18.03 Supplementary Notes

Spring 2010

Contents

0. Preface

This packet collects notes I have produced while teaching 18.03, Ordinary Differential Equations, at MIT in 1996, 1999, 2002, 2004, 2006, 2008, and 2010. They are intended to serve several rather different purposes, supplementing but not replacing the course textbook.

In part they try to increase the focus of the course on topics and perspectives which will be found useful by engineering students, while maintaining a level of abstraction, or breadth of perspective, sufficient to bring into play the added value that a mathematical treatment offers.

For example, in this course we use complex numbers, and in particular the complex exponential function, more intensively than Edwards and Penney do, and several of the sections discuss aspects of them. This ties in with the "Exponential Response Formula," which seems to me to be a linchpin for the course. It leads directly to an understanding of amplitude and phase response curves. It has a beautiful extension covering the phenomenon of resonance. It links the elementary theory of linear differential equations with the use of Fourier series to study LTI system responses to periodic signals, and to the weight function appearing in Laplace transform techniques. It allows a direct path to the solution to standard sinusoidally driven LTI equations which are often solved by a form of undetermined coefficients, and leads to the expression of the sinusoidal solution in terms of gain and phase lag, more useful and enlightening than the expression as a linear combination of sines and cosines.

As a second example, I feel that the standard treatments of Laplace transform in ODE textbooks are wrong to sacrifice the conceptual content of the transformed function, as captured by its pole diagram, and I discuss that topic. The relationship between the modulus of the transfer function and the amplitude response curve is the conceptual core of the course. Similarly, standard treatments of generalized functions, impulse response, and convolution, typically all occur entirely within the context of the Laplace transform, whereas I try to present them as useful additions to the student's set of tools by which to represent natural events.

In fact, a further purpose of these notes is to try to uproot some aspects of standard textbook treatments which I feel are downright misleading. All textbooks give an account of beats which is mathematically artificial and nonsensical from an engineering perspective. I give a derivation of the beat envelope in general, a simple and revealing

use of the complex exponential. Textbooks stress silly applications of the Wronskian, and I try to illustrate what its real utility is. Textbooks typically make the theory of first order linear equations seem quite unrelated to the second order theory; I try to present the first order theory using standard linear methods. Textbooks generally give an inconsistent treatment of the lower limit of integration in the definition of the one-sided Laplace transform, and I try at least to be consistent.

A final objective of these notes is to give introductions to a few topics which lie just beyond the limits of this course: damping ratio and logarithmic decrement; the L^2 or root mean square distance in the theory of Fourier series; the exponential expression of Fourier series; the Gibbs phenomenon; the Wronskian; a discussion of the ZSR/ZIR decomposition; the Laplace transform approach to more general systems in mechanical engineering; and a treatment of a class of "generalized functions," which, while artificially restrictive from a mathematical perspective, is sufficient for all engineering applications and which can be understood directly, without recourse to distributions. These essays are not formally part of the curriculum of the course, but they are written from the perspective developed in the course, and I hope that when students encounter them later on, as many will, they will think to look back to see how these topics appear from the 18.03 perspective.

I want to thank my colleagues at MIT, especially the engineering faculty, who patiently tutored me in the rudiments of engineering: Steve Hall, Neville Hogan, Jeff Lang, Kent Lundberg, David Trumper, and Karen Willcox, were always on call. Arthur Mattuck, Jean Lu, and Lindsay Howie read early versions of this manuscript and offered frank advice which I have tried to follow. I am particularly indebted to Arthur Mattuck, who established the basic syllabus of this course. He has patiently tried to tutor me in how to lecture and how to write (with only moderate success I am afraid). He also showed me the approach to the Gibbs phenomenon included here. My thinking about teaching ODEs has also been influenced by the the pedagogical wisdom and computer design expertise of Hu Hohn, who built the computer manipulatives ("Mathlets") used in this course. They can be found at `http://www-math.mit.edu/daimp`. Assorted errors and infelicities were caught by students in 18.03 and by Professor Sridhar Chitta of MIST, Hyderabad, India, and I am grateful to them all. Finally, I am happy to record my indebtedness to the Brit and Alex d'Arbeloff Fund for Excellence, which provided the stimulus and the support over several years to rethink the contents of this course, and to produce new curricular material.

1. Notation and language

1.1. **Dependent and independent variables.** Most of what we do will involve *ordinary* differential equations. This means that we will have only *one* independent variable. We may have several quantities depending upon that one variable, and we may wish to represent them together as a vector-valued function.

Differential equations arise from many sources, and the independent variable can signify many different things. Nonetheless, very often it represents time, and the dependent variable is some dynamical quantity which depends upon time. For this reason, in these notes we will pretty systematically use t for the independent variable, and x for the dependent variable.

Often we will write simply x, to denote the entire function. The symbols x and $x(t)$ are synonymous, when t is regarded as a variable.

We generally denote the derivative with respect to t by a dot:

$$\dot{x} = \frac{dx}{dt}\,,$$

and reserve the prime for differentiation with respect to a spatial variable. Similarly,

$$\ddot{x} = \frac{d^2x}{dt^2}\,.$$

1.2. **Equations and Parametrizations.** In analytic geometry one learns how to pass back and forth between a description of a set by means of an *equation* and by means of a *parametrization*.

For example, the *unit circle*, that is, the circle with radius 1 and center at the origin, is defined by the equation

$$x^2 + y^2 = 1\,.$$

A *solution* of this equation is a value of (x, y) which satisfies the equation; the set of solutions of this equation is the unit circle.

This solution set is the same as the set parametrized by

$$x = \cos\theta\,, \quad y = \sin\theta\,, \quad 0 \le \theta < 2\pi\,.$$

The set of *solutions* of the equation is the set of *values* of the parametrization. The angle θ is the *parameter* which specifies a solution.

An **equation** is a *criterion*, by which one can decide whether a point lies in the set or not. $(2, 0)$ does not lie on the circle, because it

doesn't satisfy the equation, but $(1, 0)$ does, because it does satisfy the equation.

A **parametrization** is an *enumeration*, a listing, of all the elements of the set. Usually we try to list every element only once. Sometimes we only succeed in picking out some of the elements of the set; for example

$$y = \sqrt{1 - x^2}\,, \quad -1 \le x \le 1$$

picks out the upper semicircle. For emphasis we may say that some enumeration gives a *complete parametrization* if every element of the set in question is named; for example

$$y = \sqrt{1 - x^2}\,, \quad -1 \le x \le 1\,, \quad \text{or} \quad y = -\sqrt{1 - x^2}\,, \quad -1 < x < 1\,,$$

is a complete parametrization of the unit circle, different from the one given above in terms of cosine and sine.

Usually the process of "solving" and equation amounts to finding a parametrization for the set defined by the equation. You could call a parametrization of the solution set of an equation the "general solution" of the equation. This is the language used in Differential Equations.

1.3. **Parametrizing the set of solutions of a differential equation.** A *differential* equation is a stated relationship between a function and its derivatives. A *solution* is a function satisfying this relationship. (We'll amend this slightly at the end of this section.)

For a very simple example, consider the differential equation

$$\ddot{x} = 0\,.$$

A *solution* is a function which satisfies the equation. It's easy to write down many such functions: any function whose graph is a straight line satisfies this ODE.

We can enumerate all such functions: they are

$$x(t) = mt + b$$

for m and b arbitrary real constants. This expression gives a *parametrization* of the set of solutions of $\ddot{x} = 0$. The constants m and b are the *parameters*. In our parametrization of the circle we could choose θ arbitrarily, and analogously now we can choose m and b arbitrarily; for any choice, the function $mt + b$ is a solution.

Warning: If we fix m and b, say $m = 1$, $b = 2$, we have a specific line in the (t, x) plane, with equation $x = t + 2$. One can parametrize this line easily enough; for example t itself serves as a parameter, so the

points $(t, t+2)$ run through the points on the line as t runs over all real numbers. This is an *entirely different* issue from the parametrization of solutions of $\ddot{x} = 0$. Be sure you understand this point.

1.4. **Solutions of ODEs.** The basic existence and uniqueness theorem for ODEs is the following. Suppose that $f(t, x)$ is continuous in the vicinity of a point (a, b). Then there exists a solution to $\dot{x} = f(t, x)$ defined in some interval containing a, and it's unique provided $\partial f/\partial x$ exists.

Here an "interval" is a collection I of real numbers such that if a and b are in I then so is every number between a and b.

There are certainly subtleties here. But some things are obvious. The "uniqueness" part of this theorem says that knowing $x(t)$ for one value $t = a$ is supposed to pick out a single solution: there's supposed to be only one solution with a given "initial value." Well, look at the ODE $\dot{x} = 1/t$. The solutions can be found by simply integrating: $x = \ln|t| + c$. This formula makes it look as though *the* solution with $x(1) = 0$ is $x = \ln|t|$. But in fact there is no reason to prefer this to the following function, which is also a solution to this initial value problem, for any value of c:

$$x(t) = \begin{cases} \ln t & \text{for } t > 0\,, \\ \ln(-t) + c & \text{for } t < 0\,. \end{cases}$$

The gap at $t = 0$ means that the values of $x(t)$ for $t > 0$ have no power to determine the values for $t < 0$.

For this reason it's best to declare that a *solution* to an ODE *must be defined on an entire interval*. The graph has to be a connected curve.

Thus it is more proper to say that the solutions to $\dot{x} = 1/t$ are

$$\ln(t) + c \quad \text{for} \quad t > 0$$

and

$$\ln(-t) + c \quad \text{for} \quad t < 0\,.$$

The *single formula* $\ln|t| + c$ actually describes *two solutions* for each value of c, one defined for $t > 0$ and the other for $t < 0$. The solution with $x(1) = 0$ is $x(t) = \ln t$, with domain of definition the interval consisting of the positive real numbers.

2. Modeling by first order linear ODEs

2.1. **The savings account model.** Modeling a savings account gives a good way to visualize the significance of many of the features of a general first order linear ordinary differential equation.

Write $x(t)$ for the number of dollars in the account at time t. It accrues interest at an interest rate I. This means that at the end of an interest period (say Δt years—perhaps $\Delta t = 1/12$, or $\Delta t = 1/365$) the bank adds $I \cdot x(t) \cdot \Delta t$ dollars to your account:

$$x(t + \Delta t) = x(t) + Ix(t)\Delta t\,.$$

I has units $(\text{years})^{-1}$. Unlike bankers, mathematicians like to take things to the limit: rewrite our equation as

$$\frac{x(t + \Delta t) - x(t)}{\Delta t} = Ix(t)\,,$$

and suppose that the interest period is made to get smaller and smaller. In the limit as $\Delta t \to 0$, we get

$$\dot{x} = Ix$$

—a differential equation.

In this computation, there was no assumption that the interest rate was constant in time; it could well be a function of time, $I(t)$. In fact it could have been a function of both time and the existing balance, $I(x, t)$. Banks often do make such a dependence—you get a better interest rate if you have a bigger bank account. If x is involved, however, the equation is no longer "linear," and we will not consider that case further here.

Now suppose we make contributions to this savings account. We'll record this by giving the *rate* of savings, q. This rate has units dollars per year, so if you contribute every month then the monthly payments will be $q\,\Delta t$ with $\Delta t = 1/12$. This payment also adds to your account, so, when we divide by Δt and take the limit, we get

$$\dot{x} = Ix + q.$$

Once again, your rate of payment may not be constant in time; we might have a function $q(t)$. Also, you may withdraw money from the bank, at some rate measured in dollars per year; this will contribute a negative term to $q(t)$, and exert a downward pressure on your bank account.

What we have, then, is the general first order linear ODE:

$$\dot{x} - I(t)x = q(t). \tag{1}$$

2.2. **Linear insulation.** Here is another example of a linear ODE. The linear model here is not as precise as in the bank account example.

A cooler insulates my lunchtime rootbeer against the warmth of the day, but ultimately heat penetrates. Let's see how you might come up with a mathematical model for this process. You can jump right to (2) if you want, but I would like to spend a minute talking about how one might get there, so that you can carry out the analogous process to model other situations.

The first thing to do is to identify relevant parameters and give them names. Let's write t for the time variable, $x(t)$ for the temperature inside the cooler, and $y(t)$ for the temperature outside.

Let's assume **(a)** that the insulating properties of the cooler don't change over time—we're not going to watch this process for so long that the aging of the cooler itself becomes important! These insulating properties probably do depend upon the inside and outside temperatures themselves. Insulation affects the *rate of change* of the temperature: the rate of change at time t of temperature inside depends upon the temperatures in side and outside at time t. This gives us a first order differential equation of the form

$$\dot{x} = F(x, y)$$

Time for the next simplifying assumption: **(b)** that this rate of change depends only on the *difference* $y - x$ between the temperatures, and not on the temperatures themselves. This means that

$$\dot{x} = f(y - x)$$

for some function f of *one* variable. If the temperature inside the cooler equals the temperature outside, we expect no change. This means that $f(0) = 0$.

Now, *any* reasonable function has a "tangent line approximation," and since $f(0) = 0$ we have

$$f(z) \simeq kz\,.$$

When $|z|$ is fairly small, $f(z)$ is fairly close to kz. (From calculus you know that $k = f'(0)$, but we won't use that here.) When we *replace* $f(y-x)$ by $k(y-x)$ in the differential equation, we are **"linearizing"**

the equation. We get the ODE

$$\dot{x} = k(y - x)\,,$$

which is a *linear* equation (first order, inhomogeneous, constant coefficient). The new assumption we are making, in justifying this final simplification, is that **(c)** we will only use the equation when $z = y - x$ is reasonably small—small enough so that the tangent line approximation is reasonably good.

We can write this equation as

$$\dot{x} + kx = ky. \tag{2}$$

The system—the cooler—is represented by the left hand side, and the input signal—the outside temperature—is represented by the right hand side. This is *Newton's law of cooling*.

The constant k is the **coupling constant** mediating between the two temperatures. It will be large if the insulation is poor, and small if it's good. If the insulation is perfect, then $k = 0$. The factor of k on the right might seem odd, but it you can see that it is forced on us by checking units: the left hand side is measured in degrees per hour, so k is measured in units of $(\text{hours})^{-1}$.

We can see some general features of insulating behavior from this equation. For example, the times at which the inside and outside temperatures coincide are the times at which the inside temperature is at a critical point:

$$\dot{x}(t_1) = 0 \qquad \text{exactly when} \qquad x(t_1) = y(t_1). \tag{3}$$

2.3. **System, signal, system response.** A first order linear ODE is in **standard form** when it's written as

$$\dot{x} + p(t)x = q(t). \tag{4}$$

In the bank account example, $p(t) = -I(t)$. This way of writing it reflects a useful "systems and signals" perspective on differential equations, one which you should develop. The left hand side of (4) describes the *system*—the bank, in this instance. Operating without outside influence (that is, without contributions or withdrawals), the system is described by the *homogeneous linear equation*

$$\dot{x} + p(t)x = 0.$$

The right hand side of (4) describes the outside influences. You are "driving" the system, with your contributions and withdrawals. These

constitute the "input signal," to which the bank system reacts. Your bank balance, x, is the "system response."

For our purposes, a **system** is represented by some combination of the dependent variable x and its derivatives; a **signal** is a dependent variable, that is, a function of the independent variable t. Both the input signal q and the system response x are signals. We tend to write the part of the ODE representing the system on the *left*, and the part representing the input signal on the *right*. For more detail on this perspective, see Sections 8, 14 and 24.

This way of thinking takes some getting used to. After all, in these terms the ODE (4) says: the system response x determines the signal (namely, the signal equals $\dot{x} + p(t)x$). The ODE (or more properly the differential *operator*) that represents the system takes the system response and gives you back the input signal—the reverse of what you might have expected. But that is the way it works; the equation gives you conditions on x which make it a response of the system. In a way, the whole objective of solving an ODE is to "invert the system" (or the operator that represents it).

We might as well mention some other bits of terminology, while we're at it. In the equation (4), the function $p(t)$ is a **coefficient** of the equation (the only one in this instance—higher order linear equations have more), and the equation is said to have "constant coefficients" if $p(t)$ is constant. In different but equivalent terminology, if $p(t)$ is a constant then we have a **linear time-invariant**, or **LTI**, system.

