1 Imagine: Imaginary Numbers Are For Real

I assume that you have been exposed to a quantity like "$i = \sqrt{-1}$" before; and you know that numbers like "i times a real number" are called "imaginary" numbers. I presume your very first exposure to them was through the quadratic formula. When applying it to an equation like $x^2 + 2x + 2 = 0$, it produces something like $\frac{1}{2}(-2 \pm \sqrt{-4}) = -1 \pm \sqrt{-1}$. When graphing $x^2 + 2x + 2$, you see that it does not vanish for any real number. The impression that such an introduction would almost inavoidably have left on you is:

> *"Well, heck, so they have invented quantities that don't actually exist and are thus justly called imaginary; all for the purpose of serving as solutions to equations that do not have solutions. Today Math has gone to become Black Magic. . . "*

The purpose of this introduction is threefold: (a) To deconstruct such a motivation, (b) to give a more compelling motivation for the sensibility of imaginary numbers, (c) to construct complex numbers in a mathematically sound way.

Deconstructing the poor motivation: If inventing quantities for the purpose of serving as solutions to equations that hitherto have not had any solutions were a valid method, we could with the same right invent an algebraic quantity ∞ as a solution to the equation $0 \cdot x = 1$. We could then do some simple algebra like "$2 = 2 \cdot 1 = 2 \cdot (0 \cdot \infty) = (2 \cdot 0) \cdot \infty = 0 \cdot \infty = 1$" to conclude that $2 = 1$.

Mind you that this 'algebraic quantity ∞' I have invented here for the sake of showing its failure is not the same as the symbol ∞ in calculus; the calculus symbol ∞ does not partake of the full family of algebraic operations.

Conclusion: When we introduce new 'numbers' we must justify that they satisfy the rules of calculation (specifically the *field axioms* outlined below).

Redeeming the algebraic motivation somehow:

The cubic formula makes a stronger case for the legitimacy of imaginary numbers than the quadratic formula. It says:

A solution to the cubic equation $x^3 + px + q = 0$ is given by Cardano's formula:

$$x_1 = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

Hwk: *Use this formula to obtain the solution $x_1 = -2$ for $x^3 - x + 6 = 0$; then check it both numerically (with step by step intermediate evaluation) and by explicit exact algebra, plugging it in. Warning: If you skip the 'step by step' part and feed the formula into some software package for black box numerical evaluation, something may go awry (software dependent) behind the black box, leaving you clueless about the source of the problem.*

The interesting thing happens when $x^3 + px + q = 0$ has *three* real solutions. Which one does the formula select? The paradoxical outcome is: When $x^3 + px + q = 0$ has three real solutions, then the square root inside Cardano's formula is imaginary! Try it out:

Hwk: *Find the real solutions to $x^3 - 7x + 6 = 0$ by guess work. Use Cardano's formula on the equation, but abstain from attempts to evaluate the cubic roots of complex numbers.*

Hwk: *Use real variable calculus to determine a precise (necessary and sufficient) condition on p, q for $x^3 + px + q$ to have three distinct real solutions. Namely, we need a maximum with positive value and a minimum with negative value. (Those who had 300 and/or 341 should use an appropriately rigorous style of exposition. Confirm that this condition is equivalent to the term under the square root in Cardano's formula being negative.)*

Once we introduce complex numbers and give proper definitions of the roots of complex numbers, Cardano's formula can be expanded to give two more solutions:

$$x_2 = \frac{-1+i\sqrt{3}}{2}\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}} + \frac{-1-i\sqrt{3}}{2}\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}$$
$$x_3 = \frac{-1-i\sqrt{3}}{2}\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}} + \frac{-1+i\sqrt{3}}{2}\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}$$

Just as in the first homework, the combination of nested roots mysteriously conspired to get a rational number out of two irrational expressions, in the case of three real solutions, the formulas conspire to get real numbers out of three manifestly non-real expressions.

So in this example, imaginary quantities seem to be working behind the scenery to produce perfectly good real solutions. If you dismiss imaginary numbers and say, 'I only care for real solutions', you disable the formula that produces them exactly in the case when your desire for real solutions is met with the best possible abundance of real solutions!

> Be forewarned however: Taking roots of complex numbers requires insight that you do not have at this point. Odds are if you try it naively you run into contradictions, due to the fact that there should be three complex solutions to the equation $x^3 = a$ and there is no preferred way of choosing one and calling it *the* cube root of a. Some essential information about the meaning of the cube root has been left unspecified in Cardano's formulas. When I plugged Cardano's formula into MATHEMATICA (Version 5.2) for numerical evaluation of Hwk 1, I got a wrong answer! This was due to incompatible conventions about the proper choice of cube roots. Mathematica chooses $\frac{1}{2}(-1+i\sqrt{3})$ as cube root of -1, whereas we probably prefer to choose -1, which was the tacitly intended choice in the present exposition.

