rho)$. *Then $f_\pm$ has a unique periodic orbit C, with $0 \in C$ but $1 \notin C$. C is also a periodic orbit of f_-, but f_+ has no periodic orbit.*
Furthermore, $\Lambda_- = C$, while $\Lambda_+ = \emptyset$ and $\Lambda_\pm = C \cup O_1$, where O_1 is the orbit of 1.

(ii) $b_-(a,\rho) < b < b_+(a,\rho)$. *Then $f_\pm$ has a unique periodic orbit C, with $0, 1 \notin C$. Furthermore, $\Lambda_\pm = \Lambda_+ = \Lambda_- = C$.*

(iii) $b = b_+(a,\rho)$. *As in* (i), *interchanging* 0 *and* 1 *and indices* $+$ *and* $-$.

In all three cases, every orbit of $f_\pm$ converges to C, so $\omega_{f_\pm}(x) = \omega_{f_-}(x) = \omega_{f_+}(x) = C$ for every $x \in [0,1]$.

Proof. The rotation number $\rho(f_\pm)$ equals ρ if and only if $b_-(a,\rho) \leqslant b \leqslant b_+(a,\rho)$ by Theorem 6.5. In this case, $f_\pm$ has a periodic orbit C by Theorem 8.1. Furthermore, C is unique by Corollary 3.4, and by (8.1)–(8.2) and (6.26)–(6.27), $0 \in C \iff \psi(\rho) = 0 \iff b = b_-(a,\rho)$ and $1 \in C \iff \psi(\rho+) = 0 \iff b = b_+(a,\rho)$. Hence, $\tau \in C$ if and only if $b = b_-(a,\rho)$ or $b = b_+(a,\rho)$. In other words, we are in Case 1a in Section 4 in (ii), and in Case 1b in (i) and (iii). □

Remark 8.3. Theorem 8.2 shows that if $\rho(f_\pm)$ is rational, then $\omega_{f_\pm}(x) \subseteq \Lambda_\pm$ for all x, with equality in Case (ii), but strict inclusion in (i) and (iii).

In contrast, we have $\omega_{f_-}(x) \supseteq \Lambda_-$ for all x, with equality in Cases (i) and (ii), but strict inclusion in (iii), when $\Lambda_- = \emptyset$, and similarly for f_+.

Theorem 8.4. *If the dynamical system $f_\pm$ has a rational rotation number, then $f_\pm$ has a universal limit cycle C. Thus every orbit of $f_\pm$ converges to C. Furthermore, the symbolic sequence of every orbit is eventually periodic.*

Proof. The first statement follows from Theorem 8.2, and it implies the second by definition. Thus, again by the definitions, if $(x_n)_0^\infty$ is any orbit, there exists a periodic orbit $(y_n)_0^\infty$ (started at a suitable point $y_0 \in C$) such that $x_n - y_n \to 0$ as $n \to \infty$. By (2.18) this implies, with obvious notation, $\varepsilon_n^x - \varepsilon_n^y \to 0$, and thus $\varepsilon_n^x = \varepsilon_n^y$ for all large n since $\varepsilon_n^x, \varepsilon_n^y \in \{0,1\}$. Consequently, the symbolic sequence for the orbit $(x_n)_0^\infty$ equals from some point on the symbolic sequence for $(y_n)_0^\infty$, which is periodic. $\square$

Example 8.5. By Theorem 6.5, or Lemma 6.4 and (6.18)–(6.20), the rotation number is 0 if and only if $0 \leqslant b \leqslant 1-a$, i.e., if and only if $a+b \leqslant 1$. This is the simple case studied already in Examples 2.1, 2.4 and 2.5. We see from Theorems 8.1 and 8.2, or directly as in these examples, that in this case (and only in this case) there is a fixed point, i.e., a periodic cycle of length 1, and that every orbit converges to the fixed point. The cases $b = 0$ and $b = 1-a$ discussed in Examples 2.5 and 2.4 are the cases (i) and (iii) in Theorem 8.2.

Theorem 2.14 shows that when $\rho = 0$, at most a finite number of the symbols ε_i are non-zero. In fact, it is easy to see that there can be at most one non-zero symbol.

8.1. **A sufficient condition for a rational rotation number.** By Theorems 7.1 and 7.2, or by the earlier results by Bugeaud and Conze [6] and Laurent and Nogueira [21] discussed in Section 7, the rotation number is rational for 'most' values of the parameters (a,b). Explicit examples with a rational rotation number can easily be produced using Theorem 6.5. Another large class of parameter values with a rational rotation number is given by the following theorem by Laurent and Nogueira [21, Theorem 3], which we quote for later reference; their proof is based on a number theoretic result by Loxton and van der Poorten [22, Theorem 7], combined with results by [6] (our (6.24)–(6.25) and Theorem 6.5).

Theorem 8.6 (Laurent and Nogueira [21])**.** *If a and b are algebraic numbers, then the dynamical system $f_\pm$ has a rational rotation number.* $\square$

9. Irrational rotation number

We now consider the case when $f_\pm$ has an irrational rotation number $\rho = \rho(f_\pm)$. By Theorem 8.1(ii), $f_\pm$ has no periodic orbit. Hence, this is Case 2 in Section 4; we proceed to verify the claims there.

By Lemma 6.4 and (6.4), $\phi_\rho(0) = \psi(\rho) = 0$, and thus, see (6.2), $\phi_\rho(1) = 1$. Moreover, ϕ_ρ is strictly increasing, by Lemma 6.1, and thus ϕ_ρ gives a bijection of $[0,1)$ onto $\Lambda_0 := \phi_\rho([0,1)) \subset [0,1)$.

It follows from (6.9) that $f_-(\Lambda_0) = \Lambda_0$, and that f_- restricted to Λ_0 is a bijection, which is conjugated by ϕ_ρ to the rotation $x \mapsto \{x+\rho\}$ on $[0,1)$.

By Lemma 6.1(iii), the set of discontinuities of ϕ_ρ in $[0,1]$ is

$$D_\rho \cap [0,1] = \bigl\{\{m\rho\} : m \geqslant 1\bigr\}. \tag{9.1}$$

This set is countably infinite, and dense in [0,1]; note also that $0, 1 \notin D_\rho$. Let $x_i := \{i\rho\}$, so $D_\rho \cap [0,1] = \{x_i\}_1^\infty$, and let $\xi_i := \phi_\rho(x_i-)$ and $\eta_i := \phi_\rho(x_i)$. Since ϕ_ρ is strictly increasing and right-continuous (Lemma 6.1), it follows that

$$\Lambda_0 = \phi_\rho([0,1)) = [0,1) \setminus \bigcup_{i=1}^{\infty} [\xi_i, \eta_i) \tag{9.2}$$

and

$$\overline{\Lambda_0} = [0,1] \setminus \bigcup_{i=1}^{\infty} (\xi_i, \eta_i) = \{\phi_\rho(x), \phi_\rho(x-) : x \in [0,1]\}. \tag{9.3}$$

It follows from (6.1) that the gap (ξ_i, η_i) has length

$$\eta_i - \xi_i = (1-a)a^{i-1}, \qquad i = 1, 2, \dots \tag{9.4}$$

Hence, the sum of the lengths of the gaps is 1, so $\overline{\Lambda_0}$ has Lebesgue measure 0. In fact, it has Hausdorff dimension 0, see Theorem 7.3.

Note also that (6.9) implies $f_-(\phi_\rho(1-\rho)) = 0$, and thus $\tau = \phi_\rho(1-\rho)$. In particular, $\tau \in \Lambda_0$; furthermore, $\tau \neq \eta_i$ for $i \geqslant 1$, and consequently, $\tau \notin [\xi_i, \eta_i]$. Since $f_-(\eta_i) = \eta_{i+1}$, by (6.9) again, it follows that for every $i \geqslant 1$, f_- maps $[\xi_i, \eta_i]$ linearly onto $[\xi_{i+1}, \eta_{i+1}]$; furthermore, $f_\pm = f_+ = f_-$ on each such interval. Finally, (6.9) and (6.12) (with $x = 0$) imply

$$f_\pm(0) = \eta_1 \qquad \text{and} \qquad f_\pm(1) = \xi_1. \tag{9.5}$$

This describes the dynamics of $f_\pm$ on $[0,1] \setminus \Lambda_0$ completely. It follows easily, by induction, that

$$f_-^n([0,1)) = [0,1) \setminus \bigcup_{i=1}^{n} [\xi_i, \eta_i), \tag{9.6}$$

$$f_+^n((0,1]) = (0,1] \setminus \bigcup_{i=1}^{n} (\xi_i, \eta_i], \tag{9.7}$$

$$f_\pm^n([0,1]) = [0,1] \setminus \bigcup_{i=1}^{n} (\xi_i, \eta_i). \tag{9.8}$$

Remark 9.1. As shown above, $\tau \in \Lambda_\pm$, and thus also $0, 1 \in \Lambda_\pm$ whenever $\rho(f_\pm)$ is irrational, see (2.9).

Theorem 9.2. *Suppose that $f_\pm$ has an irrational rotation number $\rho = \rho(f_\pm)$. Then*

$$\Lambda_\pm = \overline{\Lambda_0} = \{\phi_\rho(x), \phi_\rho(x-) : x \in [0,1]\}, \tag{9.9}$$

$$\Lambda_- = \Lambda_0 = \{\phi_\rho(x) : x \in [0,1)\}, \tag{9.10}$$

$$\Lambda_+ = \Lambda_1 := \{\phi_\rho(x-) : x \in (0,1]\} = \overline{\Lambda_0} \setminus \{0, \eta_1, \eta_2, \dots\}. \tag{9.11}$$

Furthermore, the limit sets $\omega_{f_\pm}(x) = \omega_{f_-}(x) = \omega_{f_+}(x) = \Lambda_\pm$ for every $x \in [0,1]$.

For any orbit $(x_n)_0^\infty$, the distance $d(x_n, \Lambda_\pm) \leqslant a^n$ for every $n \geqslant 0$; hence the orbits converge to $\Lambda_\pm$ uniformly (and geometrically).

Proof. First, (9.9)–(9.11) follow from (9.6)–(9.8) and (9.2)–(9.3).

For the limit sets, consider first f_-. Suppose first that $x \in \Lambda_0$. Then $x = \phi_\rho(t)$ for some $t \in [0,1)$, and thus $f_-^n(x) = f_-^n(\phi_\rho(t)) = \phi_\rho(\{t+n\rho\}) \in \Lambda_0$. Hence, $\omega_{f_-}(x) \subseteq \overline{\Lambda_0}$. On the other hand, for any $y = \phi_\rho(u) \in \Lambda_0$, there exists a subsequence (n_k) such that $t_{n_k} := \{t+n_k\rho\} \to u$ with $t_{n_k} \geqslant u$; since ϕ_ρ is right-continuous, this implies $f_-^{n_k}(x) \to \phi_\rho(u) = y$. Hence, $\omega_{f_-}(x) \supseteq \Lambda_0$. Since $\omega_{f_-}(x)$ is closed by (2.11), this implies $\omega_{f_-}(x) \supseteq \overline{\Lambda_0}$, and thus $\omega_{f_-}(x) = \overline{\Lambda_0} = \Lambda_\pm$.

On the other hand, if $x \in [0,1) \setminus \Lambda_0$, then $x \in [\xi_i, \eta_i)$ for some i. Since f_- is a linear contraction on each interval $[\xi_i, \eta_i]$, it follows that

$$f_-^n(\eta_i) - f_-^n(x) = a^n(\eta_i - x) \to 0 \tag{9.12}$$

as $n \to \infty$; hence the orbit of x is asymptotic to the orbit of $\eta_i \in \Lambda_0$, and thus $\omega_{f_-}(x) = \omega_{f_-}(\eta_i) = \overline{\Lambda_0} = \Lambda_\pm$ in this case too.

Finally, for $x = 1$, recall from (9.5) that $f_-(1) = \xi_1 \in [0,1)$. Thus $\omega_{f_-}(1) = \omega_{f_-}(\xi_1) = \Lambda_\pm$. Hence $\omega_{f_-}(x) = \Lambda_\pm$ for every $x \in [0,1]$.

By symmetry (Remark 2.3), also $\omega_{f_+}(x) = \Lambda_\pm$ for every $x \in [0,1]$.

The description of the orbits in the beginning of Section 4 shows that every orbit for $f_\pm$ is an orbit for f_- or for f_+. Hence, for any $x \in [0,1]$, $\omega_{f_\pm}(x) = \omega_{f_-}(x) \cup \omega_{f_+}(x) = \Lambda_\pm$.

Now, let $(x_n)_0^\infty$ be an arbitrary orbit. If $x_0 \in \Lambda_\pm$, then $x_n \in \Lambda_\pm$ for every n, and thus $d(x_n, \Lambda_\pm) = 0$. On the other hand, if $x_0 \in [0,1] \setminus \Lambda_\pm \subset [0,1) \setminus \Lambda_0$, then for every $n \geqslant 1$, (9.12) implies $d(x_n, \Lambda_\pm) \leqslant d(x_n, f_-^n(\eta_i)) \leqslant a^n$. □

Remark 9.3. In particular, if $\rho(f_\pm)$ is irrational, then, for any x, $\omega_{f_\pm}(x) = \Lambda_\pm$, while $\omega_{f_-}(x) \supsetneq \Lambda_-$ and $\omega_{f_+}(x) \supsetneq \Lambda_+$. Cf. the case of a rational rotation number in Remark 8.3.

