Math 202 notes

Jason Riedy

20 October, 2008

Contents

Notes also available as PDF.

- exponents, roots, and irrationals
- decimals and percentages
- floating-point arithmetic

exponents, roots, and irrationals

- exponents, rules, etc.

- extending to negative exponents: rationals
- extending to rational exponents leads to roots
- roots to/from exponents

1 Real numbers

We won't *define* the real numbers. That requires more time than we can allow here. We will simply use the reals, denoted $\mathbb{R}$, as more than the rationals. This was the state of affairs until around 1872 when Richard Dedekind finally discovered a way to construct real numbers formally.

So the reals fit into our system of sets on the very top,

$$\text{natural numbers } \subsetneq \text{ whole numbers } \subsetneq \mathbb{J} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}.$$

Look up the term "Dedekind cut" for more on actually defining real numbers.

2 Exponents and roots

We will cover:

- definition for positive integer exponents,
- rules,
- zero exponents,
- negative exponents, and
- rational exponents and roots.

2.1 Positive exponents

We've already used positive exponents when discussing the digit representation of numbers:

$$\begin{aligned}
10 &= 10 &&= 10^1 \\
100 &= 10 \cdot 10 &&= 10^2 \\
1000 &= 10 \cdot 10 \cdot 10 &&= 10^3 \\
10000 &= 10 \cdot 10 \cdot 10 \cdot 10 &&= 10^4 \\
\vdots & \quad \vdots && \quad \vdots
\end{aligned}$$

In general, for any number (integer, rational, or real), the number raised to an integer exponent is defined as:

$$\begin{aligned} a^1 &= a, \\ a^2 &= a \cdot a, \\ a^3 &= a \cdot a \cdot a, \\ &\vdots \\ a^k &= \overbrace{a \cdot a \cdot a \cdot \ldots \cdot a}^{k}. \end{aligned}$$

For example,

$$\begin{aligned} 2^3 &= 8, \text{ and} \\ \left(\frac{2}{3}\right)^2 &= \frac{4}{9}. \end{aligned}$$

Negative numbers have signs that bounce around:

$$\begin{aligned} (-5)^1 &= -5, \\ (-5)^2 &= 25, \\ (-5)^3 &= -125, \text{ and} \\ (-5)^4 &= 625. \end{aligned}$$

With the symbolic definition, we can show other properties of exponentiation:

$$\begin{aligned} (ab)^3 &= (ab) \cdot (ab) \cdot (ab) \\ &= (a \cdot a \cdot a) \cdot (b \cdot b \cdot b) \text{ (by commutativity and associativity)} \\ &= a^3 \cdot b^3. \end{aligned}$$

In general,

$$(ab)^k = a^k b^k.$$

For example,

$$1000 = 10^3 = (2 \cdot 5)^3 = 2^3 \cdot 5^3 = 8 \cdot 125.$$

Or when multiplying numbers raised to powers, we have that exponents add as in

$$\begin{aligned} a^k \cdot a^m &= \overbrace{a \cdot \ldots \cdot a}^{k} \cdot \overbrace{a \cdot \ldots \cdot a}^{m} \\ &= \overbrace{a \cdot \cdots \cdot a}^{k+m} \\ &= a^{k+m}. \end{aligned}$$

For example,

$$10^2 \cdot 10^3 = 100 \cdot 1000 = 100000 = 10^5.$$

And numbers raised to powers multiple times multiply exponents as in

$$(a^k)^m = \overbrace{a^k \cdot a^k \cdot a^k \cdot \ldots \cdot a^k}^{m} = a^{km}.$$

For example,

$$100^2 = (10^2)^2 = 10^4 = 10000.$$

2.2 Zero exponent

Consider the following relationship between integer exponents and division:

$$\begin{aligned} a^3 &= a^4/a, \\ a^2 &= a^3/a, \text{ and} \\ a^1 &= a^2/a. \end{aligned}$$

Reasoning *inductively*, we suspect that

$$a^0 = a^1/a = 1.$$

Using the rule above for adding exponents along with the additive identity property that $k + 0 = k$, we can *deduce* that

$$a^k = a^{k+0} = a^k \cdot a^0.$$

So for any $a \neq 0$,

$$a^0 = 1 \quad \text{when } a \neq 0.$$

Why can't we define this for $a = 0$? $0^k = 0$ for any integer $k > 0$. So $0 = 0^k = 0^k \cdot 0^0$ does not help to define 0^0; we're left with $0 = 0 \cdot 0^0$. Because $0 \cdot x = 0$ for any x, 0^0 can be anything.

