UNIT DISTANCE PRESERVING MAPPINGS OF THE PLANE INTO ITSELF

Gundorph K. Kristiansen

December 25th 2007

Abstract.

Consider a norm-derived metric of the plane. Let f map the plane into itself in such a way that any two points with distance 1 from each other are mapped on two points with the same property. If the norm is the Euclidean one, f is an isometry ([1]). I prove that f is an isometry if only the unit circle (i.e. the set of points with distance 1 from the origin) is not a parallelogram.

General considerations

Points P in the plane will usually be denoted and treated as complex numbers. Mostly, it will not be necessary to distinguish between the point P and the vector $\overrightarrow{OP}$ from the origin to the point.

For instance, $\|z\|$ stands for the norm of the vector connecting the origin with the point z.

The metric of the plane is determined by the unit circle S, i.e. the set of points z with $\|z\| = 1$. Accordingly, we shall denote such a plane by $\mathbb{C}(S)$.

The set of points Q with $\|\overrightarrow{PQ}\| = 1$ will be denoted $S(P)$.

Our goal in the present paper is to investigate distance 1 preserving mappings of a plane $\mathbb{C}(S)$ into itself.

We shall need some properties of unit circles.

The unit circle S is the boundary of the open unit disk U, the set of points with norm less than one.

The open unit disk U is convex, bounded, and symmetric with respect to the origin, which belongs to U.

On the other hand, any subset of the plane with these properties is the open unit disk for some norm. To see this, we first note that if $P \in U \setminus \{0\}$, we can define the positive real number t_m as $\sup\{t \in \mathbb{R} \mid tP \in U\}$. Then $t_m P \in S$. But there can be only one positive real number t with $tP \in S$. In fact, assume that t' were another such number. Because of the convexity of U we must then have $t' > t_m$. Let now N be a neighbourhood of O contained in U. Evidently, we can find a point $Q \in U$ so close to $t'P$ that the convex hull of $N \cup \{Q\}$ contains $t_m P$, and we have a contradiction. Thus, we can define $\|P\| = 1/t_m$. Similarly, for P outside U we

1991 *Mathematics Subject Classification.* 52A10.

Key words and phrases. Normed plane, Beckman-Quarles.

can find exactly one positive number t such that $tP \in S$, and we define $\|P\| = 1/t$. The function $\|\cdot\|$ is easily seen to be a norm with S as unit circle.

The remarks above are just applications of the theory of Minkowski functionals (see for instance [2, page 24]).

Now, some auxiliary results,

Lemma 1. *Let P_1, P_2 and P_3 be three collinear points, all belonging to S. Then their convex hull, an interval I, also belongs to S.*

Proof.

We can assume that $I = [P_1, P_3]$, and that P_2 is an interior point of I. Let $P_4 \in I$ be a point belonging to U. As above we can define a real number $t_m > 1$ such that $P_4' = t_m P_4 \in S$. Next choose a real number $t \in (0, 1)$ close to 1 such that the triangle with vertices tP_1, tP_3, and tP_4', all belonging to U, contains P_2 in its interior, a contradiction.

We have seen that for each angle θ we can define exactly one positive real number $r(\theta)$ such that $r(\theta)e^{i\theta} \in S$.

For each θ there is a line L_P (in general not unique) through $P = r(\theta)e^{i\theta}$ such that L_P does not contain any points of U. In fact, assume that there were no such line. So consider the lines $L(\phi) = \{P + te^{i\phi} \mid t \in \mathbb{R}\}$. There is no point in $L(\theta) \cap U$ with a positive value of t, but there is in $L(\theta + \pi) \cap U$. And so we can define ϕ_m as the supremum of the angles $\phi \in (\theta, \theta + \pi)$ such that the positive half of $L(\phi)$ has no point in common with U. According to assumption the negative part has, and for ϕ slightly less than ϕ_m both parts of $L(\phi)$ have points in common with U, contradicting Lemma 1.

Actually, because U is bounded, and because a certain neighbourhood of the origin belongs to U, there is a positive number ϵ such that for any P the angle between L_P and $[O, P]$ is greater than ϵ.

Let $P = r(\theta)e^{i\theta}$. Let $Q = r(\theta')e^{i\theta'}$, where the variable θ' tends towards θ, for definiteness increasing. When θ' is sufficiently close to θ, the point Q will lie between the lines $\{P + te^{i(\theta - \pi + \epsilon)} \mid t \in \mathbb{R}\}$ and $\{P + te^{i(\theta - \epsilon)} \mid t \in \mathbb{R}\}$. Thus Q approaches P, $r(\theta)$ is continuous, and S is a Jordan curve.

Remark. When, in the following, without further explanation, two Jordan curves J_1 and J_2 are stated to intersect, the reasoning behind is always that one can find one point of J_2, say, inside J_1, and another point of J_2 outside J_1.

On notation. When discussing positions of points along a curve we shall often use inequalities like $\arg\theta_1 < \arg\theta_2$. This will be taken to mean that we can find values of the arguments such that $\arg\theta_1 < \arg\theta_2 < \arg\theta_1 + \pi$.

Orient L_P in the usual way, i.e. the positive half of L_P is the one intersected by the line through O and Q when $\arg Q$ is slightly greater than $\arg P$. Consider L_Q for such a point, and assume that $L_Q \neq L_P$. Then the point $R = L_Q \cap L_P$ cannot lie on the negative part of L_P. We may have $R = P$, in which case Q must be on the same (the left) side of L_P as O is, and the angle from L_P to L_Q is positive. Otherwise R belongs to the positive part of L_P, and P is situated to the same side of L_Q as O. Again the angle from L_P to L_Q must be positive.

Note that this result does not (in case of non-uniqueness) depend on which lines L_P and L_Q are chosen.

There is a similar result concerning angles between chords:

Lemma 2. *Let* z_1, z_2, z_3 *be points on* S *with*

$$\arg z_1 < \arg z_2 < \arg z_3 \leq \arg z_1 + \pi.$$

Then

$$\arg(z_2 - z_1) \leq \arg(z_3 - z_1) \leq \arg(z_3 - z_2). \tag{1}$$

Proof.
If a point in (z_1, z_3) belongs to S, the whole interval $[z_1, z_3]$ is contained in S, and we have equality in (1) (because of the restriction $\arg z_3 \leq \arg z_1 + \pi$ the point z_2 belongs to (z_1, z_3)). Otherwise, $(z_1, z_3) \subset U$, and z_2 and O are on opposite sides of $[z_1, z_3]$. We can orient the coordinate axes such that $\arg(z_3 - z_1) = 0$. Then $\Im z_2 < \Im z_3 = \Im z_1 < 0$, from which follows strict inequality in (1).

In the following, the notions distance and length will always be those induced by the norm for which S is the unit circle.

Lemma 3. *Let* P_1 *and* P_2 *be two points with distance at most equal to* 2. *Then the set* $M = S(P_1) \cap S(P_2)$ *is symmetric with respect to the midpoint of the interval* $[P_1, P_2]$. *There are now the following possibilities:*

(1) *M consists of two points with sum $P_1 + P_2$;*
(2) *M consists of a single interval. In this case $\|\overrightarrow{P_1P_2}\| = 2$;*
(3) *M consists of two intervals parallel to an edge of the unit circle and parallel to $\overrightarrow{P_1P_2}$. The length of each interval is equal to the length of the edge minus $\|\overrightarrow{P_1P_2}\|$. Every point in one of the intervals has distance 2 from every point in the other interval.*

Proof.
To see that M is symmetric with respect to the midpoint of the interval $[P_1, P_2]$, do a simple calculation: If $Q \in M$, also $P_1 + (P_2 - Q) = P_2 + (P_1 - Q) \in M$, and the midpoint of the interval $[Q, P_1 + P_2 - Q]$ is $(P_1 + P_2)/2$.

It suffices to consider the case where M consists of more than two points.

Assume first that $(P_1 + P_2)/2 \in M$. From Lemma 1 it follows that if $Q \in M$, the whole interval $[Q, P_1 + P_2 - Q]$ belongs to M. But a halfline from P_1 just missing $(P_1 + P_2)/2$ can intersect M in only one point. And so M consists of a single interval through $(P_1 + P_2)/2$.

Otherwise $\|\overrightarrow{P_1P_2}\| < 2$, and no point of the straight line L through P_1 and P_2 belongs to M. Imagine L as horizontal.

We can assume the existence of two points Q_1 and Q_2 belonging to M and situated above L.

Now $S(P_2)$ contains, in addition to the points Q_j also the points

$$R_j = P_2 + (Q_j - P_1) = Q_j + (P_2 - P_1) \quad (j = 1, 2),$$

obtainable by translating Q_j by the vector $\overrightarrow{P_1P_2}$.

