university of groningen

faculty of mathematics and natural sciences

Beyond Complex Numbers

Bachelor's Thesis in Mathematics

June 2011

Student : René Herman
Supervisor : Prof. dr. Jaap Top

Contents

Chapter 1

Beyond complex numbers

1.1 Introduction

While at best marginally familiar to most current students of mathematics, the history of *quaternions* makes for an important footnote of early modern mathematics. Perhaps one of the most important in the sense that it can be viewed to have stood as part of the early development of modern abstract algebra, which has since then steered mathematics towards the pure and standalone science of today.

The story begins with the development of complex analysis in the early nineteenth century, when results by among others Augustin Cauchy, Bernhard Riemann and Karl Weierstrass make ever more apparent the inherent elegance of the subject of complex analysis. Among the people involved in this then new and exciting field of mathematics is also Irish physicist, astronomer and mathematician William Rowan Hamilton who, following earlier work by Caspar Wessel, Jean-Robert Argand and Carl Friedrich Gauss, in 1835 completes his *Theory of Couplets* which amounts to the view of complex numbers as ordered pairs of real numbers, or points on a complex plane, so familiar to us today.

Given the success of the planar view little comes more natural to a mathematician than the idea of next extending the notions from the two-dimensional plane to three-dimensional space and Hamilton subsequently sets out to do just that, hoping to construct a *Theory of Triplets* to parallel the success of complex numbers and, perhaps, eventually complex analysis itself.

Natural as it may be though, this turns out to also be naive. In modern terminology the complex numbers form an algebraic structure we call *a real division algebra* and as we shall see, no such three-dimensional structure exists. In fact, Ferdinand Frobenius shows in 1877 that the one-dimensional real numbers $\mathbb{R}$, two-dimensional complex numbers $\mathbb{C}$ and the four-dimensional quaternions $\mathbb{H}$ that Hamilton eventually does construct are (up to isomorphism) the only finite-dimensional associative real division algebras, and in the latter case only at the cost of losing commutativity. In 1898 Adolf Hurwitz then shows that only one more finite-dimensional real division algebra $\mathbb{O}$ of eight-dimensional *octonions*

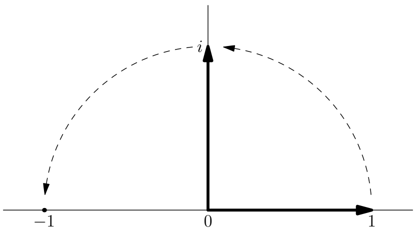

Figure 1.1: The complex plane

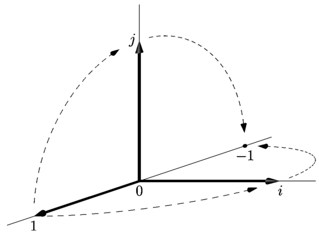

Figure 1.2: A triplet space?

results if we forego even associativity and require only the in complex analysis vital concept of having available an *absolute value* or *modulus*.

Proving these results by Frobenius and Hurwitz will be the substance of this first chapter, while we will in the next chapter see that erecting an analogue of complex analysis for the quaternions and octonions proves to be an in fact only marginally viable undertaking.

1.2 Quaternions

As said, with his *Theory of Couplets* identifying a complex number $x + iy$ with the point (x, y) in $\mathbb{R}^2$ freshly developed, Hamilton sets out to now conversely identify a point (x, y, z) in $\mathbb{R}^3$ with a new type of number $x + iy + jz$, hoping to parallel the success of complex numbers.

As an extension of the complex numbers he implicitly requires that the new triplet space needs to embed the complex plane as its $(1, i)$ plane in the same way that $\mathbb{R}^3$ embeds $\mathbb{R}^2$ which means $1i = i1 = i$ and $i^2 = -1$ same as for the complex numbers. He furthermore requires $1j = j1 = j$ simply per definition of 1 and since, as figures 1.1 and 1.2 demonstrate, multiplication of 1 by i amounts to a ninety degree counterclockwise rotation about the origin in the complex plane in the same way that multiplication of 1 by j amounts to this same rotation in the $(1, j)$ plane, he requires $j^2 = -1$ as well.

Already in trying to decide what to do with the product ij he runs into the fundamental problem of all this though. With the concepts hardly even explicitly available at the time, Hamilton also wants to simply assume distributivity and associativity but if we set $ij = x + iy + jz$ and left-multiply by i we then obtain

$$-j = i(x + iy + jz) = ix - y + ijz = ix - y + (x + iy + jz)z$$

which is to say $xz - y + i(x + yz) + j(z^2 + 1) = 0$. By perpendicularity of 1, i and j therefore $xz - y = x + yz = z^2 + 1 = 0$ which is impossible for $z \in \mathbb{R}$. It follows that ij can not in fact be an element of the triplet space and thereby that the triplet space is not closed under multiplication — something which clearly won't do for an analogue of the very algebraically clean complex numbers.

However, even an innocent formulation such as *a space being closed under multiplication* hints at the sort of modern algebraic footing which was at the time still unavailable and Hamilton in fact spends quite some time stuck at this point.

To a modern reader it is readily apparent that ij is simply a fourth linearly independent element but not only hadn't the concept of linear spaces nor linear independence been developed yet, the entire notion of a fourth dimension was a still decidedly esoteric one. It is therefore not until 1843 that Hamilton realizes that he needs to add a fourth dimension for, as he puts it in a letter to his friend John Graves, *the purpose of calculating with triplets*[1].

However, as mathematically significant the acceptance of a fourth dimension may itself have been at the time, acceptance of a specific consequence seems in retrospect more significant still. Having developed his Theory of Couplets, Hamilton is very much aware that the norm on $\mathbb{R}^2$ functioning as a multiplicative absolute value on $\mathbb{C}$ is one of the most important properties of the complex numbers, seeing as how it provides the basic ingredient of analytic concepts such as *limit* and *derivative*.

Therefore, in the same way that for a complex number $z = x + iy$ the definition

$$|z| = \sqrt{x^2 + y^2}$$

together with the computationally natural product

$$(x_1 + iy_1)(x_2 + iy_2) = x_1x_2 - y_1y_2 + i(x_1y_2 + y_1x_2)$$

means $|z_1z_2| = |z_1|\,|z_2|$ for all complex numbers z_1 and z_2, Hamilton requires that for a quaternion $q = t + ix + jy + kz$ the definition

$$|q| = \sqrt{t^2 + x^2 + y^2 + z^2}$$

together with the similarly natural product needs to mean $|q_1q_2| = |q_1|\,|q_2|$ for all quaternions q_1 and q_2. He had already noticed before that for a triplet (now a special type of quaternion) $q = t+ix+jy$ the computationally natural product means

$$q^2 = (t + ix + jy)(t + ix + jy) = t^2 - x^2 - y^2 + i(2tx) + j(2ty) + (ij + ji)xy$$

whereas by the above definition of absolute value

$$|q|^2 = t^2 + x^2 + y^2 = \sqrt{(t^2 - x^2 - y^2)^2 + (2tx)^2 + (2ty)^2}$$

so that the requirement $\left|q^2\right| = |q|^2$ very strongly suggests $ij + ji = 0$. Moreover, now that ij lies in an actual fourth direction he at this point definitively needs $ij \neq 0$ and from $ji = -ij \neq 0$ needs to thereby accept noncommutativity of his new quaternions.

At the time, this was a still largely unheard of thing to do and, we feel, the perhaps biggest contribution Hamilton made to mathematics consists of not simply discarding quaternions then and there. As we shall see later, it takes Graves only two months from hearing of them to come up with the *octonions* that forego even associativity, a word which may not even have existed up to that point in time, and which shows the quaternions to have been an important early inroad into modern abstract algebra.

[1] On Quaternions: Letter to John T. Graves, Esq. [4]

	i	j	k
i	-1	k	$-j$
j	$-k$	-1	i
k	j	$-i$	-1

Table 1.1

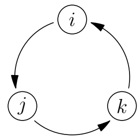

Figure 1.3

Be that as it may, Hamilton now has all the parts he needs. Setting $k = ij$ and thereby $ji = -ij = -k$ he notices

$$k^2 = (ij)(ij) = -(ij)(ji) = -i(jj)i = ii = -1$$

whereby k shows itself to be just like i and j and with $1k = k1 = k$ and

$$kj = (ij)j = i(jj) = -i \qquad jk = j(ij) = (ji)j = -kj = i$$
$$ik = i(ij) = (ii)j = -j \qquad ki = (ij)i = i(ji) = -ik = j$$

he completes the rules of quaternion multiplication, summarised as

$$i^2 = j^2 = k^2 = -1 \qquad ij = k = -ji \qquad jk = i = -kj \qquad ki = j = -ik \tag{1.1}$$

or in their most compact form as $i^2 = j^2 = k^2 = ijk = -1$. Also note the above tabular and mnemonic formats.

Identifying the new space of quaternions $\mathbb{H}$ (as we denote it now in his honour) with $\mathbb{R}^4$, he declares two quaternions

$$q_1 = t_1 + ix_1 + jy_1 + kz_1 \qquad \text{and} \qquad q_2 = t_2 + ix_2 + jy_2 + kz_2$$

to be equal if and only if $t_1 = t_2$, $x_1 = x_2$, $y_1 = y_2$ and $z_1 = z_2$ and endows them with the regular componentwise addition

$$(t_1 + ix_1 + jy_1 + kz_1) + (t_2 + ix_2 + jy_2 + kz_2) = \\ (t_1 + t_2) + i(x_1 + x_2) + j(y_1 + y_2) + k(z_1 + z_2) \tag{1.2}$$

and his desired computationally natural product

$$(t_1 + ix_1 + jy_1 + kz_1)(t_2 + ix_2 + jy_2 + kz_2) = \\ (t_1t_2 - x_1x_2 - y_1y_2 - z_1z_2) \ + \ i(t_1x_2 + x_1t_2 + y_1z_2 - z_1y_2) \ + \\ j(t_1y_2 - x_1z_2 + y_1t_2 + z_1x_2) \ + \ k(t_1z_2 + x_1y_2 - y_1x_2 + z_1t_2) \tag{1.3}$$

Then, after carefully verifying that with the desired absolute value

$$|t + ix + jy + kz| = \sqrt{t^2 + x^2 + y^2 + z^2} \tag{1.4}$$

he now indeed has $|q_1q_2| = |q_1|\,|q_2|$ for all quaternions q_1 and q_2, Hamilton finally declares victory over years of contemplating the subject.

We note associativity of the product (1.3), verification of which is a straightforward if rather tedious process. We will also show this rigorously later when we reconstruct the quaternions in a more structured way. For now, we will only quickly list a few properties so as to establish basic familiarity.

Firstly note that

$$(t_1 + ix_1 + j0 + k0) + (t_2 + ix_2 + j0 + k0) = (t_1 + t_2) + i(x_1 + x_2) + j0 + k0$$

and

$$(t_1 + ix_1 + j0 + k0)(t_2 + ix_2 + j0 + k0) = \\ (t_1t_2 - x_1x_2) + i(t_1x_2 + x_1t_2) + j0 + k0$$

so that $\mathbb{C}$ embeds as naturally into $\mathbb{H}$ as $\mathbb{R}$ in turn embeds naturally into $\mathbb{C}$.

With the absolute value (1.4) no other than the regular Euclidean norm on $\mathbb{R}^4$ it of course shares all the properties of a norm so that we have all in all, same as for $\mathbb{R}$ and $\mathbb{C}$

nonnegativity	$\lvert q\rvert \geq 0$ for all $q \in \mathbb{H}$
nondegeneracy	$\lvert q\rvert = 0$ if and only if $q = 0$
triangle inequality	$\lvert q_1 + q_2\rvert \leq \lvert q_1\rvert + \lvert q_2\rvert$ for all $q_1, q_2 \in \mathbb{H}$
multiplicativity	$\lvert q_1 \cdot q_2\rvert = \lvert q_1\rvert\,\lvert q_2\rvert$ for all $q_1, q_2 \in \mathbb{H}$

We define for the quaternion $q = t + ix + jy + kz \in \mathbb{H}$ the *conjugate* of q to be

$$\bar{q} = t - ix - jy - kz \tag{1.5}$$

Clearly $\bar{\bar{q}} = q$ and by easy direct verification, for all $q_1, q_2 \in \mathbb{H}$

$$\overline{q_1 + q_2} = \overline{q_1} + \overline{q_2} \qquad \text{and} \qquad \overline{q_1 \cdot q_2} = \overline{q_2} \cdot \overline{q_1} \tag{1.6}$$

Moreover, same as for the complex numbers

$$q\bar{q} = \bar{q}q = t^2 + x^2 + y^2 + z^2 = |q|^2 \tag{1.7}$$

so that together with nondegeneracy of the absolute value any $q \neq 0$ admits a unique inverse

$$q^{-1} = |q|^{-2}\ \bar{q} \tag{1.8}$$

satisfying $qq^{-1} = 1 = q^{-1}q$ and enabling the concept of division in $\mathbb{H}$. We do of course need separate left and right quotients $q_2^{-1}q_1$ and $q_1q_2^{-1}$ of two general quaternions q_1 and q_2 due to general noncommutativity of $\mathbb{H}$. By its nondegenerate and multiplicative absolute value, $\mathbb{H}$ is clearly without zero divisors.

As is easily directly verified we have for $q = t + ix + jy + kz \in \mathbb{H}$

$$\begin{aligned} t &= -\frac{1}{4}(-q + iqi + jqj + kqk) & x &= -\frac{i}{4}\,(q - iqi + jqj + kqk) \\ y &= -\frac{j}{4}\,(q + iqi - jqj + kqk) & z &= -\frac{k}{4}\,(q + iqi + jqj - kqk) \end{aligned} \tag{1.9}$$

With $q \in \text{Cen}(\mathbb{H}) = \{q \in \mathbb{H} \,|\, qr = rq \text{ for all } r \in \mathbb{H}\}$ therefore

$$t = -\frac{1}{4}(-q+iqi+jqj+kqk) = -\frac{1}{4}(-q+iiq+jjq+kkq) = \frac{1}{4}(q+q+q+q) = q$$

so that $\text{Cen}(\mathbb{H}) \subseteq \mathbb{R}$. By construction or easy verification conversely $qr = rq$ for all $r \in \mathbb{R}$ whereby we conclude that $\text{Cen}(\mathbb{H}) = \mathbb{R}$.

We note at this point that this also means that setting $q = t + ix + jy + kz$ versus $q = t + xi + yj + zk$ amounts to taste only. The former is more common in an analytic context whereas the latter is used more in an algebraic one.

The *real* and *pure* parts of $q = t + ix + jy + kz$ are defined to be

$$\mathrm{Re}(q) = \frac{1}{2}(q + \bar{q}) = t \qquad \text{and} \qquad \mathrm{Pu}(q) = \frac{1}{2}(q - \bar{q}) = ix + jy + kz \tag{1.10}$$

The latter is sometimes also referred to as the *imaginary* part and denoted $\mathrm{Im}(q)$ but is unlike its complex counterpart not real but quaternion-valued; with $\mathbb{C} \subset \mathbb{H}$ we will therefore reserve Im for use in the complex context.

Historically these quantities were denoted $\mathrm{Sc}(q)$ and $\mathrm{Ve}(q)$ for the *scalar* and *vector* part respectively, and we note that both words actually originate in this quaternionic context. In fact, as expressed in his aforementioned statement of introducing the fourth dimension *for the purpose of calculating with triplets* Hamilton at least for now still considers quaternions to be tools for constructing a three-dimensional calculus, and with modern vector calculus still undeveloped at the time many of the early developments in the field trace their origin to this quaternionic past.

For example, with $v_1 = ix_1 + jy_1 + kz_1$ and $v_2 = ix_2 + jy_2 + kz_2$ we have

$$v_1v_2 = -x_1x_2 - y_1y_2 - z_1z_2 + i(y_1z_2 - z_1y_2) + j(z_1x_2 - x_1z_2) + k(x_1y_2 - y_1x_2)$$

and by combining modern notation with quaternionic history therefore

$$\langle v_1, v_2 \rangle = -\,\mathrm{Sc}(v_1v_1) \qquad \text{and} \qquad v_1 \times v_2 = \mathrm{Ve}(v_1v_2)$$

from which we see the alternative names *scalar product* and *vector product* for the inner and cross product take form. We moreover note that the utility of these products in respectively determining the angle between vectors and constructing a third perpendicular vector should be taken to sufficiently explain Hamilton's choice of identifying the *last* three dimensions of $\mathbb{H}$ with $\mathbb{R}^3$ rather than the first three as he set out for originally back when they were still triplets. Our use of the standard names i, j and k for the three basis vectors of $\mathbb{R}^3$ is also still a result of this decision and one of the most visible remnants of the period now that modern vector calculus has fully replaced any quaternionic approach.

Note that whereas -1 has exactly two square roots i and $-i$ in $\mathbb{C}$ it has an infinite number of them in $\mathbb{H}$, since

$$(ix + jy + kz)^2 = -x^2 - y^2 - z^2 = -(x^2 + y^2 + z^2) = -1$$

for any point (x, y, z) of the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$. This has the effect that there are also an infinite number of possible choices for i, j and k, with any first one chosen freely from among all unit vectors $(0, x, y, z) \in \mathbb{R}^4$ and the other two as to that first and each other perpendicular unit vectors among the same. We will see this effect in the upcoming proof of the theorem of Frobenius in the sense that we need to avoid suggesting unicity and instead stress trivial isomorphism with any other choice.

Before getting there though, we will need to formally introduce the concept of *an algebra*, which is to say a linear space together with a bilinear product.

1.3 Algebras

Definition 1.1. A *linear space* or *vector space* V over a field F is a set V together with an element $0_V \in V$ and two operations

$$+\colon V\times V\to V,\ (x,y)\mapsto x+y \qquad \cdot\colon F\times V\to V,\ (a,x)\mapsto ax=a\cdot x$$

called *addition* and *scalar multiplication* such that

(V1)	$\forall x,y\in V$	$x+y=y+x$
(V2)	$\forall x,y,z\in V$	$(x+y)+z=x+(y+z)$
(V3)	$\forall x\in V$	$x+0_V=x$
(V4)	$\forall x\in V$	$x+(-1_F)x=0_V$
(V5)	$\forall a\in F,\ x,y\in V$	$a(x+y)=ax+ay$
(V6)	$\forall a,b\in F,\ x\in V$	$(a+b)x=ax+bx$
(V7)	$\forall a,b\in F,\ x\in V$	$a(bx)=(ab)x$
(V8)	$\forall x\in V$	$1_F\cdot x=x$

With $x,y\in V$ and $a\in F,\ a\neq 0_F$: $-x:=(-1_F)x$, $y-x:=y+(-x)$, $\frac{x}{a}:=a^{-1}x$.

Definition 1.2. An *algebra* V over a field F is a linear space V over F together with a third operation

$$\cdot\colon V\times V\to V,\ (x,y)\mapsto xy=x\cdot y$$

called *multiplication* such that

(A1)	$\forall x,y,z\in V$	$x(y+z)=xy+xz$
(A2)	$\forall x,y,z\in V$	$(x+y)z=xz+yz$
(A3)	$\forall a,b\in F,\ x,y\in V$	$(ax)(by)=(ab)(xy)$

The *dimension* of the algebra is the dimension of the underlying linear space.

Definition 1.3. An algebra V is said to be *commutative* if

(A4) $\quad \forall x,y\in V \qquad xy=yx$

and said to be *associative* if

(A5) $\quad \forall x,y,z\in V \qquad x(yz)=(xy)z$

It is called *alternative* if less generally

(A6) $\quad \forall x,y\in V \qquad x(xy)=(xx)y$ and $(xy)y=x(yy)$

and is said to be *unital* if there exists an element $1_V\in V$ such that

(A7) $\quad \forall x\in V \qquad 1_V\cdot x=x=x\cdot 1_V$

Definition 1.4. By the *center* of an algebra V is meant the set

$\mathrm{Cen}(V)=\{x\in V\mid xy=yx \text{ for all } y\in V\}$

Lemma 1.5. *For any unital algebra V over a field F we have $F\cdot 1_V\subseteq \mathrm{Cen}(V)$.*

Proof. For all $a\in F$ and $x\in V$

$$\begin{aligned}(a\cdot 1_V)x=(a\cdot 1_V)(1_F\cdot x)&=(a\cdot 1_F)(1_V\cdot x)=ax\\ &=(1_F\cdot a)(x\cdot 1_V)=(1_F\cdot x)(a\cdot 1_V)=x(a\cdot 1_V)\end{aligned}$$

Definition 1.6. A *division algebra* is an algebra $V \neq \{0_V\}$ such that for all $y, z \in V$, $y \neq 0_V$ the two equations $xy = z$ and $yx = z$ each have a unique solution $x \in V$.