3. Solutions of first order linear ODEs

3.1. **Homogeneous and inhomogeneous; superposition.** A first order linear equation is **homogeneous** if the right hand side is zero:

$$\dot{x} + p(t)x = 0\,. \tag{1}$$

Homogeneous linear equations are separable, and so the solution can be expressed in terms of an integral. The general solution is

$$x = \pm\, e^{-\int p(t)dt} \qquad \text{or} \quad x = 0\,. \tag{2}$$

Question: Where's the constant of integration here? Answer: The indefinite integral is only defined up to adding a constant, which becomes a positive factor when it is exponentiated.

We also have the option of replacing the indefinite integral with a definite integral. The lower bound will be some value of t at which the ODE is defined, say a, while the upper limit should be t, in order to define a function of t. This means that I have to use a different symbol for the variable inside the integral—say τ, the Greek letter "tau." The general solution can then be written as

$$x = c\, e^{-\int_a^t p(\tau)d\tau}\,, \quad c \in \mathbb{R}\,. \tag{3}$$

This expression for the general solution to (1) will often prove useful, even when it can't be integrated in elementary functions. Note that the constant of integration is also an initial value: $c = x(a)$.

I am not requiring $p(t)$ to be constant here. If it is, then we can evaluate the integral and find the familiar solution $x = ce^{-pt}$.

These formulas tell us something important about a function $x = x(t)$ which satisfies (1): either $x(t) = 0$ for *all* t, or $x(t) \neq 0$ for *all* t: either x is the zero function, or it's never zero. This is a consequence of the fact that the exponential function never takes on the value zero.

Even without solving it, we can observe an important feature of the solutions of (1):

If x_h is a solution, so is cx_h for any constant c.

The subscripted h is for "homogeneous." This can be verified directly by assuming that x_h is a solution and then checking that cx_h is too. Conversely, if x_h is *any* nonzero solution, then the *general* solution is cx_h: *every* solution is a multiple of x_h. This is because of the uniqueness theorem for solutions: for any choice of initial value $x(a)$, I can find c

so that $cx_h(a) = x(a)$ (namely, $c = x(a)/x_h(a)$), and so by uniqueness $x = cx_h$ for this value of c.

Now suppose the input signal is nonzero, so our equation is

$$\dot{x} + p(t)x = q(t)\,. \tag{4}$$

Suppose that in one way or another we have found a solution x_p to (4). *Any* single solution will do. We will call it a "particular" solution. Keeping the notation x_h for a nonzero solution to the corresponding *homogeneous* equation (1), we can calculate that $x_p + cx_h$ is again a solution to (4).

Exercise 3.1.1. Verify this.

In fact,

$$\boxed{\text{The general solution to (4) is } x_p + cx_h} \tag{5}$$

since any initial condition can be achieved by judicious choice of c. This formula shows how the constant of integration, c, occurs in the general solution of a *linear* equation. It tends to show up in a more complicated way if the equation is nonlinear.

I want to emphasize that despite being called "particular," the solution x_p can be *any* solution of (4); it need not be *special* in any way for it to serve in (5).

There's a slight generalization: suppose x_1 is a solution to

$$\dot{x} + p(t)x = q_1(t)$$

and x_2 is a solution to

$$\dot{x} + p(t)x = q_2(t)$$

—same coefficient $p(t)$, so the same system, but two different input signals. Then (for any constants c_1, c_2) $c_1x_1 + c_2x_2$ is a solution to

$$\dot{x} + p(t)x = c_1q_1(t) + c_2q_2(t)\,.$$

In our banking example, if we have two bank accounts with the same interest rate, and contribute to them separately, the sum of the accounts will be the same as if we combined them into one account and contributed the sum to the combined account. This is the **principle of superposition**.

The principle of superposition lets us break up the input signal into bitesized pieces, solve the corresponding equations, and add the solutions back together to get a solution to the original equation.

3.2. **Variation of parameters.** Now we try to solve the general first order linear equation,

$$\dot{x} + p(t)x = q(t)\,. \tag{6}$$

As we presented it above, the procedure for solving this breaks into two parts. We first find a nonzero solution, say x_h, of the **associated homogeneous equation**

$$\dot{x} + p(t)x = 0 \tag{7}$$

—that is, (6) with the right hand side replaced by zero. *Any* nonzero solution will do, and since (7) is separable, finding one is a matter of integration. The general solution to (7) is then cx_h for a constant c. The constant c "parametrizes" the solutions to (7).

The second step is to somehow find some single solution to (6) itself. We have not addressed this problem yet. One idea is to hope for a solution of the form vx_h, where v now is *not* a constant (which would just give a solution to the *homogeneous* equation), but rather some function of t, which we will write as $v(t)$ or just v.

So let's make the substitution $x = vx_h$ and study the consequences. When we make this substitution in (6) and use the product rule we find

$$\dot{v}x_h + v\dot{x}_h + pvx_h = q\,.$$

The second and third terms sum to zero, since x_h is a solution to (7), so we are left with a differential equation for v:

$$\dot{v} = x_h^{-1}q\,. \tag{8}$$

This can be solved by direct integration once again. Write v_p for a particular solution to (8). A particular solution to our original equation (6) is then given by $x_p = v_p x_h$.

By superposition, the general solution is $x = x_p + cx_h$. You can also see this by realizing that the general solution to (8) is $v = v_p + c$, so the general solution x is $vx_h = x_p + cx_h$.

Many people like to remember this in the following form: the general solution to (6) is

$$\boxed{x = x_h \int x_h^{-1} q\, dt} \tag{9}$$

since the general solution to (8) is $v = \displaystyle\int x_h^{-1}q\, dt$. Others just make the substitution $x = vx_h$ and do the calculation.

Example. The inhomogeneous first order linear ODE we wish to solve is

$$\dot{x} + tx = (1+t)e^t \,.$$

The associated homogeneous equation is

$$\dot{x} + tx = 0,$$

which is separable and easily leads to the nonzero solution $x_h = e^{-t^2/2}$. So we'll try for a solution of the original equation of the form $x = ve^{-t^2/2}$. Substituting this into the equation and using the product rule gives us

$$\dot{v}e^{-t^2/2} - vte^{-t^2/2} + vte^{-t^2/2} = (1+t)e^t \,.$$

The second and third terms cancel, as expected, leaving us with $\dot{v} = (1+t)e^{t+t^2/2}$. Luckily, the derivative of the exponent here occurs as a factor, so this is easy to integrate: $v_p = e^{t+t^2/2}$ (plus a constant, which we might as well take to be zero since we are interested only in finding one solution). Thus a particular solution to the original equation is $x_p = v_p x_h = e^t$. It's easy to check that this is indeed a solution! By (5) the general solution is $x = e^t + ce^{-t^2/2}$.

This method is called "variation of parameter." The "parameter" is the constant c in the expression cx_h for the general solution of the associated homogeneous equation. It is allowed to vary with time in an effort to come up with a solution of the given *inhomogeneous* equation. The method of variation of parameter is equivalent to the method of integrating factors described in Edwards and Penney; in fact x_h^{-1} is an integrating factor for (6). Either way, we have broken the original problem into two problems each of which can be solved by direct integration.

3.3. **Continuation of solutions.** There is an important theoretical outcome of the method of Variation of Parameters. To see the point, consider first the *nonlinear* ODE $\dot{x} = x^2$. This is separable, with general solution $x = 1/(c-t)$. There is also a "missing solution" $x = 0$ (which corresponds to $c = \infty$).

As we pointed out in Section 1, the statement that $x = 1/(c-t)$ is a solution is somewhat imprecise. This equation actually defines *two* solutions: one defined for $t < c$, and another defined for $t > c$. These are *different* solutions. One becomes asymptotic to $t = c$ as $t \uparrow c$; the other becomes asymptotic to $t = c$ as $t \downarrow c$. Neither of these solutions can be extended to a solution defined at $t = c$; both solutions "blow up" at $t = c$. This pathological behavior occurs *despite* the fact that

the ODE itself doesn't exhibit any special pathology at $t = c$ for any value of c.

With the exception of the constant solution, *no solution can be defined for all time*, despite the fact that the equation is perfectly well defined for all time.

Another thing that may happen to solutions of nonlinear equations is illustrated by the equation $\dot{x} = -x/y$. This is separable, and in implicit form the general solution is $x^2 + y^2 = c^2$, $c > 0$: circles centered at the origin. To get a *function* as a solution, one must restrict to the upper half plane or to the lower half plane: $y = \pm\sqrt{c^2 - x^2}$. In any case, these solutions can't be extended to all time, once again, but now for a different reason: they come up to a point at which the tangent line becomes vertical (at $x = \pm c$), and the solution function doesn't extend past that point.

The situation for *linear* equations is quite different. The fact that continuous functions are integrable (from calculus) shows that if $f(t)$ is defined and continuous on an interval, then all solutions to $\dot{x} = f(t)$ extend over the same interval. Because the solution to (6) is achieved by two direct integrations, we obtain the following result, which stands in contrast to the situation typical of nonlinear equations.

Theorem: If p and q are defined (and reasonably well-behaved) for all t between a and b, then any solution to $\dot{x} + p(t)x = q(t)$ defined somewhere between a and b extends to a solution defined on the entire interval from a to b.

3.4. **Final comments on the bank account model.** Let us solve (1) in the special case in which I and q are both constant. In this case the equation

$$\dot{x} - Ix = q$$

is separable; we do not need to use the method of variation of parameters or integrating factors. Separating,

$$\frac{dx}{x + q/I} = I\,dt$$

so integrating and exponentiating,

$$x = -q/I + ce^{It}\,, \quad c \in \mathbb{R}\,.$$

Let's look at this formula for a moment. There is a constant solution, namely $x = -q/I$. I call this the *credit card solution*. I owe the bank q/I dollars. They "give" me interest, at the rate of I times the

value of the bank account. Since that value is negative, what they are doing is charging me: I am using the bank account as a loan, and my "contributions" amount to interest payments on the loan, and exactly balance the interest charges. The bank balance never changes. This steady state solution has large magnitude if my rate of payments is large, or if the interest is small.

In calling this the credit card solution, I am assuming that $q > 0$. If $q < 0$, then the constant solution $x = -q/I$ is positive. What does this signify?

If $c < 0$, I owe the bank more than can be balanced by my payments, and my debt increases exponentially. Let's not dwell on this unfortunate scenario, but pass quickly to the case $c > 0$, when some of my payments are used to pay off the principal, and ultimately to add to a positive bank balance. That balance then proceeds to grow approximately exponentially.

In terms of the initial condition $x(0) = x_0$, the solution is

$$x = -q/I + (x_0 + q/I)e^{It} .$$

4. Sinusoidal solutions

Many things in nature are periodic, even sinusoidal. We will begin by reviewing terms surrounding periodic functions. If an LTI system is fed a periodic input signal, we have a right to hope for a periodic solution. Usually there is exactly one periodic solution, and often all other solutions differ from it by a "transient," a function that dies off exponentially. This section begins by setting out terms and facts about periodic and sinusoidal functions, and then studies the response of a first order LTI system to a sinusoidal signal. This is a special case of a general theory described in Sections 10 and 14.

4.1. **Periodic and sinusoidal functions.** A function $f(t)$ is **periodic** if there is a number $a > 0$ such that

$$f(t + a) = f(t)$$

for all t. It repeats itself over and over, and has done since the world began. The number a is a **period**. Notice that if a is a period then so is $2a$, and $3a$, and so in fact is any positive integral multiple of a. If $f(t)$ is continuous and not constant, there is a smallest period, called the *minimal period* or simply *the* period, and is often denoted by P. If the independent variable t is a distance rather than a time, the period is also called the *wavelength*, and denoted in physics by the Greek letter "lambda," λ.

A periodic function of time has a *frequency*, too, often written using the Greek letter "nu," ν. The frequency is the reciprocal of the minimal period:

$$\nu = 1/P.$$

This is the number of cycles per unit time, and its units are, for example, $(\text{sec})^{-1}$.

Since many periodic functions are closely related to sine and cosines, it is common to use the **angular** or **circular frequency**, denoted by the Greek letter "omega," ω. This is 2π times the frequency:

$$\omega = 2\pi\nu.$$

If ν is the number of *cycles per second*, then ω is the number of *radians per second*. In terms of the angular frequency, the period is

$$P = \frac{2\pi}{\omega}.$$

The **sinusoidal functions** make up a particular class of periodic functions, namely, those which can be expressed as a cosine function

which as been *amplified, shifted* and *compressed*:

$$\boxed{f(t) = A\cos(\omega t - \phi)} \tag{1}$$

The function (1) is periodic of period $2\pi/\omega$ and frequency $\omega/2\pi$, and circular frequency ω.

The parameter A (or, better, $|A|$) is the **amplitude** of (1). By replacing ϕ by $\phi + \pi$ if necessary, we may always assume $A \geq 0$, and we will usually make this assumption.

The number ϕ is the **phase lag** (relative to the cosine). It is measured in radians or degrees. The **phase shift** is $-\phi$. In many applications, $f(t)$ represents the response of a system to a signal of the form $B\cos(\omega t)$. The phase lag is then usually positive—the system response lags behind the signal—and this is one reason why we choose to favor the *lag* and not the *shift* by assigning a notation to it. Some engineers prefer to use ϕ for the phase shift, i.e. the negative of our ϕ. You will just have to check and see which convention is in use.

The phase lag can be chosen to lie between 0 and 2π. The ratio $\phi/2\pi$ is the fraction of a full period by which the function (1) is shifted to the right relative to $\cos(\omega t)$: $f(t)$ is $\phi/2\pi$ radians behind $\cos(\omega t)$.

Here are the instructions for building the graph of (1) from the graph of $\cos t$. First *amplify*, or vertically expand, the graph by a factor of A; then *shift* the result to the right by ϕ units; and finally *compress* it horizontally by a factor of ω.

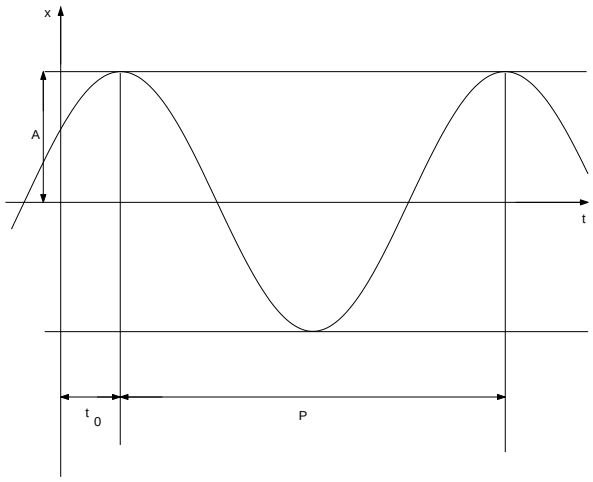


FIGURE 1. Parameters of a sinusoidal function

One can also write (1) as

$$f(t) = A\cos(\omega(t - t_0)),$$

where $\omega t_0 = \phi$, or

$$t_0 = \frac{\phi}{2\pi} P \tag{2}$$

t_0 is the **time lag**. It is measured in the same units as t, and represents the amount of time $f(t)$ lags behind the compressed cosine signal $\cos(\omega t)$. Equation (2) expresses the fact that t_0 makes up the same fraction of the period P as the phase lag ϕ does of the period of the cosine function.

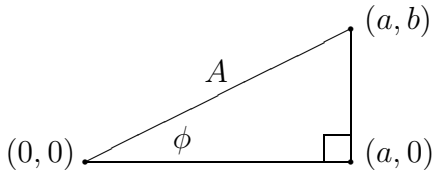
There is a fundamental trigonometric identity, illustrated in the Mathlet `Trigonometric Id`, which rewrites the shifted and scaled cosine function $A\cos(\omega t - \phi)$ as a *linear combination* of $\cos(\omega t)$ and $\sin(\omega t)$:

$$\boxed{A\cos(\omega t - \phi) = a\cos(\omega t) + b\sin(\omega t)} \tag{3}$$

The numbers a and b are determined by A and ϕ: in fact,

$$\boxed{a = A\cos(\phi)\,, \quad b = A\sin(\phi)}$$

This is the familiar formula for the cosine of a difference. Geometrically, (a, b) is the pair of coordinates of the point on the circle with radius A and center at the origin, making an angle of ϕ counterclockwise from the positive x axis.