Remark 9.4. It is easy to see that when $\rho(f_\pm)$ is irrational, $\Lambda_\pm$ is a Cantor set, i.e., a totally disconnected perfect compact set (and thus homeomorphic to the Cantor cube $\{0,1\}^\infty$). In fact, $\Lambda_\pm$ is compact and non-empty, and totally disconnected since it has measure 0 and thus does not contain any open interval. Finally, if $x \in \Lambda_\pm$, then $x \in \omega_{f_\pm}(x)$ by Theorem 9.2, so there exists an orbit (x_n) with $x_0 = x$ and a subsequence $x_{n_k} \to x$. Then each $x_n \in \Lambda_\pm$ since $\Lambda_\pm$ is invariant, and $x_n \neq x$ for $n \geqslant 1$ since there is no periodic orbit; hence x is not isolated in $\Lambda_\pm$.

Remark 9.5. When ρ is irrational, as shown above, $0, 1, \tau \in \Lambda_\pm = \omega_{f_\pm}(x)$ for any x. Hence, since each x has at most two orbits, any orbit comes arbitrarily close to the discontinuity point τ (on both sides), as well as to 0 and 1, infinitely often.

10. The invariant measure

If $\rho(f_\pm)$ is rational, so there exists a periodic orbit C by Theorem 8.1, then there is an obvious invariant probability measure μ on C, viz. the uniform measure with mass $1/|C|$ at each point. This measure μ is invariant under $f_\pm$ in the sense that if $1 \notin C$ it is invariant under f_- and if $0 \notin C$ then it is invariant under f_+; recall that at least one of these cases occurs, see Theorem 8.2.

Suppose now that $\rho(f_\pm)$ is irrational. Then we construct an invariant probability measure μ as the image measure of the Lebesgue measure on $[0,1]$ under the map ϕ_ρ, where $\rho := \rho(f_\pm)$. Then $\phi_\rho : [0,1] \to \Lambda_\pm$, see (9.9), and thus μ is a probability measure on $\Lambda_\pm$. Since ϕ_ρ is strictly increasing by Lemma 6.1, μ is in this case a continuous measure, i.e., each point has measure 0. Moreover, (6.5) holds by Lemma 6.4, so Lemma 6.2 applies, and it follows from (6.9) that μ is invariant under f_-; μ is invariant under f_+ too since μ has no point mass at τ.

Theorem 10.1. *Let $(x_i)_0^\infty$ be an arbitrary orbit of $f_\pm$. Then the empirical measure $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{x_i}$ converges weakly to the invariant μ as $n \to \infty$.*

Proof. If $\rho = \rho(f_\pm)$ is rational, this follows from the fact that the orbit converges to the limit cycle C, see Theorem 8.4.

Thus suppose that ρ is irrational. Then the orbit visits 1 at most once, and if it does, it suffices to consider the part of the orbit after 1. Hence, we may assume that $x_0 \in [0,1)$ and that $x_n = f_-^n(x_0)$.

If $x_0 \in \Lambda_0$, so $x_0 = \phi_\rho(t)$ for some $t \in [0,1)$ (see (9.10)), then (6.9) implies $x_i = \phi_\rho(\{t + i\rho\})$, and hence $\mu_n := \frac{1}{n}\sum_0^{n-1}\delta_{x_i}$ is the image under ϕ_ρ of the measure $\nu_n := \frac{1}{n}\sum_0^{n-1}\delta_{\{t+i\rho\}}$. As $n \to \infty$, the measures ν_n converge weakly to the uniform measure λ on $[0,1)$, and since ϕ_ρ is measurable and λ-a.e. continuous (by Lemma 6.1), it follows that $\mu_n \to \mu$ weakly, see [2, Theorem 5.1].

If $x_0 \in [0,1) \setminus \Lambda_0$, then there exists as in the proof of Theorem 9.2 an $\eta_i \in \Lambda_0$ such that (9.12) holds. We have just shown that the theorem holds for the orbit starting at η_i, and then (9.12) implies that the same holds for the orbit starting at x_0. □

Corollary 10.2. *The invariant measure μ has center of mass $\int_0^1 x\,\mathrm{d}\mu = \chi := (b - \rho(f_\pm))/(1-a)$.*

Proof. With μ_n as in the proof of Theorem 10.1, $\int_0^1 x\,\mathrm{d}\mu_n \to \int_0^1 x\,\mathrm{d}\mu$ by Theorem 10.1, and $\int_0^1 x\,\mathrm{d}\mu_n \to \chi$ by Theorem 5.1. □

Theorem 10.3. *The measure μ is the only probability measure on $[0,1]$ that is invariant under f_- or f_+.*

Proof. Suppose that ν is such a probability measure, invariant under, say, f_-. Let X_0 be a random point in $[0,1]$ with the distribution ν, and let $X_n := f_-^n(X_0)$. Then X_n is a sequence of random variables, each having the same distribution ν.

Let $h \in C[0,1]$ be an arbitrary continuous function on $[0,1]$. Then Theorem 10.1 shows that

$$\frac{1}{n}\sum_{i=0}^{n-1} h(X_i) \to \int h\,\mathrm{d}\mu. \tag{10.1}$$

The random variables on the left-hand side are uniformly bounded, so by dominated convergence,

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\,h(X_i)=\mathbb{E}\Bigl(\frac{1}{n}\sum_{i=0}^{n-1}h(X_i)\Bigr)\to\int h\,\mathrm{d}\mu. \tag{10.2}$$

On the other hand, each X_i has distribution ν, so $\mathbb{E}\,h(X_i)=\int h\,\mathrm{d}\nu$. Consequently, $\int h\,\mathrm{d}\nu=\int h\,\mathrm{d}\mu$, which, since h is arbitrary, means $\nu=\mu$. □

11. Phragmén's election method

11.1. **Definition of Phragmén's method.** Phragmén's election method can be described in several different, but equivalent, ways. For our purposes it is convenient to use the following, which is based on Phragmén's original formulation (in French) in [26]; see also [27, 28, 29], [20] and Section 11.2 below for different formulations and motivations.

Phragmén's election method. *Assume that each ballot has some* voting power *t; this number is the same for all ballots and will be determined later. A candidate needs total voting power* 1 *in order to be elected. The voting power of a ballot may be used by the candidates on that ballot, and it may be divided among several of the candidates on the ballot. During the procedure described below, some of the voting power of a ballot may be already assigned to already elected candidates; the remaining voting power of the ballot is free.*

The seats are distributed one by one.

For each seat, each remaining candidate may use all the free voting power of each ballot that includes the candidate. (I.e., the full voting power t except for the voting power already assigned from that ballot to candidates already elected.) The ballot voting power t that would give the candidate voting power 1 *is computed, and the candidate requiring the smallest voting power t is elected. All free (i.e., unassigned) voting power on the ballots that contain the elected candidate is assigned to that candidate, and these assignments remain fixed throughout the election.*

The computations are then repeated for the next seat for the remaining candidates (resulting in a new voting power t), and so on.

Ties are broken by lot or by some other method. The required voting power t increases for each seat, except in some cases of a tie where t may remain the same.

11.2. **An algorithmic version of Phragmén's method.** For any set σ of candidates (parties in the party version), let v_σ be the number of votes for the set σ. Hence the total number of votes for candidate (party) i is

$$W_i^0:=\sum_{\sigma\ni i}v_\sigma. \tag{11.1}$$

Phragmén's method is often formulated in the following algoritmical form, where W_i^0 is reduced to a *reduced vote* W_i when some candidates on ballots containing i already have been elected:

For each set σ with $v_\sigma>0$ (i.e., each group of identical ballots), we assign dynamically a *place number* q_σ, which is a real non-negative number that can be interpreted as the (fractional) number of seats elected so far by these ballot; the sum of the place numbers is always equal to the number of seats already allocated. The place numbers are assigned and the seats are allocated recursively by the following rules.

(i) Initially all place numbers $q_i = 0$.
(ii) The reduced vote for candidate i is defined as

$$W_i := \frac{\sum_{\sigma\ni i} v_\sigma}{1+\sum_{\sigma\ni i} q_\sigma}, \tag{11.2}$$

i.e., the total number of votes for the candidate divided by 1 + their total place number.
(iii) The candidate i with the largest W_i is elected to the next seat, breaking ties by lot or some other method. (In the original version, only unelected candidates are considered. In the party version, repetitions are allowed.)
(iv) If i is elected, then q_σ is updated for every $\sigma \ni i$ (i.e., for the ballots that contributed to the election of i); the new value is

$$q'_\sigma := \frac{v_\sigma}{W_i}. \tag{11.3}$$

q_σ remains unchanged when $\sigma \not\ni i$.
Repeat from (ii).

It is easily verified from (11.2) that (iv) increases $\sum_\sigma q_\sigma$ by 1, so by induction, $\sum_\sigma q_\sigma$ equals the number of elected, as claimed above.

For a proof that this really yields the same result as the definition in Section 11.1, see e.g. [20]; we remark here only that the connection is that the voting power t required to elect candidate i in the previous version equals $1/W_i$ with W_i given by (11.2), and that q_σ is the total voting power already assigned to previously elected on the ballots of type σ.

11.3. **Phragmén's method as a dynamical system.** Phragmén's method (in the party version) can be regarded as a dynamical system as follows.

Let $\mathcal{P}$ be the set of parties (or candidates, in the original version), and let as above v_σ be the number of votes for the set σ of parties. (We regard these numbers as fixed.) Define W_i^0 by (11.1). We may ignore parties that do not appear on any ballot, and thus we assume that $W_i^0 > 0$ for every $i \in \mathcal{P}$. Let

$$\Pi := \{\sigma \subseteq \mathcal{P} : v_\sigma > 0 \text{ and } \sigma \neq \emptyset\}, \tag{11.4}$$

the family of all nonempty sets of parties with at least one vote for the set. (I.e., the different types of ballots that occur. We ignore blank votes, i.e., $\sigma = \emptyset$, since they do not affect the outcome.)

We use the formulation of Phragmén's method in Section 11.1, and let $x_\sigma = x_\sigma(n)$ be the free voting power of each ballot σ when n candidates have been elected. Let $\mathbf{x} = \mathbf{x}(n) = (x_\sigma)_{\sigma\in\Pi}$ be the vector of free voting powers. Let $\mathbf{1} := (1)_{\sigma\in\Pi}$ be the vector with all components 1. The description in Section 11.1 now can be formalized as follows:

(i) Initialize all $x_\sigma := 0$.

(ii) A party (candidate) i can use a voting power

$$V_i(\mathbf{x}) = V_i((x_\sigma)_\sigma) := \sum_{\sigma \ni i} v_\sigma x_\sigma. \tag{11.5}$$

For each $i \in \mathcal{P}$, find $\Delta_i := \Delta_i(\mathbf{x})$ such that $V_i(\mathbf{x} + \Delta_i \mathbf{1}) = 1$, i.e.,

$$\sum_{\sigma \ni i} v_\sigma (x_\sigma + \Delta_i) = 1. \tag{11.6}$$

(iii) Find i^* such that Δ_{i^*} is minimal, i.e., $\Delta_{i^*} = \min_{i \in \mathcal{P}} \Delta_i$.
Output i^* as the next elected.

(iv) Update $\mathbf{x}$ to

$$x'_\sigma := \begin{cases} x_\sigma + \Delta_{i^*}, & i^* \notin \sigma, \\ 0, & i^* \in \sigma. \end{cases} \tag{11.7}$$

Repeat from (ii).

In the original version, candidates that are elected are not considered further, but in the party version there is no such restriction.

We can regard (ii)–(iv) as a function f, taking a vector $\mathbf{x}$ to a new vector $f(\mathbf{x}) = (x'_\sigma)_\sigma$; a natural state space is

$$K := \bigl\{\mathbf{x} = (x_\sigma)_\sigma \in [0,\infty)^\Pi : V_i(\mathbf{x}) \leqslant 1 \ \forall i \in \mathcal{P}\bigr\}. \tag{11.8}$$

If $\mathbf{x} \in K$ and $\sigma \in \Pi$, take any $i \in \sigma$; then $V_i(\mathbf{x}) \leqslant 1$ and thus $x_\sigma \leqslant 1/v_\sigma < \infty$ by (11.5). Consequently, K is closed and bounded, i.e., K is a compact subset of $\mathbb{R}^\Pi$. Note that the equation (11.6) is a linear equation in Δ_i, with positive coefficient W_i^0; thus the equation has a unique solution $\Delta_i(\mathbf{x})$. Moreover, $\Delta_i(\mathbf{x}) \geqslant 0$ for $\mathbf{x} \in K$.