Lemma 1.7. *A division algebra is without zero divisors.*

Proof. Let V be a division algebra, $x, y \in V$ and suppose $xy = 0_V$ with $y \neq 0_V$. Since $0_V \cdot y = 0_V$ by definition of 0_V we have $x = 0_V$ by unicity.

Lemma 1.8. *A finite-dimensional algebra $V \neq \{0_V\}$ without zero divisors is a division algebra.*

Proof. Let $y \in V$, $y \neq 0_V$. The linear transformation $T : V \to V$, $x \mapsto xy$ has kernel $N(T) = \{0_V\}$ by V being without zero divisors. T is therefore injective and as a linear transformation from a finite-dimensional linear space to itself therefore bijective. $xy = T(x) = z$ therefore has the unique solution $x = T^{-1}(z) \in V$. By considering $T : V \to V$, $x \mapsto yx$ instead, $yx = z$ similarly has.

Lemma 1.9. *An alternative division algebra is unital.*

Proof. Let V be an alternative division algebra and $y \in V$, $y \neq 0_V$. Let $x = 1_V \in V$ be the unique solution of $xy = y$. Since $0_V \cdot y = 0_V \neq y$ we have $1_V \neq 0_V$. By alternativity and by being without zero divisors we obtain

$$1_V^2 \cdot y = 1_V(1_V \cdot y) = 1_V \cdot y \iff (1_V^2 - 1_V)y = 0_V \iff 1_V^2 = 1_V$$

Therefore, for all $x \in V$

$$1_V(1_V \cdot x - x) = 1_V(1_V \cdot x) - 1_V \cdot x = 1_V^2 \cdot x - 1_V \cdot x = 1_V \cdot x - 1_V \cdot x = 0_V$$

so that $1_V \cdot x = x$ by V being without zero divisors. In the same way,

$$(x \cdot 1_V - x)1_V = (x \cdot 1_V)1_V - x \cdot 1_V = x \cdot 1_V^2 - x \cdot 1_V = x \cdot 1_V - x \cdot 1_V = 0_V$$

so that $x \cdot 1_V = x$.

We concern ourselves exclusively with finite-dimensional *real algebras* which is to say finite-dimensional algebras over the field of real numbers $\mathbb{R}$. Together with the regular 0 and regular addition and multiplication $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are such of dimension 1, 2 and 4 respectively by trivial verification of the axioms. $\mathbb{R}$ and $\mathbb{C}$ are commutative whereas $\mathbb{H}$ is not, and all three are associative and unital. All three are without zero divisors by their multiplicative and nondegenerate absolute values, and are therefore division algebras.

We will now be proving that $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are in fact (up to isomorphism) also the *only* finite-dimensional associative real division algebras.

1.4 Frobenius

Theorem 1.10 (Frobenius)**.** *If $\mathbb{D}$ is a finite-dimensional associative real division algebra, then it is isomorphic to $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$.*

The proof that we present is a split off version of the first part of (a slight adaptation of) an elegant proof due to A. Oneto [13] of the generalised theorem of Frobenius for alternative division algebras, which is itself a generalised version of an elegant proof due to R.S. Palais [10] of this original theorem.

Other than by implying existence of a unit in the first step, alternativity will for now only play the role of being implicitly assumed at a number of places throughout the proof and *associativity* that of being assumed in the final step only. We will get back to the issue later when we adopt all but the final step of this proof unchanged as the first part of the proof of the generalised theorem.

Proof. $\mathbb{D}$ is unital by being alternative and we denote $0 = 0_{\mathbb{D}}$ and $1 = 1_{\mathbb{D}}$. Let $R = \mathbb{R}1$ be the natural inclusion of $\mathbb{R}$ in $\mathbb{D}$. R is clearly a subspace of $\mathbb{D}$ and trivially isomorphic to $\mathbb{R}$. For all $a \in R$ and $x \in \mathbb{D}$ $ax = xa$ by lemma 1.5.

(1) If $x \in \mathbb{D}$ then $x^2 \in R + Rx$.

Proof. Let $n = \dim \mathbb{D}$ and $x \in \mathbb{D}$. The set of powers $\{1, x, x^2, \dots\} \subseteq \mathbb{D}$ has from 1 to n linearly independent elements meaning that for any $n+1$ elements $x^{i_0}, x^{i_1}, \dots, x^{i_n}$ with $i_m < i_{m+1}$ for all $m < n$ there exist $\lambda_0, \lambda_1, \dots, \lambda_{n-1} \in \mathbb{R}$ not all equal $0_{\mathbb{R}}$ such that

$$\lambda_0 x^{i_0} + \lambda_1 x^{i_1} + \cdots + \lambda_{n-1} x^{i_{n-1}} + x^{i_n} = 0$$

or if we denote by $\Phi_x : R[X] \to \mathbb{D}$ the evaluation homomorphism[2] $f \mapsto f(x)$,

$$\Phi_x \left(\lambda_0 1 X^{i_0} + \lambda_1 1 X^{i_1} + \cdots + \lambda_{n-1} 1 X^{i_{n-1}} + X^{i_n} \right) = 0$$

The argument to Φ_x is a nonconstant monic polynomial in $R[X] \cong \mathbb{R}[X]$ which by the fundamental theorem of algebra factors into irreducible quadratic and linear polynomials in $R[X]$. That is, for some a_i, b_i, $c_i \in R$

$$\Phi_x \left(\prod (a_i + b_i X + X^2) \prod (c_i + X) \right) = 0$$

and by Φ_x being a homomorphism therefore

$$0 = \prod \Phi_x(a_i + b_i X + X^2) \prod \Phi_x(c_i + X) = \prod (a_i + b_i x + x^2) \prod (c_i + x)$$

As a division algebra, $\mathbb{D}$ is without zero divisors so either $a_i + b_i x + x^2 = 0$ for one or more i, meaning $x^2 = -a_i - b_i x \in R + Rx$, or $c_i + x = 0$ for one or more i, meaning $x^2 = c_i^2 \in R \subseteq R + Rx$, as it was to show.

We will from this point on no longer explicitly remark on $\mathbb{D}$ being without zero divisors. We need the fact throughout though.

[2] Note that the fact that it *is* a homomorphism needs $ax = xa$ for all $a \in R$.

So as to not drown ourselves in notational detail we will identify $a = a_{\mathbb{R}} \cdot 1 \in R$ with $a_{\mathbb{R}} \in \mathbb{R}$ in taking for example $a > 0$ to mean $a_{\mathbb{R}} > 0$ and setting

$$\frac{1}{a} := \frac{1_{\mathbb{R}}}{a_{\mathbb{R}}} \cdot 1 \qquad \text{and} \qquad \sqrt{a} := \sqrt{a_{\mathbb{R}}} \cdot 1$$

We moreover let lemma 1.5 enable us to pull elements of R, especially $-1 \in R$, through products without hesitation or mention. We do note that for $x \in \mathbb{D}$ the product xa is strictly speaking defined only for $a \in R$ which warrants not being *completely* implicit about R versus $\mathbb{R}$ but the reader should feel free to project an innate understanding of $\mathbb{R}$ onto R.

The first point above will function as a lemma and the remainder of the proof consists of a set of consecutive subproofs that we will in the end gather up to arrive at the conclusion.

(2) If $\mathbb{D} \neq R$ then there exists an element $i \in \mathbb{D}$ such that $i^2 = -1$.

Proof. Let $x \in \mathbb{D} \setminus R$. By the above, $x^2 \in R + Rx$, say $x^2 = a + bx$ for $a, b \in R$.

$$\left(x - \frac{b}{2}\right)^2 = a + \frac{b^2}{4}$$

Since $a + \frac{b^2}{4} \geq 0$ would mean $x = \frac{b}{2} \pm \sqrt{a + \frac{b^2}{4}} \in R$ we have $a + \frac{b^2}{4} < 0$. Setting

$$0 \neq c = \sqrt{-\left(a + \frac{b^2}{4}\right)} \in R \qquad \text{and} \qquad i = \frac{1}{c}\left(x - \frac{b}{2}\right) \in \mathbb{D}$$

we obtain $i^2 = -1$.

We assume from this point on that $\mathbb{D} \neq R$ and $i \in \mathbb{D}$ is such that $i^2 = -1$.

(3) $C = \{x \in \mathbb{D} \,|\, xi = ix\}$ is isomorphic to $\mathbb{C}$.

Proof. We claim that $C = R + Ri$, which is trivially isomorphic to $\mathbb{C}$.

($\subseteq$) Let $x \in C$. $R \subseteq C$ by lemma 1.5 so either $x \in R \subseteq R + Ri$ or $x \notin R$ and then by the above

$$\left(x - \frac{b}{2}\right)^2 = -c^2$$

for some $b, c \in R$. Together with $xi = ix$ and $i^2 = -1$ we then have

$$\left(x - \frac{b}{2} + ci\right)\left(x - \frac{b}{2} - ci\right) = \left(x - \frac{b}{2}\right)^2 + c^2 = 0$$

and thereby $x = \frac{b}{2} \pm ci \in R + Ri$ also in that case.

($\supseteq$) Let $x \in R + Ri$, say $x = a + bi$ for $a, b \in R$. Trivially $xi = ix$, so $x \in C$.

(4) $\mathbb{D} = C \oplus C^-$ in which $C^- = \{x \in \mathbb{D} \,|\, xi = -ix\}$.

Proof. Consider the linear transformation $T : \mathbb{D} \to \mathbb{D}$, $x \mapsto ixi$ and note that

$$C = \{x \in \mathbb{D} \,|\, xi = ix\} = \{x \in \mathbb{D} \,|\, ixi = -x\} = \{x \in \mathbb{D} \,|\, T(x) = -x\}$$

and

$$C^- = \{x \in \mathbb{D} \,|\, xi = -ix\} = \{x \in \mathbb{D} \,|\, ixi = x\} = \{x \in \mathbb{D} \,|\, T(x) = x\}$$

which shows C and C^- to be eigenspaces of T belonging to eigenvalues -1 and 1 respectively. Furthermore note that $T^2 = I$ from

$$T^2(x) = T(ixi) = i(ixi)i = (ii)x(ii) = x$$

This firstly shows eigenvalues of T to be simple eigenvalues (since a nontrivial Jordan-block would mean $T^n \neq I$ for any n) and secondly that -1 and 1 are the *only* possible eigenvalues. Therefore $\mathbb{D} = \bigoplus_{\lambda \in \sigma(T)} E_\lambda = C \oplus C^-$ as it was to show.

(5) If $\mathbb{D} \neq C$ then there exists an element $j \in C^-$ such that $j^2 = -1$.

Proof. If $\mathbb{D} = C \oplus C^- \neq C$ then $C^- \neq \{0\}$ meaning there exists an $x \in C^-$, $x \neq 0$. Using $xi = -ix$ we obtain

$$x^2 i = (xx)i = x(xi) = -x(ix) = -(xi)x = (ix)x = i(xx) = ix^2$$

However, by (1) above $x^2 \in R + Rx$, say $x^2 = a + bx$ for $a, b \in R$, so also

$$x^2 i = (a + bx)i = ai + bxi = ai - bix = ix^2 - 2bix$$

Comparing and using $x \neq 0$ we see $b = 0$ and thereby $x^2 = a \in R$. Moreover, since $a \geq 0$ would imply $x = \sqrt{a} \in R$ and thereby $x \in R \cap C^- = \{0\}$, we must have $a < 0$. C^- is closed under scalar multiplication as a subspace and setting

$$j = \frac{1}{\sqrt{-a}}\, x \in C^-$$

we obtain $j^2 = -1$.

We assume from this point on that $\mathbb{D} \neq C$ and $j \in C^-$ is such that $j^2 = -1$.

(6) $C + Cj$ is isomorphic to $\mathbb{H}$.

Proof. Set $k = ij$. From $j \in C^-$ we have $ji = -ij = -k$ and further obtain

$$k^2 = (ij)(ij) = -(ij)(ji) = -i(jj)i = ii = -1$$

and

$$kj = (ij)j = i(jj) = -i \qquad jk = j(ij) = (ji)j = -kj = i$$
$$ik = i(ij) = (ii)j = -j \qquad ki = (ij)i = i(ji) = -ik = j$$

Together we have exactly the defining relations (1.1) of $\mathbb{H}$

$$i^2 = j^2 = k^2 = -1 \qquad ij = k = -ji \qquad jk = i = -kj \qquad ki = j = -ik$$

whereby $C + Cj = R + Ri + (R + Ri)j = R + Ri + Rj + Rk$ is trivially isomorphic to $\mathbb{H}$.

(7) $\mathbb{D} = C + Cj$.

Proof. Clearly $C + Cj \subseteq \mathbb{D}$ and we need to show $\mathbb{D} \subseteq C + Cj$. Since $\mathbb{D} = C + C^-$ by (4) it moreover suffices to show that $C^- \subseteq Cj$.

Let to this end $y \in C^-$ and $x = -yj$. Then $x \in C$ since

$$xi = -(yj)i = -y(ji) = y(ij) = (yi)j = -(iy)j = -i(yj) = ix$$

by definition of x, associativity and $y \in C^-$. We furthermore have

$$xj = -(yj)j = y$$

by definition of x and alternativity and therefore $y \in Cj$ as it was to show.

We now only need to gather up these arguments to end the proof. Specifically, if $\mathbb{D}$ is a finite-dimensional associative real division algebra it could firstly be R and thereby trivially isomorphic to $\mathbb{R}$. If it is not R then it can by (4) be C which is isomorphic to $\mathbb{C}$ by (2) and (3). If it is also not C then it is by (7) $C + Cj$ which is by (5) and (6) isomorphic to $\mathbb{H}$, ending the proof.

We quickly note at this point that we see associativity explicitly featured in the above last point (7) whereas points (1) to (6) in fact use only alternativity as we will show in more detail later.

1.5 The Cayley-Dickson construction

Step (6) in the proof of the theorem of Frobenius shows that we have a rather more structured view of $\mathbb{H}$ available than the one resulting from the historic development by which we have introduced it.

That is, analogous to our view of the complex numbers $\mathbb{C}$ as $\mathbb{R} + \mathbb{R}i$ we can by

$$a + bi + cj + dk = a + bi + (c + di)j$$

consider the quaternions $\mathbb{H}$ to be nothing other than $\mathbb{C} + \mathbb{C}j$, thereby hinting at a generic process of constructing a new algebra of twice the dimension from an existing one. Indeed such a generic process exists, and it requires no more of the parent algebra than availability of conjugation.

Definition 1.11. A *$*$-algebra* V over a field F is an algebra V over F together with an F-linear operation

$$*\colon V \to V,\ x \mapsto x^*$$

called *conjugation* such that

(C1)	$\forall x \in V$	$x^{**} := (x^*)^* = x$
(C2)	$\forall x, y \in V$	$(xy)^* = y^* x^*$

Remark 1.12. Note that $0_V^* = (0_F \cdot 0_V)^* = 0_F \cdot 0_V^* = 0_V$.

Lemma 1.13. *If V is a $*$-algebra then so is $W = V \oplus V$ with multiplication and conjugation inductively defined by respectively*

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) \quad \textit{and} \quad (x_1, x_2)^* = (x_1^*, -x_2)$$

Proof. W is by definition of the direct sum the linear space $W = V \times V$ together with $0_W = (0_V, 0_V)$ and addition and scalar multiplication inductively defined by respectively

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2) \qquad \text{and} \qquad a(x_1, x_2) = (ax_1, ax_2)$$

As to it being an algebra, for all $a, b \in F$ and $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in W$

(A1) $$\begin{aligned}(x_1, x_2)[(y_1, y_2) + (z_1, z_2)] &= (x_1, x_2)(y_1 + z_1, y_2 + z_2)\\ &= (x_1(y_1 + z_1) - (y_2 + z_2)x_2^*, x_1^*(y_2 + z_2) + (y_1 + z_1)x_2)\\ &= (x_1 y_1 + x_1 z_1 - y_2 x_2^* - z_2 x_2^*, x_1^* y_2 + x_1^* z_2 + y_1 x_2 + z_1 x_2)\\ &= (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) + (x_1 z_1 - z_2 x_2^*, x_1^* z_2 + z_1 x_2)\\ &= (x_1, x_2)(y_1, y_2) + (x_1, x_2)(z_1, z_2)\end{aligned}$$

(A2) $$\begin{aligned}[(x_1, x_2) + (y_1, y_2)](z_1, z_2) &= (x_1 + y_1, x_2 + y_2)(z_1, z_2)\\ &= ((x_1 + y_1)z_1 - z_2(x_2 + y_2)^*, (x_1 + y_1)^* z_2 + z_1(x_2 + y_2))\\ &= (x_1 z_1 + y_1 z_1 - z_2 x_2^* - z_2 y_2^*, x_1^* z_2 + y_1^* z_2 + z_1 x_2 + z_1 y_2)\\ &= (x_1 z_1 - z_2 x_2^*, x_1^* z_2 + z_1 x_2) + (y_1 z_1 - z_2 y_2^*, y_1^* z_2 + z_1 y_2)\\ &= (x_1, x_2)(z_1, z_2) + (y_1, y_2)(z_1, z_2)\end{aligned}$$

(A3) $$\begin{aligned}(a(x_1, x_2))(b(y_1, y_2)) &= (ax_1, ax_2)(by_1, by_2)\\ &= ((ax_1)(by_1) - (by_2)(ax_2)^*, (ax_1)^*(by_2) + (by_1)(ax_2))\\ &= ((ax_1)(by_1) - (by_2)(ax_2^*), (ax_1^*)(by_2) + (by_1)(ax_2))\\ &= ((ab)x_1 y_1 - (ba)y_2 x_2^*, (ab)x_1^* y_2 + (ba)y_1 x_2)\\ &= ((ab)x_1 y_1 - (ab)y_2 x_2^*, (ab)x_1^* y_2 + (ab)y_1 x_2)\\ &= (ab)(x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) = (ab)((x_1, x_2)(y_1, y_2))\end{aligned}$$

and as to it being a $*$-algebra, for all $(x_1, x_2), (y_1, y_2) \in W$

(C1) $(x_1, x_2)^{**} = (x_1^*, -x_2)^* = (x_1, x_2)$

(C2) $$\begin{aligned}&((x_1, x_2)(y_1, y_2))^* = (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2)^*\\ &= ((x_1 y_1 - y_2 x_2^*)^*, -x_1^* y_2 - y_1 x_2) = (y_1^* x_1^* - x_2 y_2^*, -y_1 x_2 - x_1^* y_2)\\ &= (y_1^*, -y_2)(x_1^*, -x_2) = (y_1, y_2)^*(x_1, x_2)^*\end{aligned}$$

Remark 1.14. $\dim W = \dim(V \oplus V) = \dim V + \dim V = 2 \dim V$.

This process of constructing a new $*$-algebra W from an existing $*$-algebra V is called *the Cayley-Dickson process* after mathematicians Arthur Cayley and Leonard Dickson who first investigated it and W is called the *Cayley-Dickson double* of V.

Applying the process iteratively for any starting $*$-algebra V we are provided with an infinite sequence of $*$-algebras doubling in dimension at each step.

Let from this point on V be a $*$-algebra and W its Cayley-Dickson double.

Lemma 1.15. *If $x^* = x$ for all $x \in V$ then W is commutative.*

Proof. V is firstly itself commutative by, for all $x, y \in V$

$$\text{(A4)} \quad xy = (xy)^* = y^* x^* = yx$$

and therefore, for all $(x_1, x_2), (y_1, y_2) \in W$

$$\begin{aligned}\text{(A4)} \quad (x_1, x_2)(y_1, y_2) &= (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) \\ &= (x_1 y_1 - y_2 x_2, x_1 y_2 + y_1 x_2) \\ &= (y_1 x_1 - x_2 y_2, y_1 x_2 + x_1 y_2) \\ &= (y_1 x_1 - x_2 y_2^*, y_1^* x_2 + x_1 y_2) = (y_1, y_2)(x_1, x_2)\end{aligned}$$

Remark 1.16. A $*$-algebra V for which $x^* = x$ for all $x \in V$ is sometimes said to be *real* but we reserve use of the adjective *real* for an algebra over the field of real numbers $\mathbb{R}$.

