A geometric approach to the diophantine Frobenius problem

Christian Blatter

ABSTRACT. It turns out that all instances of the diophantine Frobenius problem for three coprime a_i have a common geometric structure which is independent of arithmetic coincidences among the a_i. By exploiting this structure we easily obtain Johnson's formula for the largest non-representable z, as well as a formula for the number of such z. A procedure is described which computes these quantities in $O\bigl(\log(\max a_i)\bigr)$ steps.

1. INTRODUCTION

For an n-tuple $\mathbf{a}=(a_1,a_2,\ldots,a_n)$ of positive integers we denote by $T:=T(\mathbf{a})$ the set of natural numbers z that can be written in the form

$$z=\sum_{k=1}^n x_k\,a_k\ , \qquad x_k\in\mathbb{N}:=\{0,1,2\ldots\}\ ,$$

and by $F:=F(\mathbf{a})$ the set of natural numbers z that cannot be so represented. If $\gcd(a_1,\ldots,a_n)=1$ then it is easily seen that all sufficiently large numbers z are in T. It follows that in this case F is a finite set, so there is a largest non-representable number $g(\mathbf{a}):=\max F$. To compute this number and maybe even the cardinality $N(\mathbf{a})$ of F in terms of a_1, ..., a_n constitutes the so-called *diophantine Frobenius problem*. We recommend [4], printed in 2005, as a comprehensive source of material about this problem; the bibliography alone contains about 500 items.

The case $n=2$ was first considered and solved by Sylvester [6], [7]. He proved:

Proposition 1. *If a_1, a_2 are >1 and coprime then*

$$g(a_1,a_2)=a_1\,a_2-a_1-a_2\ , \qquad N(a_1,a_2)=\frac{(a_1-1)(a_2-1)}{2}\ . \tag{1}$$

The proof of this result follows from inspection of Fig. 1 and is given at the beginning of the next section.

This paper deals with the case $n=3$. We shall give a natural geometric description of the set F from which an explicit answer to the Frobenius-3-problem can immediately be read off. In order to formulate our result we introduce the quantities

$$l_3:=\min\{l\in\mathbb{N}_{>0}\,|\,l\,a_3\in T(a_1,a_2)\}\qquad \circlearrowleft\ .$$

Here and in the sequel the sign $\circlearrowleft$ indicates that there are three such formulae in all, whereby the other two are obtained by cyclic permutation $1\to2\to3\to1$ of the indices. About the l_i the following can be said right away (cf. [2], Theorem 3 and eq. 26):

Proposition 2. *Assume that the three numbers a_1, a_2, a_3 are pairwise prime and that $l_i \geq 2$ $(1 \leq i \leq 3)$. Then the minimal representation*

$$l_3\, a_3 = x_{31}\, a_1 + x_{32}\, a_2\ , \quad x_{31}, x_{32} \in \mathbb{N} \qquad \circlearrowleft \tag{2}$$

of $l_i\, a_i$ $(1 \leq i \leq 3)$ is uniquely determined, and one has

$$x_{ij} \geq 1 \qquad (\text{all } i \neq j)\ . \tag{3}$$

Furthermore the l_i are coupled to the x_{ij} through

$$l_3 = x_{13} + x_{23} \qquad \circlearrowleft\ . \tag{4}$$

We now state our main result; it will be proven in section 3:

Theorem 3. *Assume that the three numbers a_1, a_2, a_3 are pairwise prime and that $l_i \geq 2$ $(1 \leq i \leq 3)$. Then*

$$g(\mathbf{a}) = l_1\, l_2\, l_3 + \max\{x_{12}x_{23}x_{31},\ x_{21}x_{32}x_{13}\} - \sum\nolimits_i a_i\ ; \tag{5}$$

$$N(\mathbf{a}) = \frac{1}{2}\Big(\sum_i (l_i - 1)a_i\ -\ l_1\, l_2\, l_3 + 1\Big)\ . \tag{6}$$

The assumption $l_i \geq 2$ means that none of the a_i is "superfluous". If, e.g., $l_3 = 1$ then $F(a_1, a_2, a_3) = F(a_1, a_2)$; this case is handled in Proposition 1. When the given a_i are not pairwise prime then there is a way to get rid of common factors, see [2], Theorem 2. Our formula (5), resp. its preliminary version (8), appears as Theorem 4 in [2] and on p. 35 of [4]. Note, however, that the proof given in [4] uses heavy algebraic machinery and is deferred to a later chapter.

Example 1. Let $a_1 := 2n-1$, $a_2 := 2n$, $a_3 := 2n+1$ for an $n \geq 2$. As $a_2 \equiv 1$ and $a_3 \equiv 2 \pmod{a_1}$ the smallest multiple of a_1 in $T(a_2, a_3)$ is $a_2 + (n-1)a_3$; similarly the smallest multiple of a_3 in $T(a_1, a_2)$ is $na_1 + a_2$, and obviously the minimal representation of a_2 is $2a_2 = a_1 + a_3$. Altogether we have

$$l_1 = n+1,\ x_{12} = 1,\ x_{13} = n-1;\ l_2 = 2,\ x_{21} = 1,\ x_{23} = 1;\ l_3 = n,\ x_{31} = n,\ x_{32} = 1\ ;$$

so Theorem 3 gives

$$g(\mathbf{a}) = 2n(n+1) + \max\{n, n-1\} - 6n = 2n^2 - 3n\ ,$$
$$N(\mathbf{a}) = \frac{1}{2}\big(n(2n-1) + 2n + (n-1)(2n+1) - 2n(n+1) + 1\big) = n^2 - n\ .$$

For the computation of the l_i and the x_{ij} we propose the so called Lagrange algorithm – a kind of two-dimensional euclidean algorithm modeled after a Gram-Schmidt-process – which takes $O\big(\log(\max a_i)\big)$ steps. The resulting procedure is developped in sections 4 and 5 of this paper. In [4] several other algorithms for $g(\mathbf{a})$ are described, notably the algorithm of Rødseth [5] which is an improved version of an earlier continued fraction algorithm by Selmer & Beyer.

