Shrinking Lattice Polyhedra

John Cremona Department of Mathematics University of Exeter

Susan Landau*
Department of Mathematics
Wesleyan University

Abstract

This paper treats the following geometric problem: Given vertices x_1, \ldots, x_n of a polyhedron in the integer lattice in k dimensions, can that polyhedron be "shrunk" to a similar one (i.e., one whose sides remain in the same ratio as the original polyhedron) while still remaining on the integer lattice. We give a necessary condition for this to be possible, which depends on the parity of k; for $k \leq 4$ we show that the condition is sufficient, and we also give algorithms to do the shrinking. In two dimensions, the algorithm only involves computing greatest common divisors over the Gaussian integers and is polynomial time. For k=3 and 4 the algorithms involve computing g.c.d.s in the algebra of Hurwitz quaternions. This gives a polynomial time algorithm for k=4, but because the algorithm in three dimensions relies on determining the square factors of an integer, it is at present exponential. The proofs are remarkably simple and are quite computational in nature.

1 Introduction

This paper treats the following geometric problem: Given vertices x_1, \ldots, x_n of a polyhedron in the integer lattice in k dimensions, can that polyhedron be "shrunk" to a similar one (i.e., one whose sides remain in the same ratio as the original polyhedron) while still remaining on the integer lattice.

^{*}Supported in part by an MSRI fellowship and NSF grant DCR-8402175.

We give a necessary condition for this to be possible, which depends on the parity of k; for $k \leq 4$ we show that the condition is sufficient, and we also give algorithms to do the shrinking. In two dimensions, the algorithm only involves computing greatest common divisors over the Gaussian integers and is polynomial time. For k=3 and 4 the algorithms involve computing g.c.d.s in the algebra of Hurwitz quaternions. This gives a polynomial time algorithm for k=4, but because the algorithm in three dimensions relies on determining the square factors of an integer, it is at present exponential. The proofs are remarkably simple and are quite computational in nature.

The idea behind the algorithms is that there is a natural correspondence between the points of the integer lattice and elements of a certain algebraic structure: the Gaussian integers in the two dimensional case, Hurwitz quaternions in the three and four dimensional cases. Then the shrinking problem can be viewed algebraically. The solutions given rely on the existence of Euclidean algorithms in the rings of Gaussian integers and Hurwitz quaternions respectively.

In section 2 we make some preliminary geometrical observations common to all dimensions, and derive a necessary condition for shrinking to be possible. This depends, somewhat surprisingly, on the parity of the dimension k. Then in the last three sections we present the solution for the two, three and four dimensional cases respectively.

2 Geometrical preliminaries

A general reference for this section is [7, §§65,70].

A similarity of \mathbb{R}^k is a linear map $T: \mathbb{R}^k \to \mathbb{R}^k$ which preserves angles. If T has matrix M (with respect to the standard basis of \mathbb{R}^k) this is equivalent to $M^tM = \lambda I$ for some non-zero scalar λ . Moreover, λ is necessarily positive, since for any $x \in \mathbb{R}^k$ with $Mx \neq 0$, we have $|Mx|^2 = \lambda |x|^2$, so that $\lambda = |Mx|^2/|x|^2$. The (positive) quantity $\sqrt{\lambda}$ is the scale factor of T (or of M), and T is the product of the orthogonal transformation $\frac{1}{\sqrt{\lambda}}M$ and a scaling by $\sqrt{\lambda}$. A similarity is direct or orientation-preserving if it has positive determinant.

As with any linear map, a similarity T is determined by the images of any k linearly independent points in \mathbb{R}^k . If these points and their images all have rational coordinates, then clearly the matrix M of T has rational entries. When k is odd, this imples that $\sqrt{\lambda}$ is also rational:

Lemma 2.1. Let k be odd, and let M in $M_k(\mathbb{Q})$ satisfy $M^tM = \lambda I$, with $\lambda \neq 0$. Then $\lambda = \mu^2$ for some μ in \mathbb{Q} with $\mu^k = \det M$. Hence $\mu^{-1}M$ is in $\mathsf{SO}(k,\mathbb{Q})$, the group of $k \times k$ orthogonal matrices over \mathbb{Q} with determinant 1.

Proof. We have $(\det M)^2 = \lambda^k$, so $|\det M| = \lambda^{k/2}$. Since $\det M$ is rational and k is odd, this implies that $\mu = \sqrt{\lambda}$ is rational, and $\mu^k = |\det M|$. Replacing μ by $-\mu$ if necessary we can ensure that $\mu^k = \det M$, again since k is odd.

A general angle-preserving map $\mathbb{R}^k \to \mathbb{R}^k$ consists of a translation followed by a similarity, and is determined by the images of any k+1 affinely independent points (that is, of any k+1 points which do not all lie on a hyperplane of dimension k-1).

Our strategy will be as follows: given n points x_1, \ldots, x_n with integer coordinates in \mathbb{R}^k , we must find an angle-preserving transformation R of \mathbb{R}^k with minimal scale factor such that each Rx_i again has integer coordinates. First assume that 0 (the origin) is one of the given points. Then without loss of generality we may assume that R(0) = 0, so that R is a similarity; again without loss, we may presume R to be direct. Having found the solution to this problem of finding the best direct similarity, we must then determine the translation with which to precede such a similarity in order to achieve the best shrinkage. For this it is always optimal to translate so that one of the given points is at the origin: since similarities preserve 0, for any other choice we are effectively including 0 as an extra n + 1st point in the given configuration, which could not give a better shrinkage. Hence we will assume from now on that the origin is one of the given points, and hence that the transformation sought is a direct similarity.

Since the transformations we seek are linear (after the initial translation), we could replace the given points in \mathbb{Z}^k with a set of (at most) k points which generate the same subgroup of the lattice \mathbb{Z}^k , thus reducing the problem to one involving at most k points in k dimensions. This shows that the smallest scale factor is not a function of the particular set of points given, but of the subgroup of \mathbb{Z}^k which they generate (assuming that 0 is in the set). So properties of the particular geometrical configuration, such as convexity, are irrelevant in the solution.

For vectors x and y in \mathbb{R}^k we denote the usual scalar product of x and y by $x \cdot y$, and (when k = 3) their vector or cross product by $x \times y$. Also, for a positive integer g the square part of g is the largest integer whose square

divides g (for example, the square part of 12 is 2). A number is said to be square-free if its square part is 1.