Lemma 1.17. *If V is commutative and associative then W is associative.*

Proof. Let $(x_1, x_2), (y_1, y_2), (z_1, z_2) \in W$.

$$\begin{aligned}\text{(A5)} \quad & (x_1, x_2)[(y_1, y_2)(z_1, z_2)] = (x_1, x_2)(y_1 z_1 - z_2 y_2^*, y_1^* z_2 + z_1 y_2) \\ &= (x_1(y_1 z_1 - z_2 y_2^*) - (y_1^* z_2 + z_1 y_2) x_2^*, \\ &\qquad x_1^*(y_1^* z_2 + z_1 y_2) + (y_1 z_1 - z_2 y_2^*) x_2) \\ &= (x_1 y_1 z_1 - x_1 z_2 y_2^* - y_1^* z_2 x_2^* + z_1 y_2 x_2^*, \\ &\qquad x_1^* y_1^* z_2 + x_1^* z_1 y_2 + y_1 z_1 x_2 - z_2 y_2^* x_2) \\ &= (x_1 y_1 z_1 - y_2 x_2^* z_1 - z_2 y_2^* x_1 + z_2 x_2^* y_1^*, \\ &\qquad y_1^* x_1^* z_2 - x_2 y_2^* z_2 + z_1 x_1^* y_2 + z_1 y_1 x_2) \\ &= ((x_1 y_1 - y_2 x_2^*) z_1 - z_2(y_2^* x_1 + x_2^* y_1^*), \\ &\qquad (y_1^* x_1^* - x_2 y_2^*) z_2 + z_1(x_1^* y_2 + y_1 x_2)) \\ &= ((x_1 y_1 - y_2 x_2^*) z_1 - z_2(x_1^* y_2 + y_1 x_2)^*, \\ &\qquad (x_1 y_1 - y_2 x_2^*)^* z_2 + z_1(x_1^* y_2 + y_1 x_2)) \\ &= (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2)(z_1, z_2) = [(x_1, x_2)(y_1, y_2)](z_1, z_2)\end{aligned}$$

Lemma 1.18. *If V is unital with unit 1_V then $1_V^* = 1_V$.*

Proof. $1_V^* = 1_V^* \cdot 1_V = 1_V^*(1_V^*)^* = (1_V^* \cdot 1_V)^* = (1_V^*)^* = 1_V$.

Lemma 1.19. *If V is unital with unit 1_V then W is unital with unit $(1_V, 0_V)$.*

Proof. Let $(x_1, x_2) \in W$. Note again that $0_V^* = 0_V$ for any $*$-algebra V.

$$\begin{aligned}\text{(A7)} \quad (1_V, 0_V)(x_1, x_2) &= (1_V \cdot x_1, 1_V^* \cdot x_2) = (1_V \cdot x_1, 1_V \cdot x_2) \\ &= (x_1, x_2) \\ &= (x_1 \cdot 1_V, 1_V \cdot x_2) = (x_1, x_2)(1_V, 0_V)\end{aligned}$$

Now, starting with the commutative and associative real unital $*$-algebra $V = \mathbb{R}$ with trivial conjugation $x^* = x$, its Cayley-Dickson double is $W = \mathbb{R} \oplus \mathbb{R}$ with regular addition and scalar multiplication and multiplication

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) = (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1)$$

Setting

$$1 = (1_\mathbb{R}, 0_\mathbb{R}) = 1_W \qquad \text{and} \qquad i = (0_\mathbb{R}, 1_\mathbb{R})$$

we obtain

$$i^2 = (0_\mathbb{R}, 1_\mathbb{R})(0_\mathbb{R}, 1_\mathbb{R}) = (-1_\mathbb{R}, 0_\mathbb{R}) = -(1_\mathbb{R}, 0_\mathbb{R}) = -1$$

and together with $\mathbb{R}1 \subseteq \mathrm{Cen}(W)$ by lemma 1.5, W shows itself to be trivially isomorphic to $\mathbb{C}$. Lemma 1.15 expresses the familiar fact that $\mathbb{C}$ is commutative and 1.17 the familiar fact that $\mathbb{C}$ is associative. Also note that the induced conjugation $(x_1, x_2) = (x_1, -x_2)$ on W is no other than the regular conjugation on $\mathbb{C}$.

Repeating the process, we start with the commutative and associative real unital $*$-algebra $V = \mathbb{C}$ with conjugation $x^* = \bar{x}$ and Cayley-Dickson double $W = \mathbb{C} \oplus \mathbb{C}$ again with regular addition and scalar multiplication and multiplication

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2) = (x_1 y_1 - \bar{x}_2 y_2, \bar{x}_1 y_2 + x_2 y_1)$$

Setting

$$1 = (1_\mathbb{C}, 0_\mathbb{C}) = 1_W \qquad \text{and} \qquad i = (i_\mathbb{C}, 0_\mathbb{C}) \qquad \text{and} \qquad j = (0_\mathbb{C}, 1_\mathbb{C})$$

we obtain

$$i^2 = (i_\mathbb{C}, 0_\mathbb{C})(i_\mathbb{C}, 0_\mathbb{C}) = (i_\mathbb{C} i_\mathbb{C}, 0_\mathbb{C}) = (-1_\mathbb{C}, 0_\mathbb{C}) = -(1_\mathbb{C}, 0_\mathbb{C}) = -1$$
$$j^2 = (0_\mathbb{C}, 1_\mathbb{C})(0_\mathbb{C}, 1_\mathbb{C}) = (-\bar{1}_\mathbb{C}, 0_\mathbb{C}) = (-1_\mathbb{C}, 0_\mathbb{C}) = -(1_\mathbb{C}, 0_\mathbb{C}) = -1$$

and

$$ji = (0_\mathbb{C}, 1_\mathbb{C})(i_\mathbb{C}, 0_\mathbb{C}) = (0_\mathbb{C}, i_\mathbb{C}) = -(0_\mathbb{C}, -i_\mathbb{C}) = -(i_\mathbb{C}, 0_\mathbb{C})(0_\mathbb{C}, 1_\mathbb{C}) = -ij$$

Lemma 1.17 says that W is associative and setting $k = ij$ we now obtain in the same way as in step (6) of the proof of the theorem of Frobenius the defining relations (1.1) of $\mathbb{H}$

$$i^2 = j^2 = k^2 = -1 \qquad ij = k = -ji \qquad jk = i = -kj \qquad ki = j = -ik$$

Together with $\mathbb{R}1 \subseteq \mathrm{Cen}(W)$ by lemma 1.5, W therefore shows itself to be trivially isomorphic to $\mathbb{H}$. The induced conjugation $(x_1, x_2)^* = (\bar{x}_1, -x_2)$ is moreover again no other than the regular conjugation on $\mathbb{H}$.

Specifically note that we now proved that $\mathbb{H}$ is in fact associative, whereas we previously only mentioned this being straightforward to verify. We also proved that $\mathbb{H}$ is not commutative simply by for example $ij = -ji$ but have a more structured method available as well, which will moreover show why we cannot expect to usefully continue this doubling process forever.

Let from this point on V be a *unital* $*$-algebra and W its Cayley-Dickson double.

Lemma 1.20. *We cannot have $(x_1, x_2)^* = (x_1, x_2)$ for all $(x_1, x_2) \in W$.*

Proof. $(0_V, 1_V)^* = (0_V, -1_V) = -(0_V, 1_V) \neq (0_V, 1_V)$.

Starting with the commutative and associative real unital $*$-algebra $\mathbb{R}$ with trivial conjugation $x^* = x$, the above lemma expresses the fact that in the first step of the Cayley-Dickson process we lose the property of trivial conjugation as indeed we know to be the case for $\mathbb{C}$.

Lemma 1.21. *If W is commutative then $x^* = x$ for all $x \in V$.*

Proof. Let $x \in V$. $(x^*, 0_V) = (0_V, x)(0_V, -1_V) = (0_V, -1_V)(0_V, x) = (x, 0_V)$.

The above lemma reverses lemma 1.15 for unital $*$-algebras. Given that the first step already lost trivial conjugation, we now know that after the second step we lose commutativity and specifically that $\mathbb{H}$ is therefore indeed not commutative.

Lemma 1.22. *If W is associative then V is commutative and associative.*

Proof. Let $x, y, z \in V$.

$$\text{(A4)} \quad \begin{aligned}(xy, 0_V) &= (x, 0_V)(y, 0_V) = [(0_V, x^*)(0_V, -1_V)]\,(y, 0_V) \\ &= (0_V, x^*)\,[(0_V, -1_V)(y, 0_V)] = (0_V, x^*)(0_V, -y) = (yx, 0_V)\end{aligned}$$

$$\text{(A5)} \quad \begin{aligned}((xy)z, 0_V) &= (xy, 0_V)(z, 0_V) = [(x, 0_V)(y, 0_V)]\,(z, 0_V) \\ &= (x, 0_V)\,[(y, 0_V)(z, 0_V)] = (x, 0_V)(yz, 0_V) = (x(yz), 0_V)\end{aligned}$$

Reversing lemma 1.17 for unital $*$-algebras, the above lemma now says that having lost commutativity at the second step we lose associativity at the third so that we know that the next algebra in the sequence can no longer be associative.

We will see shortly that it does remain a division algebra whereby the proved theorem of Frobenius of course already implied as much. We will also see that it is alternative, and that it is in fact the *last* division algebra in the sequence.

1.6 Octonions

We have up to now concentrated on associativity but recall that associativity in fact played an only implicit role in the development of quaternions, with availability of a multiplicative absolute value the guiding principle.

Let us therefore now firstly introduce a norm on these $*$-algebras. We will also specialise to real algebras at this point.

Definition 1.23. A real $*$-algebra V is said to be *nicely normed* if it is unital and

(N1)	$\forall x \in V$	$x + x^* \in \mathbb{R}1_V$
(N2)	$\forall x \in V,\ x \neq 0_V$	$xx^* = x^*x \in \mathbb{R}^+1_V$

We define $\mathrm{Re}(x)1_V := \frac{1}{2}(x + x^*)$ and $\|x\|^2\, 1_V := xx^*$ for $x \in V$.

Remark 1.24. If V is nicely normed it is trivially verified that

$$\langle x, y\rangle = \operatorname{Re}(xy^*)$$

is an inner product on V whereby $\|x\| = \sqrt{\langle x, x\rangle}$ is in fact a norm.

Lemma 1.25. *If V is nicely normed then $\|x^*\| = \|x\|$ for all $x \in V$.*

Proof. $\|x^*\|^2 1_V = x^*(x^*)^* = x^*x = xx^* = \|x\|^2 1_V$.

Lemma 1.26. *If V is nicely normed then W is nicely normed.*

Proof. W is unital with unit $1_W = (1_V, 0_V)$ by lemma 1.19. Let $(x_1, x_2) \in W$.

(N1) $$\begin{aligned}(x_1, x_2) + (x_1, x_2)^* &= (x_1, x_2) + (x_1^*, -x_2) = (x_1 + x_1^*, 0_V)\\ &= (2\operatorname{Re}(x_1)1_V, 0_V) = 2\operatorname{Re}(x_1)1_W\end{aligned}$$

(N2) $$\begin{aligned}(x_1, x_2)(x_1, x_2)^* &= (x_1, x_2)(x_1^*, -x_2) = (x_1x_1^* + x_2x_2^*, -x_1^*x_2 + x_1^*x_2)\\ &= (\|x_1\|^2 1_V + \|x_2\|^2 1_V, 0_V) = (\|x_1\|^2 + \|x_2\|^2)1_W\end{aligned}$$

Similarly $(x_1, x_2)^*(x_1, x_2) = (\|x_1\|^2 + \|x_2\|^2)1_W = (x_1, x_2)(x_1, x_2)^*$. If moreover $(x_1, x_2) \neq (0_V, 0_V) = 0_W$ then at least one of $\|x_1\|^2$ and $\|x_2\|^2$ is positive whereby $\|x_1\|^2 + \|x_2\|^2$ is positive.

Remark 1.27. Note that this also showed that $\|(x_1, x_2)\|^2 = \|x_1\|^2 + \|x_2\|^2$.

Lemma 1.28. *If V is associative and nicely normed then W is alternative.*

Proof. Let $(x_1, x_2), (y_1, y_2) \in W$. As to the left alternative law we have

(A6) $$\begin{aligned}(x_1, x_2)[(x_1, x_2)(y_1, y_2)] &= (x_1, x_2)(x_1y_1 - y_2x_2^*, x_1^*y_2 + y_1x_2)\\
&= (x_1(x_1y_1 - y_2x_2^*) - (x_1^*y_2 + y_1x_2)x_2^*,\\
&\qquad x_1^*(x_1^*y_2 + y_1x_2) + (x_1y_1 - y_2x_2^*)x_2)\\
&= (x_1x_1y_1 - x_1y_2x_2^* - x_1^*y_2x_2^* - y_1x_2x_2^*,\\
&\qquad x_1^*x_1^*y_2 + x_1^*y_1x_2 + x_1y_1x_2 - y_2x_2^*x_2)\\
&= (x_1x_1y_1 - (x_1 + x_1^*)y_2x_2^* - y_1x_2x_2^*,\\
&\qquad x_1^*x_1^*y_2 + (x_1^* + x_1)y_1x_2 - y_2x_2^*x_2)\\
&= (x_1x_1y_1 - 2\operatorname{Re}(x_1)y_2x_2^* - \|x_2\|^2 y_1,\\
&\qquad x_1^*x_1^*y_2 + 2\operatorname{Re}(x_1)y_1x_2 - \|x_2\|^2 y_2)\\
&= (x_1x_1y_1 - y_2x_2^*(x_1 + x_1^*) - x_2x_2^*y_1,\\
&\qquad x_1^*x_1^*y_2 + y_1(x_1^* + x_1)x_2 - x_2x_2^*y_2)\\
&= (x_1x_1y_1 - y_2x_2^*x_1 - y_2x_2^*x_1^* - x_2x_2^*y_1,\\
&\qquad x_1^*x_1^*y_2 + y_1x_1^*x_2 + y_1x_1x_2 - x_2x_2^*y_2)\\
&= ((x_1x_1 - x_2x_2^*)y_1 - y_2(x_2^*x_1 + x_2^*x_1^*),\\
&\qquad (x_1^*x_1^* - x_2x_2^*)y_2 + y_1(x_1^*x_2 + x_1x_2))\\
&= (x_1x_1 - x_2x_2^*, x_1^*x_2 + x_1x_2)(y_1, y_2) = [(x_1, x_2)(x_1, x_2)](y_1, y_2)\end{aligned}$$

and similarly as to the right alternative law.

Lemma 1.29. *If V is finite-dimensional, nicely normed and alternative then V is a division algebra.*

Proof. Lemma 1.8 says that we need to show that V is without zero divisors. Let therefore $x, y \in V$, $y \neq 0_V$ and suppose that $xy = 0_V$. Using the left alternative law we obtain

$$\begin{aligned} \|x\|^2 y &= (x^*x)y = ((x+x^*)x)y - (xx)y = 2\operatorname{Re}(x)(xy) - x(xy) \\ &= 2\operatorname{Re}(x) \cdot 0_V - x \cdot 0_V = 0_V \end{aligned}$$

Taking the norm of both sides we have $\|x\|^2 = 0_{\mathbb{R}}$ and therefore $x = 0_V$, as it was to prove.

$\mathbb{R}$ with trivial conjugation $x^* = x$ is clearly nicely normed with $xx^* = x^*x = x^2$ and $\|x\| = |x| = \sqrt{x^2}$ whereby its entire sequence of Cayley-Dickson doubles is nicely normed per the above. We are moreover now guaranteed yet another division algebra in the sequence after $\mathbb{H}$ so let us pick things up where we left off in the previous section.

That is, we let $V = \mathbb{H}$ be the nicely normed associative real unital $*$-algebra with conjugation $x^* = \bar{x}$ and Cayley-Dickson double $W = \mathbb{H} \oplus \mathbb{H}$ with regular addition and scalar multiplication and multiplication

$$(x_1, x_2)(y_1, y_2) = (x_1y_1 - y_2x_2^*, x_1^*y_2 + y_1x_2) = (x_1y_1 - y_2\bar{x}_2, \bar{x}_1y_2 + y_1x_2)$$

Setting

$$1 = i_0 = (1_{\mathbb{H}}, 0_{\mathbb{H}}) = 1_W \qquad i_1 = (i_{\mathbb{H}}, 0_{\mathbb{H}}) \qquad i_2 = (j_{\mathbb{H}}, 0_{\mathbb{H}}) \qquad i_3 = (k_{\mathbb{H}}, 0_{\mathbb{H}})$$

we add $i_4 = (0_{\mathbb{H}}, 1_{\mathbb{H}})$ and complete a basis for the eight-dimensional W with

$$\begin{aligned} i_5 = i_1i_4 &= (i_{\mathbb{H}}, 0_{\mathbb{H}})(0_{\mathbb{H}}, 1_{\mathbb{H}}) = (0_{\mathbb{H}}, \bar{\imath}_{\mathbb{H}}) = (0_{\mathbb{H}}, -i_{\mathbb{H}}) = -(0_{\mathbb{H}}, i_{\mathbb{H}}) \\ i_6 = i_2i_4 &= (j_{\mathbb{H}}, 0_{\mathbb{H}})(0_{\mathbb{H}}, 1_{\mathbb{H}}) = (0_{\mathbb{H}}, \bar{\jmath}_{\mathbb{H}}) = (0_{\mathbb{H}}, -j_{\mathbb{H}}) = -(0_{\mathbb{H}}, j_{\mathbb{H}}) \\ i_7 = i_3i_4 &= (k_{\mathbb{H}}, 0_{\mathbb{H}})(0_{\mathbb{H}}, 1_{\mathbb{H}}) = (0_{\mathbb{H}}, \bar{k}_{\mathbb{H}}) = (0_{\mathbb{H}}, -k_{\mathbb{H}}) = -(0_{\mathbb{H}}, k_{\mathbb{H}}) \end{aligned}$$

We briefly note that in the case of the quaternions writing $k = ij$ caused

$$k_{\mathbb{H}} = i_{\mathbb{H}} \cdot j_{\mathbb{H}} = (i_{\mathbb{C}}, 0_{\mathbb{C}})(0_{\mathbb{C}}, 1_{\mathbb{C}}) = (0_{\mathbb{C}}, \bar{\imath}_{\mathbb{C}}) = (0_{\mathbb{C}}, -i_{\mathbb{C}}) = -(0_{\mathbb{C}}, i_{\mathbb{C}})$$

in the same way that the minus signs in i_5, i_6 and i_7 above are introduced. Were we to simply reverse their definition we would forego the minus signs but the current definitions best serve the algebraic notational style of writing for example $\mathbb{R}i$ versus $i\mathbb{R}$. This is clearly otherwise irrelevant.

Straightforward direct calculations such as for example

$$\begin{aligned} i_4^2 &= (0_{\mathbb{H}}, 1_{\mathbb{H}})(0_{\mathbb{H}}, 1_{\mathbb{H}}) = (-1_{\mathbb{H}}\bar{1}_{\mathbb{H}}, 0_{\mathbb{H}}) = (-1_{\mathbb{H}}, 0_{\mathbb{H}}) = -(1_{\mathbb{H}}, 0_{\mathbb{H}}) = -1 \\ i_5i_1 &= -(0_{\mathbb{H}}, i_{\mathbb{H}})(i_{\mathbb{H}}, 0_{\mathbb{H}}) = -(0_{\mathbb{H}}, i_{\mathbb{H}}i_{\mathbb{H}}) = -(0_{\mathbb{H}}, -1_{\mathbb{H}}) = (0_{\mathbb{H}}, 1_{\mathbb{H}}) = i_4 \end{aligned}$$

provide us with the below multiplication table 1.2 for W which together with $\mathbb{R}1 \subseteq \operatorname{Cen}(W)$ and distributivity lays down the rules of multiplication for this so constructed eight-dimensional alternative real division algebra that has become known as the space of *octonions* $\mathbb{O}$.

	i_1	i_2	i_3	i_4	i_5	i_6	i_7
i_1	-1	i_3	$-i_2$	i_5	$-i_4$	$-i_7$	i_6
i_2	$-i_3$	-1	i_1	i_6	i_7	$-i_4$	$-i_5$
i_3	i_2	$-i_1$	-1	i_7	$-i_6$	i_5	$-i_4$
i_4	$-i_5$	$-i_6$	$-i_7$	-1	i_1	i_2	i_3
i_5	i_4	$-i_7$	i_6	$-i_1$	-1	$-i_3$	i_2
i_6	i_7	i_4	$-i_5$	$-i_2$	i_3	-1	$-i_1$
i_7	$-i_6$	i_5	i_4	$-i_3$	$-i_2$	i_1	-1

Table 1.2: Octonion multiplication

Analogously to $\mathbb{C}$ and $\mathbb{H}$, the induced conjugation on $\mathbb{O}$ is given simply by

$$\bar{o} = a_0 - a_1 i_1 - a_2 i_2 - a_3 i_3 - a_4 i_4 - a_5 i_5 - a_6 i_6 - a_7 i_7$$

for an octonion $o = a_0 + a_1 i_1 + a_2 i_2 + a_3 i_3 + a_4 i_4 + a_5 i_5 + a_6 i_6 + a_7 i_7 \in \mathbb{O}$.

It is these octonions, which he called *octaves*, that John Graves discovered in December 1843 only two months after being told about the quaternions by Hamilton. They were one of the first ever uses of the concept of nonassociativity and are as such interesting from an historic viewpoint. Graves at the time communicated their discovery back to Hamilton who offered to publish about them but who, being absorbed by quaternions, kept putting it off until Arthur Cayley rediscovered them in 1845.