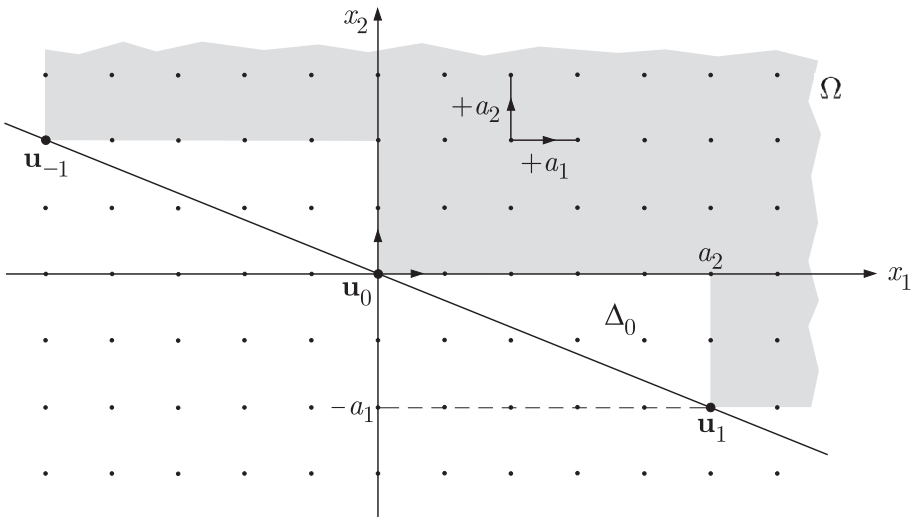

Fig. 1

2. PRELIMINARIES

Proof of Proposition 1. We draw in the (x_1, x_2)-plane the directed graph Γ with vertex set $\mathbb{Z}^2$ and edges of unit length connecting neighboring lattice points in the direction of increasing x_1 resp. x_2, see Fig. 1. For given a_1, $a_2 \in \mathbb{N}_{>1}$ we define the *height function*

$$f(x_1, x_2) \;:=\; a_1\, x_1 + a_2\, x_2 \;.$$

Two points in $\mathbb{Z}^2$ have the same height iff they differ by a vector $\mathbf{u} \in L := \mathbb{Z}^2 \cap f^{-1}(0)$. Since a_1 and a_2 are coprime the set L is the one-dimensional lattice formed by the vectors $\mathbf{u}_k := (ka_2, -ka_1),\ k \in \mathbb{Z}$.

An integer $z \in \mathbb{N}$ can be represented in the form $z = x_1\, a_1 + x_2\, a_2$ with x_1, $x_2 \in \mathbb{N}$ iff there is a directed edge path in Γ connecting a point $\mathbf{u}_k \in L$ with a point $\mathbf{x} \in \mathbb{Z}^2$ of height $f(\mathbf{x}) = z$. Now the lattice points that can be reached from a given $\mathbf{u}_k \in L$ lie in the set $Q_k := \mathbf{u}_k + \mathbb{R}^2_{\geq 0}$, a first quadrant with origin at $\mathbf{u}_k$, and the set of all possible end points of such paths consists of the lattice points in the union $\Omega := \bigcup_k Q_k$ of these quadrants.

The lattice points of positive height that cannot be reached from one of the points $\mathbf{u}_k$ are the interior lattice points of the rectangular triangles Δ_k with vertices $\mathbf{u}_k$, $\mathbf{u}_{k+1}$, $\mathbf{u}_k + (a_2, 0)$. The largest occurring height in such a Δ_k is given by the first formula (1), and using symmetry one concludes that each Δ_k contains exactly $(a_1 - 1)(a_2 - 1)/2$ lattice points in its interior, which all have different heights. □

For later purposes we note the following: Any lattice point in Δ_0 can be connected by an admissible path to the point $(a_2, 0)$ of height a_1a_2. It follows that a number $z > 0$ is in $F(a_1, a_2)$ iff there are integers k_1, $k_2 \geq 1$ such that $z = a_1a_2 - k_1a_1 - k_2a_2$.

Proof of Proposition 2. First we show (3). Assume, e.g., that $x_{13} = 0$. Then we have $l_1a_1 = x_{12}a_2$, and as a_1, a_2 are coprime it follows that $l_1 \geq a_2$. On the other hand, from $l_3 \geq 2$, i.e., $a_3 \in F(a_1, a_2)$ it follows that there are integers k_1, $k_2 \geq 1$ with

$$a_3 = a_1a_2 - k_1a_1 - k_2a_2 = (a_2 - k_1)a_1 - k_2a_2 \;.$$

Whence we would have $(a_2 - k_1)a_1 = k_2 a_2 + a_3$ which would imply $l_1 < a_2$ – a contradiction.

We next show (4). Let the x_{ij} be determined such that (2) holds. Then the quantities

$$\mu_3 := l_3 - x_{13} - x_{23} \qquad \circlearrowleft$$

satisfy $\mu_1 a_1 + \mu_2 a_2 + \mu_3 a_3 = 0$. If the μ_i do not all vanish then up to a permutation of the a_i we must have one of the following:

$$\text{(a)} \qquad \mu_1 > 0\,, \quad \mu_2 < 0\,, \quad \mu_3 \le 0\,; \qquad\qquad \text{(b)} \qquad \mu_1 > 0\,, \quad \mu_2 > 0\,, \quad \mu_3 < 0\,.$$

In case (a) it follows that

$$(l_1 - x_{21} - x_{31})a_1 = \mu_1 a_1 = (-\mu_2)a_2 + (-\mu_3)a_3\,,$$

contradicting the definition of l_1. In case (b), from

$$(x_{13} + x_{23} - l_3)a_3 = -\mu_3 a_3 = \mu_1 a_1 + \mu_2 a_2$$

it follows by definition of l_3 that $x_{13} + x_{23} \ge 2l_3$, whence, e.g., $x_{13} \ge l_3$. By definition of the x_{ij} we now have the representation

$$(l_1 - x_{31})a_1 = (x_{12} + x_{32})a_2 + (x_{13} - l_3)a_3$$

which again contradicts the definition of l_1.