From the defining equation of a similarity we see that $(Rx) \cdot (Ry) = \lambda x \cdot y$ for all vectors x and y in \mathbb{R}^k . If for $1 \leq i \leq n$ we have both $x_i \in \mathbb{Z}^k$ and $Rx_i \in \mathbb{Z}^k$, it follows that λ is a rational number whose denominator divides each of the scalar products $x_i \cdot x_j$. Moreover, when k is odd this denominator is a square, by Lemma 2.1. This gives the following necessary condition for a scaling by a factor λ to be possible, and a lower bound on the possible shrinkage of a given configuration.

Lemma 2.2. Let $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{Z}^k$. Set $g = \gcd_{\mathbb{Z}}(\{x \cdot y \mid x, y \in X\})$. Let R be a similarity with scale factor $\sqrt{\lambda}$ such that $R(X) \subset \mathbb{Z}^k$; write $\lambda = a/b$ with $\gcd(a,b) = 1$. Then b divides g, and moreover if k is odd then b is a square divisor of g. Hence the smallest such λ satisfies $\sqrt{\lambda} \geq 1/d$, where $d = \sqrt{g}$ when k is even, and d is the square part of g when k is odd. In particular, no shrinkage of X is possible if g = 1, or if k is odd and g is square-free.

In the following sections we will show that for $k \leq 4$ the necessary condition is sufficient and the lower bound is actually attained. We sum this up in the following theorem.

Theorem 2.3. Let $X = \{x_1, x_2, ..., x_n\} \subset \mathbb{Z}^k$, where $k \leq 4$. Set $g = \gcd_{\mathbb{Z}}(\{x \cdot y \mid x, y \in X\})$ and let d be the square part of g. Then there exists a similarity R with $R(X) \subset \mathbb{Z}^k$ and scale factor $1/\sqrt{g}$ (when k is even) or 1/d (when k is odd). This achieves the best possible shrinkage of the given configuration X.

For completeness, we give here the proof for k=1. The idea is essentially the same as in the higher dimensional cases, but is almost trivial. We are given integers x_1, x_2, \ldots, x_n and we must find the smallest positive real number r such that each rx_i is again integral. If $d = \gcd_{\mathbb{Z}}(x_1, \ldots, x_n)$ this is equivalent to the single condition $rd \in \mathbb{Z}$, so the smallest such r is clearly r = 1/d. The map $x \mapsto rx$ is a direct similarity with scale factor 1/d. This proves the theorem for the case k = 1, since it is clear that $g = \gcd(\{x_ix_j\}) = d^2$. To shrink a one-dimensional configuration $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{Z} \subset \mathbb{R}$ (not necessarily including 0) we therefore compute $d = \gcd_{\mathbb{Z}}(x_2 - x_1, \ldots, x_n - x_1)$ and map $x_i \mapsto (x_i - x_1)/d$ for $1 \le i \le n$. The computation is dominated by the single g.c.d. computation which takes $O(n \log N)$ steps where $N = \max(|x_1|, |x_2|, \ldots, |x_n|)$.

Next we turn to the two-dimensional case. Here it turns out that the solution is almost identical to that for the one-dimensional case just given, replacing the integers by the ring of Gaussian integers, which also has a Euclidean algorithm.

3 The 2-dimensional case

We will use the fact that the set $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ of Gaussian integers forms a commutative ring with a Euclidean algorithm; thus any two Gaussian integers α , β have a greatest common divisor which is well-defined up to multiplication by ± 1 or $\pm i$. We define $\gcd_{\mathbb{Z}[i]}(\alpha, \beta)$ to be this g.c.d., normalized to lie in the first quadrant; thus $\gcd_{\mathbb{Z}[i]}(\alpha, \beta) = 1$ if and only if α and β have no non-trivial common factor.

Identify a point (x, y) in \mathbb{R}^2 with the complex number z = x + iy, where $i = \sqrt{-1}$. Points with integer coordinates then correspond to Gaussian integers, and points with rational coordinates to elements of the field $\mathbb{Q}(i)$. The n given integer points x_1, \ldots, x_n thus correspond to n Gaussian integers $\alpha_1, \ldots, \alpha_n$.

Our main result is:

Theorem 3.1. The polygon defined by x_1, \ldots, x_n can be shrunk if and only if $\gcd_{\mathbb{Z}[i]}(\alpha_2 - \alpha_1, \ldots, \alpha_n - \alpha_1) \neq 1$. The maximal shrinkage factor can be computed in $O(n \log N)$ steps, where $N = \max(|\alpha_1|, \ldots, |\alpha_n|)$.

The idea of the proof is as follows: viewing the points x_i as elements of the ring of Gaussian integers, any shrinkage will be achieved by dividing the points by some fixed Gaussian integer. We can thus find the maximal shrinkage by computing the g.c.d. of the given points, after first shifting the origin. We prove this by a succession of lemmas.

Lemma 3.2. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be a direct similarity. Then T has matrix $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$ with x, y in \mathbb{R} and $x^2 + y^2 \neq 0$.

Proof. The matrix M of T satisfies $M^tM = \text{scalar}$ and $\det M > 0$, by the definitions in the previous section. An easy computation shows that M must then have the given form.

Lemma 3.3. The set $\left\{ \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \mid x, y \in \mathbb{Q} \right\}$ of matrices forms a field isomorphic to $\mathbb{Q}(i)$ via the mapping $\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \mapsto x + yi$.

Proof. Standard exercise.

Any fixed nonzero complex number α determines a transformation T_{α} of \mathbb{R}^2 , namely $z\mapsto \alpha z$ (identifying \mathbb{R}^2 with \mathbb{C} as always). If $\alpha=x+iy$ then T_{α} has matrix $\begin{pmatrix} x & -y \\ y & x \end{pmatrix}$, so that T_{α} is a direct similarity with scale factor $|\alpha|=\sqrt{x^2+y^2}$. Conversely, any direct similarity has the form T_{α} by Lemma 3.2. Moreover it is clear that T_{α} preserves the rational points $\mathbb{Q}^2\subset\mathbb{R}^2$ if and only if x and y are rational, i.e., if and only if $x\in\mathbb{Q}(i)$.

