Isometries and Linearity*

Rajendra Bhatia

Let f be a transformation on the Euclidean space such that $f(0) = 0$, and f preserves distances. Then f is linear, and this makes it easier to analyze f. The Mazur–Ulam theorem generalizes this to maps between real normed linear spaces. We discuss this theorem and its proofs.

The author is Professor of Mathematics at Ashoka University and Professor Emeritus at the Indian Statistical Institute.

1. Introduction

Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two real normed linear spaces; i.e., X and Y are vector spaces over $\mathbb{R}$ with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. A map $f : X \to Y$ is called an *isometry* if for all $x, y \in X$ we have

$$\|f(x) - f(y)\|_Y = \|x - y\|_X. \tag{1}$$

It is an immediate consequence of this definition that an isometry is injective and continuous. A map $f : X \to Y$ is said to be *linear* if

$$\begin{aligned} f(x+y) &= f(x) + f(y), \quad x, y \in X, & (2a)\\ \text{and} \quad f(\alpha x) &= \alpha f(x), \quad x \in X, \alpha \in \mathbb{R}. & (2b) \end{aligned} \tag{2}$$

The property (1) defining an isometry does not carry any whiff of linearity. However, it turns out that, subject to two small caveats, every isometry is linear.

While a linear map is obliged to carry the null vector 0 to 0, an isometry is not. Every translation $T_b(x) = x + b$, where b is any vector in X is an isometry on X and $T_b(0) = b$. The remedy for this is simple. If f is an isometry, then so is the map defined by $g(x) = f(x) - f(0)$, and $g(0) = 0$. So we may restrict ourselves to isometries between real normed linear spaces that carry 0 to 0,

Keywords
Isometries, linearity, Mazur–Ulam theorem, inner product spaces, additive functions.

*Vol.27, No.9, DOI: https://doi.org/10.1007/s12045-022-1448-5

and ask whether such maps are linear. The answer turns out to be yes for a large class of norms (to be made precise below) that includes norms generated by inner products. The extra condition that takes care of all normed linear spaces is surjectivity. This is the assertion of the famous Mazur–Ulam theorem proved in 1932 by S. Mazur and S. Ulam, Polish mathematicians belonging to the fabled school of S. Banach, one of the founders of Functional Analysis.

The Mazur–Ulam Theorem: *Let X, Y be real normed linear spaces and let $f : X \to Y$ be a surjective isometry such that $f(0) = 0$. Then f is linear.*

Let X, Y be real normed linear spaces and let $f : X \to Y$ be a surjective isometry such that $f(0) = 0$. Then f is linear.

Before proceeding further, we remark that the theorem is not valid for complex linear spaces. The complex conjugation map $f(z) = \bar{z}$ is bijective, isometric and $f(0) = 0$. But f is not a linear map of $\mathbb{C}$ onto itself.

In this article, we provide first a proof of the Mazur–Ulam theorem in the special case of inner product spaces, and then another proof that works for these spaces and many more. The assumption that f is surjective is not required for these special cases. We then give examples to show that the assumption is necessary in the general case and provide a full proof of the theorem.

A word of notation: we will use the symbol $||\cdot||$ for any norm on the space X or Y (dropping the subscripts as in $||\cdot||_X$ and $||\cdot||_Y$). Subscripts will be used for special norms, like $||\cdot||_p$, that will be defined as they occur in the discussion.

2. Inner Product Spaces

To prove the Mazur–Ulam theorem, we have to extract linearity out of a statement about norms. For one kind of norm associated with an inner product, linearity enters through this latter object.

Let X be a real vector space. An *inner product* on X associates to each pair of vectors x, y in X a real number $\langle x, y\rangle$, called the inner product of x and y, that obeys the following rules:

(i) $\langle x, y\rangle = \langle y, x\rangle$,

(ii) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ and $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ for all $\alpha \in \mathbb{R}$,

(iii) $\langle x, x\rangle \geq 0$ for all x, and $\langle x, x\rangle = 0$ if and only if $x = 0$.