As (4) is true for all possible choices of x_{21} and x_{31} consistent with the definition of l_2 and l_3, and as these choices can be made independently, it follows that there is in fact no choice at all, which means that the x_{ij} are indeed uniquely determined. □

3. PROOF OF THE MAIN RESULT

We now come to the proof of Theorem 3. Inspired by the proof of Sylvester's result for $n = 2$ we embed the problem into the following geometric setup: Consider the integer lattice $\mathbb{Z}^3$ in euclidean (x_1, x_2, x_3)-space $\mathbb{R}^3$. We use $\mathbb{Z}^3$ as vertex set of a directed graph Γ whose edges are the segments of unit length connecting neighboring lattice points in the direction of increasing x_1, resp. x_2, x_3. Given a_1, a_2, a_3, we again define the *height function*

$$f(x_1, x_2, x_3) := a_1\, x_1 + a_2\, x_2 + a_3\, x_3$$

which on the one hand is just a linear functional on $\mathbb{R}^3$ and on the other hand assigns a height $f(\mathbf{x})$ to each lattice point $\mathbf{x} \in \mathbb{Z}^3$. The kernel $H := f^{-1}(0)$ of f is a plane through the origin of $\mathbb{R}^3$ and contains the *Frobenius lattice* $L := H \cap \mathbb{Z}^3$ of integer solutions to the equation $f(\mathbf{x}) = 0$.

Lemma 4. (a) *Let $m_1 a_1 + m_2 a_2 = 1$ with $m_i \in \mathbb{Z}$. Then the vectors $\mathbf{e}_1 := (a_2, -a_1, 0)$, $\mathbf{e}_2 := (a_3 m_1, a_3 m_2, -1)$ form a basis of L.*

(b) *The fundamental domain of the lattice L, when projected to the plane $x_i = 0$, has area a_i $(1 \le i \le 3)$.*

Proof. (a) One easily checks that $f(\mathbf{e}_1) = f(\mathbf{e}_2) = 0$, which means that $\mathbf{e}_1,\ \mathbf{e}_2 \in L$. On the other hand, let $\mathbf{u} = (u_1, u_2, u_3)$ be an arbitrary point of L. Then $\mathbf{u} + u_3\mathbf{e_2} = (u_1', u_2', 0) \in L$ which implies $a_1u_1' + a_2u_2' = 0$. Since a_1, a_2 are coprime it follows that $(u_1', u_2', 0) = k\mathbf{e}_1$ for a $k \in \mathbb{Z}$, whence we have $\mathbf{u} = k\mathbf{e}_1 - u_3\mathbf{e}_2$.

(b) It suffices to compute the vector product

$$\mathbf{e}_1 \times \mathbf{e_2} = \big(a_1, a_2, (m_1a_1 + m_2a_2)a_3\big) = (a_1, a_2, a_3)\ . \qquad \square$$

The three lattice vectors

$$\mathbf{f}_1 := (l_1, -x_{12}, -x_{13}),\ \mathbf{f}_2 := (-x_{21}, l_2, -x_{23}),\ \mathbf{f}_3 := (-x_{31}, -x_{32}, l_3)\ \in L$$

encoding the data l_i, x_{ij} will play a special rôle. We shall call any vector of the form $\mathbf{f}_i$ or $-\mathbf{f}_i$ a *basic vector* and any set of three essentially different vectors among the $\pm\mathbf{f}_i$ a *solution basis* for the Frobenius problem at hand.