Lemma 3.4. Let $\alpha_1, \ldots, \alpha_n$ be a finite set of points in \mathbb{Z}^2 (identified with $\mathbb{Z}[i]$). The similarity T_{α} takes each α_j to an integral point if and only if α is in the fractional ideal of $\mathbb{Q}(i)$ inverse to the ideal generated by $\alpha_1, \ldots, \alpha_n$.

Proof. We require $\alpha \alpha_j \in \mathbb{Z}[i]$ for j = 1, ..., n. This is equivalent to $\alpha A \subseteq \mathbb{Z}[i]$, where A is the ideal $(\alpha_1, ..., \alpha_n)$, which in turn is equivalent to $\alpha \in A^{-1}$.

For the best shrinkage we must find $\alpha \in A^{-1}$ with minimal norm: a generator of A^{-1} achieves this. Write $A = (\gamma)$ where $\gamma = \gcd_{\mathbb{Z}[i]}(\alpha_1, \ldots, \alpha_n)$; then $A^{-1} = (\gamma)^{-1} = (\beta)$ with $\beta = \gamma^{-1}$. Now T_{β} is the best shrinkage, and it is trivial if and only if $|\beta| = 1$, which is equivalent to $\gcd_{\mathbb{Z}[i]}(\alpha_1, \ldots, \alpha_n) = 1$.

Hence we have a solution to the two-dimensional shrinking problem. To find the best origin-preserving shrinkage, compute γ , the g.c.d. of the given points $\alpha_1, \ldots, \alpha_n$ viewed as Gaussian integers, and then divide each α_j by γ . The best choice of origin is, as observed in the previous section, to place it at one of the given points. So the best shrinkage is finally obtained by shifting the origin to α_1 , computing γ as the g.c.d. of $\alpha_2 - \alpha_1, \ldots, \alpha_n - \alpha_1$, and then dividing by γ .

We can now present the algorithm.

Algorithm: Shrink Polygons

Input: $\{x_j = (a_j, b_j)\}$ * a set of n points which defines a polygon in \mathbb{Z}^2 *; Output: $\{y_j = (a_j, b_j)\}$ * a set of n points which defines a similar polygon of minimum size in \mathbb{Z}^2 *;

- (1) BEGIN
- (2) FOR j = 1 TO n DO: $\alpha_j \leftarrow (a_j a_1) + i(b_j b_1)$;
- (3) $\gamma \leftarrow \gcd_{\mathbb{Z}[i]}(\alpha_2, \dots, \alpha_n);$
- (4) FOR j = 1 TO n DO: $a_j + ib_j \leftarrow \alpha_j/\gamma$;
- (5) END.

All that remains to be shown is that the running time is $O(n \log N)$ steps, where N is the maximum norm of the α_j . But this is clear, since the running time of the algorithm is dominated by the g.c.d. calculation on line (3). By [6] this computation can be done in $O(n \log N)$ steps. This completes the proof of Theorem 3.1.

We note that we made no assumptions regarding convexity or holes; thus our algorithm works for all polygons in \mathbb{Z}^2 .

Finally we must show that the shrinkage obtained by the algorithm agrees with that claimed in Theorem 2.3. For this it suffices to show that if for some prime number p we have $x_j \cdot x_k \in p\mathbb{Z}$, then $\gcd_{\mathbb{Z}[i]}(\alpha_1, \ldots, \alpha_n) \neq 1$. Here as before we set $\alpha_j = a_j + ib_j$ where $x_j = (a_j, b_j)$.

If p = 2 the conditions imply that either $a_j \equiv b_j \equiv 0 \pmod{2}$ for all j, in which case 2 divides each α_j , or $a_j \equiv b_j \equiv 1 \pmod{2}$ for all j, in which case 1 + i divides each α_j .

If $p \equiv 3 \pmod{4}$ then $a^2 + b^2 \equiv 0 \pmod{p}$ implies that $a \equiv b \equiv 0 \pmod{p}$ so that the given conditions imply that each α_i is divisible by p.

If $p \equiv 1 \pmod{4}$ then $p = \pi \overline{\pi}$ for some $\pi \in \mathbb{Z}[i]$, and $N(\alpha_j) = \alpha_j \overline{\alpha_j} = a_j^2 + b_j^2 \equiv 0 \pmod{p}$ implies that each α_j is divisible either by π or by $\overline{\pi}$ (or by both). Finally the condition $a_j a_k + b_j b_k \equiv 0 \pmod{p}$ implies that each $\alpha_j \overline{\alpha_k}$ is divisible by p, so that at least one of π or $\overline{\pi}$ divides all of the α_j .

(Here we have been using well-known properties of the ring $\mathbb{Z}[i]$, including unique factorization: see [2, §3.8] for details).

4 The 3-dimensional case

In both three and four dimensions our algorithm will use quaternions, making use of both their geometrical and arithmetical properties.

Let $\mathbb{H} = \{x+yi+zj+tk \mid x,y,z,t\in\mathbb{R}\}$, the 4-dimensional real algebra of quaternions. The quaternion q=x+yi+zj+tk has a conjugate $\overline{q}=x-yi-zj-tk$, a trace $Tr(q)=q+\overline{q}=2x$ and a norm $N(q)=q\overline{q}=x^2+y^2+z^2+t^2$. We will identify \mathbb{R}^3 with the subspace V of \mathbb{H} spanned by i,j and k (the quaternions with zero trace), via $(x,y,z)\mapsto xi+yj+zk$. It is well-known that the group of rotations of \mathbb{R}^3 can be parametrized by non-zero quaternions using the conjugation action (see [5, pp.42–44]). Specifically, any non-zero quaternion q induces a transformation R_q from V to V, which is in fact a rotation: $v\mapsto qvq^{-1}$. In fact, every rotation of V arises in this way, though we will not need this.

The second important ingredient in the algorithm is the ring S of Hurwitz quaternions which lies in \mathbb{H} ; it is spanned over \mathbb{Z} by the four elements $\frac{1}{2}(1+i+j+k)$, i, j and k. We define

$$S_0 = S \bigcap V = \{ai + bj + ck \mid a, b, c \in \mathbb{Z}\},\$$

so that under the above identification S_0 is identified with \mathbb{Z}^3 . The crucial fact for the efficiency of our algorithm is that the non-commutative ring S has a right-Euclidean algorithm. We summarise here for convenience the facts about quaternion arithmetic which we will use below, giving references to $[1, \S 91]$ for details.