Note that using (i), we see that the condition (ii) implies

$$\begin{aligned} \langle x, y + z\rangle &= \langle x, y\rangle + \langle x, z\rangle, \\ \text{and } \langle x, \alpha y\rangle &= \alpha\langle x, y\rangle. \end{aligned}$$

Thus an inner product $\langle x, y\rangle$ is linear in each of the variables x and y. An inner product gives rise to a norm on X defined by

$$||x|| = \langle x, x\rangle^{1/2}.$$

The standard example of an inner product space is $\mathbb{R}^n$ with the usual Euclidean inner product

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i,$$

and the associated norm

$$||x||_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}. \tag{3}$$

In the rest of this section, the symbol $||\cdot||$ stands for a norm on X or Y arising from an inner product $\langle\cdot,\cdot\rangle$.

Let $f : X \to Y$ be an isometry such that $f(0) = 0$. Then $||f(x)|| = ||x||$ for all x. Using the identity

$$||x - y||^2 = ||x||^2 + ||y||^2 - 2\langle x, y\rangle,$$

we see that

$$\langle f(x), f(y)\rangle = \langle x, y\rangle;$$

i.e., f preserves inner products. Now let x, y, z be any three elements of X. Then

$$\begin{aligned} \langle f(x + y), f(z)\rangle &= \langle x + y, z\rangle \\ &= \langle x, z\rangle + \langle y, z\rangle \\ &= \langle f(x), f(z)\rangle + \langle f(y), f(z)\rangle \\ &= \langle f(x) + f(y), f(z)\rangle \end{aligned}$$

From this, it follows that

$$\langle f(x+y) - f(x) - f(y), f(z)\rangle = 0 \tag{4}$$

This is true for all z. Choose, successively, $z = x + y$, x and y in (4), and then combine the three equations obtained to get

$$\langle f(x+y) - f(x) - f(y), f(x+y) - f(x) - f(y)\rangle = 0;$$

i.e.,

$$\|f(x+y) - f(x) - f(y)\| = 0.$$

This shows that $f(x+y) = f(x)+f(y)$. A similar argument shows that $f(\alpha x) = \alpha f(x)$ for all $\alpha \in \mathbb{R}$.

We have shown that every isometry f between inner product spaces, with $f(0) = 0$, is linear.

3. Continuous Additive Functions are Linear

The two conditions that define linearity are independent; i.e., there exist functions f that satisfy one but not the other condition.

The two conditions that define linearity are independent; i.e., there exist functions f that satisfy one but not the other condition. Functions that satisfy the requirement ($2a$) are called *additive*. We now show that a continuous additive function is linear. In other words, a continuous function satisfying ($2a$) also satisfies ($2b$).

Putting $x = y = 0$ in ($2a$) we see that $f(0) = 2f(0)$, and hence $f(0) = 0$. Then choosing $y = -x$, we get $f(-x) = -f(x)$ for all x. From ($2a$) it also follows that $f(mx) = mf(x)$ for all natural numbers m, and hence for all integers m. Next, observe that $f(x) = f\left(m\frac{x}{m}\right) = mf\left(\frac{x}{m}\right)$. So $f\left(\frac{x}{m}\right) = \frac{1}{m}f(x)$ for every nonzero integer m. From these observations we can conclude that $f(\alpha x) = \alpha f(x)$ for every rational number α. Since f is continuous, this is true for every real α too. We have shown that every continuous additive function is linear.

The next observation is crucial for the proof of the Mazur-Ulam Theorem. Suppose $f : X \to Y$ is a map such that $f(0) = 0$ and

$$f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}, \tag{5}$$

for all $x, y \in X$. Choosing $y = 0$, this gives $f(2x) = 2f(x)$ for all x. Plugging this back into (5) we get $f(x+y) = \frac{1}{2}[f(2x)+f(2y)] = f(x) + f(y)$. Thus f is additive. Hence, if it is also continuous, then it is linear.

