This page first defines the sinusoidal functions that can be used to decompose waves on the sea surface into sums of sines and cosines. Sampling of sea surfaces at a finite number of points is then discussed.

Sinusoidal Wave Representations

The sea surface can be described using a Cartesian coordinate system. Let $\mathbf{x}$ be a unit vector chosen to point in the downwind direction, and let $\mathbf{z}$ be directed upward (from the water to the air) normal to the mean sea surface at $z = 0$. Then $\mathbf{y} = \mathbf{z} \times \mathbf{x}$ is in the cross-wind direction. $z(x, y, t)$ is the surface elevation in meters at spatial location (x, y) and time t.

Fourier proved in the early 1800's that, under very general mathematical conditions, an arbitrary function can be written as a sum of sines and cosines of different amplitudes and wavelengths. The sea surface elevation $z(x, y, t)$ is usually such a function. An exception is a sea surface with waves curling over and breaking on a beach. There can then be multiple air-water surfaces for a given (x, y, t): the point on the surface that sees the full sky, the point above a surfer's head as the wave curls over her, and the point on the water that supports the surf board. However, for a sea surface without breaking waves, there is only one air-water surface for a given location and time, which can be described as a sum of sinusoids.

The terms and notation needed to describe propagating sinusoidal waves are defined as follows:

- A is the *amplitude* (in units of meters) of the wave. This is one-half the distance from the trough to the crest of a sinusoidal wave. Oceanographers often talk about the *wave height*, H, which is the distance from a wave trough to a wave crest. Wave height is what most people visualize when they talk about how "big" a wave is. For a sinusoidal wave, $H = 2A$.

- T is the *temporal period* (seconds), usually called just the period. This is the time it takes for one wave crest to pass by a fixed point, i.e. for the argument of a sinusoid to go through 2π radians.

- $f = 1/T$ is the *temporal frequency* (1/seconds), usually called just the frequency. This is how many waves pass by a fixed point per second.

- $\omega = 2\pi/T = 2\pi f$ is the *angular frequency* (radians/second). This is the number of waves passing a fixed point in 2π seconds.

- Λ is the *wavelength*, or spatial period, (meters). This is the distance from one wave crest to the next.

- $\nu = 1/\Lambda$ is the *wavenumber* or *spatial frequency* (1/meters). This is the number of wavelengths per meter.

- $k = 2\pi/\Lambda = 2\pi\nu$ is the *angular wavenumber* or *angular spatial frequency* (radians/meter). This is the number of wavelengths in 2π meters.

- ϕ is the phase (nondimensional). This fixes the location of a wave crest relative to the origin of a coordinate system.

There is no uniformity in the literature. People often refer to "the frequency" without specifying whether they mean f or ω, and "wavenumber" or "spatial frequency" may mean either ν or k. People use $\nu, \tilde{\nu}$, or σ for the wavenumber, some use σ for ω, and so on. You just have to figure it out on a case by case basis. For pedagogic purposes, we will start with wavenumber ν as the measure of spatial frequency, and then switch to the more common angular spatial frequency k. This presentation will use ω as the measure of temporal frequency.

Now consider a one-dimensional sea surface with elevation $z(x,t)$ over a spatial region of length L, e.g. $0 \le x \le L$ or $-L/2 \le x \le L/2$. According to Fourier, $z(x,t)$ can be written as a sum of sinusoids, the n^{th} one of which is

$$z_n(x,t) = A_n \cos\left(\frac{2\pi n x}{L} + \phi_n \pm \frac{2\pi n t}{T}\right), \tag{1}$$

where $n = 0, 1, 2, ...$ is simply an index for which sinusoid is being considered. (Note that we could just as well write a sine here, which corresponds to a different value for the phase.) For $n = 0$ this reduces to a constant offset $z_0(x,t) = A_0 \cos(\phi_0)$, which is usually taken to be mean sea level and set to a reference value of 0 via setting $A_0 = 0$. For the moment, take $t = 0$, i.e., we have a snapshot of the ocean surface at time zero. This sinusoid has a wavelength of $\Lambda_n = L/n$, i.e., the cosine returns to the same value after a distance of $x = L/n$. It is a pure cosine in the chosen coordinate system if the phase is 0 or an even integer multiple of $\pm\pi$. If the phase is an odd integer multiple of $\pm\pi/2$ it is a sine. If time is included, the wave is propagating in the $+x$ direction if the time term is $-2\pi nt/T$; the cosine returns to its initial value after a time of $t = T/n$. A wave propagating in the $-x$ direction is described by a $+2\pi nt/T$ term. The physical angular frequency $\omega_n \equiv 2\pi n/T$ is always positive, but mathematically we can write just $+\omega_n t$ in the equation and then view a wave propagating in the $+x$ direction as having a negative temporal frequency.