Let, for $j = 1, 2$, L_j be the line through Q_j and R_j.

Assume that the two parallel lines L_1 and L_2 do not coincide. For definiteness, let L_1 separate L_2 and L.

$S(P_2)$ is symmetric with respect to P_2 and therefore also contains the points $Q'_j = 2P_2 - Q_j$ and $R'_j = 2P_2 - R_j$.

Consider the parallelogram with vertices Q_2, R_2, Q'_2, and R'_2. The interior of this parallogram must belong to $U(P_2)$ and therefore cannot contain points of $S(P_2)$ like Q_1 and R_1. But then these two points must belong to the boundary of the parallelogram, i.e. we must have $Q_1 \in [Q_2, R'_2]$. Now, $R'_2 = 2P_2 - (Q_2 + P_2 - P_1) = P_1 + P_2 - Q_2$, and so, according to Lemma 1, we have the first case, contrary to assumption.

That L_1 and L_2 coincide, means that the intervals $[P_1, P_2]$ and $[Q_1, Q_2]$ are parallel to the same edge of $S(P_2)$. Actually, this edge contains each interval $[Q_j, R_j]$, and so its length must equal the sum of the lengths of $[P_1, P_2]$ and the maximal interval $[Q_1, Q_2]$.

We introduce a coordinatesystem with x-axis along the line through P_1 and P_2, and origin at $(P_1 + P_2)/2$. Let the component of M situated above the x-axis stretch from $x = b$ to $x = c$. The component below the x-axis will then have x-values between $-c$ and $-b$. To find the distance between a point in the first interval (at $x = d_1$, say) and a point in the second interval at $x = d_2$, we find a line through the origin parallel to the line connecting these two points. This line will obvously intersect the first interval at $x = (d_1 - d_2)/2 \in [b, c]$, i.e. at a point of M, and so have length 1. This means that the sought distance must be 2, q.e.d.

A trivial but useful consequence of Lemma 3 is

Lemma 4. *If an S-edge has length 2, the unit circle is a parallelogram.*

Proof.

Let $[A, B]$ be the edge of length 2. Then also $[-B, -A]$ belongs to S and has length 2. The two points A and $-B$ both have distance 2 from both points B and $-A$. Thus, according to Lemma 3, A and $-B$ have distance 2 from any point in $[B, -A]$, and this interval belongs to S. The same is true for $[A, -B]$, and the lemma is proved.

Lemma 5. *The distance between a fixed point $z \in S$ and a variable point $z' \in S$ does not decrease when $\arg z'$ increases from $\arg z$ to $\pi + \arg z$.*

Proof.

Let z_1 and z_2 be points on S with

$$\arg z < \arg z_1 < \arg z_2 \leq \arg z + \pi,$$

but with

$$\|z_2 - z\| < \|z_1 - z\|. \tag{2}$$

Let $j \in \{1, 2\}$. Then

$$\arg\left(\frac{z_j - z}{z}\right) = \arg\left(\frac{z_j}{z} - 1\right) \in (0, \pi]$$

and also

$$\arg\left(\frac{z_j - z}{z_j}\right) = \arg\left(1 - \frac{z}{z_j}\right) \geq 0.$$

Thus,

$$\arg z_j \le \arg(z_j - z) \le \arg z + \pi. \tag{3}$$

Now compare the triangle with corners z, z_1, z_2 with the triangle with corners

$$O, \frac{z_1 - z}{\|z_1 - z\|}, \frac{z_2 - z}{\|z_2 - z\|}.$$

If we had equality in (2), these two triangles would be similar, and the two S-chords $[z_1, z_2]$, $[\frac{z_1-z}{\|z_1-z\|}, \frac{z_2-z}{\|z_2-z\|}]$ would be parallel. As it is, the angle from the first chord to the second one is negative. But Lemma 2 together with (3) shows this to be false. Thus (2) cannot be satisfied, q.e.d.

Lemma 6. *A closed curve S' (whose interior contains U properly) exists, such that for each point $R \in S'$ there are two points P and Q in $S \cap S(R)$, such that the distance $\|\overrightarrow{PQ}\|$ equals* 1.

Proof. We first consider a point z tracing the unit circle in the positive direction, i.e. $\arg z$ increases from 0 to 2π. In distance 1 from z we find a point $z' \in S$ with $\arg z < \arg z' < \arg z + \pi$. The problem is that sometimes z' is not uniquely determined. And it may also happen that several values of z give the same value of z'.

To handle this situation I find it convenient to regard $(\arg z, \arg z')$ as a point on a torus T^2 (T is the unit circle in Euclidean geometry). We shall show that the set of points $\{(\arg z, \arg z') \mid \|z - z'\| = 1\}$ can be parametrized as a curve on T^2.

We choose $t = \arg z + \arg z'$.

Note that the set $E = \{z \in S \mid z' \text{ is not unique}\}$ is finite. In fact, according to Lemma 3 the line from O to a point $z \in E$ must be parallel to an edge of S, such that each point of a non-empty subinterval of this edge has distance 1 from z. It also follows from Lemma 3 that the length of the edge is greater than 1. The curve S is easily shown to be rectifiable (with length at most 8), and thus the number of such edges is finite (in Lemma 9 below it is shown that this number is at most four), and so is the cardinality of E.

Similarly, the set $E' = \{z' \in S \mid z \text{ is not unique}\}$ is finite. It also follows from the above that, regarded as sets of points, the two sets E and E' are equal.

Let $z \in E$. When z', with $\arg z'$ increasing, runs through the subinterval in which $\|z - z'\| = 1$, the function $\Phi(t) = (\arg z, \arg z')$ is trivially defined and continuous in the corresponding interval of t-values, which is traced in the direction of increasing t.

Similarly, when $z' \in E'$. Here z runs through an interval with $\arg z$ increasing.

Let us now consider pairs (z, z') with $z \notin E$ and $z' \notin E'$. Since $z \notin E$, z' is uniquely determined. We can also show that in the neighbourhood of such points z, the angle $\arg z'$ is a strictly increasing function of $\arg z$. Otherwise we could find two points z_1 and z_2 with $\arg z_1 < \arg z_2$ and $\arg z_2' \le \arg z_1'$. But then, according to Lemma 5,

$$1 = \|z_2 - z_2'\| \le \|z_2 - z_1'\| < \|z_1 - z_1'\| = 1,$$

where the strict inequality is due to $z_1' \notin E'$.

Assume that $z_n \longrightarrow z_0$ through $S \setminus E$, for definiteness with $\arg z_n$ decreasing. Then the sequence $\arg z_n'$ is also decreasing, and, if $z_0 \notin E$, for all n we have

$\arg z'_n \geq \arg z'_0$. But if $z'_n \to w \neq z'_0$, we would, because of the continuity of the norm function have $\|z_0 - w\| = 1$, a contradiction. If $z_0 \in E$, the point w must be the point on S with distance 1 from z and $\arg z_0 < \arg w < \arg z + \pi$ whose argument is maximal. Similarly, the point w obtained as limit for a sequence z'_n corresponding to a sequence z_n approaching a point $z_0 \in E$ counter-clockwise has minimal value of its argument.

We conclude that in a neighbourhood of a point $z \notin E$ the angle $\arg z'$ is a continuous and non-decreasing function of $\arg z$. Thus t is here a continuous and strictly increasing function of $\arg z$. The inverse function is then also continuous, so that the function pair $(\arg z, \arg z')$, which, as we have seen earlier, is continuous for t-values interior to the intervals corresponding to $z \in E$ and $z' \in E'$, is continuous everywhere. Since, as we have seen, z is a continuous function of $\arg z$ (and z' of $\arg z'$) both z and z' become continuous functions of t and describe the unit circle when t runs through an interval $[0, 4\pi)$. The point $z + z'$ describes a closed curve S', which obviously satisfies our requirements (the proof that the interior of S' contains U properly is postponed to the next section).

Lemma 7. *If we can find four points $P_j \quad (j = 1, 2, 3, 4)$ such that the distance between any two points among them is equal to 1, the unit circle is a parallelogram.*

Proof.

Assume first that $S(P_1) \cap S(P_2) = \{P_3, P_4\}$. Then, according to Lemma 3, we have $P_3 + P_4 = P_1 + P_2$. If, at the same time, we had $M = S(P_1) \cap S(P_3) = \{P_2, P_4\}$, we would also have $P_3 + P_1 = P_2 + P_4$, implying $P_4 = P_1$, which is incompatible with $\|P_1 - P_4\| = 1$. Thus M must contain more than two points, and so, according to Lemma 3, consist of two intervals parallel to $[P_1, P_3]$. But P_2 and P_4 must then belong to the same component of M, as otherwise, again according to Lemma 3, their distance would be 2. The S-edge to which $[P_1, P_3]$ and $[P_2, P_4]$ are parallel, must have maximal length 2. And so, according to Lemma 4, S is a parallelogram.

