THE THREE GAP THEOREM AND THE SPACE OF LATTICES

JENS MARKLOF AND ANDREAS STRÖMBERGSSON

Abstract. The three gap theorem (or Steinhaus conjecture) states that there are at most three distinct gap lengths in the fractional parts of the sequence $\alpha, 2\alpha, \ldots, N\alpha$, for any integer N and real number α. The statement was proved in the 1950s independently by various authors, see e.g. [N.B. Slater, Gaps and steps for the sequence $n\theta$ mod 1, *Proc. Camb. Phil. Soc.* **63** (1967), 1115–1123]. This hand-out presents a different proof using the space of two-dimensional Euclidean lattices.

For fixed $\alpha \in \mathbb{R}$, let $\xi_k = \{k\alpha\}$ be the fractional part of $k\alpha$. We are interested in the gaps between the elements of the sequence $(\xi_k)_{k=1}^N$ on $\mathbb{R}/\mathbb{Z}$. [These gaps are, in other words, the lengths of the N intervals that $\mathbb{R}/\mathbb{Z}$ is partitioned into by $(\xi_k)_{k=1}^N$. Shifting by $-\alpha$, this is the same as the lengths of the N intervals that $\mathbb{R}/\mathbb{Z}$ is partitioned into by $(\xi_k)_{k=0}^{N-1}$, and therefore the same as the lengths of the N intervals that $[0,1]$ is partitioned into by $(\xi_k)_{k=1}^{N-1}$.] The gap between ξ_k and its *next* neighbour on $\mathbb{R}/\mathbb{Z}$ (this is not necessarily the *nearest* neighbour, as the gap to the element preceding ξ_k may be the smaller one) is

$$\begin{aligned} s_{k,N} &= \min\{(\ell - k)\alpha + n \geq 0 \mid (\ell, n) \in \mathbb{Z}^2,\ 0 < \ell \leq N,\ \ell \neq k\} \\ &= \min\{m\alpha + n \geq 0 \mid (m,n) \in \mathbb{Z}^2,\ -k < m \leq N-k,\ m \neq 0\} \\ &= \min\{m\alpha + n \geq 0 \mid (m,n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\},\ -k < m \leq N-k\}. \end{aligned} \tag{1}$$

The last minimum is taken over a larger set than that in the second line, where the additional elements correspond to $m = 0$ and $n \neq 0$. For these values

$$\min\{m\alpha + n \geq 0\} = 1, \tag{2}$$

which means they do not contribute to the minimum in (1). This justifies the last equality in (1). Now

$$s_{k,N} = \frac{1}{N} \min\left\{ y \geq 0 \;\middle|\; (x,y) \in \mathcal{L} \setminus \{\mathbf{0}\},\ -\frac{k}{N} < x \leq 1 - \frac{k}{N} \right\}. \tag{3}$$

where

$$\mathcal{L} = \mathbb{Z}^2 M, \qquad M = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} N^{-1} & 0 \\ 0 & N \end{pmatrix}. \tag{4}$$

Date: July 29, 2015.

This hand-out has been prepared for the course "Applications of homogeneous dynamics: from number theory to statistical mechanics," ICTP Trieste, 27-31 July 2015. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 291147. A.S. is supported by a grant from the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine, and also by the Swedish Research Council Grant 621-2011-3629.

More generally, for every $M \in G := \mathrm{SL}(2,\mathbb{R})$ we obtain a lattice $\mathcal{L} = \mathbb{Z}^2 M$ of covolume one. If

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{5}$$

a basis of this lattice is given by

$$\boldsymbol{b}_1 = \boldsymbol{e}_1 M = (a,b), \qquad \boldsymbol{b}_2 = \boldsymbol{e}_2 M = (c,d), \tag{6}$$

where $\boldsymbol{e}_1 = (1,0)$, $\boldsymbol{e}_2 = (0,1)$ for the standard basis of $\mathbb{Z}^2$. All other bases of $\mathcal{L}$ with the same orientation can be obtained by replacing M by γM provided $\gamma \in \Gamma := \mathrm{SL}(2,\mathbb{Z})$. The space of lattices can in this way be identified with the coset space $\Gamma\backslash G$. The "modular group" Γ is a lattice in G.