In the formula either or both of a and b can be negative; (a, b) can be any point in the plane.

I want to stress the importance of this simple observation. Perhaps it's more striking when read from right to left: *any* linear combination of $\cos(\omega t)$ and $\sin(\omega t)$ is not only *periodic*, of period $2\pi/\omega$—this much is obvious—but even *sinusoidal*—which seems much less obvious. And the geometric descriptions of the amplitude A and phase lag ϕ is very useful. Remember them:

$$\boxed{A \text{ and } \phi \text{ are the polar coordinates of } (a, b)}$$

If we replace ωt by $-\omega t + \phi$ in (3), then $\omega t - \phi$ gets replaced by $-\omega t$ and the identity becomes $A\cos(-\omega t) = a\cos(-\omega t+\phi)+b\sin(-\omega t+\phi)$. Since the cosine is even and the sine is odd, this is equivalent to

$$A\cos(\omega t) = a\cos(\omega t - \phi) - b\sin(\omega t - \phi) \tag{4}$$

which is often useful as well. The relationship between a, b, A, and ϕ is always the same.

4.2. **Periodic solutions and transients.** Let's return to the model of the cooler, described in Section 2.2: $x(t)$ is the temperature inside the cooler, $y(t)$ the temperature outside, and we model the cooler by the first order linear equation with constant coefficient:

$$\dot{x} + kx = ky.$$

Let's suppose the outside temperature varies sinusoidally (warmer in the day, cooler at night). (This involves choosing units for temperature so that the *average* temperature is zero.) By setting our clock so that the highest temperature occurs at $t = 0$, we can thus model $y(t)$ by

$$y(t) = y_0\cos(\omega t)$$

where $y_0 = y(0)$ is the daily high temperature. So our model is

$$\dot{x} + kx = ky_0\cos(\omega t). \tag{5}$$

The equation (5) can be solved by the standard method for solving first order linear ODEs (integrating factors, or variation of parameter). In fact, we will see in Section 10 that since the right hand side is sinusoidal there is an explicit and direct way to write down the solution using complex numbers. Here's a different approach, which one might call the "method of optimism."

Let's look for a *periodic* solution; not unreasonable since the driving function is periodic. Even more optimistically, let's hope for a sinusoidal function. At first you might hope that $A\cos(\omega t)$ would work, for suitable constant A, but that turns out to be too much to ask, and doesn't reflect what we already know from our experience with temperature: the temperature inside the cooler tends to lag behind the ambient temperature. This lag can be accommodated by means of the formula:

$$x_p = gy_0\cos(\omega t - \phi). \tag{6}$$

We have chosen to write the amplitude here as a multiple of the ambient high temperature y_0. The multiplier g and the phase lag ϕ are numbers which we will try to choose so that x_p is indeed a solution. We use the

subscript p to indicate that this is a Particular solution. It is also a Periodic solution, and generally will turn out to be the only periodic solution.

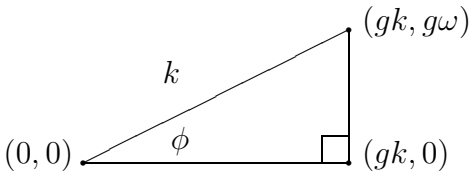
We can and will take ϕ between 0 and 2π, and $g \geq 0$: so gy_0 is the amplitude of the temperature oscillation in the cooler. The number g is the ratio of the maximum temperature in the cooler to the maximum ambient temperature; it is called the **gain** of the system. The angle ϕ is the **phase lag**. Both of these quantities depend upon the coupling constant k and the circular frequency of the input signal ω.

To see what g and ϕ must be in order for x_p to be a solution, we will use the alternate form (4) of the trigonometric identity. The important thing here is that there is only one pair of numbers (a, b) for which this identity holds: they are the rectangular coordinates of the point with polar coordinates (A, ϕ).

If $x = gy_0 \cos(\omega t - \phi)$, then $\dot{x} = -gy_0\omega \sin(\omega t - \phi)$. Substitute these values into the ODE:

$$gy_0 k \cos(\omega t - \phi) - gy_0\omega \sin(\omega t - \phi) = ky_0 \cos(\omega t).$$

I have switched the order of the terms on the left hand side, to make comparison with the trig identity (4) easier. Cancel the y_0. Comparing this with (4), we get the triangle



From this we read off

$$\tan\phi = \omega/k \tag{7}$$

and

$$g = \frac{k}{\sqrt{k^2 + \omega^2}} = \frac{1}{\sqrt{1 + (\omega/k)^2}}. \tag{8}$$

Our work shows that with these values for g and ϕ the function x_p given by (6) *is* a solution to (5).

Incidentally, the triangle shows that the gain g and the phase lag ϕ in this first order equation are related by

$$g = \cos\phi. \tag{9}$$

According to the principle of superposition, the general solution is

$$x = x_p + ce^{-kt}, \tag{10}$$

since e^{-kt} is a nonzero solution of the homogeneous equation $\dot{x}+kx = 0$.

You can see why you need the extra term ce^{-kt}. Putting $t = 0$ in (6) gives a specific value for $x(0)$. We have to do something to build a solution for initial value problems specifying different values for $x(0)$, and this is what the additional term ce^{-kt} is for. But this term dies off exponentially with time, and leaves us, for large t, with the same solution, x_p, independent of the initial conditions. In terms of the model, the cooler did start out at refrigerator temperature, far from the "steady state." In fact the periodic system response has average value zero, equal to the average value of the signal. No matter what the initial temperature $x(0)$ in the cooler, as time goes by the temperature function will converge to $x_p(t)$. This long-term lack of dependence on initial conditions confirms an intuition. The exponential term ce^{-kt} is called a **transient**. The general solution, in this case and in many others, is a periodic solution plus a transient.

I stress that *any* solution can serve as a "particular solution." The solution x_p we came up with here is special not because it's a particular solution, but rather because it's a *periodic solution*. In fact (assuming $k > 0$) it's the *only* periodic solution.

4.3. **Amplitude and phase response.** There is a lot more to learn from the formula (6) and the values for g and ϕ given in (7) and (8). The terminology applied below to solutions of the first order equation (5) applies equally well to solutions of second and higher order equations. See Section 14 for further discussion, and the Mathlet `Amplitude and Phase: First Order` for a dymamic illustration.

Let's fix the coupling constant k and think about how g and ϕ vary as we vary ω, the circular frequency of the signal. Thus we will regard them as functions of ω, and we may write $g(\omega)$ and $\phi(\omega)$ in order to emphasize this perspective. We are supposing that the *system* is constant, and watching its response to a variety of different input signals. Graphs of $g(\omega)$ and $-\phi(\omega)$ for values of the coupling constant $k = .25, .5, .75, 1, 1.25, 1.5$ is displayed in Figure 2.

These graphs are essentially **Bode plots**. Technically, the Bode plots displays $\log g(\omega)$ and $-\phi(\omega)$ against $\log \omega$.

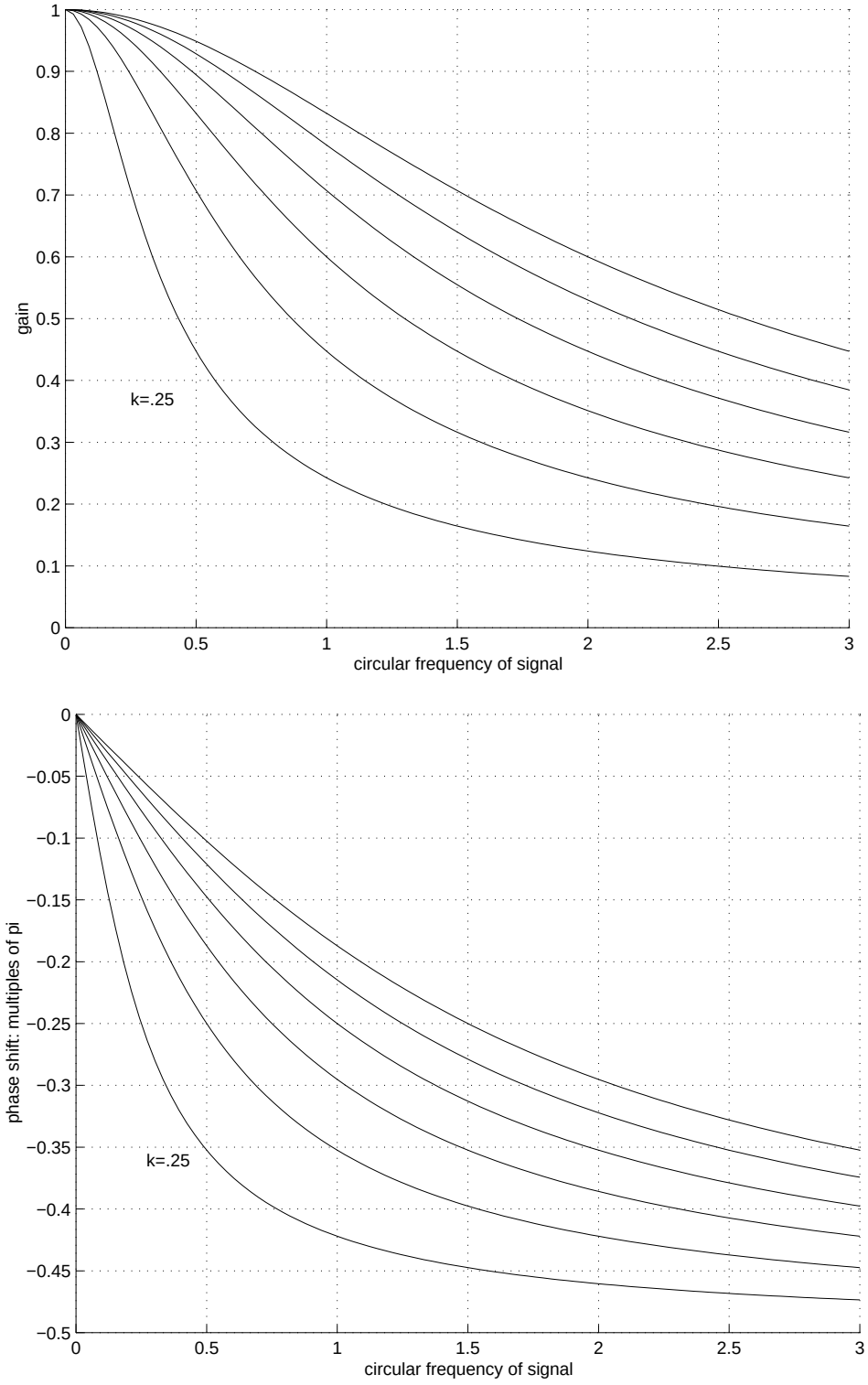


FIGURE 2. First order amplitude response curves

5. The algebra of complex numbers

We use complex numbers for more purposes in this course than the textbook does. This chapter tries to fill some gaps.

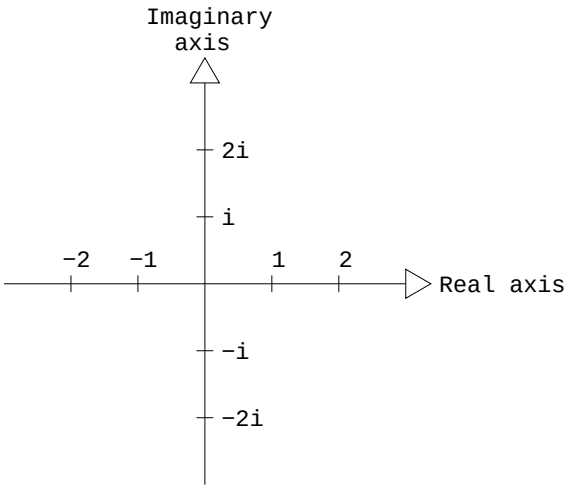
5.1. **Complex algebra.** A "complex number" is an element (a, b) of the plane.

Special notation is used for vectors in the plane when they are thought of as complex numbers. We think of the real numbers as lying in the plane as the horizontal axis: the real number a is identified with the vector $(a, 0)$. In particular 1 is the basis vector $(1, 0)$.

The vector $(0, 1)$ is given the symbol i. Every element of the plane is a linear combination of these two vectors, 1 and i:

$$(a, b) = a + bi.$$

When we think of a point in the plane as a complex number, we always write $a + bi$ rather than (a, b).



The real number a is called the **real part** of $a + bi$, and the real number b is the **imaginary part** of $a + bi$. Notation:

$$\mathrm{Re}\,(a + bi) = a\,, \quad \mathrm{Im}\,(a + bi) = b.$$

A complex number is **purely imaginary** if it lies on the vertical or **imaginary axis**. It is a real multiple of the complex number i. A complex number is **real** if it lies on the horizontal or **real axis**. It is a real multiple of the complex number 1.

The only complex number which is both real and purely imaginary is 0, the origin.

Complex numbers are **added** by vector addition. Complex numbers are **multiplied** by the rule

$$i^2 = -1$$

and the standard rules of arithmetic.

This means we "FOIL" out products. For example,

$$(1+2i)(3+4i) = 1\cdot 3 + 1\cdot 4i + (2i)\cdot 3 + (2i)\cdot(4i) = \cdots$$

—and then use commutativity and the rule $i^2 = -1$:

$$\cdots = 3 + 4i + 6i - 8 = -5 + 10i.$$

The real part of the product is the product of real parts *minus* the product of imaginary parts. The imaginary part of the product is the sum of the crossterms.

We will write the set of all real numbers as $\mathbb{R}$ and the set of all complex numbers as $\mathbb{C}$. Often the letters z, w, v, and s, and r are used to denote complex numbers. The operations on complex numbers satisfy the usual rules:

Theorem. If v, w, and z are complex numbers then

$$z + 0 = z\,, \qquad v + (w + z) = (v + w) + z\,, \qquad w + z = z + w,$$

$$z \cdot 1 = z\,, \qquad v(wz) = (vw)z\,, \qquad wz = zw$$

$$(v + w)z = vz + wz.$$

This is easy to check. The vector negative gives an additive inverse, and, as we will see below, every complex number except 0 has a multiplicative inverse.

Unlike the real numbers, the set of complex numbers doesn't come with a notion of greater than or less than.

Exercise 5.1.1. Rewrite $((1+\sqrt{3}i)/2)^3$ and $(1+i)^4$ in the form $a+bi$.

5.2. **Conjugation and modulus.** The **complex conjugate** of a complex number $a+bi$ is the complex number $a-bi$. Complex conjugation reflects a complex number across the real axis. The complex conjugate of z is written $\bar{z}$:

$$\boxed{\overline{a+bi} = a - bi}$$

Theorem. Complex conjugation satisfies:

$$\bar{\bar{z}} = z, \qquad \overline{w+z} = \bar{w} + \bar{z}, \qquad \overline{wz} = \bar{w}\bar{z}.$$

A complex number z is real exactly when $\bar{z} = z$ and is purely imaginary exactly when $\bar{z} = -z$. The real and imaginary parts of a complex number z can be written using complex conjugation as

$$\operatorname{Re} z = \frac{z + \bar{z}}{2}, \qquad \operatorname{Im} z = \frac{z - \bar{z}}{2i}. \tag{1}$$

Again this is easy to check.

Exercise 5.2.1. Show that if $z = a + bi$ then

$$z\bar{z} = a^2 + b^2.$$

This is the square of the distance from the origin, and so is a nonnegative real number, nonzero as long as $z \neq 0$. Its nonnegative square root is the **absolute value** or **modulus** of z, written

$$|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}.$$

Thus

$$\boxed{z\bar{z} = |z|^2} \tag{2}$$

Exercise 5.2.2. Show that $|wz| = |w||z|$. Since this notation clearly extends its meaning on real numbers, it follows that if r is a positive real number then $|rz| = r|z|$, in keeping with the interpretation of absolute value as distance from the origin.

Any nonzero complex number has a multiplicative inverse: as $z\bar{z} = |z|^2$, $z^{-1} = \bar{z}/|z|^2$. If $z = a + bi$, this says

$$\boxed{\frac{1}{(a + bi)} = \frac{a - bi}{a^2 + b^2}}$$

This is "rationalizing the denominator."