It follows from our assumption that P_3 and P_4 are on opposite sides of the line through P_1 and P_2. Since the four points apparently are vertices of a parallelogram Π, $[P_1, P_2]$ and $[P_3, P_4]$ must be diagonals, and $[P_1, P_4]$ and $[P_3, P_2]$ are parallel.

If we did not have $S(P_1) \cap S(P_2) = \{P_3, P_4\}$, rôles would be interchanged, and we would find that $[P_1, P_2]$ and $[P_3, P_4]$ were sides of a parallelogram. Again there would be an S-edge of maximal length, and S would be a parallelogram.

Lemma 8. *It is possible to construct a hexagon with all vertices on the unit circle and with all sides of length 1. One vertex may be specified arbitrarily.*

Only if the unit circle is a parallelogram, is it possible to inscribe in it polygons with three, four, five, seven or eight sides, all of length 1.

Proof.

The hexagon construction is illustrated in Figure 1.

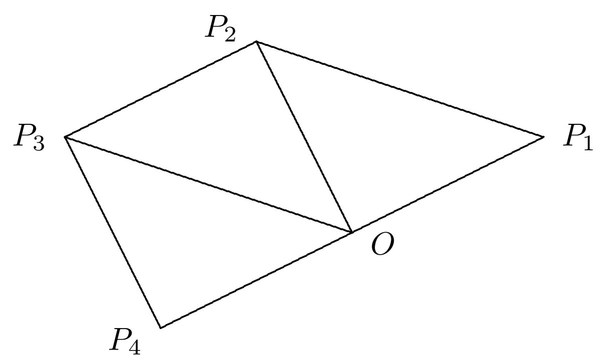

Figure 1

Here the point O is the centre of a unit circle S in the given norm, and P_1 is an arbitrary point on S. The point $P_4 \in S$ is its opposite point, and $P_2 \in S$ is chosen as a point having distance 1 from P_1. The point $P_3 \in S$ is the point found by going from O in a direction parallel to $\overrightarrow{P_1P_2}$. Then $\overrightarrow{P_3P_2}$ becomes parallel to $\overrightarrow{OP_1}$ (and so also to $\overrightarrow{P_4O}$) and is easily seen to have length 1. Finally, $\overrightarrow{P_4P_3}$ becomes parallel to $\overrightarrow{OP_2}$ and has length equal to 1. Let P_5 and P_6 be the points on S opposite to P_2 and P_3, respectively. Then the hexagon $P_1P_2P_3P_4P_5P_6P_1$ is the one we were looking for.

According to Lemma 3 each vertex is uniquely determined by the previous one, and the hexagon is the only possible inscribed polygon with all sides of length 1 (in the following just called an inscribed polygon), unless S has an edge of length greater than one.

We use Figure 1 for the further analysis of inscribed polygons.

The procedure used in the construction of Figure 1 can be formalized in the following way:

For $j \geq 2$ we put

$$P_{j+1} = P_j - P_{j-1}. \tag{4}$$

For $j \geq 3$ we substitute in (4) the expression obtained from (4) by replacing j by $j-1$. This gives $P_{j+1} = -P_{j-2}$. With j understood modulo 6 this shows that (4) is, in fact, valid for all j.

In the following we shall discuss unit circles S containing at least one edge (and so, because of the symmetry with respect to the origin, at least 2 edges) of length greater than 1.

Then P_{j+1} is not, as in (4), uniquely determined from P_j and P_{j-1}, but may, for $j \geq 2$, be replaced by

$$P'_{j+1} = P_{j+1} + t_{j+1}P'_j, \tag{5}$$

where (4) has been changed to

$$P_{j+1} = P'_j - P'_{j-1}, \tag{4'}$$

and we have put $P'_j = P_j$ for $j = 1, 2$.

In (5), the real number t_{j+1} cannot be chosen arbitrarily.

We first note that (5) can be applied with $t_{j+1} \neq 0$ only when the intersection $S \cap S(P'_j)$ contains the interval $[P_{j+1}, P'_{j+1}]$. Thus S must have an edge E_{j+1} containing the interval $[P_{j+1} - P'_j, P_{j+1} + t_{j+1}P'_j]$ for $t_{j+1} > 0$, and the interval $[P_{j+1} - P'_j + t_{j+1}P'_j, P_{j+1}]$ for $t_{j+1} < 0$. Both intervals have length $1 + |t_{j+1}|$, so that we must have $|t_{j+1}| \leq 1$ with equality only if S is a parallelogram. Note that $P_{j+1} - P'_j = -P'_{j-1}$, so that $-P'_{j-1} \in E_{j+1}$, and $P'_{j-1} \in -E_{j+1}$, where $-E_{j+1} \subset S$ is the interval symmetric to E_{j+1} with respect to the origin. If $t_{j+1} > 0$, E_{j+1} contains P_{j+1} and P_{j+2} as interior points. If $t_{j+1} < 0$, E_{j+1} contains P'_{j+1} and $-P'_{j-1}$ as interior points.

Next we show that if, for some $j \geq 3$, we have $t_j > 0$, then we cannot have $t_{j+1} \neq 0$. In fact, E_j contains P_{j+1} as an interior point, and so P_{j+1} cannot simultaneously belong to another S-edge.

But assume that, for some $j \geq 3$, we have $t_j < 0$ and $t_{j+1} > 0$. Then E_j contains the interval $[P_{j+1}, P_j]$, and the general point in this interval can be written $P'_j - uP'_{j-1}$, where $t_j \leq u \leq 1]$. We find the point of intersection between this interval and the line through O and P'_{j+1}:

$$P'_j - uP'_{j-1} = k(P'_j - P'_{j-1} + t_{j+1}P'_j).$$

We find $k = u = 1/(1 + t_{j+1}) < 1$. Thus S separates O and P'_{j+1}, an impossibility. And so $t_{j+1} > 0$ implies $t_j = 0$, q.e.d.

We see that if, for some $j \geq 3$, we have $t_{j+1} > 0$, we must have $t_j = t_{j+2} = 0$.

In the following, we shall always assume that the vertices of the inscribed polygon are indexed in such a way that for all j we have $\arg P'_j < \arg P'_{j+1}$.

The point P_3 may have distance 1 from P_1. This is possible iff S is a parallelogram (Lemma 7). We then have an inscribed triangle.

Next we examine the possibility of inscribed quadrangles and pentagons. Here the average increase $\arg P'_{j+1} - \arg P'_j$ is larger than obtained by the construction in Figure 1. In particular, we can choose indexes such that $P'_3 \neq P_3$. Now use (5) for $j = 2$, divided by P_3. Since $\Im(P_2/P_3) < 0$, we must choose $t_3 < 0$ to make $\Im(P'_3/P_3) > 0$ and thus $\arg P'_3 > \arg P_3$. Then, according to the above, $t_4 \leq 0$. We have

$$P'_3/P_1 = (1 + t_3)(P_2/P_1) - 1,$$

and so $\arg P_1 < \arg P'_3 \leq \arg P_1 + \pi$ with equality only if $t_3 = -1$,

$$P'_4/P'_3 = 1 + t_4 - P_2/P'_3,$$

showing that $\arg P'_3 < \arg P'_4$.

We conclude that if $t_3 > -1$ the line L through O and P'_3 separates P_1 from P'_4. However, the interval $[O, P'_3]$ belongs to $S(P'_4)$, and P'_4 has distance greater than 1 to P_1, unless $t_3 = -1$, which makes $P'_3 = -P_1$ and necessitates that S is a parallelogram. Then the line through O and P_2 separates P'_4 from P_1 unless $t_4 = -1$, making $P'_4 = -P_2$. Actually, E_3 contains the interval $[P_3, -P_1 - P_2]$, and so $S(P_1)$ contains $[P_2, -P_2]$ and thus P'_4.

With respect to inscribed pentagons, we continue the reasoning above, but we are now looking for a point P'_5 in the intersection $S \cap S(P'_4) \cap S(P_1)$.

First consider the possibility $t_4 = 0$, i.e. $P'_4 = P_4$. We have $P_4 \in E_3$ and $P_1 \in -E_3$, and their distance is 2. This is Case 2 of Lemma 3. The intersection

$S(P_4)\cap S(P_1)$ is an interval I containing O and with centre at the point $(P_1+P_4)/2$, which must belong to the line through O and P_2. Thus the only candidate for P_5' is $P_5 = -P_2$, which makes the whole interval $[-P_2, P_2]$ a part of $S(P_1)$. So S is a parallelogram, and the construction works for all $t_3 \in [-1, 0]$.