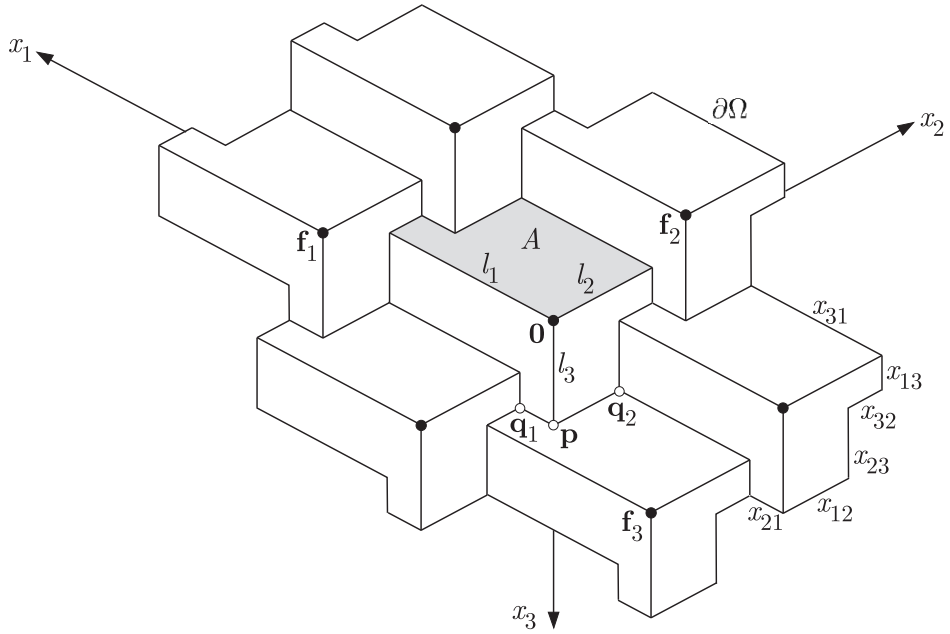

Fig. 2

The connection of the three-dimensional structure described so far with the diophantine Frobenius problem is the following: A natural number z is in $T(\mathbf{a})$ iff there is a lattice point $\mathbf{u} \in L$ and a directed edge path in Γ beginning at $\mathbf{u}$ and ending in a point $\mathbf{x} \in \mathbb{Z}^3$ of height $f(\mathbf{x}) = z$. Now the lattice points that can be reached from a given $\mathbf{u} \in L$ lie in the set $O_\mathbf{u} := \mathbf{u} + \mathbb{R}^3_{\geq 0}$, an octant with origin at $\mathbf{u}$, and the set of all possible end points of such paths consists of the lattice points in the union $\Omega := \bigcup_{\mathbf{u}\in\mathbf{L}} O_\mathbf{u}$ of these octants. It follows that $T(\mathbf{a}) = f(\mathbb{Z}^3 \cap \Omega)$. The boundary $\partial\Omega$ is L-periodic; it consists of three L-shapes per octant and touches the plane H in the points of L. It looks like a washboard and is depicted in Fig. 2, *as seen from below*. The proof of Theorem 3 essentially consists in understanding this figure.

Consider, e.g., the positive x_3-axis. It is an edge of the octant $O_{\mathbf{0}}$ and belongs to the boundary of Ω until it is intercepted at $\mathbf{p} = (0, 0, u_3)$ by a wall $x_3 = \text{const.}$ belonging to another octant $O_{\mathbf{u}}$, $\mathbf{u} = (u_1, u_2, u_3) \in L$. For this to happen it is necessary that $u_1 \le 0$, $u_2 \le 0$, $u_3 > 0$, and that there is no point $\mathbf{u}' = (u_1', u_2', u_3') \in L$ with $u_1' \le 0$, $u_2' \le 0$ and $u_3' < u_3$. This means

$$\begin{aligned} u_3 &= \min\bigl\{u \in \mathbb{N}_{>0} \;\bigm|\; \exists\, u_1, u_2 \in \mathbb{Z}_{\le 0} \colon\; a_1 u_1 + a_2 u_2 + a_3 u = 0\bigr\} \\ &= \min\bigl\{u \in \mathbb{N}_{>0} \;\bigm|\; u\, a_3 \in T(a_1, a_2)\bigr\}\,, \end{aligned} \tag{7}$$

from which we deduce

$$u_3 = l_3\,, \quad -u_1 = x_{31}\,, \quad -u_2 = x_{32}\,,$$

i.e., $(u_1, u_2, u_3) = \mathbf{f}_3$; and similarly for the other l_i, x_{ij}.

The lattice points $\mathbf{x}$ of positive height that cannot be reached from a point $\mathbf{u} \in L$ are the interior lattice points contained in the region W enclosed between H and $\partial\Omega$. As seen in the figure, the restriction $f \restriction \partial\Omega$ takes local maxima at the points $\mathbf{q}_1$, $\mathbf{q}_2$, and the interior lattice points of maximal height are $\mathbf{q}_1 - (1,1,1)$ or $\mathbf{q}_2 - (1,1,1)$ and their equivalents $\bmod L$. It follows that the maximal non-realizable height $g(\mathbf{a})$ is given by

$$g(\mathbf{a}) = \max\{f(\mathbf{q}_1), f(\mathbf{q}_2)\} - \sum\nolimits_i a_i = l_3\, a_3 + \max\{x_{21}a_1, x_{12}a_2\} - \sum\nolimits_i a_i\ . \tag{8}$$

Since the three L-shapes have areas a_i by Lemma 4(b), we deduce from Fig. 2 that the a_i satisfy

$$a_1 = x_{12}l_3 + x_{13}x_{32}\,, \quad a_2 = x_{21}l_3 + x_{23}x_{31}\,, \quad a_3 = l_1l_2 - x_{12}x_{21} \qquad \circlearrowleft\ . \tag{9}$$

Substituting these expressions into (8) one arrives at the symmetric formula (5):

$$\begin{aligned} g(\mathbf{a}) &= l_1l_2l_3 - l_3x_{12}x_{21} + \max\{x_{21}x_{12}l_3 + x_{21}x_{13}x_{32}, x_{12}x_{21}l_3 + x_{12}x_{23}x_{31}\} - \sum\nolimits_i a_i \\ &= l_1l_2l_3 + \max\{x_{21}x_{13}x_{32}, x_{12}x_{23}x_{31}\} - \sum\nolimits_i a_i\ . \end{aligned}$$

We now come to the proof of formula (6). We have to count the number z_0 of interior lattice points in the quotient $\hat{W} := W/L$. As $\hat{W}$ does not have a simple description in terms of inequalities we are going to determine z_0 "from the outside" by means of a three-dimensional analog of Pick's area formula. Let z_1 denote the total number of relative interior lattice points in the three L-shapes; similarly, let z_2 be the total number of relative interior lattice points on the reentrant edges of $\hat{W}$ and z_3 be the number of such points on the protruding edges of $\hat{W}$. Then we have the following formula:

Lemma 5. $$\mathrm{vol}(\hat{W}) = z_0 + \frac{1}{2}z_1 + \frac{3}{4}z_2 + \frac{1}{4}z_3 + \frac{7}{4}\,,$$

whereby the last term incorporates the contribution of the corners of $\hat{W}$.

Proof. We perform a "Gedankenexperiment" used already in [1] for a proof of Pick's area formula. Imagine that at time 0 a unit of heat is concentrated at each point of $\mathbb{Z}^3$.