For $M \in G$ and $0 < t \leq 1$ define

$$F(M,t) = \min\left\{ y \geq 0 \;\middle|\; (x,y) \in \mathbb{Z}^2 M \setminus \{\mathbf{0}\},\ -t < x \leq 1-t \right\}. \tag{7}$$

Then

$$s_{k,N} = \frac{1}{N} F\left(\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} N^{-1} & 0 \\ 0 & N \end{pmatrix}, \frac{k}{N} \right). \tag{8}$$

We first check F is well defined.

Proposition 1. *F is well defined as a function $\Gamma\backslash G \times (0,1] \to \mathbb{R}_{\geq 0}$.*

Proof. We first show that

$$\left\{ y \geq 0 \;\middle|\; (x,y) \in \mathbb{Z}^2 M \setminus \{\mathbf{0}\},\ -t < x \leq 1-t \right\} \tag{9}$$

is non-empty for every $M \in G$, $t \in (0,1]$. Let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{10}$$

and assume first that $a = 0$. Then $c \neq 0$ and $b = -1/c$, and (9) becomes

$$\left\{ bm + dn \geq 0 \;\middle|\; (m,n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\},\ -t < cn \leq 1-t \right\} \supset |b|\mathbb{N}, \tag{11}$$

which is non-empty. If $a \neq 0$, we have

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & ba^{-1} \\ 0 & 1 \end{pmatrix}, \tag{12}$$

and so (9) equals

$$\left\{ y + ba^{-1}x \geq 0 \;\middle|\; (x,y) \in \mathbb{Z}^2 \begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix} \setminus \{\mathbf{0}\},\ -t < x \leq 1-t \right\}. \tag{13}$$

Since $-t < x \leq 1-t$ implies $|x| \leq 1$, the set in (13) contains the set

$$\begin{aligned} &\left\{ y + ba^{-1}x \;\middle|\; (x,y) \in \mathbb{Z}^2 \begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix} \setminus \{\mathbf{0}\},\ -t < x \leq 1-t,\ y \geq |ba^{-1}| \right\} \\ &\quad = \left\{ bm + dn \;\middle|\; (m,n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\},\ -t < am + cn \leq 1-t,\ n \geq |b| \right\}. \end{aligned} \tag{14}$$

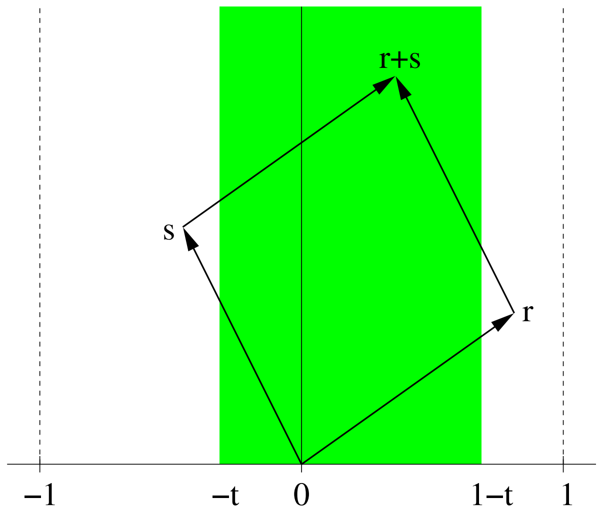

FIGURE 1. Illustration of the lattice configuration in the proof of Proposition 2

If c/a is rational, there exist $(m,n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}$ with $n \geq |b|$ such that $am + cn = 0$. If c/a is irrational, then the set $\{am + cn | (m,n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\},\ n \geq |b|\}$ is dense in $\mathbb{R}$. Therfore, in both cases, (14) is nonempty, and the minimum of (9) exists due to the discreteness of $\mathbb{Z}^2 M$.