4. Algebraic and Metric Midpoints

Let x, y be two points in a vector space X. The *algebraic midpoint* of x, y is the vector $\frac{x+y}{2}$. In the usual Euclidean space $\mathbb{R}^n$ this is also the unique point z which is at half the distance $\frac{1}{2}\|x-y\|_2$ from both x and y. More generally, given any norm on X, let us say z is a *metric midpoint* of x and y, if

$$\|z-x\| = \|z-y\| = \frac{1}{2}\|x-y\|. \tag{6}$$

Let x, y be two points in a vector space X. The *algebraic midpoint* of x, y is the vector $\frac{x+y}{2}$.

It is easy to see that an algebraic midpoint is a metric midpoint. If $f : X \to Y$ is an isometry, then it would carry a metric midpoint of x, y in X to a metric midpoint of$f(x)$ and $f(y)$ in Y.

In Section 3 we saw that an isometry $f : X \to Y$ with $f(0) = 0$ would be linear if it satisfies (5), i.e., if it carries algebraic midpoints in X to algebraic midpoints in Y. So, if every metric midpoint in the space Y were an algebraic midpoint, we could conclude that f is linear. And here, there is a twist in the tale: there are normed linear spaces where metric midpoints are not unique.

A very simple example is provided by the ∞-*norm* on $\mathbb{R}^n$ defined as

$$\|x\|_\infty = \max_{1\le i\le n} |x_i|. \tag{7}$$

Let $n = 2$ and consider the vectors $x = (-1, 0)$, $y = (1, 0)$. Then $\|x-y\|_\infty = 2$. The algebraic midpoint of x and y is $(0, 0)$. This is at distance 1 from both x and y, but so is every vector $z = (0, t)$ where $|t| < 1$. Thus every point in the line segment joining $(0, -1)$ and $(0, 1)$ is a midpoint of x and y.

Another example comes from the 1-*norm* on $\mathbb{R}^n$ defined as

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|. \tag{8}$$

Let $n = 2$ and choose $x = (1, 0)$ and $y = (0, 1)$. Then $||x - y||_1 = 2$. The algebraic midpoint of x and y is $\left(\frac{1}{2}, \frac{1}{2}\right)$. Every point in the line segment joining $(0, 0)$ and $(1, 1)$ is a metric midpoint of x and y.

5. Strictly Convex Norms

A norm on X is called *strictly convex* if whenever x and y are distinct points in X with $||x|| = ||y|| = 1$ then $\left\|\frac{x+y}{2}\right\| <1$.

These are norms for which there are no metric midpoints other than the algebraic midpoint. More precisely, a norm on X is called *strictly convex* if whenever x and y are distinct points in X with $||x|| = ||y|| = 1$ then $\left\|\frac{x+y}{2}\right\| <1$. A norm arising from an inner product obeys the parallelogram law

$$||x + y||^2 + ||x - y||^2 = 2(||x||^2 + ||y||^2),$$

and is, therefore, strictly convex.

Let $|| \cdot ||$ be a strictly convex norm on X and x, y, z any three points in X satisfying the relations in (6). Then

$$\left\|\frac{(z - x) + (y - z)}{2}\right\| = \frac{1}{2}||y - x|| = ||z - x|| = ||z - y||$$

The strict convexity now implies that $z-x = y-z$. In other words, $z = \frac{x+y}{2}$. Thus a metric midpoint is also an algebraic midpoint.

For $1 \leq p < \infty$, the p-norm on $\mathbb{R}^n$ is defined as

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

The special cases $p = 1, 2$, have been seen in (3) and (8). The norm $|| \cdot ||_\infty$ in (7) is the limiting case

$$||x||_\infty = \lim_{p\to\infty} ||x||_p.$$

It can be seen that all norms in this family except the ones corresponding to $p = 1, \infty$, are strictly convex. (For $1 < p < \infty$, there is equality in the Minkowski inequality $||x + y||_p \leq ||x||_p + ||y||_p$ if and only if $x = ty$ for some positive number t.)