Examples:

$$\begin{aligned} 5^0 &= 1 \\ (-73)^0 &= 1 \\ 0^0 &\text{ is undefined}\ldots \end{aligned}$$

2.3 Negative exponents

Continuing *inductively* for $a \neq 0$,

$$\begin{aligned} a^0 &= 1, \\ a^{-1} &= a^0/a = \frac{1}{a}, \text{ and} \\ a^{-2} &= a^{-1}/a = \frac{1}{a} \cdot \frac{1}{a} = \frac{1}{a^2}. \end{aligned}$$

Again, we can use the fact that exponents add to derive this *deductively*:

$$1 = a^0 = a^{k+-k} = a^k \cdot a^{-k},$$

and so a^{-k} is the multiplicative inverse of a^k, and we previously showed that to be $\frac{1}{a^k}$. We have shown that

$$a^{-k} = \frac{1}{a^k}$$

for all $a \neq 0$.

For example:

$$2^{-2} = \frac{1}{2^2} = \frac{1}{4}$$

is the inverse of

$$2^2 = 4.$$

Also,

$$\begin{aligned} \left(\frac{2}{3}\right)^{-1} &= \frac{1}{\frac{2}{3}} \\ &= \frac{3}{2} \end{aligned}$$

is the multiplicative inverse of

$$\frac{2}{3}.$$

And

$$\begin{aligned} \left(\frac{2}{3}\right)^{-2} &= \frac{1}{(\frac{2}{3})^2} \\ &= \frac{1}{\frac{4}{9}} \\ &= \frac{9}{4} \end{aligned}$$

is the multiplicative inverse of

$$\left(\frac{2}{3}\right)^2 = \frac{2}{3}.$$

2.4 Rational exponents and roots

So we've played with division and exponents. Consider now reasoning *inductively* using the multiplication rule for exponents:

$$\begin{aligned} a^4 &= (a^2)^2, \\ a^2 &= (a^1)^2, \text{ and so} \\ a^1 &= (a^{\frac{1}{2}})^2. \end{aligned}$$

We call $a^{\frac{1}{2}}$ the square root of a and write $\sqrt{a}$.

But $\sqrt{a}$ is only defined some of the time. Over integers, there clearly is no integer b such that $b^2 = 2$, so $\sqrt{2}$ is not defined **over the integers** and fractional exponents are **not closed** over integers.

Also, the product of two negative numbers is positive, and the product of two positive numbers is positive, so there is no real number whose square is negative. Hence for real a,

$$\sqrt{a} \text{ is undefined for } a < 0.$$

Remember that $(-b)^2 = (-1)^2 \cdot b^2 = b^2$, so the square root may be either positive or negative!

$$\begin{aligned} 2^2 &= 4, \\ (-2)^2 &= 4, \text{ hence} \\ \sqrt{4} &= \pm 2. \end{aligned}$$

In most circumstances, $\sqrt{a}$ means the positive root, often called the **principal square root**. When you hit a square-root key or apply a square root in a spreadsheet, you get the principal square root.

Other rationals provide other roots:

$$a^1 = (a^{\frac{1}{3}})^3$$

is the cube root,

$$\sqrt[3]{a} = a^{\frac{1}{3}}.$$

Here, though, $(-a)^3 = (-1)^3 \cdot a^3 = -(a^3)$, and there is no worry about the sign of the cube root.

Using $(a^k)^m = a^{km}$, we also have

$$a^{\frac{2}{3}} = \sqrt[3]{a^2} = (\sqrt[3]{a})^2$$

The exponential operator can be defined on more than just the rationals, but we won't go there. However, remember that I mentioned the rationals are *dense* in the reals. There is a rational number close

2.5 Irrational numbers

There are more reals than rationals. This is a very non-obvious statement. To justify it, we will

- prove that $\sqrt{2}$ is not rational, and
- generalize that proof to other roots.

Remember the table to show that there are as many integers as rationals? You cannot construct one for the reals. I might show that someday. It's shockingly simple but still a mind-bender. But for now, a few simple examples suffice to make the point.

Theorem: The number $\sqrt{2}$ is not rational.