The only remaining possibility is $t_4 < 0$. Now $[P_4, P_4']$ is part of $[P_5, P_4] \subset E_4$, which is parallel to $[O, P_3']$ and contains $-P_2$. The S-edge $-E_4$ goes through P_2 and intersects the S-edge $-E_3$ through P_1 at $-P_4$. The usual calculation confirms that the interval $[P_1, P_2]$ intersects the interval $I = [O, -P_4']$ at an interior point of I unless $t_4 = -1$ and $P_4' = -P_2$. For the moment disregarding this last possibility, we see that the distance from P_4' to P_1 is less than 2, and so we do not have Case 2 of Lemma 3. We are not interested in Case 2 either, since we want a point $P_5' \in S$. Thus the only possibility is that $S(P_4') \cap S(P_1)$ consists of the two points O and $P_4' + P_1$, the latter point being the only candidate for P_5'. When t_4 decreases from 0 to -1 the point $P_4' + P_1$ moves linearly from t_3P_2, which is clearly inside S, to the point $-P_3$, which must be P_5'. But this presupposes $t_4 = -1$, so that S is a parallelogram, just as in the case $P_4' = -P_2$ above.

Note that if S is a parallelogram, all inscribed pentagons have three consecutive vertices as S-edge midpoints.

Finally consider inscribed polygons with more than six vertices. Here the average increase $\arg P_{j+1}' - \arg P_j'$ is smaller than obtained by the construction in Figure 1. In particular, we can choose indexes such that $\arg P_3' < \arg P_3$, i.e. have $t_3 > 0$. Then, according to the above, $t_4 = 0$, $P_4 = P_3' - P_2$, $P_5 = P_4 - P_3' = -P_2$, so that not very much has been gained. We may, however, choose $t_5 \neq 0$.

But start with $t_5 = 0$ (no extra S-edges in addition to the old ones $\pm E_3$, where E_3 contains $[-P_1, P_3']$). We have $P_6 = P_5 - P_4 = -P_3'$. Whatever value t_6 has, the point $P_6' = P_6 + t_6P_5$ belongs to the straight line L through $-P_3$ and P_1. But L contains the S-edge $-E_3$ to which P_6' must belong. The point P_7 with $\arg P_6' < \arg P_7 < \arg P_1$ must belong to $-E_3$ also, and so P_6' must have distance 2 from P_1, i.e. $P_6' = P_1 - 2P_2$. Since $P_6' = -P_3' - t_6P_2 = P_1 - (1 + t_3 + t_6)P_2$, we must have $t_3 + t_6 = 1$. And since we have an S-edge of length 2, according to Lemma 4 the unit circle must be a parallelogram.

Next consider the possibility $t_5 < 0$. This means that S has an edge E_5 containing $[P_6, P_5]$ and having $-P_3'$ as an interior point. But $-P_3'$ should also belong to the S-edge $-E_3$, which is impossible.

So we are left with the possibility $t_5 > 0$. Now $t_6 = 0$, and $P_6 = P_5' - P_4$ is an interior point on the S-edge E_5 containing the interval $[-P_3', P_5']$. The arc on S from P_6 to P_1 is contained in the union of the S-edges E_5 and $-E_3$. The unit circle $S(P_5')$ contains the interval $[O, P_4]$, the continuation of which intersects the S-edge $-E_3 = [-P_3', P_1]$ in the interior point $-P_4$. Thus $S(P_6)$ contains the interval $[-P_4, O]$, and P_6 and P_1 are on opposite sides of the line containing this interval. And so $\|P_6 - P_1\| > 1$.

An upper bound for $\|P_6 - P_1\|$ is found as the length of the arc on S from P_6 to P_1, i.e. $1 + t_3 + t_5$. Three more edges in the inscribed polygon requires $t_3 = t_5 = 1$, i.e. S must be a parallelogram. Then $P_3' = 2P_2 - P_1$, while $P_4 = P_2 - P_1 = P_3$, $P_5' = P_3' - 2P_2 = -P_1$, $P_6 = -P_2$, and $P_7 = P_1 - P_2 = -P_3$. If we form $P_7 - P_6$, we get P_1 again, and we have a heptagon. But with $t_7 = 1$ we obtain $P_7' = -P_2 - P_3 = -P_3'$ and $P_8 = -P_3$, an octagon. Actually, the octagon is just the unit circle, but with the edge midpoints promoted to vertices.

In general, if we are satisfied with a heptagon, we must have the vertex $P_7' =$

$P_1 - P_2$ and $\|P_6 - P_7'\| = 1$. But

$$P_6 = P_5' - P_4 = P_5 + (t_5 - 1)P_4 = - P_2 + (t_5 - 1)(P_3 + (t_3 - 1)P_2)$$
$$=((-1 + (t_5 - 1)t_3)P_2 + (1 - t_5)P_1,$$

and

$$P_7' - P_6 = t_5P_1 + (1 - t_5)t_3P_2 = N\left(\frac{t_5}{N}P_1 + \frac{(1-t_5)t_3}{N}P_2\right),$$

where

$$N = t_5 + (1 - t_5)t_3 = 1 - (1 - t_5)(1 - t_3).$$

Thus $P_7' - P_6$ equals N times a point on the chord $[P_1, P_2]$. If the distance between P_7' and P_6 has to be 1, we must have $1 \leq N$, which can only be true if one of t_3, t_5 is equal to one. Furthermore, the chord $[P_1, P_2]$ must actually be part of S. The two conditions can be realized. Each heptagon has two of the vertices being consecutive vertices of the parallelogram S, three of the vertices midpoints of the S-edges adjoining the two S-vertices mentioned, and the remaining two vertices of the heptagon being points with distance 1 chosen on the fourth S-edge.

Lemma 9. *A unit circle can have at most four edges of length greater than* 1.

Proof.

Assume that we have a unit circle S with six edges of length greater than 1. Try by alteration of S to increase the smallest edge-length.

Let the edges be $[P_j, Q_j]$ ($j = 1, 2, \ldots, 6$), where, for all j, $\arg P_j < \arg Q_j \leq \arg P_{j+1}$, $P_{j+3} = -P_j$, $Q_{j+3} = -Q_j$, and j is understood modulo 6.

We wish to simplify the unit circle while not decreasing the lengths of the six edges mentioned. Here we must remember that the unit of length in a particular direction is found by drawing a halfline L in this direction from the origin. Let P be the point of intersection $L \cap S$. Then $\overrightarrow{OP}$ is the unit vector in the given direction.

We shall at each step replace S by a new unit circle S', such that the edges of length greater than 1 are obtained by translation of the similar edges of S, and such that the points of S' are inside or on S. In this way the edges are not changed, but the unit vectors are either unchanged or shorter than in S. This means that the lengths of the six edges are not decreased when S is replaced by S'.

At each step we get rid of two "gaps"(arcs from Q_j to P_{j+1} for some value of j) in S. In fact, the arc from $-P_{j+1}$ (in the counter-clockwise direction) to Q_j is moved by the vector $\frac{1}{2}\overrightarrow{Q_jP_{j+1}}$, while the arc from P_{j+1} to $-Q_j$ is moved by $-\frac{1}{2}\overrightarrow{Q_jP_{j+1}}$. The new unit circle S' is strictly inside the old one S, except possibly at the points $\pm\frac{1}{2}(Q_j + P_{j+1})$, which are common to S and S', if S is linear from Q_j to P_{j+1}. The new unit circle is easily seen to satisfy the requirements (convexity of U', symmetry with respect to the origin, because two points with sum zero are mapped onto two points with sum zero).

After three such steps we end up with a unit circle which is a hexagon, and it suffices to derive a contradiction for unit circles of this type.

We denote the vertices in S by $P_1, \ldots, P_6$, taken in counter-clockwise order. Choose the coordinate axes such that $\arg P_1 = 0$. Assume that $\Im(P_2) \geq \Im(P_3)$. Let the unit vectors in the directions $\overrightarrow{P_3P_2}$, $\overrightarrow{P_1P_2}$, and $\overrightarrow{P_4P_3}$ be $\overrightarrow{OQ_{32}}$, $\overrightarrow{OQ_{12}}$, and $\overrightarrow{OQ_{43}}$, respectively.

We have

$$\overrightarrow{OQ_{43}} = \overrightarrow{P_4P_3}/\|\overrightarrow{P_4P_3}\|,$$

and so $\Im(Q_{43}) < \Im(P_3)$. This means that Q_{43} must belong to the open S-edge (P_1, P_2). It also follows that $\arg Q_{32} > 0$, so that $Q_{32} \in (P_1, P_2)$. Since $\|\overrightarrow{P_3P_2}\| > 1$, the point $P_2 - Q_{32}$ must belong to (P_3, P_2). As Q_{32} belongs to (P_1, P_2) and has distance 1 from P_2, we must have $P_2 - Q_{32} = Q_{12}$.