Exercise 5.2.3. Compute i^{-1}, $(1 + i)^{-1}$, and $\dfrac{1 + i}{2 - i}$. What is $|z^{-1}|$ in terms of $|z|$?

Exercise 5.2.4. Since rules of algebra hold for complex numbers as well as for real numbers, the quadratic formula correctly gives the roots of a quadratic equation $x^2 + bx + c = 0$ even when the "discriminant" $b^2 - 4c$ is negative. What are the roots of $x^2 + x + 1$? Of $x^2 + x + 2$? The quadratic formula even works if b and c are not real. Solve $x^2 + ix + 1 = 0$.

5.3. **The fundamental theorem of algebra.** Complex numbers remedy a defect of real numbers, by providing a solution for the quadratic equation $x^2 + 1 = 0$. It turns out that you don't have to worry that someday you'll come across a weird equation that requires numbers even more complex than complex numbers:

Fundamental Theorem of Algebra. Any nonconstant polynomial (even one with complex coefficients) has a complex root.

Once you have a single root, say r, for a polynomial $p(x)$, you can divide through by $(x - r)$ and get a polynomial of smaller degree as quotient, which then also has a complex root, and so on. The result is that a polynomial $p(x) = ax^n + \cdots$ of degree n factors completely into linear factors over the complex numbers:

$$p(x) = a(x - r_1)(x - r_2) \cdots (x - r_n).$$

6. The complex exponential

The exponential function is a basic building block for solutions of ODEs. Complex numbers expand the scope of the exponential function, and bring trigonometric functions under its sway.

6.1. **Exponential solutions.** The function e^t is *defined* to be the solution of the initial value problem $\dot{x} = x$, $x(0) = 1$. More generally, the chain rule implies the

Exponential Principle:

For any constant w, e^{wt} is the solution of $\dot{x} = wx$, $x(0) = 1$.

Now look at a more general constant coefficient homogeneous linear ODE, such as the second order equation

$$\ddot{x} + c\dot{x} + kx = 0. \tag{1}$$

It turns out that there is always a solution of (1) of the form $x = e^{rt}$, for an appropriate constant r.

To see what r should be, take $x = e^{rt}$ for an as yet to be determined constant r, substitute it into (1), and apply the Exponential Principle. We find

$$(r^2 + cr + k)e^{rt} = 0.$$

Cancel the exponential (which, conveniently, can never be zero), and discover that r must be a root of the polynomial $p(s) = s^2 + cs + k$. This is the characteristic polynomial of the equation. See Section 10 for more about this. The **characteristic polynomial** of the linear equation with constant coefficients

$$a_n \frac{d^n x}{dt^n} + \cdots + a_1 \frac{dx}{dt} + a_0 x = 0$$

is

$$p(s) = a_n s^n + \cdots + a_1 s + a_0 \,.$$

Its roots are the **characteristic roots** of the equation. We have discovered the

Characteristic Roots Principle:

(2)	e^{rt} is a solution of a constant coefficient homogeneous linear differential equation exactly when r is a root of the characteristic polynomial.

Since most quadratic polynomials have two distinct roots, this normally gives us two linearly independent solutions, $e^{r_1 t}$ and $e^{r_2 t}$. The general solution is then the linear combination $c_1 e^{r_1 t} + c_2 e^{r_2 t}$.

This is fine if the roots are real, but suppose we have the equation

$$\ddot{x} + 2\dot{x} + 2x = 0 \tag{3}$$

for example. By the quadratic formula, the roots of the characteristic polynomial $s^2 + 2s + 2$ are the complex conjugate pair $-1 \pm i$. We had better figure out what is meant by $e^{(-1+i)t}$, for our use of exponentials as solutions to work.

6.2. **The complex exponential.** We don't yet have a definition of e^{it}. Let's hope that we can define it so that the Exponential Principle holds. This means that it should be the solution of the initial value problem

$$\dot{z} = iz\,, \quad z(0) = 1\,.$$

We will probably have to allow it to be a *complex valued* function, in view of the i in the equation. In fact, I can produce such a function:

$$z = \cos t + i \sin t\,.$$

Check: $\dot{z} = -\sin t + i \cos t$, while $iz = i(\cos t + i \sin t) = i \cos t - \sin t$, using $i^2 = -1$; and $z(0) = 1$ since $\cos(0) = 1$ and $\sin(0) = 0$.

We have now justified the following definition, which is known as **Euler's formula:**

$$\boxed{e^{it} = \cos t + i \sin t} \tag{4}$$

In this formula, the left hand side is *by definition* the solution to $\dot{z} = iz$ such that $z(0) = 1$. The right hand side writes this function in more familiar terms.

We can reverse this process as well, and express the trigonometric functions in terms of the exponential function. First replace t by $-t$ in (4) to see that

$$e^{-it} = \overline{e^{it}}\,.$$

Then put $z = e^{it}$ into the formulas (5.1) to see that

$$\boxed{\cos t = \frac{e^{it} + e^{-it}}{2}\,, \quad \sin t = \frac{e^{it} - e^{-it}}{2i}} \tag{5}$$

We can express the solution to

$$\dot{z} = (a + bi)z\,, \quad z(0) = 1$$

in familiar terms as well: I leave it to you to check that it is

$$z = e^{at}(\cos(bt) + i\sin(bt)).$$

We have discovered what e^{wt} must be, if the Exponential principle is to hold true, for any complex constant $w = a + bi$:

(6) $$\boxed{e^{(a+bi)t} = e^{at}(\cos bt + i \sin bt)}$$

Let's return to the example (3). The root $r_1 = -1 + i$ leads to

$$e^{(-1+i)t} = e^{-t}(\cos t + i \sin t)$$

and $r_2 = -1 - i$ leads to

$$e^{(-1-i)t} = e^{-t}(\cos t - i \sin t)\,.$$

We probably really wanted a *real* solution to (3), however. For this we have the

Reality Principle:

(7) If z is a solution to a homogeneous linear equation with real coefficients, then the real and imaginary parts of z are too.

We'll explain why this is in a minute, but first let's look at our example (3). The real part of $e^{(-1+i)t}$ is $e^{-t}\cos t$, and the imaginary part is $e^{-t}\sin t$. Both are solutions to (3).

In practice, you should just use the following consequence of what we've done:

Real solutions from complex roots:

If $r_1 = a + bi$ is a root of the characteristic polynomial of a homogeneous linear ODE whose coefficients are constant and real, then

$$e^{at}\cos(bt) \quad \text{and} \quad e^{at}\sin(bt)$$

are solutions. If $b \neq 0$, they are independent solutions.

To see why the Reality Principle holds, suppose z is a solution to a homogeneous linear equation with real coefficients, say

(8) $$\ddot{z} + p\dot{z} + qz = 0$$

for example. Let's write x for the real part of z and y for the imaginary part of z, so $z = x + iy$. Since q is real,

$$\mathrm{Re}\,(qz) = qx \quad \text{and} \quad \mathrm{Im}\,(qz) = qy.$$

Derivatives are computed by differentiating real and imaginary parts separately, so (since p is also real)

$$\mathrm{Re}\,(p\dot{z}) = p\dot{x} \quad \text{and} \quad \mathrm{Im}\,(p\dot{z}) = p\dot{y}.$$

Finally,

$$\mathrm{Re}\,\ddot{z} = \ddot{x} \quad \text{and} \quad \mathrm{Im}\,\ddot{z} = \ddot{y}$$

so when we break down (8) into real and imaginary parts we get

$$\ddot{x} + p\dot{x} + qx = 0\,, \quad \ddot{y} + p\dot{y} + qy = 0$$

—that is, x and y are solutions of the same equation (8).

6.3. **Polar coordinates.** The expression

$$e^{it} = \cos t + i \sin t$$

parametrizes the unit circle in the complex plane. As t increases from 0 to 2π, the complex number $\cos t + i \sin t$ moves once counterclockwise around the circle. The parameter t is just the radian measure counterclockwise from the positive real axis.

More generally,

$$z(t) = e^{(a+bi)t} = e^{at}(\cos(bt) + i \sin(bt)).$$

parametrizes a curve in the complex plane. What is it?

Begin by looking at some values of t. When $t = 0$ we get $z(0) = 1$ no matter what a and b are. When $t = 1$ we get

$$e^{a+bi} = e^{a}(\cos b + i \sin b). \tag{9}$$

The numbers a and b determine the polar coordinates of this point in the complex plane. The absolute value (=magnitude) of $\cos(b)+i\sin(b)$ is 1, so (since $|wz| = |w||z|$ and $e^{at} > 0$)

$$|e^{a+bi}| = e^{a}.$$

This is the radial distance from the origin.

The polar angle—the angle measured counterclockwise from the positive x axis—is called the **argument** of the complex number z, and is written $\mathrm{Arg}z$. According to (9), the argument of e^{a+bi} is simply b. As usual, the argument of a complex number is only well defined up to adding multiples of 2π.

The other polar coordinate—the distance from the origin—is the **modulus** or **absolute value** of the complex number z, and is written $|z|$. According to (9), the modulus of e^{a+bi} is e^{a}.

Any complex number except for zero can be expressed as e^{a+bi} for some a, b. You just need to know a polar expression for the point in the plane.

Exercise 6.3.1. Find expressions of $1, i, 1+i, (1+\sqrt{3}i)/2$, as complex exponentials.

For general t,

$$e^{(a+bi)t} = e^{at}(\cos(bt) + i\sin(bt)) \tag{10}$$

parametrizes a spiral (at least when $b \neq 0$). If $a > 0$, it runs away from the origin, exponentially, while winding around the origin (counterclockwise if $b > 0$, clockwise if $b < 0$). If $a < 0$, it decays exponentially towards the origin, while winding around the origin. Figure 3 shows a picture of the curve parametrized by $e^{(1+2\pi i)t}$.

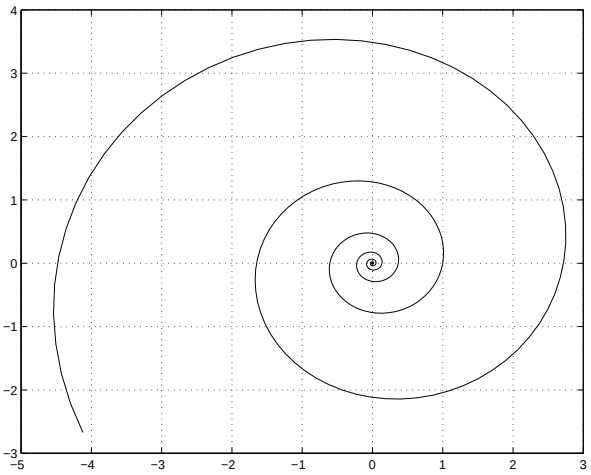


FIGURE 3. The spiral $z = e^{(1+2\pi i)t}$

If $a = 0$ equation (10) parametrizes a circle. If $b = 0$, the curve lies on the positive real axis.

6.4. **Multiplication.** Multiplication of complex numbers is expressed very beautifully in these polar terms. We already know that

$$\textbf{Magnitudes Multiply:} \qquad |wz| = |w||z|. \tag{11}$$

To understand what happens to arguments we have to think about the product $e^r e^s$, where r and s are two complex numbers. This is a major test of the reasonableness of our definition of the complex

exponential, since we know what this product ought to be (and what it *is* for r and s real). It turns out that the notation is well chosen:

Exponential Law:

(12) $\boxed{\text{For any complex numbers } r \text{ and } s,\ e^{r+s} = e^r e^s}$

This fact comes out of the uniqueness of solutions of ODEs. To get an ODE, let's put t into the picture: we claim that

(13) $$e^{r+st} = e^r e^{st}.$$

If we can show this, then the Exponential Law as stated is the case $t = 1$. Differentiate each side of (13), using the chain rule for the left hand side and the product rule for the right hand side:

$$\frac{d}{dt}e^{r+st} = \frac{d(r+st)}{dt}e^{r+st} = se^{r+st}\,, \quad \frac{d}{dt}(e^r e^{st}) = e^r \cdot se^{st}.$$

Both sides of (13) thus satisfy the IVP

$$\dot{z} = sz\,, \quad z(0) = e^r,$$

so they are equal.

In particular, we can let $r = i\alpha$ and $s = i\beta$:

(14) $$e^{i\alpha}e^{i\beta} = e^{i(\alpha+\beta)}.$$

In terms of polar coordinates, this says that

(15) **Angles Add:** $\quad \mathrm{Arg}(wz) = \mathrm{Arg}(w) + \mathrm{Arg}(z).$

Exercise 6.4.1. Compute $((1+\sqrt{3}i)/2)^3$ and $(1+i)^4$ afresh using these polar considerations.

Exercise 6.4.2. Derive the addition laws for cosine and sine from Euler's formula and (14). Understand this exercise and you'll never have to remember those formulas again.

6.5. **Roots of unity and other numbers.** The polar expression of multiplication is useful in finding roots of complex numbers. Begin with the sixth roots of 1, for example. We are looking for complex numbers z such that $z^6 = 1$. Since *moduli multiply*, $|z|^6 = |z^6| = |1| = 1$, and since moduli are nonnegative this forces $|z| = 1$: all the sixth roots of 1 are on the unit circle. *Arguments add*, so the argument of a sixth root of 1 is an angle θ so that 6θ is a multiple of 2π (which are the angles giving 1). Up to addition of multiples of 2π there are six such angles: $0, \pi/3, 2\pi/3, \pi, 4\pi/3$, and $5\pi/3$. The resulting points on the unit circle divide it into six equal arcs. From this and some geometry or trigonometry it's easy to write down the roots as $a + bi$: ± 1 and

$(\pm 1 \pm \sqrt{3}i)/2$. In general, the nth roots of 1 break the circle evenly into n parts.

Exercise 6.5.1. Write down the eighth roots of 1 in the form $a + bi$.

Now let's take roots of numbers other than 1. Start by finding a single nth root z of the complex number $w = re^{i\theta}$ (where r is a positive real number). Since magnitudes multiply, $|z| = \sqrt[n]{r}$. Since angles add, one choice for the argument of z is θ/n: one nth of the way up from the positive real axis. Thus for example one square root of $4i$ is the complex number with magnitude 2 and argument $\pi/4$, which is $\sqrt{2}(1 + i)$. To get all the nth roots of w notice that you can multiply one by any nth root of 1 and get another nth root of w. Angles add and magnitudes multiply, so the effect of this is just to add a multiple of $2\pi/n$ to the angle of the first root we found. There are n distinct nth roots of any nonzero complex number $|w|$, and they divide the circle with center 0 and radius $\sqrt[n]{r}$ evenly into n arcs.

Exercise 6.5.2. Find all the cube roots of -8. Find all the sixth roots of $-i/64$.

We can use our ability to find complex roots to solve more general polynomial equations.

Exercise 6.5.3. Find all the roots of the polynomials $x^3 + 1$, $ix^2 + x + (1 + i)$, and $x^4 - 2x^2 + 1$.

7. Beats

7.1. **What beats are.** Musicians tune their instruments using "beats." Beats occur when two very nearby pitches are sounded simultaneously. We'll make a mathematical study of this effect, using complex numbers.

We will study the sum of two sinusoidal functions. We might as well take one of them to be $a\sin(\omega_0 t)$, and adjust the phase of the other accordingly. So the other can be written as $b\sin((1+\epsilon)\omega_0 t - \phi)$: amplitude b, circular frequency written in terms of the frequency of the first sinusoid as $(1+\epsilon)\omega_0$, and phase lag ϕ.

We will take $\phi = 0$ for the moment, and add it back in later. So we are studying

$$x = a\sin(\omega_0 t) + b\sin((1+\epsilon)\omega_0 t).$$

We think of ϵ as a small number, so the two frequencies are relatively close to each other.