Hamilton then tried to remedy the situation by announcing Graves' priority but it was already too late and the octaves became known as *the Cayley numbers* instead. With Graves' interest in algebra having put Hamilton on the path to quaternions in the first place, it would have been a fitting tribute to this early pioneer of abstract algebra had they beared his name. These days we call them octonions by analogy with quaternions.

Nonassociativity of the octonions is easily directly verified by for example

$$(i_1 \cdot i_2) i_4 = i_3 \cdot i_4 = i_7 \neq -i_7 = i_1 \cdot i_6 = i_1 (i_2 \cdot i_4)$$

and we will at this point recap what we have done.

We have, assuming known no more than commutativity and associativity of $\mathbb{R}$, proved that $\mathbb{R}$ and $\mathbb{C}$ are commutative and associative real division algebras, that $\mathbb{H}$ is a noncommutative associative real division algebra and that $\mathbb{O}$ is a noncommutative nonassociative alternative real division algebra.

We have by the theorem of Frobenius seen that $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are up to isomorphism the only finite-dimensional associative real division algebras and therefore that $\mathbb{R}$ and $\mathbb{C}$ are up to isomorphism the only finite-dimensional commutative and associative real division algebras. We will next be proving the generalised theorem of Frobenius which says that $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are up to isomorphism the only finite-dimensional alternative real division algebras.

However, recall once more a multiplicative absolute value being the guiding principle for the development of the quaternions and note that we have not yet actually shown $\mathbb{O}$ to possess such an absolute value. We will do this now.

Definition 1.30. A real $*$-algebra V is said to be *absolute-valued* if it is nicely normed and, for all $x, y \in V$

$$\|xy\| = \|x\| \, \|y\|$$

We sometimes denote $|x| = \|x\|$ in that case.

Lemma 1.31. *If V is absolute-valued and associative then W is absolute-valued.*

Proof. From section 6.5 of Kantor and Solodovnikov [11].

Let V be absolute-valued and associative and let $(x_1, x_2), (y_1, y_2) \in W$. We have by remark 1.27 firstly

$$\|(x_1, x_2)\|^2 \, \|(y_1, y_2)\|^2 = (\|x_1\|^2 + \|x_2\|^2)(\|y_1\|^2 + \|y_2\|^2)$$

and secondly

$$\begin{aligned}
\|(x_1, x_2)(y_1, y_2)\|^2 \, 1_V &= \|(x_1 y_1 - y_2 x_2^*, x_1^* y_2 + y_1 x_2)\|^2 \, 1_V \\
&= (\|x_1 y_1 - y_2 x_2^*\|^2 + \|x_1^* y_2 + y_1 x_2)\|^2) 1_V \\
&= (x_1 y_1 - y_2 x_2^*)(x_1 y_1 - y_2 x_2^*)^* + (x_1^* y_2 + y_1 x_2)(x_1^* y_2 + y_1 x_2)^* \\
&= (x_1 y_1 - y_2 x_2^*)(y_1^* x_1^* - x_2 y_2^*) \; + (x_1^* y_2 + y_1 x_2)(y_2^* x_1 + x_2^* y_1^*) \\
&= \; x_1 y_1 y_1^* x_1^* - x_1 y_1 x_2 y_2^* - y_2 x_2^* y_1^* x_1^* + y_2 x_2^* x_2 y_2^* \\
&\qquad + x_1^* y_2 y_2^* x_1 + x_1^* y_2 x_2^* y_1^* + y_1 x_2 y_2^* x_1 + y_1 x_2 x_2^* y_1^*) \\
&= \; (\|x_1 y_1\|^2 + \|y_2 x_2^*\|^2 + \|x_1^* y_2\|^2 + \|y_1 x_2\|^2) 1_V \\
&\qquad - x_1 y_1 x_2 y_2^* - y_2 x_2^* y_1^* x_1^* + x_1^* y_2 x_2^* y_1^* + y_1 x_2 y_2^* x_1 \\
&= \; (\|x_1\|^2 \|y_1\|^2 + \|y_2\|^2 \|x_2^*\|^2 + \|x_1^*\|^2 \|y_2\|^2 + \|y_1\|^2 \|x_2\|^2) 1_V \\
&\qquad - x_1 y_1 x_2 y_2^* - y_2 x_2^* y_1^* x_1^* + x_1^* y_2 x_2^* y_1^* + y_1 x_2 y_2^* x_1 \\
&= \; (\|x_1\|^2 \|y_1\|^2 + \|x_2\|^2 \|y_2\|^2 + \|x_1\|^2 \|y_2\|^2 + \|x_2\|^2 \|y_1\|^2) 1_V \\
&\qquad - x_1 y_1 x_2 y_2^* - y_2 x_2^* y_1^* x_1^* + x_1^* y_2 x_2^* y_1^* + y_1 x_2 y_2^* x_1 \\
&= \; (\|x_1\|^2 + \|x_2\|^2)(\|y_1\|^2 + \|y_2\|^2) 1_V \\
&\qquad - x_1 y_1 x_2 y_2^* - y_2 x_2^* y_1^* x_1^* + x_1^* y_2 x_2^* y_1^* + y_1 x_2 y_2^* x_1
\end{aligned}$$

We therefore need to show $x_1^* y_2 x_2^* y_1^* - x_1 y_1 x_2 y_2^* + y_1 x_2 y_2^* x_1 - y_2 x_2^* y_1^* x_1^* = 0_V$.

Assume x_1 to have zero imaginary part, $x_1 = \mathrm{Re}(x_1) 1_V$. Then $x_1^* = x_1$ and

$$\begin{aligned}
&x_1^* y_2 x_2^* y_1^* - x_1 y_1 x_2 y_2^* + y_1 x_2 y_2^* x_1 - y_2 x_2^* y_1^* x_1^* \\
&\qquad = \mathrm{Re}(x_1)(y_2 x_2^* y_1^* - y_1 x_2 y_2^* + y_1 x_2 y_2^* - y_2 x_2^* y_1^*) = \mathrm{Re}(x_1) \cdot 0_V = 0_V
\end{aligned}$$

Next assume x_1 to be purely imaginary, $\mathrm{Re}(x_1) = 0_{\mathbb{R}}$. Then $x_1^* = -x_1$ and

$$\begin{aligned}
&x_1^* y_2 x_2^* y_1^* - x_1 y_1 x_2 y_2^* + y_1 x_2 y_2^* x_1 - y_2 x_2^* y_1^* x_1^* \\
&\qquad = -x_1 y_2 x_2^* y_1^* - x_1 y_1 x_2 y_2^* + y_1 x_2 y_2^* x_1 + y_2 x_2^* y_1^* x_1 \\
&\qquad = (y_2 x_2^* y_1^* + y_1 x_2 y_2^*) x_1 - x_1 (y_2 x_2^* y_1^* + y_1 x_2 y_2^*) \\
&\qquad = 2 \, \mathrm{Re}(y_2 x_2^* y_1^*) x_1 - 2 \, \mathrm{Re}(y_2 x_2^* y_1^*) x_1 = 0_V
\end{aligned}$$

For a general x_1 we write $x_1 = \mathrm{Re}(x_1) 1_V + (x_1 - \mathrm{Re}(x_1) 1_V)$ and use linearity of the conjugate.

$\mathbb{R}$ with trivial conjugation $x^* = x$ is clearly not only nicely normed but absolute-valued with $xx^* = x^*x = x^2$ and $\|x\| = |x| = \sqrt{x^2}$ and we therefore have now shown $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ to all four be absolute valued alternative real division algebras.

The next algebra in the sequence, called $\mathbb{S}$ for the algebra of *sedenions*, is no longer a division algebra by introducing zero divisors. For example,

$$\begin{aligned}(i_3, i_2)(i_6, -i_7) &= (i_3 \cdot i_6 + i_7 \cdot i_2, i_3 \cdot i_7 + i_6 \cdot i_2) \\ &= (i_5 - i_5, -i_4 + i_4) = (0_{\mathbb{O}}, 0_{\mathbb{O}}) = 0_{\mathbb{S}}\end{aligned}$$

Zero divisors clearly lay to rest any hope of a multiplicative and nondegenerate absolute value and moreover, if a $*$-algebra V has zero divisors then so does its Cayley-Dickson double W by, if $xy = 0_V$ for some $x, y \in V$, $x \neq 0_V$, $y \neq 0_V$

$$(x, 0_V)(y, 0_V) = (xy, 0_V) = (0_V, 0_V) = 0_W$$

$\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are therefore the only division algebras in this otherwise infinite sequence of doubles.

1.7 Alternativity

Before getting to the generalised theorem of Frobenius in the next section we will need to quickly derive a few important results on alternativity for use in the proof. Let $\mathbb{D}$ be any alternative real division algebra (therefore unital) and denote $0 = 0_{\mathbb{D}}$ and $1 = 1_{\mathbb{D}}$.

Definition 1.32 (Associator)**.** We define for $x, y, z \in \mathbb{D}$

$$A(x, y, z) = x(yz) - (xy)z$$

The associator is clearly $\mathbb{R}$-linear in all three variables and identically zero if $\mathbb{D}$ is associative.

Lemma 1.33 (Antisymmetry of the associator)**.** *For all* $x, y, z \in \mathbb{D}$

$$A(x, y, z) = -A(y, x, z) = -A(x, z, y)$$

Proof. By the left alternative law and linearity of the associator we obtain

$$\begin{aligned}0 &= A(x + y, x + y, z) \\ &= A(x, x, z) + A(x, y, z) + A(y, x, z) + A(y, y, z) = A(x, y, z) + A(y, x, z)\end{aligned}$$

thereby proving the first equality. Similarly by the right alternative law

$$\begin{aligned}0 &= A(x, y + z, y + z) \\ &= A(x, y, y) + A(x, y, z) + A(x, z, y) + A(x, z, z) = A(x, y, z) + A(x, z, y)\end{aligned}$$

thereby proving the second equality.

Corollary 1.34. $A(x, y, z) = A(y, z, x) = A(z, x, y) = -A(z, y, x)$.

Lemma 1.35 (The flexible law)**.** *For all* $x, y \in \mathbb{D}$

$$(xy)x = x(yx)$$

and we will denote this common value by xyx.

Proof. $A(x, y, x) = -A(x, x, y) = 0$.

Lemma 1.36 (Anticommutativity)**.** *If* $xy = -yx$ *for* $x, y \in \mathbb{D}$ *then for all* $z \in \mathbb{D}$

$$x(yz) = -y(xz) \qquad \textit{and} \qquad (zx)y = -(zy)x$$

Proof of left part. By $A(x, y, z) = -A(y, x, z)$ and $xy + yx = 0$ we obtain

$$\begin{aligned} 0 &= A(x, y, z) + A(y, x, z) + (xy + yx)z \\ &= x(yz) - (xy)z + y(xz) - (yx)z + (xy)z + (yx)z = x(yz) + y(xz) \end{aligned}$$

Proof of right part. By $A(z, x, y) = -A(z, y, x)$ and $xy + yx = 0$ we obtain

$$\begin{aligned} 0 &= A(z, x, y) + A(z, y, x) - z(xy + yx) \\ &= z(xy) - (zx)y + z(yx) - (zy)x - z(xy) - z(yx) = -(zx)y - (zy)x \end{aligned}$$

Lemma 1.37 (The Moufang identities)**.** *For all* $x, y, z \in \mathbb{D}$

$$z(x(zy)) = (zxz)y \qquad (zx)(yz) = z(xy)z \qquad ((xz)y)z = x(zyz)$$

Proof.

We will prove all three identities, respectively called the left, middle and right Moufang identity, if only because it makes for a somewhat more memorably structured proof but will *use* the middle one only, when it will be referred to simply as the Moufang identity or even just Moufang.

Proof of left Moufang identity.

$$\begin{aligned} z(x(zy)) - (zxz)y &= z(x(zy)) - (zx)(zy) + (zx)(zy) - (zxz)y \\ &= A(z, x, zy) + A(zx, z, y) = A(z, x, zy) + A(z, y, zx) \end{aligned}$$

Let $T(x, y, z) = A(z, x, zy) + A(z, y, zx)$. Then $T(y, x, z) = T(x, y, z)$ and

$$\begin{aligned} T(x, x, z) &= 2A(z, x, zx) = -2A(z, zx, x) = -2[z((zx)x) - (z(zx))x] \\ &= -2[z(z(xx)) - ((zz)x)x] = -2[(zz)(xx) - (zz)(xx)] = 0 \end{aligned}$$

for all $x, y, z \in \mathbb{D}$ by both left and right alternative laws. T moreover inherits linearity from A so that

$$\begin{aligned} 0 &= T(x + y, x + y, z) \\ &= T(x, x, z) + T(x, y, z) + T(y, x, z) + T(y, y, z) = 2T(x, y, z) \end{aligned}$$

and thereby $z(x(zy)) - (zxz)y = T(x, y, z) = 0$ as it was to show.

Proof of middle Moufang identity.

$$\begin{aligned}(zx)(yz) - z(xy)z &= z(x(yz)) - z(xy)z - z(x(yz)) + (zx)(yz) \\ &= zA(x, y, z) - A(z, x, yz) \\ &= zA(y, z, x) + A(z, yz, x) \\ &= z[y(zx) - (yz)x] + z((yz)x) - (zyz)x \\ &= z(y(zx)) - (zyz)x = 0\end{aligned}$$

in which the last step is the left Moufang identity with the roles of x and y reversed.

Proof of right Moufang identity.

$$\begin{aligned}((xz)y)z - x(zyz) &= ((xz)y)z - (x(zy))z + (x(zy))z - x(zyz) \\ &= -A(x, z, y)z - A(x, zy, z) \\ &= -A(z, y, x)z + A(zy, x, z) \\ &= -[z(yx) - (zy)x]z + (zy)(xz) - ((zy)x)z \\ &= (zy)(xz) - z(yx)z = 0\end{aligned}$$

in which the last step is the middle Moufang identity, again with the roles of x and y reversed.

Remark 1.38. Note that while we proved that alternativity implies the Moufang identities above, the left and right Moufang identities

$$z(x(zy)) = (zxz)y \qquad \text{and} \qquad ((xz)y)z = x(zyz)$$

conversely imply in any unital algebra V the left and right alternative laws

$$z(zy) = (zz)y \qquad \text{and} \qquad (xz)z = x(zz)$$

by simply taking $x = 1_V$ in the left case and $y = 1_V$ in the right.

Lemma 1.39 (Powerassociativity)**.** *Defining inductively $x^1 = x$, $x^{n+1} = x^n x$ for $n \geq 1$ we have for all $x \in \mathbb{D}$ and $m \geq 1, n \geq 1$*

$$x^m x^n = x^{m+n}$$

Proof. We firstly prove, by induction on n, $xx^n = x^{n+1}$ for all $n \geq 1$.

For $n = 1$ we have $xx^1 = xx = x^1 x = x^2 = x^{1+1}$ and for the induction step,

$$xx^{n+1} = x(x^n x) = (xx^n)x = x^{n+1}x = x^{n+2}$$

by respectively definition, the flexible law, the induction hypothesis and definition again, as it was required to show.

Now proving the lemma by induction on m, the base case $m = 1$ is exactly the foregoing. In the induction step, for $n = 1$ the result is true by definition and for $n > 1$ we have

$$\begin{aligned}x^{m+1}x^n = (xx^m)x^n = (xx^m)(x^{n-1}x) &= x(x^m x^{n-1})x \\ &= xx^{m+n-1}x = xx^{m+n} = x^{m+n+1}\end{aligned}$$

by respectively the foregoing, definition, Moufang, induction hypothesis, definition and lastly the foregoing again, as it was required to show.

1.8 Generalised Frobenius

We are now ready for the proof of the generalised theorem of Frobenius, generally attributed to Max Zorn.

Theorem 1.40 (Generalised Frobenius)**.** *If* $\mathbb{D}$ *is a finite-dimensional alternative real division algebra, then it is isomorphic to* $\mathbb{R}$*,* $\mathbb{C}$*,* $\mathbb{H}$ *or* $\mathbb{O}$*.*

Proof. As mentioned at the time we have in fact already presented the first part of this proof as the proof of the original theorem of Frobenius. Since $\mathbb{O}$ is not associative we could of course have presented the original theorem as an immediate corollary to this more general theorem as well but thought it useful to present the first part of the proof without needing to worry about nonassociativity. We will revisit it now so as to make sure it still works in the alternative context. We will not fully repeat it here but ask the reader to refer to it while we comment.

$\mathbb{D}$ is still unital by being alternative and without zero divisors by being a division algebra. The notational identification of R with $\mathbb{R}$ is also still in effect.

(1) If $x \in \mathbb{D}$ then $x^2 \in R + Rx$.

Proof. The powers of x remain well-defined by powerassociativity of $\mathbb{D}$ and while the polynomial product should now be written out using brackets, nothing changes in the conclusion which only needs that $\mathbb{D}$ is without zero divisors.

(2) If $\mathbb{D} \neq R$ then there exists an element $i \in \mathbb{D}$ such that $i^2 = -1$.

Proof. Step (2) does not feature an issue of associativity.

(3) $C = \{x \in \mathbb{D} \,|\, xi = ix\}$ is isomorphic to $\mathbb{C}$.

Proof. Step (3) does not feature an issue of associativity.

(4) $\mathbb{D} = C \oplus C^-$ in which $C^- = \{x \in \mathbb{D} \,|\, xi = -ix\}$.

Proof. We are justified in writing and manipulating ixi by the flexible law.

(5) If $\mathbb{D} \neq C$ then there exists an element $j \in C^-$ such that $j^2 = -1$.

Proof. In the expression for $x^2 i$ we see the left and right alternative laws and the flexible law featured.

(6) $C + Cj$ is isomorphic to $\mathbb{H}$.

Proof. In the expression for k^2 we see Moufang and in the others the left and right alternative laws and flexible law featured.

The final step of the former proof explicitly featured associativity and will now be replaced by the first step of the second half of the proof.

(7) $\quad H = C \oplus (C^- \cap H)$ in which $H = \{x \in \mathbb{D} \,|\, xk = (xi)j\}$.

Proof. We adjust subproof (4) by now considering T to be a linear transformation from H to H instead of from $\mathbb{D}$ to $\mathbb{D}$.

That is, consider the linear transformation $T : H \to H$, $x \mapsto ixi$. Since $C + Cj$ is according to (6) isomorphic to $\mathbb{H}$, we have $C \subseteq C + Cj \subseteq H$. Therefore

$$C = C \cap H = \{x \in H \,|\, xi = ix\} = \{x \in H \,|\, ixi = -x\} = \{x \in H \,|\, T(x) = -x\}$$

meaning that C still shows itself to be the eigenspace of T belonging to the eigenvalue -1. Together with

$$C^- \cap H = \{x \in H \,|\, xi = -ix\} = \{x \in H \,|\, ixi = x\} = \{x \in H \,|\, T(x) = x\}$$

the conclusion $H = C \oplus (C^- \cap H)$ therefore follows in the same manner as in subproof (4). It however remains to show that T is in fact well-defined, which is to say that we need to show $T(H) \subseteq H$.

Let to this end $y \in T(H)$, say $y = T(x) = ixi$ for $x \in H$. We firstly obtain

$$yk = (ixi)k = -((ix)k)i = ((ix)(-k))i = ((ix)(ji))i = (i(xj)i)i = -i(xj)$$

by anticommutativity of i and k for the first equality and respectively Moufang and alternativity for the last two. By alternativity only

$$(yi)j = ((ixi)i)j = -(ix)j$$

By definition of H and antisymmetry of the associator, for any $x \in H$

$$0 = A(x, i, j) = -A(i, x, j) = -i(xj) + (ix)j \tag{1.11}$$

so that $yk = -i(xj) = -(ix)j = (yi)j$ and therefore $y \in H$ as it was to show.

(8) $\quad H = C + Cj$.

Proof. We already noted in subproof (7) above that $C + Cj \subseteq H$ so that it remains to show that $H \subseteq C + Cj$. Since $H = C + (C^- \cap H)$ by (7) it moreover suffices to show that $C^- \cap H \subseteq Cj$.

Let to this end $y \in C^- \cap H$ and $x = -yj$. Then $x \in C$ since

$$xi = -(yj)i = (yi)j = -(iy)j = -i(yj) = ix$$

by respectively definition of x, anticommutativity of j and i, $y \in C^-$, $y \in H$ together with (1.11) above, and lastly again definition of x. We furthermore have

$$xj = -(yj)j = y$$

by definition of x and alternativity and therefore $y \in Cj$ as it was to show.

(9) $\mathbb{D} = H \oplus H^-$ in which $H^- = \{x \in \mathbb{D} \,|\, xk = -(xi)j\}$.