Ties are possible in (iii); in that case we choose i^* by lot or by some other method. We regard the method as indeterminate in that case. We formalize this by defining, for $i \in \mathcal{P}$,

$$K_i := \bigl\{\mathbf{x} \in K : \Delta_i(\mathbf{x}) \leqslant \Delta_j(\mathbf{x}) \ \forall j \in \mathcal{P}\bigr\}, \tag{11.9}$$

i.e., the set of free voting powers where i can be chosen as i^*. Then (iv) (with $i^* = i$) defines a function $f_i : K_i \to K$, and f is the union of these functions. Note that $K = \bigcup_i K_i$, so f is defined everywhere on K, but f is multivalued at points in the intersection $K_i \cap K_j$ of two (or more) domains. (Cf. [8], where multivalued functions of this type are studied in the case when each f_i is a contraction.)

Note that the result is the same if all vote numbers v_σ are multiplied by the same positive constant. We may thus divide by the total number of votes and thus replace the numbers of votes by their proportions; we keep the notation v_σ but may thus without loss of generality assume $\sum_\sigma v_\sigma = 1$. Moreover, we allow v_σ to be arbitrary real numbers in $[0,1]$ (with sum 1). (In a real election, the proportions are of course rational numbers, but we may imagine that we have weighted votes, where voters have different weigths that are arbitrary positive real numbers.)

The general case seems quite difficult to analyse, so we consider in the sequel the case of only two parties.

Remark 11.1. The dynamical system just described is in general not locally contractive for the standard Euclidean metric on $K \subset [0,\infty)^\Pi$ (or for the

ℓ^1 or ℓ^∞ metric, say), not even for two parties; see (11.25) below for a counterexample.

11.4. **Phragmén's method for two parties.** With two parties A and B, the possible votes are A, B and AB (and blank votes, but they may be ignored as said above). For convenience, we may assume as above that v_σ is the proportion of votes on σ, and thus that they sum to 1; furthermore we change notation and denote these proportions by $\alpha := v_A$, $\beta := v_B$ and $\zeta := v_{AB} = 1 - \alpha - \beta$.

By symmetry, we may assume $\alpha \geqslant \beta \geqslant 0$. The cases $\beta = 0$ and $\alpha = \beta$ are simple, see Examples 11.2 and 11.3. We may thus assume $\alpha > \beta > 0$. We shall show that it then is possible to transform the dynamical system in Section 11.3 into the system $f_\pm = \{\{ax + b\}, \{ax + b\}_+\}$ studied above, for some a and b.

We do the transformation in several steps. First, note that we do not use all of the set K in (11.8). In fact, when A is elected we put $x_A = x_{AB} = 0$, and when B is elected we put $x_B = x_{AB} = 0$. Hence, both f_A and f_B map K into the subset, with $\mathbf{x} = (x_A, x_B, x_{AB})$,

$$K' := K \cap \big(\{(x,0,0) : x \geqslant 0\} \cup \{(0,y,0) : y \geqslant 0\}\big) \tag{11.10}$$

and thus it suffices to consider the action of f_A and f_B on K'.

There are thus two cases:

(i) Suppose that $\mathbf{x} = (x, 0, 0)$. If the voting power of each ballot is increased by Δ, then A has available voting power, cf. (11.5)–(11.6),

$$V_A(\mathbf{x}+\Delta\mathbf{1}) = v_A(x+\Delta)+v_{AB}\Delta = (\alpha+\zeta)\Delta+\alpha x = (1-\beta)\Delta+\alpha x, \tag{11.11}$$

and thus A requires additional voting power

$$\Delta_A = \frac{1-\alpha x}{1-\beta}. \tag{11.12}$$

On the other hand, B has available voting power

$$V_B(\mathbf{x} + \Delta\mathbf{1}) = v_B\Delta + v_{AB}\Delta = (\beta + \zeta)\Delta = (1 - \alpha)\Delta, \tag{11.13}$$

so B requires voting power

$$\Delta_B = \frac{1}{1-\alpha}. \tag{11.14}$$

Since $\alpha > \beta$ by assumption, $\Delta_B > 1/(1 - \beta) \geqslant \Delta_A$; hence the next seat goes to A, updating $(x, 0, 0)$ to (x', y', z') with $x' = z' = 0$ and

$$y' = \Delta_A = \frac{1-\alpha x}{1-\beta}. \tag{11.15}$$

(ii) Suppose that $\mathbf{x} = (0, y, 0)$. Arguing as above, we find that the additional voting power required for the two parties are

$$\Delta_A = \frac{1}{\alpha+\zeta} = \frac{1}{1-\beta}, \tag{11.16}$$

$$\Delta_B = \frac{1-\beta y}{\beta+\zeta} = \frac{1-\beta y}{1-\alpha}. \tag{11.17}$$

Thus, there are two subcases: (In case of equality in (11.18) and (11.21), we are in the indeterminate case when both alternatives are possible; the same applies to all transformations below.)

(a) A is elected if

$$\frac{1}{1-\beta} \leqslant \frac{1-\beta y}{1-\alpha}, \tag{11.18}$$

or, equivalently,

$$\beta y \leqslant 1 - \frac{1-\alpha}{1-\beta} = \frac{\alpha-\beta}{1-\beta}. \tag{11.19}$$

The free voting powers are updated to $(0, y', 0)$ where

$$y' := y + \Delta_A = y + \frac{1}{1-\beta}. \tag{11.20}$$

(b) B is elected if

$$\frac{1}{1-\beta} \geqslant \frac{1-\beta y}{1-\alpha}, \tag{11.21}$$

or, equivalently,

$$\beta y \geqslant 1 - \frac{1-\alpha}{1-\beta} = \frac{\alpha-\beta}{1-\beta}. \tag{11.22}$$

The free voting powers are updated to $(x', 0, 0)$ with

$$x' := \Delta_B = \frac{1-\beta y}{1-\alpha}. \tag{11.23}$$

11.4.1. *First dynamical system.* Since $x_{AB} = 0$ on K', we may ignore x_{AB} and write the elements of K' as (x_A, x_B). Phragmén's method can thus be formulated as a dynamical system, operating on vectors $(x, y) \in ([0,\infty) \times \{0\}) \cup (\{0\} \times [0,\infty))$ by the function $(x,y) \mapsto f_1(x,y)$ given by

(i) If $y = 0$, then output A and let

$$f_1(x, 0) := \Bigl(0, \frac{1-\alpha x}{1-\beta}\Bigr). \tag{11.24}$$

(iia) If $x = 0$ and $\beta y \leqslant \frac{\alpha-\beta}{1-\beta}$, then output A and let

$$f_1(0, y) := \Bigl(0, y + \frac{1}{1-\beta}\Bigr). \tag{11.25}$$

(iib) If $x = 0$ and $\beta y \geqslant \frac{\alpha-\beta}{1-\beta}$, then output B and let

$$f_1(0, y) := \Bigl(\frac{1-\beta y}{1-\alpha}, 0\Bigr). \tag{11.26}$$

The system starts in $(0,0)$, and thus begins with (i) or (iia) which both give the same result when $x = y = 0$.

11.4.2. *Second dynamical system.* We can simplify the analysis by noting that an election of B, by (11.26) always gives case (i) and thus election of A for the next seat. Let us consider these two seat assignments as a combined move. The combination thus start as in (iib) above with $\mathbf{x} = (0, y)$, where $\beta y \geqslant (\alpha-\beta)/(1-\beta)$. First B is elected, leaving by (11.26) each ballot A with a free voting power $x' = (1-\beta y)/(1-\alpha)$. Secondly, A is elected, leaving by (11.24) each ballot B with a free voting power

$$y'' = \frac{1-\alpha x'}{1-\beta} = \frac{1-\alpha-\alpha(1-\beta y)}{(1-\alpha)(1-\beta)} = \frac{1-2\alpha+\alpha\beta y}{(1-\alpha)(1-\beta)}. \tag{11.27}$$

Using this combination instead of (iib) above, each case yields a vector of the form $(0, y)$. We can thus simplify the dynamical system to the following, acting on a single variable $y \geqslant 0$ (starting with $y = 0$) by the function f_2 given by

(i) If $\beta y \geqslant \frac{\alpha-\beta}{1-\beta}$, then output BA and let

$$f_2(y) := \frac{1 - 2\alpha + \alpha\beta y}{(1-\alpha)(1-\beta)}. \tag{11.28}$$

(ii) If $\beta y \leqslant \frac{\alpha-\beta}{1-\beta}$, then output A and let

$$f_2(y) := y + \frac{1}{1-\beta}. \tag{11.29}$$

11.4.3. *Third dynamical system.* We simplify further by replacing y by $z := (1-\beta)y$, noting that

$$\beta y \geqslant \frac{\alpha-\beta}{1-\beta} \iff \beta z \geqslant \alpha - \beta \iff z \geqslant \frac{\alpha}{\beta} - 1.$$

This yields an equivalent dynamical system acting on a variable $z \geqslant 0$ (starting with $z = 0$) by the function f_3 given by

(i) If $z \geqslant \frac{\alpha}{\beta} - 1$, then output BA and let

$$f_3(z) := \frac{1-2\alpha}{1-\alpha} + \frac{\alpha\beta}{(1-\alpha)(1-\beta)} z. \tag{11.30}$$

(ii) If $z \leqslant \frac{\alpha}{\beta} - 1$, then output A and let

$$f_3(z) := z + 1. \tag{11.31}$$

11.4.4. *Fourth dynamical system.* We replace z by $w := \alpha/\beta - z$ and obtain the dynamical system (starting with $w = \alpha/\beta$) given by the function f_4 defined by:

(i) If $w \leqslant 1$, then output BA and let

$$\begin{aligned} f_4(w) &:= \frac{\alpha}{\beta} - \frac{1-2\alpha}{1-\alpha} - \frac{\alpha\beta}{(1-\alpha)(1-\beta)}\Big(\frac{\alpha}{\beta} - w\Big) \\ &= \frac{\alpha}{\beta} + \frac{\alpha}{1-\alpha} - 1 - \frac{\alpha^2}{(1-\alpha)(1-\beta)} + \frac{\alpha\beta}{(1-\alpha)(1-\beta)} w. \end{aligned} \tag{11.32}$$

(ii) If $w \geqslant 1$, then output A and let

$$f_4(w) := w - 1. \tag{11.33}$$

In other words,

$$f_4(w) = \begin{cases} aw + b^*, & w \leqslant 1, \quad (11.34a) \\ w - 1, & w \geqslant 1, \quad (11.34b) \end{cases}$$

where

$$a = \frac{\alpha\beta}{(1-\alpha)(1-\beta)} = \frac{\alpha\beta}{(\alpha+\zeta)(\beta+\zeta)} \in (0, 1], \tag{11.35}$$

$$b^* = \frac{\alpha}{\beta} + \frac{\alpha}{1-\alpha} - 1 - \frac{\alpha^2}{(1-\alpha)(1-\beta)} = \frac{\alpha-\beta}{\beta} + \frac{\alpha(1-\alpha-\beta)}{(1-\alpha)(1-\beta)} > 0. \tag{11.36}$$

Note that $a < 1$ unless $\zeta = 0$ (in which case Phragmén's method reduces to D'Hondt's, as said above). On the other hand, b^* can be arbitrarily large; we define $b := \{b^*\} \in [0, 1)$ and $b_0 := \lfloor b^* \rfloor$.

Note also that $0 < f_4(0) = b^* < f_4(1-) = a + b^*$ and that

$$a + b^* = \frac{\alpha - \beta}{\beta} + \frac{\alpha(1 - \alpha - \beta) + \alpha\beta}{(1 - \alpha)(1 - \beta)} = \frac{\alpha}{\beta(1 - \beta)} - 1. \tag{11.37}$$

11.4.5. *Final (fifth) dynamical system.* We can reformulate the dynamical system once more by combining each BA move (11.34a) with all following A moves (11.34b). This yield the dynamical system acting on $w \in [0, 1]$ by the function $f_5 : [0, 1] \to [0, 1]$ given by

$$f_5(w) := \{f_4(w)\} = \{aw + b^*\} = \{aw + b\} \tag{11.38}$$

with the output BA^k where

$$k := 1 + \lfloor f_4(w) \rfloor = 1 + \lfloor aw + b^* \rfloor = 1 + b_0 + \lfloor aw + b \rfloor, \tag{11.39}$$

except that in the indeterminate case when $aw + b^*$ is an integer, we also allow $f_5(w) = \{aw + b\}_+ = 1$ with $k := aw + b^*$.

Thus $f_5(w) = f_\pm(w)$, the multi-valued function studied in the present paper, with a and $b := \{b^*\}$ given by (11.35)–(11.36). Furthermore, (11.39) can be written, using (11.38) and defining the symbol $\varepsilon \in \{0, 1\}$ as in (2.18),

$$k := 1 + b_0 + aw + b - f_5(w) = 1 + b_0 + \varepsilon. \tag{11.40}$$

Note that this includes both possibilities in the indeterminate case.

The dynamical system really starts with $w = \alpha/\beta$, which outputs A $\lfloor \alpha/\beta \rfloor$ times before the first B (or possibly one less, if α/β is an integer), so in the version using f_5, we start with an initial output A^ℓ with $\ell := \lfloor \alpha/\beta \rfloor$ and then run the dynamical system $f_\pm$ starting with $w = w_0 := \{\alpha/\beta\}$ (possibly modified if α/β is an integer); the output is by (11.40) given by $BA^{1+b_0+\varepsilon_i}$ for each symbol ε_i in the symbolic sequence. In other words, after the initial A's, the output is obtain from the symbolic sequence by the substitutions

$$0 \to BA^{b_0+1}, \qquad 1 \to BA^{b_0+2}. \tag{11.41}$$

Example 11.2. The case $\alpha > \beta = 0$ was excluded above. In this case, it is easily seen that every seat goes to A. Thus $n_A = n$ for any n. In particular, $n_A/n \to p_A = 1$. (This can be seen as (1.2) with $b_0 = \infty$.)

