Contents

1. Introduction

1.1. Introduction

A **signal** is defined as “something” that carries information. In general, this “something” is a pattern of variations of a physical quantity that can be manipulated, stored, or transmitted by physical processes. Sadly, this definition is not particularly informative, since it encompasses a wide range of eterogeneous physical phenomena. A notable property of signals is that they can be represented or encoded in many equivalent ways, convertible into one another.

The most natural language to describe signals is mathematics. Signals can have one or more than one dimension, depending on how many variables are needed to describe them. Some examples of one-dimensional signals are:

- Sound, like music or human speech;
- A sensor’s output, like those of a thermal sensor or of a motion sensor;
- Physiological signals, like EEGs;
- Financial data, like market trends and exchange rates.

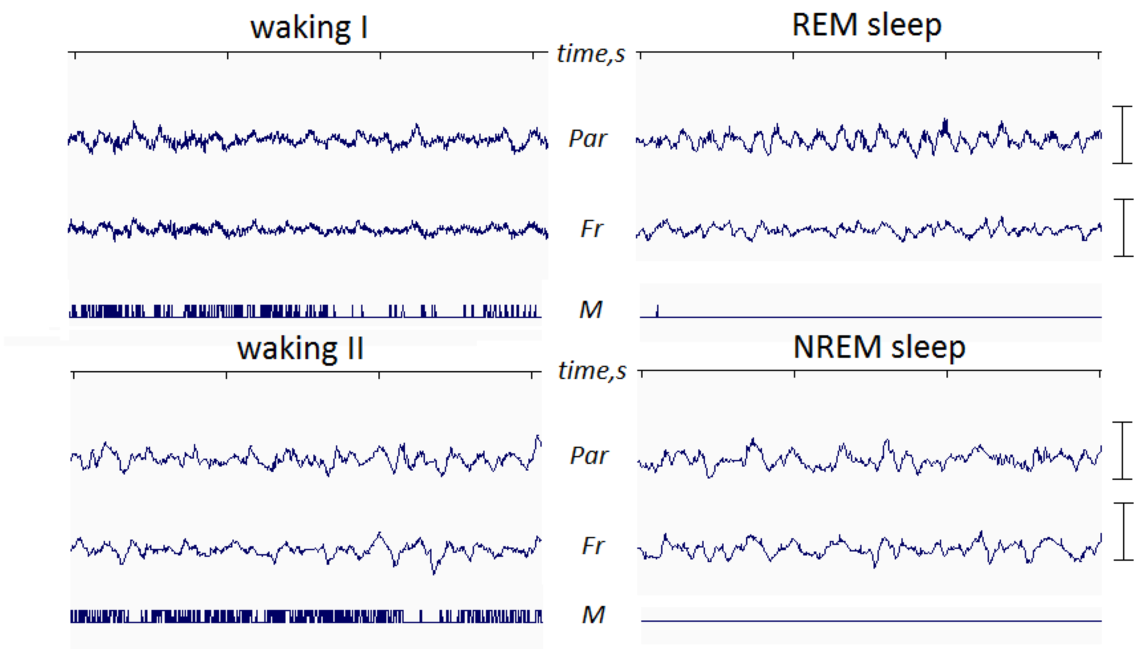


Figure 1: EEG of a mouse during different stages of sleep. [Original image by Andrii Cherninskyi, licensed under CC BY-SA 4.0.]

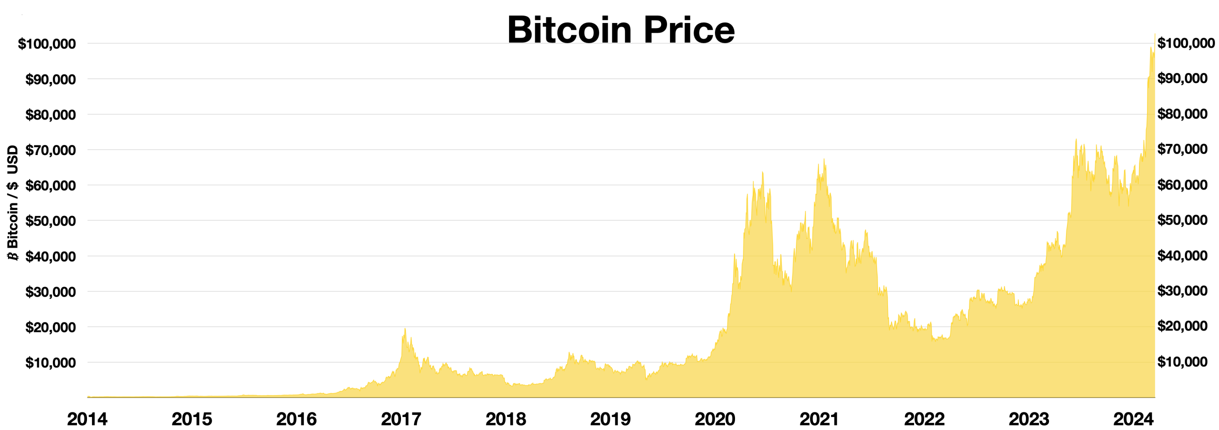


Figure 2: Bitcoin/US Dollar exchange date from 2014 to 2024. [Original image by Wikideas1, licensed under CC0. Original file in webp format.]

From one dimensional signals it is possible to generalize to multidimensional signals. Examples of two dimensional signals are images (photographs, thermal captures, radiographies, ...). Examples of three dimensional signals are 3D models (point clouds, meshes, ...).

A **system** is any process or apparatus that has a signal as input, performs some manipulation on such signal and then returns another signal as output. The output signal can be the original signal but in a different representation or a completely different signal altogether.

Many signals, such as sound, are naturally thought of as a pattern of variations in time. The evolution of a signal with respect to time is described by what's called the **time waveform**, a function $s(t)$ with a single independent variable t, representing time, and whose output is a displacement or disturbance of sort. $s(t)$ can be of arbitrary complexity, and may not be possible to write it as a closed-form expression, but it exists nonetheless. As a matter of fact, it is possible to directly refer to a signal by its waveform.

