Aliasing and Non-Linear Instability

Note: There exist several sign errors within the course textbook in its mathematical formulation for aliasing. Any discrepancies between the notes below and the course textbook should be reconciled in favor of the notes below, which are believed to be correct.

Aliasing and Non-Linear Instability

Our consideration of numerical instability to this point has emphasized determining the stability criteria for linear forcing terms. Generally speaking, linear numerical stability is said to exist when the amplitude of the solution does not grow exponentially with time (e.g., $\left|e^{\omega_I t}\right| < 1$).

The primitive equations, however, contain non-linear forcing terms. Consider, for instance, a one-dimensional advection equation for the zonal velocity *u*:

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x}$$

The stability of this equation can be evaluated as before, from which a stability criterion may be obtained. This stability criterion is dependent upon the chosen combination of temporal and spatial finite differencing schemes and must be adhered to in order to ensure numerical stability.

However, there exists a second potential source of non-linear instability that must be considered when determining numerical stability. **Aliasing** occurs when two waves represented on a model grid interact and produce fictitious waves and an erroneous redistribution of energy within wave space. Aliasing can arise in any model that discretizes the primitive equations with finite difference approximations in an Eulerian framework. In the following, we develop a framework for aliasing that is independent of the chosen temporal and spatial finite differencing schemes.

Note that aliasing does not impact models that use semi-Lagrangian methods, wherein non-linear terms are encapsulated within the total derivative-based (i.e., flow-following) formulation of the primitive equations. Aliasing also does not impact models that use spectral methods, wherein model variables and their partial derivatives are treated analytically and potentially troublesome wave interactions are not permitted. The absence of aliasing with such methods is one of several reasons why they have gained widespread use in operational numerical weather prediction.

The byproduct of aliasing is the accumulation of erroneous wave energy at short wavelengths (generally speaking, $\leq 4\Delta x$), which can lead to the model solution becoming unstable with time. Previously, we demonstrated that short wavelengths have large truncation error (poorly-resolved), are among those whose amplitudes grow most rapidly if the linear numerical stability criterion is violated, and are associated with significant departures of phase speed and group velocity from the true advective velocity (numerical dispersion). Thus, aliasing is yet another reason why shorter wavelengths are particularly problematic for grid-based numerical weather prediction and thus why implicit or explicit damping of such wavelengths can be beneficial.

Analytic Framework for Aliasing

We follow the example of the course text. Consider the non-linear one-dimensional advection equation above. As we have done previously, let us assume a wave-like solution for *u*. For simplicity, assume that this wave-like solution is given by the linear superposition of cosine waves, e.g.,

$$u = \sum_{m=0}^{\infty} a_m \cos(k_m x)$$

rather than sine and cosine waves as before. Here, wavenumber $k_m = 2\pi m/L$, where *m* is a zonal wavenumber and *L* is the domain length. Note the slight difference in how this *k* is defined relative to that in our lecture on linear numerical stability, where $k = 2\pi/L$ (where *L* was wavelength). Here, k_m is defined specific to a given wavelength. The ratio of *m* to *L* is the inverse wavelength, such that the ratio of *L* to *m* defines the wavelength (e.g., *m* = 1 defines a wave with wavelength *L*, *m* = 2 defines a wave with wavelength *L*/2, etc.). In other words, *m* is the number of waves over the domain length *L*. Thus, this formulation for k_m is functionally equivalent to that for *k* before.

The first partial derivative of *u* with respect to *x* can be obtained analytically and is given by:

$$\frac{\partial u}{\partial x} = -\sum_{m=0}^{\infty} a_m k_m \sin(k_m x)$$

Consequently,

$$-u\frac{\partial u}{\partial x} = \left(\sum_{m=0}^{\infty} a_m \cos(k_m x)\right)\left(\sum_{n=0}^{\infty} a_n k_n \sin(k_n x)\right)$$

Note that the indices *m* and *n* may be switched without changing the result. The separate notation for each term (*m* for *u*, *n* for its partial derivative) is used to indicate that a wave in *u* of a given wavelength may interact with a wave in $\partial u/\partial x$ of another wavelength.

Or, expanding the summation notation,

$$-u\frac{\partial u}{\partial x} = \left(a_0 + a_1 \cos(k_1 x) + a_2 \cos(k_2 x) + \ldots + a_\infty \cos(k_\infty x)\right) * \\ \left(a_1 k_1 \sin(k_1 x) + a_2 k_2 \sin(k_2 x) + \ldots + a_\infty k_\infty \sin(k_\infty x)\right)$$

In the above, for $m = 0$, $\cos(k_m x) = \cos(0) = 1$, so that $a_0 \cos(k_0 x) = a_0$. For $n = 0$, $\sin(k_n x) = \sin(0) = 0$, so that $a_0 k_0 \sin(k_0 x) = 0$.