Example 11.3. The case $\alpha = \beta$ was also excluded above. In this case, if $\alpha = \beta > 0$ and $\zeta > 0$, it is easily seen that the first seat goes to either A or B, and all following seats alternate between the two parties; hence $|n_A - n_B| \leqslant 1$. In particular, $n_A/n \to p_A = 1/2$.

In the extreme case $\alpha = \beta = 1/2$ and $\zeta = 0$, there is a tie at every second seat; the first two seats go to either AB or BA, and the same holds for each following pair of seats; however, the order within each pair is arbitrary. Hence Theorem 1.2(iii) does not hold if, for example, ties are resolved by lot. (However, it holds if ties always are resolved in favour of, say, A.) Nevertheless, in any case we still have $|n_A - n_B| \leqslant 1$.

In the opposite extreme case $\alpha = \beta = 0$, so all votes are for AB (and thus $\zeta = 1$), every seat is a tie. If the ties are resolved by lot, then almost surely the proportion $n_A/n \to p_A = 1/2$, but other resolution rules may give e.g. all seats to A (or B).

Example 11.4. The case $\zeta = 0$ is not excluded above; if $\alpha > \beta > 0$ and $\zeta = 0$, then Phragmén's method is still described by the dynamical system f_5 and (11.41). However, in this case (11.35) yields $a = 1$, and thus $f_\pm(x) = \{x+b\}$ (or $\{x+b\}_+$), which is the limiting case of a rotation on the circle mentioned in Remark 2.2. Our results in the preceding sections do not include this (simple) case, but it is easy to see from (2.18) that Theorem 2.14 still holds, with the rotation number $\rho = b$.

Furthermore, since now $\alpha + \beta = 1$, (1.5) yields

$$b^* = \frac{\alpha - \beta}{\beta} + \frac{\alpha(1-\alpha-\beta)}{\beta\alpha} = \frac{1-2\beta}{\beta} = \frac{1}{\beta} - 2. \tag{11.42}$$

and thus $b = \{b^*\} = \{1/\beta\}$. Since the dynamical system starts with $w = \{\alpha/\beta\} = \{(1-\beta)/\beta\} = \{1/\beta\}$, it follows that $f_\pm^n(w) = \{(n+1)/\beta\}$ or $\{(n+1)/\beta\}_+$; hence, if $\beta = p/q$ is rational, then there is a choice at each p:th iteration. Hence, if e.g. the choices are made by lot, the orbit is a.s. *not* periodic. (We are in an orbit that is periodic except that each p:th term is either 0 or 1, but these may be chosen arbitrarily.) This is in stark contrast to the case $a < 1$ studied in the present paper, see for example Lemma 3.1 and Theorem 8.4, and we see that Theorem 1.2(iii) does not hold when $\zeta = 0$. (Note that in this case, $\rho = b \in \mathbb{Q} \iff \beta \in \mathbb{Q}$ by (11.42).)

Note that the same behaviour was found for $\zeta = 0$ and $\alpha = \beta$ in Example 11.3.

11.5. **Proof of Theorem 1.2.** We consider several cases, and begin with the main case. By symmetry, it suffices to consider $\alpha \geqslant \beta$.

Case 1: $\alpha > \beta > 0$ and $\zeta > 0$. In this case, Phragmén's election method is described by the dynamical system $f_5 = f_\pm$ as described above. Note that $a < 1$ by (11.35). Let $S_m := \sum_{i=0}^{m-1} \varepsilon_i$, where ε_i is the symbolic sequence defined in Section 2.7. Let $m \geqslant 0$ and suppose that at some stage of the election, $n_B = m$. This means that we are in the mth iteration of the dynamical system; in other words, we have so far made m substitutions (11.41), except that the last may be incomplete. Taking into account also the initial string of A's, we obtain

$$n_A = \sum_{i=0}^{m-1} (b_0 + 1 + \varepsilon_i) + O(1) = (b_0+1)m + S_m + O(1). \tag{11.43}$$

Consequently, letting $\rho = \rho(f_\pm)$ be the rotation number of (1.3), Theorem 2.14 yields

$$n_A = (b_0+1)m + \rho m + O(1), \tag{11.44}$$

which together with our assumption $n_B = m$ yields

$$n = n_A + n_B = (2 + b_0 + \rho)m + O(1) \tag{11.45}$$

and thus

$$n_B = m = \frac{n}{2 + b_0 + \rho} + O(1). \tag{11.46}$$

Consequently,

$$\frac{n_B}{n} = \frac{1}{2 + b_0 + \rho} + O\Big(\frac{1}{n}\Big), \tag{11.47}$$

which shows both the existence of the limit p_B as $n \to \infty$, and its value (1.2) in (ii). Furthermore, obviously $n_A/n \to p_A := 1 - p_B$,

(i) follows from (11.46).

Finally, if ρ is rational, then the symbolic sequence is eventually periodic by Theorem 8.4, and thus so is the sequence of awarded seats by (11.41), showing (iii).

This completes the proof in Case 1.

Case 2: $\alpha > \beta > 0$ and $\zeta = 0$. As said in Example 11.4, we can use the dynamical system f_5 above in this case too; the only difference from the preceding case is that now (1.4) yields $a = 1$, but Theorem 2.14 still holds and (i) and (ii) follow as above. However, as noted in Example 11.4, (iii) does not always hold.

In this case, all votes are for A or B, and Phragmén's method reduces to D'Hondt's. The results can also easily be shown directly, see e.g. [19]. Note that in this case, $\rho = b$ and thus, by (1.6)–(1.7) and (11.42), $2 + b_0 + \rho = 2 + b^* = \beta^{-1}$; hence (1.2) yields $p_B = \beta$. In other words, when $\zeta = 0$, the proportion of seats for a party converges to its proportion of the votes, as said earlier.

Case 3: $\alpha > \beta = 0$. Trivial by Example 11.2, with $p_A = 1$ and $p_B = 0$.

Case 4: $\alpha = \beta > 0$. By Example 11.3, (i) holds, with $p_A = p_B = 1/2$, and if $\zeta > 0$, then also (iii) holds. Furthermore, (11.37) yields

$$a + b^* = \frac{1}{1-\beta} - 1 = \frac{\beta}{1-\beta} = \frac{\alpha}{\alpha+\zeta} \leqslant 1. \tag{11.48}$$

In particular, $b^* < 1$ and thus $b_0 = 0$. Furthermore, $a + b \leqslant 1$, and thus the rotation number $\rho = 0$, see Example 8.5. Consequently, (1.2) holds too. $\square$

11.6. **Further results.** We combine Theorem 1.2 with the result by Laurent and Nogueira [21] on rational rotation numbers quoted above as Theorem 8.6, and obtain the following.

Theorem 11.5. *Consider the party version of Phragmén's election method with two parties. If, with notation as in Theorem 1.2, the proportions α, β, ζ are algebraic numbers (in particular, if they are rational), and $0 < \zeta < 1$, then the sequence of awarded seats is eventually periodic. In particular, the proportions n_A/n and n_B/n of seats given to each party converge to rational numbers.*

Proof. By symmetry, we may assume $\alpha \geqslant \beta$. The case $\beta = 0$ is trivial by Example 11.2 (all seats go to A); hence we may assume $\alpha \geqslant \beta > 0$, so Theorem 1.2(ii) applies. The numbers a and b^* in (1.4)–(1.5) are algebraic, and thus so is b by (1.6). Furthermore, $0 < a < 1$ since $\zeta > 0$. Hence, Theorem 8.6 applies and shows that ρ is rational. The proof is completed by Theorem 1.2(iii). $\square$

Remark 11.6. Of course, in a real election, with integer numbers of votes, the proportions of votes are always rational. (Unless votes are weighted, and even then the proportions are rational or algebraic unless some weight is transcendental.) However, we are studying an idealized mathematical situation (where we may let $n \to \infty$), and then it is natural to allow arbitrary real numbers α and β (with $\alpha, \beta \geqslant 0$ and $\alpha + \beta \leqslant 1$).

Example 11.7. When is $p_A = p_B = 1/2$? By symmetry we may assume $\alpha \geqslant \beta$. Then $\beta > 0$ is necessary by Example 11.2, and thus (1.2) shows that $p_B = 1/2$ if and only if $b_0 + \rho = 0$, i.e., if and only if $b_0 = 0$ and $\rho = 0$. By Example 8.5, $\rho = 0 \iff a + b \leqslant 1$, and thus, using also (1.6)–(1.7) and (11.37),

$$p_B = \frac{1}{2} \iff b_0 = 0 \text{ and } a + b \leqslant 1 \iff a + b^* \leqslant 1 \iff \alpha \leqslant 2\beta(1-\beta). \tag{11.49}$$

By symmetry, if $\alpha \leqslant \beta$, then $p_B = 1/2 \iff \beta \leqslant 2\alpha(1-\alpha)$.

We may note that if $\alpha \geqslant \beta$, then either $\alpha \leqslant \frac{1}{2}$ and then $\beta \leqslant \alpha \leqslant 2\alpha(1-\alpha)$, or $\alpha \geqslant \frac{1}{2}$ and then $\beta \leqslant 1 - \alpha \leqslant 2\alpha(1-\alpha)$; thus $\beta \leqslant 2\alpha(1-\alpha)$ always holds when $\alpha \leqslant \beta$. Hence, using symmetry again, we see that

$$p_B = \frac{1}{2} \iff \alpha \leqslant 2\beta(1-\beta) \text{ and } \beta \leqslant 2\alpha(1-\alpha), \tag{11.50}$$

as always excluding the case $\alpha = \beta = 0$.

Given ζ with $0 \leqslant \zeta < 1$, a simple calculation using (11.49) shows that

$$p_B = \frac{1}{2} \iff \frac{3 - \sqrt{1+8\zeta}}{4} \leqslant \alpha \leqslant \frac{1 - 4\zeta + \sqrt{1+8\zeta}}{4}. \tag{11.51}$$

If $p_B = \frac{1}{2}$ and $\zeta > 0$, then the sequence of awarded seats is eventually periodic by Theorem 1.2; furthermore, (11.41) shows that the sequence is eventually alternating between the two parties. In fact, in this simple special case, the sequence alternates from the beginning.

Theorem 11.8. *Consider the party version of Phragmén's election method with two parties, with the notations in Theorem 1.2. If the conditions in* (11.50) *hold and* $0 < \zeta < 1$*, then the seats are awarded alternatingly to A and B (starting with A if $\alpha > \beta$, and with B if $\beta > \alpha$).*

Proof. The assumptions imply $\alpha, \beta > 0$, and the case $\alpha = \beta$ follows by Example 11.3; hence we may, again using symmetry, assume $\alpha > \beta > 0$. Then Phragmén's method is described by the dynamical system $f_5 = f_\pm$ above, starting at $w_0 := \{\alpha/\beta\}$, after an initial $A^{\lfloor \alpha/\beta \rfloor}$. We have, using (11.50), $\beta < \alpha \leqslant 2\beta(1-\beta) < 2\beta$ and thus $1 < \alpha/\beta < 2$. Hence $\lfloor \alpha/\beta \rfloor = 1$, and $0 < w_0 < 1$. Thus the first seat goes to A, and then we run f_5 starting at w_0. We have $0 < a < 1$, and $a + b \leqslant 1$ since $\rho = 0$ (see Example 8.5 or 11.7). In the case, at most one symbol $\varepsilon_i \neq 0$, see Examples 2.1, 2.4, 2.5 and Section 2.7; furthermore, it is easy to see that a non-zero ε_i can occur only in an orbit starting at 1 (if $a + b = 1$) or 0 (if $b = 0$), but this is not the case here since $0 < w_0 < 1$. Thus, $\varepsilon_i = 0$ for all i, and thus (11.41) shows that the output sequence is $A(BA)^\infty$. □

If $p_B = 1/2$ and $\zeta = 0$, then (11.50) (or Example 11.4) implies that $\alpha = \beta = 1/2$; this case is treated in Example 11.3. As shown there, the sequence of elected seats is not necessarily periodic in this case, because of ties. Hence, Theorem 11.8 does not extend to $\zeta = 0$.

Remark 11.9. The result in Theorem 11.8 is both surprising and unsatisfactory from the point of view of applications. For example, if 40% of the votes are for A, 30% for B and 30% for AB, then Theorem 11.8 applies and shows that the seats are awarded $ABAB\dots$; hence, for any even number of seats, A and B get equally many, in spite of the fact that A has substantially more votes than B.