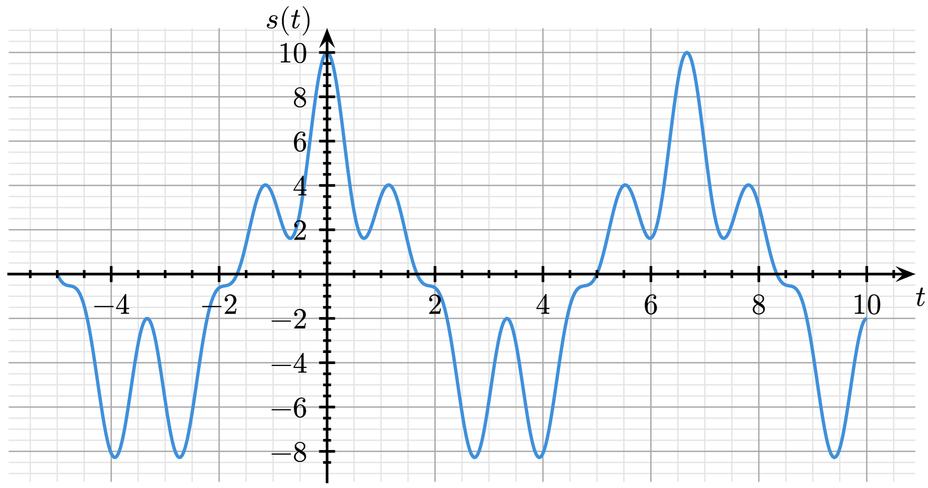


Figure 3: The plot of a continuous time signal.

Most real-world signals vary continuously, meaning that their time waveform has the entire number line as its domain. For this reason they are called **continuous time signals**. However, most systems and in particular all digital systems cannot operate with continuous quantities, only discrete quantities. For this reason, it is convenient to convert a continuous time signal into a **discrete time signal**, by quantizing or discretizing its wave form.

The most intuitive way to quantize a time signal is by sampling it at isolated, equally spaced points in time[1]. The newly obtained signal is still a function s of time, but having $\mathbb{Z}$ instead of $\mathbb{R}$ as domain. To distinguish between a continuous time waveform and a discrete time waveform, the latter uses square brackets instead of round brackets. $s[n]$ is related to $s(n)$ in the following way:

$$s[t] = \begin{cases} s(nT_s) & \text{if } n \in \mathbb{Z} \\ \text{undefined} & \text{otherwise} \end{cases}$$

Where T_s is the **sampling period**, the time interval between one instance of sampling and the next. Without knowing what the sampling period is, $s[n]$ is a mere vector of numbers with no semantics.

[1]It's also possible to have unequally spaced samples, but the mathematical underpinning is hard to tackle.

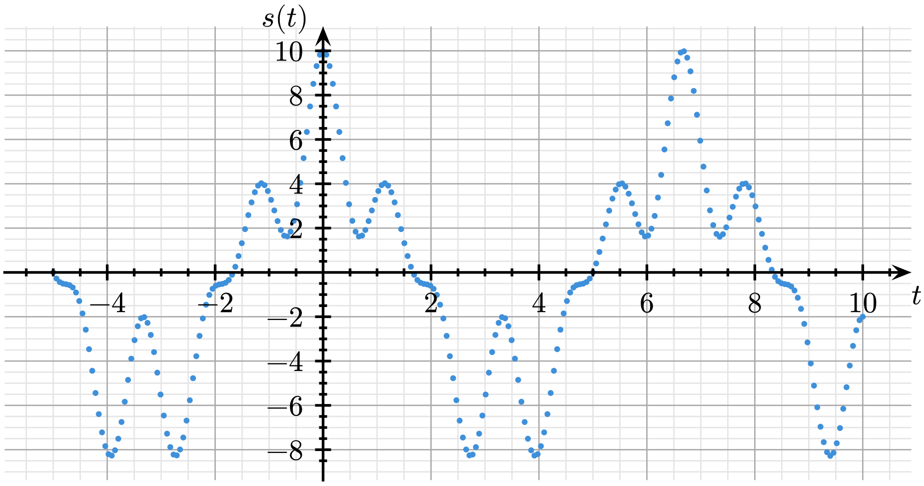


Figure 4: The signal in Figure 3, sampled with a sampling period of 0.06

Quantizing a signal necessarely entail a loss of information, because all (countably infinite) points between two sampled time instants are lost. However, if the number of sampled points is sufficient, the original signal can be reconstructed with a surprinsingly high degree of accuracy. The sampling period should not be too small, or the number of points would require too much memory to be stored. There has to be some tradeoff between sampling accuracy (quality) and space occupied (quantity). There is no silver bullet when choosing a sampling period: the choice is problem-dependent.

Not all signals can be thought of as being time dependent. For example, (still) images clearly do not depend on time. A better representation for an image would be a function of two independent variables $p(x, y)$, representing the spatial coordinates. The output of the function is the intensity of the color, having chosen an appropriate encoding. As for signals depending on time, sampling is also possible for images.

Systems, not only signals, can be represented as functions. Consider a system that has continuous signals both as argument and as return value: this is referred to as a **continuous-time system**. A one-dimensional continuous-time system can be represented as a function T that has a continuous signal $x(t)$ as input and another continuous signal $y(t)$ as output:

$$y(t) = \mathrm{T}\{x(t)\}$$

Consider instead a system that has discrete signals both as argument and as return value: this is referred to as a **discrete-time system**. A one-dimensional discrete-time system can be represented as a function T that has a discrete signal $x(t)$ as input and another discrete signal $y(t)$ as output:

$$y[n] = \mathrm{T}\{x[n]\}$$

Alongside the mathematical representation, systems are also represented using **block diagrams**, diagrams where the each rectangle (block) denotes a sub-component of a system and the arrows denote the flow of operation. To represent a continuous-time system that has a one-dimensional signal $x(t)$ as input and another one-dimensional signal $y(t)$ as output, one would do the following:

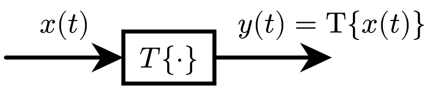


Figure 5: Block diagram representation of $\mathrm{T}\{x(t)\}$.

One example of a system is a **sampler**: a sampler has a continuous signal as input and, given a certain sampling period, returns the vector of sampled points as output. A sampler is often referred to as an **ideal continuous-to-discrete converter** because no real-world sampler can possibly compute the value of the signal at each point in time with perfect accuracy (it is still a valid theoretical model, however). The block diagram representation of a sampler would look like this:

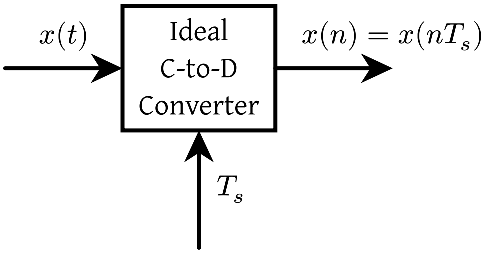


Figure 6: Block diagram representation of a sampler.