Proof. Consider similarly to subproof (4) the linear transformation $T : \mathbb{D} \to \mathbb{D}$, $x \mapsto ((xi)j)k$ and note that

$$H = \{x \in \mathbb{D} \,|\, xk = (xi)j\} = \{x \in \mathbb{D} \,|\, x = -((xi)j)k\} = \{x \in \mathbb{D} \,|\, T(x) = -x\}$$

and

$$H^- = \{x \in \mathbb{D} \,|\, xk = -(xi)j\} = \{x \in \mathbb{D} \,|\, x = ((xi)j)k\} = \{x \in \mathbb{D} \,|\, T(x) = x\}$$

By a series of invocations of anticommutativity and alternativity we see

$$\begin{aligned} T^2(x) = T(((xi)j)k) &= (((((xi)j)k)i)j)k = -(((((xi)j)k)i)k)j \\ &= (((((xi)j)k)k)i)j = -(((xi)j)i)j = (((xi)j)j)i = -(xi)i = x \end{aligned}$$

so that again $T^2 = I$ and $\mathbb{D} = H \oplus H^-$ in the same manner as in subproof (4).

The next point functions as a lemma for the remaining three points.

(10) If $x \in H^-$ then x anticommutes with i, j and k.

Proof. Let $x \in H^-$. Right-multiplying $xk = -(xi)j$ by $-k$ we have $x = ((xi)j)k$ and therefore

$$kx = k((xi)j)k = (k(xi))(jk) = (k(xi))i = ((ij)(xi))i = (i(jx)i)i = -i(jx)$$

by Moufang and alternativity. By antisymmetry of the associator therefore

$$kx = -\frac{1}{2}A(i,j,x) = -\frac{1}{2}A(x,i,j) = -\frac{1}{2}(xk - (xi)j) = -xk$$

Similarly, from $x = ((xi)j)k = -((xi)k)j$ we obtain $xj = (xi)k$ and

$$\begin{aligned} jx = -j((xi)k)j = -(j(xi))(kj) = (j(xi))i \\ = -((ik)(xi))i = -(i(kx)i)i = i(kx) \end{aligned}$$

by Moufang and alternativity and thereby

$$jx = \frac{1}{2}A(i,k,x) = \frac{1}{2}A(x,i,k) = \frac{1}{2}(-xj - (xi)k) = -xj$$

Lastly by $i = jk$, anticommutativity of k and x, j and x, j and k, definition of H^- and alternativity

$$ix = (jk)x = -(jx)k = (xj)k = -(xk)j = ((xi)j)j = -xi$$

(11) If $\mathbb{D} \neq H$ then there exists an element $l \in H^-$ such that $l^2 = -1$.

Proof. If $\mathbb{D} = H \oplus H^- \neq H$ then $H^- \neq \{0\}$ meaning there exists an $x \in H^-$, $x \neq 0$. By (10) $H^- \subseteq C^-$ so that we can simply repeat subproof (5).

That is, using alternativity, $xi = -ix$ and flexibility we obtain

$$x^2 i = (xx)i = x(xi) = -x(ix) = -(xi)x = (ix)x = i(xx) = ix^2$$

However, by (1) we have $x^2 \in R + Rx$, say $x^2 = a + bx$ for $a, b \in R$, so also

$$x^2 i = (a + bx)i = ai + bxi = ai - bix = ix^2 - 2bix$$

Comparing and using $x \neq 0$ we see $b = 0$ and thereby $x^2 = a \in R$. Moreover, since $a \geq 0$ would imply $x = \sqrt{a} \in R$ and thereby $x \in R \cap H^- = \{0\}$, we must have $a < 0$. H^- is closed under scalar multiplication as a subspace and setting

$$l = \frac{1}{\sqrt{-a}} x \in H^-$$

we obtain $l^2 = -1$.

Assume from this point on that $\mathbb{D} \neq H$ and $l \in H^-$ is such that $l^2 = -1$.

(12) $H + Hl$ is isomorphic to $\mathbb{O}$.

Proof. We set $i_0 = 1$, $i_1 = i$, $i_2 = j$, $i_3 = k$, $i_4 = l$, $i_5 = il$, $i_6 = jl$ and $i_7 = kl$.

Straightforward calculations using anticommutativity, alternativity and Moufang such as

$$\begin{aligned} i_5 i_1 &= (il)i = -(li)i = l = i_4 \\ i_5 i_5 &= (il)(il) = -(li)(il) = -li^2 l = l^2 = -1 \end{aligned}$$

now again provide us with the defining set of relations for the octonions as collected in table 1.2, whereby

$$\begin{aligned} H + Hl &= C + Cj + (C + Cj)l \\ &= R + Ri + (R + Ri)j + (R + Ri + (R + Ri)j)l \\ &= R + Ri + Rj + Rk + (R + Ri + Rj + Rk)l \\ &= R + Ri_1 + Ri_2 + Ri_3 + Ri_4 + Ri_5 + Ri_6 + Ri_7 \end{aligned}$$

is trivially isomorphic to $\mathbb{O}$.

(13) $\mathbb{D} = H + Hl$.

Proof. Clearly $H+Hl \subseteq \mathbb{D}$ and we need to show $\mathbb{D} \subseteq H+Hl$. Since $\mathbb{D} = H+H^-$ by (9) it moreover suffices to show that $H^- \subseteq Hl$.

Let to this end $y \in H^-$ and $x = -yl$. Then $x \in H$ since

$$xk = -(yl)k = (yk)l = -((yi)j)l = ((yi)l)j = -((yl)i)j = (xi)j$$

by respectively definition of x, anticommutativity of l and k, $y \in H^-$, anticommutativity of j and l, of i and l and lastly definition of x again. We furthermore have

$$xl = -(yl)l = y$$

by definition of x and alternativity and therefore $y \in Hl$ at is was to show.

As before, we now only need to gather up these arguments to end the proof.

Specifically, if $\mathbb{D}$ is a finite-dimensional alternative real division algebra it could firstly be R and thereby trivially isomorphic to $\mathbb{R}$. If it is not R then it can by (4) be C which is isomorphic to $\mathbb{C}$ by (2) and (3). If it is also not C then it can by (9) be H which is isomorphic to $\mathbb{H}$ by (5), (6), (7) and (8). Finally, if it also not H then it is by (13) $H + Hl$ which by (10), (11) and (12) is isomorphic to $\mathbb{O}$, ending the proof.

1.9 Hurwitz and sums of squares

While we haven't mentioned it up to now, the history of especially the octonions is intertwined with a somewhat more number-theoretical sounding problem.

Note that with the absolute value on $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ no other than the regular Euclidean norm on $\mathbb{R}$, $\mathbb{R}^2$, $\mathbb{R}^4$ and $\mathbb{R}^8$ respectively, the squared form of the fundamental relation $|xy| = |x|\,|y|$ is

$$\langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle$$

for the regular Euclidean inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$. If we set $z = xy$ we therefore obtain a relation of the form

$$\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) = \sum_{i=1}^n z_i^2$$

which expresses that the product of two sums of n squares is again a sum of n squares, here for $n = 1$, 2, 4 and 8, and it was in the context of *looking* for a relation of this type for 8 squares that Arthur Cayley rediscovered the octonions in 1845.

As it turns out, this number-theoretical subject and alternative division algebras are not just related but very intimately related.

Definition 1.41. A *normed algebra* V is a real algebra V for which an inner product exists such that $\langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle$ for all $x, y \in V$. We define the norm on V by $\|x\| = \sqrt{\langle x, x\rangle}$.

While this is the usual definition, the reader is cautioned that terminology is not completely stable across authors, with some preferring to call these algebras *absolute-valued* instead.

We ourselves have previously used the adjective *absolute-valued* for a nicely normed (and therefore unital by definition) real $*$-algebra with multiplicative norm and have remarked by way of remark 1.24 that the norm on any nicely normed algebra derives from an inner product

$$\langle x, y\rangle = \mathrm{Re}(xy^*)$$

An absolute-valued algebra is therefore clearly a normed algebra through

$$\langle xy, xy\rangle = |xy|^2 = (|x|\,|y|)^2 = |x|^2\,|y|^2 = \langle x, x\rangle\langle y, y\rangle$$

Specifically $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are normed algebras, and a theorem of Hurwitz that we will now be proving as a corollary to the generalised theorem of Frobenius says that they are up to isomorphism also the *only* finite-dimensional normed unital algebras.

Lemma 1.42. *A finite-dimensional normed algebra $V \neq \{0_V\}$ is a division algebra.*

Proof. Let $x, y \in V$, $y \neq 0_V$. If $xy = 0_V$ then $\langle x, x\rangle\langle y, y\rangle = \langle xy, xy\rangle = 0$. Since $\langle y, y\rangle \neq 0$ we must have $\langle x, x\rangle = 0$ meaning $x = 0_V$. It follows that V is without zero divisors and therefore that it is a division algebra by lemma 1.8.

Lemma 1.43. *A normed unital algebra is alternative.*

Proof. From section 10.1 of Ebbinghaus et al. [12].

If in the primary relation for a normed algebra V, for all $x, y \in V$

$$\langle xy, xy\rangle = \langle x, x\rangle\langle y, y\rangle$$

we replace x by $x + x'$ we obtain using bilinearity and symmetry of the inner product

$$\begin{aligned}\langle (x+x')y, (x+x')y\rangle &= \langle x+x', x+x'\rangle\langle y, y\rangle \\ &= [\langle x, x\rangle + 2\langle x, x'\rangle + \langle x', x'\rangle]\,\langle y, y\rangle \\ &= \langle xy, xy\rangle + 2\langle x, x'\rangle\langle y, y\rangle + \langle x'y, x'y\rangle\end{aligned}$$

whereas using bilinearity and symmetry *only* we have

$$\begin{aligned}\langle (x+x')y, (x+x')y\rangle &= \langle xy + x'y, xy + x'y\rangle \\ &= \langle xy, xy\rangle + 2\langle xy, x'y\rangle + \langle x'y, x'y\rangle\end{aligned}$$

Comparing, we see that $\langle xy, x'y\rangle = \langle x, x'\rangle\langle y, y\rangle$ for all $x, x', y \in V$.

If in this result we now replace y with $y + y'$ we obtain similarly first

$$\begin{aligned}\langle x(y+y'), x'(y+y')\rangle &= \langle x, x'\rangle\langle y+y', y+y'\rangle \\ &= \langle x, x'\rangle\,[\langle y, y\rangle + 2\langle y, y'\rangle + \langle y', y'\rangle] \\ &= \langle xy, x'y\rangle + 2\langle x, x'\rangle\langle y, y'\rangle + \langle xy', x'y'\rangle\end{aligned}$$

and then second

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

so that $\langle xy, x'y'\rangle + \langle xy', x'y\rangle = 2\langle x, x'\rangle\langle y, y'\rangle$ for all $x, x', y, y' \in V$.

Setting first $x' = x$, $y' = z$ and second $x' = z$, $y' = y$ we get, for all $x, y, z \in V$

$$\langle xy, xz\rangle = \langle x, x\rangle\langle y, z\rangle \tag{1.12}$$

$$\langle xy, zy\rangle = \langle x, z\rangle\langle y, y\rangle \tag{1.13}$$

whereas setting third $x' = 1_V$, $y' = z$ and fourth $x' = z$, $y' = 1_V$ we obtain

$$\langle xy, z\rangle + \langle xz, y\rangle = 2\langle x, 1_V\rangle\langle y, z\rangle \tag{1.14}$$

$$\langle xy, z\rangle + \langle x, zy\rangle = 2\langle y, 1_V\rangle\langle x, z\rangle \tag{1.15}$$

If in (1.14) we replace y by xy and in (1.15) x by xy and use (1.12) and (1.13) respectively we obtain

$$\langle x(xy), z\rangle + \langle x, x\rangle\langle y, z\rangle = 2\langle x, 1_V\rangle\langle xy, z\rangle$$

$$\langle (xy)y, z\rangle + \langle x, z\rangle\langle y, y\rangle = 2\langle y, 1_V\rangle\langle xy, z\rangle$$

Rearranging using bilinearity and symmetry we therefore have, for all $x, y, z \in V$

$$\langle x(xy) + \langle x, x\rangle y - 2\langle x, 1_V\rangle xy, z\rangle = 0$$

$$\langle (xy)y + \langle y, y\rangle x - 2\langle y, 1_V\rangle xy, z\rangle = 0$$

and therefore by nondegeneracy, for all $x, y \in V$

$$x(xy) = 2\langle x, 1_V\rangle xy - \langle x, x\rangle y \quad \text{and} \quad (xy)y = 2\langle y, 1_V\rangle xy - \langle y, y\rangle x \tag{1.16}$$

Setting $y = 1_V$ in the first of these and $x = 1_V$ in the second therefore

$$xx = 2\langle x, 1_V\rangle x - \langle x, x\rangle \qquad \text{and} \qquad yy = 2\langle y, 1_V\rangle y - \langle y, y\rangle \tag{1.17}$$

and right multiplying by y respectively left multiplying by x therefore also

$$(xx)y = 2\langle x, 1_V\rangle xy - \langle x, x\rangle y \quad \text{and} \quad x(yy) = 2\langle y, 1_V\rangle xy - \langle y, y\rangle x \tag{1.18}$$

Comparing (1.16) and (1.18) we see the left and right alternative laws emerge, as it was to show.

Theorem 1.44 (Hurwitz)**.** *If $\mathbb{D}$ is a finite-dimensional normed unital algebra, then it is isomorphic to $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{O}$.*

Proof. $\mathbb{D}$ is a finite-dimensional alternative real division algebra by the above two lemmas so that this is now an immediate corollary to the generalised theorem of Frobenius.

Now, returning to the problem of the sums of squares, let us first make things a bit more precise[3]. Suppose the objective is to find all values n for which there exists a bilinear form

$$\Phi = (\Phi_1, \ldots, \Phi_n) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$

such that for all $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$

$$\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) = \sum_{i=1}^n \Phi_i^2(x, y)$$

We already know $n = 1$, 2, 4 and 8 to qualify by letting $\Phi(x, y) = xy$ be the multiplication on $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ respectively.

Now suppose we have a (finite) value n and bilinear form Φ. Let V be the real vector space $V = \bigoplus_{i=1}^n \mathbb{R}$ together with multiplication $xy = \Phi(x, y)$. Trivially

(A1) $x(y + z) = \Phi(x, y + z) = \Phi(x, y) + \Phi(x, z) = xy + xz$

(A2) $(x + y)z = \Phi(x + y, z) = \Phi(x, z) + \Phi(y, z) = xz + yz$

(A3) $(ax)(by) = \Phi(ax, by) = (ab)\Phi(x, y) = (ab)(xy)$

whereby V is a real algebra. It is moreover a normed algebra under the usual inner product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ by

$$\langle xy, xy\rangle = \langle \Phi(x, y), \Phi(x, y)\rangle = \sum_{i=1}^n \Phi_i^2(x, y) = \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) = \langle x, x\rangle\langle y, y\rangle$$

If V were necessarily unital we could now by the theorem of Hurwitz conclude that V is isomorphic to $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{O}$ and specifically that $n = \dim(V)$ can be only 1, 2, 4 or 8, but V is not in fact necessarily unital with respect to this multiplication $xy = \Phi(x, y)$. The same conclusion still holds, however.

To see this, let $y \in V$, $y \neq 0_V$ and $u = \|y\|^{-1} y$ and define linear transformations

$$L\colon V \to V,\ x \mapsto ux \qquad \text{and} \qquad R\colon V \to V,\ x \mapsto xu$$

L and R are norm preserving by $\|L(x)\| = \|ux\| = \|u\|\,\|x\| = \|x\|$ and similarly for R. Any norm preserving linear transformation $T\colon V \to V$ is invertible by

$$T(x) = 0_V \iff \|T(x)\| = 0 \iff \|x\| = 0 \iff x = 0_V$$

so that T is injective and as a linear transformation from a finite-dimensional linear space to itself therefore bijective. T^{-1} is moreover norm preserving by

$$\left\|T^{-1}(x)\right\| = \left\|T(T^{-1}(x))\right\| = \|x\|$$

Now define a new multiplication on V in terms of the old multiplication by

$$x \circ y = R^{-1}(x)L^{-1}(y)$$

[3]We proceed along the lines of section 18.2 of Kantor and Solodovnikov [11].

It is once again trivially verified that V is a real algebra also with the new multiplication by

$$\text{(A1)}\quad x \circ (y+z) = R^{-1}(x)L^{-1}(y+z) = R^{-1}(x)\left[L^{-1}(y) + L^{-1}(z)\right]$$
$$= R^{-1}(x)L^{-1}(y) + R^{-1}(x)L^{-1}(z) = x \circ y + x \circ z$$

$$\text{(A2)}\quad (x+y) \circ z = R^{-1}(x+y)L^{-1}(z) = \left[R^{-1}(x) + R^{-1}(y)\right] L^{-1}(z)$$
$$= R^{-1}(x)L^{-1}(z) + R^{-1}(y)L^{-1}(z) = x \circ z + y \circ z$$

$$\text{(A3)}\quad (ax) \circ (by) = R^{-1}(ax)L^{-1}(by) = (ab)R^{-1}(x)L^{-1}(y) = (ab)(x \circ y)$$

and that it is a normed algebra under the same inner product by

$$\langle x \circ y, x \circ y \rangle = \|x \circ y\|^2 = \left\|R^{-1}(x)L^{-1}(y)\right\|^2$$
$$= \left\|R^{-1}(x)\right\|^2 \left\|L^{-1}(y)\right\|^2 = \|x\|^2 \|y\|^2 = \langle x, x \rangle \langle y, y \rangle$$

It is therefore also a division algebra and we claim V to be unital with respect to the new multiplication with unit u^2. We have after all

$$uL^{-1}(u^2) = L(L^{-1}(u^2)) = u^2 \qquad \text{and} \qquad R^{-1}(u^2)u = R(R^{-1}(u^2)) = u^2$$

so that both $L^{-1}(u^2) = u$ and $R^{-1}(u^2) = u$ by unicity and thereby

$$\text{(A7)}\quad u^2 \circ x = R^{-1}(u^2)L^{-1}(x) = uL^{-1}(x) = L(L^{-1}(x)) = x \quad \text{and}$$
$$x \circ u^2 = R^{-1}(x)L^{-1}(u^2) = R^{-1}(x)u = R(R^{-1}(x)) = x$$

as it was to show.

It follows that to any finite-dimensional normed algebra V there corresponds a normed *unital* algebra of the same dimension so that still $n = \dim(V)$ can be only 1, 2, 4 or 8 by the theorem of Hurwitz, confirming the promised intimate relation between the problem of the sums of squares and the only four finite-dimensional alternative real division algebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$.

Note lastly that although no one-dimensional normed nonunital algebra exists we can by starting with the normed unital algebra $V = \mathbb{C}$, $\mathbb{H}$ or $\mathbb{O}$ easily construct a normed nonunital algebra by equipping it with for example the multiplication

$$x \circ y = x\bar{y}$$

It is again trivially verified that V with this new multiplication and regular absolute value is a normed algebra. It is not unital since if $e \circ x = x = x \circ e$ for an element $e \in V$ and all $x \in V$ then for all $x \in V$

$$x = e \circ x = \overline{x \circ e} = \bar{x}$$

which is definitely not the case. That is, normed nonunital algebras do in fact exist and we can as such not leave out the word *unital* from the formulation of the theorem of Hurwitz.

In the context of the problem of the sums of squares this means that although the problem was proved to be solveable only for $n = 1$, 2, 4 and 8 the solution for $n = 2$, 4 and 8 in terms of the regular multiplication on $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ is not unique.

1.10 Finite-dimensional

The reader will have noticed the specification of finite-dimensionality in the theorems of Frobenius and Hurwitz and while finite-dimensional algebras are all that we are personally interested in given the context of an analogue of complex numbers and, perhaps, complex analysis we should in closing remark on the significance, or rather insignificance, of the phrase.

We have used finite-dimensionality in the first step of the proof of the theorems of Frobenius and here and there in the equivalence of being a finite-dimensional division algebra and being without zero divisors. As it turns out, it is not fundamentally significant though and an advanced result proved independently by Rauol Bott and John Milnor [8] and Michel Kervaire [9] in 1958 enables dropping finite-dimensionality from these theorems all together.

Theorem 1.45 (Bott-Milnor-Kervaire)**.** *A real division algebra can have dimension 1, 2, 4 or 8 only.*

Generally called the $(1, 2, 4, 8)$-Theorem, this result to date resists a purely algebraic proof and is obtained instead as a corollary to a topological property called the parallelizability of the n-sphere.

Although we will not further belabour the point, we note that it is strictly speaking really also only at this point that a remark in the introduction to this chapter that a three-dimensional real division algebra does not exist has been substantiated.

What we have proved rigorously in this first chapter is that $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are up to isomorphism the only finite-dimensional alternative (or normed and unital) real division algebras. In the next chapter we will come to further appreciate the importance of alternativity.

