Introduction to the mathematics of DSP

Topics: Complex numbers, probability and statistics, arithmetic of sinusoids

It is not possible to understand digital signal processing (DSP) without understanding some of the mathematics behind it. A little actually does go a long way, and fortunately, some of what is needed you already know from high school or earlier.

COMPLEX NUMBERS

> "The Divine Spirit found a sublime outlet in that wonder of analysis, that portent of the ideal world, that amphibian between being and not-being, which we call the imaginary root of negative unity" –Leibnitz (1646-1716).

It is often assumed that complex numbers arose to "solve" quadratic equations that had no real roots, i.e., $x^2+1=0$. However, this is not true. Such quadratic equations were simply deemed to have no solution, in the same way that $x \cdot 0 = 3$ has no solution. It was only when mathematicians discovered that without complex numbers, there was no way to obtain *real* solutions to *cubic* equations that they began to take complex numbers seriously.

Consider a general cubic equation $x^3+bx^2+cx=d$. Since cubics of the form $x^3=d$ are easy to solve, it is reasonable to expect that the next easiest is $x^3+px=q$. Perhaps the first to solve this type of equation were three 16th century Italian mathematicians: S. dal Ferro, N. Fontana (nicknamed "Tartaglia" due to his stammer) and G. Cardano. Cardano published the treatise *Ars Magna* (1545) with the first complete description of the method. Before we consider this special case, it is worth noting that the general cubic reduces to this with the substitution $x=y-b/3$.

Now to solve $x^3+px=q$, note that we have two coefficients p and q, which allow us to set up two independent equations. Suppose we let $x=s-t$ for two numbers s,t. Then we must have $s^3-t^3=q$ and $3st=p$ in order to have a solution. Solving these two equations for s gives us the equation $s^6-qs^3-p^3/27=0$. We can solve this by letting $u=s^3$ and applying the quadratic formula to the equation $u^2-qu-p^3/27=0$. Then $s=\sqrt[3]{u}$ and $t=\dfrac{p}{3s}$.

It would seem we are done, except that if we try this with a simple equation like $x^3-3x=0$, then we run into a problem right away. We know the solutions here: $x=0$ and $x=\pm\sqrt{3}$. But if we solve as above, we get that $s=\sqrt[3]{\sqrt{-1}}$, and therefore

$$x = \sqrt[3]{\sqrt{-1}} + \frac{1}{\sqrt[3]{\sqrt{-1}}}$$

What on earth is the right hand side? It is not zero, and so therefore it must be $\pm\sqrt{3}$. But how could this be? We will see later that the right hand side is $\sqrt{3}$, but for right now it is worth considering what the 16th century mathematicians must be been faced with. *To get a real answer to a real equation, we must use an imaginary number*[1]. There is no avoiding $j = \sqrt{-1}$. In mathematics j is as important as the numbers 1, 0, e, and π.

EULER

The Swiss mathematician Leonhard Euler (1707-1783) is one the most prolific and influential mathematicans ever. His collected works span 74 volumes, including nearly 400 papers completed after he lost sight in both eyes. One of his greatest discoveries is that

$$e^{j\theta} = \cos\theta + j\sin\theta$$

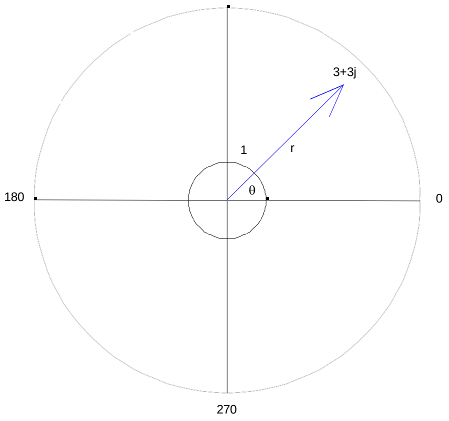
This equation can be proven by comparing the Taylor series of functions on both sides. This leads immediately to the famous theorem, which connects the five most important numbers in mathematics:

$$e^{j\pi} + 1 = 0$$

Euler's result lead to the polar representation, illustrated in Figure 1:

$$z = x + jy = re^{j\theta} \qquad r = \sqrt{x^2 + y^2} \qquad \theta = \tan^{-1}(y/x)$$

Figure 1: Polar representation of the number $z = 3 + 3j$



---

[1] Of course we could have solved this equation without using the general formula, but that begs the question of what the formula is good for.