Generally, the product of any two waves can be expressed as:

$$a_m a_n k_n \sin(k_n x)\cos(k_m x) \qquad \text{or, equivalently,} \qquad a_n a_m k_m \sin(k_m x)\cos(k_n x)$$

We can simplify this expression. Note that sin c cos d, where c and d are generic variables, can be expressed using a trigonometric identity as follows:

$$\sin c \cos d = \frac{\sin(c+d) + \sin(c-d)}{2}$$

For $c = k_n x$ and $d = k_m x$, we obtain:

$$a_m a_n k_n \sin(k_n x)\cos(k_m x) = \frac{1}{2} a_m a_n k_n \sin((k_n + k_m)x) \sin((k_n - k_m)x)$$

Or, substituting for k_n and k_m,

$$\frac{1}{2} a_m a_n k_n \sin\left(\frac{2\pi}{L}(n+m)x\right) \sin\left(\frac{2\pi}{L}(n-m)x\right)$$

There exist two waves defined by the above – the $n + m$ wave and the $n - m$ wave. Note, as before, that the indices m and n may be swapped without changing the result.

In physical space, where all wavenumbers are possible, this is not a problem. However, on a model grid, only waves of wavelength $2\Delta x$ and larger may be represented. Recall that the ratio of L to m describes a wave's wavelength. Consider a one-dimensional model grid with j_{max} grid points, such that $j_{max}\Delta x = L$. Thus, for the $2\Delta x$ wave, we can determine m as follows:

$$\frac{j_{\max}\Delta x}{m} = 2\Delta x$$

Solving for m, we obtain $j_{max}/2$. This represents the *maximum* value of $n + m$ that may be represented on a model grid. You can prove this by considering other wavelengths longer than $2\Delta x$ in the above – e.g., for the $3\Delta x$ wave, m equals $j_{max}/3$, which is smaller than $j_{max}/2$.

Thus, the following inequality must hold in order for the $n + m$ wave, defined by the product of u and $\partial u/\partial x$, to be represented on the model grid:

$$n + m \leq \frac{j_{\max}}{2}$$

Or, stated in the inverse, the following inequality describes the case where the $n + m$ wave, defined by the product of u and $\partial u/\partial x$, is unable to be represented on the model grid:

$$n + m > \frac{j_{\max}}{2}$$

Let us consider this unresolvable wave in more detail. The inequality describing this wave can alternatively be written as:

$$n + m = j_{\max} - s$$

Here, s is some generic wavenumber, where $s < \frac{j_{\max}}{2}$. Thus, all values of $j_{max} - s$ are greater than $j_{max}/2$. Considering only the $n + m$ wave, if we substitute this relationship for $n + m$, we obtain:

$$\sin\left(\frac{2\pi}{L}(n+m)x\right) = \sin\left(\frac{2\pi}{L}(j_{\max} - s)x\right)$$

However, because we previously defined $L = j_{max}\Delta x$, we can also substitute for L in the above. Further, the position x along the wave is equal to the product of the grid index j and the grid spacing Δx, such that we obtain:

$$\sin\left(\frac{2\pi}{j_{\max}\Delta x}(j_{\max} - s)j\Delta x\right)$$

Simplifying the terms inside of the sin function, we obtain:

$$\sin\left(2\pi j\frac{(j_{\max} - s)}{j_{\max}}\right) = \sin\left(2\pi j - \frac{2\pi js}{j_{\max}}\right)$$

We can now apply another trigonometric identity,

$$\sin(c - d) = \sin c \cos d - \cos c \sin d$$

Doing so, we obtain:

$$\sin\left(2\pi j - \frac{2\pi js}{j_{\max}}\right) = \sin(2\pi j)\cos\left(\frac{2\pi js}{j_{\max}}\right) - \cos(2\pi j)\sin\left(\frac{2\pi js}{j_{\max}}\right)$$

However, for all grid indices j (which are positive integers), $\sin(2\pi j) = 0$ and $\cos(2\pi j) = 1$. Thus, the above expression simplifies to the following:

$$-\sin\left(\frac{2\pi js}{j_{\max}}\right)$$

Noting again that $x = j\Delta x$ and $L = j_{max}\Delta x$, this can be rewritten as:

$$-\sin\left(\frac{2\pi s}{L}x\right)$$

Because this expression results from the unresolvable $m + n$ wave, we state that the unresolvable wave shows up on the model grid as one that has wavenumber s, where $s = j_{\max} - (n+m)$.