Consider the line L through O parallel to (P_1, P_2). It intersects S in the points $\pm Q_{12}$. As we have just seen, Q_{12} belongs to (P_3, P_2).

The vector $\overrightarrow{OQ_{43}}$ is a unit vector in the direction $\overrightarrow{P_6P_1}$, and since the length of the latter vector is greater than 1, we must have $P_1 - Q_{43} \in (P_6, P_1)$. But, as we have seen, $Q_{43} \in (P_1, P_2)$, and so $P_1 - Q_{43} = -Q_{12}$. This means that $-Q_{12} \in (P_6, P_1)$, while, as shown above, $Q_{12} \in (P_3, P_2)$, which would make $-Q_{12} \in (P_5, P_6)$. This contradiction establishes the lemma.

Properties of unit distance preserving maps

In the following we shall consider only norms for which the unit circle is not a parallelogram. When it is natural to emphasize this, we shall write "$S \neq$ par".

Let f map a normed plane $\mathbb{C}(S)$ into itself in such a way that any two points with distance 1 from each other are mapped on two points with the same property.

The set of vertices of a triangle whose sides all have length 1 will be called a unit triangle. We note that a unit triangle is mapped by f on a unit triangle.

Next we use the curve S' of Lemma 6 in the following way: Assume that a point P has distance 1 from a point $z + z' \in S'$, where, as usual, z and z' are points of S with $\|z - z'\| = 1$. If we had $f(P) = f(O)$, this point would have distance 1 from each member of the unit triangle $\{f(z), f(z'), f(z + z')\}$, contradicting Lemma 7.

A consequence is

Lemma 10. *A mapping f preserving distance 1 cannot map two points whose distance belongs to the interval $(0, 2]$, onto one point.*

Proof. Again we let z and z' be points of S with $\|z - z'\| = 1$. Then

$$\begin{aligned} \|z + z'\| &\leq \|z\| + \|z'\| = 2 \\ \|z + z'\| &\geq 2\|z\| - \|(z - z')\| = 1. \end{aligned} \tag{6}$$

A point in S' cannot also belong to S, since the point O would then have distance 1 from each point in a unit triangle (even if S were a parallelogram, we could not have $S = S'$: if z and z' belong to the same S-edge, we have $\|z + z'\| = 2$. This concludes the proof of Lemma 6). Thus we have,

$$R \in S' \implies 1 < \|\overrightarrow{OR}\| \leq 2. \tag{7}$$

To prove that f cannot map two points whose distance belongs to the interval $[1, 2]$, onto one point, it suffices to consider points P whose distance from O belongs to the half open interval $(1, 2]$.

Then a point Q_1 defined by $\overrightarrow{OQ_1} = (1 + 1/\|\overrightarrow{OP}\|)\overrightarrow{OP}$ belongs to $S(P)$ and lies outside S', while the point $Q_2 \in S(P)$, defined by $\overrightarrow{OQ_2} = (1 - 1/\|\overrightarrow{OP}\|)\overrightarrow{OP}$, has a distance from O which belongs to the half open interval $(0, 1]$, and so Q_2 is inside S'. Thus $S(P)$ intersects S' and $f(P) \neq f(O)$.

We are left with the more difficult case where f is assumed to map two points whose distance belongs to the interval $(0, 1)$ onto one point.

So we shall assume that $0 < \|z - z'\| < 1$, and that $f(z) = f(z')$. This implies that $f(S(z) \cup S(z')) \subset S(f(z))$. We shall show that this is impossible.

For some $m \in \mathbb{N}$ we shall construct a sequence $z_0, z_1, \ldots, z_{6m+1}$ in $S(z) \cup S(z')$ with $\|z_{j-1} - z_j\| = 1$ for $j = 1, 2, \ldots, 6m + 1$, and $z_{6m+1} = z_0$. But we shall have $f(z_{6m+1}) \neq f(z_0)$, a contradiction.

We shall need some facts concerning $M = S(z) \cap S(z')$. We know from Lemma 3 that M is symmetric with respect to the point $(z + z')/2$ and has two components. Each component is a closed interval. Take first the case where this interval reduces to a point, so that $M = \{\zeta_1, \zeta_2\}$. We have $\zeta_1 + \zeta_2 = z + z'$, and thus

$$\begin{aligned} 2(z - \zeta_1) =&(z - z') + (\zeta_2 - \zeta_1) \\ 2 \leq&\|z - z'\| + \|\zeta_2 - \zeta_1\| \\ 1 <&\|\zeta_2 - \zeta_1\|. \end{aligned} \tag{8}$$

When the components of M are intervals of non-zero length, the distance from any point in one component to any point in the other component is 2.

We now choose a particular point $\zeta_1 \in M$, namely that point where a further continuation in the positive direction along $S(z)$ would lead into $U(z')$.

Notation: For simplicity, when talking about points on a unit circle $S(z)$, we shall say that a point z_1 is to the left of another point z_2, if $\arg z_1 < \arg z_2$, in other words, if z_2 follows z_1 when we move in the counter-clockwise direction along $S(z)$.

Our first strategy: Choose z_0 on the arc A of points of $S(z)$ with distance less than 1 from and to the left of ζ_1. Each of the following points is chosen with distance 1 from the previous point and always proceeding in the counter-clockwise direction. The point z_1 is chosen on $S(z') \setminus S(z)$ (according to (8) this is always possible), and the following points again on $S(z)$, until we for some j have $z_j \in A$. Then z_{j+1} is chosen on $S(z') \setminus S(z)$ and so on. We shall try to find values of m and of z_0 with $z_{6m+1} = z_0$. We shall show below that this may not always be possible, but then a modified strategy will work.

The sequence $\Psi = (z_j)$, chosen as indicated above, satisfies the following condition:

For $j = 1, 2, \ldots, 6m$ we have $1 \leq \|z_{j-1} - z_{j+1}\| \leq 2$, which ensures that $f(z_{j-1}) \neq f(z_{j+1})$.

The condition is clearly fulfilled if the three points all belong to $S(z)$. But it is true also if z_j belongs to $S(z')$. Let, for definiteness, $j = 1$, and assume that $\|z_0 - z_2\| < 1$.

Now z' belongs to the arc from z_2 to z_0 on $S(z_1)$ inside $S(z)$, while ζ_1 belongs to the arc from z_0 to z_2 on $S(z)$, but also to the arc from z_1 through ζ_1 on $S(z')$. This arc intersects with the above mentioned arc on $S(z_1)$ from z_2 to z_0 in a point w. Then, according to Lemma 5, $1 = \|z' - w\| \leq \|z_2 - z_0\| < 1$, a contradiction.

In spite of this simplification, the sequence $\Phi = (P_j)$, where $\forall j : P_j = f(z_j)$, of points on $S(f(z))$ can be rather irregular. In fact, we must take into account the possibility that the sequence contains turning points where, for instance, for some j we have not only $\arg P_j > \arg P_{j-1}$, but also $\arg P_j > \arg P_{j+1}$ (we have here, for notational convenience, chosen $f(z)$ as the origin of the image plane). Thus even if our goal is to show that $P_0 \neq P_{6m+1}$, matters become less complicated if we decide to show the more general statement that we cannot have $P_0 = P_{2p+1}$ for any positive integer p.

According to Lemma 8 the equality $P_0 = P_{2p+1}$ is impossible for $P = 1, 2, 3$, if Φ does not contain turning points. To tackle the general situation, we use Lemma 9, from which one can infer that if there are more than two possibilities forP_{j+1} when P_j is given, then P_j must belong to an exceptional set E consisting of at most four points, which we shall call $\pm R_1$ and $\pm R_2$, and then the interval $[O, P_j]$ is parallel to an S-edge of length greater than 1. If we have chosen p minimal with the property that the equality $P_0 = P_{2p+1}$ is possible, then clearly a turning point in Φ must belong to E.

For p minimal (which we shall assume in the following) we discuss various cases:

First, assume that neither P_0, P_1, or P_2 belong to E. Since $S \cap S(P_1)$ has only two members, and one of them is P_0, according to Case 1 of Lemma 3 we have $P_1 = P_0 + P_2$. Similarly, $P_3 = P_2 - P_1 = -P_0 \notin E$. In other words, the rest of the sequence Φ can be found by using the formula (4) for all $j \geq 2$. Clearly, Φ is cyclic with period 6, and we cannot have $P_0 = P_{2p+1}$.