Proof. Suppose $\sqrt{2}$ were a rational number. Then

$$\sqrt{2} = \frac{a}{b}$$

for some integers a and b. We will show that any such a and b, two must divide both and so $(a, b) \geq 2$. Previously, we explained that any fraction can be reduced to have $(a, b) = 1$. Proving that $(a, b) \geq 2$ shows that we *cannot* write $\sqrt{2}$ as a fraction.

Now if $\sqrt{2} = \frac{a}{b}$, then $2 = \frac{a^2}{b^2}$ and $2b^2 = a^2$. Because $2 \mid 2b^2$, we also know that $2 \mid a^2$. In turn, $2 \mid a^2$ and 2 being prime imply that $2 \mid a$ and thus $a = 2q$ for some integer q.

With $a = 2q$, $a^2 = 4q^2$. And with $a^2 = 2b^2$, $2b^2 = 4q^2$ or $b^2 = 2q^2$. Now $2 \mid b$ as well as $2 \mid a$, so $(a, b) \geq 2$. □

Theorem: Suppose x and n are positive integers and that $\sqrt[n]{x}$ is rational. Then $\sqrt[n]{x}$ is an integer.

Proof. Because $\sqrt[n]{a}$ is rational and positive, there are positive integers a and b such that

$$\sqrt[n]{x} = \frac{a}{b}.$$

We can assume further that the fraction is in lowest terms, so $(a, b) = 1$. Now we show that $b = 1$.

As in the previous proof, $\sqrt[n]{x} = \frac{a}{b}$ implies that $x \cdot b^n = a^n$.

If $b \geq 1$, there is a prime p that divides b. And as before, $p \mid b$ implies $p \mid a$, contradicting the assumption that $(a, b) = 1$. Thus $b = 1$ and $\sqrt[n]{x}$ is an integer. □

With decimal expansions, we will see that rational numbers have repeating expansions. Irrational numbers have decimal expansions that never repeat. There are some fascinating properties of the expansions

Irrational numbers come in two kinds, **algebraic** and **transcendental**. We won't go into the difference in detail, but numbers like $\sqrt{2}$ are algebraic, and numbers like π and e are transcendental.

3 Decimal expansions and percentages

Remember positional notation:

$$1\,234 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0.$$

Given negative exponents, we can expand to the right of 10^0. General English notation uses a **decimal point** to separate the integer portion of the number from the rest.

So with the same notation,

$$\begin{gathered} 1\,234.567 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0 \\ +5 \cdot 10^{-1} + 6 \cdot 10^{-2} + 7 \cdot 10^{-3}. \end{gathered}$$

Operations work in exactly the same digit-by-digit manner as before. When any position goes over 9, a factor of 10 **carries** into the next higher power of 10. If any digit becomes negative, a factor of 10 is **borrowed** frrom the next higher power of 10.

Other languages use a comma to separate the integer from the rest and also use a period to mark off powers of three on the other side, for example

$$1,234.567 = 1.234,567.$$

You may see this if you play with "locales" in various software packages. Obviously, this can lead to massive confusion among travellers. (A price of 1.234 is **not** less than 2 but rather greater than 1000.)

Typical international mathematical and science publications use a period to separate the integer and use a space to break groups of three:

$$1,234.567 = 1\,234.567.$$

3.1 Representing rationals with decimals

What is the part to the right of the decimal point? It often is called the **fractional part** of the number, giving away that it is a representation of a fraction.

Here we consider the decimal representation of rational numbers $\frac{1}{a}$ for different integers a. We will see that the expansions fall into two categories:

1. some **terminate** after a few digits, leaving the rest zero; and
2. some **repeat** a trailing section of digits.

For rational numbers, these are the only two possibilities.

We can find the decimal expansions by long division.

Two simple examples that terminate:

$$\begin{array}{r|l} & 0.5 \\ \cline{2-2} 2 & 1.0 \\ & -1.0 \end{array} \qquad\qquad \begin{array}{r|l} & 0.2 \\ \cline{2-2} 5 & 1.0 \\ & -1.0 \end{array}$$

Note that $2 \mid 10$ and $5 \mid 10$, so both expansions terminate immediately with $\frac{1}{2} = .5$ and $\frac{1}{5} = .2$.