1.2. Sinusoids

Signals can have an arbitrary complicated equation, but such equations can be constructed from the ground up starting from simple building blocks[2].

The simplest one-dimensional continuous time signals are the **sinusoidal signals**, or **sinusoids** for short. The equation describing a sinusoidal signal has the following general form:

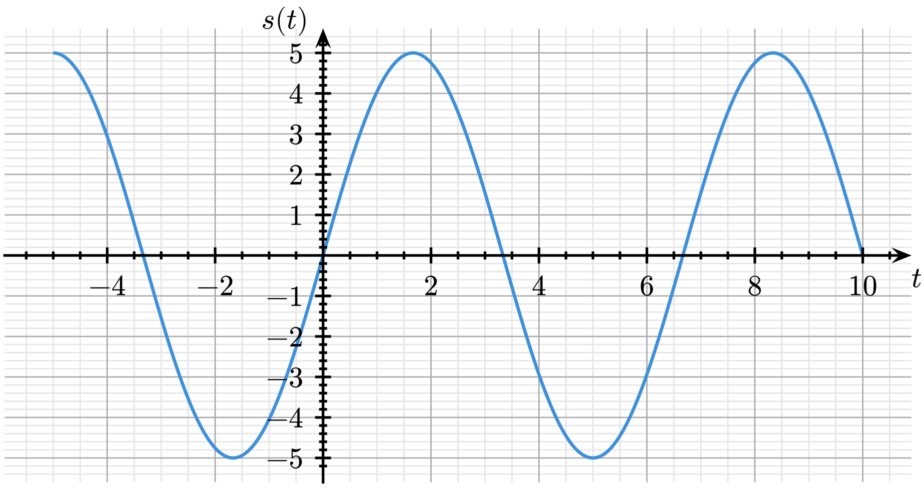
$$s(t) = A\cos(\omega t + \varphi)$$

Where $\cos$ is the trigonometric cosine function. Note that sinusoids are periodical functions (the cosine function multiplied by a constant), hence $s(t) = s(t + 2\pi)$ for any time instant t. As for the other components:

- A is the **amplitude**, the maximum value that the signal can attain (the height of any "spike"). Since $\cos$ oscillates between $+1$ and -1, a sinusoid oscillates between $-A$ and $+A$;
- ω is the **radian frequency**, the number of oscillations that the signal makes every 2π seconds;
- φ is the **phase**, the displacement from 0.

The radian frequency is the **frequency** multiplied by 2π. The frequency f is the number of oscillations that the signal makes every second. The **period** T is the time the signal takes to make an entire oscillation. The frequency and the period are the reciprocal of each other.

$$f = 2\pi\omega \qquad T = \frac{1}{f} = \frac{2\pi}{\omega}$$



[2]Note that, again, there is no difference between a signal and the equation that models it, because the equation captures all information related to the signal. As a matter of fact, the expressions "the signal having equation $s(t)$" and "the signal $s(t)$" are equivalent

Figure 7: Plot of the sinusoidal signal $s(t) = 5\cos(0.3\pi t + 1.5\pi)$.

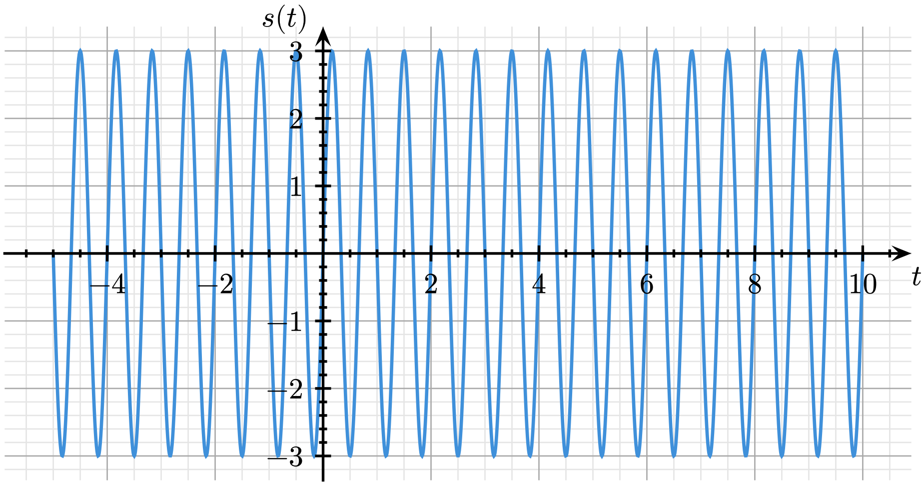


Figure 8: Plot of the sinusoidal signal $s(t) = 3\cos(3\pi t - 0.5\pi)$.

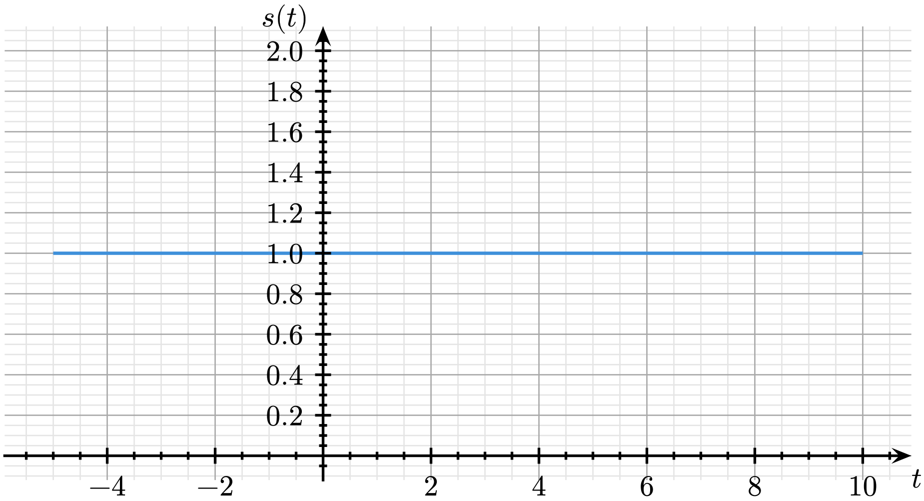


Figure 9: Plot of the sinusoidal signal $s(t) = \cos(0) = 1$ (A valid sinusoid nonetheless).