Next, assume that there is a $j \geq 1$ with $P_j \in E$, but that E has only two members ($E = \{\pm R_1\}$, say). Then we have a choice between $P_{j+1} = P_{j-1} + t_{j+1}P_j$ and $P_{j+1} = P_j - P_{j-1} + t_{j+1}P_j$. In the first case t_{j+1} must be non-zero, and in both cases $1 + |t_{j+1}|$ may at most equal the length of the S-edge to which $[O, P_j]$ is parallel, in particular $|t_{j+1}| < 1$. In any case, P_{j+1}/P_j cannot be real, and so $P_{j+1} \notin E$. Thus we must have $P_{j+2} = P_{j+1} - P_j$, which cannot belong to E either, so that $P_{j+3} = P_{j+2} - P_{j+1} = -P_j$, which again belongs to E.

But then $P_{j+3n} = (-1)^n P_j$ for all integers n (also negative) with $j + 3n \geq 0$. And, according to the above, P_{j+3n+1} equals either $P_{j+3n-1} + t_{j+3n+1}P_{j+3n}$ or $P_{j+3n} - P_{j+3n-1} + t_{j+3n+1}P_{j+3n}$, while $P_{j+3n+2} = P_{j+3n+1} - P_{j+3n}$.

It also follows that a ratio of the form P_{j+3n_1+k}/P_{j+3n_2}, where $k = 1$ or 2, cannot be equal to 1 (it is not even real).

It suffices then, to establish the contradiction, to consider equality of two points $P_{j+6n_1+k_1}$ and $P_{j+6n_2+k_2}$, where $\{k_1, k_2\} \in \{\{1, 4\}, \{2, 5\}, \{1, 2\}, \{4, 5\}\}$. In all these cases the point $P = P_{j+6n_1+k_1} = P_{j+6n_2+k_2}$ must have distance 1 from both points $\pm R_1$. But if, for instance, P_{j+6n+1} has distance 1 from P_{j+6n+3}, the existence of the four points O, P_{j+6n+1}, P_{j+6n+2}, and P_{j+6n+3} contradicts Lemma 7, since, according to assumption, $S \neq$ par. A similar argument can be applied to the other cases.

The general situation when $E = \{\pm R_1, \pm R_2\}$, is considerably more complicated.

We first note, however, that if we have three consecutive points not in E, then no member of Φ is contained in E. This follows from (4), which can also be read $P_{j-1} = P_j - P_{j+1}$, since $\{P_j, P_{j+1}, P_{j+2}\} \cap E = \emptyset$ then is seen to imply $P_{j-1} = -P_{j+2} \notin E$.

In the following we shall assume that Φ contains members of E.

The following observations may be of interest:

Let, for some j, P_j and P_{j+1} be outside E. Then, as above, two applications of (4) give $P_{j-1} = -P_{j+2}$, where now the two points P_{j-1} and P_{j+2} both belong to E.

For no j can we have $P_{j+1} \in \{P_j, -P_j\}$. This is due to the fact that $\|P_j - P_{j+1}\| = 1$, since, for instance, $P_j + P_{j+1} = 0$ would imply

$$1 = \|P_j - P_{j+1}\| = \| - 2P_{j+1}\| = 2,$$

which, as an equality in the real number field, is untrue.

For no j can we have $P_{j+2} \in \{P_j, -P_j\}$. First, the equality $P_j = P_{j+2}$ contradicts the minimality of p. Secondly, the equality $P_j = -P_{j+2}$ would imply that the point P_{j+1} should have distance 1 from the three collinear points O and $\pm P_j$, which, according to Lemmata 1 and 4, would be incompatible with our assumption that $S \neq$ par.

More generally, for any integers j and n with $0 \leq j < j + 2n \leq 2p$, we cannot have $P_{j+2n} = P_j$, since this would contradict the minimality of p.

For no j can we have $P_j = P_{j+3}$, since then the four points O, P_j, P_{j+1}, and P_{j+2} would be the vertices of a complete unit distance graph, contradicting Lemma 7.

It follows from the above that $p \geq 2$. It is also clear that if $P_j \in E$ (let, for definiteness, $P_j \in \{\pm R_1\}$), and $P_{j+1} \in E$ also, we must have $P_{j+1} \in \{\pm R_2\}$, while then P_{j+2} cannot belong to E at all. Thus we can have at most two consecutive points belonging to E.

Another standard argument we shall use in the following, is that we cannot have

$$\|P_j - P_k\| = \|P_j + P_k\| = 1$$

for any points P_j, P_k on S. In fact, P_j would then have distance 1 from each of the three collinear points $O, \pm P_k$, which would contradict Lemma 4, since $S \neq$ par.

Assume that we have $P_0 = P_5$.

If neither of P_2, P_3 were members of E, we would have $P_4 = -P_1$. Then

$$\|P_0 - P_1\| = \|P_0 + P_1\| = 1,$$

an impossibility.

Assume now that both P_2 and P_3 are members of E. Then none of the other points are. And so we have $P_0 = P_1 + P_4$, and $P_2 = -P_4$, which is impossible.

Let $P_2 \in E$, and $P_3 \notin E$. If $P_4 \notin E$, we have $P_0 = P_5 = -P_2$, impossible. Otherwise $P_4 \in E \setminus \{\pm P_2\}$. Since, as we have seen, $P_4 \neq -P_1$, none of the other points belongs to E. Then, by an argument similar to that above, $P_2 = -P_4$, a contradiction.

The situation where $P_3 \in E$, and $P_2 \notin E$ is handled by a procedure symmetrical to that above.

This concludes the proof that $p \geq 3$.

It will be necessary to check also the cases $p = 3$ and $p = 4$, which is done below. That this also suffices, is seen from the following argument:

As we know, there can be a "gap"of at most two ordinary points between two exceptional points, and we cannot have two such gaps after another, since this gives a cycle of six points which can be eliminated. Thus the longest index-distance D possible between two exceptional points, of which one is a repetition of the other, is $3 + 2 + 3 + 2 = 10$. Let $p \geq 5$, and let $P_j = P_{j+D}$ for some j. Then the original cycle with $2p + 1$ elements is the union of an even and an odd cycle. We can then discard the even cycle, which would contradict the minimality of p.

Before we embark on the detailed investigation of the two mentioned values of p, we mention a simple observation: we noted above that we could not have two consecutive gaps of non-exceptional points of length 2. But we cannot have two consecutive gaps of length 1 either. Assume to the contrary, that P_0, P_2, and P_4

belong to E, while P_1 and P_3 do not. Then we can put $P_0 = R_1$ and $P_2 = R_2$. We cannot have $P_4 = R_1$, since this would give a cycle of length 4 contradicting the minimality of p. If we had $P_4 = -R_1$, we would have $P_1 = R_1 + R_2$ and $P_3 = R_2 - R_1$, and we would get our usual contradiction in norm.

Now assume that we have a cycle of length 7 or 9.

First, let the maximal gap be 1 (it is clear that it cannot be 0). We may then put $P_0 = R_1$ and $P_2 = R_2$. As we just showed, we cannot have $P_3 \notin E$.

But if $P_3 \in E$, we must have $P_3 = -R_1$. Then $P_4 \notin E$, and $P_5 \in E$. And so $P_5 = -R_2$. We cannot have $P_6 = P_0$, and so P_6 cannot belong to E. But this is again the excluded possibility.

If the maximal gap is 2, we again let $P_0 = R_1$, while now both points P_1 and P_2 are outside E. Then $P_3 = -R_1$, and we must consider two cases:

In the first case, $P_4 \in E$. We put $P_4 = R_2$. Then P_5 and P_6 are not in E, and $P_7 = -R_2$ (excluding $P_0 = P_7$). If $p = 4$, i.e. $P_9 = P_0 = R_1$, we cannot have $P_8 \in E$, and so $P_8 = R_1 - R_2$. But $P_4 - P_3 = R_1 + R_2$, and we have incompatible norms.

In the second case, $P_4 \notin E$. But P_5 must belong to E, and we may put $P_5 = R_2$. To avoid a cycle of 6, we cannot have P_6 in E. As we have seen, we cannot have P_7 in E either, and so $P_8 = -R_2$ (again excluding $p = 3$). But now $P_9 - P_8 = R_1 + R_2$, while $P_4 = R_2 - R_1$, again an impossibility.

The sequence Φ having been taken care of, we look at the sequence $\Psi = (z_j)$. The important thing is here that when z_0 runs through the arc A, we want, for a fixed vakue of m, the endpoint z_{6m+1} to run through a similar (connected) arc. The only difficulty is here the values of z_j for which z_{j+1} is not uniquely determined. We must then keep z_j fixed while z_{j+1} runs through the possible values. It follows from earlier results that then each succeeding z_k (fixed k) traces a (continuous) curve.