What does this mean? Let us consider the interaction of two waves, for instance a $2\Delta x$ wave and a $4\Delta x$ wave. For the $2\Delta x$ wave, $m = j_{max}/2$. For the $4\Delta x$ wave, $n = j_{max}/4$. Please see the discussion at the bottom of page three of these notes to recall the basis for these definitions. In this case,

$$m + n = \frac{j_{\max}}{2} + \frac{j_{\max}}{4} = \frac{3j_{\max}}{4}$$

This defines a wave with wavelength $\frac{4}{3}\Delta x$, which <u>cannot</u> be resolved on the model grid. But,

$$s = j_{\max} - (n+m) = j_{\max} - \frac{3j_{\max}}{4} = \frac{j_{\max}}{4}$$

This defines a wave with wavelength $4\Delta x$, which <u>can</u> be resolved on the model grid! The unresolvable wave *actually is resolved* on the model grid, but in a non-physical way: it is *aliased* to a wavelength that is resolvable. Stated differently, the energy associated with the wave that is unresolved is *folded* (to borrow a term from radar meteorology) over the shortest-resolvable wave (the $2\Delta x$ wave) into a wave that is resolved on the model grid.

Let us consider the idea of folding in a bit more detail. Consider a model grid with $j_{max} = 24$ grid points. We can obtain the values of m and n for the $2\Delta x$ and $4\Delta x$ waves on this grid as follows. Recall that $L = j_{max}\Delta x = 24\Delta x$ and the ratio of L to m (or n) defines the wavelength of the wave. Thus, for the $2\Delta x$ wave,

$$\frac{L}{m} = 2\Delta x \rightarrow \frac{24\Delta x}{m} = 2\Delta x \rightarrow m = 12$$

And, for the $4\Delta x$ wave,

$$\frac{L}{n} = 4\Delta x \rightarrow \frac{24\Delta x}{n} = 4\Delta x \rightarrow n = 6$$

Thus, $m + n = 12 + 6 = 18$. As a result, $s = j_{max} – (n + m) = 24 – 18 = 6$. Since this is equal to n, the wave with wavenumber s in this case is the $4\Delta x$ wave, as before. The unresolved wavenumber was 6 greater than the maximum-resolvable wavenumber (12, defined by the $2\Delta x$ wave), while the wavenumber to which it is aliased is 6 smaller than the maximum-resolvable wavenumber. This is the manifestation of folding over the shortest-resolvable wavelength, which is illustrated in Figure 1 below.

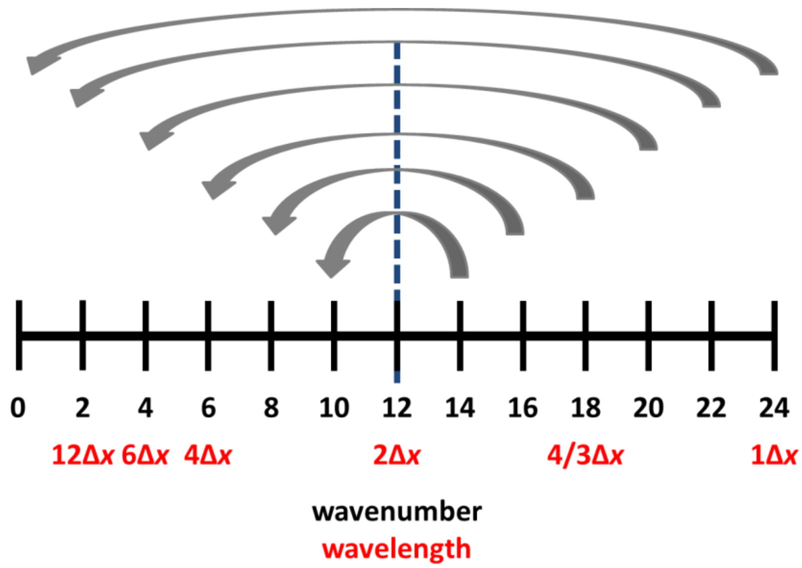

Figure 1. Conceptual illustration of how the interaction of two waves with $m + n > j_{max}/2$ produces aliasing, manifest as the folding of wave energy across the shortest-resolvable wavelength, for $j_{max} = 24$. Here, the unresolvable wavenumber that results from the interaction of two waves is folded over the lowest-resolvable wave (the $2\Delta x$ wave) to a resolved wavelength. Adapted from Warner (2011), their Figure 3.32.