With this interpretation of $\omega_n t$, Eq. likesection1 is conveniently written as

$$z_n(x,t) = A_n \cos(k_n x + \phi_n + \omega_n t), \tag{2}$$

where $k_n = 2\pi n/L = 2\pi/\Lambda_n$ is the angular wavenumber of the n^{th} wave of wavelength $\Lambda_n = L/n$. Likewise, $\omega_n = 2\pi n/T = 2\pi/T_n =$ is the angular frequency of the n^{th} wave with period $T_n = T/n$. It may be intuitively easier to think of wavelengths per meter than of wavelengths per 2π meters, but the convenience of writing k rather than $2\pi/\Lambda$ or $2\pi\nu$ makes k rather than ν the spatial frequency variable used in most publications. A similar comment holds for ω vs. $2\pi/T = 2\pi f$, so ω rather than f is the common choice for the temporal frequency variable.

If we take $t = 0$, or just combine the time term with the phase, expanding the cosine in Eq. (likesection2) gives an equivalent form

$$z_n(x) = a_n \cos(k_n x) + b_n \sin(k_n x), \tag{3}$$

where $a_n = A_n \cos\phi_n$ and $b_n = -A_n \sin\phi_n$, $n = 1, 2, ...$, $a_0 = A_0$, and $b_0 = 0$. This equation can be written in terms of complex numbers:

$$z_n(x) = c_{+n} e^{ik_n x} + c_{-n} e^{-ik_n x} \,, \tag{4}$$

where $i = \sqrt{-1}$ and

$$c_{+n} = (a_n - ib_n)/2, \quad c_{-n} = (a_n + ib_n)/2, \quad \text{and} \quad c_0 = a_0/2 \,. \tag{5}$$

Recalling that $e^{\pm i\theta} = \cos\theta \pm i \sin\theta$, we see that the imaginary parts of Eq. (likesection4) add to zero, so that $z_n(x)$ is still a real variable even though it is written in complex form. Also note that $c^*_{+n} = c_{-n}$, where c^* denotes the complex conjugate. (Complex conjugation means replace i by $-i$ in all terms.) Although two complex numbers in general contain four independent real numbers, these c_{+n} and c_{-n} pairs contain only two independent numbers, a_n and b_n.

We thus have three ways to describe a sinusoidal wave: (1) The cosine of Eq. (likesection2) defined by an amplitude and a phase; (2) the cosine and sine of Eq. (likesection3) defined by two amplitudes, or (3) the complex exponentials of Eq. (likesection4) defined by two complex numbers containing two independent real numbers. These equations all give the same $z_n(x)$, so we are free to choose whichever form is mathematically convenient for the problem at hand. For visualization of sea surfaces, the real forms (likesection2) or (likesection3) are convenient, but the calculations on the next pages are most conveniently carried out using complex exponentials.

Returning now to Fourier, linear wave theory says that we can write the sea surface elevation as a Fourier series

$$z(x) = \sum_{n=0}^{\infty} z_n(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(k_n x) + b_n \sin(k_n x) \right] \,, \tag{6}$$

which in conjunction with Eq. (likesection4) can be rewritten as

$$z(x) = \sum_{n=-\infty}^{\infty} c_n e^{ik_n x} \,. \tag{7}$$

As previously noted, the a_0 or c_0 term is usually set to 0. This equation is the mathematical heart of our subsequent descriptions sea surfaces.

Although the physics that leads to a given sea surface in general involves non-linear interactions between waves of different wavelengths and periods, the shape of the resulting sea surface can be written as a linear sum of sinusoids of different frequencies. Owing to the orthogonality of sines and cosines for different n values as defined above, these wave components $z_n(x)$ are independent of each other. (Orthogonality means that $\int_0^{2\pi} \cos(mx)\cos(nx)dx = 0$ if $m \neq n$, etc.) Thus a description of a time-dependent surface based on an expansion like that of Eq. (likesection6) (including the time-dependent terms as seen in Eq. (likesection2)) cannot be used to predict the nonlinear evolution of a sea surface from a given initial condition. To do that, you must return to the world of hydrodynamics in all of its nonlinear glory. For our present purposes, we only want to model the shapes of sea surfaces, not predict their hydrodynamic development from an initial state, so linear wave representations are adequate.

Sampling

We next consider the implications of *sampling* the sea surface at a finite number of discrete locations or times. It is usually the case when we want to measure the sea surface elevation that we either make measurements at discrete locations $x_r, r = 0, 1, ..., N-1$ at a given time, or at discrete times $t_r, r = 0, 1, ..., N-1$ at a given location.