One case admits a simple discussion, namely when the two amplitudes are equal: $a = b$. Then the trig identity

$$\sin(\alpha+\beta) + \sin(\alpha-\beta) = 2\cos(\beta)\sin(\alpha)$$

with $\alpha = (1+\epsilon/2)\omega_0 t$ and $\beta = \epsilon\omega_0 t/2$ gives us the equation

$$x = a\sin(\omega_0 t) + a\sin((1+\epsilon)\omega_0 t) = 2a\cos\left(\frac{\epsilon\omega_0 t}{2}\right)\sin\left(\left(1+\frac{\epsilon}{2}\right)\omega_0 t\right).$$

(The trig identity is easy to prove using complex numbers: Compute

$$e^{i(\alpha+\beta)} + e^{i(\alpha-\beta)} = (e^{i\beta} + e^{-i\beta})e^{i\alpha} = 2\cos(\beta)e^{i\alpha}$$

using (6.5); then take imaginary parts.)

We might as well take $a > 0$. When ϵ is small, the period of the cosine factor is much longer than the period of the sine factor. This lets us think of the product as a wave of circular frequency $(1+\epsilon/2)\omega_0$—that is, the average of the circular frequences of the two constituent waves—giving the audible tone, whose amplitude is modulated by multiplying it by

$$g(t) = 2a\left|\cos\left(\frac{\epsilon\omega_0 t}{2}\right)\right|. \tag{1}$$

The function $g(t)$ the "envelope" of x. The function $x(t)$ oscillates rapidly between $-g(t)$ and $+g(t)$.

To study the more general case, in which a and b differ, we will study the function made of complex exponentials,

$$z = ae^{i\omega_0 t} + be^{i(1+\epsilon)\omega_0 t}.$$

The original function x is the imaginary part of z.

We can factor out $e^{i\omega_0 t}$:

$$z = e^{i\omega_0 t}(a + be^{i\epsilon\omega_0 t}).$$

This gives us a handle on the magnitude of z, since the magnitude of the first factor is 1. Using the formula $|w|^2 = w\bar{w}$ on the second factor, we get

$$|z|^2 = a^2 + b^2 + 2ab\cos(\epsilon\omega_0 t).$$

The imaginary part of a complex number z lies between $-|z|$ and $+|z|$, so $x = \operatorname{Im} z$ oscillates between $-|z|$ and $+|z|$. The function $g(t) = |z(t)|$, i.e.

$$g(t) = \sqrt{a^2 + b^2 + 2ab\cos(\epsilon\omega_0 t)}, \tag{2}$$

thus serves as an "envelope," giving the values of the peaks of the oscillations exhibited by $x(t)$.

This envelope shows the "beats" effect. It reaches maxima when $\cos(\epsilon\omega_0 t)$ does, i.e. at the times $t = 2k\pi/\epsilon\omega_0$ for whole numbers k. A single beat lasts from one maximum to the next: the period of the beat is

$$P_b = \frac{2\pi}{\epsilon\omega_0} = \frac{P_0}{\epsilon}$$

where $P_0 = 2\pi/\omega_0$ is the period of $\sin(\omega_0 t)$. The maximum amplitude is then $a + b$, i.e. the sum of the amplitudes of the two constituent waves; this occurs when their phases are lined up so they reinforce. The minimum amplitude occurs when the cosine takes on the value -1, i.e. when $t = (2k+1)\pi/\epsilon\omega_0$ for whole numbers k, and is $|a - b|$. This is when the two waves are perfectly out of sync, and experience destructive interference.

Figure 4 is a plot of beats with $a = 1, b = .5, \omega_0 = 1, \epsilon = .1, \phi = 0$, showing also the envelope.

Now let's allow ϕ to be nonzero. The effect on the work done above is to replace $\epsilon\omega_0 t$ by $\epsilon\omega_0 t - \phi$ in the formulas (2) for the envelope $g(t)$. Thus the beat gets shifted by the same phase as the second signal.

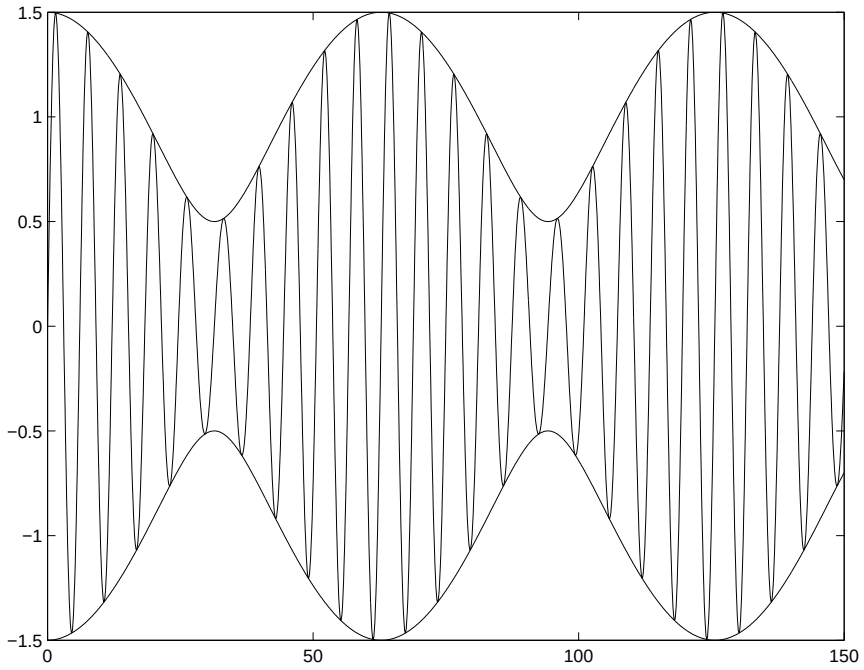


FIGURE 4. Beats, with envelope

If $b \neq 1$ it is not very meaningful to compute the pitch, i.e. the frequency of the wave being modulated by the envelope. It lies somewhere between the two initial frequencies, and it varies periodically with period P_b.

7.2. **What beats are not.** Many differential equations textbooks present beats as a system response when a harmonic oscillator is driven by a signal whose frequency is close to the natural frequency of the oscillator. This is true as a piece of mathematics, but it is almost never the way beats occur in nature. The reason is that if there is any damping in the system, the "beats" die out very quickly to a steady sinusoidal solution, and it is that solution which is observed.

Explicitly, the Exponential Response Formula (Section 12, equation 3) shows that the equation

$$\ddot{x} + \omega_n^2 x = \cos(\omega t)$$

has the periodic solution

$$x_p = \frac{\cos(\omega t)}{\omega^2 - \omega_n^2}$$

unless $\omega = \omega_n$. If ω and ω_n are close, the amplitude of the periodic solution is large; this is "near resonance." Adding a little damping won't change that solution very much, but it will convert homogeneous solutions from sinusoids to *damped* sinusoids, i.e. transients, and rather quickly any solution becomes indistinguishable from x_p.

So beats do not occur this way in engineering situations. But they do occur. They are used for example in reconstructing an amplitude-modulated signal from a frequency-modulated ("FM") radio signal. The radio receiver produces a signal at a fixed frequency ν, and adds it to the received signal, whose frequency differs slightly from ν. The result is a beat, and the beat frequency is the audible frequency.

Differential equations textbooks also always arrange initial conditions in a very artificial way, so that the solution is a sum of the periodic solution x_p and a homogeneous solution x_h having exactly the same amplitude as x_p. They do this by imposing the initial condition $x(0) = \dot{x}(0) = 0$. This artifice puts them into the simple situation $a = b$ mentioned above. For the general case one has to proceed as we did, using complex exponentials.

8. RLC circuits

8.1. **Series RLC Circuits.** Electric circuits provide an important example of linear, time-invariant differential equations, alongside mechanical systems. We will consider only the simple series circuit pictured below.

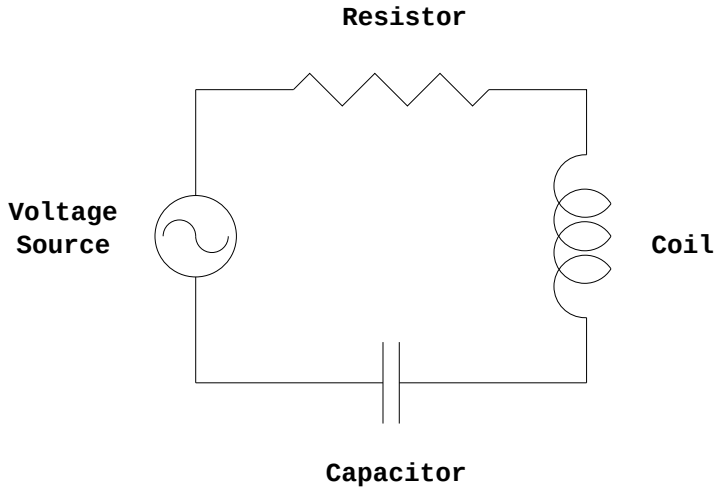


FIGURE 5. Series RLC Circuit

The Mathlet `Series RLC Circuit` exhibits the behavior of this system, when the voltage source provides a sinusoidal signal.

Current flows through the circuit; in this simple loop circuit the current through any two points is the same at any given moment. Current is denoted by the letter I, or $I(t)$ since it is generally a function of time.

The current is created by a force, the "electromotive force," which is determined by voltage *differences*. The voltage *drop* across a component of the system *except for the power source* will be denoted by V with a subscript. Each is a function of time. If we orient the circuit consistently, say clockwise, then we let

$V_L(t)$ denote the voltage drop across the coil
$V_R(t)$ denote the voltage drop across the resistor
$V_C(t)$ denote the voltage drop across the capacitor
$V(t)$ denote the voltage *increase* across the power source

"Kirchoff's Voltage Law" then states that

$$V = V_L + V_R + V_C \tag{1}$$

The circuit components are characterized by the relationship between the current flowing through them and the voltage drop across them:

$$\begin{aligned} &\text{Coil :} && V_L = L\dot{I} \\ &\text{Resistor :} && V_R = RI \\ &\text{Capacitor :} && \dot{V}_C = (1/C)I \end{aligned} \tag{2}$$

The constants here are the "inductance" L of the coil, the "resistance" R of the resistor, and the *inverse* of the "capacitance" C of the capacitor. A very large capacitor, with C large, is almost like no capacitor at all; electrons build up on one plate, and push out electrons on the other, to form an uninterrupted circuit. We'll say a word about the actual units below.

To get the expressions (2) into comparable form, differentiate the first two. Differentiating (1) gives $\dot{V}_L + \dot{V}_R + \dot{V}_C = \dot{V}$, and substituting the values for $\dot{V}_L$, $\dot{V}_R$, and $\dot{V}_C$ gives us

$$\boxed{L\ddot{I} + R\dot{I} + (1/C)I = \dot{V}} \tag{3}$$

This equation describes how I is determined from the impressed voltage V. It is a second order linear time invariant ODE. Comparing it with the familiar equation

$$\boxed{m\ddot{x} + b\dot{x} + kx = F} \tag{4}$$

governing the displacement in a spring-mass-dashpot system reveals an analogy between the two types of system:

Mechanical	**Electrical**
Mass	Coil
Damper	Resistor
Spring	Capacitor
Driving force	Time derivative of impressed voltage
Displacement	Current

8.2. **A word about units.** There is a standard system of units called the International System of Units, SI, formerly known as the mks (meter-kilogram-second) system. In terms of those units, (3) is correct when:

inductance L is measured in henries, H
resistance R is measured in ohms, Ω
capacitance C is measured in farads, F
voltgage V is measured in volts, also denoted V
current I is measured in amperes, A

Balancing units in the equation shows that

$$\frac{\text{henry} \cdot \text{ampere}}{\text{sec}^2} = \frac{\text{ohm} \cdot \text{ampere}}{\text{sec}} = \frac{\text{ampere}}{\text{farad}} = \frac{\text{volt}}{\text{sec}}$$

Thus one henry is the same as one volt-second per ampere.

The analogue for mechanical units is this:

mass m is measured in kilograms, kg
damping constant b is measured in kg/sec
spring constant k is measured in kg/sec^2
applied force F is measured in newtons, N
displacement x is measured in meters, m

Here

$$\text{newton} = \frac{\text{kg} \cdot \text{m}}{\text{sec}^2}$$

so another way to describe the units in which the spring constant is measured in is as newtons per meter—the amount of force it produces when stretched by one meter.

9. Normalization of solutions

9.1. **Initial conditions.** The general solution of any homogeneous linear second order ODE

$$\ddot{x} + p(t)\dot{x} + q(t)x = 0 \tag{1}$$

has the form $c_1x_1 + c_2x_2$, where c_1 and c_2 are constants. The solutions x_1, x_2 are often called "basic," but this is a poorly chosen name since it is important to understand that there is absolutely nothing special about the solutions x_1, x_2 in this formula, beyond the fact that *neither is a multiple of the other.*

For example, the ODE $\ddot{x} = 0$ has general solution $at + b$. We can take $x_1 = t$ and $x_2 = 1$ as basic solutions, and have a tendency to do this or else list them in the reverse order, so $x_1 = 1$ and $x_2 = t$. But equally well we could take a pretty randomly chosen pair of polynomials of degree at most one, such as $x_1 = 4 + t$ and $x_2 = 3 - 2t$, as basic solutions. This is because for any choice of a and b we can solve for c_1 and c_2 in $at + b = c_1x_1 + c_2x_2$. The only requirement is that neither solution is a multiple of the other. This condition is expressed by saying that the pair $\{x_1, x_2\}$ is *linearly independent.*

Given a basic pair of solutions, x_1, x_2, there is a solution of the initial value problem with $x(t_0) = a, \dot{x}(t_0) = b$, of the form $x = c_1x_1 + c_2x_2$. The constants c_1 and c_2 satisfy

$$a = x(t_0) = c_1x_1(t_0) + c_2x_2(t_0)$$

$$b = \dot{x}(t_0) = c_1\dot{x}_1(t_0) + c_2\dot{x}_2(t_0).$$

For instance, the ODE $\ddot{x} - x = 0$ has exponential solutions e^t and e^{-t}, which we can take as x_1, x_2. The initial conditions $x(0) = 2, \dot{x}(0) = 4$ then lead to the solution $x = c_1e^t + c_2e^{-t}$ as long as c_1, c_2 satisfy

$$2 = x(0) = c_1e^0 + c_2e^{-0} = c_1 + c_2,$$

$$4 = \dot{x}(0) = c_1e^0 + c_2(-e^{-0}) = c_1 - c_2,$$

This pair of linear equations has the solution $c_1 = 3, c_2 = -1$, so $x = 3e^t - e^{-t}$.

9.2. **Normalized solutions.** Very often you will have to solve the same differential equation subject to several different initial conditions. It turns out that one can solve for just *two* well chosen initial conditions, and then the solution to *any other* IVP is instantly available. Here's how.

Definition 9.2.1. A pair of solutions x_1, x_2 of (1) is *normalized at* t_0 if

$$x_1(t_0) = 1, \qquad x_2(t_0) = 0,$$
$$\dot{x}_1(t_0) = 0, \qquad \dot{x}_2(t_0) = 1.$$

By existence and uniqueness of solutions with given initial conditions, there is always exactly one pair of solutions which is normalized at t_0.

For example, the solutions of $\ddot{x} = 0$ which are normalized at 0 are $x_1 = 1, x_2 = t$. To normalize at $t_0 = 1$, we must find solutions—polynomials of the form $at + b$—with the right values and derivatives at $t = 1$. These are $x_1 = 1, x_2 = t - 1$.

For another example, the "harmonic oscillator"

$$\ddot{x} + \omega_n^2 x = 0$$

has basic solutions $\cos(\omega_n t)$ and $\sin(\omega_n t)$. They are normalized at 0 only if $\omega_n = 1$, since $\dfrac{d}{dt}\sin(\omega_n t) = \omega_n \cos(\omega_n t)$ has value ω_n at $t = 0$, rather than value 1. We can fix this (as long as $\omega_n \neq 0$) by dividing by ω_n: so

$$\cos(\omega_n t)\,, \qquad \omega_n^{-1}\sin(\omega_n t) \tag{2}$$

is the pair of solutions to $\ddot{x} + \omega_n^2 x = 0$ which is normalized at $t_0 = 0$.