We must give a more detailed prescription for the elements of Ψ. Note that the arc A is outside $U(z')$. For a given $z_0 \in A$ we let z_1 be the point of $S(z')$ where $S(z_0)$ enters the disk $D(z')$. Since $S(z)$ from ζ_1 onwards has an arc inside $S(z')$, and since the distance between the two components of M is greater than 1, the continuation of $S(z_0)$ beyond z_1 enters $S(z)$ later, at a point z_1'.

We define the points $z_2 = z + z_1 - z_0$ and $z_2' = z + z_1' - z_0$. Both points belong to $S(z)$, and

$$\arg(z_2 - z) = \arg(z_1 - z_0) < \arg(z_1' - z_0) = \arg(z_2' - z).$$

Thus, the journey z_0, z_1, z_2 is really a detour compared to z_0, z_1', z_2'. This means that

$$\arg(z_6 - z) < \arg(z_0 - z).$$

Obviously, by choosing z_0 sufficiently close to the left endpoint of A, we can obtain z_6 as close to z_0 as we wish. However, this is not at all what we want. We would rather that by suitable choice of z_0 we could have

$$\arg(z_7 - z) = \arg(z_0 - z),$$

as this would solve our problem with $m = 1$.

But if this is not possible, then any choice of $z_0 \in A$ yields

$$\arg(z_7 - z) > \arg(z_0 - z),$$

and either z_6 or z_7 belongs to A.

Let us now start with an arbitrary point $z_0 \in A$. The points $z_{6m} \quad (m = 0, 1, \dots)$ define a monotonically decreasing sequence $(\arg(z_{6m} - z))$. If, for some m, we have $\|z_{6m} - z_0\| \geq 1$, we may, by modification of z_0, obtain $z_{6m+1} = z_0$.

However, there is the possibility that the sequence (z_{6m}) gets stuck at the left endpoint a_{min} of A, and we must modify the arc A. The new arc is called A'. It is obtained by moving A somewhat in the positive direction along $S(z)$. In particular, the distance between the endpoints of A' is 1. We must still have the distance between the right endpoint of A' and the other component of M (the one that ζ_1 does not belong to) greater than 1. Otherwise everything works as before, except that the sequence (z_{6m}) cannot now converge towards a point in the closure of A, and it is possible to obtain $z_{6m+1} = z_0$. In fact, if the original arc A does not work, for every $z_0 \in A$ we shall have

$$\arg(a_{min} - z) < \arg(z_{6m} - z) < \arg(z_0 - z)$$

for all $m \in \mathbb{N}$, and $\lim_{m \to \infty} z_{6m} = a_{min}$. We then define the point $a_{mid} \in A$ as the point obtained as z_6 if $z_0 = \zeta_1$. When z_{6m} traces the arc $A \setminus A'$, z_{6m+1} traces the arc $A' \setminus A$, and z_{6m+7} runs through an arc B from a_{mid} to a point b_{mid} to its right. If we choose the point z_0 in the interior of B, for a certain sufficiently large m we shall have z_{6m+7} to the left of z_0. Keeping m constant, we let z_0 move to the left. Then also z_{6m} moves to the left, but does not get close to a_{min}. Thus z_{6m+7} does not approach a_{mid} and so must equal z_0 at some point, q.e.d.

To simplify, origins are, in the following, chosen such that $f(0) = 0$. However, for arbitrary $z_0 \in \mathbb{C}$ we may consider the function $g(z) = f(z_0 + z) - f(z_0)$. We have $g(0) = 0$, and g also preserves distance 1. And so, any result arrived at for f is valid for g. Thus we obtain the "long form" of the result, containing the extra parameter z_0.

We also have

Lemma 11. *Let z and z' be arbitrary points. Let the mapping f preserve distance 1. Then $\|f(z) - f(z')\| \leq \max\{-[-\|z - z'\|], 2\}$. If, in addition, z has the property that there are arbitrary large positive numbers N such that $f(Nz) = Nf(z)$, then $\|f(z)\| \leq \|z\|$.*

Proof.

If $\|z - z'\| \leq 2$, we can choose a point $z'' \in S(z) \cap S(z')$. Then

$$\|f(z) - f(z')\| \leq \|f(z) - f(z'')\| + \|f(z'') - f(z')\| = 2.$$

Otherwise we can define the number $n \in \mathbb{N}$ such that $n + 1 < \|z' - z\| \leq n + 2$ (in fact, $n = -[2 - \|z' - z\|]$), and the points $z^{(j)} = z + j(z' - z)/\|z' - z\|$ for $j = 0, 1, \dots, n$. So $\|z' - z^{(n)}\| = \|z' - z\| - n \in (1, 2]$. Then, according to the above, $\|f(z') - f(z^{(n)})\| \leq 2$, and as $\|f(z) - f(z^{(n)})\| \leq \sum_{j=1}^{n} \|f(z^{(j)} - f(z^{(j-1)})\| = n$, we have indeed $\|f(z) - f(z')\| \leq \max\{-[-\|z - z'\|], 2\}$.

Dividing the inequality $\|f(Nz)\| \leq \|Nz\| + 2$ by N, and letting N tend towards infinity, we see that the second part of the lemma follows.

Let the unit circle S contain the segments $[a, b]$ and $[-b, -a]$.

For simplicity, the unit circle in the image plane is also denoted S. We have $f(S) \subset S$, and, according to Lemma 10, the restriction to S of f is injective.

The following remark will be useful in the following:

Because of the local injectivity of f, the image of a non-empty interval I has cardinality c, in particular contains more than two points, which may permit us to use Lemma 3 when considering mappings of intervals into intersections of unit circles.

We have $[a,b] = S \cap S(a+b)$, and so

$$f([a,b]) \subset S \cap S(f(a+b)). \tag{9}$$

Thus an S-edge $[a,b]$ is mapped into either an S-edge ($[c,d]$, say) or the union of two S-edges ($[c,d]$ and $[-d,-c]$). This is still a statement about the local properties of the mapping f. To go beyond this we can use the long form of (9), which can be written

$$f([z_0+a, z_0+b]) \subset S(f(z_0)) \cap S(f(a+b+z_0)). \tag{9'}$$

We are particularly interested in putting $z_0 = t(b-a)$, with real t, which would enable us to say something about the mapping by f of the line through O and $b-a$. If $|t|$ is small, the intervals $[a,b]$ and $[z_0+a, z_0+b]$ overlap, and so do their images. Then at least one edge of $S(f(z_0))$ (namely either $[f(z_0)+c, f(z_0)+d]$ or $[f(z_0)-d, f(z_0)-c]$) overlaps with $[c,d]$, which implies that $f(z_0)$ belongs to either the line L through O parallel to $[c,d]$ or to one of the two lines parallel to L in distance 2 from this line. But this is in fact true for any z_0 of the form $t(b-a)$ with real t. We first see by induction that such a point must have an image belonging to a line parallel to L in a distance from L which is an even integer. Next we consider the particular case $z_0 = e$, where $e = (b-a)/(\|b-a\|)$. Here $e \in S$, and so $f(e)$ must belong to L. The point $f(2e)$ has distance 1 from $f(e)$, is different from O, and belongs to L. But then we must have $f(2e) = 2f(e)$. By induction we see that all points $f(ne)$ with n integral belong to L, and that we have

$$f(ne) = nf(e) \text{ for } n \in \mathbb{Z}. \tag{10_e}$$

Any point $t(b-a)$ with t real has distance less than 1 from some point ne with n integral. According to Lemma 11 its image has distance at most 2 from the point $f(ne)$ and so from the line L, which is what we wanted to prove.

In the following we shall meet vectors w satisfying a condition (10_w), which is just (10_e) with e replaced by w. Let us consider such a vector w. First we shall show that if w can be shown to satisfy

$$f(-w) = -f(w), \tag{11_w}$$

then (10_w) follows.

Actually, the long form of (11_w) can be written

$$f(z_0+w) + f(z_0-w) = 2f(z_0). \tag{$11'_w$}$$

Put $z_0 = (n \pm 1)w$ here to prove (10_w) for $|n| \geq 2$ by induction.