Hwk: *Make a mathematically correct and useful statement out of the power law $(a^b)^c = a^{bc}$ for real numbers. By mathematically correct, I mean that quantifiers like e.g. 'for every real number' or 'for every positive number' or 'for every integer' should be given with each symbol, and that with these quantifiers a true staement must result. By useful, I mean that the desirable cases should be covered and there is room for interpretation of what is desirable. Include at least one example of real a, b, c for which $(a^b)^c$ and a^{bc} are both defined, but are not equal.*

CONSTRUCTION AND FIELD AXIOMS; NO ROOTS YET:

On the set $\mathbb{R}^2$, we define addition as the usual vector addition $(a, b) + (c, d) := (a + c, b + d)$ and multiplication by $(a, b) \cdot (c, d) := (ac - bd, ad + bc)$. We often use the letters z and w for elements of $\mathbb{R}^2$ in the context of this algebra, and we denote the set $\mathbb{R}^2$ as $\mathbb{C}$. Elements of $\mathbb{C}$ are called complex numbers. The first coordinate a of (a, b) is called the real part of (a, b); the second coordinate is called the imaginary part.

One can then prove that these operations satisfy the usual rules of algebra with real numbers, known as *field axioms*:

(C+) For any $z, z' \in \mathbb{C}$, it holds $z + z' = z' + z$.
(A+) For any $z, z', z'' \in \mathbb{C}$, it holds $(z + z') + z'' = z + (z' + z'')$.
(N+) There exists an element $\mathbf{0} \in \mathbb{C}$ such that $z + \mathbf{0} = z$ for every $z \in \mathbb{C}$. [It is an easy consequence of (C+) that there can be only one such element.]
(I+) For each $z \in \mathbb{C}$ there exists an element $w \in \mathbb{C}$ such that $z + w = \mathbf{0}$. [It is then an easy consequence of the axioms so far that w is uniquely determined by z. With this being said, we may define $-z$ to be the unique element w satisfying $z + w = \mathbf{0}$.]

(C·) For any $z, z' \in \mathbb{C}$, it holds $z \cdot z' = z' \cdot z$.
(A·) For any $z, z', z'' \in \mathbb{C}$, it holds $(z \cdot z') \cdot z'' = z \cdot (z' \cdot z'')$.
(N·) There exists an element $\mathbf{1} \in \mathbb{C}$ such that $z \cdot \mathbf{1} = z$ for every $z \in \mathbb{C}$. [It is an easy consequence of (C·) that there can be only one such element.]
(01) $\mathbf{0} \neq \mathbf{1}$
(I·) For each $z \in \mathbb{C} \setminus \{\mathbf{0}\}$ there exists an element $w \in \mathbb{C}$ such that $z \cdot w = \mathbf{1}$. [It is then an easy consequence of the axioms so far that w is uniquely determined by z. With this being said, we may define z^{-1} to be the unique element w satisfying $z \cdot w = \mathbf{1}$.]

(D) For each $z, z', z'' \in \mathbb{C}$ it holds $(z + z') \cdot z'' = z \cdot z'' + z' \cdot z''$.

Each of these properties can be proved in a straightforward manner from similar properties for real numbers and the definition of the operations. Specifically, one can identify $(0, 0)$ as the element $\mathbf{0}$ stipulated in (N+), and $(1, 0)$ as the element $\mathbf{1}$ stipulated in (N·).

Concerning (I·), it can be verified that $(a, b)^{-1}$ is actually $(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2})$.

Next we observe that $(a, 0) + (b, 0) = (a + b, 0)$ and $(a, 0) \cdot (b, 0) = (ab, 0)$. In other words, if we restrict to the elements $(a, 0) \in \mathbb{C}$ they obey exactly the same arithmetic as the corresponding elements $a \in \mathbb{R}$, so we identify $(a, 0)$ with a in the notation. Algebra folks will prefer to state this property in the words: The mapping $a \mapsto (a, 0)$ is an injective field homomorphism from $\mathbb{R}$ to $\mathbb{C}$. With this identification, $\mathbb{R}$ becomes a subset of $\mathbb{C}$. We can now assign the symbol i to the special complex number $(0, 1) \in \mathbb{C}$. Then, for real a, b we have $a + b \cdot i = (a, 0) + (b, 0) \cdot (0, 1) = (a, 0) + (0, b) = (a, b)$. Moreover $i \cdot i = (0, 1) \cdot (0, 1) = (-1, 0) = -(1, 0) = -1$.

After this boring procedure, we write complex numbers always as $a + bi$ instead of (a, b), and we can be assured that calculating with these expressions according to the same algebra as real numbers, subject to the extra rule that $i \cdot i = -1$, is mathematically well founded. Note that $(-i) \cdot (-i)$ is also equal to -1, so it is not clear which of $i, -i$ should be called $\sqrt{-1}$ and which $-\sqrt{-1}$.