Example 11.10. When is $p_B = 1/3$? This cannot happen if $\beta > \alpha$ or if $\beta = 0$; thus $\alpha \geqslant \beta > 0$. Hence, (1.2) yields $b_0 + \rho = 1$, and thus (recalling that b_0 is an integer), $b_0 = 1$ and $\rho = 0$. Again, by Example 8.5, $\rho = 0 \iff a + b \leqslant 1$. Furthermore, by (1.6)–(1.7), $b^* = b_0 + b$, and thus, using (1.5) and (11.37), for $\alpha \geqslant \beta$,

$$\begin{aligned} p_B = \frac{1}{3} &\iff b_0 = 1 \text{ and } a + b \leqslant 1 \iff b^* \geqslant 1 \text{ and } a + b^* \leqslant 2 \\ &\iff \alpha - 2\beta - \alpha^2 + 2\alpha\beta + 2\beta^2 - 3\alpha\beta^2 \geqslant 0 \text{ and } \alpha \leqslant 3\beta(1-\beta). \end{aligned} \tag{11.52}$$

Example 11.11. When is $p_B = 2/5$? We need $\alpha > \beta > 0$. Furthermore, (1.2) yields $b_0 + \rho = 1/2$, i.e., $b_0 = 0$ and $\rho = 1/2$. Assume $\zeta > 0$, so $0 < a < 1$. Using (6.39) in Example 6.7, we obtain, assuming $\alpha \geqslant \beta$,

$$\rho = \frac{2}{5} \iff \frac{1}{1+a} \leqslant b^* \leqslant \frac{1+a-a^2}{1+a} \iff 1 \leqslant (1+a)b^* \leqslant 1 + a - a^2, \tag{11.53}$$

with a and b^* given by (1.4) and (1.5). This can be expressed as two polynomial inequalities in α and β, with one polynomial of degree 5 and one of degree 4; we omit the details.

Similarly, for any given rational $p \in (0, \frac{1}{2})$, one can see that $p_B = p$ is equivalent to a few polynomial inequalities in α and β, but it seems that the degrees of the polynomials increase with the denominator of p.

12. Thiele's method

12.1. **Definition of Thiele's method.** Thiele's election method has a simple (and rather intuitive) formulation:

Thiele's election method. *Seats are awarded sequentially, and in each round, each ballot is counted as* $1/(\bar{n}+1)$ *for each name on it, where* $\bar{n}$ *is the number of candidates on that ballot that already have been elected.*

As with Phragmén's method, we consider the party version, where each ballot contains a set of parties, and each party may get an arbitrary number of seats; then $\bar{n}$ is counted with repetitions, i.e., $\bar{n}$ is the number of seats that so far have been awarded to the parties on the ballot.

We can rephrase Thiele's method in the following form, similar to the formulation of Phragmén's method in Section 11.2. As above, let v_σ be the number of votes for the set σ of candidates (parties). The numbers n_σ defined below will be the numbers of already elected on the different ballots (denoted $\bar{n}$ in the description above).

(i) Initially all $n_\sigma = 0$.

(ii) The reduced vote for candidate i is defined as

$$W_i := \sum_{\sigma \ni i} \frac{v_\sigma}{1+n_\sigma}. \tag{12.1}$$

(iii) The candidate i with the largest W_i is elected to the next seat, breaking ties by lot or some other method. (In the original version, only unelected candidates are considered. In the party version, repetitions are allowed.)

(iv) If i is elected, then n_σ is updated for every $\sigma \ni i$ (i.e., for the ballots that contributed to the election of i); the new value is

$$n'_\sigma := n_\sigma + 1. \tag{12.2}$$

n_σ remains unchanged when $\sigma \not\ni i$.
Repeat from (ii).

The difference from Phragmén's method is thus that the reduction of votes in (12.1) is done in a different way.

Remark 12.1. A ballot voting for all parties will give the same contribution to everyone, and thus does not influence the result. In other words, with Thiele's method, ballots containing all parties can be ignored, just as blank votes.

12.2. **Main results for Thiele's method.** We assume as in Section 11.3 that we are given a set $\mathcal{P}$ of parties, and some numbers v_σ of votes on the sets $\sigma \subseteq \mathcal{P}$. We let $n \geqslant 1$ seats by distributed by Thiele's method, and let n_i be the number of seats received by party $i \in \mathcal{P}$. We also let $p_i := n_i/n$, the fraction of the seats received by i, and we define for a set $\sigma \subseteq \mathcal{P}$ the sums

$$n_\sigma := \sum_{i\in\sigma} n_i, \qquad p_\sigma := \sum_{i\in\sigma} p_i = n_\sigma/n. \tag{12.3}$$

(These quantities all depend on n, but we do not show this in the notation.)

We let $N = |\mathcal{P}|$, the number of parties, and assume for notational convenience that $\mathcal{P} = \{1, \dots, N\}$. We let $\mathbf{p} = \mathbf{p}_n := (p_1, \dots, p_N)$, the vector of proportions of seats given to the different parties. Note that $\mathbf{p}$ belongs to the simplex

$$\mathfrak{S} = \mathfrak{S}_N := \Bigl\{(x_1, \dots, x_N) : x_i \geqslant 0 \text{ and } \sum_{i=1}^N x_i = 1\Bigr\}. \tag{12.4}$$

Let $\mathfrak{S}^\circ := \bigl\{(x_1, \dots, x_N) \in \mathfrak{S} : x_i > 0 \text{ for all } i\bigr\}$, the corresponding open simplex.

The following two theorems give conditions that guarantee that the vector $\mathbf{p}$ converges, and provide a method to find the limit by solving a system of (non-linear) equations. Theorem 12.2 is more general, but its condition may be less easy to verify; Theorem 12.3 has a simple condition that still covers most cases of interest. Furthermore, we give the even more general Theorem 12.5 below, with a different characterization of the limit. The proofs of the results below are given in the next subsection.

Theorem 12.2. *Consider Thiele's method for a set $\mathcal{P} = \{1, \dots, N\}$ of N parties with some given numbers of votes $\{v_\sigma\}_{\sigma\subseteq\mathcal{P}}$. For a vector $(x_1, \dots, x_N)$, define*

$$x_\sigma := \sum_{i\in\sigma} x_i, \qquad \sigma \subseteq \mathcal{P}. \tag{12.5}$$

If, using (12.5)*, the system of $N-1$ equations*

$$\sum_{\sigma\ni 1} \frac{v_\sigma}{x_\sigma} = \sum_{\sigma\ni 2} \frac{v_\sigma}{x_\sigma} = \dots = \sum_{\sigma\ni N} \frac{v_\sigma}{x_\sigma} \tag{12.6}$$

has a unique solution $\mathbf{x}_0$ in the open simplex $\mathfrak{S}^\circ$, then $\mathbf{p}_n \to \mathbf{x}_0$ as $n \to \infty$.

Note that if $\mathbf{x} \in \mathfrak{S}^\circ$, or more generally $\mathbf{x} \in \mathfrak{S}$, then

$$\sum_{i=1}^{N} x_i = 1, \tag{12.7}$$

which together with (12.6) yields a system of N non-linear equations in the N unknowns x_i.

Theorem 12.3. *Consider Thiele's method for a set $\mathcal{P}$ of N parties with some given numbers of votes $\{v_\sigma\}_{\sigma\subseteq\mathcal{P}}$. Suppose that every party gets some individual vote, i.e.,*

$$v_{\{i\}} > 0 \text{ for every } i \in \mathcal{P}. \tag{12.8}$$

Then the system (12.6) *has a unique solution $\mathbf{x}_0$ in $\mathfrak{S}^\circ$, and $\mathbf{p}_n \to \mathbf{x}_0$ as $n \to \infty$. Moreover, $\mathbf{x}_0$ is a smooth function of the vote numbers v_σ as long as* (12.8) *holds.*

The limit $\mathbf{x}_0$ in these theorems can also be characterized as the solution to an optimization problem, which futhermore allows for a more general result.

Using the notations (12.5) and (11.4), and the standard convention $0^0 = 1$, define the function, for $x_1, \dots, x_N \geqslant 0$,

$$\Psi(x_1, \dots, x_N) := \prod_{\sigma\neq\emptyset} x_\sigma^{v_\sigma} = \prod_{\sigma\in\Pi} x_\sigma^{v_\sigma}. \tag{12.9}$$

It is immediate that Ψ is a continuous function $[0,\infty)^N \to [0,\infty)$. Let M be the maximum of ψ on the compact set $\mathfrak{S}$, and let

$$\mathcal{M} := \{\mathbf{x} \in \mathfrak{S} : \Psi(\mathbf{x}) = M\} \tag{12.10}$$

be the set where the maximum is attained.

Lemma 12.4. (i) *$\mathcal{M}$ is a non-empty compact convex subset of $\mathfrak{S}$.*

(ii) *If $\mathbf{x} \in \mathfrak{S}^\circ$, then*

$$\mathbf{x} \in \mathcal{M} \iff \text{(12.6) holds.} \tag{12.11}$$

(iii) *If* (12.6) *has a unique solution $\mathbf{x}_0$ in $\mathfrak{S}^\circ$, then $\mathcal{M} = \{\mathbf{x}_0\}$, i.e., $\mathbf{x}_0$ is the only point in $\mathfrak{S}$ where the maximum of Ψ is attained.*

The limit $\mathbf{x}_0$ in Theorems 12.2–12.3 is thus the unique maximum point of Ψ on $\mathfrak{S}$. The following, more general, theorem gives a (weaker) result also in the case when the maximum point is not unique.

Theorem 12.5. *Consider Thiele's method for N parties $1,\dots,N$, with some given numbers of votes v_σ, for $\sigma\subseteq\mathcal{P}=\{1,\dots,N\}$. Then, as $n\to\infty$, $\mathbf{p}_n\to\mathcal{M}$, in the sense that the (Euclidean) distance $d(\mathbf{p}_n,\mathcal{M})\to 0$. In particular, if $\mathcal{M}$ consists of a single point, i.e., $\mathcal{M}=\{\mathbf{x}_0\}$ for some $\mathbf{x}_0\in\mathfrak{S}$, then $\mathbf{p}_n\to\mathbf{x}_0$.*

12.3. **Proofs.** We consider the votes $\{v_\sigma\}$ as fixed. Explicit and implicit constants below generally depend on $\{v_\sigma\}$.

We define for $x_1,\dots,x_N\geqslant 0$, recalling (12.9) and with the convention $0\cdot\infty=0$,

$$\psi(x_1,\dots,x_N):=\log\Psi(x_1,\dots,x_N)=\sum_{\sigma\in\Pi}v_\sigma\log x_\sigma. \tag{12.12}$$

Note that ψ may take the value $-\infty$. Since Ψ is a continuous function $[0,\infty)^N\to[0,\infty)$, $\psi=\log\Psi$ is a continuous function $[0,\infty)^N\to[-\infty,\infty)$ (with the standard topology); furthermore, ψ is concave.

The partial derivatives of ψ are

$$\partial_i\psi:=\frac{\partial\psi}{\partial x_i}=\sum_{\sigma\ni i}\frac{v_\sigma}{x_\sigma}. \tag{12.13}$$

(If $\psi(\mathbf{x})=-\infty$, we regard the sum in (12.13) as a definition of $\partial_i\psi(\mathbf{x})$.) These derivatives are finite (and smooth) in $(0,\infty)^N$, but may be infinite on the boundary; more precisely, $\partial_i\psi=+\infty$ when $x_\sigma=0$ for some $\sigma\in\Pi$ with $i\in\sigma$.

We are mainly interested in the behaviour of ψ on the simplex $\mathfrak{S}$. However, the partial derivatives ∂_i are along directions poining out of $\mathfrak{S}$; we thus also consider directional derivatives in $\mathfrak{S}$. Let e_i, $i=1,\dots,N$, be the unit vectors and define, for $\mathbf{x}=(x_1,\dots,x_N)\in\mathfrak{S}$,

$$e_i^*=e_i^*(\mathbf{x}):=e_i-\sum_{j=1}^N x_je_j \tag{12.14}$$

which is parallel to $\mathfrak{S}$ and can be seen as a projection of e_i to the hyperplane H of vectors tangent to $\mathfrak{S}$, and the corresponding directional derivative

$$\partial_i^*:=\partial_{e_i^*}=\partial_i-\sum_{j=1}^N x_j\partial_j. \tag{12.15}$$

Equivalently, for any differentiable function f on $(0,\infty)^N$ and $\mathbf{x}\in\mathfrak{S}$,

$$\partial_i^*f(\mathbf{x})=\partial_if(\mathbf{x})-\frac{df(t\mathbf{x})}{dt}\Big|_{t=1}. \tag{12.16}$$

Note that the vectors e_i^* span the hyperplane H, and thus the operators ∂_i^* span the $(N-1)$-dimensional space of directional derivatives parallel to $\mathfrak{S}$; moreover, they satsfy the linear relation

$$\sum_{i=1}^N x_i\partial_i^*=0. \tag{12.17}$$

As said above, $\partial_i\psi(\mathbf{x})$ may be $+\infty$ (but not $-\infty$). Furthermore, it follows from (12.13) that $x_i\partial_i\psi(\mathbf{x})=O(1)$. Hence $\partial_i^*\psi(\mathbf{x})$ is well-defined by (12.15) for every $\mathbf{x} \in \mathfrak{S}$, with

$$\partial_i^* \psi(\mathbf{x}) = \partial_i \psi(\mathbf{x}) + O(1) \in (-\infty, \infty]. \tag{12.18}$$

Moreover, if $v^* := \sum_\sigma v_\sigma$, then by (12.12), for any $t > 0$,

$$\psi(t\mathbf{x}) = \psi(\mathbf{x}) + v^* \log t. \tag{12.19}$$

Hence, by (12.16), for $\mathbf{x} \in \mathfrak{S}^\circ$,

$$\partial_i^* \psi(\mathbf{x}) = \partial_i \psi(\mathbf{x}) - v^*. \tag{12.20}$$

(More generally, (12.20) holds for all $\mathbf{x} \in \mathfrak{S}$ with $\psi(\mathbf{x}) > -\infty$, but not necessarily everywhere on the boundary of $\mathfrak{S}$.)