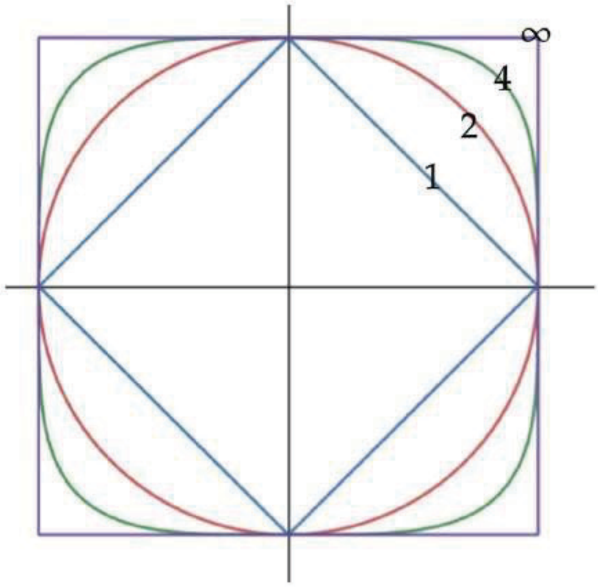

Figure 1. Unit balls corresponding to p-norms for $p = 1, 2, 4$ and ∞.

A discussion of strict convexity will take us away from our main theme, but a little digression might be in order here. Let $n = 2$ and draw the unit balls for the $\|\cdot\|_p$ norms for some different values of p. The ones for $p = 1, 2, 4$ and ∞ are shown here. All are convex sets. The boundaries of the balls for $p = 1$ and ∞ contain line segments, and for other values of p, they don't. A compact convex set in $\mathbb{R}^n$ is called strictly convex if every point on its boundary is an extreme point. It can be seen that a norm on $\mathbb{R}^n$ is strictly convex if and only if its unit ball is a strictly convex set.

Returning to isometries, we have shown that if $f : X \to Y$ is an isometry with $f(0) = 0$, and the norm on Y is strictly convex, then f is linear. It is not necessary to assume here that f is surjective.

The example we present now illustrates that without the assumption of surjectivity on f or of strict convexity on Y, an isometry f may fail to be linear.

Let $X = \mathbb{R}$ and $Y = \mathbb{R}^2$ with the ∞-norm. Let $f : X \to Y$ be the map $f(t) = (t, |t|)$. Then f is isometric, $f(0) = 0$, but f is not linear. With the same X and Y the map $f(t) = (t, \sin t)$ provides another example.

6. The General Case

We now give a proof of the Mazur–Ulam theorem in the general case. If we show that f preserves algebraic midpoints (i.e., (5) is valid), then it would follow that f is linear. The original proof of Mazur and Ulam [6] proceeded by obtaining a metric characterization of the algebraic midpoint from which (5) followed. Subsequent expositions (see, e.g., the classic [3], or the more recent text [5]) followed essentially the same idea. In 2012 B. Nica [7], expanding on an earlier idea of J. Väisälä [10], published a beautifully simplified proof. That is the proof we present.

We will need the notion of a reflection through a point. Let z be a given point of X. The *reflection through* z is the map ρ_z defined on X by $\rho_z(x) = 2z - x$. (Note that x and $\rho_z(x)$ are equidistant from z but located on "opposite sides" of it; hence the name reflection.) It is clear that ρ_z is a surjective isometry on X.