D1: Every pair of elements α and β of S has a greatest common right divisor γ which is characterised by the properties that γ divides both α and β on the right, and that every common right divisor of α and β is a right divisor of γ . We denote the greatest common right divisor or g.c.r.d. of α and β by $gcrd(\alpha, \beta)$. [1, §91,Theorem 2]

D2: The g.c.r.d. is uniquely determined up to multiplication on the left by a unit (invertible element) of S. [1, §91,Theorem 2]

D3: If $\gamma = \gcd(\alpha, \beta)$ then there exist δ and η in S such that $\gamma = \delta\alpha + \eta\beta$. [1, §91, Theorem 2]

D4: If the norm of $\alpha \in S$ is divisible by an integer n > 1 then $gcrd(\alpha, n)$ is not a unit. [1, §91,Lemma3]

D5: The g.c.r.d. of two quaternions in S may be computed by a Euclidean algorithm which uses a simple generalisation of the usual (integer) division algorithm to S. [1, §91,Lemmas 1 and 2]

Similar properties hold for the greatest common left divisor, or g.c.l.d., which will be used in the next section.

As in the case of two dimensions, we are given a polyhedron with vertices x_1, \ldots, x_n in \mathbb{Z}^3 , and we must find y_1, \ldots, y_n in \mathbb{Z}^3 which are the vertices of a "similar" polyhedron which is as small as possible. We may assume that the x_r do not all lie on a plane, otherwise the problem is degenerate.

We know from Lemma 2.2 that if shrinkage by a factor d is possible then d^2 must divide each of the numbers $x_r \cdot x_s$, and hence divide their greatest common divisor g. We now prove Theorem 2.3 for k = 3, which states that the necessary condition is in fact sufficient. For this it suffices to show that if g is divisible by p^2 for some prime $p \in \mathbb{Z}$, then we can shrink by the factor p. We therefore assume that $x_r \cdot x_s \equiv 0 \pmod{p}$ for all r, s, and we must find R in $SO(3, \mathbb{Q})$ such that $Rx_r \in (p\mathbb{Z})^3$ for $r = 1, \ldots, n$; the required similarity is then $M = p^{-1}R$.

The trivial case occurs when each x_r is already in $(p\mathbb{Z})^3$, and then we may take R = I (the 3×3 identity matrix) and $M = p^{-1}I$. This is necessarily the case when p = 2, since $a^2 + b^2 + c^2 \equiv 0 \pmod{4}$ implies that $a \equiv b \equiv c \equiv 0 \pmod{2}$.

We may now assume that p is odd. Then our problem may be translated into the language of quaternions as follows. The given points $x_r = (a_r, b_r, c_r) \in \mathbb{Z}^3$ correspond to quaternions $\alpha_r = a_r i + b_r j + c_r k \in S_0$, and we must find π in $\mathbb{H}^* = \mathbb{H} - \{0\}$ such that $\pi \alpha_r \pi^{-1} \in pS_0$ for all r. It will turn out that we may take π to be $\gcd(p, \alpha_1, \alpha_2, \ldots, \alpha_n)$.

Let X be the $n \times 3$ matrix with x_r in row r. Our condition (2) implies that $X^t X \equiv 0 \pmod{p^2}$. It follows that the rank of X modulo p is 0 or 1 (for if $\operatorname{rk}_p X = \operatorname{rk}_p X^t \geq 2$, then $\operatorname{rk}_p X^t X \geq 1$). The case where $\operatorname{rk}_p X = 0$ is the trivial case which was handled earlier, so we may assume that $\operatorname{rk}_p X = 1$.

Without loss of generality, we thus have $x_1 = (a_1, b_1, c_1) \not\equiv (0, 0, 0)$ (mod p). For each $r \geq 2$ we may assume (replacing x_r by $x_r + x_1$ if necessary) that $x_r \not\equiv (0, 0, 0) \pmod{p}$ also. The preceding remark about rank implies that each $x_r \equiv \lambda_r x_1 \pmod{p}$ for some integers λ_r . Thus as well as $x_r \cdot x_s \equiv 0 \pmod{p^2}$ for all r and s, we also have $x_r \times x_s \equiv 0 \pmod{p}$. Hence p divides $\alpha_r \alpha_s$ for all r and s, since the 'real' part of $\alpha_r \alpha_s$ is $-x_r \cdot x_s$ while the 'imaginary' or vector part is $x_r \times x_s$. In addition, we know that p^2 divides $N(\alpha_r)$ for all r.

The desired result now follows from the following result concerning quaternions, which implies the existence of $\pi \in S$ with $N(\pi) = p$ such that each α_r is of the form $\overline{\pi}\beta_r\pi$ with $\beta_r \in S_0$. Then the rotation R_{π} (given by $\alpha \mapsto \pi\alpha\pi^{-1}$)

sends each α_r to $\pi \alpha_r \pi^{-1} = \pi \overline{\pi} \beta_r \pi \pi^{-1} = p \beta_r$, as required.

Proposition 4.1. Let p be an odd prime.