Let $m := \log M = \max_{\mathfrak{S}} \psi$. Then (12.10) can be written

$$\mathcal{M} = \{\mathbf{x} \in \mathfrak{S} : \psi(\mathbf{x}) = m\}. \tag{12.21}$$

Since the function ψ is concave (and the set $\mathfrak{S}$ convex), if $\mathbf{x} \in \mathfrak{S}^\circ$, then

$$\mathbf{x} \in \mathcal{M} \iff \partial_i^* \psi(\mathbf{x}) = 0 \text{ for every } i. \tag{12.22}$$

For x on the boundary $\partial\mathfrak{S}$, we still have an implication $\Leftarrow$, but not necessarily in the opposite direction, see Example 12.14 below.

Proof of Lemma 12.4. (i): $\mathcal{M}$ is non-empty and compact by the definition (12.10) because Ψ is continuous and $\mathfrak{S}$ is compact. Furthermore, $\mathcal{M}$ is convex by (12.21) because ψ is a concave function.

(ii): If (12.6) holds, then by (12.13), $\partial_1 \psi(\mathbf{x}) = \dots = \partial_N \psi(\mathbf{x})$. If also $\mathbf{x} \in \mathfrak{S}^\circ$, so (12.7) holds, then (12.15) yields $\partial_i^* \psi(\mathbf{x}) = 0$ for every i, and thus $\mathbf{x} \in \mathcal{M}$ by (12.22).

Conversely, if $\mathbf{x} \in \mathcal{M} \cap \mathfrak{S}^\circ$, then (12.22) and (12.20) yield $\partial_i \psi(\mathbf{x}) = v^*$ for all i, and thus (12.13) shows that (12.6) holds.

(iii): If (12.6) has a unique solution $\mathbf{x}_0$ in $\mathfrak{S}^\circ$, then (ii) shows that $\mathcal{M} \cap \mathfrak{S}^\circ = \{\mathbf{x}_0\}$. Since $\mathcal{M}$ is convex by (i), it follows that $\mathcal{M} = \{\mathbf{x}_0\}$. □

An important link between the seat assignments by Thiele's method and the function ψ is given by the following lemma. We define

$$n_* := \min_{\sigma \in \Pi} n_\sigma, \tag{12.23}$$

$$p_* := \min_{\sigma \in \Pi} p_\sigma = n_*/n. \tag{12.24}$$

Lemma 12.6. *For every party i,*

$$W_i = \frac{1}{n} \partial_i \psi(\mathbf{p}) + O\Big(\frac{1}{n_*^2}\Big). \tag{12.25}$$

Moreover

$$\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) = \frac{1}{n} \max_i \partial_i^* \psi(\mathbf{p}_n) + O\Big(\frac{1}{n_*^2}\Big). \tag{12.26}$$

Proof. First, for any i, (12.1) yields

$$W_i = \sum_{\sigma \ni i} \Big(\frac{v_\sigma}{n_\sigma} + O\Big(\frac{v_\sigma}{n_\sigma^2}\Big)\Big) = \sum_{\sigma \ni i} \frac{v_\sigma}{n p_\sigma} + O\Big(\frac{1}{n_*^2}\Big), \tag{12.27}$$

which by (12.13) shows (12.25).

Suppose that the $(n+1)$:th seat goes to party ℓ. Let, with $\mathbf{n} := (n_1, \dots, n_N)$,

$$\Delta\mathbf{p} := \mathbf{p}_{n+1} - \mathbf{p}_n = \frac{\mathbf{n} + e_\ell}{n+1} - \frac{\mathbf{n}}{n} = \frac{e_\ell - \mathbf{p}_n}{n+1} \tag{12.28}$$

and note that $|\Delta\mathbf{p}| = O(1/n)$.

It follows from (12.13) that

$$\partial_i\psi(\mathbf{x}) = O\left(\frac{1}{\min_{\sigma\in\Pi} x_\sigma}\right), \tag{12.29}$$

$$\partial_i^2\psi(\mathbf{x}) = O\left(\frac{1}{(\min_{\sigma\in\Pi} x_\sigma)^2}\right). \tag{12.30}$$

Thus, for $\mathbf{x}$ on the line segment between $\mathbf{p}_n$ and $\mathbf{p}_{n+1}$, since $\mathbf{p}_{n+1} \geqslant \frac{n}{n+1}\mathbf{p}_n$,

$$\partial_i^2\psi(\mathbf{x}) = O\left(\frac{1}{p_*^2}\right) = O\left(\frac{n^2}{n_*^2}\right). \tag{12.31}$$

Hence, a Taylor expansion yields, using (12.28) and (12.15),

$$\begin{aligned}\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) &= \psi(\mathbf{p} + \Delta\mathbf{p}) - \psi(\mathbf{p}) = \Delta\mathbf{p}\cdot\nabla\psi(\mathbf{p}) + O\left(\frac{n^2}{n_*^2}|\Delta\mathbf{p}|^2\right)\\ &= \frac{\partial_\ell\psi(\mathbf{p}) - \sum_{j=1}^N p_j\partial_j\psi(\mathbf{p})}{n+1} + O\left(\frac{1}{n_*^2}\right)\\ &= \frac{\partial_\ell^*\psi(\mathbf{p})}{n+1} + O\left(\frac{1}{n_*^2}\right).\end{aligned} \tag{12.32}$$

Furthermore, by (12.29),

$$\frac{\partial_\ell^*\psi(\mathbf{p})}{n+1} - \frac{\partial_\ell^*\psi(\mathbf{p})}{n} = -\frac{\partial_\ell^*\psi(\mathbf{p})}{n(n+1)} = O\left(\frac{1}{n^2 p_*}\right) = O\left(\frac{1}{n_*^2}\right). \tag{12.33}$$

Consequently, (12.32) yields

$$\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) = \frac{1}{n}\partial_\ell^*\psi(\mathbf{p}) + O\left(\frac{1}{n_*^2}\right). \tag{12.34}$$

Furthermore, by the definition of Thiele's method, $W_\ell = \max_i W_i$, and thus (12.25) yields

$$\frac{1}{n}\partial_\ell^*\psi(\mathbf{p}) = \max_i W_i + O\Bigl(\frac{1}{n_*^2}\Bigr) = \max_i \frac{1}{n}\partial_i^*\psi(\mathbf{p}) + O\Bigl(\frac{1}{n_*^2}\Bigr), \tag{12.35}$$

which yields (12.26) by (12.34). □

Lemma 12.7. *Let $U \subset \mathfrak{S}$ be an open neighbourhood of $\mathcal{M}$. Then there exists $c_1 > 0$ such that for every $\mathbf{x} \in \mathfrak{S} \setminus U$, there exists i with $\partial_i^*\psi(\mathbf{x}) \geqslant c_1$.*

Proof. Let

$$g(\mathbf{x}) := \max_{1\leqslant i\leqslant N} \partial_i^*\psi(\mathbf{x}). \tag{12.36}$$

The assertion is equivalent to $g(\mathbf{x}) \geqslant c_1$ for $x \notin U$. We first show $g(\mathbf{x}) > 0$.

Suppose that $\mathbf{x} \in \mathfrak{S}$ with $g(\mathbf{x}) \leqslant 0$. Then $\partial_i^*\psi(\mathbf{x}) \leqslant 0$ for every i. It follows from (12.17) that then $x_i\partial_i^*\psi(\mathbf{x}) = 0$ for every i, so $\partial_i^*\psi(\mathbf{x}) = 0$ for every i such that $x_i > 0$.

Let $\mathbf{y} := (y_1, \dots, y_N)$ be any point in $\mathfrak{S}$, and let $h(t) := \psi(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$. Then h is a concave function on $[0, 1]$, and, using (12.15) and $\sum_i (y_i - x_i) = 1 - 1 = 0$,

$$h'(0) = \sum_{i=1}^{N} (y_i - x_i)\partial_i \psi(\mathbf{x}) = \sum_{i=1}^{N} (y_i - x_i)\partial_i^* \psi(\mathbf{x}). \tag{12.37}$$

If $x_i > 0$, then $\partial_i^* \psi(\mathbf{x}) = 0$ as just seen. Furthermore, if $x_i = 0$, then $y_i - x_i \geqslant 0$ and $\partial_i^* \psi(\mathbf{x}) \leqslant 0$. It follows that every term in the final sum in (12.37) is $\leqslant 0$, and thus $h'(0) \leqslant 0$. Since h is concave, this implies $\psi(\mathbf{y}) = h(1) \leqslant h(0) = \psi(\mathbf{x})$.

We have shown that if $\mathbf{x} \in \mathfrak{S}$ and $g(\mathbf{x}) \leqslant 0$, then $\psi(\mathbf{x}) \geqslant \psi(\mathbf{y})$ for every $\mathbf{y} \in \mathfrak{S}$, and thus $\mathbf{x} \in \mathcal{M}$. Equivalently, if $\mathbf{x} \notin \mathcal{M}$, then $g(\mathbf{x}) > 0$.

To complete the proof, it suffices to show that g is continuous on $\mathfrak{S}$ (with values in $[0, \infty]$). This is not quite trivial, since the individual $\partial_i^* \psi$ in general are not, because $x_j \partial_j \psi(\mathbf{x})$ by (12.13) is discontinuous at $x_j = 0$ if $v_{\{j\}} > 0$. We let $\mathbf{x} \in \mathfrak{S}$ and consider two cases.

(i) If $\partial_i \psi(\mathbf{x}) < \infty$ for every i, then by (12.13) and (12.15), this holds in a neighbourhood V of $\mathbf{x}$, and in V furthermore every $\partial_i \psi$ and every $\partial_i^* \psi$ is continuous. Hence, g is continuous at $\mathbf{x}$.

(ii) If $\partial_i \psi(\mathbf{x}) = \infty$ for some i, suppose that $\mathbf{y} \to \mathbf{x}$ with $\mathbf{y} \in \mathfrak{S}$. Then $\partial_i \psi(\mathbf{y}) \to \infty$ and thus, using (12.18),

$$g(\mathbf{y}) \geqslant \partial_i^* \psi(\mathbf{y}) = \partial_i \psi(\mathbf{y}) + O(1) \to \infty = g(\mathbf{x}). \tag{12.38}$$

Hence, g is continuous at $\mathbf{x}$ in this case too.

Consequently, g is continuous everywhere in $\mathfrak{S}$, and since we have shown that $g > 0$ on the compact set $\mathfrak{S} \setminus U \subseteq \mathfrak{S} \setminus \mathcal{M}$, the result follows. □

Lemma 12.8. *As $n \to \infty$, $n_* \to \infty$.*

Proof. Suppose not. Then, since each n_σ is non-decreasing, there exists $\sigma \in \Pi$ such that $n_\sigma = O(1)$. Let

$$\Pi_0 := \{\sigma \in \Pi : n_\sigma = O(1)\}, \tag{12.39}$$

$$\mathcal{E} := \bigcup_{\sigma \in \Pi_0} \sigma = \bigl\{i : \exists \sigma \in \Pi_0 \text{ with } i \in \sigma\bigr\}. \tag{12.40}$$

Then $\mathcal{E}$ is a non-empty set of parties, and if $i \in \mathcal{E}$, then there exists σ with $i \in \sigma \in \Pi_0$ and thus $n_i \leqslant n_\sigma = O(1)$. In other words, after some time, no further seat goes to any party in $\mathcal{E}$.

On the other hand, if $i \in \mathcal{E}$, take again $\sigma \in \Pi_0$ with $i \in \sigma$. Then by (12.1),

$$W_i \geqslant \frac{v_\sigma}{1 + n_\sigma} \geqslant c, \tag{12.41}$$

for some $c > 0$. On the other hand, if $i \notin \mathcal{E}$, then $n_\sigma \to \infty$ for every $\sigma \in \Pi$ such that $i \in \sigma$, and thus (12.1) yields $W_i \to 0$. This implies that if n is large enough, then $W_i < c$ for every $i \notin \mathcal{E}$, so by (12.41), the party i with the largest W_i is a party in $\mathcal{E}$, and thus every seat, for large n, goes to a party in $\mathcal{E}$. This contradiction proves the lemma. □

Lemma 12.9. *Let $U \subset \mathfrak{S}$ be an open neighbourhood of $\mathcal{M}$. Then there exists n_0 and $c_2 > 0$ such that for all $n \geqslant n_0$, either $\mathbf{p}_n \in U$ or*

$$\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) \geqslant c_2/n. \tag{12.42}$$

Proof. Let c_1 be as in Lemma 12.7 and let $c_2 := c_1/2$. We assume $\mathbf{p} = \mathbf{p}_n \notin U$ and use (12.26). We consider two cases.