Chapter 2

Beyond complex analysis

2.1 Introduction

While the previous chapter showed that real division algebras such as $\mathbb{R}$ and $\mathbb{C}$ are quite exceptional, we have of course two more of them in $\mathbb{H}$ and $\mathbb{O}$ and with the original goal in all of this having been not only an analogue of complex numbers but explicitly also of complex analysis we may therefore still entertain a notion of a quaternionic or even octonionic analysis. We will however in this chapter find either to be an only marginally viable concept.

In constructing any sort of analogue of real and complex analysis, having available the concept of *differentiability* would seem to be the first order of business and given its definition in both the real and complex case in terms of existence of a limit

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

we will first of all want to take a look at this concept of *limit*. We let in this chapter $\mathbb{D}$ stand for either $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ or $\mathbb{O}$ and consider the identification of $\mathbb{R}$ with $\mathbb{R}1_{\mathbb{D}}$ to be implicit.

2.2 Limits and continuity

Recall from real analysis that for $D \subseteq \mathbb{R}$ an open interval, $\xi \in D$, $f \colon D \to \mathbb{R}$ a real-valued function of a real variable and $y \in \mathbb{R}$ we take the symbolism

$$\lim_{x \to \xi} f(x) = y$$

to mean that for all $\varepsilon \in \mathbb{R},\ \varepsilon > 0$ there exists a $\delta \in \mathbb{R}, \delta > 0$ such that

$$x \in D,\ 0 < |x - \xi| < \delta \implies |f(x) - y| < \varepsilon$$

whereas we call $\lim_{x \to \xi} f(x)$ non-existent if no such $y \in \mathbb{R}$ exists.

The same symbolism features unchanged in complex analysis for a domain (a non-empty, open, connected set) $D \subseteq \mathbb{C}$, $f\colon D \to \mathbb{C}$ a complex-valued function of a complex variable and a value $y \in \mathbb{C}$ as long as we understand that $|\cdot|$ then signifies the complex modulus, and we desire to extend this concept of limit to a quaternion-valued function of a quaternionic variable and an octonion-valued function of an octonionic variable as well.

We already have available a well-behaved quaternionic and octonionic modulus (by design, as we've seen in the previous chapter) and can therefore simply adopt the concept unchanged as we do in the complex case but our first stop will need to be reviewing the proof of the following foundational theorem for issues of commutativity and associativity.

Theorem 2.1. *Let $\xi \in D$ and $f, g\colon D \to \mathbb{D}$ be functions such that*

$$\lim_{x\to\xi} f(x) = y \qquad \text{and} \qquad \lim_{x\to\xi} g(x) = z$$

for certain $y, z \in \mathbb{D}$. Then we have

$$(i) \quad \lim_{x\to\xi} (f(x) + g(x)) = y + z \qquad (ii) \quad \lim_{x\to\xi} (f(x)g(x)) = yz$$

$$(iii) \quad \lim_{x\to\xi} (g(x)^{-1} f(x)) = z^{-1}y \qquad (iv) \quad \lim_{x\to\xi} (f(x)g(x)^{-1}) = yz^{-1}$$

provided, in cases (iii) and (iv), that $z \neq 0$.

Proof of (i). We write

$$f(x) + g(x) - (y + z) = (f(x) - y) + (g(x) - z) \tag{2.1}$$

Let $\varepsilon > 0$ be given and $\delta_1 > 0$ and $\delta_2 > 0$ be such that

$$x \in D,\ 0 < |x - \xi| < \delta_1 \implies |f(x) - y| < \varepsilon/2 \quad \text{and} \tag{2.2}$$

$$x \in D,\ 0 < |x - \xi| < \delta_2 \implies |g(x) - z| < \varepsilon/2 \tag{2.3}$$

Then for $x \in D,\ 0 < |x - \xi| < \delta = \min(\delta_1, \delta_2)$ we have

$$\begin{aligned} |f(x) + g(x) - (y + z)| &= |(f(x) - y) + (g(x) - z)| \\ &\leq |f(x) - y| + |g(x) - z| < \varepsilon/2 + \varepsilon/2 = \varepsilon \end{aligned}$$

Proof of (ii). Using commutativity with $-1 \in \mathbb{R}$ only we write

$$f(x)g(x) - yz = f(x)(g(x) - z) + (f(x) - y)z. \tag{2.4}$$

Let $\delta_1 > 0$ be such that

$$x \in D,\ 0 < |x - \xi| < \delta_1 \implies |f(x) - y| < 1$$

Since $|f(x)| = |f(x) - y + y| \leq |f(x) - y| + |y|$ we then also have

$$x \in D,\ 0 < |x - \xi| < \delta_1 \implies |f(x)| < 1 + |y| \tag{2.5}$$

Let $\varepsilon > 0$ be given, $\delta_2 > 0$ be such that

$$x \in D,\ 0 < |x - \xi| < \delta_2 \implies |g(x) - z| < \frac{\varepsilon}{2(1 + |y|)} \tag{2.6}$$

and $\delta_3 > 0$ be such that, allowing for $z = 0$ through a random $\delta_3 > 0$,

$$x \in D,\ 0 < |x - \xi| < \delta_3 \implies |f(x) - y|\,|z| < \frac{\varepsilon}{2} \tag{2.7}$$

Then for $x \in D,\ 0 < |x - \xi| < \delta = \min(\delta_1, \delta_2, \delta_3)$ we have

$$\begin{aligned} |f(x)g(x) - yz| &= |f(x)(g(x) - z) + (f(x) - y)z| \\ &\le |f(x)|\,|g(x) - z| + |f(x) - y|\,|z| < (1 + |y|)\frac{\varepsilon}{2(1 + |y|)} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

These standard proofs of the first and second statements, then, show themselves to be unexcitingly generic but the standard proof of the third and fourth statements does in fact require slight vigilance. Before continuing we will therefore now first state a basic result on behalf of the octonionic case.

Lemma 2.2. *The left and right* alternative laws *satisfied by all* $x, y \in \mathbb{D}$

$$x(xy) = (xx)y \qquad \textit{and} \qquad (xy)y = x(yy) \tag{2.8}$$

are equivalent to, respectively, the left and right cancellation laws

$$x^{-1}(xy) = y \qquad \textit{and} \qquad (xy)y^{-1} = x \tag{2.9}$$

for $x, y \in \mathbb{D}$, $x \neq 0$ *respectively* $y \neq 0$.

Proof. We will prove that the alternative laws (2.8) are equivalent to

$$\bar{x}(xy) = |x|^2\, y \qquad \text{and} \qquad (xy)\bar{y} = x\,|y|^2 \tag{2.10}$$

after which setting $x^{-1} = |x|^{-2}\,\bar{x}$ for $x \neq 0$ respectively $y^{-1} = \bar{y}\,|y|^{-2}$ for $y \neq 0$ and observing that (2.8) is automatically satisfied when either $x = 0$ or $y = 0$ proves the lemma.

Proof of left part. Setting $\lambda = 2\,\mathrm{Re}\,x = x + \bar{x} \in \mathbb{R}$ we write

$$\bar{x}(xy) = (\lambda - x)(xy) = \lambda(xy) - x(xy) = (\lambda x)y - x(xy) \tag{2.11}$$

whereas on the other hand we have

$$|x|^2\, y = (\bar{x}x)y = ((\lambda - x)x)y = (\lambda x - xx)y = (\lambda x)y - (xx)y \tag{2.12}$$

Assuming the left part of (2.8) we make equal the right-hand sides of these thereby proving the left part of (2.10) and conversely assuming the left part of (2.10) we make equal the left-hand sides thereby proving the left part of (2.8).

Proof of right part. Analogous by setting $\lambda = 2\,\mathrm{Re}\,y = y + \bar{y} \in \mathbb{R}$.

Armed with this lemma we will now continue the proof of Theorem 2.1.

Proof of (iii) and (iv). We will in fact prove the special case statement

$$\lim_{x \to \xi} g(x)^{-1} = z^{-1} \tag{2.13}$$

after which the second statement proves both the third and fourth.

Proof of special case statement. Using our just proved lemma in the first step and commutativity with $-1 \in \mathbb{R}$ in the third we write

$$\begin{aligned} g(x)^{-1} - z^{-1} &= \left[g(x)^{-1}z\right] z^{-1} - z^{-1} \\ &= \left[g(x)^{-1}z - 1\right] z^{-1} = \left[g(x)^{-1}(z - g(x))\right] z^{-1} \end{aligned} \tag{2.14}$$

Noting that we are assuming $z \neq 0$, we can let $\delta_1 > 0$ be such that

$$x \in D,\ 0 < |x - \xi| < \delta_1 \implies |g(x) - z| < \frac{|z|}{2}$$

Since $|g(x)| \leq \frac{|z|}{2}$ would lead to the impossible inequality

$$|z| = |z - g(x) + g(x)| \leq |z - g(x)| + |g(x)| < \frac{|z|}{2} + \frac{|z|}{2} = |z|$$

it follows that we then have $|g(x)| > \frac{|z|}{2}$ and therefore

$$x \in D,\ 0 < |x - \xi| < \delta_1 \implies \frac{1}{|g(x)|} < \frac{2}{|z|} \tag{2.15}$$

Now let $\varepsilon > 0$ be given and $\delta_2 > 0$ be such that

$$x \in D,\ 0 < |x - \xi| < \delta_2 \implies |g(x) - z| < \frac{|z|^2 \varepsilon}{2} \tag{2.16}$$

Then for $x \in D,\ 0 < |x - \xi| < \delta = \min(\delta_1, \delta_2)$ we have

$$\begin{aligned} \left|g(x)^{-1} - z^{-1}\right| &= \left|g(x)^{-1}(z - g(x))\right| \left|z^{-1}\right| = \left|g(x)^{-1}\right| |z - g(x)| \left|z^{-1}\right| \\ &= \frac{1}{|g(x)|}\ |g(x) - z|\ \frac{1}{|z|}\ <\ \frac{2}{|z|}\ \frac{|z|^2 \varepsilon}{2}\ \frac{1}{|z|} = \varepsilon \end{aligned}$$

proving the special case statement (2.13) and thereby the third and fourth statements of Theorem 2.1.

Corollary 2.3. *Let $\xi \in D$ and $f\colon D \to \mathbb{D}$ be a function such that*

$$\lim_{x \to \xi} f(x) = y$$

for a certain $y \in \mathbb{D}$. Then for any $c \in \mathbb{D}$ we have

$$\lim_{x \to \xi} cf(x) = cy \qquad \text{and} \qquad \lim_{x \to \xi} f(x)c = yc$$

In particular $\lim_{x \to \xi} -f(x) = -y$.

Proof. Follows directly from the second statement of Theorem 2.1 by noting that a constant function $f\colon D \to \mathbb{D},\ x \mapsto c$ has $\lim_{x \to \xi} f(x) = c$.

Alternativity has thus equipped us with a functional limit and, same as in the real case, we therefore immediately get the important analytic concept of *continuity* for free.

Definition 2.4. A function $f\colon D \to \mathbb{D}$ is said to be *continuous* at $\xi \in D$ if

$$\lim_{x \to \xi} f(x) = f(\xi)$$

It is said to be *continuous on* D or simply *continuous* if it is continuous at all $\xi \in D$.

Theorem 2.5. *Let $f, g\colon D \to \mathbb{D}$ be functions continuous at $\xi \in D$. Then so are the functions*

$$\begin{array}{llll} (i) & x \mapsto f(x) + g(x) & (ii) & x \mapsto f(x)g(x) \\ (iii) & x \mapsto g(x)^{-1} f(x) & (iv) & x \mapsto f(x)g(x)^{-1} \end{array}$$

provided, in cases (*iii*) *and* (*iv*), *that* $g(\xi) \neq 0$.

Proof. Follows immediately from Theorem 2.1.

By noting that a constant function $x \mapsto c$ and the identity function $x \mapsto x$ are both trivially continuous this last theorem already allows us to recursively construct a few interesting classes of continuous functions, from monomials to polynomials to rational functions.

Of more current interest to us though is that our limit has paved the way for the introduction of the sought after concept of differentiability.

2.3 Differentiability

Definition 2.6. A function $f\colon D \to \mathbb{D}$ is said to be *left differentiable* at $x \in D$ if the limit

$$\frac{d}{dx} f = \lim_{h \to 0} h^{-1}(f(x+h) - f(x)) \tag{2.17}$$

exists and is said to be *right differentiable* at x if the limit

$$f \frac{d}{dx} = \lim_{h \to 0} (f(x+h) - f(x))h^{-1} \tag{2.18}$$

exists. It is said to be *left (right) differentiable on* D or simply *left (right) differentiable* if it is left (right) differentiable at all $x \in D$.

These notions of left and right differentiability obviously coincide with each other and agree with the standard definition for the commutative $\mathbb{R}$ and $\mathbb{C}$, in which cases we call them simply *differentiability*, whereas they as we shall see need not coincide for the noncommutative $\mathbb{H}$ and $\mathbb{O}$.

One way in which they *are* always equivalent is that either sided differentiability implies continuity. Of some interest here is that the use of the cancellation laws once more emphasizes the role of the alternativity of $\mathbb{O}$.

Theorem 2.7. *If a function $f\colon D \to \mathbb{D}$ is either left or right differentiable at $\xi \in D$ it is continuous at ξ.*

Proof. Noting use of the left and right cancellation laws respectively we have in the case of left differentiability

$$\lim_{x\to\xi} f(x) = \lim_{h\to 0} f(\xi+h) = \lim_{h\to 0} \left[f(\xi) + h\left(h^{-1}\left[f(\xi+h) - f(\xi)\right]\right)\right]$$
$$= f(\xi) + \lim_{h\to 0} h \cdot \lim_{h\to 0} h^{-1}\left[f(\xi+h) - f(\xi)\right] = f(\xi) + 0 \cdot \tfrac{d}{d\xi} f = f(\xi)$$

and similarly in the case of right differentiability

$$\lim_{x\to\xi} f(x) = \lim_{h\to 0} f(\xi+h) = \lim_{h\to 0} \left[f(\xi) + \left(\left[f(\xi+h) - f(\xi)\right] h^{-1}\right) h\right]$$
$$= f(\xi) + \lim_{h\to 0} \left[f(\xi+h) - f(\xi)\right] h^{-1} \cdot \lim_{h\to 0} h = f(\xi) + f \tfrac{d}{d\xi} \cdot 0 = f(\xi)$$

Be that as it may, at this point and given our functional concepts of limits and continuity, including continuous functions, what our concept of differentiability now needs are differentiable functions.

In real analysis the two directions only in which h can approach 0 (from the left and from the right) lets differentiability amount to no more than a mild smoothness condition but from complex analysis we recall that the additional dimension available from the complex plane already leads to the significantly sterner condition of having to satisfy a pair of partial differential equations known as the *Cauchy-Riemann equations*.

With the quaternions and octonions making available another two respectively six dimensions we may as such certainly expect to experience further restrictions on the class of differentiable functions — but what we needn't immediately expect is the following result, first pioneered by G. Scheffers [18] in 1893 and proved by N.M. Krylov [19] in 1947 and more generally by his student A.S. Meilikhson [20] a year later.

Theorem 2.8 (Meilikhson)**.** *Let $D \subseteq \mathbb{H}$ be a domain. A function $f\colon D \to \mathbb{H}$ is left (right) differentiable on D if and only if it is of the form*

$$f(q) = a + qb \qquad\qquad (f(q) = a + bq) \tag{2.19}$$

on D for some constants $a, b \in \mathbb{H}$.

That is, while the usefulness of complex analysis is testament to the fact that complex differentiability does not in fact overly restrict the class of complex differentiable functions, it appears that already in the quaternionic case we are left with nothing but linear functions and, hence, with little more than an academic exercise.

M.S. Marinov [29] proves the same result for the octonionic case but we will limit ourselves to the quaternions in what follows. The octonionic proof is analogous to the quaternionic one and moreover, given Meilikhson's result and the embedding of $\mathbb{H}$ in $\mathbb{O}$ this same result is of course the best one could hope for in the octonionic case anyway, while having only linear differentiable functions available is already quite a bit too restrictive for our tastes.

Specifically noteworthy though is the fact that the octonionic case does not *further* restrict the class of differentiable functions, again pointing to alternativity rather than associativity as the important concept in all of this.

2.4 Meilikhson

We first of all quickly note that the left- and right-linear functions (2.19) do, for a general domain D, denote two different classes of functions. Specifically, let $D = \mathbb{H}$ and assume that both

$$f(q) = a + qb \qquad \text{and} \qquad f(q) = c + dq$$

for constants $a, b, c, d \in \mathbb{H}$. By firstly setting $q = 0$ we obtain $c = a$ and by secondly setting $q = 1$ and subtracting we get $d = b$. Comparing the general forms we then have $qb = bq$ for all $q \in \mathbb{H}$ and therefore $b \in \mathrm{Cen}(\mathbb{H}) = \mathbb{R}$ which shows us to be looking at a subset of linear functions only.

Secondly we note that the *if* part is obvious since if $f\colon D \to \mathbb{H}$ is of the form

$$f(q) = a + qb \qquad \text{or} \qquad f(q) = a + bq$$

then we have respectively

$$\frac{d}{dq} f = \lim_{h\to 0} \; h^{-1}(f(q+h) - f(q)) = \lim_{h\to 0} \; h^{-1}(hb) = b$$
$$f\frac{d}{dq} = \lim_{h\to 0} \; (f(q+h) - f(q))h^{-1} = \lim_{h\to 0} \; (bh)h^{-1} = b$$

confirming left respectively right differentiability at any $q \in D$.

It is the *only if* part that will formally need some work but the idea is in fact quite simple and the result was in that sense already known to Hamilton himself.

Specifically, take even the simplest quadratic function $f(q) = q^2$. Then

$$\begin{aligned}\frac{d}{dq} f &= \lim_{h\to 0} h^{-1}((q+h)(q+h) - q^2) = \lim_{h\to 0} h^{-1}(qh + hq + h^2)\\ &= \lim_{h\to 0} h^{-1}qh + q + \lim_{h\to 0} h = q + \lim_{h\to 0} h^{-1}qh\end{aligned}$$

and similarly $f\frac{d}{dq} = q + \lim_{h\to 0} \; hqh^{-1}$.

Existence of these limits means they must be independent of the path that h takes towards 0 but if we set $q = t + ix + jy + kz$ and let $h = \Delta t \to 0$ along the real axis we obtain

$$\frac{d}{dq} f = f\frac{d}{dq} = q + \lim_{\Delta t\to 0} \frac{\Delta t}{\Delta t} q = q + q = 2q$$

whereas letting $h = i \cdot \Delta x \to 0$ along the i axis we have

$$\frac{d}{dq} f = f\frac{d}{dq} = q + \lim_{\Delta x\to 0} -\frac{\Delta x}{\Delta x} iqi = q - iqi$$

and in the same way for $h = j \cdot \Delta y \to 0$ along the j axis and $h = k \cdot \Delta z \to 0$ along the k axis

$$\frac{d}{dq} f = f\frac{d}{dq} = q - jqj \qquad\qquad \frac{d}{dq} f = f\frac{d}{dq} = q - kqk$$

Comparing these results, we see we must have $q = -iqi = -jqj = -kqk$ and thereby $q = \frac{1}{4}(q - iqi - jqj - kqk) = t \in \mathbb{R}$.

However, D can as a non-empty subset of $\mathbb{H}$ not be a subset of $\mathbb{R}$ (any ball around any point $q \in \mathbb{R}$ leaves $\mathbb{R}$) proving that f is neither left nor right differentiable on all of D.

This same directional limit argument is at the heart of the general proof as well, and in the same way as it is at the heart of the classic Cauchy-Riemann equations. We will therefore now quickly review those to prepare ourselves for the quaternionic case.

Let $D \subseteq \mathbb{C}$ be a domain and $f\colon D \to \mathbb{C}$ a complex-valued function of a complex variable. We have defined f to be differentiable at $z \in D$ if the limit

$$f'(z) = \frac{df}{dz} = \frac{d}{dz}f = f\frac{d}{dz} = \lim_{h\to 0}\frac{f(z+h) - f(z)}{h}$$

exists which means it must be independent of the path that h takes towards 0. Writing $z = x + iy$ and $f(z) = u(x, y) + i\,v(x, y)$ where u and v are two real-valued functions of two real variables we firstly let $h = \Delta x \to 0$ along the real axis to obtain

$$\frac{df}{dz} = \lim_{\Delta x\to 0}\frac{u(x+\Delta x, y) - u(x, y)}{\Delta x} + i\lim_{\Delta x\to 0}\frac{v(x+\Delta x, y) - v(x, y)}{\Delta x} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}$$

and secondly let $h = i\cdot\Delta y \to 0$ along the imaginary axis to obtain

$$\begin{aligned}\frac{df}{dz} &= \lim_{\Delta y\to 0}\frac{u(x, y+\Delta y) - u(x, y)}{i\cdot\Delta y} + i\lim_{\Delta y\to 0}\frac{v(x, y+\Delta y) - v(x, y)}{i\cdot\Delta y}\\ &= \lim_{\Delta y\to 0}\frac{v(x, y+\Delta y) - v(x, y)}{\Delta y} - i\lim_{\Delta y\to 0}\frac{u(x, y+\Delta y) - u(x, y)}{\Delta y} = \frac{\partial v}{\partial y} - i\frac{\partial u}{\partial y}\end{aligned}$$

where in both cases we understand the partial derivatives to be evaluated at the point $(x, y) \in \mathbb{R}^2$. Comparing the real and imaginary parts of these expressions we then immediately see the Cauchy-Riemann equations emerge as a condition for complex differentiability of f at z:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \qquad \text{and} \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{2.20}$$

or, notationally identifying f with the complex-valued function of two real variables $(x, y) \mapsto f(x + iy)$ more compactly

$$\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y} = 0 \tag{2.21}$$

It is a basic fact from complex analysis that if we assume continuity of these partial derivatives, (2.20) or (2.21) is not only necessary but in fact also sufficient for differentiability of f at z.