- (a) Let $\alpha \in S pS$, and suppose that $p | N(\alpha)$. Then $\pi = \gcd(p, \alpha)$ satisfies $N(\pi) = p$.
- (b) Let $\alpha \in S_0 pS_0$, and suppose that $p^2 | N(\alpha)$. Set $\pi = \gcd(p, \alpha)$. Then $\alpha = \overline{\pi}\beta\pi$ for some $\beta \in S_0$.
- (c) Suppose that α_1 and α_2 satisfy the conditions of part (b), and also that $p | \alpha_1 \alpha_2$. Set $\pi = \gcd(p, \alpha_1)$. Then there exist $\beta_1, \beta_2 \in S_0$ such that $\alpha_r = \overline{\pi} \beta_r \pi$ for r = 1, 2.
- *Proof.* (a) By **D4** we have $N(\pi) > 1$. Writing $p = \eta \pi$ with $\eta \in S$ we then have $N(\eta) = 1$ or p; but if $N(\eta) = 1$ then α is divisible by p, giving a contradiction.
- (b) By part (a) we have $\alpha = \gamma \pi$ with $\gamma \in S$. Applying part (a) again to $\overline{\gamma}$ we also have $\overline{\gamma} = \delta \pi_1$ with π_1 , $\delta \in S$ and $N(\pi_1) = p$. Then on the one hand $\alpha = \overline{\pi_1} \overline{\delta} \pi$; on the other hand, $\alpha = -\overline{\alpha} = -\overline{\pi} \delta \pi_1$. The uniqueness property **D2** now gives $\pi_1 = \eta \pi$ for some unit $\eta \in S$, and hence $\alpha = \overline{\pi}(\overline{\eta} \overline{\delta})\pi$. This gives the desired result with $\beta = \overline{\eta} \overline{\delta}$: the fact that β has trace 0 follows from the relation $p\beta = \pi \alpha \pi^{-1}$.
- (c) Applying part (b) separately to α_1 and α_2 , we have $\alpha_r = \overline{\pi_r}\beta_r\pi_r$ with $\beta_r \in S$ and $N(\pi_r) = p$, where $\pi_r = \gcd(p, \alpha_r)$, for r = 1, 2. By **D3** there exist ξ_r , $\eta_r \in S$ such that $\xi_r\alpha_r + \eta_r p = \pi_r$ for r = 1, 2. Then

$$\pi_1 \overline{\pi_2} \equiv (\xi_1 \alpha_1) \overline{(\xi_2 \alpha_2)} = \xi_1 \alpha_1 \overline{\alpha_2} \overline{\xi_2} = -\xi_1 \alpha_1 \alpha_2 \overline{\xi_2} \equiv 0 \pmod{p}$$

so that $\pi_1\overline{\pi_2} = \delta p$ for some $\delta \in S$ with $N(\delta) = 1$. Since $p = \pi_2\overline{\pi_2}$ this gives $\pi_1 = \delta \pi_2$, and so $\alpha_2 = \overline{\pi_2}\beta_2\pi_2 = \overline{\pi_1}(\delta\beta_2\overline{\delta})\pi_1$. Hence the desired result with $\pi = \pi_1$, on replacing β_2 by $\overline{\delta}\beta_2\delta$.

As well as proving Theorem 2.3 for k=3, we have now shown how to go about shrinking by a prime p whose square divides the g.c.d. of the scalar products of the x_r . We pick a vector (a,b,c) in our list which is not trivially a multiple of p, set $\alpha = ai + bj + ck$ and compute $\pi = \gcd(p,\alpha)$. Then the rotation R_{π} defined by $\alpha \mapsto \pi \alpha \pi^{-1}$ maps each x_r into $p\mathbb{Z}^3$, after which the factor of p can be divided out.

However in practice one need not shrink by one prime at a time, as the next result shows.

Proposition 4.2. Let $x_1, x_2, \ldots, x_n \in \mathbb{Z}^3$. Write $x_r = (a_r, b_r, c_r)$ for $r = 1, 2, \ldots, n$. Suppose that

- (1) $x_r \cdot x_s \in d^2 \mathbb{Z}$ for $1 \leq r \leq s \leq n$ and some odd integer d;
- (2) $gcd(a_1, b_1, c_1, a_2, \dots, c_n) = 1.$

Set $\alpha_r = a_r i + b_r j + c_r k$ and $\pi = \gcd(d, \alpha_1, \alpha_2, \dots, \alpha_n)$. Then $N(\pi) = d$, and each α_r has the form $\overline{\pi}\beta_r\pi$ with $\beta_r \in S_0$, so that $R_{\pi}(x_r) \in (d\mathbb{Z})^3$ for $r = 1, 2, \dots, n$.

Proof. Write d as a product of (not necessarily distinct) odd primes, say $d = p_1 p_2 \dots p_k$. We prove the Proposition by induction on k. The case k = 1 (so that d is an odd prime) was handled above, so suppose that $k \geq 2$ and the result holds for k - 1.

The hypotheses clearly still hold if we replace d by $d_0 = d/p_k$, which is a product of k-1 primes, so by induction we have $\alpha_r = \overline{\pi_0}\gamma_r\pi_0$ with $\gamma_r \in S_0$ and $N(\pi_0) = d_0$. Write $\gamma_r = a'_r i + b'_r j + c'_r k$ and set $y_r = (a'_r, b'_r, c'_r)$. Then $R_{\pi_0}(x_r) = d_0 y_r$; since R_{π_0} preserves the scalar product it follows that p_k^2 divides each $y_r \cdot y_s$. Secondly, if $g = \gcd(a'_1, b'_1, c'_1, a'_2, \dots, c'_n) > 1$, then g would divide each γ_r and hence each α_r , contradicting condition (2). It follows that the conditions of the Proposition are satisfied with y_r and p_k in place of x_r and d; we may apply the induction hypothesis again (with k = 1) to obtain $\gamma_r = \overline{\pi_k}\beta_r\pi_k$ with $\beta_r \in S_0$ and $N(\pi_k) = p_k$. Setting $\pi = \pi_k\pi_0$ we have, for each r, $\alpha_r = \overline{\pi}\beta_r\pi$ and $N(\pi) = d_0 p_k = d$.

To complete the proof we must show that π is indeed a g.c.r.d. of d and $\alpha_1, \alpha_2, \ldots, \alpha_n$. Certainly π is a common right divisor of these, since $\alpha_r = \overline{\pi}\beta_r\pi$ and $d = N(\pi) = \overline{\pi}\pi$. Thus if $\delta = \gcd(d, \alpha_1, \alpha_2, \ldots, \alpha_n)$ then $\delta = \eta\pi$ for some $\eta \in S$, and we must show that $N(\eta) = 1$.

Since δ is a right divisor of d we have $d = \theta \delta$ for some $\theta \in S$. Then $\overline{\pi}\pi = d = \theta \eta \pi$ so that $\theta \eta = \overline{\pi}$, or $\pi = \overline{\eta}\overline{\theta}$. Now $\delta = \eta \pi = N(\eta)\overline{\theta}$ is a right divisor of each α_r , from which it follows that $N(\eta)$ is an integer which divides each α_r . From condition (2) we must have $N(\eta) = 1$ as required.