Case 1: $n_* \geqslant n^{3/4}$. By (12.26) and Lemma 12.7,

$$\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) \geqslant \frac{c_1}{n} + O(n_*^{-2}) = \frac{c_1}{n} + O\bigl(n^{-3/2}\bigr), \tag{12.43}$$

which is larger than c_2/n for large n.

Case 2: $n_* < n^{3/4}$. Let $\sigma \in \Pi$ with $n_\sigma = n_*$. By (12.13), for any $i \in \sigma$,

$$\partial_i\psi(\mathbf{p}) \geqslant \frac{v_\sigma}{p_\sigma} = n\frac{v_\sigma}{n_\sigma} \tag{12.44}$$

and thus by (12.26) and (12.18), using also Lemma 12.8,

$$\begin{aligned}\psi(\mathbf{p}_{n+1}) - \psi(\mathbf{p}_n) &\geqslant \frac{1}{n}\partial_i^*\psi(\mathbf{p}) + O\bigl(n_*^{-2}\bigr) = \frac{1}{n}\partial_i\psi(\mathbf{p}) + O\bigl(n^{-1}\bigr) + O\bigl(n_*^{-2}\bigr)\\ &\geqslant \frac{v_\sigma}{n_\sigma} + O\bigl(n^{-1}\bigr) + O\bigl(n_*^{-2}\bigr) = \frac{v_\sigma}{n_*} + o\bigl(n_*^{-1}\bigr). \end{aligned} \tag{12.45}$$

For large n, the right-hand side is at least $v_\sigma/(2n_*) \geqslant c_2/n$. □

We can now prove the theorems showing convergence of the proportions $\mathbf{p}_n$ for Thiele's method.

Proof of Theorem 12.5. Let $\varepsilon > 0$, and let $U := \{\mathbf{x} \in \mathfrak{S} : d(\mathbf{x}, \mathcal{M}) < \varepsilon\}$. If $\mathbf{x} \notin U$, then $\mathbf{x} \notin \mathcal{M}$ and thus $\psi(\mathbf{x}) < m$; hence, by compactness, there exists $\delta > 0$ such that $\psi(\mathbf{x}) \leqslant m - \delta$ for $x \in \mathfrak{S} \setminus U$.

Let $U_1 := \{\mathbf{x} \in \mathfrak{S} : \psi(\mathbf{x}) > m - \delta\}$ and $U_2 := \{\mathbf{x} \in \mathfrak{S} : \psi(\mathbf{x}) > m - \delta/2\}$. Then U_1 and U_2 are open in $\mathfrak{S}$ and

$$\mathcal{M} \subset U_2 \subset \overline{U}_2 \subset U_1 \subseteq U. \tag{12.46}$$

In particular, the two compact sets $\overline{U}_2$ and $\mathfrak{S}\setminus U_1$ are disjoint, and thus have a positive distance η. In other words, $\eta > 0$ and if $\mathbf{x} \in \mathfrak{S}$ with $d(\mathbf{x}, \overline{U}_2) < \eta$, then $\mathbf{x} \in U_1$. We apply Lemma 12.9 to U_2.

First, we claim that $\mathbf{p}_n \in U_2$ for infinitely many n. In fact, if this is false, then by Lemma 12.9, (12.42) holds for all large n. Since $\sum_n c_2/n = \infty$, this would imply $\psi(\mathbf{p}_n) \to \infty$, which is a contradiction because $\psi(\mathbf{x}) \leqslant 0$ when $\mathbf{x} \in \mathfrak{S}$.

Next, suppose that $n \geqslant n_0$ and that $\mathbf{p}_n \in U_1$. There are two cases.

(i) If $\mathbf{p}_n \notin U_2$, then (12.42) holds, and thus

$$\psi(\mathbf{p}_{n+1}) > \psi(\mathbf{p}_n) > m - \delta; \tag{12.47}$$

hence $\mathbf{p}_{n+1} \in U_1$.

(ii) If $\mathbf{p}_n \in U_2$, then we use (12.28) which implies $|\Delta\mathbf{p}| \to 0$. Hence, provided n is large enough, $|\Delta\mathbf{p}| < \eta$, which together with $\mathbf{p}_n \in U_2$ and the definition of η implies $\mathbf{p}_{n+1} = \mathbf{p}_n + \Delta\mathbf{p} \in U_1$.

We have thus shown that, in any case, if n is large enough and $\mathbf{p}_n \in U_1$, then $\mathbf{p}_{n+1} \in U_1$. Since we also have shown that $\mathbf{p}_n \in U_2 \subset U_1$ for arbitrarily large n, it follows that for all sufficiently large n, $\mathbf{p}_n \in U_1 \subset U$, and thus $d(\mathbf{p}_n, \mathcal{M}) < \varepsilon$. □

Proof of Theorem 12.2. The assumption that $\mathbf{x}_0$ is a unique solution of (12.6) in $\mathfrak{S}^\circ$ implies $\mathcal{M} = \{\mathbf{x}_0\}$ by Lemma 12.4(iii). Hence, the result follows from Theorem 12.5. □

Proof of Theorem 12.3. When (12.8) holds, $\Psi(\mathbf{x}) = 0$ as soon as some $x_i = 0$; hence the maximum M of Ψ can not be attained on the boundary of $\mathfrak{S}$ so $\mathcal{M} \subset \mathfrak{S}^\circ$. Furthermore, along any straight line in $\mathfrak{S}^\circ$, each term in the sum (12.12) is smooth and concave, and at least one of the terms has a strictly negative second derivative. (For example the term with $\sigma = \{i\}$ for any i such that x_i varies along the line.) It follows that ψ is strictly concave in $\mathfrak{S}^\circ$, and thus the maximum set $\mathcal{M}$ cannot contain more than one point. Hence $\mathcal{M} = \{\mathbf{x}_0\}$ for some point $\mathbf{x}_o \in \mathfrak{S}^\circ$. It follows by Lemma 12.4(ii) that $\mathbf{x}_0$ is the unique solution of (12.6) in $\mathfrak{S}^\circ$, and $\mathbf{p}_n \to \mathbf{x}_0$ follows by Theorem 12.5 (or Theorem 12.2).

Finally, use $x_1, \dots, x_{N-1}$ as coordinates on $\mathfrak{S}^\circ$ and write $\bar\psi(x_1, \dots, x_{N-1}) := \psi(x_1, \dots, x_{N-1}, 1 - x_1 - \dots - x_N)$. Then the maximum point $\mathbf{x}_0$ is given by

$$D\bar\psi := \Big(\frac{\partial\bar\psi}{\partial x_i}\Big)_{i=1}^{N-1} = 0. \tag{12.48}$$

Moreover, the function $\bar\psi$ is concave in $\mathfrak{S}^\circ$, with a strictly negative second derivative along any line as shown above; in other words, the Hessian matrix $\big(\frac{\partial^2\bar\psi}{\partial x_i \partial x_j}\big)_{i,j=1}^{N-1}$ is negative definite, and thus non-singular at every point. It follows from the implicit function theorem that the solution $\mathbf{x}_0$ of (12.48) is a smooth function of the parameters v_σ. □

12.4. **Examples and further results.**

Example 12.10 (Two parties)**.** Suppose that there are two parties, A and B, and assume $v_A, v_B > 0$. The equation (12.6) is

$$\frac{v_A}{x_A} + \frac{v_{AB}}{x_{AB}} = \frac{v_B}{x_B} + \frac{v_{AB}}{x_{AB}}, \tag{12.49}$$

which simplifies to $v_A/x_A = v_B/x_B$, so the system (12.6)–(12.7) has the unique solution $x_A = v_A/(v_A + v_B)$, $x_B = v_A/(v_A + v_B)$. Theorem 12.2 applies and thus $n_A/n \to x_A = v_A/(v_A + v_B)$; see also Theorem 12.3. This also follows from Remark 12.1, which for two parties says that we can ignore the ballots AB, leaving only ballots A and B, and then Thiele's method reduces to D'Hondt's for which the result is well known.

We continue with some examples with three parties.

Example 12.11. Supose that there are three parties A, B, C, and 5 votes: 1 A, 1 B, 1 C, 1 AB, 1 AC. Then (12.6) is

$$\frac{1}{x_A} + \frac{1}{x_A + x_B} + \frac{1}{x_A + x_C} = \frac{1}{x_B} + \frac{1}{x_A + x_B} = \frac{1}{x_C} + \frac{1}{x_A + x_C}. \tag{12.50}$$

Theorem 12.3 applies, and thus (12.6)–(12.7) has a unique solution in $\mathfrak{S}^\circ$, which by symmetry has to satisfy $x_B = x_C$. Hence (12.6) simplifies to

$$\frac{1}{x_A} + \frac{1}{x_A + x_B} = \frac{1}{x_B}. \tag{12.51}$$

Furthermore, $x_B = x_C = (1 - x_A)/2$, and we obtain

$$\frac{1}{x_A} + \frac{2}{1 + x_A} = \frac{2}{1 - x_A} \tag{12.52}$$

which yields the quadratic equation $5x_A^2 - 1 = 0$. Hence the maximum point $\mathbf{x}_0$ is given by $x_A = 1/\sqrt{5}$, $x_B = x_C = \frac{1}{2}(1 - 1/\sqrt{5})$. Theorem 12.2 yields $\mathbf{p} \to \mathbf{x}_0 = \frac{1}{2\sqrt{5}}(2, \sqrt{5} - 1, \sqrt{5} - 1)$.

The fact that the proportions converge to irrational numbers shows that even in this simple example, there is no ultimate periodicity in the seat assignment.

Problem 12.12. Is the sequence of seats assigned to A quasiperiodic in some sense? Can it be described explicitly?

More generally, a similar calculation shows that if the number of votes for A is changed to an arbitrary $v_A > 0$ (with the other votes kept the same), then $n_A/n \to x_A = \sqrt{v_A/(v_A + 4)}$.

In the general (non-symmetric) case with 3 parties, (12.6)–(12.7) lead (using Maple) to a quartic equation for x_A, where the coefficients are polynomials of degree 3 in the vote numbers v_σ. We spare the reader the general formula, and give a numerical example.

Example 12.13. Supose that there are three parties A, B, C, and 9 votes: 1 A, 2 B, 3 C, 1 AB, 1 AC, 1 BC. Theorem 12.3 applies and shows $\mathbf{p} \to \mathbf{x}_0$ for some solution $\mathbf{x}_0 = (x_A, x_B, x_C)$ to (12.6)–(12.7). Maple yields that x_A is a root of

$$135x_A^4 - 161x_A^3 - 22x_A^2 + 64x_A - 10 = 0. \tag{12.53}$$

Numerically, $x_A = 0.1797714258$, $x_B = 0.341215728$, $x_C = 0.4790128462$.

Example 12.14 (An exceptional case)**.** Supose that there are three parties A, B, C, and 6 votes: 2 A, 2 B, 1 AC, 1 BC. Then (12.6) is

$$\frac{2}{x_A} + \frac{1}{x_A + x_C} = \frac{2}{x_B} + \frac{1}{x_B + x_C} = \frac{1}{x_A + x_C} + \frac{1}{x_B + x_C} \tag{12.54}$$

and it is easily found that the unique solution that also satisfies (12.7) is $(\frac{2}{3}, \frac{2}{3}, -\frac{1}{3})$. This solution lies outside $\mathfrak{S}$, so Theorem 12.2 does not apply. However, Theorem 12.5 still applies, and it is easily verified that the maximum set $\mathcal{M}$ consist of the single point $(\frac{1}{2}, \frac{1}{2}, 0)$; thus $p_{An} \to \frac{1}{2}$, $p_{Bn} \to \frac{1}{2}$ and $p_{Cn} \to 0$.

In fact, it is easily seen from (12.1) that C will never get any seat, since always at least one of W_A and W_B is larger than W_C. Furthermore, each pair of sets goes to either A, B or B, A, and thus for any even number of seats n, $p_{An} = p_{Bn} = \frac{1}{2}$ and $p_{Cn} = 0$ exactly.

Example 12.15 (Another exceptional case)**.** Suppose that there are three parties A, B, C and two votes: 1 A and 1 BC. In this case, $\Psi(x_A, x_B, x_C) = x_A(x_B + x_C) = x_A(1 - x_A)$, and it is easy to see that $\mathcal{M} = \{(\frac{1}{2}, x_B, \frac{1}{2} - x_B) : x_B \in [0, \frac{1}{2}]\}$, a line segment.

Indeed, in this case, of each pair of seats, one goes to A and the other to either B or C. If ties are resolved by lot, almost surely $\mathbf{p}_n \to (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, but for other tie-breaking rules, other limits in $\mathcal{M}$ are possible, and $\mathbf{p}_n$ may even oscillate without a limit, for example if a tie for seat n is resolved in favour of B when $\lfloor \log_2 n \rfloor$ is even, and in favour of C otherwise.