Next, we consider the long form of (10_w), which is

$$f(z_0+nw) - f(z_0) = n(f(z_0+w) - f(z_0)) \text{ for } n \in \mathbb{Z}. \tag{$10'_w$}$$

Replace here z_0 with z_0' and take the difference between the new equation and $(10'_w)$:

(12_w)

$$f(z_0'+nw) - f(z_0+nw) = f(z_0') - f(z_0) + n((f(z_0'+w) - f(z_0')) - (f(z_0+w) - f(z_0))).$$

According to Lemma 11 the norm of the lhs of (12_w) is at most $\max\{\|z_0 - z_0'\| + 1, 2\}$, which is independent of n. Thus, taking norms in (12_w)) and dividing by n gives in the limit $n \to \infty$ that

$$(13_w) \qquad f(z_0'+w) - f(z_0') = f(z_0+w) - f(z_0),$$

i.e. $f(z_0+w) - f(z_0)$ is independent of z_0 and so equals $f(w)$. More generally,

$$(14_w) \qquad f(z_0+nw) = f(z_0) + nf(w) \text{ for } n \in \mathbb{Z}.$$

We now consider a special point z_1, characterized by the equations

$$(15) \qquad \|z_1\| = \|z_1 + e\| = 1.$$

Actually any point in the intersection $M = S \cap S(-e)$ can be taken for z_1. However, only in the case where $\|b-a\| > 1$ does M consist of more than two points. If z_1 satisfies (15) then also $z_1' = -z_1 - e$ does. Using (14_e) with $n = 1$, the image of z_1 satisfies

$$(16) \qquad \|f(z_1)\| = \|f(z_1) + f(e)\| = 1.$$

Thus $f(M)$ belongs to a set $N = S \cap S(-f(e))$. In particular, because of the local injectivity of f, we must have $\|d-c\| > 1$, if $\|b-a\| > 1$.

If $\|d-c\| \le 1$ (implying $\|b-a\| \le 1$), and if $f(z_1)$ is a solution of (16), $-f(z_1) - f(e)$ is the other one. But then injectivity implies that

$$(17) \qquad f(-z_1-e) = -f(z_1) - f(e) \text{ i.e. } f(-z_1) = -f(z_1),$$

and so (14_{z_1}) is satisfied.

A similar analysis can be carried through for all S-edges. One of the implications of the results above is that the direction of the pair of S-edges into which the edge $[a,b]$ is mapped, is determined by $\pm f(e)$. Taken together with the injectivity of the restriction of f to S, this means that the edge $[-b,-a]$ is mapped into the same pair of edges $\pm[c,d]$ as $[a,b]$, while any other edge-pair is mapped into a different edge-pair. Actually (see (14_e)), any line parallel to $[a,b]$ is mapped into a line parallel to $[c,d]$. We can say that f induces an injection of the set of pairs of S-edges into the set of pairs of S-edges. We also saw that the set of pairs of S-edges longer than 1 were mapped into the similar set of edge-pairs.

We shall say that a point z on S is of type 1 if it has the properties of points z_1 and $z_1 + e$ above, i.e. if there is a point $z' \in S$ such that $\|z - z'\| = 1$, with $z - z'$ parallel to an S-edge which is mapped into a pair of S-edges of length equal to or less than 1. The point z will then satisfy the equation (14_z).

A point $z \in S$ will be said to be of type 2 if it has the properties of points z_1 and $z_1 + e$ above, but if the corresponding S-edge has length greater than 1. Here both z an z' belong to the S-edge.

The remaining points on $z \in S$ will be said to be of type 3. Let $z' \in S$ be a point with distance 1 from such a point z. Put $z'' = z - z'$. Then $\|z''\| = \|z'' - z\| = 1$, and $z'' \neq z'$. Furthermore, $\|-z'\| = \|z'' - (-z')\| = 1$, and we see that the hexagon with vertices z', z, z'', $-z'$, $-z$, and $-z''$ is inscribed in S and has all sidelengths equal to 1. Now consider the images of these points. We have $\|f(z)\| = \|f(z')\| = \|f(z') - f(z)\| = 1$. Since we have required that the image plane should have the same metric (in fact the same S) as the original plane, the injection of the (finite!) set of S-edges of length greater than 1 into the similar set of S-edges is in fact a bijection. Thus, if, for instance, the vector $f(z')$ were parallel to a long S-edge $[c, d]$, it would be equal to the image of one of the two unit vectors $\pm e$ parallel to the corresponding long S-edge $[a, b]$. Because of the injectivity of f on S, this would give $z' = \pm e$, and so $z - z''$ should be parallel to a long S-edge, contrary to the definition of type 3 points. Thus the image of the mentioned regular inscribed hexagon is again a regular inscribed hexagon, and we have $f(-z) = -f(z)$, so that (14_z) is valid also for points z of type 3.

Since all S-edges have length less than 2 (otherwise S would be a parallelogram), there is, on each S-edge $[a, b]$ of length greater than 1, an interval I of non-zero length containing the midpoint $(a + b)/2$, such that each point in I is of type 1 or 3. We shall use this later.

We have

Theorem 1. *We can find a linear transformation ϕ such that $\phi(f(z)) = z$ for all points $z \in S$.*

Proof.

The set of points whose type is 1 or 3, has cardinality c. Choose two of them (w_1 and w_2) as basis over the reals. Let z be an arbitrary point of type 1 or 3. We then have an expansion

$$z = a_1 w_1 + a_2 w_2 \tag{18}$$

with real numbers a_1 and a_2.

If these numbers are rational, (18) can be rewritten as

$$dz = n_1 w_1 + n_2 w_2 \tag{18'}$$

with integers d, n_1 and n_2. Using (14_w) with $w = z, w_1$, or w_2, we find that

$$df(z) = n_1 f(w_1) + n_2 f(w_2), \tag{18''}$$

so that in this case the linear relation (18) is inherited by the images.

But this is, as we shall see, true in general. Starting from (18) we use a well known argument which runs as follows:

We define the function g from $\mathbb{N}$ into $[0, 1]^2$ by

$$g(q) = (qa_1 - [qa_1], qa_2 - [qa_2]). \tag{19}$$

Dividing $[0,1]^2$ into N^2 subsquares of the form $[(p-1)/N, p/N] \times [(k-1)/N, k/N]$ we see that we can choose a subsquare such that it contains two points $g(q_1)$ and $g(q_2)$ with $1 \le q_1 < q_2 \le N^2+1$. Thus

$$\|(q_1-q_2)z - ([q_1a_1]-[q_2a_1])w_1 - ([q_1a_2]-[q_2a_2])w_2\| \le (1/N)(\|w_1\|+\|w_2\|).$$

This means that we can find a sequence of triples (m_N, n_N, p_N) of integers, such that the sequence $t_N = m_N w_1 + n_N w_2 + p_N z$ tends towards zero. The same is, according to the second part of Lemma 11, true for the sequence of numbers $f(t_N)$, and even faster for the sequence $f(t_N)/p_N$, which, in the limit $N \to \infty$, proves our point.

Thus if the linear mapping ϕ is chosen such that

$$\phi(f(w)) = w \tag{20}$$

for the two basis points, this equation is valid also for the rest of the points.

Consider now a point z of type 2. Let it belong to an interval $[a,b]$ of length greater than 1, and let $I = (b-e, a+e)$ be the interval, contained in $[a,b]$, of points of type 1 or 3. Here $e = (b-a)/\|b-a\|$. The length of I is $2-\|b-a\|$.

Let $z \in [a,b] \setminus I$.

Define the positive integer

$$n = [\|z-(a+b)/2\|/(2-\|b-a\|)] + 1.$$

It is then possible to find two points d_1 and d_2 in I such that

$$\|d_2-d_1\| = \|z-(a+b)/2\|/n,$$

and such that

$$z = (a+b)/2 + n(d_2-d_1). \tag{21}$$

Using the relevant equations (14_w) and (21) we see that (11_z) and thus (14_z) is true. We can then repeat the argument above for points of type 2. We conclude that (20) is satisfied for all $z \in S$.

Finally we have

Theorem 2. *The Beckman–Quarles Theorem is valid for any normed plane, except when the set of points with norm equal to one is a parallelogram.*

Proof. Any point in the plane is a sum of points belonging to S (see, for instance, the proof of Lemma 11). Thus, with the notation of Theorem 1, the mapping $\phi \circ f$ is simply the identity, q.e.d.

Concluding remarks

When I, during a visit at Université de Montréal in 1974, discussed the construction of Figure 1 with Hwang, he said that he thought he had seen it before. I think that he was right, and that most of the lemmata proved in the present paper are probably what is commonly called "folklore". Nevertheless, I think it is a good thing to have proper proofs published (I have later seen that many of these simple truths have been published by Chilakamarri (see [3])). However my main theorem is, as far as I know, not "well known" and not earlier published.

This paper was slightly modified on 16 July 2014.

References

1. F. S. Beckman and D. A. Quarles, Jr., *On isometries of Euclidean spaces*, Proc. Amer. Math. Soc. **4** (1953), 810–815.
2. Walter Rudin, *Functional analysis*, TMH Edition, New Delhi, 1974.
3. K. B. Chilakamarri, *Unit distance graphs in Minkowski metric spaces*, Geometriae Dedicata **37** (1991), 345–356.

Fjordparken 23A, DK–4000 Roskilde, Denmark
E-mail address: gkkristi@post6.tele.dk