We wish to show that f satisfies (5) for all x and y. Suppose there are points x, y for which (5) is not true. Let

$$\Delta_f(x, y) = \left\| f\left(\frac{x+y}{2}\right) - \frac{f(x) + f(y)}{2} \right\|.$$

This is the measure of f's "defect from linearity". We have

$$\begin{aligned}
\Delta_f(x, y) &= \left\| \frac{1}{2}\left[f\left(\frac{x+y}{2}\right) - f(x) \right] + \frac{1}{2}\left[f\left(\frac{x+y}{2}\right) - f(y) \right] \right\| \\
&\leq \frac{1}{2}\left\| f\left(\frac{x+y}{2}\right) - f(x) \right\| + \frac{1}{2}\left\| f\left(\frac{x+y}{2}\right) - f(x) \right\| \\
&= \frac{1}{2}\left\| \frac{x+y}{2} - x \right\| + \frac{1}{2}\left\| \frac{x+y}{2} - y \right\| \\
&= \frac{1}{2}\|x - y\|.
\end{aligned}$$

Thus $\Delta_f(x, y)$ is bounded above by $\frac{1}{2}\|x - y\|$ for every isometry f.

Given f, x, and y as above, we will create another surjective isometry g on X for which $\Delta_g(x, y) = 2\Delta_f(x, y)$. For this let $z = \frac{f(x)+f(y)}{2}$ and let $\rho = \rho_z$, the reflection through the point z in the space Y. Thus $\rho(u) = f(x) + f(y) - u$ for every u in Y. If f is a surjective isometry, then it is invertible. Define g as the composite map $g = f^{-1}\rho f$. Being a composition of surjective isometries, g is a surjective isometry. Note that

$$g(x) = f^{-1}\rho(f(x)) = f^{-1}(f(x) + f(y) - f(x)) = f^{-1}(f(y)) = y$$

Similarly $g(y) = x$. So, we have

$$\begin{aligned} \Delta_g(x, y) &= \left\| g\left(\frac{x+y}{2}\right) - \frac{g(x) + g(y)}{2} \right\| \\ &= \left\| f^{-1}\left(f(x) + f(y) - f\left(\frac{x+y}{2}\right)\right) - \frac{x+y}{2} \right\|. \end{aligned}$$

Since f is an isometry, this gives

$$\begin{aligned} \Delta_g(x, y) &= \left\| f(x) + f(y) - f\left(\frac{x+y}{2}\right) - f\left(\frac{x+y}{2}\right) \right\| \\ &= 2\left\| \frac{f(x) + f(y)}{2} - f\left(\frac{x+y}{2}\right) \right\| \\ &= 2\Delta_f(x, y). \end{aligned}$$

Iterating this process, we can successively create bijective isometries on X whose defects from linearity are $2\Delta_f(x, y), 4\Delta_f(x, y), \ldots, 2^n\Delta_f(x, y)$. But we have also seen that for each isometry, this defect has to be bounded by $\frac{1}{2}\|x - y\|$. This is possible only if $\Delta_f(x, y) = 0$.

We have shown that f satisfies (5) and hence is linear.

7. Some Extras

The Mazur–Ulam theorem says that a surjective isometry $f : X \to Y$ between real normed linear spaces has the form $f(x) = Ax + b$, where $A : X \to Y$ is a linear operator, and b is an element of Y. Particularly interesting is the case $X = Y = \mathbb{R}^n$ with the usual Euclidean norm. Then A must be an orthogonal matrix; i.e. $A^T A = I$. In the special situation $n = 2$, A must be one of the two kinds

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \text{ or } A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

The Mazur–Ulam theorem says that a surjective isometry $f : X \to Y$ between real normed linear spaces has the form $f(x) = Ax + b$, where $A : X \to Y$ is a linear operator, and b is an element of Y.

The first one represents the rotation by an angle θ in the anticlockwise direction. The second one is a reflection through the line at angle $\theta/2$ with the x-axis.

We saw that an additive function f on $\mathbb{R}$ must be of the form $f(x) = \alpha x$ for some $\alpha \in \mathbb{R}$ provided f is continuous. The same conclusion can be obtained with conditions weaker than continuity. For example, it is enough to assume that f is measurable.