Euler's theorem also shows how to find the roots of complex numbers. What is the $\sqrt[3]{\sqrt{-1}} = \sqrt[3]{j}$? By Euler, we see that $j = e^{j(\pi/2)}$, and therefore $\sqrt[3]{j} = \left[e^{j(\pi/2)}\right]^{1/3} = e^{j(\pi/6)}$. However, there is more. How many cube roots can a number have? We know that any number has two square roots, $x^2 = a$ has two solutions. Therefore, there should be three cube roots for each number. By Figure 1, each number has a polar representation. Since angles are equivalent modulo 2π , every complex number can be represented in one of an infinite number of ways:

$$z = re^{j(\theta + 2\pi k)} \qquad k = 0, \pm 1, \pm 2, ...$$

Therefore the cube roots of a complex number are

$$z = r^{1/3} e^{j(\theta + 2\pi k)/3} = r^{1/3} e^{j(\theta/3 + 2\pi k/3)} \qquad k = 0, \pm 1, \pm 2, \cdots$$

Although there are infinitely many *k* values, there are only three distinct angles on the right hand side of the equation above:

$$\frac{\theta}{3}, \frac{\theta}{3} + \frac{2\pi}{3}, \frac{\theta}{3} + \frac{4\pi}{3}$$

Therefore, there are exactly three cube roots for every complex number.

Using this result, we can solve the cubic $x^3 - 3x = 0$ by finding the value of

$$x = \sqrt[3]{\sqrt{-1}} + \frac{1}{\sqrt[3]{\sqrt{-1}}}$$

Using $\sqrt[3]{\sqrt{-1}} = e^{j\pi/6}$, we obtain that $x = e^{j(\pi/6)} + e^{-j(\pi/6)} = 2\cos(\pi/6) = \sqrt{3}$. This is one of the solutions for the cubic.

Exercises:

1. Show that the other two solutions to $x^3 - 3x = 0$ can be found by using the other two cube roots of $\sqrt{-1}$ in the expression for *x* above.

2. Show that if *z* is any complex number such that $z \neq 1$, then $\sum_{k=0}^{N-1} z^k = \frac{z^N - 1}{z - 1}$

PROBABILITY AND STATISTICS

> Suppose you are shown three doors, numbered one, two and three. Behind one door there is a million dollars, and but behind the other two there is only one penny. Each door is equally likely to hold the million dollars, so there is no significance to the numbering one, two and three. You can choose one door to open, and you get to keep whatever is behind that door. You choose door number one. Before you can open it, a genie who knows what is behind the doors appears, and tells you that you can change your mind after viewing what is behind door number three. It turns out to be a penny. Should you change your mind and pick door number two?
>
> *-- the "Monty Hall" problem*

The answer turns out to be yes, you should pick door two, because your chances of winning the million dollars are 2/3, whereas they are only 1/3 if you stick with door one. Why is that? Before answering this question, we turn to the subject of probability theory.

A random variable is a quantity whose value we can't know beforehand. The *mean* of a random variable X is its average value, which is estimated by

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

This estimate gets better with N, and eventually converges to the true mean μ of x. The *standard deviation* of x is a measure of its spread about the mean, and it is estimated by

$$s = \sqrt{\frac{\sum_{k=0}^{N-1} (x_k - \bar{x})^2}{N-1}}$$

Note that the denominator contains *N-1*, and not *N*. This is to provide an unbiased (accurate on average) estimate since we do not know the true mean of X. This estimate also converges to the true standard deviation σ as N grows larger. For example, suppose we have the following N = 10 observations of a certain random variable:

1, 3, –1, 5, 6, –9, 11, 0, 1, 2

Then $\bar{x} = 1.9$ and $s = 5.19$.