Technically speaking, it's also possible to use the sine instead of the cosine to construct sinusoidal signals, since $\sin(\omega t) = \cos(\omega t - \pi/2)$ for any ω and t. However, the cosine often takes precedence because $\cos(0) = 1$, while $\sin(0) = 0$, making it much easier in computations.

Plotting a sinusoidal function is relatively straightforward, since knowing its shape inside a single period gives the shape of the entire function. Knowing the expression of a sinusoid $A\cos(\omega t + \varphi)$, to plot it it is necessary to:

- Compute its period;
- Find any value of t that results in a peak;
- Find any value of t that results in 0;

Since ω is known, the period T is just $2\pi/\omega$. To find any t that results in a peak, let it be t_p, it suffices to observe how the peaks of a sinusoid are equal to their amplitude. This means that t_p is given by imposing $s(t_p) = A$:

$$\cancel{A} = \cancel{A}\cos(\omega t_p + \varphi) \Rightarrow \cos(\omega t_p + \varphi) = 1 \Rightarrow \omega t_p + \varphi = 2\pi k \quad \text{with } k \in \mathbb{Z}$$

Since there is no difference in choosing one peak over another, the simplest choice is $k = 0$. Solving for t_p:

$$\omega t_p + \varphi = 2\pi 0 \Rightarrow \omega t_p + \varphi = 0 \Rightarrow t_p = -\frac{\varphi}{\omega}$$

To find a t that results in 0, let it be t_c, it suffices to impose $s(t_c) = 0$ and solve for t_c:

$$0 = A\cos(\omega t_c + \varphi) \Rightarrow \cos(\omega t_c + \varphi) = 0 \Rightarrow \omega t_c + \varphi = \frac{\pi}{2} + 2\pi k \quad \text{with } k \in \mathbb{Z}$$

Choosing $k = 0$ once again:

$$\omega t_c + \varphi = \frac{\pi}{2} + 2\pi 0 \Rightarrow \omega t_c + \varphi = \frac{\pi}{2} \Rightarrow \omega t_c = \frac{\pi}{2} - \varphi \Rightarrow t_c = \frac{\pi}{2\omega} - \frac{\varphi}{\omega}$$

But $-\varphi/\omega$ is the time resulting in a peak. Moreover, since $\omega = 2\pi f$:

$$t_c = \frac{\pi}{2\omega} - \frac{\varphi}{\omega} = \frac{\cancel{\pi}}{4\cancel{\pi} f} + t_p = \frac{1}{4f} + t_p = \frac{T}{4} + t_p$$

Exercise 1.2.1: Plot the sinusoid $s(t) = 20\cos(0.6\pi t - 0.4\pi)$

Solution: The period is given by:

$$T = \frac{2\pi}{\omega} = \frac{2\cancel{\pi}}{0.6\cancel{\pi}} = \frac{2}{0.6} \approx 3.333$$

The time resulting in a peak is given by:

$$t_p = -\frac{\varphi}{\omega} = -\frac{-0.4\cancel{\pi}}{0.6\cancel{\pi}} = \frac{0.4}{0.6} \approx 0.666$$

The time resulting in a 0 crossing is given by:

$$t_c = \frac{T}{4} + t_p = \frac{3.333}{4} + 0.666 \approx 0.834 + 0.666 = 1.5$$

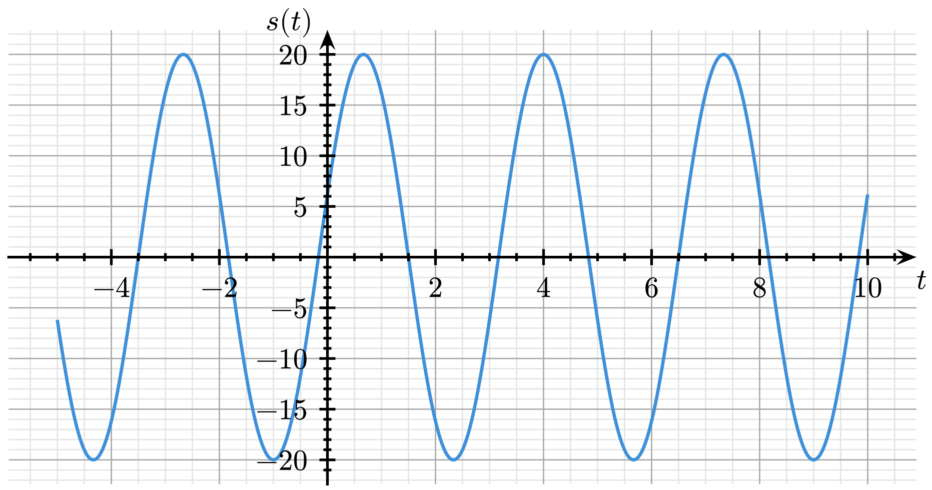


Figure 10: Plot of the sinusoidal signal $s(t) = 20\cos(0.6\pi t - 0.4\pi)$.

□

Note that sketching any sinusoid by hand can be done with continuous strokes, while plotting it with a computer necessarely requires a discretization of the sinusoid. A digital plot is nothing but the quantized sinusoid with a sampling period so small that the function appears continuous.

If the equation of a sinusoid is known, there is no need to introduce a system convert from continuous to discrete: it is sufficient to evaluate the function at sufficiently many distinct and equally spaced time instants. Given a sinusoid $s(t) = A\cos(\omega t + \varphi)$ and a sampling period T_s, the vector of samples is given by solving $s(nT_s) = A\cos(\omega n T_s + \varphi)$ for a sufficient number of n.

In general, a computer plotting device performs what's called a **linear interpolation**, connecting adjacent points with a straight line: for sufficiently small segments, a straight line and a smooth curve are indistinguishable. Intuitively, having a greater number of points will result in a better approximation. In turn, a smaller sampling period will increase the number of points, because more points can "fit" into one period. This also means that a sinusoid with a smaller period (higher frequency) requires a smaller sampling period to achieve the same accuracy. As it will be clear later, there are techniques that go beyond linear interpolation and that, under the right assumptions, can reconstruct the original sinusoid with perfect accuracy.