Here is another example. The equation $\ddot{x} - x = 0$ has linearly independent solutions e^t, e^{-t}, but these are not normalized at any t_0 (for example because neither is ever zero). To find x_1 in a pair of solutions normalized at $t_0 = 0$, we take $x_1 = ae^t + be^{-t}$ and find a, b such that $x_1(0) = 1$ and $\dot{x}_1(0) = 0$. Since $\dot{x}_1 = ae^t - be^{-t}$, this leads to the pair of equations $a + b = 1, a - b = 0$, with solution $a = b = 1/2$. To find $x_2 = ae^t + be^{-t}$ $x_2(0) = 0, \dot{x}_2(0) = 1$ imply $a + b = 0, a - b = 1$ or $a = 1/2, b = -1/2$. Thus our normalized solutions x_1 and x_2 are the *hyperbolic sine* and *cosine* functions:

$$\cosh t = \frac{e^t + e^{-t}}{2}, \qquad \sinh t = \frac{e^t - e^{-t}}{2}\,.$$

These functions are important precisely because they occur as normalized solutions of $\ddot{x} - x = 0$.

Normalized solutions are always linearly independent: x_1 can't be a multiple of x_2 because $x_1(t_0) \neq 0$ while $x_2(t_0) = 0$, and x_2 can't be a multiple of x_1 because $\dot{x}_2(t_0) \neq 0$ while $\dot{x}_1(t_0) = 0$.

Now suppose we wish to solve (1) with the general initial conditions.

If x_1 and x_2 are a pair of solutions normalized at t_0, then the solution x with $x(t_0) = a$, $\dot{x}(t_0) = b$ is

$$x = ax_1 + bx_2 \,.$$

The constants of integration *are* the initial conditions.

If I want x such that $\ddot{x} + x = 0$ and $x(0) = 3, \dot{x}(0) = 2$, for example, we have $x = 3\cos t + 2\sin t$. Or, for an other example, the solution of $\ddot{x} - x = 0$ for which $x(0) = 2$ and $\dot{x}(0) = 4$ is $x = 2\cosh(t) + 4\sinh(t)$. You can check that this is the same as the solution given above.

Exercise 9.2.2. Check the identity

$$\cosh^2 t - \sinh^2 t = 1 \,.$$

9.3. **ZSR and ZIR.** There is an interesting way to decompose the solution of a linear initial value problem which is appropriate to the *inhomogeneous* case and which arises in the system/signal approach. Two distinguishable bits of data determine the choice of solution: the initial condition, and the input signal.

Suppose we are studying the initial value problem

$$\ddot{x} + p(t)\dot{x} + q(t)x = f(t)\,, \quad x(t_0) = x_0\,, \quad \dot{x}(t_0) = \dot{x}_0\,. \tag{3}$$

There are two related initial value problems to consider:

[**ZSR**] The *same* ODE but with *rest* initial conditions (or "zero state"):

$$\ddot{x} + p(t)\dot{x} + q(t)x = f(t)\,, \quad x(t_0) = 0\,, \quad \dot{x}(t_0) = 0\,.$$

Its solution is called the **Zero State Response** or **ZSR.** It depends entirely on the input signal, and assumes zero initial conditions. We'll write x_f for it, using the notation for the input signal as subscript.

[**ZIR**] The associated *homogeneous* ODE with the *given* initial conditions:

$$\ddot{x} + p(t)\dot{x} + q(t)x = 0\,, \quad x(t_0) = x_0\,, \quad \dot{x}(t_0) = \dot{x}_0\,.$$

Its solution is called the the **Zero Input Response**, or **ZIR**. It depends entirely on the initial conditions, and assumes null input signal. We'll write x_h for it, where h indicates "homogeneous."

By the superposition principle, the solution to (3) is precisely

$$x = x_f + x_h.$$

The solution to the initial value problem (3) is the sum of a ZSR and a ZIR, in exactly one way.

Example 9.3.1. Drive a harmonic oscillator with a sinusoidal signal:

$$\ddot{x} + \omega_n^2 x = a\cos(\omega t)$$

(so $f(t) = a\cos(\omega t)$) and specify initial conditions $x(0) = x_0$, $\dot{x}(0) = \dot{x}_0$. Assume that the system is not in resonance with the signal, so $\omega \neq \omega_n$. Then the Exponential Response Formula (Section 10) shows that the general solution is

$$x = a\frac{\cos(\omega t)}{\omega_n^2 - \omega^2} + b\cos(\omega_n t) + c\sin(\omega_n t)$$

where b and c are constants of integration. To find the ZSR we need to find $\dot{x}$, and then arrange the constants of integration so that both $x(0) = 0$ and $\dot{x}(0) = 0$. Differentiate to see

$$\dot{x} = -a\omega\frac{\sin(\omega t)}{\omega_n^2 - \omega^2} - b\omega_n\sin(\omega_n t) + c\omega_n\cos(\omega_n t)$$

so $\dot{x}(0) = c\omega_n$, which can be made zero by setting $c = 0$. Then $x(0) = a/(\omega_n^2 - \omega^2) + b$, so $b = -a/(\omega_n^2 - \omega^2)$, and the ZSR is

$$x_f = a\,\frac{\cos(\omega t) - \cos(\omega_n t)}{\omega_n^2 - \omega^2}\,.$$

The ZIR is

$$x_h = b\cos(\omega_n t) + c\sin(\omega_n t)$$

where this time b and c are chosen so that $x_h(0) = x_0$ and $\dot{x}_h(0) = \dot{x}_0$. Thus (using (2) above)

$$x_h = x_0\cos(\omega_n t) + \dot{x}_0\frac{\sin(\omega_n t)}{\omega_n}\,.$$

Example 9.3.2. The same works for linear equations of any order. For example, the solution to the bank account equation (Section 2)

$$\dot{x} - Ix = c\,, \quad x(0) = x_0,$$

(where we'll take the interest rate I and the rate of deposit c to be constant, and $t_0 = 0$) can be written as

$$x = \frac{c}{I}(e^{It} - 1) + x_0 e^{It}.$$

The first term is the ZSR, depending on c and taking the value 0 at $t = 0$. The second term is the ZIR, a solution to the homogeneous equation depending solely on x_0.

10. Operators and the exponential response formula

10.1. **Operators.** Operators are to functions as functions are to numbers. An operator takes a function, does something to it, and returns this modified function. There are lots of examples of operators around:

—The *shift-by-a operator* (where a is a number) takes as input a function $f(t)$ and gives as output the function $f(t-a)$. This operator shifts graphs to the right by a units.

—The *multiply-by-h(t) operator* (where $h(t)$ is a function) multiplies by $h(t)$: it takes as input the function $f(t)$ and gives as output the function $h(t)f(t)$.

You can go on to invent many other operators. In this course the most important operator is:

—The *differentiation operator*, which carries a function $f(t)$ to its derivative $f'(t)$.

The differentiation operator is usually denoted by the letter D; so $Df(t)$ is the function $f'(t)$. D carries f to f'; for example, $Dt^3 = 3t^2$. Warning: you can't take this equation and substitute $t = 2$ to get $D8 = 12$. The only way to interpret "8" in "$D8$" is as a *constant* function, which of course has derivative zero: $D8 = 0$. The point is that in order to know the function $Df(t)$ at a particular value of t, say $t = a$, you need to know more than just $f(a)$; you need to know how $f(t)$ is changing near a as well. This is characteristic of operators; in general you have xto expect to need to know the *whole* function $f(t)$ in order to evaluate an operator on it.

The *identity operator* takes an input function $f(t)$ and returns the *same* function, $f(t)$; it does nothing, but it still gets a symbol, I.

Operators can be added and multiplied by numbers or more generally by functions. Thus $tD+4I$ is the operator sending $f(t)$ to $tf'(t)+4f(t)$.

The single most important thing associated with the concept of operators is that they can be *composed* with each other. I can hand a function off from one operator to another, each taking the output from the previous and modifying it further. For example, D^2 differentiates twice: it is the second-derivative operator, sending $f(t)$ to $f''(t)$.

We have been studying ODEs of the form $m\ddot{x} + b\dot{x} + kx = q(t)$. The left hand side is the effect of an operator on the function $x(t)$, namely, the operator $mD^2 + bD + kI$. This *operator* describes the *system* (composed for example of a mass, dashpot, and spring).

We'll often denote an operator by a single capital letter, such as L. If $L = mD^2 + bD + kI$, for example, then our favorite ODE,

$$m\ddot{x} + b\dot{x} + kx = q$$

can be written simply as

$$Lx = q.$$

At this point m, b, and k could be functions of t.

Note well: the operator does NOT take the signal as input and return the system response, but rather the reverse: $Lx = q$, the operator takes the response and returns the signal. In a sense the system is better modeled by the "inverse" of the operator L. In rough terms, solving the ODE $Lx = q$ amounts to inverting the operator L.

Here are some definitions. A **differential operator** is one which is algebraically composed of D's and multiplication by functions. The **order** of a differential operator is the highest derivative appearing in it. $mD^2+bD+kI$ is an example of a second order differential operator.

This example has another important feature: it is *linear*. An operator L is *linear* if

$$L(cf) = cLf \qquad \text{and} \qquad L(f + g) = Lf + Lg.$$

10.2. **LTI operators and exponential signals.** We will study almost exclusively linear differential operators. They are the operators of the form

$$L = a_n(t)D^n + a_{n-1}(t)D^{n-1} + \cdots + a_0(t)I.$$

The functions $a_0, \ldots, a_n$ are the **coefficients** of L.

In this course we focus on the case in which the coefficients are *constant*; each a_k is thus a *number*, and we can form the **characteristic polynomial** of the operator,

$$p(s) = a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0.$$

The operator is **Linear** and **Time Invariant**: an **LTI** operator. The original operator is obtained from its characteristic polynomial by formally replacing the indeterminate s here with the differentiation operator D, so we may write

$$L = p(D).$$

The characteristic polynomial completely determines the operator, and many properties of the operator are conveniently described in terms of its characteristic polynomial.

Here is a first example of the power of the operator notation. Let r be any constant. (You might as well get used to thinking of it as a possibly *complex* constant.) Then

$$De^{rt} = re^{rt}.$$

(A fancy expression for this is to say that r is an *eigenvalue* of the operator D, with corresponding *eigenfunction* e^{rt}.) Iterating this we find that

$$D^k e^{rt} = r^k e^{rt}.$$

We can put these equations together, for varying k, and evaluate a general LTI operator

$$p(D) = a_n D^n + a_{n-1} D^{n-1} + \cdots + a_0 I$$

on e^{rt}. The operator D^k pulls r^k out as a factor, and when you add them all up you get the value of the polynomial $p(s)$ at $s = r$:

$$p(D)e^{rt} = p(r)e^{rt}. \tag{1}$$

It is crucial here that the operator be time invariant: If the coefficients a_k are not constant, then they don't just pull outside the differentiation; you need to use the product rule instead, and the formulas become more complicated—see Section 12.

Multiplying (1) by $a/p(r)$ we find the important

Exponential Response Formula: A solution to

$$p(D)x = ae^{rt} \tag{2}$$

is given by the

$$\boxed{x_p = a\frac{e^{rt}}{p(r)}} \tag{3}$$

provided only that $p(r) \neq 0$.

The Exponential Response Formula ties together many different parts of this course. Since the most important signals are exponential, and the most important differential operators are LTI operators, this single formula solves most of the ODEs you are likely to face in your future.

The function x_p given by (3) is the *only* solution to (2) which is a multiple of an exponential function. If r has the misfortune to be a root of $p(s)$, so that $p(r) = 0$, then the formula (3) would give a zero in the denominator. The conclusion is that there are *no* solutions which are multiples of exponential functions. This is a "resonance" situation. In this case we can still find an explicit solution; see Section 12 for this.

Example 10.2.1. Let's solve

$$2\ddot{x} + \dot{x} + x = 1 + 2e^t \,. \tag{4}$$

This is an inhomogeneous linear equation, so the general solution as the form $x_p + x_h$, where x_p is any particular solution and x_h is the general homogeneous solution. The characteristic polynomial is $p(s) = 2s^2 + s + 1$, with roots $(-1 \pm \sqrt{7}\, i)/4$ and hence general homogeneous solution is given by $x_h = e^{-t/4}(a\cos(\sqrt{7}\, t/4) + b\sin(\sqrt{7}\, t/4))$, or, in polar expression, $A\cos(\sqrt{7}\, t/4 - \phi)$.

The inhomogeneous equation is $p(D)x = 1 + 2e^t$. The input signal is a linear combination of 1 and e^t, so, again by superposition, if x_1 is a solution of $p(D)x = 1$ and x_2 is a solution of $p(D)x = e^t$, then a solution to (4) is given by $x_p = x_1 + 2x_2$.

The constant function 1 is exponential: $1 = e^{rt}$ with $r = 0$. Thus $p(D)x = 1$ has for solution $1/p(0) = 1$. This is easily checked without invoking the Exponential Response Formula! So take $x_1 = 1$.

Similarly, we can take $x_2 = e^t/p(1) = e^t/4$. Thus

$$x_p = 1 + 2e^t/4 \,.$$

Remark 10.2.2. The quantity

$$W(s) = \frac{1}{p(s)}$$

that occurs in the Exponential Response Formula (3) is the **transfer function** of the system. One usually encounters this in the context of the Laplace transform, but it has a clear interpretation for us already: for any given r, one response of the system to the exponential signal e^{rt} is simply $W(r)e^{rt}$ (as long as $p(r) \neq 0$).

The transfer function is sometimes called the **system function** (e.g. by Oppenheim and Willsky) or the **complex gain**, and it is often written as $H(s)$.

10.3. **Sinusoidal signals.** Being able to handle exponential signals is even more significant than you might think at first, because of the richness of the *complex* exponential. To exploit this richness, we have to allow complex valued functions of t. The main complex valued function we have to consider is the complex exponential function $z = e^{wt}$, where w is some fixed complex number. We know its derivative, by the Exponential Principle (Section 6.1): $\dot{z} = we^{wt}$.

Here's how we can use this. Suppose we want to solve

$$2\ddot{x} + \dot{x} + x = 2\cos(t/2). \tag{5}$$

Step 1. Find a complex valued equation with an exponential signal of which this is the real part.

There is more than one way to do this, but the most natural one is to view $2\cos(t/2)$ as the real part of $2e^{it/2}$ and write down

$$2\ddot{z} + \dot{z} + z = 2e^{it/2}\,. \tag{6}$$

This is a *new* equation, different from the original one. Its solution deserves a different name, and we have chosen one for it: z. This introduction of a new variable name is an essential part of Step 1. The real part of a solution to (6) is a solution to (5): $\operatorname{Re} z = x$.

(If the input signal is sinusoidal, it is some shift of a cosine. This can be handled by the method described below in Section 10.5. Alternatively, if it is a sine, you can write the equation as the imaginary part of an equation with exponential input signal, and proceed as below.)

Step 2. Find a particular solution z_p to the new equation.

By the Exponential Response Formula (3)

$$z_p = 2\frac{e^{it/2}}{p(i/2)}\,.$$

Compute:

$$p(i/2) = 2(i/2)^2 + i/2 + 1 = (1+i)/2$$

so

$$z_p = 4\frac{e^{it/2}}{1+i}\,. \tag{7}$$

Step 3. Extract the real (or imaginary) part of z_p to recover x_p. The result will be a sinusoidal function, and there are good ways to get to either expression for a sinusoidal function.

Rectangular version. Write out the real and imaginary parts of the exponential and rationalize the denominator:

$$z_p = 4\frac{(1-i)(\cos(t/2) + i\sin(t/2))}{1+1}\,.$$

The real part is

$$x_p = 2\cos(t/2) + 2\sin(t/2)\,, \tag{8}$$

and there is our solution!

Polar version. To do this, write the factor

$$\frac{2}{p(i/2)} = \frac{4}{1+i} = 2(1-i)$$

which shows up in the Exponential Response Formula in polar form:

$$\frac{2}{p(i/2)} = ge^{-i\phi}\,,$$

so g is the magnitude and $-\phi$ is the angle. (We use $-\phi$ instead of ϕ is because we will want to wind up with a phase *lag*.) The magnitude is

$$g = \frac{4}{|1+i|} = 2\sqrt{2}\,.$$

The angle ϕ is the argument of the denominator $p(i/2) = 1+i$, which is $\pi/4$. Thus

$$z_p = ge^{-\phi i}e^{it/2} = 2\sqrt{2}e^{(t/2-(\pi/4))i}\,.$$

The real part is now exactly

$$x_p = 2\sqrt{2}\cos(t/2-\pi/4)\,.$$

These two forms of the sinusoidal solution are related to each other by the relation (3) in Section 4. The polar form has the advantage of exhibiting a clear relationship between the input signal and the sinusoidal system response: the amplitude is multiplied by a factor of $\sqrt{2}$—this is the **gain**—and there is a phase lag of $\pi/4$ behind the input signal. In Section 10.5 we will observed that these two features persist for *any* sinusoidal input signal with circular frequency 1/2.