Example 12.16. Supose that there are three parties A, B, C, and only votes for combinations of two parties, with $v_{AB}, v_{AC}, v_{BC} > 0$ and $v_{AB} + v_{AC} + v_{BC} = 1$. Then (12.6) is

$$\frac{v_{AB}}{x_A + x_B} + \frac{v_{AC}}{x_A + x_C} = \frac{v_{AB}}{x_A + x_B} + \frac{v_{BC}}{x_B + x_C} = \frac{v_{AC}}{x_A + x_C} + \frac{v_{BC}}{x_B + x_C} \quad (12.55)$$

which yields

$$\frac{v_{AB}}{x_A + x_B} = \frac{v_{AC}}{x_A + x_C} = \frac{v_{BC}}{x_B + x_C}. \quad (12.56)$$

The equations (12.56) and (12.7) have the unique solution

$$(x_A, x_B, x_C) = \big(v_{AB} + v_{AC} - v_{BC}, v_{AB} + v_{BC} - v_{AC}, v_{AC} + v_{BC} - v_{AB}\big). \quad (12.57)$$

If the three numbers on the right-hand side of (12.57) are positive, then $\mathbf{x}_0 := (x_A, x_B, x_C) \in \mathfrak{S}^\circ$, so Theorem 12.2 applies and shows that $\mathbf{p} \to \mathbf{x}_0$ given by (12.57). Note that Theorem 12.3 does not apply, but nevertheless we have the same conclusions, with a limit $\mathbf{x}_0$ that is a smooth function of the vote numbers by (12.57). Hence the condition (12.8) is not necessary for good behaviour.

Suppose now that, say, $v_{AB} = v_{AC} + v_{BC}$. Then $\mathbf{x}_0$ given by (12.57) has one coordinate 0 and lies thus on the boundary $\partial\mathfrak{S}$; nevertheless $\partial_i^* \psi(\mathbf{x}_0) = 0$ for every i, e.g. by (12.20), and as remarked after (12.22), this implies $\mathbf{x}_0 \in \mathcal{M}$. Furthermore, ψ is strictly concave on $\mathfrak{S}$, and thus $\mathcal{M} = \{\mathbf{x}_0\}$. Thus Theorem 12.5 applies and yields $\mathbf{p} \to \mathbf{x}_0$ in this case too.

Finally, suppose that $v_{AB} > v_{AC} + v_{BC}$. Then (12.57) would yield $x_C < 0$, so (12.6) has no solution in $\mathfrak{S}$. It is easy too see that the maximum of ψ on $\mathfrak{S}$ is attained on the part of the boundary with $x_C = 0$, and then we have $\Psi(x_A, x_B, 0) = x_A^{v_{AC}} x_B^{v_{BC}} (x_A + x_B)^{v_{AB}} = x_A^{v_{AC}} x_B^{v_{BC}}$, leading to the same equation as in Example 12.10 (with some changes in notation) and $\mathcal{M} = \{\mathbf{x}_0\}$ with

$$\mathbf{x}_0 = \Big(\frac{v_{AC}}{v_{AC} + v_{BC}}, \frac{v_{BC}}{v_{AC} + v_{BC}}, 0\Big). \quad (12.58)$$

Again, Theorem 12.5 applies and yields $\mathbf{p} \to \mathbf{x}_0$.

In fact, it is easy to see that if $v_{AB} \geqslant v_{AC} + v_{BC}$, then party C will never get any seat by Thiele's method (we omit the details). Hence, we may in this case ignore C and the result follws by Example 12.10. (Note that in the case $v_{AB} = v_{AC} + v_{BC}$, (12.57) and (12.58) yield the same result.)

Example 12.17. Supose that there are three parties A, B, C, and no votes AC or BC. Then (12.6) becomes

$$\frac{v_A}{x_A} = \frac{v_B}{x_B}, \qquad \frac{v_A}{x_A} + \frac{v_{AB}}{x_A + x_B} = \frac{v_C}{x_C} \quad (12.59)$$

which leads to

$$\frac{v_A + v_B + v_{AB}}{x_A + x_B} = \frac{v_C}{x_C} \quad (12.60)$$

and thus to $x_C = v_C / v^*$ and, e.g., $x_A = \frac{v_A}{v_A + v_B}(1 - x_C)$. This is generalized in Theorem 12.18.

Theorem 12.18. *Suppose that the N parties are partitioned into a number of blocks $\mathcal{P}_1, \mathcal{P}_2, \dots,$ and that each voter votes for a subset of one of the blocks. Assume also, for simplicity, assume that* (12.8) *holds. Then* $\mathbf{p}$ *converges to a limit $\mathbf{x} = (x_1, \dots, x_N)$ such that if $\mathcal{P}_j$ is one of the blocks, $q_j = \sum_{\sigma \subseteq \mathcal{P}_j} v_\sigma / v^*$ is the proportion of votes for this block, and x'_i is the asymptotic proportion of seats assigned to party i if the election is restricted to the parties in $\mathcal{P}_j$ only (with the same votes for them), then $x_i = q x'_i$ for every party $i \in \mathcal{P}_j$.*

Proof. It suffices to consider the case of two blocks, $\mathcal{P}_1$ and $\mathcal{P}_2$. For $j = 1, 2$, let $N_j := |\mathcal{P}_j|$, the number of parties in block $\mathcal{P}_j$, let $v^*_j := \sum_{\sigma \subseteq \mathcal{P}_j} v_\sigma$, the number of votes for block $\mathcal{P}_j$, let $z_j := \sum_{i \in \mathcal{P}_j} x_i$ and, for $i \in \mathcal{P}_j$, $y_i := x_i / z_j$. If $\sigma \subseteq \mathcal{P}_j$, then thus $x_\sigma = \sum_{i \in \sigma} z_j y_i = z_j y_\sigma$. Consequently, if $\mathbf{y}^j := (y_i)_{i \in \mathcal{P}_j} \in \mathfrak{S}_{N_j}$, and ψ^j denotes ψ defined as in (12.12) but for the votes $\sigma \subseteq \mathcal{P}_j$ only, then

$$\psi(\mathbf{x}) = \sum_j \sum_{\sigma \subseteq \mathcal{P}_j} v_\sigma (\log z_j + \log y_i) = \sum_j v^*_j \log z_j + \sum_j \psi^j(\mathbf{y}^j). \qquad (12.61)$$

Evidently, this is maximized by maximizing each $\psi^j(\mathbf{y}^j)$ separately, and maximizing $v^*_1 \log z_1 + v^*_2 \log z_2$ subject to $z_1 + z_2 = 1$; the latter leads to $z_j = v^*_j / v^*$, and the result follows. □

Remark 12.19. Theorem 12.18 is very natural and satisfying. It means for example that a party cannot influence its shares of seats by tactically splitting into several parts and distributing votes among combinations of them in some clever way.

However, this asymptotic result does *not* hold for small numbers of seats. In fact, one of the main problems with Thiele's method when used in Sweden in the 1910's (see the Historical note in Section 1) was the possibility of such manoeuvres. An example by [32] (see also [20]) is 3 seats and the votes 37 ABC, 13 KLM; in this case Thiele's method reduces to D'Hondt's, and gives two seats to ABC and one to KLM. However, if the ABC voters split their votes as 1 A, 9 AB, 9 AC, 9 B, 9 C then all three seats go to ABC. (Of course, Tenow considers individual candidates and not parties, but for this example the result is the same.)

It thus seems that Thiele's method behaves better asymptotically than for a finite number of seats.

References

[1] M. L. Balinski and H. P. Young, *Fair Representation.* 2nd ed., Brookings Institution Press, Washington, D.C., 2001.

[2] P. Billingsley, *Convergence of Probability Measures.* Wiley, New York, 1968.

[3] J. Brémont, Dynamics of injective quasi-contractions, *Ergodic Theory & Dynam. Systems* **26** (2006), 19–44.

[4] H. Bruin and J.H.B. Deane, Piecewise contractions are asymptotically periodic, *Proc. Amer. Math. Soc.* **137**(4) (2009), 1389–1395.

[5] Y. Bugeaud, Dynamique de certaines applications contractantes, linéaires par morceaux, sur $[0, 1)$. *C. R. Acad. Sci. Paris Sér. I Math.* **317** (1993), no. 6, 575–578.
[6] Y. Bugeaud, J.-P. Conze, Calcul de la dynamique de transformations linéaires contractantes mod 1 et arbre de Farey, *Acta Arithm.* **LXXXVIII**.3 (1999), 201–218.
[7] Y. Bugeaud and J.-P. Conze, Dynamics of some contracting linear functions modulo 1, *Noise, Oscillators and Algebraic Randomness (Chapelle des Bois, 1999)*, 379–387, Lecture Notes in Phys., **550**, Springer, Berlin, 2000.
[8] E. Catsigeras, P. Guiraud, A. Meyroneinc and E. Ugalde, On the asymptotic properties of piecewise contracting maps, *Dyn. Sys.* **31** (2016), no. 2, 107–135.
[9] R. Coutinho, *Dinâmica Simbólica Linear*, Ph.D. Thesis, Instituto Superior Técnico, Universidade de Lisboa, 1999.
[10] R. Coutinho, B. Fernandez, R. Lima, and A. Meyroneine, Discrete time piecewise affine models of genetic regulatory networks, *J. Math. Biol.* **52** (2006), 524–570.
[11] V. D'Hondt, *Question électorale: La représentation proportionnelle des partis, par un électeur*. Bruylant, Brussels, 1878.
[12] V. D'Hondt, *Système pratique et raisonné de représentation proportionnelle*, Muquardt, Brussels, 1882.
[13] E. J. Ding and P. C. Hemmer, Exact treatment of mode locking for a piecewise linear map. *J. Statist. Phys.* **46** (1987), no. 1-2, 99–110.
[14] O. Feely and L. O. Chua, The effect of the integrator leak in Σ–Δ modulation, *IEEE Transactions on Circuits and Systems* **38** (1991), 1293–1305.
[15] S. von Friesen, G. Appelberg, I. Bendixson, E. Phragmén, *Betänkande angående ändringar i gällande bestämmelser om den proportionella valmetoden.* Commission report, 3 December 1913, Stockholm, 1913.
[16] M. Gallagher and P. Mitchell (Eds.), *The Politics of Electoral Systems.* Oxford Univ. Press, Oxford, 2005.
[17] J.-M. Gambaudo and C. Tresser, On the dynamics of quasi-contractions, *Bull. Braz. Math. Soc.* **19**(1) (1988), 61–114.
[18] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers.* 4th ed., Oxford, at the Clarendon Press, 1960.
[19] S. Janson, Asymptotic bias of some election methods. *Annals of Operations Research*, **215** (2014), no. 1, 89–136.
[20] S. Janson, Phragmén's and Thiele's election methods, preprint, 2016: `arXiv:1611.08826`.
[21] M. Laurent and A. Nogueira, Rotation number of interval contracted rotations, preprint 2017: `arXiv:1704.05130v2`.
[22] J. H. Loxton and A. J. van der Poorten, Arithmetic properties of certain functions in several variables III, *Bull. Austral. Math. Soc.* **16** (1977), 15–47.
[23] X. Mora and M. Oliver, Eleccions mitjançant el vot d'aprovació. El mètode de Phragmén i algunes variants. *Butlletí de la Societat Catalana de Matemàtiques* **30** (2015), no. 1, 57–101.

[24] A. Nogueira and B. Pires, Dynamics of piecewise contractions of the interval, *Ergodic Theory & Dynam. Systems* **35**(7) (2015), 2198–2215.

[25] A. Nogueira, B. Pires and R.A. Rosales, Topological dynamics of piecewise λ-affine maps, *Ergodic Theory & Dynam. Systems*, to appear.

[26] E. Phragmén, Sur une méthode nouvelle pour réaliser, dans les élections, la représentation proportionelle des partis. *Öfversigt av Kongl. Vetenskaps-Akademiens Förhandlingar 1894*, N:o 3, Stockholm, 133–137.

[27] E. Phragmén, *Proportionella val. En valteknisk studie.* Svenska spörsmål 25, Lars Hökersbergs förlag, Stockholm, 1895.

[28] E. Phragmén, Sur la théorie des élections multiples, *Öfversigt av Kongl. Vetenskaps-Akademiens Förhandlingar 1896*, N:o 3, Stockholm, 181–191.

[29] E. Phragmén, Till frågan om en proportionell valmetod. *Statsvetenskaplig Tidskrift* **2** (1899), nr 2, 297–305. `http://cts.lub.lu.se/ojs/index.php/st/article/view/1949`

[30] F. Pukelsheim, *Proportional Representation. Apportionment Methods and Their Applications*, Springer, Cham, Switzerland, 2014.

[31] A. Sainte-Laguë, La représentation proportionnelle et la méthode des moindres carrés. *Ann. Sci. École Norm. Sup.* (3) **27** (1910), 529–542. Summary: *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, **151** (1910), 377–378.

[32] N. B. Tenow, Felaktigheter i de Thieleska valmetoderna. *Statsvetenskaplig Tidskrift* 1912, 145–165.

[33] T. N. Thiele, Om Flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger* 1895, København, 1895–1896, 415–441.

[34] P. Veerman, Symbolic dynamics of order-preserving orbits, *Physica D* **29** (1987), 191–201.

Svante Janson and Anders Öberg, Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

E-mail address: `svante.janson@math.uu.se; anders@math.uu.se`