We introduce subtraction and division by $z - w := z + (-w)$, $z/w := z \cdot w^{-1}$ (the latter assuming $w \neq 0$). We define z^n for positive integers n inductively; for $z \neq 0$, we define $z^0 = 1$ and $z^{-n} := (z^{-1})^n$.

While in later developments, a convention will be employed that defines $\sqrt{-1}$ to be i rather than $-i$, it is probably preferrable for now not to use the notation $\sqrt{-1}$ at all and stick with i instead. So for the moment (in defiance of 1.13 in the textbook) we will consider $\sqrt{-1}$ as a notation that has not been properly defined yet. Be aware that naive algebra with this symbol may run into contradictions and should therefore be avoided:

$$¿ \quad -1 = i \cdot i = \sqrt{-1} \cdot \sqrt{-1} \stackrel{?}{=} \sqrt{(-1)(-1)} = \sqrt{1} = 1 \quad ?$$

So $\sqrt{a}$ is, for now, only defined when a is a nonnegative real number, and the quadratic formula should have a footnote saying 'read $\pm\sqrt{b^2 - 4ac}$ as $\pm i\sqrt{4ac - b^2}$ when the former is undefined'. For the same reason, we do *not* presently define powers z^a of complex z for other exponents a than integers. We'll come to discuss this matter further when we discuss roots in more detail.

No Order:

We also do *not* (now or ever) define the relations $<$, $>$, $\leq$, $\geq$ for complex numbers. There cannot be a field together with an order relation $>$ satisfying the usual ordered field axioms (that I skip here) that at the same time would contain an element x satisfying $x^2 + 1 = 0$.

Complex conjugates, arguments, absolute values:

For a complex $z = a + bi$ with a, b real, we call a the real part and b the imaginary part of z, in symbols $a = \operatorname{Re} z$ and $b = \operatorname{Im} z$. Note that the imaginary part is the real quantity b, not the imaginary quantity ib.

The complex conjugate $\bar{z}$ of $z = a + bi$ with real a, b is defined to be $a - ib$. Note the following rules, which can be easily checked:

$\overline{z + w} = \bar{z} + \bar{w}$
$\overline{z - w} = \bar{z} - \bar{w}$
$\overline{z \cdot w} = \bar{z} \cdot \bar{w}$
$\overline{z/w} = \bar{z}/\bar{w}$
$\bar{\bar{z}} = z$

In the language of abstract algebra, complex conjugacy is a field automorphism of $\mathbb{C}$ and it is its own inverse. For this reason, in view of $\bar{i} = -i$, there is no way of distinguishing i and $-i$ merely by properties of algebra. In more elementary words, any correct algebraic formula involving complex numbers remains true if i gets consistently replaced with $-i$, in all explicit as well as implicit occurrences; here replacing i with $-i$ in implicit occurrences means to replace any complex number by its complex conjugate.

If we represent the complex number $z = a + bi \in \mathbb{C}$ as the point or vector (a, b) in the plane $\mathbb{R}^2$ (per our original definition of complex numbers), we can associate with z also its polar coordinates: We denote by $|z| := \sqrt{a^2 + b^2} = \sqrt{z\bar{z}}$ the absolute value of z, and if we write $a = |z| \cos\phi$, $b = |z| \sin\phi$, we call ϕ 'the' argument of z, written $\arg z$.

$\arg z$ is only defined for $z \neq 0$, and even then it is only defined up to an integer multiple of 2π. (Algebra buffs may view $\arg z$ as an element of the quotient group $(\mathbb{R}, +)/(2\pi\mathbb{Z}, +)$, but to the relief of non-algebraists, we will not insist on abstract algebra terminology.)

Ok, I cannot omit this joke here: Complex numbers have arguments and absolute values, but they are neither cantankerous, nor bigoted;-)

The following new trig function cis is useful to introduce here:

$$\operatorname{cis}\phi := \cos\phi + i\sin\phi$$

[If all the pathetic precal and cal textbooks out there would just refrain from introducing the utterly useless miscreant functions secant and cosecant, they could actually do some good with the saved time by introducing cis instead.]

What makes cis useful is the simple formula (that you can check by using the addition theorems of sin and cos)

$$\operatorname{cis}(\phi + \theta) = \operatorname{cis}\phi \cdot \operatorname{cis}\theta\ .$$

So cis behaves like an exponential function. We will later see that $\operatorname{cis}\phi = e^{i\phi}$, once we have defined e^z for $z \in \mathbb{C}$. You should have seen this identity in the sophomore DiffEq class. By that time, the notation cis will then hardly be needed any more. But in the meanwhile we can use this distinct notation to explore various approaches to this fundamental identity without implicitly assuming it in the notation $e^{i\phi}$ already.