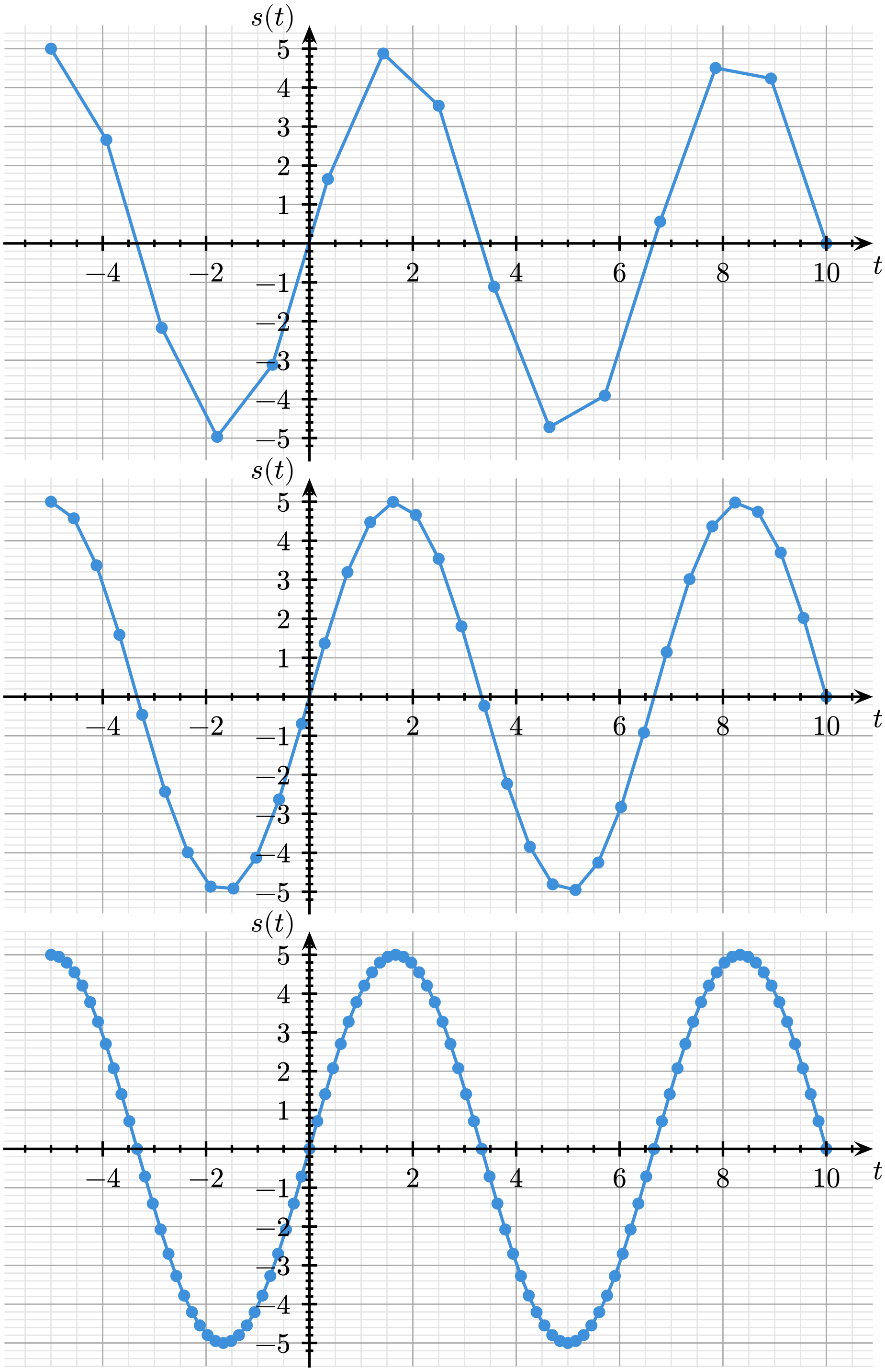


Figure 11: Plot of $s(t) = 5\cos(0.3\pi t + 1.5\pi)$ with three different choices of the sampling period.

If so desired, it's also possible to go the other way around, from the plot of a sinusoid to its equation. To do so:

- Determine the period, which is the time interval between any two points having the same value of $s(t)$. The simplest choices are either two adjacent 0 crossings or two adjacent peaks;
- Compute the radian frequency from the period as $2\pi/T$;
- Determine the amplitude, which is just the height of any peak;
- Compute the phase by using the formula $t = -\varphi/\omega$ backwards, solving for φ instead of t. The value of t is t_p, the time instant closest to 0 that results in a peak.

Exercise 1.2.2: What is the equation of the following sinusoid?

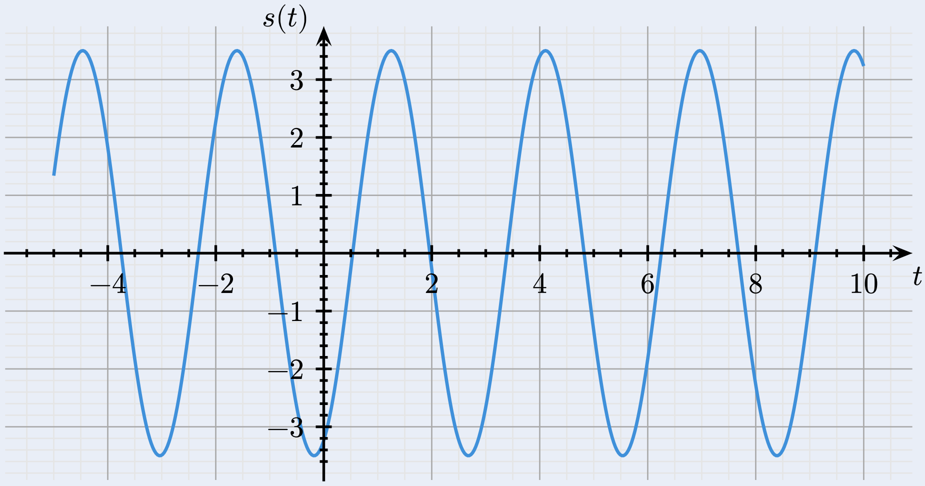


Figure 12: Plot of a sinusoid, whose equation is to be determined.

Solution:

- The sinusoid has its first peak (starting from 0) at about 1.25 and its second peak at about 4.1, hence its period is $4.1 - 1.25 = 2.85$;
- Its radian frequency is $2\pi/2.85 \approx 0.7\pi$;
- The amplitude is about 3.5;
- The phase is given by $\varphi = -\omega t_p = -0.7\pi \cdot 1.25 = -0.875$.