This heat will be distributed all over space by heat conduction, and at time ∞ it will be equally distributed in space with density 1. In particular, the amount of heat contained in $\hat{W}$ will be $\mathrm{vol}(\hat{W})$. Where does this amount of heat come from? For symmetry reasons there is absolutely no flux across the unit squares of $\partial\Omega$, and, again by symmetry, the net flux across H/L is 0 as well. As a consequence, the final amount of heat within $\hat{W}$ comes from the interior lattice points, counted by z_0, and from the lattice points on the boundary of $\hat{W}$. The lattice points counted by z_1 send half their heat into $\hat{W}$, whereas the corresponding factor is $\frac{3}{4}$ for the points counted by z_2 and $\frac{1}{4}$ for the points counted by z_3. Furthermore $\hat{W}$ possesses two protruding corners $\mathbf{q}_1$ and $\mathbf{q}_2$ which contribute $\frac{1}{8}$ each, three "L-corners" contributing $\frac{3}{8}$, and finally the reentrant corner on H which contributes $\frac{3}{8}$ as well. □

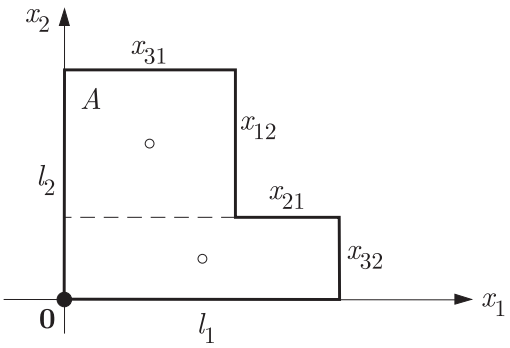

Fig. 3

In order to compute $\mathrm{vol}(\hat{W})$ directly we use the L-shape A in the plane $x_3 = 0$ as fundamental domain. Fig. 3 shows A, as seen from the positive x_3-direction. The vertical prism K determined by A and the plane

$$H: \qquad x_3 = -\frac{1}{a_3}(a_1x_1 + a_2x_2)$$

is a representative for $\hat{W}$. To compute the volume of K we split A into two rectangles and work with the heights of K in their midpoints. We obtain

$$\mathrm{vol}(\hat{W}) = \frac{x_{31}x_{12}}{a_3}\Big(a_1\frac{x_{31}}{2} + a_2\frac{2x_{32}+x_{12}}{2}\Big) + \frac{l_1x_{32}}{a_3}\Big(a_1\frac{l_1}{2} + a_2\frac{x_{32}}{2}\Big)\,,$$

which using (2) and (9) can be brought into the following symmetric form:

$$\mathrm{vol}(\hat{W}) = \frac{1}{2}\left(\sum\nolimits_i l_i\,a_i \; - l_1l_2l_3\right)\;. \tag{10}$$

We now compute the quantities z_1, z_2 and z_3. – The horizontal L-shape A in Fig. 3 has area a_3 by Lemma 4(b) and $2(l_1 + l_2)$ lattice points on its boundary. Therefore by Pick's area formula for the plane the number of interior lattice points on A is given by $a_3 - (l_1 + l_2) + 1$, and we obtain

$$z_1 = \sum\nolimits_i a_i - 2\sum\nolimits_i l_i + 3\;. \tag{11}$$

Inspection of Fig. 2 shows that

$$z_2 = \sum\nolimits_i (l_i - 1) = \sum\nolimits_i l_i - 3\,, \qquad z_3 = \sum\nolimits_{i\neq j} (x_{ij} - 1) = \sum\nolimits_i l_i - 6\,. \tag{12}$$

Introducing (10), (11) and (12) into Lemma 5 we get

$$\begin{aligned} z_0 &= \mathrm{vol}(\hat{W}) - \frac{1}{2}z_1 - \frac{3}{4}z_2 - \frac{1}{4}z_3 - \frac{7}{4} \\ &= \frac{1}{2}\Big(\sum\nolimits_i l_i\, a_i \; - l_1 l_2 l_3\Big) - \frac{1}{2}\Big(\sum\nolimits_i a_i - 2\sum\nolimits_i l_i + 3\Big) - \frac{3}{4}\Big(\sum\nolimits_i l_i - 3\Big) \\ &\quad - \frac{1}{4}\Big(\sum\nolimits_i l_i - 6\Big) - \frac{7}{4} \\ &= \frac{1}{2}\Big(\sum\nolimits_i (l_i - 1)a_i - l_1 l_2 l_3 + 1\Big)\;. \end{aligned}$$

This concludes the proof of Theorem 3. □

4. FINDING A SOLUTION BASIS

In order to make Theorem 3 useful we have to establish a procedure to compute the quantities l_i, x_{ij}. Our argument takes place in the plane H. Fig. 4 shows H as seen from the tip of the vector $\mathbf{a}$, the points of L are again marked by bullets. The three planes $x_i = 0$ intersect H in three lines g_i through the origin which altogether divide H into six sectors of various widths. The line g_3, spanned by the vector $\mathbf{v}_3 := (-a_2, a_1, 0)$, is at the same time a level line of the linear function x_3 restricted to H. Equation (7) can now be interpreted as follows: In order to find l_3 we have to translate the line $g_3\colon\ x_3 = 0$ in the direction of increasing x_3 (marked by an arrow in Fig. 4) until it hits for the first time a lattice point in the sector $x_1 \leq 0 \ \wedge\ x_2 \leq 0$. The lattice point obtained in this way is the point $\mathbf{f}_3$. Translating similarly the lines g_1 and g_2 one obtains the lattice points $\mathbf{f}_1$, $\mathbf{f}_2$ in the appropriate sectors.