In our algorithm we will first divide out by the 'trivial' shrinkage factor $f = \gcd(a_1, b_1, c_1, a_2, \ldots, c_n)$. Then, using Proposition 4.2, we compute the non-trivial 'rotational' shrinkage, which will be of the form $N(\pi)^{-1}R_{\pi}$ for some $\pi \in S$ with odd norm. Writing $\pi = x + yi + zj + tk$, the rotation R_{π} has matrix

$$R_{\pi} = \frac{1}{N(\pi)} \begin{pmatrix} x^2 + y^2 - z^2 - t^2 & 2(yz - xt) & 2(xz + yt) \\ 2(xt + yz) & x^2 - y^2 + z^2 - t^2 & 2(zt - xy) \\ 2(yt - xz) & 2(xy + zt) & x^2 - y^2 - z^2 + t^2 \end{pmatrix}.$$

Hence the final transformation has the form $x \mapsto f^{-1}N(\pi)^{-1}R_{\pi}(x)$ for some $\pi \in S$ and $f \in \mathbb{Z}$.

The question of the running time of the algorithm will be considered below.

Algorithm: Shrink Polyhedra

Input: $\{x_r = (a_r, b_r, c_r)\}$ * a set of n points which defines a polyhedron in \mathbb{Z}^3 *:

Output: $\{y_r = (a_r, b_r, c_r)\}$ * a set of n points which defines a similar polyhedron of minimum size in \mathbb{Z}^3 *;

- (1) BEGIN
- (2) FOR r = 1 TO n DO: $x_r \leftarrow (a_r, b_r, c_r) \leftarrow x_r x_1$;
- $(3) f \leftarrow \gcd_{\mathbb{Z}}(a_2, b_2, c_2, a_3, \dots, c_n);$
- (4) FOR r = 1 TO n DO: $x_r \leftarrow (a_r, b_r, c_r) \leftarrow x_r/f$;
- $(5) g \leftarrow \gcd_{\mathbb{Z}}(x_2 \cdot x_2, x_2 \cdot x_3, \dots, x_2 \cdot x_n, x_3 \cdot x_3, \dots, x_n \cdot x_n);$
- (6) $d \leftarrow \text{squarepart}(g);$
- (7) $\pi \leftarrow \gcd(d, a_1i + b_1j + c_1k, a_2i + b_2j + c_2k, \dots, a_ni + b_nj + c_nk);$
- (8) FOR r = 1 TO n DO: $(a_r, b_r, c_r) \leftarrow R_{\pi}(x_r)/d$;
- (9) END.

Proposition 4.3. The algorithm Shrink Polyhedra takes as input a set of n points x_1, x_2, \ldots, x_n in \mathbb{Z}^3 which are the vertices of a polyhedron, and returns a set of n points in \mathbb{Z}^3 which are the vertices of a similar polyhedron of minimal size. It does so in the time it takes to compute the square part of $\gcd_{\mathbb{Z}}(\{x'_r.x'_s \mid 1 \leq r \leq s \leq n\})$, where $x'_r = x_r - x_1$.

Proof. Our work is essentially done. By Theorem 2.2, shrinking can be done if and only if (after translating so that one of the given points is at the origin) each $x_r \cdot x_s \in d^2\mathbb{Z}$ for some integer d > 1. After translating the origin to x_1 in line (2), we perform the 'trivial' shrinking (which is possible if the coordinates of the points have a nontrivial common divisor) in lines (3)–(4). In lines (5)–(6) we compute the remaining shrinking factor d, and finally in lines (7)–(8) we carry out the non-trivial shrinking, using the result of Proposition 4.2. The matrix R_{π} in line (8) is the matrix of the rotation $\alpha \mapsto \pi \alpha \pi^{-1}$ given above.

The running time is dominated by the single computation of the square part of g, which occurs once in line (6). Unfortunately at present the best algorithms to do this are exponential, for they involve factoring g, and thus are of complexity $\exp(c\sqrt{\log g \log \log g})$.

The g.c.d. computation in line (3) takes $O(n \log a)$ steps where $a = \max_r(|a_r|, |b_r|, |c_r|)$. The g.c.d. computation in line (7) is a computation over the quaternions (see [2, Lemma 7.15] or [1, §1, Lemmas 1 and 2] for a discussion and [4] for a precise time bound), and the time bound for this computation is similar. The time to factor g in line (6) is exponentially greater than these computations.

Finally we note (again) that we made no assumptions concerning convexity or holes, so the algorithm works for non-convex polyhedra with holes as well.

5 The 4-dimensional case

In this section we will prove Theorem 2.3 for k = 4, and give an algorithm which achieves maximum shrinking of a given configuration of points in \mathbb{Z}^4 in polynomial time. Again we will use quaternions.

Identifying \mathbb{H} with \mathbb{R}^4 , we first observe that for fixed $h \in \mathbb{H}$ the left and right multiplication maps $\lambda_h \colon x \mapsto hx$ and $\rho_h \colon x \mapsto xh$ from \mathbb{R}^4 to \mathbb{R}^4 are in fact similarities with scale factor $\sqrt{N(h)}$. This follows immediately from the equations

$$N(hx) = N(h)N(x) = N(x)N(h) = N(xh)$$

for $x \in \mathbb{H}$, since N(x) is the square length of the vector x. In fact, every similarity of \mathbb{R}^4 has the form $x \mapsto h_1 x h_2$ for suitable nonzero quaternions h_1 and h_2 , with scale factor $N(h_1 h_2)$, but we will not need to use this.

After shifting the origin and identifying the given points with elements of S (the ring of Hurwitz quaternions), the algorithm will consist merely of first dividing on the left by the greatest common left divisor (g.c.l.d.), then dividing on the right by the greatest common right divisor (g.c.r.d.) of the resulting points. This is guaranteed to give the maximum shrinkage, achieving the bound given in Theorem 2.3. Some care is needed to ensure that the transformed points do have integer coordinates, since S includes certain quaternions whose coordinates have a denominator of 2. We start with a

lemma which essentially solves this problem. Let $S_1 = \{a + bi + cj + dk \mid a, b, c, d \in \mathbb{Z}\}$, which is a subgroup of index 2 in S.

Lemma 5.1. Let $\alpha \in S$. Then there exists a unit $\eta \in S$ such that $\eta \alpha \in S_1$ and $\alpha \eta \in S_1$.

Proof. See $[1, \S 91, \text{Lemma } 7]$.