Which means that the sought for equation is:

$$s(t) = 3.5\cos(0.7\pi t - 0.875\pi)$$

□

Most real waves can hardly be modeled by a simple sinusoid functions, since some attenuation over time or over distance has to be taken into account. Given a real value α that represents the dampening of the strength of the signal, a more accurate waveform has the following equation:

$$s(t) = A(t)\cos(\omega t + \varphi) = Ae^{-t/\alpha}\cos(\omega t + \varphi)$$

Where the amplitude has now a time dependence, instead of being a constant. Since the negative exponential is a decreasing function, the amplitude of the signal will decrease over time.

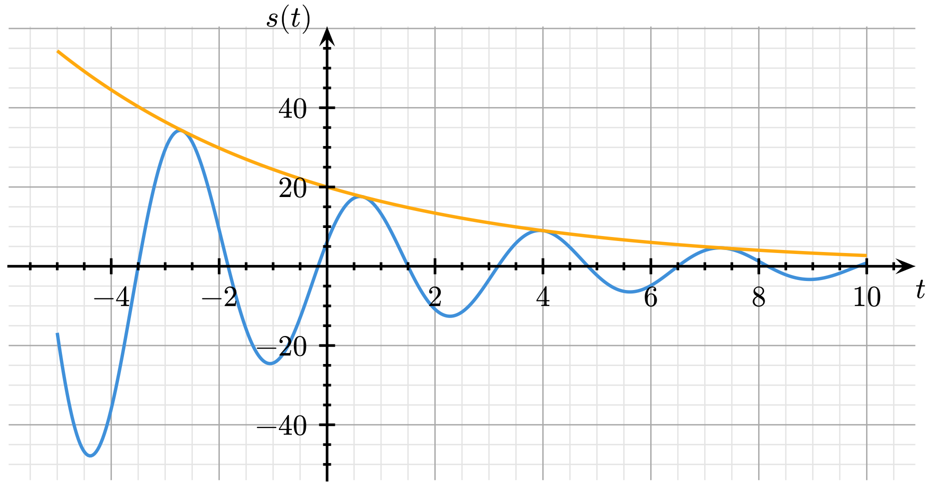


Figure 13: In blue, the plot of the sinusoidal signal $s(t) = 20e^{-t/5} \cos(0.6\pi t - 0.4\pi)$, using a sampling period of $T_s = 0.03$. In orange, the time-dependent amplitude $A(t) = 20e^{-t/5}$.

Another way to construct elaborate signals is to sum more sinusoids. The problem of summing sinusoids is that computing the sum of cosines (or sines, for that matter) is tedious. However, it's possible to rewrite the sinusoid in a slightly different form that aides in performing mathematical manipulations.

Recall how a complex number in polar form can be converted into the exponential form[3]:

$$re^{j\theta} = r\cos(\theta) + rj\sin(\theta)$$

Given an amplitude A, a radian frequency ω_0 and a phase φ, a **complex exponential signal** is defined as:

$$z(t) = Ae^{j(\omega_0 t + \varphi)} = A\cos(\omega_0 t + \varphi) + Aj\sin(\omega_0 t + \varphi)$$

Whose magnitude is the constant A and whose argument is the time-dependent expression $\omega_0 t + \varphi$. Both the real part and the imaginary part of the complex number represent a sinusoid; the two sinusoids have the same amplitude and frequency, differing only by a phase factor of $\pi/2$. This is because:

$$z(t) = A\cos(\omega_0 t + \varphi) + Aj\sin(\omega_0 t + \varphi) = A\cos(\omega_0 t + \varphi) + Aj\cos\left(\omega_0 t + \varphi - \frac{\pi}{2}\right)$$

In particular, notice how:

$$s(t) = \Re\{z(t)\} = \Re\left\{Ae^{j(\omega_0 t + \varphi)}\right\} = A\cos(\omega_0 t + \varphi)$$

Which means that the sinusoids so far considered are just the real part of complex exponential signals.

[3] It is customary to use j instead of i to denote the imaginary unit when dealing with signals, because i is often used to denote the intensity of a signal. This notation is commonplace in all fields of engineering.

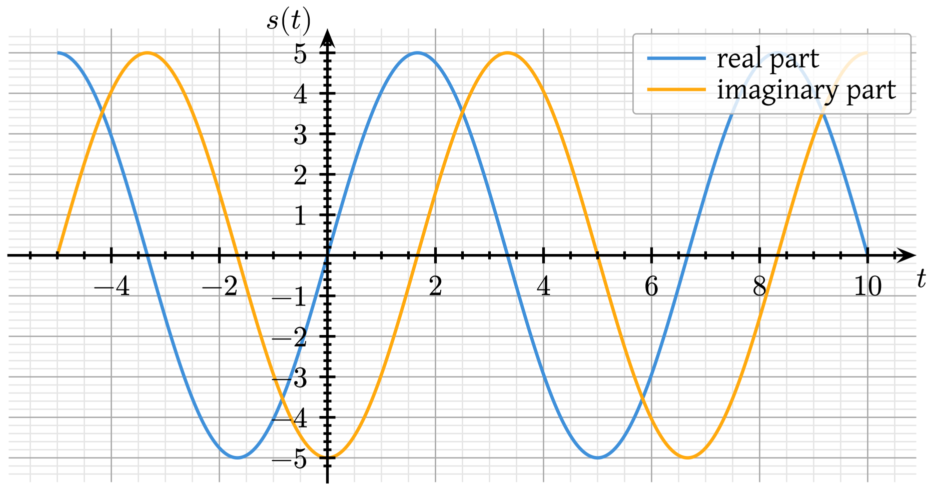


Figure 14: Plot of the complex exponential $z(t) = 5\cos(0.3\pi t + 1.5\pi) + 5j\cos(0.3\pi t + 1.5\pi - 0.5\pi)$, with both the real part and the imaginary part. The two differ by $\pi/2$.

Recall that the geometric interpretation of the multiplication of two complex numbers is a rotation in the complex plane (angles are added and magnitudes are scaled). The complex exponential signal $z(t)$ can be written as the product of two complex numbers:

$$z(t) = Ae^{j(\omega_0 t+\varphi)} = Ae^{j\omega_0 t+j\varphi} = Ae^{j\omega_0 t}e^{j\varphi} = Xe^{j\omega_0 t}$$

Where $X = Ae^{j\varphi}$ is called the **complex amplitude**, or **phasor**[4]. This means that $z(t)$ is the product of a complex constant (the phasor) and a complex-valued function dependent on time $e^{j\omega_0 t}$.