Finally, we note that $F(\,\cdot\,,t)$ is well defined on $\Gamma\backslash G$ since $F(M,t) = F(\gamma M,t)$ for all $M \in G$, $\gamma \in \Gamma$. □

The following assertion implies the classical three gap theorem.

Proposition 2. *For every given M, the function $F(M,\cdot\,)$ is piecewise constant and takes at most three distinct values. If there are three values, then the third is the sum of the first and second.*

Proof. Among all points of the set $\mathcal{L} \setminus \{\mathbf{0}\}$ with $\mathcal{L} = \mathbb{Z}^2 M$ in the region $\mathcal{A} = (-1,1) \times [0,\infty)$, let $\boldsymbol{r} = (r_1, r_2)$ be a point with minimal second coordinate r_2. See Figure 1. Let us assume $r_2 > 0$ (the case $r_2 = 0$ is treated at the end of the proof). Next let $\boldsymbol{s} = (s_1, s_2)$ be a point in $\mathcal{A} \cap \mathcal{L} \setminus \mathbb{Z}\boldsymbol{r}$ with s_2 minimal. Then $s_2 \geq r_2 > 0$.

The choice of $\boldsymbol{r}$ and $\boldsymbol{s}$ implies that the closed triangle with vertices $\mathbf{0}, \boldsymbol{r}, \boldsymbol{s}$ does not contain an additional lattice point. It follows that the parallelogram with vertices $\mathbf{0}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{r}+\boldsymbol{s}$ does not contain an additional lattice point, and hence $\boldsymbol{r}, \boldsymbol{s}$ form a basis of $\mathcal{L}$.

Note that r_1 and s_1 must have opposite signs, i.e. $r_1 s_1 < 0$, since otherwise $\boldsymbol{s} - \boldsymbol{r} \in \mathcal{A}$ with a second coordinate that is smaller than s_2, contradicting the assumed minimality of s_2. It follows that if we set $\mathcal{J}_{\boldsymbol{r}} = (0,1] \cap (-r_1, 1-r_1]$ and $\mathcal{J}_{\boldsymbol{s}} = (0,1] \cap (-s_1, 1-s_1]$, then one of these intervals is of the form $(0,q]$ and the other is of the form $(q',1]$, for some $q, q' \in (0,1)$. Now in view of definition (7), we obtain

$$F(M,t) = \begin{cases} r_2 & \text{if } t \in \mathcal{J}_{\boldsymbol{r}} \\ s_2 & \text{if } t \in \mathcal{J}_{\boldsymbol{s}} \setminus \mathcal{J}_{\boldsymbol{r}} \\ r_2 + s_2 & \text{if } t \in (0,1] \setminus (\mathcal{J}_{\boldsymbol{r}} \cup \mathcal{J}_{\boldsymbol{s}}). \end{cases} \tag{15}$$

(Here the set $(0,1] \setminus (\mathcal{J}_{\boldsymbol{r}} \cup \mathcal{J}_{\boldsymbol{s}})$ may be empty.) Thus, for any fixed M, the function $F(M,\cdot\,)$ can only take one of the three values $r_2, s_2, r_2 + s_2$.

Now consider the remaining case $r_2 = 0$. Let us then also require that $\boldsymbol{r}$ is a primitive lattice point, and again let $\boldsymbol{s} = (s_1, s_2)$ be a point in $\mathcal{A} \cap \mathcal{L} \setminus \mathbb{Z}\boldsymbol{r}$ with s_2 minimal (then $s_2 > 0$). If $|r_1| \leq 1/2$ then $F(M,t) = 0$ for all $t \in (0,1]$. On the other hand, if $|r_1| > 1/2$ then $F(M,t) = s_2$ for $t \in (1-|r_1|, |r_1|]$ and $F(M,t) = 0$ for all other t in $(0,1]$. □

School of Mathematics, University of Bristol, Bristol BS8 1TW, U.K.
j.marklof@bristol.ac.uk

Department of Mathematics, Box 480, Uppsala University, SE-75106 Uppsala, Sweden
astrombe@math.uu.se