Actually, all fractions with a denominator consisting of powers of 2 and five have terminating expansions. For example,

$$\begin{aligned} \frac{1}{2^2} &= \frac{1}{4} = 0.25, \\ \frac{1}{5^3} &= \frac{1}{125} = 0.008, \text{ and} \\ \frac{1}{2 \cdot 5^2} &= \frac{1}{50} = 0.02. \end{aligned}$$

What if the denominator a in $\frac{1}{a}$ does not divide 10, or $a \nmid 10$? Then the expansion does not terminate, but it does **repeat**. If the denominator has no factors of 2 or 5, it repeats immediately.

Examples of repeating decimal expansions:

$$\begin{array}{r|l} & 0.33\ldots \\ \cline{2-2} 3 & 1.000 \\ & -.9 \\ & .10 \\ & -9 \\ & 10 \end{array} \qquad\qquad \begin{array}{r|l} & 0.1428571\ldots \\ \cline{2-2} 7 & 1.00000000 \\ & -.7 \\ & 30 \\ & -28 \\ & 20 \\ & -14 \\ & 60 \\ & -56 \\ & 40 \\ & -35 \\ & 50 \\ & -49 \\ & 10 \\ & -7 \\ & 3 \end{array}$$

We write these with a bar over the repeating portion, as in

$$\frac{1}{3} = 0.\overline{3}, \text{ and}$$
$$\frac{1}{7} = 0.\overline{142857}.$$

We say that $0.\overline{3}$ has a **period** of 1 and $0.\overline{142857}$ has a period of 6.

We could write $0.2 = 0.2\overline{0}$, but generally we say that this **terminates** once we reach the repeting zeros.

If the denominator a contains factors of 2 or 5, the repeating portion occurs a number of places after the decimal. For example, consider $\frac{1}{6} = \frac{1}{2\cdot 3}$ and $\frac{1}{45} = \frac{1}{5\cdot 9}$:

$$\begin{array}{r|l} & 0.166\ldots \\ \hline 6 & 1.0000 \\ & -.6 \\ & 40 \\ & -36 \\ & 40 \end{array} \qquad \begin{array}{r|l} & 0.022\ldots \\ \hline 45 & 1.0000 \\ & -.90 \\ & 100 \\ & - 90 \\ & 10 \end{array}$$

So the decimal representations are

$$\frac{1}{6} = 0.1\overline{6}, \text{ and}$$
$$\frac{1}{45} = 0.0\overline{2}.$$

3.1.1 The hard way to determine the period of a repeating fraction

Note that for all non-negative integer k,

$$10^k \equiv 0 \pmod{2},$$
$$10^k \equiv 0 \pmod{5}, \text{ and}$$
$$10^k \equiv 1 \pmod{3}.$$

These tell us that the expansions have periods of 0, 0, and 1.

For seven,

$$\begin{aligned}
10^0 &\equiv 1 \pmod 7,\\
10^1 &\equiv 3 \pmod 7,\\
10^2 &\equiv 2 \pmod 7,\\
10^3 &\equiv 6 \pmod 7,\\
10^4 &\equiv 4 \pmod 7,\\
10^5 &\equiv 5 \pmod 7, \text{ and}\\
10^6 &\equiv 1 \pmod 7,
\end{aligned}$$

so the period is of length 7.

For 45,

$$\begin{aligned}
10^0 &\equiv 1 \pmod{45},\\
10^1 &\equiv 10 \pmod{45}, \text{ and}\\
10^2 &\equiv 10 \pmod{45}.
\end{aligned}$$

This is a little more complicated, but the pattern shows that there is one initial digit before hitting a repeating pattern, exactly like the expansion $\frac{1}{45} = 0.0\overline{2}$.

In each case, we are looking for the **order** of 10 modulo the denominator. Finding an integer with a large order modulo another integer is a building block in RSA encryption used in SSL (the `https` prefix in URLs).

3.2 The repeating decimal expansion may not be unique!

One common stumbling block for people is that the repeating decimal expansion is not unique.

Let

$$n = 0.\overline{9} = 0.9999\overline{9}.$$

Then multiplying n by 10 shifts the decimal over one but does not alter the pattern, so

$$10n = 9.\overline{9} = 9.9999\overline{9}.$$

Given

$$\begin{aligned}
10n &= 9.\overline{9}, \text{ and}\\
n &= 0.\overline{9},
\end{aligned}$$

we can subtract n from the former.

$$9n = 9.\overline{9} - 0.\overline{9} = 9.$$

With $9n = 9$, we know $n = 1$. Thus $1 = 0.\overline{9}$!