By writing $\theta(t) = \omega_0 t + \varphi$, that is by expliciting the time dependence of the angle, it's also possible to write a complex exponential signal as:

$$z(t) = Xe^{j\omega_0 t} = Ae^{j\theta(t)}$$

At a given time instant t, the value of the complex exponential signal $z(t)$ is a complex number whose magnitude is A and whose angle is $\theta(t)$.

Consider the representation of $z(t) = Ae^{j\theta(t)}$ in the complex plane: as t increases, the complex number only changes in angle but not in magnitude, as the time dependency is only present in the angle. This means that the corresponding vector in the complex plane keeps rotating without ever changing in magnitude. This is why a complex exponential signal is also called **rotating phasor**.

The "speed" at which the vector rotates, meaning how much area of the plane is traversed as time increases, depends on the radian frequency ω_0: the higher ω_0, the "faster" the rotation. Moreover, the sign of the radian frequency determines the direction of the rotation: if ω_0 is positive, the rotation is counterclockwise, since the angle θ increases; if ω_0 is negative, the rotation is clockwise, since the angle θ decreases. Rotating phasors are said to have **positive frequency** if they rotate counterclockwise, and **negative frequency** if they rotate clockwise.

A rotating phasor makes one complete revolution every time the angle $\theta(t)$ changes by 2π radians. The time it takes to make one revolution is also equal to the period T_0 of the complex exponential signal, so:

$$\omega_0 T_0 = (2\pi f_0)T_0 = 2\pi \Rightarrow T_0 = \frac{1}{f_0}$$

Notice that the phase φ defines where the phasor is pointing when $t = 0$.

[4]The term phasor is common in electrical circuit theory

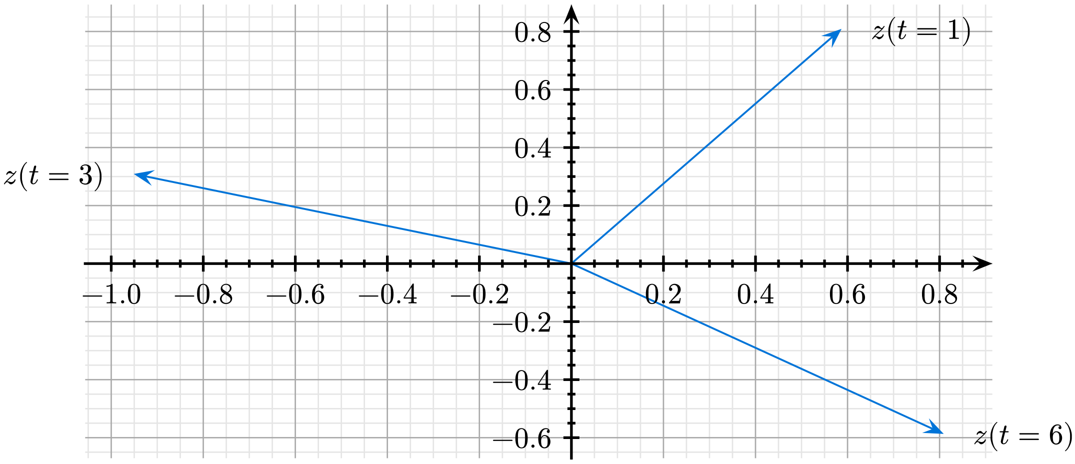


Figure 15: From the complex exponential signal $z(t) = 5e^{j(0.3\pi t + 1.5\pi)}$, discarding the phasor one gets $e^{j0.3\pi t}$.

Recall the inverse Euler formula for the cosine:

$$\cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2} = \frac{e^{j\theta} + \left(e^{j\theta}\right)^*}{2}$$

Since a sinusoid is in the form $s(t) = A\cos(\omega_0 t + \varphi)$, it's also possible to write is as:

$$s(t) = A\cos(\omega_0 t + \varphi) = A\left(\frac{e^{j(\omega_0 t + \varphi)} + e^{-j(\omega_0 t + \varphi)}}{2}\right) = \frac{A}{2}e^{j(\omega_0 t + \varphi)} + \frac{A}{2}e^{-j(\omega_0 t + \varphi)} =$$

$$= \frac{A}{2}e^{j\omega_0 t}e^{j\varphi} + \frac{A}{2}e^{-j\omega_0 t}e^{-j\varphi} = \frac{1}{2}Xe^{j\omega_0 t} + \frac{1}{2}\left(Xe^{j\omega_0 t}\right)^* = \frac{1}{2}z(t) + \frac{1}{2}z^*(t) = \mathfrak{R}\{z(t)\}$$

This formula has an interesting interpretation. The sinusoid $s(t)$ is actually composed of a positive frequency complex exponential $\frac{1}{2}Xe^{j\omega_0 t}$ and a negative frequency complex exponential $\frac{1}{2}\left(Xe^{j\omega_0 t}\right)^*$. The two have the same amplitude, the same phase in modulus and the same radian frequency in modulus. In other words, any sinusoid can be represented as the sum of two complex rotating phasors that are rotating in opposite directions (the angles have opposite sign) starting from phasors that are complex conjugates of each other.

As already hinted at, complex exponential signals allow one to compute the sum of sinusoids with ease. This is remarkably true when summing sinusoids having the same radian frequency.

Theorem 1.2.1 (Phasor addition rule): Let $A_k \cos(\omega_0 t + \varphi_k)$ with $k \in \{1, 2, ..., n\}$ be a family of n sinusoids, all having the same radian frequency. Then:

$$\sum_{k=1}^{n} A_k \cos(\omega_0 t + \varphi_k) = A\cos(\omega_0 t + \varphi)$$

Where:

$$A = \left\| \sum_{k=1}^{n} A_k e^{j\varphi_k} \right\| \qquad \varphi = \arg\left(\sum_{k=1}^{n} A_k e^{j\varphi_k}\right)$$

Proof: Recall that, for any sinusoid:

$$A\cos(\omega_0 t+\varphi)=\Re\left\{Ae^{j(\omega_0 t+\varphi)}\right\}$$

Then:

$$\sum_{k=1}^{n} A_k \cos(\omega_0 t+\varphi_k)=\sum_{k=1}^{n}\Re\left\{A_k e^{j(\omega_0 t+\varphi_k)}\right\}=\sum_{k=1}^{n}\Re\left\{A_k e^{j\omega_0 t}e^{j\varphi_k}\right\}$$