Note for later use that the argument also shows that, given (2.21)

$$\frac{df}{dz} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right) \tag{2.22}$$

Generalizing to the quaternionic case, we have defined the quaternion-valued function of a quaternionic variable $f\colon D \to \mathbb{H}$ to be left (right) differentiable at $q \in D \subseteq \mathbb{H}$ if the limit

$$\frac{d}{dq}f = \lim_{h\to 0} h^{-1}(f(q+h) - f(q)) \qquad \left(f\frac{d}{dq} = \lim_{h\to 0}(f(q+h) - f(q))h^{-1}\right)$$

exists. Writing $q = t + ix + jy + kz$ and

$$f(q) = s(t,x,y,z) + i\,u(t,x,y,z) + j\,v(t,x,y,z) + k\,w(t,x,y,z)$$

where s, u, v and w are now four real-valued functions of four real variables we proceed in the same way again and first let $h = \Delta t \to 0$ along the real axis, second $h = i \cdot \Delta x \to 0$ along the i-axis, third $h = j \cdot \Delta y \to 0$ along the j axis and finally $h = k \cdot \Delta z \to 0$ along the k-axis to straightforwardly obtain

$$\begin{aligned}
\frac{d}{dq}f &= \frac{\partial s}{\partial t} + i\frac{\partial u}{\partial t} + j\frac{\partial v}{\partial t} + k\frac{\partial w}{\partial t}\\
\frac{d}{dq}f &= \frac{\partial u}{\partial x} - i\frac{\partial s}{\partial x} + j\frac{\partial w}{\partial x} - k\frac{\partial v}{\partial x}\\
\frac{d}{dq}f &= \frac{\partial v}{\partial y} - i\frac{\partial w}{\partial y} - j\frac{\partial s}{\partial y} + k\frac{\partial u}{\partial y}\\
\frac{d}{dq}f &= \frac{\partial w}{\partial z} + i\frac{\partial v}{\partial z} - j\frac{\partial u}{\partial z} - k\frac{\partial s}{\partial z}
\end{aligned}
\qquad
\left(\begin{aligned}
f\frac{d}{dq} &= \frac{\partial s}{\partial t} + i\frac{\partial u}{\partial t} + j\frac{\partial v}{\partial t} + k\frac{\partial w}{\partial t}\\
f\frac{d}{dq} &= \frac{\partial u}{\partial x} - i\frac{\partial s}{\partial x} - j\frac{\partial w}{\partial x} + k\frac{\partial v}{\partial x}\\
f\frac{d}{dq} &= \frac{\partial v}{\partial y} + i\frac{\partial w}{\partial y} - j\frac{\partial s}{\partial y} - k\frac{\partial u}{\partial y}\\
f\frac{d}{dq} &= \frac{\partial w}{\partial z} - i\frac{\partial v}{\partial z} + j\frac{\partial u}{\partial z} - k\frac{\partial s}{\partial z}
\end{aligned}\right)$$

where we of course now understand these partial derivatives to be evaluated at the point $(t,x,y,z) \in \mathbb{R}^4$. By comparing parts we then arrive at the twelve partial differential equations that will together be referred to as *the left (right) quaternionic Cauchy-Riemann equations.*

Theorem 2.9. *Let $D \subseteq \mathbb{H}$ be a domain, $q = t+ix+jy+kz \in D$ and $f\colon D \to \mathbb{H}$ the quaternion-valued function of a quaternionic variable*

$$f(q) = s(t,x,y,z) + i\,u(t,x,y,z) + j\,v(t,x,y,z) + k\,w(t,x,y,z)$$

in which s, u, v and w are four real-valued functions of four real variables. If f is left (right) differentiable at q, the left (right) quaternionic Cauchy-Riemann equations

$$\begin{aligned}
\frac{\partial s}{\partial t} &= \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = \frac{\partial w}{\partial z}\\
\frac{\partial s}{\partial x} &=-\frac{\partial u}{\partial t} =-\frac{\partial v}{\partial z} = \frac{\partial w}{\partial y}\\
\frac{\partial s}{\partial y} &= \frac{\partial u}{\partial z} =-\frac{\partial v}{\partial t} =-\frac{\partial w}{\partial x}\\
\frac{\partial s}{\partial z} &=-\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x} =-\frac{\partial w}{\partial t}
\end{aligned}
\qquad
\left(\begin{aligned}
\frac{\partial s}{\partial t} &= \frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = \frac{\partial w}{\partial z}\\
\frac{\partial s}{\partial x} &=-\frac{\partial u}{\partial t} = \frac{\partial v}{\partial z} =-\frac{\partial w}{\partial y}\\
\frac{\partial s}{\partial y} &=-\frac{\partial u}{\partial z} =-\frac{\partial v}{\partial t} = \frac{\partial w}{\partial x}\\
\frac{\partial s}{\partial z} &= \frac{\partial u}{\partial y} =-\frac{\partial v}{\partial x} =-\frac{\partial w}{\partial t}
\end{aligned}\right)$$

are satisfied at the point $(t,x,y,z) \in \mathbb{R}^4$.

Proof. The rows in these representations are nothing but the columns of the expressions above.

While the reader will have to compensate for the reuse of symbols, we of course recognize the classic Cauchy-Riemann equations (2.20) in the upper-left quadrant of either left or right quaternionic Cauchy-Riemann equations in the representations above, in accordance with the embedding of $\mathbb{C}$ in $\mathbb{H}$ and the equivalence of left and right differentiability in this subset.

Like before in the complex case, notationally identifying f with the quaternion-valued function of four real variables $(t, x, y, z) \mapsto f(t + ix + jy + kz)$, we may condense the equations into the equivalent but rather more memorable form, for the left version

$$\frac{\partial f}{\partial t} + i\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial f}{\partial t} + j\frac{\partial f}{\partial y} = 0 \qquad \frac{\partial f}{\partial t} + k\frac{\partial f}{\partial z} = 0 \tag{2.23}$$

and for the right version

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}i = 0 \qquad \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}j = 0 \qquad \frac{\partial f}{\partial t} + \frac{\partial f}{\partial z}k = 0 \tag{2.24}$$

and we will be using them in this form, albeit to immediately split them into complex equations again.

The proof that we will present is due to A. Sudbery [26] and uses a few basic results from the theory of complex-valued functions of several complex-variables, a theory not widely known at the undergraduate level. The interested reader unacquainted with it is referred to any one of the many available sources on the subject such as H. Cartan [25] or S.G. Krantz [27]; here, we will simply state the few results that we need directly.

First a few basics from regular complex analysis though.

Definition 2.10. Let $D \subseteq \mathbb{C}$ be a non-empty open set. A function $f : D \to \mathbb{C}$ is said to be *analytic* at $z \in D$ if it is differentiable throughout some neighbourhood

$$B_\varepsilon(z) = \big\{\zeta \in \mathbb{C} \;\big|\; |z - \zeta| < \varepsilon\big\} \subseteq D$$

of z and is said to *analytic on D* or simply *analytic* if it is analytic at all $z \in D$.

Theorem 2.11. *Let $D \subseteq \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ an analytic function. Then f has continuous derivatives of all orders on D.*

Theorem 2.12. *Let $D \subseteq \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ an analytic function. If $f'(z) = 0$ everywhere on D, then $f(z)$ is constant on D.*

Corollary 2.13. *Let $D \subseteq \mathbb{C}$ be a domain and $f : D \to \mathbb{C}$ an analytic function. If $f''(z) = 0$ everywhere on D, then $f(z)$ is linear on D.*

This next theorem is in fact a corollary to a non-trivial result from the theory of several complex variables known as *Hartogs' theorem* (on separate analyticity) but for our purposes we may consider it to simply define analyticity of a function of several complex variables.

Theorem 2.14. *Let $D \subseteq \mathbb{C}^n$ be a domain. A function $f : D \to \mathbb{C}$ is analytic if and only if it is analytic in each variable separately.*

By analytic in each variable separately we mean that for each $j = 1, \ldots, n$ and for each fixed set of points $z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_n$, the complex-valued function of a single complex variable

$$\zeta \mapsto f(z_1, \ldots, z_{j-1}, \zeta, z_{j+1}, \ldots, z_n)$$

is analytic on the set

$$D(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_n) = \big\{\zeta \in \mathbb{C} \;\big|\; (z_1, \ldots, z_{j-1}, \zeta, z_{j+1}, \ldots, z_n) \in D\big\}$$

Note that with D open as a subset of $\mathbb{C}^n$, $D(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_n)$ is obviously open as a subset of $\mathbb{C}$ and that if D is *convex* $D(z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_n)$ also is, it being the intersection of the convex set D with a copy of the convex complex plane (alternatively, by it being the kind of straight line by which convexity is *defined*). It is therefore then certainly connected, and thereby a domain as a subset of $\mathbb{C}$.

We will now first introduce a bit of notation. Let $D \subseteq \mathbb{C}^n$ be a non-empty open set, $z = x + iy \in D$. Writing $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, we denote for a function $f \colon D \to \mathbb{C}$

$$\frac{\partial f}{\partial z_j} = \frac{1}{2}\left(\frac{\partial f}{\partial x_j} - i\frac{\partial f}{\partial y_j}\right) \qquad\qquad \frac{\partial f}{\partial \bar{z}_j} = \frac{1}{2}\left(\frac{\partial f}{\partial x_j} + i\frac{\partial f}{\partial y_j}\right)$$

and, understanding the real partial derivatives to be evaluated at the point $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, call these expressions the *complex partial derivative* respectively *complex conjugate partial derivative* of f with respect to z_j at z.

This formulation may appear slightly puzzling at first glance but is in fact to a large extent dictated by notational consistency. Specifically, if we take just $n=1$ and pretend that z and $\bar{z}$ are normal, independent variables, we set

$$x(z, \bar{z}) = \frac{z + \bar{z}}{2} \qquad\qquad y(z, \bar{z}) = \frac{z - \bar{z}}{2i}$$

to identify f with the function $(z, \bar{z}) \mapsto f(x(z, \bar{z}) + iy(z, \bar{z}))$ and write using the regular two dimensional chain rule familiar from real analysis

$$\frac{\partial f}{\partial z} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial z} = \frac{\partial f}{\partial x}\frac{\partial}{\partial z}\left(\frac{z + \bar{z}}{2}\right) + \frac{\partial f}{\partial y}\frac{\partial}{\partial z}\left(\frac{z - \bar{z}}{2i}\right) = \frac{1}{2}\left(\frac{\partial f}{\partial x} - i\frac{\partial f}{\partial y}\right)$$
$$\frac{\partial f}{\partial \bar{z}} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \bar{z}} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \bar{z}} = \frac{\partial f}{\partial x}\frac{\partial}{\partial \bar{z}}\left(\frac{z + \bar{z}}{2}\right) + \frac{\partial f}{\partial y}\frac{\partial}{\partial \bar{z}}\left(\frac{z - \bar{z}}{2i}\right) = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right)$$

These so defined complex partial derivatives behave in all the expected ways but do note that since z and $\bar{z}$ are obviously anything *but* independent variables they are strictly speaking still to be considered notational devices only.

Using this notation, classic Cauchy-Riemann (2.21) and (2.22) take the particularly compact form

$$\frac{\partial f}{\partial \bar{z}} = 0 \qquad\qquad \text{and} \qquad\qquad \frac{df}{dz} = \frac{\partial f}{\partial z} \tag{2.25}$$

and we will be using this formulation in the upcoming proof.

Theorem 2.15. *Let $D \subseteq \mathbb{C}^n$ be a domain. An analytic function $f\colon D \to \mathbb{C}$ has continuous complex partial derivatives of all orders throughout D.*

This theorem parallels theorem 2.11 from the single variable case. Note that the continuity of course includes the conjugate complex partial derivatives; one is continuous if and only if the other is. Mixed continuous complex partial derivatives may, moreover, be applied in any order same as in real analysis.

Finally, we have available the important uniqueness principle, unaltered from the case of a single complex variable.

Theorem 2.16. *Let $D \subseteq \mathbb{C}^n$ be a domain, $U \subseteq D$ a non-empty open set and $f, g\colon D \to \mathbb{C}$ analytic functions. If $f(z) = g(z)$ for all $z \in U$ then $f(z) = g(z)$ for all $z \in D$.*

We are now ready for the proof of Meilikhson's theorem. We showed the *if* part to be satisfied already at the start of the section and, having given both left and right versions of the quaternionic Cauchy-Riemann equations above, we further prove directly the left version of the theorem only, rendering the right version to obvious analogy. That is, we prove the following.

Theorem 2.17 (Meilikhson)**.** *If a function $f\colon D \to \mathbb{H}$ is left differentiable on D, it is of the form $f(q) = a + qb$ on D for some constants $a, b \in \mathbb{H}$.*

Proof. Let $q = t + ix + jy + kz \in D$. f is left differentiable at q meaning that the quaternionic Cauchy-Riemann equations (2.23) hold at the point (t, x, y, z).

Identifying $\mathbb{H}$ with $\mathbb{C}^2$, we set $v = t + ix \in \mathbb{C}$ and $w = y - iz \in \mathbb{C}$ to write $q = v + jw$ and $f(q) = g(v, w) + j\,h(v, w)$ where g and h are complex-valued functions of two complex variables. Under the usual notational identifications we write

$$\frac{\partial f}{\partial t} = \frac{\partial g}{\partial t} + j\frac{\partial h}{\partial t} \qquad \frac{\partial f}{\partial x} = \frac{\partial g}{\partial x} + j\frac{\partial h}{\partial x} \qquad \frac{\partial f}{\partial y} = \frac{\partial g}{\partial y} + j\frac{\partial h}{\partial y} \qquad \frac{\partial f}{\partial z} = \frac{\partial g}{\partial z} + j\frac{\partial h}{\partial z}$$

and thereby

$$\begin{aligned}
\frac{\partial f}{\partial t} + i\frac{\partial f}{\partial x} &= \left(\frac{\partial g}{\partial t} + j\frac{\partial h}{\partial t}\right) + i\left(\frac{\partial g}{\partial x} + j\frac{\partial h}{\partial x}\right) = \left(\frac{\partial g}{\partial t} + i\frac{\partial g}{\partial x}\right) + j\left(\frac{\partial h}{\partial t} - i\frac{\partial h}{\partial x}\right)\\
\frac{\partial f}{\partial t} + j\frac{\partial f}{\partial y} &= \left(\frac{\partial g}{\partial t} + j\frac{\partial h}{\partial t}\right) + j\left(\frac{\partial g}{\partial y} + j\frac{\partial h}{\partial y}\right) = \left(\frac{\partial g}{\partial t} - \frac{\partial h}{\partial y}\right) + j\left(\frac{\partial h}{\partial t} + \frac{\partial g}{\partial y}\right)\\
\frac{\partial f}{\partial t} + k\frac{\partial f}{\partial z} &= \left(\frac{\partial g}{\partial t} + j\frac{\partial h}{\partial t}\right) + k\left(\frac{\partial g}{\partial z} + j\frac{\partial h}{\partial z}\right) = \left(\frac{\partial g}{\partial t} - i\frac{\partial h}{\partial z}\right) + j\left(\frac{\partial h}{\partial t} - i\frac{\partial g}{\partial z}\right)
\end{aligned}$$

which splits the three quaternionic equations (2.23) into the six complex equations

$$\begin{aligned}
&\frac{\partial g}{\partial t} + i\frac{\partial g}{\partial x} = 0 \qquad &&\frac{\partial g}{\partial t} - \frac{\partial h}{\partial y} = 0 \qquad &&\frac{\partial g}{\partial t} - i\frac{\partial h}{\partial z} = 0\\
&\frac{\partial h}{\partial t} - i\frac{\partial h}{\partial x} = 0 \qquad &&\frac{\partial h}{\partial t} + \frac{\partial g}{\partial y} = 0 \qquad &&\frac{\partial h}{\partial t} - i\frac{\partial g}{\partial z} = 0
\end{aligned}$$

or formulated more explicitly

$$\begin{aligned}\frac{\partial g}{\partial t} &= -i\frac{\partial g}{\partial x} = \frac{\partial h}{\partial y} = i\frac{\partial h}{\partial z}\\ \frac{\partial h}{\partial t} &= i\frac{\partial h}{\partial x} = -\frac{\partial g}{\partial y} = i\frac{\partial g}{\partial z}\end{aligned}$$

In terms of *complex* partial derivatives we then firstly have

$$\begin{aligned}\frac{\partial g}{\partial \bar{v}} &= \frac{1}{2}\left(\frac{\partial g}{\partial t} + i\frac{\partial g}{\partial x}\right) = 0 &\qquad \frac{\partial g}{\partial w} &= \frac{1}{2}\left(\frac{\partial g}{\partial y} + i\frac{\partial g}{\partial z}\right) = 0\\ \frac{\partial h}{\partial v} &= \frac{1}{2}\left(\frac{\partial h}{\partial t} - i\frac{\partial h}{\partial x}\right) = 0 &\qquad \frac{\partial h}{\partial \bar{w}} &= \frac{1}{2}\left(\frac{\partial h}{\partial y} - i\frac{\partial h}{\partial z}\right) = 0\end{aligned}$$

showing by the Cauchy-Riemann formulation (2.25) g to be separately analytic as a function of v and as a function of $\bar{w}$ and h to be separately analytic as a function of $\bar{v}$ and as a function of w. By theorem 2.14 therefore g to be analytic as a function of v and $\bar{w}$ and h to be analytic as a function of $\bar{v}$ and w, and by theorem 2.15 therefore either to have continuous complex partial derivatives of all orders so that we can freely interchange mixed partial derivatives.

Secondly we have

$$\begin{aligned}\frac{\partial g}{\partial v} &= \frac{1}{2}\left(\frac{\partial g}{\partial t} - i\frac{\partial g}{\partial x}\right) = \frac{1}{2}\left(\frac{\partial h}{\partial y} + i\frac{\partial h}{\partial z}\right) = \frac{\partial h}{\partial w}\\ \frac{\partial g}{\partial \bar{w}} &= \frac{1}{2}\left(\frac{\partial g}{\partial y} - i\frac{\partial g}{\partial z}\right) = -\frac{1}{2}\left(\frac{\partial h}{\partial t} + i\frac{\partial h}{\partial x}\right) = -\frac{\partial h}{\partial \bar{v}}\end{aligned} \tag{2.26}$$

and thereby then

$$\begin{aligned}\frac{\partial^2 g}{\partial v^2} &= \frac{\partial}{\partial v}\left(\frac{\partial g}{\partial v}\right) = \frac{\partial}{\partial v}\left(\frac{\partial h}{\partial w}\right) = \frac{\partial}{\partial w}\left(\frac{\partial h}{\partial v}\right) = 0\\ \frac{\partial^2 g}{\partial \bar{w}^2} &= \frac{\partial}{\partial \bar{w}}\left(\frac{\partial g}{\partial \bar{w}}\right) = -\frac{\partial}{\partial \bar{w}}\left(\frac{\partial h}{\partial \bar{v}}\right) = -\frac{\partial}{\partial \bar{v}}\left(\frac{\partial h}{\partial \bar{w}}\right) = 0\\ \frac{\partial^2 h}{\partial \bar{v}^2} &= \frac{\partial}{\partial \bar{v}}\left(\frac{\partial h}{\partial \bar{v}}\right) = -\frac{\partial}{\partial \bar{v}}\left(\frac{\partial g}{\partial \bar{w}}\right) = -\frac{\partial}{\partial \bar{w}}\left(\frac{\partial g}{\partial \bar{v}}\right) = 0\\ \frac{\partial^2 h}{\partial w^2} &= \frac{\partial}{\partial w}\left(\frac{\partial h}{\partial w}\right) = \frac{\partial}{\partial w}\left(\frac{\partial g}{\partial v}\right) = \frac{\partial}{\partial v}\left(\frac{\partial g}{\partial w}\right) = 0\end{aligned}$$

If we for the moment assume that D is convex, then by separate analyticity and the remark following theorem 2.14 we can apply corollary 2.13 to find that g is linear as a function of v and linear as a function of $\bar{w}$ and h is linear as a function of $\bar{v}$ and linear as a function of w:

$$\begin{aligned}g(v, w) &= \alpha + \beta v + \gamma\bar{w} + \delta v\bar{w}\\ h(v, w) &= \varepsilon + \zeta\bar{v} + \eta w + \theta\bar{v}w\end{aligned}$$

for some α, β, γ, δ, ε, ζ, η, $\theta \in \mathbb{C}$.