This is a consequence of sums over infinite sequences, a very interesting and useful topic for another course. But this technique is useful for proving that rationals have repeating expansions.

3.3 Rationals have terminating or repeating expansions

Theorem: A decimal expansion that repeats (or terminates) represents a rational number.

Proof. Let n be the number represented by a repeating decimal expansion. Without loss of generality, assume that $n > 0$ and that the integer portion is zero. Now let that expansion have d initial digits and then a period of length p. Here we let a terminating decimal be represented by trailing 0 digits with a period of 1.

For example, let $d = 4$ and $p = 5$. Then n looks like

$$n = 0\,.\,d_1d_2d_3d_4\overline{p_1p_2p_3p_4p_5}.$$

Then $10^d n$ leaves the repeating portion to the right of the decimal. Following our example $d = 4$ and $p = 5$,

$$10^4 n = d_1d_2d_3d_4\,.\,\overline{p_1p_2p_3p_4p_5}.$$

Because it repeats, $10^{d+p}n$ has the same pattern to the right of the decimal. In our running example,

$$10^{4+5}n = d_1d_2d_3d_4p_1p_2p_3p_4p_5\,.\,\overline{p_1p_2p_3p_4p_5}.$$

So $10^{d+p}n - 10^d n$ has zeros to the right of the decimal and is an integer k. In our example,

$$k = 10^{4+5}n - 10^4 n = d_1d_2d_3d_4p_1p_2p_3p_4p_5 - d_1d_2d_3d_4.$$

We assumed $n > 0$, so the difference above is a positive integer. The fractional parts cancel out.

Now $n = \frac{k}{10^{d+p} - 10^d}$ is one integer over another and thus is rational. □

Theorem: All rational numbers have repeating or terminating decimal expansions.

Proof. This is a very different style of proof, using what we have called the **pidgeonhole principle**. Without loss of generality, assume the rational number of interest is of the form $\frac{1}{d}$ for some positive integer d.

At each step in long division, there are only d possible remainers. If some remainder is 0, the expansion terminates.

If no remainder is 0, then there are only $d-1$ possible remainders that appear. If the expansion is taken to length d, some remainder must appear twice. Because of the long division procedure, equal remainders leave equal sub-problems, and thus the expansion repeats. □

3.4 Therefore, irrationals have non-repeating expansions.

So we know that any repeating or terminating decimal expansion represents a rational, and that all rationals have terminating or repeating decimal expansions.

Thus, we have the following:
Corollary: A number is rational if and only if it has a repeating decimal expansion.

So if there is no repeating portion, the number is *irrational.* One example,

$$0.101001000100001\cdots,$$

has an increasing number of zero digits between each one digit. This number is irrational.

It's beyond our scope to prove that π is irrational, but it is. Thus the digits of π do not repeat.

3.5 Percentages as rationals and decimals

Percentage comes from *per centile*, or part per 100. So a direct numerical equivalent to 85% is

$$85\% = \frac{85}{100} = .85.$$

We can expand fractions to include decimals in the numerator and denominator. The decimals are just rationals in another form, and we already explored "complex fractions" with rational numerators and denominators.

So we can express decimal percentages,

$$85.75\% = \frac{85.75}{100} = .8575.$$

Everything else "just works". To convert a fraction into a percentage, there are two routes. One is to convert the denominator into 100:

$$\frac{1}{2} = \frac{50}{100} = 50\%.$$

Another is to produce the decimal expansion and then multiply that by 100:

$$\frac{1}{7} = 0.\overline{142857} = 14.2857\overline{142857}\%.$$

Converting a percentage into a proper fraction required dropping the percentage into the numerator and then manipulating it appropriately:

$$85.75\% = \frac{85.75}{100} = \frac{\frac{8575}{100}}{100} = \frac{8575}{10000} = \frac{343}{400}.$$

4 Fixed and floating-point arithmetic

So far we have considered *infinite* expansions, ones that are not limited to a set number of digits. Computers (and calculators) cannot store infinite expansions that do not repeat, and those that do require more overhead than they are worth.