Remarks

- 1. It follows that every set of elements of S has both a g.c.r.d. and a g.c.l.d. in S_1 .
- 2. The unit group G of S has order 24 ([1, page 153]) and the subgroup $H = G \cap S_1 = \{\pm 1, \pm i, \pm j, \pm k\}$ has order 8. Since H clearly preserves S_1 , in implementing Lemma 5.1 we need only try two possible values of η if α is not already in S_1 , namely representatives of the two other cosets of H in G. For these we may take η and $\overline{\eta}$, where $\eta = \frac{1}{2}(1+i+j+k)$. In implementing the g.c.r.d. algorithm, we therefore assume that where possible the delivered value of $\pi = \gcd(\alpha, \beta)$ will satisfy $\pi^{-1}\alpha, \pi^{-1}\beta \in S_1$. Similarly for the g.c.l.d. algorithm. Clearly this will not affect the order of the running time of these algorithms.

Now let x_1, x_2, \ldots, x_n be a set of points in \mathbb{Z}^4 with the property that $x_r \cdot x_s \in d\mathbb{Z}$ for all r, s. We must show that there exists a similarity R of \mathbb{R}^4 with scale factor $1/\sqrt{d}$ such that $R(x_r) \in \mathbb{Z}^4$ for all r. As usual it suffices to consider the case when d = p is a prime. The following proposition now gives the desired result.

Proposition 5.2. Let $x_r = (a_r, b_r, c_r, d_r) \in \mathbb{Z}^4$ for r = 1, 2, ..., n. Suppose that some prime p divides $x_r \cdot x_s$ for all r, s. Set $\alpha_r = a_r + ib_r + jc_r + kd_r \in S_1$. Then there exists $\pi \in S$ such that $N(\pi) = p$, and either $\pi^{-1}\alpha_r \in S_1$ for all r, or $\alpha_r \pi^{-1} \in S_1$ for all r. Hence either $\lambda_{\pi^{-1}}(x_r) \in \mathbb{Z}^4$ for all r or $\rho_{\pi^{-1}}(x_r) \in \mathbb{Z}^4$ for all r.

Proof. Let X be the $4 \times n$ matrix with x_r in row r. We then have $X^t X \equiv 0 \pmod{p}$, from which it follows that the rank $\operatorname{rk}_p(X)$ of X modulo p is at most 2. If $\operatorname{rk}_p(X) = 0$ then each x_r is already in $p\mathbb{Z}^4$, so that shrinkage by a factor of p is trivially possible. Hence we may assume that $x_1 \notin p\mathbb{Z}^4$ and that for $r \geq 3$ we have $x_r \equiv \mu_r x_1 + \nu_r x_2 \pmod{p}$ for some integers μ_r and ν_r .

Now any common left or right divisor of p, α_1 and α_2 will automatically divide α_r for $r \geq 3$. For simplicity write $\alpha = \alpha_1$ and $\beta = \alpha_2$. The given conditions now become

$$N(\alpha) \equiv N(\beta) \equiv Tr(\alpha \overline{\beta}) \equiv 0 \pmod{p},$$

and it suffices to show that α and β have a common factor π of norm p either on the right or on the left, with the additional property that $\alpha \pi^{-1}$, $\beta \pi^{-1} \in S_1$, or $\pi^{-1}\alpha$, $\pi^{-1}\beta \in S_1$ respectively.

First consider the case where p = 2. The elements 1 + i, 1 + j and 1 + k all have norm 2 and divide every element of S with even norm on both left and right. To ensure that $\pi^{-1}\alpha$ and $\pi^{-1}\beta$ are both in S_1 we must investigate further. Elementary calculation shows that (for $\alpha \in S_1$):

$$(1+i)^{-1}\alpha \in S_1$$
 if $\alpha \equiv 0, 1+i+j+k, 1+i$, or $j+k$ (mod $2S_1$);
 $(1+j)^{-1}\alpha \in S_1$ if $\alpha \equiv 0, 1+i+j+k, 1+j$, or $i+k$ (mod $2S_1$);
 $(1+k)^{-1}\alpha \in S_1$ if $\alpha \equiv 0, 1+i+j+k, 1+k$, or $i+j$ (mod $2S_1$).

Now the condition that $2|Tr(\alpha\overline{\beta})$ gives us the further information that α and β both lie in one of the three sets $\{0, 1+i+j+k, 1+i, j+k\}$, $\{0, 1+i+j+k, 1+j, i+k\}$, and $\{0, 1+i+j+k, 1+k, i+j\}$ when reduced modulo 2. Hence at least one of 1+i, 1+j or 1+k satisfies the required condition on π .

Now let p be an odd prime. If p divides both α and β then the result is trivial, since we have $p = \pi \overline{\pi}$ for some π in S by $[1, \S 91, Lemma 7]$.

If p divides α but not β then, by Proposition 4.1(a), β has a right divisor of norm p, which clearly also divides α .

Now assume that neither α nor β is divisible by p. The quotient algebra A=S/pS can be viewed as the quaternion algebra $\left(\frac{-1,-1}{F}\right)$ over the field $F=\mathbb{Z}/p\mathbb{Z}$. Such an algebra is necessarily split, that is, isomorphic to the 2×2 matrix algebra M(2,F). [A is simple by [3, page 14] and non-commutative; now $A\cong M(2,F)$ follows from Wedderburn's structure theorem for simple algebras, together with Wedderburn's theorem on the non-existence of finite non-commutative division algebras.] Under this isomorphism, the quaternion norm corresponds to the determinant, and the quaternion trace to the matrix trace. It then follows from elementary linear algebra (see Lemma 5.3 below) that either $\alpha\overline{\beta}\equiv 0\pmod{p}$ or $\overline{\beta}\alpha\equiv 0\pmod{p}$.

Suppose that $\alpha \overline{\beta} \equiv 0 \pmod{p}$. We have $\alpha = \alpha_1 \pi_1$ and $\beta = \beta_1 \pi_2$ where $\pi_1 = \gcd(p, \alpha)$ and $\pi_2 = \gcd(p, \beta)$ both have norm p by Proposition 4.1(a).

As in Proposition 4.1(c) we now deduce that p divides $\pi_2\overline{\pi_1}$, hence that $\pi_2\pi_1^{-1}$ is a unit, so that π_1 is a common right divisor of α and β .