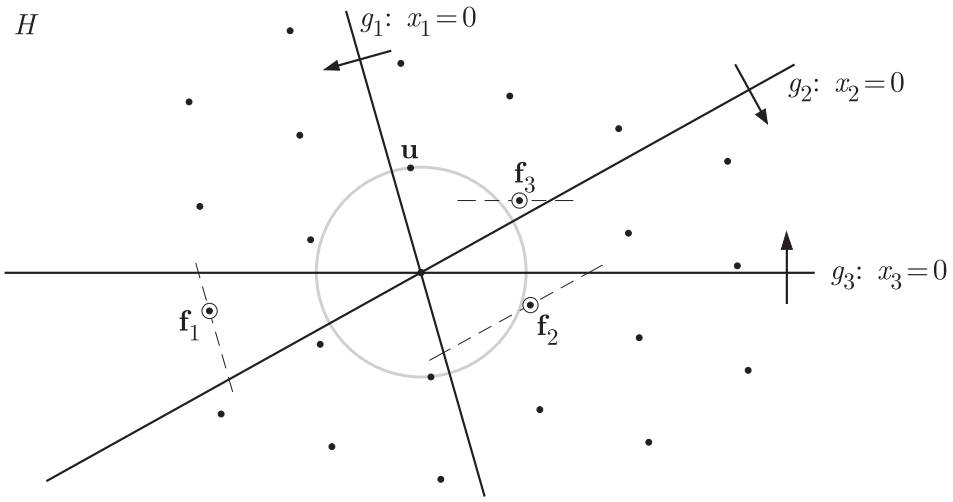

Fig. 4

Being minimizers of some sort the $\mathbf{f}_i \in L$ tend to be short. Now there exists a revered algorithm (attributed to Lagrange, Gauss and others, see [3]) which finds the shortest vector $\mathbf{u}$ of the lattice L, and we plan to make use of this algorithm. But if $\mathbf{u}$ happens to lie in a sector of width $> 60°$, as in Fig. 4, there is no guarantee that $\mathbf{u}$ coincides with the basic vector $\mathbf{f}_i$ (or $-\mathbf{f}_i$) in that sector. For this reason we change the metric in such a way that the three lines $x_i = 0$ intersect at angles of $60°$.

Lemma 6. *For a suitable scalar product* $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^\mathsf{T} Q\, \mathbf{y}$ *the three lines* $x_i = 0$ *in* H *intersect at angles of* $60°$.

Proof. The three directions in question are

$$\mathbf{v}_1 := (0, -a_3, a_2)\,, \quad \mathbf{v}_2 := (a_3, 0, -a_1)\,, \quad \mathbf{v}_3 := (-a_2, a_1, 0)\,.$$

Consider now the linear map $P : \mathbb{R}^3 \to \mathbb{R}^2$ given by the matrix

$$P := \begin{bmatrix} -a_1 & a_2 & 0 \\ 0 & 0 & \sqrt{3}a_3 \end{bmatrix}\,.$$

The kernel of P is spanned by the vector $(a_2, a_1, 0) \notin H$, therefore the restriction $P \restriction H$ maps H bijectively onto the euclidean plane $E := \mathbb{R}^2$. One easily computes

$$P\mathbf{v}_1 = 2a_2a_3\Big(-\frac{1}{2}, \frac{\sqrt{3}}{2}\Big)\,, \quad P\mathbf{v}_2 = 2a_1a_3\Big(-\frac{1}{2}, -\frac{\sqrt{3}}{2}\Big)\,, \quad P\mathbf{v}_3 = 2a_1a_2\,(1, 0)\,,$$

which shows that in the image plane the lines g_i intersect at angles of $60°$.

Pulling back the euclidean scalar product in E to H one obtains there the new scalar product

$$\langle \mathbf{x}, \mathbf{y} \rangle := (P\mathbf{x})^\mathsf{T}\, P\mathbf{y} = \mathbf{x}^\mathsf{T} P^\mathsf{T}\, P\, \mathbf{y} = \mathbf{x}^\mathsf{T}\, Q\, \mathbf{y}\,,$$

where the integer matrix $Q := P^\mathsf{T} P$ is given by

$$Q = \begin{bmatrix} a_1^2 & -a_1a_2 & 0 \\ -a_1a_2 & a_2^2 & 0 \\ 0 & 0 & 3a_3^2 \end{bmatrix}\,.$$

□

In what follows, $|\cdot|$ denotes the norm corresponding to the scalar product $\langle \cdot, \cdot \rangle$.

Lagrange's algorithm (to be described in the next section) produces a *reduced basis* $(\mathbf{u}, \mathbf{v})$ of the Frobenius lattice L. This means that $\mathbf{u}$ is a shortest nonzero vector in L and that $\mathbf{v}$ is a shortest vector in $L \setminus \mathbb{Z}\,\mathbf{u}$; in particular, $|\mathbf{u}| \le |\mathbf{v}|$.

Theorem 7. *Under the hypotheses of Theorem* 3, *let* $(\mathbf{u}, \mathbf{v})$ *be a reduced basis of the Frobenius lattice* L. *Assume that* $u_1 \le 0$, $u_2 \le 0$, $u_3 > 0$ *and put* $\lambda := v_3/u_3$. *Then the three vectors*

$$\mathbf{u}\,, \quad \mathbf{v}_- := \mathbf{v} - \lceil\lambda\rceil\mathbf{u}\,, \quad \mathbf{v}_+ := \mathbf{v} - \lfloor\lambda\rfloor\mathbf{u}$$

form a solution basis for the Frobenius problem determined by the a_i.