But the sum of the real part of n complex numbers is the real part of their sum, hence:

$$\sum_{k=1}^{n} A_k \cos(\omega_0 t+\varphi_k)=\sum_{k=1}^{n}\Re\left\{A_k e^{j\omega_0 t}e^{j\varphi_k}\right\}=\Re\left\{\sum_{k=1}^{n} A_k e^{j\omega_0 t}e^{j\varphi_k}\right\}=$$
$$=\Re\left\{e^{j\omega_0 t}\sum_{k=1}^{n} A_k e^{j\varphi_k}\right\}$$

The sum of complex numbers is itself a complex number. With $Ae^{j\varphi}=\sum_{k=1}^{n} A_k e^{j\varphi_k}$:

$$\sum_{k=1}^{n} A_k \cos(\omega_0 t+\varphi_k)=\Re\left\{e^{j\omega_0 t}\sum_{k=1}^{n} A_k e^{j\varphi_k}\right\}=\Re\left\{e^{j\omega_0 t}Ae^{j\varphi}\right\}=$$
$$=\Re\left\{Ae^{j(\omega_0 t+\varphi)}\right\}=A\cos(\omega_0 t+\varphi)$$

$\square$

The only caveat to using Theorem 1.2.1 is that the sum $\sum_{k=1}^{n} A_k e^{j\varphi_k}$, in order to be computed, requires the k phasors to be converted in rectangular form.

Exercise 1.2.3: Compute the sum of the two sinusoids:

$$s_1(t)=1.7\cos\left(20\pi t+\frac{70\pi}{180}\right) \qquad s_2(t)=1.9\cos\left(20\pi t+\frac{200\pi}{180}\right)$$

Solution: $s_1(t)$ and $s_2(t)$, written as the real part of a complex exponential, are:

$$s_1(t)=\Re\left\{1.7e^{j(20\pi t+70\pi/180)}\right\}=\Re\left\{1.7e^{j20\pi t}e^{j70\pi/180}\right\}$$
$$s_2(t)=\Re\left\{1.9e^{j(20\pi t+200\pi/180)}\right\}=\Re\left\{1.9e^{j20\pi t}e^{j200\pi/180}\right\}$$

Which gives the two phasors:

$$X_1=1.7e^{j\frac{70\pi}{180}}=1.7\left(\cos\left(\frac{70\pi}{180}\right)+j\sin\left(\frac{70\pi}{180}\right)\right)\approx 1.7(0.34+j0.94)=0.58+j1.60$$
$$X_2=1.9e^{j\frac{200\pi}{180}}=1.9\left(\cos\left(\frac{200\pi}{180}\right)+j\sin\left(\frac{200\pi}{180}\right)\right)\approx 1.9(-0.94-j0.34)=-1.79-j0.65$$

And their sum is:

$$X=X_1+X_2=(0.58+j1.60)+(-1.79-j0.65)=-1.21+j0.95$$

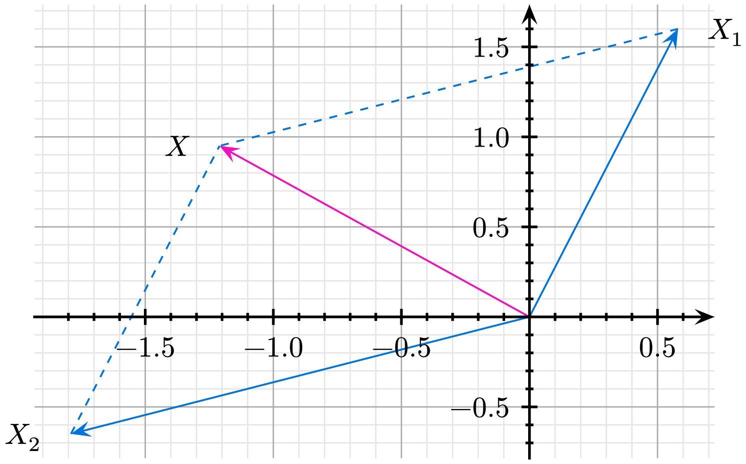


Figure 16: The two phasors X_1 and X_2 and their sum X, plotted on the complex plane.

The magnitude and the argument of X are:

$$\|X\| = \sqrt{(-1.21)^2 + (0.95)^2} \approx 1.54 \qquad \arg(X) = \tan^{-1}\left(\frac{0.95}{-1.21}\right) \approx \frac{142\pi}{180}$$

Which means that:

$$s(t) = s_1(t) + s_2(t) = 1.54\cos\left(20\pi t + \frac{142\pi}{180}\right)$$

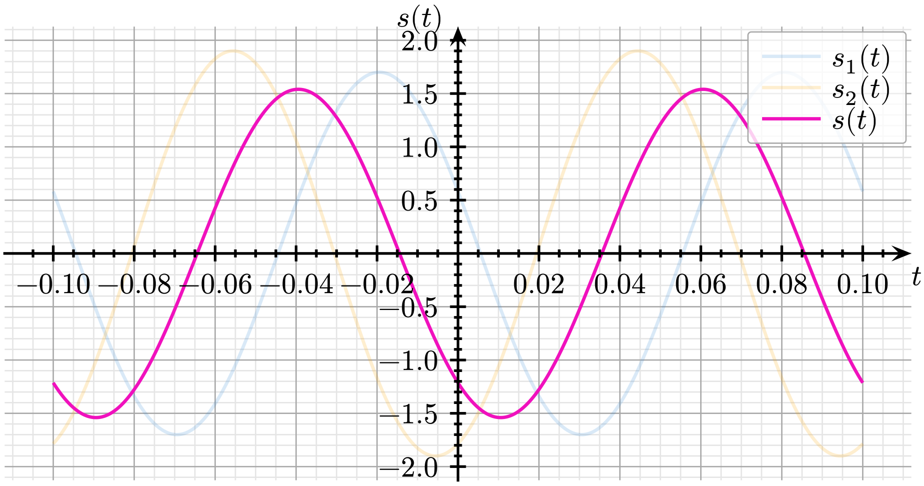


Figure 17: Plot of the sinusoids $s_1(t) = 1.7\cos(20\pi t + 70\pi/180)$ and $s_2(t) = 1.9\cos(20\pi t + 200\pi/180)$ and of their sum $s(t) = 1.54\cos(20\pi t + 142\pi/180)$.

$\square$