Example 10.3.1. The harmonic oscillator with sinusoidal forcing term:

$$\ddot{x} + \omega_n^2 x = A\cos(\omega t)\,.$$

This is the real part of the equation

$$\ddot{z} + \omega_n^2 z = Ae^{i\omega t}\,,$$

which we can solve directly from the Exponential Response Formula: since $p(i\omega) = (i\omega)^2 + \omega_n^2 = \omega_n^2 - \omega^2$,

$$z_p = A\frac{e^{i\omega t}}{\omega_n^2 - \omega^2}$$

as long as the input frequency is different from the natural frequency of the harmonic oscillator. Since the denominator is *real*, the real part of z_p is easy to find:

$$x_p = A\frac{\cos(\omega t)}{\omega_n^2 - \omega^2}\,. \tag{9}$$

Similarly, the sinusoidal solution to

$$\ddot{y} + \omega_n^2 y = A\sin(\omega t)$$

is the imaginary part of z_p,

$$y_p = A\frac{\sin(\omega t)}{\omega_n^2 - \omega^2}\,. \tag{10}$$

This solution puts in precise form some of the things we can check from experimentation with vibrating systems. When the frequency of the signal is smaller than the natural frequency of the system, $\omega < \omega_n$, the denominator is positive. The effect is that the system response is a *positive* multiple of the signal: the vibration of the mass is "in sync" with the impressed force. As ω increases towards ω_n, the denominator in (9) nears zero, so the amplitude of the solution grows arbitrarily large. When $\omega = \omega_n$ the system is **in resonance** with the signal; the Exponential Response Formula fails, and there is *no* periodic (or even bounded) solution. (We'll see in Section 12 how to get a solution in this case.) When $\omega > \omega_n$, the denominator is negative. The system response is a *negative* multiple of the signal: the vibration of the mass is perfectly "out of sync" with the impressed force.

Since the coefficients are constant here, a time-shift of the signal results in the same time-shift of the solution:

$$\ddot{x} + \omega_n^2 x = A\cos(\omega t - \phi)$$

has the periodic solution

$$x_p = A\frac{\cos(\omega t - \phi)}{\omega_n^2 - \omega^2}.$$

The equations (9) and (10) will be very useful to us when we solve ODEs via Fourier series.

10.4. **Damped sinusoidal signals.** The same procedure may be used to solve equations of the form

$$Lx = e^{at}\cos(\omega t - \phi_0)$$

where $L = p(D)$ is any LTI differential operator.

Example 10.4.1. Let's solve

$$2\ddot{x} + \dot{x} + x = e^{-t}\cos t$$

We found the general solution of the homogeneous equation above, in Example 10.2.1, so what remains is to find a particular solution. To do this, replace the equation by complex-valued equation of which it is the real part:

$$2\ddot{x} + \dot{x} + x = e^{(-1+i)t}$$

Then apply the Exponential Response Formula:

$$z_p = \frac{e^{(-1+i)t}}{p(-1+i)}$$

In extracting the real part of this, to get x_p, we again have a choice of rectangular or polar approaches. In the rectangular approach, we expand

$$p(-1+i) = 2(-1+i)^2 + (-1+i) + 1 = -3i$$

so $z_p = ie^{(-1+i)t}/3$, and the real part is

$$x_p = -(1/3)e^{-t}\sin(\omega t)\,.$$

In the polar approach, we write

$$\frac{1}{p(-1+i)} = ge^{-i\phi}$$

so that

$$z_p = ge^{-i\phi}e^{(-1+i)t} = ge^{-t}e^{i(t-\phi)}$$

and the real part is

$$x_p = ge^{-t}\cos(t-\phi)$$

I leave it to you to check that you get the same answer!

10.5. **Time invariance.** The fact that the coefficients of $L = p(D)$ are constant leads to an important and useful relationship bewteen solutions to $Lx = f(t)$ for various input signals $f(t)$.

Translation invariance. If L is an LTI operator, and $Lx = f(t)$, then $Ly = f(t-c)$ where $y(t) = x(t-c)$.

This is the "time invariance" of L. Here is an example of its use.

Example 10.5.1. Let's solve

$$2\ddot{x} + \dot{x} + x = 3\sin(t/2 - \pi/3) \tag{11}$$

There are many ways to deal with the phase shift in the signal. Here is one: We saw that when the input signal was $2\cos(t/2)$, the sinusoidal system response was characterized by a gain of $\sqrt{2}$ and a phase lag of $\pi/4$. By time invariance, the same is true for any sinusoidal input with the same frequency. This is the really useful way of expressing the sinusoidal solution, but we can also write it out:

$$x_p = 3\sqrt{2}\sin(t/2 - \pi/3 - \pi/4) = 3\sqrt{2}\sin(t/2 - 7\pi/12)$$

11. Undetermined coefficients

In this section we describe now to solve the constant coefficient linear ODE $p(D)x = q(t)$ in case $q(t)$ is polynomial rather than exponential. Any function can be approximated in a suitable sense by polynomial functions, and this makes polynomials a flexible and important tool.

A *polynomial* is a function of the form

$$q(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_0.$$

The smallest k for which $a_k \neq 0$ is the *degree* of $q(t)$. (The zero function is a polynomial too, but it doesn't have a degree.)

Note that $q(0) = a_0$ and $q'(0) = a_1$.

Here is the basic fact about the response of an LTI system with characteristic polynomial $p(s)$ to polynomial signals:

Theorem. (Undetermined coefficients) If $p(0) \neq 0$, and $q(t)$ is a polynomial of degree n, then

$$p(D)x = q(t)$$

has exactly one solution which is polynomial, and it is of degree n.

The best way to see this, and to see how to compute this polynomial particular solution, is by an example. Suppose we have

$$\ddot{x} + 2\dot{x} + 3x = 4t^2 + 5.$$

The theorem asserts that there is exactly one solution of the form

$$x = at^2 + bt + c,$$

where a, b, c are constants. To find them, just substitute this expression for x into the equation. It's helpful to be systematic in making this computation. Write out $x, \dot{x}$, and $\ddot{x}$, and then multiply by the coefficients, taking care to line up powers of t:

$$\begin{array}{rcccccc}
3x & = & 3at^2 & + & 3bt & + & 3c \\
2\dot{x} & = & & & 4at & + & 2b \\
\ddot{x} & = & & & & & 2a \\
\hline
4t^2 + 5 & = & 3at^2 & + & (4a + 3b)t & + & (2a + 2b + 3c)
\end{array}$$

Now we equate coefficients of corresponding powers of t. It's easiest to start with the highest power of t:

$$4 = 3a \qquad \text{so} \qquad a = 4/3,$$

$$3b = -4a = -16/3 \qquad \text{so} \qquad b = -16/9\,,$$

$$3c = 5 - 2(4/3) - 2(-16/9) \qquad \text{so} \qquad c = 53/27\,.$$

Our solution is thus

$$x = (4/3)t^2 - (16/9)t + (53/27)\,.$$

The computation in this example is about as complicated as it could get. I planned it that way so you would see the point: since it's an "upper triangular" set of linear equations, you can always solve for the coefficients one after the other.

If the constant term in the characteristic polynomial had been zero, though, there would have been trouble: there would have been nothing on the right and side of the table to give t^2. This is why the hypothesis in the theorem is needed.

There is a simple dodge we can use in case $p(0) = 0$, though. If $p(0) = 0$, then $p(D)x$ doesn't involve x itself; it involves only $\dot{x}$ and *its* derivatives. So we can regard it as an ODE (of one order less) for $\dot{x}$, solve that, and then integrate to solve for x. It may be useful to write $y = \dot{x}$ in this process, to keep your head straight.

Suppose we have

$$\ddot{x} + 2\dot{x} = 3t^2\,,$$

for example. If we write $y = \dot{x}$, this is the same as

$$\dot{y} + 2y = 3t^2\,.$$

The theorem applies to this equation, now: there is exactly one solution of the form $y = at^2 + bt + c$.

Exercise. Carry out the computation to show that this polynomial solution is $y = (3/2)t^2 - (3/2)t + 3/4$.

Now a solution to the original problem is given by an integral of y: $x = (1/2)t^3 - (3/4)t^2 + (3/4)t$. You still get a polynomial solution, but it is no longer the only polynomial solution—I can add any constant to it and get another—and its degree is larger than the degree of the input function.

These methods let you find a polynomial response to a polynomial signal for any LTI system.

Final remark: there is overlap with the case of exponential signal, since the constant function with value 1 is an exponential: $e^{0t} = 1$. Our earlier method gives the solution $e^{0t}/p(0)$ for a solution to $p(D)x = 1$, provided $p(0) \neq 0$. This the same as the solution given by the method of undetermined coefficients.

12. Resonance and the exponential shift law

12.1. **Exponential shift.** The calculation (10.1)

$$p(D)e^{rt} = p(r)e^{rt} \tag{1}$$

extends to a formula for the effect of the operator $p(D)$ on a product of the form $e^{rt}u$, where u is a general function. This is useful in solving $p(D)x = f(t)$ when the input signal is of the form $f(t) = e^{rt}q(t)$.

The formula arises from the product rule for differentiation, which can be written in terms of operators as

$$D(vu) = v\,Du + (Dv)u.$$

If we take $v = e^{rt}$ this becomes

$$D(e^{rt}u) = e^{rt}Du + re^{rt}u = e^{rt}(Du + ru)\,.$$

Using the notation I for the identity operator, we can write this as

$$D(e^{rt}u) = e^{rt}(D + rI)u. \tag{2}$$

If we apply D to this equation again,

$$D^2(e^{rt}u) = D(e^{rt}(D + rI)u) = e^{rt}(D + rI)^2u\,,$$

where in the second step we have applied (2) with u replaced by $(D + rI)u$. This generalizes to

$$D^k(e^{rt}u) = e^{rt}(D + rI)^k u.$$

The final step is to take a linear combination of D^k's, to form a general LTI operator $p(D)$. The result is the

Exponential Shift Law:

$$\boxed{p(D)(e^{rt}u) = e^{rt}p(D + rI)u} \tag{3}$$

The effect is that we have pulled the exponential outside the differential operator, at the expense of changing the operator in a specified way.

12.2. **Product signals.** We can exploit this effect to solve equations of the form

$$p(D)x = e^{rt}q(t)\,,$$

by a version of the method of variation of parameter: write $x = e^{rt}u$, apply $p(D)$, use (3) to pull the exponential out to the left of the operator, and then cancel the exponential from both sides. The result is

$$p(D + rI)u = q(t)\,,$$

a new LTI ODE for the function u, one from which the exponential factor has been eliminated.

Example 12.2.1. Find a particular solution to $\ddot{x} + \dot{x} + x = t^2 e^{3t}$.

With $p(s) = s^2 + s + 1$ and $x = e^{3t}u$, we have

$$\ddot{x} + \dot{x} + x = p(D)x = p(D)(e^{3t}u) = e^{3t}p(D + 3I)u\,.$$

Set this equal to t^2e^{3t} and cancel the exponential, to find

$$p(D + 3I)u = t^2$$

or $\dot{u} + 3u = t^2$. This is a good target for the method of undetermined coefficients (Section 11). The first step is to compute

$$p(s + 3) = (s + 3)^2 + (s + 3) + 1 = s^2 + 7s + 13\,,$$

so we have $\ddot{u} + 7\dot{u} + 13u = t^2$. There is a solution of the form $u_p = at^2 + bt + c$, and we find it is

$$u_p = (1/13)t^2 - (14/13^2)t + (85/13^3)\,.$$

Thus a particular solution for the original problem is

$$x_p = e^{3t}((1/13)t^2 - (14/13^2)t + (85/13^3))\,.$$

Example 12.2.2. Find a particular solution to $\dot{x} + x = te^{-t}\sin t$.

The signal is the imaginary part of $te^{(-1+i)t}$, so, following the method of Section 10, we consider the ODE

$$\dot{z} + z = te^{(-1+i)t}\,.$$

If we can find a solution z_p for this, then $x_p = \operatorname{Im} z_p$ will be a solution to the original problem.

We will look for z of the form $e^{(-1+i)t}u$. The Exponential Shift Law (3) with $p(s) = s + 1$ gives

$$\begin{aligned}\dot{z} + z = (D + I)(e^{(-1+i)t}u) &= e^{(-1+i)t}((D + (-1 + i)I) + I)u \\ &= e^{(-1+i)t}(D + iI)u.\end{aligned}$$

When we set this equal to the right hand side we can cancel the exponential:

$$(D + iI)u = t$$

or $\dot{u} + iu = t$. While this is now an ODE with *complex* coefficients, it's easy to solve by the method of undetermined coefficients: there is a solution of the form $u_p = at + b$. Computing the coefficients, $u_p = -it + 1$; so $z_p = e^{(-1+i)t}(-it + 1)$.

Finally, extract the imaginary part to obtain x_p:

$$z_p = e^{-t}(\cos t + i\sin t)(-it + 1)$$

has imaginary part

$$x_p = e^{-t}(-t\cos t + \sin t).$$

12.3. **Resonance.** We have noted that the Exponential Response Formula for a solution to $p(D)x = e^{rt}$ fails when $p(r) = 0$. For example, For example, suppose we have $\dot{x} + x = e^{-t}$. The Exponential Response Formula proposes a solution $x_p = e^{-t}/p(-1)$, but $p(-1) = 0$ so this fails. There is no solution of the form ce^{rt}.

This situation is called *resonance*, because the signal is tuned to a natural mode of the system.

Here is a way to solve $p(D)x = e^{rt}$ when this happens. The ERF came from the calculation

$$p(D)e^{rt} = p(r)e^{rt},$$

which is valid whether or not $p(r) = 0$. We will take this expression and *differentiate it with respect to* r, keeping t constant. The result, using the product rule and the fact that partial derivatives commute, is

$$p(D)te^{rt} = p'(r)e^{rt} + p(r)te^{rt}$$

If $p(r) = 0$ this simplifies to

$$p(D)te^{rt} = p'(r)e^{rt}\,. \tag{4}$$

Now if $p'(r) \neq 0$ we can divide through by it and see:

The Resonant Exponential Response Formula: If $p(r) = 0$ then a solution to $p(D)x = ae^{rt}$ is given by

$$\boxed{x_p = a\,\frac{te^{rt}}{p'(r)}} \tag{5}$$

provided that $p'(r) \neq 0$.

In our example above, $p(s) = s + 1$ and $r = 1$, so $p'(r) = 1$ and $x_p = te^{-t}$ is a solution.

This example exhibits a characteristic feature of resonance: the solutions grow faster than you might expect. The characteristic polynomial leads you to expect a solution of the order of e^{-t}. In fact the solution is t times this. It still decays to zero as t grows, but not as fast as e^{-t} does.

Example 12.3.1. Suppose we have a harmonic oscillator represented by $\ddot{x} + \omega_n^2 x$, or by the operator $D^2 + \omega_n^2 I = p(D)$, and drive it by the

signal $a\cos(\omega t)$. This ODE is the real part of

$$\ddot{z} + \omega_n^2 z = ae^{i\omega t}\,,$$

so the Exponential Response Formula gives us the periodic solution

$$z_p = a\frac{e^{i\omega_n t}}{p(i\omega)}\,.$$

This is fine *unless* $\omega = \omega_n$, in which case $p(i\omega_n) = (i\omega_n)^2 + \omega_n^2 = 0$; so the amplitude of the proposed sinusoidal response should be infinite. The fact is that there is *no* periodic system response; the system is in *resonance* with the signal.