Instead, computers **round** infinite results to have at most a fixed number of **significant digits**. Operations on these limited representations incur some **round-off error**, leading to a tension between computing speed and the precision of computed results. One important fact to bear in mind is that **precision does not imply accuracy**. The following is a very precise but completely in-accurate statement:

> The moon is made of Camembert cheese.

First we'll cover different rounding rules from the perspective of **fixed-point arithmetic**, or arithmetic using a set number of digits to the right of the decimal plce. Then we'll explain **floating-point arithmetic** where the decimal point "floats" through a fixed number of significant digits.

We will not cover the errors in floating-point operations, but we will cover the errors that come from typical binary representation of decimal data.

The points you need to take away from this are the following:

- Using a limited number of digits (or bits) to represent real numbers leads to some inherent, representationall error.
- Representing every-day decimal quantities in binary also incurs some representational error.

Despite the doom-like points above, floating-point arithmetic often provides results that are accurate enough. We won't be able to cover why this is, but the high-level reasons include:

- using far more digits of precision than initially appear necessary, and
- carrying intermediate results to even higher precision.

4.1 Rounding rules

Generally, computer arithmetic can be modelled as computing the **exact** result and then rounding that exact result into an economical representation.

truncation or rounding to zero With this rounding method, digits beyond the stored digits are simply dropped.

rounding half-way away from zero This is the text's method of "round half up". A number is rounded to the nearest representable number. In the half-way case, where the digits beyond the number of digits stored are $5000\cdots$, the number is rounded upwards.

rounding half-way to even This is the **preferred** method for rounding in general. A number is rounded to the nearest representable number. In the half-way case, where the digits beyond the number of digits stored are $5000\cdots$, the number is rounded so the final stored digit is **even**.

There are more rounding methods, but these suffice for our discussion. Rounding rules are hugely important in banking and finance, and there are quite a few versions required by different regulations and laws.

Examples of each rounding method above, rounding to two places after the decimal point:

initial number	truncate	round half up	round to nearest even
$\frac{1}{3} = 0.\overline{3}$	0.33	0.33	0.33
$\frac{1}{7} = 0.\overline{142857}$	0.14	0.14	0.14
0.444	0.44	0.44	0.44
0.445	0.44	0.45	0.44
0.4451	0.44	0.45	0.45
0.446	0.44	0.45	0.45
0.455	0.44	0.46	0.46

Rounding error is the absolute difference between the exact number and the rounded, stored representation. In the table above, the rounding error in representing $\frac{1}{3}$ is $|\frac{1}{3} - 0.33| = |\frac{1}{3} - \frac{33}{100}| = |\frac{100}{300} - \frac{99}{300}| = \frac{1}{300} = 0.00\overline{3}$. Note that here the rounding error is 1% of the exact result. That error is large because we use only two digits.

Note that you cannot round in stages. Consider round-to-nearest-even applied to 0.99455 and rounding to two places after the point:

Incorrect	Correct
0.99455	0.99455
0.9946	
0.995	
1.00	0.99

4.2 Floating-point representation

Consider repeatedly dividing by 10 in fixed-point arithmetic that carries two digits beyond the decimal:

$$\begin{aligned} 1 \div 10 &= 0.10, \\ 0.1 \div 10 &= 0.01, \\ 0.01 \div 10 &= 0.00. \end{aligned}$$

So $((1 \div 10) \div 10) \div 10$ evaluates to 0! This phenomenon is called **underflow**, where a number grows too small to be represented. A similar phenomenon, **overflow**, occurs when a number becomes too large to be represented. Computer arithmetics differ on how they handle over- and underflow, but generally overflow produces an ∞ symbol and underflow produces 0.

Floating-point arithmetic compensates for this by carrying a fixed number of **significant** digits rather than a fixed number of fractional digits. The position of the decimal place is carried in an **explicit, integer exponent**. This allows floating-point numbers to store a wider range and actually makes analysis of the round-off error easier.

In floating-point arithmetic,

$$\begin{aligned} 1 \div 10 &= 1.000 \cdot 10^{0}, \\ 0.1 \div 10 &= 1.000 \cdot 10^{-1}, \\ 0.01 \div 10 & 1.000 \cdot 10^{-2}, \\ &\vdots \end{aligned}$$

This continues until we run out of representable range for the integer exponents. We leave the details of floating-point underflow for another day (if you're unlucky).

4.3 Binary fractional parts

Just as integers can be converted to other bases, fractional parts can be converted as well.