Note that the standard deviation can be estimated with only one pass through the data by the rule:

$$s = \frac{1}{\sqrt{N-1}} \sqrt{\left(\sum_{k=0}^{N-1} x_k^2 \right) - \frac{1}{N} \left(\sum_{n=0}^{N-1} x_n \right)^2}$$

For theoretical work, we use a *probability density function* (pdf) to describe a random variable *X*. If $p(x)$ is the pdf of *X*, then the probability that *X* lies in any range is determined as follows:

$$\text{Probability } \{a \le X \le b\} = \int_a^b p(x)dx$$

Note that this requires that

$$p(x) \ge 0; \quad \int_{-\infty}^{\infty} p(x)dx = 1.$$

The two most important probability densities are the uniform and the Gaussian. We say that a random variable *X* is uniformly distributed between *a* and *b*, and write $X \sim U[a,b]$, if its probability density is

$$p(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & else \end{cases}$$

The second important probability density was discovered by Carl Freidrich Gauss (1777-1855), who used it to describe errors in measuring angles in astronomy. We say that a random variable is distributed as a Gaussian with mean μ and standard deviation σ if its probability density is

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Figure 2 shows the plot of a Gaussian probability density.

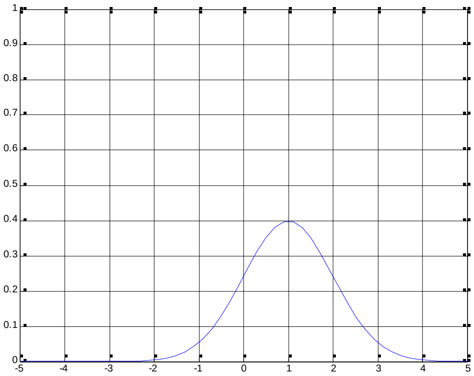


Figure 2: A Gaussian probability density with mean $\mu = 1$ and standard deviation $\sigma = 1$ is graphed.

Notice that the Gaussian favors values near the mean, giving them higher probabilities than those further away. In fact, the chances of a Gaussian random variable lying within $\pm\sigma$ of its mean μ is 0.68, or 68%, within $\pm 2\sigma$ of μ is 0.95, and within $\pm 3\sigma$ of μ is 0.998.

An important result in probability theory known as the Central Limit Theorem shows that the sum of independent random variables becomes increasingly like a Gaussian random variable as more variables are added, regardless of the probability densities of the individual variables. Since many random effects are produced by the addition of independent sources, the Gaussian distribution occurs frequently. So frequently in fact, that the Gaussian is called the "normal" distribution[2]. We say that a random variable is normally distributed, and write $X \sim N[\mu,\sigma]$, if its probability density is a Gaussian with mean μ and standard deviation σ.

It is important to think about the difference between *accuracy* of a statement, which is how close is it to the truth, and the *precision* of a statement , which is how specific is it. For example, if we place a set of temperature sensors around the room and average their readings, the value may be 25.12 degrees C. Placing more sensors results in more digits after the decimal place, or greater precision. However, the accuracy of the estimate will be improved not by adding more sensors, but by calibrating the sensors, or by arranging them carefully to avoid drafts, etc.

Exercises

3. The standard deviation of a random variable X with a probability density p(x) can be estimated by the formula

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

   Show that if $X \sim U[a,b]$, then $\sigma = \frac{b-a}{\sqrt{12}} = 0.29(b-a)$. *Hint:* first solve the problem for a = 0, b = 1.

4. Explain how you could obtain a Gaussian random variable from a uniform random number generator.

5. The root mean square (RMS) value of a random number is

$$rms = \sqrt{\frac{\sum_{k=0}^{N-1} x_k^2}{N}}$$

   If you know the estimated mean $\bar{x}$ and the estimated standard deviation s, can you determine the *rms* value? Assume that the value of N is known.

---

[2] A portrait of Gauss and the formula for a Gaussian used to appear on the German 10 mark note in the days before the Euro.