Note, however, that the interaction between two waves does *not* always result in aliasing. Consider, for instance, the interaction of two well-resolved waves on this grid: the $12\Delta x$ wave ($m = 2$) and the $8\Delta x$ wave ($n = 3$). Here, $m + n = 2 + 3 = 5$, which defines a wave with wavelength $4.8\Delta x$ that *can* be resolved on the model grid. Only where $m + n > j_{max}/2$ does aliasing result. This is generally limited to interactions between two resolved but relatively short wavelength features.

Let us continue to consider this model grid with $j_{max} = 24$ grid points. Allowable values of m and n each range from 0 to 12. There exist 42 distinct combinations of m and n that result in aliasing:

Value(s) of *n* (or *m*)	Value(s) of *m* (or *n*)
12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
11	2, 3, 4, 5, 6, 7, 8, 9, 10, 11
10	3, 4, 5, 6, 7, 8, 9, 10

9	4, 5, 6, 7, 8, 9
8	5, 6, 7, 8
7	6, 7

This listing does not count duplicates; e.g., aliasing for *n* = 11 can also occur for *m* = 12, but this case is already accounted for by *n* = 12 and *m* = 11. One could follow a similar procedure to identify the distinct combinations (totaling 49) of *m* and *n* which do not result in aliasing.

Of these 42 combinations, 30 of them result in $m + n \leq 18$: six each for *n* between 9 and 12, four for *n* = 8, and two for *n* = 7. Why are we interested in $m + n \leq 18$? Consider Figure 1. Unresolvable wavenumbers from 13 through 18 alias, or fold, to resolvable wavenumbers between 6 and 11. These identify waves with wavelengths of 2-4Δx, or those that are poorly resolved on the model grid. Thus, **aliasing preferentially results in the artificial accumulation of wave energy at short, poorly-resolved wavelengths**.

When we introduced the concept of *effective resolution* earlier in the semester, we defined it as the smallest wavelength at which the modeled kinetic energy spectrum matches that from theory and observations. At smaller but still resolvable wavelengths, ideally the modeled kinetic energy spectrum is associated with less energy than that from theory and observations. Aliasing, however, can result in an excess accumulation of wave energy in short wavelengths, as the above discussion illustrates. This can result in a modeled kinetic energy spectrum with greater energy that that from theory and observations at short wavelengths. Examples of each are provided in Figure 2.

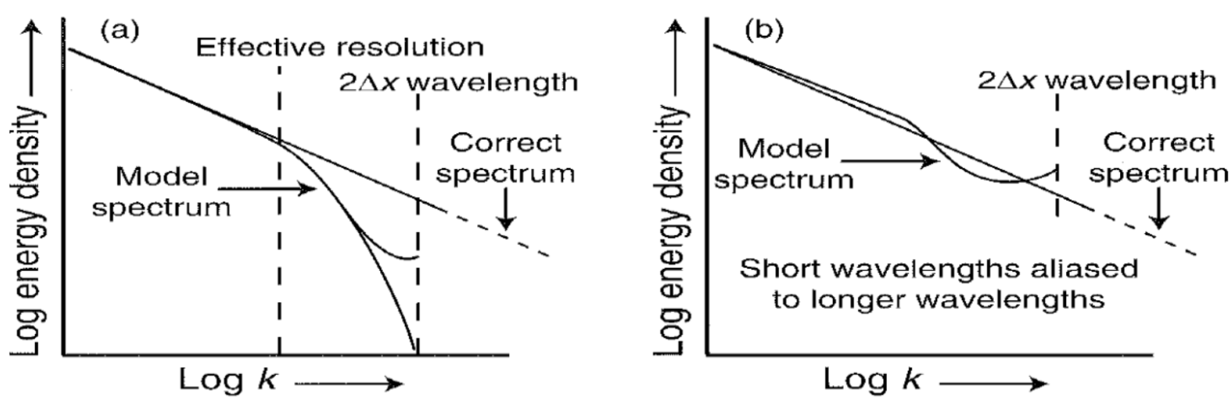

Figure 2. Examples of modeled kinetic energy spectra relative to theory and observations (i.e., the "correct spectrum") for (a) a case where numerical diffusion dampens short wavelengths (large wavenumber *k*, here expressed on a logarithmic axis) and (b) a case where aliasing is not controlled for by numerical diffusion, resulting in an excess of kinetic energy at short wavelengths (or large *k*). Reproduced from Warner (2011), their Figure 3.33.

This is problematic. We know short wavelengths have large truncation error, significant numerical dispersion, and often are those that most rapidly become unstable if the numerical stability criterion is violated. Amplifying the amount of energy contained within these wavelengths only exacerbates these problems. This is another illustrative example of the utility of numerical diffusion, whether implicit or explicit in nature, which we consider in more detail in our next lecture.