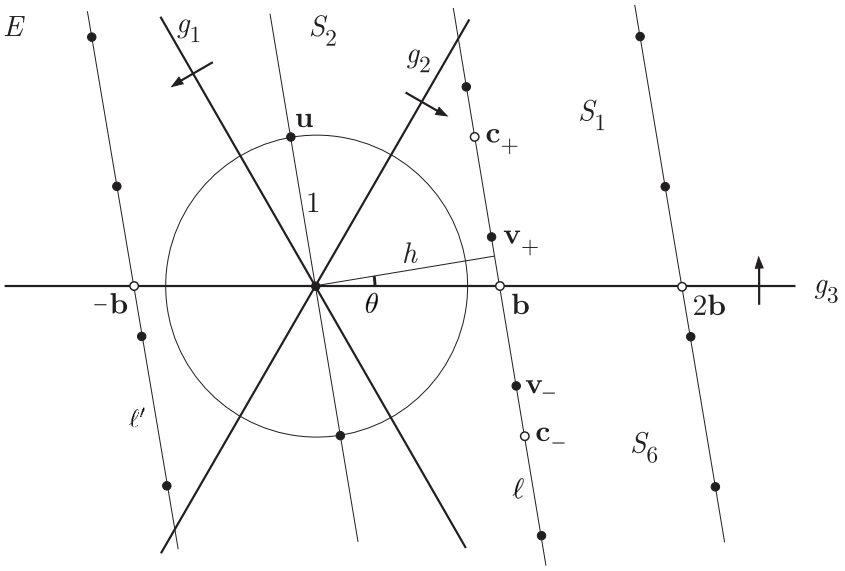

Fig. 5

Proof. We argue in the (x, y)-plane E, but omit the P in our notation. The lines g_i enclose angles of 60°, creating (closed) sectors S_k $(1 \le k \le 6)$, see Fig. 5. After scaling we have $\mathbf{u} = (-\sin\theta, \cos\theta)$ with $|\theta| \le \frac{\pi}{6}$, whence $\mathbf{u} \in S_2$, and we may assume that the lattice line $\ell \parallel \mathbf{u}$ containing $\mathbf{v}$ is to the right of $\mathbf{u}$. We shall show that the points $\mathbf{u}$, $\mathbf{v}_-$ and $\mathbf{v}_+$ are basic points in the sectors S_6, S_1 and S_2 respectively.

We begin with the remark that in fact $|\theta| < \pi/6$. Assume to the contrary that, e.g., $\theta = \pi/6$. Then $\mathbf{u} = (-1/2, \sqrt{3}/2) \in g_1$, meaning $u_1 = 0$. In this case $\mathbf{u}$ could not be basic by (3). It follows that there would have to be a lattice point in S_2 with y-coordinate $< \sqrt{3}/2$. But there is no room for such a point since the interior of the unit circle is forbidden.

Let $\mathbf{b} := \mathbf{v} - \lambda\mathbf{u} = (b, 0)$ be the point where ℓ intersects g_3. Then $b = h/\cos\theta$ where h denotes the distance from $\mathbf{0}$ to ℓ, whence $h \ge \sqrt{3}/2$. We write $(\cos\phi, \sin\phi) =: \mathbf{e}_\phi$.

Lemma 8. (a) $\langle \mathbf{e}_{\pi/6}, \mathbf{u}\rangle < \langle \mathbf{e}_{\pi/6}, \mathbf{b}\rangle$, (b) $\langle \mathbf{e}_{-\pi/6}, -\mathbf{u}\rangle < \langle \mathbf{e}_{-\pi/6}, \mathbf{b}\rangle$.

Proof. The left sides of (a) and (b) are $\sin\left(\frac{\pi}{6} - \theta\right)$ and $\sin\left(\frac{\pi}{6} + \theta\right)$ respectively, so they both are $\le \sin\left(\frac{\pi}{6} + |\theta|\right)$. On the other hand the right sides of (a) and (b) both have the same value

$$\cos\frac{\pi}{6}\, b = \cos\frac{\pi}{6}\,\frac{h}{\cos\theta} \ge \frac{3}{4\cos\theta}\ .$$

It remains to prove that for $0 \le \theta < \frac{\pi}{6}$ one has

$$2\sin\left(\frac{\pi}{6} + \theta\right)\cos\theta < \frac{3}{2}\ .$$

But here the left side can be written as $\sin\left(\frac{\pi}{6} + 2\theta\right) + \sin\left(\frac{\pi}{6}\right)$ which is $< 1 + \frac{1}{2}$. □

Put $\mathbf{c}_- := \mathbf{b} - \mathbf{u}$, $\mathbf{c}_+ := \mathbf{b} + \mathbf{u}$. Then from Lemma 8(a) it follows that $\langle \mathbf{e}_{\pi/6}, \mathbf{c}_-\rangle > 0$, whence $\mathbf{c}_- \in \mathrm{int}(S_6)$, and analogously from Lemma 8(b) it follows that $\langle \mathbf{e}_{-\pi/6}, \mathbf{c}_+\rangle > 0$,

whence $\mathbf{c}_+ \in \mathrm{int}(S_1)$. From this we conclude that the lattice points $\mathbf{v}_- \in [\mathbf{b}, \mathbf{c}_-]$ and $\mathbf{v}_+ \in [\mathbf{b}, \mathbf{c}_+]$ lie in S_6 and S_1 respectively.

As $\mathbf{c}_+ \in \mathrm{int}(S_1)$ the line ℓ intersects g_2 at a y-level $> \cos\theta$, whence all lattice points on $\ell \cap S_2$ have a larger y-level than $\mathbf{u}$, and similarly, as $-\mathbf{c}_- \in \mathrm{int}(S_3)$, the line ℓ' intersects g_1 at a y-level $> \cos\theta$, whence all lattice points on $\ell' \cap S_2$ have a larger y-level than $\mathbf{u}$. This implies that the vector $\mathbf{u}$ is basic in its sector S_2.

Note that $\langle \mathbf{e}_{\pi/6}, \mathbf{u} \rangle > 0$, whence going upwards along $\ell \cap S_1$ the distance to g_1 increases. Since $\mathbf{v}_+$ is the first lattice point met along this path, $\mathbf{v}_+$ is basic for the sector S_1, unless there were an even better lattice point on the parallel to ℓ through the point $2\mathbf{b}$. But the latter is prohibited by the inequality $\langle \mathbf{e}_{\pi/6}, \mathbf{c}_+ \rangle < \langle \mathbf{e}_{\pi/6}, 2\mathbf{b} \rangle$ which follows easily from Lemma 8(a).