ARITHMETIC OF SINUSOIDS

A simple but useful view of signal processing is that it is the "arithmetic of sinusoids by using complex numbers". A sinusoid is any signal of the form

$$x(t) = A\cos(2\pi ft + \theta).$$

As Figure 3 shows, the parameters of a sinusoid are as follows: A is the *amplitude*, f is the *frequency* in units of cycles/sec (Hertz), t is time in seconds, and θ is *phase* in radians.

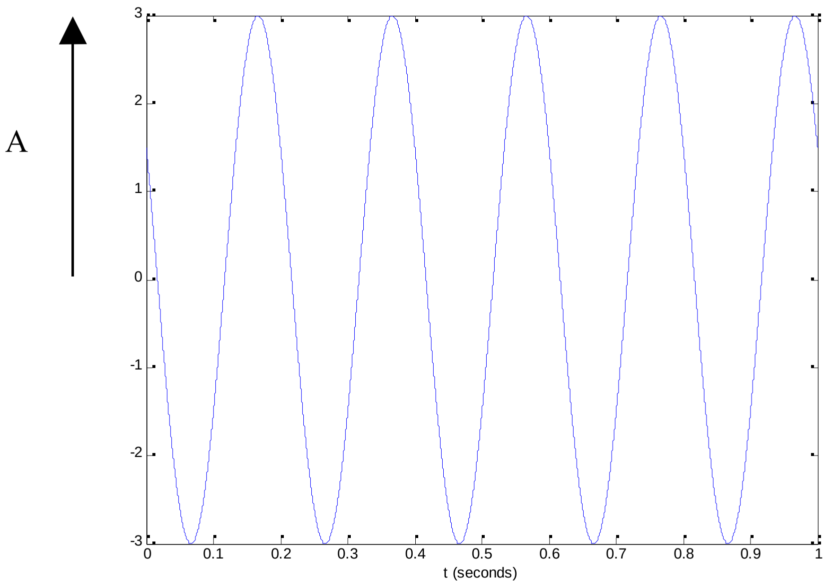


Figure 3: A sinusoid with amplitude $A = 3$, *frequency* $f = 3$ *Hz (note the 3 complete cycles in one second), and phase* $\theta = \pi/3$. *Note that at* $t = 0$, *the value is* $3\cos(\pi/3) = 1.5$.

The sinusoidal addition formula will be useful to us.

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B)$$

Note that this formula implies the sinusoidal product formulas

$$\cos(A) + \cos(B) = 2\cos(\frac{A+B}{2})\cos(\frac{A-B}{2})$$
$$\cos(A) - \cos(B) = -2\sin(\frac{A+B}{2})\sin(\frac{A-B}{2})$$

These formulas can be proven by using Euler's theorem:

$$e^{j\theta} = \cos(\theta) + j\sin(\theta)$$

They are the basis for the “arithmetic” of sinusoids.

The first arithmetic result we establish is that frequencies do not change when two or more sinusoids are added, but amplitudes and phases do. For example, adding two sinusoids of the same frequency produces a third sinusoid of the same frequency, but with a different amplitude and phase[3]. Mathematically, this can be written as

$$A_1\cos(2\pi ft+\theta_1) + A_2\cos(2\pi ft+\theta_2) = A_3\cos(2\pi ft+\theta_3)$$

A simple way to prove this is to realize that $\cos\theta = \mathrm{Re}\{e^{j\theta}\}$, and therefore

$$\begin{aligned} &A_1\cos(2\pi ft+\theta_1) + A_2\cos(2\pi ft+\theta_2) \\ &= \mathrm{Re}\{A_1 e^{j2\pi ft}e^{j\theta_1}\} + \mathrm{Re}\left\{A_2 e^{j2\pi ft}e^{j\theta_2}\right\} \\ &= \mathrm{Re}\left\{e^{j2\pi ft}\left(A_1e^{j\theta_1} + A_2e^{j\theta_2}\right)\right\} \\ &= \mathrm{Re}\left\{e^{j2\pi ft}\left(A_3e^{j\theta_3}\right)\right\} = A_3\cos(2\pi ft+\theta_3) \end{aligned}$$