In [4] Hyers and Ulam introduced the notion of an *approximate isometry* or an *ε-isometry*. This is a map $f : X \to Y$ such that

$$|\|f(x) - f(y)\| - \|x - y\|| < \varepsilon \text{ for all } x, y.$$

The motivation for this is that often distances are known only approximately. The question Hyers and Ulam raised was whether such a map is close to being an isometry. More precisely, if f is a surjective ε-isometry between real normed linear spaces such that $f(0) = 0$, then does there exist a surjective linear isometry $g : X \to Y$ such that

$$\|f(x) - g(x)\| \leq k\varepsilon \text{ for all } x \in X,$$

where k is a constant independent of f? An expository article describing this problem is [2].

The Mazur-Ulam theorem is a simple illustration of the interplay between metric and algebraic properties of objects. This is a recurring theme in analysis. See, for example, [2] and [8].

Different distances in the same space are used in different contexts.

The notion of a metric midpoint is useful in several situations. Different distances in the same space are used in different contexts. Thus on the simplest space $\mathbb{R}_+ = (0, \infty)$ consisting of positive real numbers, instead of the usual distance $|x - y|$ between x and y, one may consider another distance $d(x, y) = |\log x - \log y|$. The Richter scale in seismology, the decibel scale in acoustics, and the pH values in chemistry all use this distance. Given two points $x, y \in \mathbb{R}_+$ their metric midpoint now would be a point z such that

$$d(z, x) = d(z, y) = \frac{1}{2} d(x, y).$$

A small calculation shows that there is a unique point z satisfying this requirement, and it is $z = \sqrt{xy}$, otherwise known as the *geometric mean*. The reader may calculate the metric midpoints corresponding to a few more distances, such as

$$\begin{aligned} d(x,y) &= \left|\frac{1}{x} - \frac{1}{y}\right|, \\ d(x,y) &= |x^2 - y^2|, \\ \text{and} \quad d(x,y) &= \left|\sqrt{x} - \sqrt{y}\right|. \end{aligned}$$

The Polish school led by Stefan Banach played a central role in the development of functional analysis in the years between the two world wars. Stanislaw Mazur and Stanislaw Ulam were prominent members of this school. Every student of Functional Analysis learns about the Banach–Mazur theorem and the Gelfand–Mazur theorem. Ulam was one of the most versatile mathematicians who made major contributions to set theory, topology, analysis, group theory, number theory, combinatorics, probability, computations, and to the design of thermonuclear weapons. He has also written a very interesting autobiography titled *The Adventures of a Mathematician*.

The Polish school led by Stefan Banach played a central role in the development of functional analysis in the years between the two world wars.

Suggested Reading

[1] R. Bhatia, *Positive Definite Matrices*, Princeton Univ. Press and Hindustan Book Agency, 2007.

[2] R. Bhatia and P. Semrl, Approximate isometries on Euclidean spaces, *Amer. Math. Monthly*, Vol.104, pp.497–504, 1997.

[3] M. M. Day, *Normed Linear Spaces*, Springer, Berlin, 1958.

[4] D. H. Hyers and S. M. Ulam, On approximate isometries, *Bull. Amer. Math. Soc.*, Vol.51, pp.288–292, 1945.

[5] P. Lax, *Functional Analysis*, Wiley, 2002.

[6] S. Mazur and S. Ulam, Sur les transformations isométriques d'espaces vectoriels normés, *C. R. Acad. Sci. Paris*, Vol.194, pp.946–948, 1932.

[7] B Nica, The Mazur-Ulam theorem, *Expo. Math.*, Vol.30, pp.397–398, 2012.

[8] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge University Press, 2009.

[9] S. Ulam, *Adventures of a Mathematician*, Charles Scribner's Sons, New York, 1976.

[10] J. Väisälä, A proof of the Mazur-Ulam theorem, *Amer. Math. Monthly*, 110, pp.633–635, 2003.

Address for Correspondence
Rajendra Bhatia
Ashoka University, Rai,
Haryana 131 029, India.
Email:
rajenbhatia@gmail.com