Each position to the right of the point (no longer the *decimal* point) corresponds to a power of the base. For binary, the typical computer representation,

$$\begin{aligned}\frac{1}{2} &= 2^{-1} = 0.1_2 = 0.5,\\ \frac{1}{4} &= 2^{-2} = 0.01_2 = 0.25,\\ \frac{1}{8} &= 2^{-3} = 0.001_2 = 0.125.\end{aligned}$$

So a binary fractional part can be expanded with powers of two:

$$0.1101_2 = \frac{1}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} = 0.8125.$$

To find a binary expansion, we need to carry out long division in base 2. I won't ask you to do that.

The important part to recognize is that finite *decimal* expansions may have infinite, repeating *binary* expansions! Remember that in decimal, $2 \mid 10$ and $5 \mid 10$, so negative powers of 2 and 5 have terminating decimal expansions. In binary, only $2 \mid 2$, so only powers of 2 have terminating binary expansions.

Numbers you expect to be exact are not. Consider 0.1. Its binary expansion is

$$0.1 = 0.\overline{00011}_2.$$

A five-bit fixed-point representation would use

$$0.1 \approx 0.00011_2.$$

The error in representing this with a five-digit fixed-point representation is 0.00625, or over 6%.

In a five-bit floating-point representation,

$$0.1 \approx 1.1001_2 \cdot 2^{-4}.$$

The error here is less than 0.0024, or less than 0.24%. You can see what floating-point gains here.

Ultimately, though, in a limited binary fractional representation, adding ten dimes does not equal one dollar! This is why often programs slanted towards finance (*e.g.* spreadsheets) use a form of decimal arithmetic. On current common hardware, decimal arithmetic is implemented in software rather than hardware and is orders of magnitude slower than binary arithmetic.

5 Homework

Practice is absolutely critical in this class.

Groups are fine, turn in your own work. Homework is due in or before class on Mondays.

- Problem set 7.1 (p421):
  - 1, 2, 3, 4
  - 14
  - 25
  - 28
- Problem set 7.2 (p433):
  - 4
  - 5 (I won't get a chance to cover this, but scientific notation is a good exercise in positional notation and rounding.)
  - 9
- Problem set 7.4 (p457);
  - 2, 3, 4
  - 18
- On rounding and floating point arithmetic:
  - Round each of the following to the nearest tenth (one place after the decimal) using **round to nearest even**, **round to zero (truncation)**, and **round half-up**:
    - 86.548
    - 86.554
    - 86.55
  - Compute the following quantities with a computer or a calculator. **Write what type of computer/calculator you used and the software package if it's a computer.** Compute it as shown. Do not simplify the expression before computing it, and do not re-enter the intermediate results into the calculator or computer program. Also compute the expressions that do not include 10^{16} by hand exactly. There should be a difference between the exact result and the displayed result in some of these cases. Remember to work from the innermost parentheses outward.

  - $\overbrace{(0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1)}^{\text{10 times}} - 1$
  - $(\overbrace{(0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1 + 0.1)}^{\text{10 times}} - 1) \times 10^{16}$, where 10^{16} often is entered as 1e16. If the result overflows (signals an error) on various calculators, replace 10^{16} by 10^{8} in this and later portions.
  - $(((2 \div 3) - 1) \times 3) + 1$
  - $((((2 \div 3) - 1) \times 3) + 1) \times 10^{16}$
  - $(((6 \div 7) - 1) \times 7) + 1$
  - $((((6 \div 7) - 1) \times 7) + 1) \times 10^{16}$

  The object of this first part is to demonstrate round-off error. The first to problems, adding 0.1 repeatedly, may see no error if the device calculates in decimal. The latter four parts should see some error regardless of the base used.

- Now copy down the number displayed by the first calculation in each of the following. Re-enter it as x in the second calculation.
  - $1 \div 3$, then $1 \div 3 - x$ where x is the number displayed.
  - If you have a calculator or program with π, π, then $\pi - x$ where x is the number displayed.

  The object here is to see that the number displayed often is not the number the computer or calculator has stored.

Note that you *may* email homework. However, I don't use Microsoft$^{\text{TM}}$ products (*e.g.* Word), and software packages are notoriously finicky about translating mathematics.

If you're typing it (which I advise just for practice in whatever tools you use), you likely want to turn in a printout. If you do want to email your submission, please produce a PDF or PostScript document.