Using the relations (2.26) again, we get firstly

$$\begin{aligned}
\delta &= \frac{\partial}{\partial \bar{w}}\,(\beta + \delta\bar{w}) = \frac{\partial}{\partial \bar{w}}\left(\frac{\partial g}{\partial v}\right) = \frac{\partial}{\partial \bar{w}}\left(\frac{\partial h}{\partial w}\right) = \frac{\partial}{\partial \bar{w}}\,(\eta + \theta\bar{v}) = 0 \\
\theta &= \frac{\partial}{\partial \bar{v}}\,(\eta + \theta\bar{v}) \;\; = \frac{\partial}{\partial \bar{v}}\left(\frac{\partial h}{\partial w}\right) = \frac{\partial}{\partial \bar{v}}\left(\frac{\partial g}{\partial v}\right) \;\; = \frac{\partial}{\partial \bar{v}}\,(\beta + \delta\bar{w}) = 0
\end{aligned}$$

and then secondly

$$\begin{aligned}
\zeta = \zeta + \theta w &= \frac{\partial h}{\partial \bar{v}} = -\frac{\partial g}{\partial \bar{w}} = -(\gamma + \delta v) = -\gamma \\
\eta = \eta + \theta\bar{v} &= \frac{\partial h}{\partial w} = \frac{\partial g}{\partial v} = \beta + \delta\bar{w} = \beta
\end{aligned}$$

We therefore have

$$\begin{aligned}
g(v,w) &= \alpha + \beta v + \gamma\bar{w} \\
h(v,w) &= \varepsilon - \gamma\bar{v} + \beta w
\end{aligned}$$

and, noting that for any $z \in \mathbb{C}$ we have $jz = \bar{z}j$, by a straightforward calculation

$$\begin{aligned}
f(q) = g(v,w) + j\,h(v,w) &= \alpha + \beta v + \gamma\bar{w} + j(\varepsilon - \gamma\bar{v} + \beta w) \\
&= \alpha + j\varepsilon + (v + jw)(\beta - j\gamma) \\
&= a + qb
\end{aligned}$$

where we have set $a = \alpha + j\varepsilon$ and $b = \beta - j\gamma$, thereby proving the result in the case of a convex D.

The remainder of the argument consists of the in complex analysis familiar process of covering the general domain D by more specific, overlapping sets and invoking the uniqueness principle 2.16 on the overlaps.

In this case, the domain D can be covered by convex sets, any two of which can be connected by a chain of convex sets which overlap in pairs. Invoking the uniqueness principle on the overlaps, we see that $f(q) = a + qb$ with the same constants a and b throughout D, proving the theorem.

2.5 Analyticity

While we have up to this point concentrated on differentiability and have found the concept to be a rather limiting one in the quaternionic and octonionic cases, complex analysis is of course rather the study of *complex analytic* functions and it is just part of the charm of the subject that differentiabiliy and analyticity coincide.

Our concept of analyticity as differentiability of a function throughout some open subset is however alternatively reffered to as *holomorphicity* with, same as for a real-valued function of a real variable, analyticity of a complex-valued function of a complex variable more fundamentally defined in terms of local representability by a convergent power series.

Definition 2.18. Let $D \subseteq \mathbb{C}$ be a non-empty open set. A function $f\colon D \to \mathbb{C}$ is said to be (complex) *analytic* at $z_0 \in D$ if there exist $\epsilon > 0$ and $c_n \in \mathbb{C}$ such that

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n$$

whenever $|z - z_0| < \varepsilon$ or equivalently if $f(z+z_0) = \sum_{n=0}^{\infty} c_n z^n$ whenever $|z| < \varepsilon$.

Suppose $f(x + iy) = u(x, y) + iv(x, y)$ is analytic at $x_0 + iy_0$. Then for some $\varepsilon > 0$ and $c_n \in \mathbb{C}$

$$u(x + x_0, y + y_0) + iv(x + x_0, y + y_0) = \sum_{n=0}^{\infty} c_n (x + iy)^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n} u_{n,k} x^k y^{n-k} + i \sum_{n=0}^{\infty}\sum_{k=0}^{n} v_{n,k} x^k y^{n-k}$$

for some $u_{n,k} \in \mathbb{R}$ and $v_{n,k} \in \mathbb{R}$ whenever $\|(x, y)\| < \varepsilon$ or equivalently

$$u(x, y) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} u_{n,k} (x - x_0)^k (y - y_0)^{n-k}$$
$$v(x, y) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} v_{n,k} (x - x_0)^k (y - y_0)^{n-k}$$

whenever $\|(x, y) - (x_0, y_0)\| < \varepsilon$ which is precisely to say that both u and v are real analytic at the point (x_0, y_0).

A complex analytic function therefore implies two real analytic component functions but note that it is conversely certainly not the case that two real analytic functions of two real variables make one complex analytic function. For example

$$f(x + iy) = x$$

consists of the two everywhere real analytic component functions $u(x, y) = x$ and $v(x, y) = 0$ but is itself certainly nowhere complex analytic due to

$$\frac{\partial u}{\partial x} = 1 \neq 0 = \frac{\partial v}{\partial y}$$

Complex analyticity is a much stronger condition than real analyticity of the component functions.

This changes in the quaternionic case however. We define a quaternionic monomial of degree n in its most general form to be an expression

$$c_0 q c_1 q \cdots c_{n-1} q c_n$$

for some $c_0, \ldots, c_n \in \mathbb{H}$ and again call a function $f\colon D \to \mathbb{H}$ with $D \subseteq \mathbb{H}$ (quaternionic) analytic at $q_0 \in D$ if it is locally a convergent sum of monomials. That is, f is said to be analytic at q_0 if for some $\varepsilon > 0$ and $c_{n,k} \in \mathbb{H}$ and whenever $|q| < \varepsilon$ one has

$$f(q + q_0) = \sum_{n=0}^{\infty} c_{n,0} q c_{n,1} q \cdots c_{n,n-1} q c_{n,n} \tag{2.27}$$

Setting $q = t + ix + jy + kz$, $q_0 = t_0 + ix_0 + jy_0 + kz_0$ and

$$f(q) = s(t,x,y,z) + iu(t,x,y,z) + jv(t,x,y,z) + kw(t,x,y,z) \tag{2.28}$$

we analogously to the complex case above find that if f is quaternionic analytic at q_0 then s, u, v and w are four at (t_0, x_0, y_0, z_0) real analytic functions,

$$s(t,x,y,z) = \sum_{n=0}^{\infty} \sum_{i+j+k=0}^{n} s_{n,i,j,k}(t-t_0)^i (x-x_0)^j (y-y_0)^k (z-z_0)^{n-i-j-k}$$

and similarly for u, v and w.

However, if we are in this case conversely supplied four at (t_0, x_0, y_0, z_0) real analytic functions s, u, v and w of four real variables, we plug in the relations (1.9)

$$t = -\frac{1}{4}(-q + iqi + jqj + kqk) \quad x = -\frac{i}{4}(q - iqi + jqj + kqk)$$
$$y = -\frac{j}{4}(q + iqi - jqj + kqk) \quad z = -\frac{k}{4}(q + iqi + jqj - kqk)$$

and end up with four convergent quaternionic power series. Therefore with an at q_0 quaternionic analytic function f defined through (2.28).

It follows that a theory of quaternionic analytic functions is no other than a theory of real analytic functions of four real variables, a this time much too *broad* a class of functions to be interesting.

This important distinction with the complex case is very fundamentally due to

$$\bar{q} = -\frac{1}{2}(q + iqi + jqj + kqk)$$

which expresses the conjugate $\bar{q}$ as a quaternionic analytic function of q and enables the above relations (1.9). In the complex case no such analytic function exists since, as expressed by the Cauchy-Riemann formulation

$$\frac{\partial f}{\partial \bar{z}} = 0,$$

a function is complex analytic if and only if it depends only on z and not its conjugate. A complex analytic function thereby treats its variable $z = x + iy$ as one single entity and fundamentally not as an of x and y composed number. Conversely formulated, a complex number is (smoothly, analytically) inseparably one and it is this property of complex numbers that makes them as interesting as they are.

Like in the case of differentiability the situation is also again no different for the octonions where, with the flexible law justifying the notation,

$$\bar{o} = -\frac{1}{6}(o + i_1 o i_1 + i_2 o i_2 + i_3 o i_3 + i_4 o i_4 + i_5 o i_5 + i_6 o i_6 + i_7 o i_7)$$

and with both these fundamental inroads into complex analysis unavailable in the quaternionic and octonionic cases, we are about ready to give up on the endeavour.

2.6 Fueter

We should however not leave without at least noting that there does in fact exist a well-developed quaternionic analysis, albeit one somewhat removed from the sort of natural analogue of complex analysis that we were looking for.

We have seen the quaternionic Cauchy-Riemann equations in the form

$$\frac{\partial f}{\partial t} + i\frac{\partial f}{\partial x} = 0 \qquad \frac{\partial f}{\partial t} + j\frac{\partial f}{\partial y} = 0 \qquad \frac{\partial f}{\partial t} + k\frac{\partial f}{\partial z} = 0$$

for the left version and

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}i = 0 \qquad \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}j = 0 \qquad \frac{\partial f}{\partial t} + \frac{\partial f}{\partial z}k = 0$$

for the right, and in comparison with the classic Cauchy-Riemann equation

$$\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y} = 0$$

also a different analogy presents itself in, for a left and right version respectively

$$\frac{\partial f}{\partial t} + i\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y} + k\frac{\partial f}{\partial z} = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}i + \frac{\partial f}{\partial y}j + \frac{\partial f}{\partial z}k = 0$$

We call functions that satisfy these equations left respectively right *regular*.

As is seen by considering $f(q) = q = t + ix + jy + kz$, since as to left regularity

$$\frac{\partial f}{\partial t} + i\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y} + k\frac{\partial f}{\partial z} = 1 + i^2 + j^2 + k^2 = -2$$

and same as to right, not even linear quaternionic functions are regular in this sense, and while this may at first glance appear to now limit us to constant functions there does in fact exist an interesting class of regular functions.

Let g_i and h_i be complex analytic functions of a single complex variable and

$$f_i(t + ix + jy + kz) = g_i(t + ix) + h_i(t + ix)j$$

Then f_i is left regular since by classic Cauchy-Riemann

$$\begin{aligned}\frac{\partial f_i}{\partial t} + i\frac{\partial f_i}{\partial x} + j\frac{\partial f_i}{\partial y} + k\frac{\partial f_i}{\partial z} &= \left(\frac{\partial g_i}{\partial t} + \frac{\partial h_i}{\partial t}j\right) + i\left(\frac{\partial g_i}{\partial x} + \frac{\partial h_i}{\partial x}j\right) + j0 + k0 \\ &= \left(\frac{\partial g_i}{\partial t} + i\frac{\partial g_i}{\partial x}\right) + \left(\frac{\partial h_i}{\partial t} + i\frac{\partial h_i}{\partial x}\right)j = 0 + 0j = 0\end{aligned}$$

The function $f_i(t + ix + jy + kz) = g_i(t + ix) + jh_i(t + ix)$ is similarly right regular.

Note that with the $(1, j)$ plane just another copy of the complex plane we can of course also take two complex analytic functions g_j and h_j defined on *that* plane and with

$$f_j(t + ix + jy + kz) = g_j(t + jy) + h_j(t + jy)k$$

obtain in the same way, as to left regularity

$$\frac{\partial f_j}{\partial t} + i\frac{\partial f_j}{\partial x} + j\frac{\partial f_j}{\partial y} + k\frac{\partial f_j}{\partial z} = \left(\frac{\partial g_j}{\partial t} + \frac{\partial h_j}{\partial t}k\right) + i0 + j\left(\frac{\partial g_j}{\partial y} + \frac{\partial h_j}{\partial y}k\right) + k0$$
$$= \left(\frac{\partial g_j}{\partial t} + j\frac{\partial g_j}{\partial y}\right) + \left(\frac{\partial h_j}{\partial t} + j\frac{\partial h_j}{\partial y}\right)k = 0 + 0k = 0$$

and again similarly for $f_j(t+ix+jy+kz) = g_j(t+iy) + kh_j(t+iy)$ as to right regularity.

Repeating once more with g_k and h_k complex analytic functions on the $(1,k)$ plane and

$$f_k(t+ix+jy+kz) = g_k(t+kz) + h_k(t+kz)i$$

respectively

$$f_k(t+ix+jy+kz) = g_k(t+kz) + ih_k(t+kz)$$

we again find f_k to be left (right) regular and moreover, a sum of left (right) regular functions is clearly again left (right) regular so that specifically a function such as $f = f_i + f_j + f_k$ of the complete quaternionic variable is again left (right) regular.

Although not polynomial in a quaternionic sense, we are therefore now provided with an interesting class of functions to study and it is upon this definition of *regular* that Rudolf Fueter in the 1930s built a theory of regular functions to parallel complex analysis.

Both the in complex analysis primary Cauchy results have an analogue in this theory, in the (left) form of

$$\int_C Dq\, f = 0$$

where C is any smooth closed 3-manifold in $\mathbb{H}$ and Dq is a certain natural quaternion-valued differential 3-form, and in the form of the integral formula

$$f(q_0) = \frac{1}{2\pi^2}\int_{\partial D} \frac{(q-q_0)^{-1}}{|q-q_0|^2} Dq\, f(q)$$

for $q_0 \in D$ where D is a domain in $\mathbb{H}$.

However, said differential 3-form Dq is with $q = t+ix+jy+kz$ given by

$$Dq = dx \wedge dy \wedge dz - i\, dt \wedge dy \wedge dz - j\, dt \wedge dz \wedge dx - k\, dt \wedge dx \wedge dy$$

and together with a somewhat involved definition of a (left) quaternionic derivative that it induces through

$$d(dq \wedge dq\, f) = Dq\, f'(q)$$

uses of the theory remain quite specialised and are in that sense certainly not on par with complex analysis.

The interested reader is referred to Deavours [24] and Sudbery [26] but we will therefore, and while remarking that there have also been different and differently succesful approaches to quaternionic analysis throughout the years, leave things at that and draw our conclusions.

2.7 Conclusion

Having set out on a generalisation of $\mathbb{C}$, we have in the first chapter seen real division algebras such as $\mathbb{C}$ to in fact be quite rare.

They specifically exist only in dimension 1, 2, 4 and 8 and if we additionally require them to be alternative (or normed and unital) then $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and $\mathbb{O}$ are up to isomorphism the only ones.

In the second chapter we first encounter the importance of alternativity for the basic concept of limit and second the importance of the additional dimensions *and* of alternativity for the concept of differentiability. These concepts that we tend to take for granted for $\mathbb{R}$ and that were so succesfully extended to $\mathbb{C}$ by complex analysis are not in fact to be taken overly lightly it seems.

Finally we run into what we feel to be the most fundamental issue in all of this and which consists of the concept of *analyticity* being a remarkable one for complex functions. A quaternionic or octonionic analytic function is nothing other than a componentwise real analytic function due to quaternions and octonions being smoothly decomposable into their constituent real parts, but decomposing a complex number into its real and imaginary part requires the force of a blunt instrument.

A complex number is much more *one* in that sense and therefore also much more interesting as a mathematical entity. We eventually feel that it should therefore not actually come as a surprise that complex analysis is also special and rather uniquely interesting.

We shall no more underestimate a complex number.

Bibliography

[1] W.R. Hamilton, *Elements of Quaternions*, Edited by W.E. Hamilton, Longmans, Green and Co. 1866

[2] W.R. Hamilton, *Elements of Quaternions*, Second Edition, Volume I, Edited by C.J. Joly, Longmans, Green and Co. 1899

[3] W.R. Hamilton, *Elements of Quaternions*, Second Edition, Volume II, Edited by C.J. Joly, Longmans, Green and Co. 1901

[4] W.R. Hamilton, *On Quaternions: Letter to John T. Graves, Esq.*, Philos. Mag. 25 (1844), 489–495

[5] G. Frobenius, *Über lineare Substitutionen und bilineare Formen*, J. Reine Angew. Math. 84 (1878), 1–63

[6] A. Hurwitz, *Über die Composition der quadratischen Formen von belibig vielen Variablen*, Nachr. Gesellschaft Wissenschaft. Göttingen Math-Phys (1898), 309–316

[7] M. Zorn, *Theorie der Alternativen Ringe*, Abh. Math. Sem. Univ. Hamburg 9 (1933), 395–402

[8] R. Bott and J. Milnor, *On the parallelizability of the spheres*, Bull. Amer. Math. Soc. 64 (1958), 87–89

[9] M. Kervaire, *Non-Parallelizability of the n-sphere for $n > 7$*, Proc. Nat. Acad. Sci. USA 44 (1958), 280-283

[10] R.S. Palais, *The Classification of Real Division Algebras*, Amer. Math. Monthly 75 (1968), 366–368

[11] I.L. Kantor and A.S. Solodovnikov, *Hypercomplex Numbers (An Elementary Introduction to Algebras)*, Translated from the Russian *Giperkompleksnye Čhisla* (Nauka, Moscow, 1973), Springer-Verlag 1989, ISBN 0-387-96980-2

[12] H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel and A. Remmert, *Numbers*, Translated from the German *Zahlen* (Grundwissen Mathematik 1, Springer-Verlag 1988), Springer-Verlag 1991, Corrected third printing 1995, ISBN 0-387-97497-0

[13] A. Oneto, *Alternative Real Division Algebras of Finite Dimension*, Divulg. Mat. 10-2 (2002), 161–169

[14] J.C. Baez, *The Octonions*, Bull. Amer. Math. Soc. 39 (2002), 145–205

[15] J.H. Conway and D.A. Smith, *On Quaternions and Octonions: Their Geometry, Arithmetic and Symmetry*, A K Peters Ltd. 2003, ISBN 1-56881-134-9

[16] M. Badger, *Division algebras over the real numbers*, Midterm project for Math 2501, Abstract Algebra II, University of Pittsburgh, 2006

[17] G. Roos, *Exceptional symmetric domains*, Contemp. Math. 468 (2008), 157–189

[18] G. Scheffers, *Verallgemeinerung der Grundlagen der gewöhnlichen komplexen Funktionen*, Sitzungsber. Säch. Gesellsch. Wiss. Math.-Phys. Kl. 45 (1893), 328–342

[19] N.M. Krylov, *On quaternions of Rowan Hamilton and on the notion of monogenity*, Dokl. Akad. Nauk. SSSR 55 (1947), 799–800 (in Russian)

[20] A.S. Meilikhson, *About monogenity of quaternions*, Dokl. Akad. Nauk. SSSR 69 (1948), 431–434 (in Russian)

[21] T.M. Apostol, Calculus, Second Edition, Volume I, John Wiley & Sons, Inc. 1967, ISBN 0-471-00006-X

[22] T.M. Apostol, Calculus, Second Edition, Volume II, John Wiley & Sons, Inc. 1967, ISBN 0-471-00008-6

[23] N. Levinson and R.M. Redheffer, *Complex Variables* McGraw-Hill Inc. 1970, ISBN 0-07-037492-9

[24] C.A. Deavours, *The quaternion calculus*, Amer. Math. Monthly 80 (1973), 995–1008

[25] H. Cartan, *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Translated from the French, Reprint of the 1973 Edition, Dover Publications Inc. 1995, ISBN 0-486-68543-8

[26] A. Sudbery, *Quaternionic analysis*, Math. Proc. Camb. Phil. Soc. 85 (1979), 199–225

[27] S.G. Krantz, *Function Theory of several complex variables*, Reprint of the 1992 edition, AMS Chelsea Publishing 2001, ISBN 0-8218-2724-3

[28] E.B. Saff and A.D. Snider, *Fundamentals of Complex Analysis (with Applications to Engineering and Science)*, Third Edition, Pearson Education International 2003, ISBN 0-13-017968-X

[29] M.S. Marinov, *On the Meilikhson theorem*, Trends in Complex Analysis, Differential Geometry and Mathematical Physics (Proceedings of the 6th International Workshop on Complex Structures and Vector Fields), Editors S. Dimiev and K. Segikawa, World Scientific Publishing Co. Inc. 2003, ISBN 978-981-238-452-2, 135–140

[30] S. De Leo and P.P. Rotelli, *Quaternionic Analyticity*, Appl. Math. Lett. 16 (2003), 1077–1081