For a specific example, consider a region of sea surface $L = 10$ m long where we take $N = 8$ evenly spaced samples. This gives a measurement every $\Delta x = L/N = 1.25$ m at a given time. We are going to describe this surface as a sum of sinusoids. Figure figure1 illustrates the first few sinusoids shown as cosines. The cosines in the figure all have the same amplitude and each is offset vertically for ease of viewing. The dots show the sampled values for each cosine component.

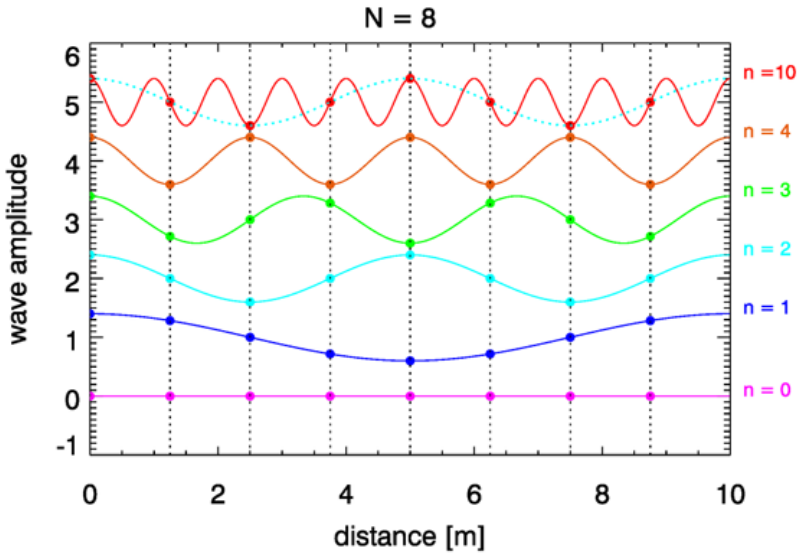

Figure 1: Illustration of sampling the harmonics of a given wave and of aliasing. The vertical dotted lines show the sampling locations, and the dots show the sampled values for the various cosine components.

The longest wave that can fit into a region of length L has a wavelength of $\Lambda = L$. This is called the fundamental wave or first harmonic. This wave is shown by the blue $n = 1$ curve in the figure. Note that the $n = 4$ cosine in the figure has a period of $2\Delta x$. The smallest wave that can be captured correctly in a sampling scheme has a period of twice the sampling interval. This called the two-point wave or two-point oscillation. Note that a sine term with period $2\Delta x$ would be sampled at arguments of $0, \pi, 2\pi, ...$ where the sine is zero and is thus not detectable. The spatial frequency $1/(2\Delta x)$ (or temporal frequency $1/(2\Delta t)$ if sampling in time at a given location) is called the spatial (or temporal) *Nyquist frequency*. Waves with higher frequencies than the Nyquist frequency are still sampled, as illustrated by the dots on the wave with $n = 10$. Note, however, that the pattern of sampled points for the $n = 10$ wave is exactly the same as for the $n = 2$ wave. If all we have are the measured points, we cannot distinguish between the $n = 10$ and the $n = 2$ waves in this example. This

illustrates the phenomenon of *aliasing*. In general, a wave with a frequency greater than the Nyquist frequency (i.e., a wavelength or period less that twice the sampling interval) is still sampled, but it appears as though the high frequency wave is a wave with a frequency less than the Nyquist frequency. The information about the high frequency (short wavelength) wave is added to or aliased into a lower frequency (longer wavelength) wave, thereby giving incorrect information about the lower frequency wave. Aliasing places a severe constraint on any sampling scheme. We can correctly sample only waves with wavelengths that lie between the size of the spatial region, L, and twice the size of the sampling interval. (In the temporal setting, the limits are the length of sampling time and twice the temporal sampling interval.)

The relation between a sampled frequency f_s greater than the Nyquist frequency, the sampling rate or Nyquist frequency $f_{\rm Ny}$, and the frequency f_a receiving the aliased signal is

$$f_a = |m f_{\rm Ny} - f_s|$$

where m is the closest integer multiple of the sampling rate to the signal being sampled. In the example of Fig. figure1, $f_s = f_{n=10} = 10$ waves per 10 meters $= 1\ {\rm m}^{-1}$, $f_{\rm Ny} = 0.4$, and either $m = 2$ or $m = 3$ gives $f_a = |2 \times 0.4 - 1| = 0.2$ (or $f_a = |3 \times 0.4 - 1| = 0.2$). Thus the $n = 10$ wave is aliased into the $n = 2$ wave, which has a frequency of 2 waves per 10 meters, or $0.2\,{\rm m}^{-1}$, just as seen in the figure. Note that many other frequencies are also aliased into this frequency. For example, a wave with $f_s = 1.4$, corresponding to $n = 14$, also gives $f_a = 0.2$ with $m = 4$, and so on.