Similarly if $\overline{\beta}\alpha \equiv 0 \pmod{p}$ then $\gcd(p,\alpha) = \gcd(p,\beta) = \pi$ with $N(\pi) = p$, either by conjugation or by a "left-handed" version of Proposition 4.1. So in this case, π is a common left divisor of α and β .

To prove the last part, we may assume that $\pi \in S_1$ by Lemma 5.1. Then (assuming that π divides α on the left, say), $\overline{\pi}\alpha \in pS \cap S_1 = pS_1$, since p is odd, so that $\pi^{-1}\alpha \in S_1$; similarly for β .

Lemma 5.3. Let F be any field, and let A and B be 2×2 matrices over F such that

$$\det(A) = \det(B) = \operatorname{tr}(AB) = 0.$$

Then either AB = 0 or BA = 0.

Proof. If A = 0 or B = 0 then the result is trivial; otherwise we have rk(A) = rk(B) = 1. Moreover det(AB) = tr(AB) = 0, so by the Cayley-Hamilton Theorem we have $(AB)^2 = 0$. Suppose that $AB \neq 0$; then rk(AB) = 1 also.

View each matrix as acting on the two-dimensional space F^2 of column vectors by left multiplication. Now $(AB)^2 = 0$ implies that $\operatorname{im}(AB) \subseteq \ker(AB)$, but since $\dim \ker(AB) = \dim \operatorname{im}(AB) = 1$, equality holds. Hence

$$\ker(B) \subseteq \ker(AB) = \operatorname{im}(AB) \subseteq \operatorname{im}(A);$$

but $\dim \ker(B) = \dim \operatorname{im}(A) = 1$, so again equality holds. Hence BA = 0.

The algorithm to do the shrinking in four dimensions can now be presented.

Algorithm: Shrink Polytopes

Input: $\{x_r = (a_r, b_r, c_r, d_r)\}$ * a set of n points which defines a polyhedron in \mathbb{Z}^4 *;

Output: $\{y_r = (a_r, b_r, c_r, d_r)\}$ * a set of n points which defines a similar polyhedron of minimum size in \mathbb{Z}^4 *;

- (1) BEGIN
- (2) FOR r = 1 TO n DO: $x_r \leftarrow (a_r, b_r, c_r, d_r) \leftarrow x_r x_1$;
- (3) FOR r = 1 TO n DO: $\alpha_r \leftarrow a_r + ib_r + jc_r + kd_r$;
- (4) $\pi_1 \leftarrow \gcd(\alpha_1, \ldots, \alpha_n);$
- (5) FOR r = 1 TO n DO: $\beta_r \leftarrow \pi_1^{-1} \alpha_r$;
- (6) $\pi_2 \leftarrow \operatorname{gcrd}(\beta_1, \dots, \beta_n);$
- (7) FOR r = 1 TO n DO: $a_r + ib_r + jc_r + kd_r \leftarrow \gamma_r \leftarrow \beta_r \pi_2^{-1}$;
- (8) FOR r = 1 TO n DO: $y_r \leftarrow (a_r, b_r, c_r, d_r)$;
- (9) END.

Proposition 5.4. The algorithm Shrink Polytopes takes as input a set of n points x_1, x_2, \ldots, x_n in \mathbb{Z}^4 which are the vertices of a four-dimensional polytope, and returns a set of n points in \mathbb{Z}^4 which are the vertices of a similar polytope of minimal size. It does so in $O(n \log N)$ steps, where $N = \max\{|x_1|, |x_2|, \ldots, |x_n|\}$.

Proof. We first shift the origin in line (2), then convert the points x_r to elements $\alpha_r \in S_1$ in line (3). By Lemma 2.2 and Proposition 5.2, shrinking is possible if and only if the α_r have a non-trivial common divisor on the right or left. So if we divide out by their g.c.l.d., and then by the g.c.r.d. of the resulting points, we are guaranteed to have shrunk the given configuration to a minimal similar one.

In line (4) we compute $\pi_1 = \gcd(\alpha_1, \alpha_2, \dots, \alpha_n)$, and then $\beta_r = \pi_1^{-1}\alpha_r \in S$ in line (5). By the second remark following Lemma 5.1 and Proposition 5.2 we may assume that each β_r is in fact in S_1 . Next in line (6) we compute $\pi_2 = \gcd(\beta_1, \beta_2, \dots, \beta_n)$, and then $\gamma_r = \beta_r \pi_2^{-1} \in S$ in line (7). Again we may assume that each $\gamma_r \in S_1$. The map $\lambda(\pi_1^{-1})\rho(\pi_2^{-1})$ which maps $\alpha_r \mapsto \pi_1^{-1}\alpha_r\pi_2^{-1} = \gamma_r$ is then a similarity which shrinks the configuration by a factor of $(N(\pi_1)N(\pi_2))^{1/2}$. The resulting points γ_r have trivial g.c.r.d. and g.c.l.d.; it follows from Proposition 5.2 that the corresponding vectors

 $y_r \in \mathbb{Z}^4$ satisfy $\gcd_{\mathbb{Z}}(\{y_r \cdot y_s\}) = 1$, so by Lemma 2.2 no further shrinkage is possible.

The running time of the algorithm is again dominated by the two g.c.d. computations on lines (4) and (6). This gives a running time of $O(n \log N)$ where N is the maximum of the $|x_r|$.

References

- [1] Dickson, L.E. Algebras and their arithmetics, Dover, New York 1960.
- [2] Herstein, I.N. *Topics in Algebra*, 2nd edn., Xerox College Publ., Lexington, Mass., 1982.
- [3] Pierce, R.S. Associative Algebras, Springer-Verlag, Berlin-Heidelberg-New York 1982.
- [4] Rabin, M. and Shallit, J., Randomized Algorithms in Number Theory, Comm. in Pure and Applied Math. 39 No.5 (1986), pp.239–256.
- [5] Rees, E.G. Notes on Geometry, Springer-Verlag, Berlin-Heidelberg-New York 1983.
- [6] Rolletschek, H. (1986) On the Number of Divisions of the Euclidean Algorithm Applied to Gaussian Integers. J. Symb. Comput. 2, pp.261–291.
- [7] Snapper, E. and Troyer, R.J. Metric Affine Geometry, Academic Press, New York, 1971.