The second result we prove explains the “beat-note” phenomonen: adding two sinusoids of different frequencies appears to be a product of sinusoids. Beat notes are useful for tuning musical instruments[4]. The specific result to prove is that

$$\begin{aligned} A_1\cos(2\pi f_1t+\theta_1) + A_2\cos(2\pi f_2t+\theta_2) = &(A_1+A_2)\cos(2\pi f_ct+\theta_c)\cos(2\pi f_dt+\theta_d) \\ &+(A_2-A_1)\sin(2\pi f_ct+\theta_c)\sin(2\pi f_dt+\theta_d) \end{aligned}$$

Here $f_c = (f_1+f_2)/2$ is the center frequency, $f_d = (f_1-f_2)/2$ is the difference, and similarly for θ_c, θ_d. This is a direct application of the sinusoidal product formulas. Note that when the center frequency is much larger than the difference, the product appears as a slowly varying amplitude modulation.

A MATLAB demonstration of a beat note, and the code used to generate it included in the Appendix.

Exercise (extra credit)

6. Prove the beat note formula above.

---

[3] On a guitar, two locations on the fretboard are “unisons” if they produce the same note. If these two locations are played simultaneously, the note does not change.

[4] If two unisons on a guitar are played simultaneously, the result should sound as one note. If the guitar is out of tune, a beat note is audible.

Musical terminology in DSP

Many DSP terms originated in music. One example is the term “octave”, which mean a doubling of frequency. Two frequencies are said to be one octave apart if one is twice the other. For exampe, 440 Hz and 880 Hz are one octave apart. Another such term is “harmonic”. To any frequency, there are second, third, fourth, ... n-th harmonics, which are frequencies that are n times the original. For example, for f = 440 Hz, the second harmonic is 880 Hz, the third harmonic is $3x440 = 1,320$ Hz, etc. Note that the second harmonic is also one octave higher. The base frequency 440 Hz is called the “fundamental”, to distinguish it from the harmonics.

Harmonics are often used to measure nonlinearity in a system. For example, consider a system whose output is $y = x + x^2$. If the input is $x(t) = \cos(2\pi ft)$, then the output is

$$\begin{aligned} y(t) &= \cos(2\pi ft) + \cos^2(2\pi ft) \\ &= \cos(2\pi ft) + \frac{1}{2} + \frac{1}{2}\cos(4\pi ft) \end{aligned}$$

Note that the output contains power both at DC (constant term) and at the second harmonic.

The term “total harmonic distortion” (THD) is often used in audio reproduction systems. It refers to the output power at all other frequencies besides the fundamental, divided by total output power, when a pure tone is input to a system:

$$THD = \frac{\text{power excluding fundamental}}{\text{total power}} x100\%$$

For the example above, the THD is

$$THD = \frac{(0.25 + 0.125)}{(0.5 + 0.25 + 0.125)} x100\% = 43\%$$

This result follows once we verify that the power in a sinusoidal voltage is

$$\frac{1}{T}\int_0^T \left(A\cos\left(2\pi ft + \theta\right)\right)^2 dt = \begin{cases} \dfrac{A^2}{2} & f \neq 0 \\ A^2\cos^2(\theta) & f = 0 \end{cases}$$

Here, $T = \frac{1}{f}$.

APPENDIX

```
% demonstration of beat note, formed by adding two
% sinusoids whose frequencies are closely spaced
%
% r kakarala
% UCBx DSP I

fs=8192;  %sampling rate
f1=999;   % two closely
f2=1001;  % spaced frequencies

t=0:1/fs:3;   % three seconds at sampling rate of 8192 Ha.
x1=cos(2*pi*f1*t);
x2=cos(2*pi*f2*t);

pause;

% play x1
soundsc(x1);
pause;

% play beat note
soundsc(x1+x2,fs);
% plot beat note

plot(t,x1+x2);
title('beat note');
xlabel('t (seconds)');
```