Similarly one has $\langle \mathbf{e}_{-\pi/6}, \mathbf{u} \rangle < 0$, and this implies that going downwards along $\ell \cap S_6$ the distance to g_2 increases. Since $\mathbf{v}_-$ is the first lattice point met along this path, $\mathbf{v}_-$ is basic for the sector S_6, unless there were an even better lattice point on the parallel to ℓ through the point $2\mathbf{b}$. But the latter is prohibited by the inequality $\langle \mathbf{e}_{-\pi/6}, \mathbf{c}_- \rangle < \langle \mathbf{e}_{-\pi/6}, 2\mathbf{b} \rangle$ which follows easily from Lemma 8(b). □

5. LAGRANGE'S ALGORITHM

Lagrange's algorithm, as it is called in [3], takes an arbitrary basis $(\mathbf{u}, \mathbf{v})$ of the Frobenius lattice L as input and in a certain number of steps arrives at a reduced basis of L. An essential accessory to the calculations is the *Gram matrix*

$$G := G(\mathbf{u}, \mathbf{v}) := \begin{bmatrix} \langle \mathbf{u}, \mathbf{u} \rangle & \langle \mathbf{u}, \mathbf{v} \rangle \\ \langle \mathbf{v}, \mathbf{u} \rangle & \langle \mathbf{v}, \mathbf{v} \rangle \end{bmatrix}$$

of the current basis $(\mathbf{u}, \mathbf{v})$; it is updated along with the basis vectors.

The following box is taken from [3]. The subscript $\leq$ to a basis indicates that one assumes $|\mathbf{u}| \leq |\mathbf{v}|$, and $\lfloor \cdot \rceil$ denotes the nearest integer function.

Input: A basis $(\mathbf{u}, \mathbf{v})_{\leq}$ with its Gram matrix $G = (g_{i,j})_{1 \leq i,j \leq 2}$.
Output: A reduced basis of L with its Gram matrix.

1. Repeat
2. $\quad \mathbf{r} := \mathbf{v} - x\,\mathbf{u}$, where $x := \lfloor g_{1,2}/g_{1,1} \rceil$.
   $\quad$ When computing x, also compute the remainder $y := g_{1,2} - x g_{1,1}$.
3. $\quad \mathbf{v} := \mathbf{u}$.
4. $\quad \mathbf{u} := \mathbf{r}$.
5. $\quad$ Update the Gram matrix as follows: swap $g_{2,2}$ and $g_{1,1}$; then let $g_{1,2} := y$ and $g_{1,1} := g_{1,1} - x(y + g_{1,2})$.
6. Until $|\mathbf{u}| \geq |\mathbf{v}|$.
7. Return $(\mathbf{v}, \mathbf{u})_{\leq}$ and its Gram matrix (setting $g_{2,1} := g_{1,2}$).

Lagrange's algorithm

We now combine this with the results of the foregoing section in order to obtain a coherent description of the computational procedure to determine the l_i, x_{ij}.

When a_1, a_2, a_3 are given, one first has to set up the basis $(\mathbf{e}_1, \mathbf{e}_2)$ of L given in Lemma 4(a). This requires $O\big(\log(\max_i a_i)\big)$ steps for the euclidean algorithm to find m_1, m_2. Using this basis as input one then runs Lagrange's algorithm and obtains a reduced basis $(\mathbf{u}, \mathbf{v})_{\leq}$ of L. As shown in [3], Theorem 3.0.3, this is accomplished in at most $O\big(\log(\max_i a_i)\big)$ loops of the algorithm. Replacing $\mathbf{u}$ by $-\mathbf{u}$, if necessary, makes $u_i > 0$, $u_j < 0$ $(j \neq i)$ for some i. Now put $\lambda := v_i/u_i$ and define $\mathbf{v}_-$, $\mathbf{v}_+$ as given in Theorem 7. The quantities l_i, x_{ij} can then be read off from the coordinates of the three vectors $\mathbf{u}$, $\mathbf{v}_-$, $\mathbf{v}_+$. Note however that the bit complexity of the whole computation is quadratic insofar as the bit-length of the input data a_i not only affects the number of required steps/loops but also the cost of each step.

Example 2. Consider the random numerical example $a_1 := 4327$, $a_2 := 6716$, $a_3 := 9237$. In 8 steps the euclidean algorithm finds $2055a_1 - 1324a_2 = 1$, and after 6 loops of Lagrange's algorithm we arrive at a reduced basis of L given by $\mathbf{u} = (-53, -47, 59)$, $\mathbf{v} = (-130, 59, 18)$. Theorem 7 then tells us that

$$\mathbf{f}_1 = (130, -59, -18)\,, \quad \mathbf{f}_2 = (-77, 106, -41)\,, \quad \mathbf{f}_3 = (-53, -47, 59)$$

is a solution basis for the given data, and by Theorem 3 we have

$$g(\mathbf{a}) = 920\,947\,, \qquad N(\mathbf{a}) = 493\,045\,.$$

References

[1] Chr. Blatter: *Another proof of Pick's area theorem.* Math. Mag. 70 (1997), 200.

[2] S.M. Johnson: *A linear diophantine problem.* Can. J. Math. 12 (1960), 390–398.

[3] P.Q. Nguyen & D. Stehlé: *Low-dimensional lattice basis reduction revisited.* ACM Transactions on Algorithms 5 (2009), 4 (Oct.), 1–48.

[4] J.L. Ramirez Alfonsin: *The diophantine Frobenius problem.* Oxford University Press 2005

[5] Ø.J. Rødseth: *On a linear diophantine problem of Frobenius.* J. Reine und Angewandte Mathematik 301 (1978), 171–178.

[6] J.J. Sylvester: *On subinvariants, i.e. semi-invariants to binary quantities of an unlimited order.* Am. J. Math. 5 (1882), 119–136.

[7] J.J. Sylvester: *Problem 7382.* Educational Times 37 (1884), 26.

Address of the author:

Christian Blatter
Department of Mathematics
Swiss Fed. Inst. of Technology
CH-8092 Zurich (Switzerland)

e-mail: `christian.blatter@math.ethz.ch`