To circumvent this problem, let's apply the Resonance Exponential Response Formula: since $p(s) = s^2 + \omega_n^2$, $p'(s) = 2s$ and $p'(i\omega_n) = 2i\omega_0$, so

$$z_p = a\frac{te^{i\omega_n t}}{2i\omega_n}\,.$$

The real part is

$$x_p = \frac{a}{2\omega_n} t\sin(\omega_n t)\,.$$

The general solution is thus

$$x = \frac{a}{2\omega_n} t\sin(\omega_n t) + b\cos(\omega_n t - \phi)\,.$$

In words, all solutions oscillate with pseudoperiod $2\pi/\omega_n$, and grow in amplitude like $at/(2\omega_n)$. When ω_n is large—high frequency—this rate of growth is small.

12.4. **Higher order resonance.** It may happen that both $p(r) = 0$ and $p'(r) = 0$. The general picture is this: Suppose that k is such that $p^{(j)}(r) = 0$ for $j < k$ and $p^{(k)}(r) \neq 0$. Then $p(D)x = ae^{rt}$ has as solution

$$x_p = a\frac{t^k e^{rt}}{p^{(k)}(r)}\,. \tag{6}$$

For instance, if $\omega = \omega_0 = 0$ in Example 12.3.1, $p'(i\omega) = 0$. The signal is now just the constant function a, and the ODE is $\ddot{x} = a$. Integrating twice gives $x_p = at^2/2$ as a solution, which is a special case of (6), since $e^{rt} = 1$ and $p''(s) = 2$.

You can see (6) in the same way we saw the Resonant Exponential Response Formula. So take (4) and differentiate again with respect to r:

$$p(D)t^2e^{rt} = p''(r)e^{rt} + p'(r)te^{rt}$$

If $p'(r) = 0$, the second term drops out and if we suppose $p''(r) \neq 0$ and divide through by it we get

$$p(D)\left(\frac{t^2 e^{rt}}{p'(r)}\right) = e^{rt}$$

which the case $k = 2$ of (6). Continuing, we get to higher values of k as well.

12.5. **Summary.** The work of this section and the last can be summarized as follows: Among the responses by an LTI system to a signal which is polynomial times exponential (or a linear combination of such) there is always one which is again a linear combination of functions which are polynomial times exponential. By the magic of the complex exponential, sinusoidal factors are included in this.

13. Natural frequency and damping ratio

There is a standard, and useful, normalization of the second order homogeneous linear constant coefficient ODE

$$m\ddot{x} + b\dot{x} + kx = 0$$

under the assumption that both the "mass" m and the "spring constant" k are positive. It is illustrated in the Mathlet `Damping Ratio`.

In the absence of a damping term, the ratio k/m would be the square of the circular frequency of a solution, so we will write $k/m = \omega_n^2$ with $\omega_n > 0$, and call ω_n the **natural circular frequency** of the system.

Divide the equation through by m: $\ddot{x} + (b/m)\dot{x} + \omega_n^2 x = 0$. Critical damping occurs when the coefficient of $\dot{x}$ is $2\omega_n$. The **damping ratio** ζ is the ratio of b/m to the critical damping constant: $\zeta = (b/m)/(2\omega_n)$. The ODE then has the form

$$\boxed{\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = 0} \tag{1}$$

Note that if x has dimensions of cm and t of sec, then ω_n had dimensions sec^{-1}, and the damping ratio ζ is "dimensionless," a number which is the same no matter what units of distance or time are chosen. Critical damping occurs precisely when $\zeta = 1$: then the characteristic polynomial has a repeated root: $p(s) = (s + \omega_n)^2$.

In general the characteristic polynomial is $s^2 + 2\zeta\omega_n s + \omega_n^2$, and it has as roots

$$-\zeta\omega_n \pm \sqrt{\zeta^2\omega_n^2 - \omega_n^2} = \omega_n(-\zeta \pm \sqrt{\zeta^2 - 1}).$$

These are real when $|\zeta| \geq 1$, equal when $\zeta = \pm 1$, and nonreal when $|\zeta| < 1$. When $|\zeta| \leq 1$, the roots are

$$-\zeta\omega_n \pm i\omega_d$$

where

$$\omega_d = \sqrt{1 - \zeta^2}\,\omega_n \tag{2}$$

is the **damped circular frequency** of the system. These are complex numbers of magnitude ω_n and argument $\pm\theta$, where $-\zeta = \cos\theta$. Note that the presence of a damping term decreases the frequency of a solution to the undamped equation—the natural frequency ω_n—by the factor $\sqrt{1 - \zeta^2}$. The general solution is

$$x = Ae^{-\zeta\omega_n t}\cos(\omega_d t - \phi) \tag{3}$$

Suppose we have such a system, but don't know the values of ω_n or ζ. At least when the system is underdamped, we can discover them by a couple of simple measurements of the system response. Let's displace the mass and watch it vibrate freely. If the mass oscillates, we are in the underdamped case. We can find ω_d by measuring the times at which x achieves its maxima. These occur when the derivative vanishes, and

$$\dot{x} = Ae^{-\zeta\omega_n t}\left(-\zeta\omega_n\cos(\omega_d t - \phi) - \omega_d\sin(\omega_d t - \phi)\right).$$

The factor in parentheses is sinusoidal with circular frequency ω_d, so successive zeros are separated from each other by a time lapse of π/ω_d. If t_1 and t_2 are the times of neighboring maxima of x (which occur at every other extremum) then $t_2 - t_1 = 2\pi/\omega_d$, so we have discovered the damped natural frequency:

$$\omega_d = \frac{2\pi}{t_2 - t_1}. \tag{4}$$

We can also measure the ratio of the value of x at two successive maxima. Write $x_1 = x(t_1)$ and $x_2 = x(t_2)$. The difference of their natural logarithms is the **logarithmic decrement**:

$$\Delta = \ln x_1 - \ln x_2 = \ln\left(\frac{x_1}{x_2}\right).$$

Then

$$x_2 = e^{-\Delta}x_1.$$

The logarithmic decrement turns out to depend only on the damping ratio, and to determine the damping ratio. To see this, note that the values of $\cos(\omega_d t - \phi)$ at two points of time differing by $2\pi/\omega_d$ are equal. Using (3) we find

$$\frac{x_1}{x_2} = \frac{e^{-\zeta\omega_n t_1}}{e^{-\zeta\omega_n t_2}} = e^{\zeta\omega_n(t_2 - t_1)}.$$

Thus, using (4) and (2),

$$\Delta = \ln\left(\frac{x_1}{x_2}\right) = \zeta\omega_n(t_2 - t_1) = \zeta\omega_n\frac{2\pi}{\omega_d} = \frac{2\pi\zeta}{\sqrt{1-\zeta^2}}.$$

From the quantities ω_d and Δ, which are directly measurable characteristics of the unforced system response, we can calculate the system parameters ω_n and ζ:

$$\zeta = \frac{\Delta/2\pi}{\sqrt{1 + (\Delta/2\pi)^2}}, \qquad \omega_n = \frac{\omega_d}{\sqrt{1-\zeta^2}} = \sqrt{1 + \left(\frac{\Delta}{2\pi}\right)^2}\,\omega_d\,. \tag{5}$$

14. Frequency response

In Section 3 we discussed the frequency response of a first order LTI operator. In Section 10 we used the Exponential Response Formula to understand the response of an LTI operator to a sinusoidal input signal. Here we will study this in more detail in case the operator is of second order, and understand how the gain and phase lag vary with the driving frequency.

A differential equation relates input signal to system response. What constitutes the "input signal" and what constitutes the "system response" are matters of convenience for the user, and are not determined by the differential equation. We will illustrate this in a couple of examples. The case of sinusoidal input signal and system response are particularly important. The question is then: what is the ratio of amplitude of the system response to that of the input signal, and what is the phase lag of the system response relative to the input signal?

We will carry this analysis out three times: first for two specific examples of mechanical (or electrical) systems, and then in general using the notation of the damping ratio.

14.1. **Driving through the spring.** The Mathlet `Amplitude and Phase: Second order I` illustrates a spring/mass/dashpot system that is driven through the spring. Suppose that y denotes the displacement of the plunger at the top of the spring, and $x(t)$ denotes the position of the mass, arranged so that $x = y$ when the spring is unstretched and uncompressed. There are two forces acting on the mass: the spring exerts a force force given by $k(y - x)$ (where k is the spring constant), and the dashpot exerts a force given by $-b\dot{x}$ (against the motion of the mass, with damping coefficient b). Newton's law gives

$$m\ddot{x} = k(y - x) - b\dot{x}$$

or, putting the system on the left and the driving term on the right,

$$m\ddot{x} + b\dot{x} + kx = ky\,. \tag{1}$$

In this example it is natural to regard y, rather than ky, as the input signal, and the mass position x as the system response.

Another system leading to the same equation is a series RLC circuit, discussed in Section 8 and illustrated in the Mathlet `Series RLC Circuit`. We consider the impressed voltage as the input signal, and the voltage drop across the capacitor as the system response. The

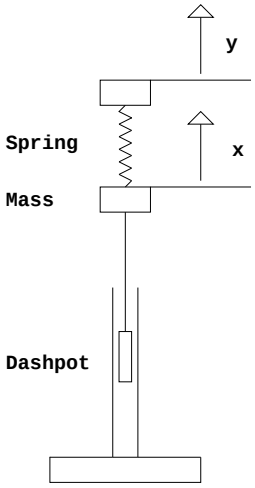


FIGURE 6. Spring-driven system

equation is then

$$L\ddot{V}_C + R\dot{V}_C + (1/C)V_C = (1/C)V$$

We will favor the mechanical system notation, but the mathematics is exactly the same in both systems.

When y is sinusoidal, say

$$y = A\cos(\omega t)\,,$$

then (putting aside the possibility of resonance) we expect a sinusoidal solution, one of the form

$$x = B\cos(\omega t - \phi)$$

The ratio of the amplitude of the system response to that of the input signal, B/A, is called the **gain** of the system. We think of the system as fixed, while the frequency ω of the input signal can be varied, so the gain is a function of ω, $g(\omega)$. Similarly, the **phase lag** ϕ is a function of ω. The entire story of the steady state system response to sinusoidal input signals is encoded in those to functions of ω, the gain and the phase lag.

There is a systematic way to work out what g and ϕ are. The equation (1) is the real part of a complex-valued differential equation:

$$m\ddot{z} + b\dot{z} + kz = Ake^{st}$$

with $s = i\omega$. The Exponential Response Formula gives the solution

$$z_p = \frac{Ak}{p(s)}\,e^{st}$$

where

$$p(s) = ms^2 + bs + k$$

(as long as $p(s) \neq 0$).

Our choice of input signal and system response correspond in the complex equation to regarding Ae^{st} as the input signal and z_p as the exponential system response. The **transfer function** is the ratio between the two:

$$W(s) = \frac{k}{p(s)}$$

so

$$z_p = W(s)Ae^{st}\,.$$

Now take $s = i\omega$. The **complex gain** is

$$W(i\omega) = \frac{k}{k - m\omega^2 + ib\omega}\,. \tag{2}$$

I claim that the polar form of the complex gain determines the gain g and the phase lag ϕ as follows:

$$W(i\omega) = ge^{-i\phi}$$

To verify this, substitute this expression into the formula for z_p—

$$z_p = g\,e^{-i\phi}Ae^{i\omega t} = gAe^{i(\omega t - \phi)}$$

—and extract the real part, to get the sinusoidal solution to (1):

$$y_p = gA\cos(\omega t - \phi)\,.$$

The amplitude of the input signal, A, has been multiplied by the gain

$$g(\omega) = |W(i\omega)| = \frac{k}{\sqrt{k^2 + (b^2 - 2mk)\omega^2 + m^2\omega^4}} \tag{3}$$

The phase lag of the system response, relative to the input signal, is $\phi = -\mathrm{Arg}(W(i\omega))$. Since $\mathrm{Arg}(1/z) = -\mathrm{Arg}(z)$, ϕ is the argument of the denominator in (2). The tangent of the argument of a complex number is the ratio of the imaginary part by the real part, so

$$\tan\phi = \frac{b\omega}{k - m\omega^2}$$

The `Amplitude and Phase: Second order I` Mathlet shows how the gain varies with ω. Often there is a choice of frequency ω for which the gain is maximal: this is "near resonance." To compute what this frequency is, we can try to *minimize* the denominator in (3). That

minimum occurs when $k^2 + (b^2 - 2mk)\omega^2 + m^2\omega^4$ is minimized, which occurs when ω is either 0 or the *resonant frequency*

$$\omega_r = \sqrt{\frac{k}{m} - \frac{b^2}{2m^2}}$$

When $b = 0$, this is the natural frequency $\omega_n = \sqrt{k/m}$ and we have true resonance; the gain becomes infinite. As we increase b, the resonant frequency decreases, till when $b = \sqrt{2mk}$ we find $\omega_r = 0$. For b less than this, practical resonance occurs only for $\omega = 0$.

14.2. **Driving through the dashpot.** Now suppose instead that we fix the top of the spring and drive the system by moving the bottom of the dashpot instead. This is illustrated in `Amplitude and Phase: Second order II`.

Suppose that the position of the bottom of the dashpot is given by $y(t)$, and again the mass is at $x(t)$, arranged so that $x = 0$ when the spring is relaxed. Then the force on the mass is given by

$$m\ddot{x} = -kx + b\,\frac{d}{dt}(y - x)$$

since the force exerted by a dashpot is supposed to be proportional to the speed of the piston moving through it. This can be rewritten

$$m\ddot{x} + b\dot{x} + kx = b\dot{y}\,. \tag{4}$$

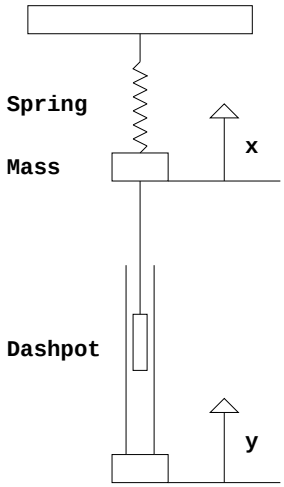


FIGURE 7. Dashpot-driven system

Again we will consider x as the system response, and the position of the back end of the dashpot, y, as the input signal. Note that the

derivative of the input signal (multiplied by b) occurs on the right hand side of the equation. Another system leading to the same mathematics is the series RLC circuit shown in the Mathlet `Series RLC Circuit`, in which the impressed voltage is the input variable and the voltage drop across the resistor is the system response. The equation is

$$L\ddot{V}_R + R\dot{V}_R + (1/C)V_R = R\dot{V}$$

Here's a frequency response analysis of this problem. We suppose that the input signal is sinusoidal:

$$y = B\cos(\omega t)\,.$$

Then $\dot{y} = -\omega B\sin(\omega t)$ so our equation is

$$m\ddot{x} + b\dot{x} + kx = -b\omega B\sin(\omega t)\,. \tag{5}$$

The periodic system response will be of the form

$$x_p = gB\cos(\omega t - \phi)$$

for some gain g and phase lag ϕ, which we now determine by making a complex replacement.

The right hand side of (5) involves the sine function, so one natural choice would be to regard it as the imaginary part of a complex equation. This would work, but we should also keep in mind that the input signal is $B\cos(\omega t)$. For that reason, we will write (5) as the *real* part of a complex equation, using the identity $\mathrm{Re}\,(ie^{i\omega t}) = -\sin(\omega t)$. The equation (5) is thus the real part of

$$m\ddot{z} + b\dot{z} + kz = bi\omega Be^{i\omega t}\,. \tag{6}$$

and the complex input signal is $Be^{i\omega t}$ (since this has real part $B\cos(\omega t)$).

The sinusoidal system response x_p of (5) is the real part of the exponential system response z_p of (6). The Exponential Response Formula gives

$$z_p = \frac{bi\omega}{p(i\omega)}Be^{i\omega t}$$

where

$$p(s) = ms^2 + bs + k$$

is the characteristic polynomial.

The complex gain is the complex number $W(i\omega)$ by which you have to multiply the complex input signal to get the exponential system response. Comparing z_p with $Be^{i\omega t}$, we see that

$$W(i\omega) = \frac{bi\omega}{p(i\omega)}\,.$$

As usual, write

$$W(i\omega) = ge^{-i\phi}$$

so that

$$z_p = W(i\omega)Be^{i\omega t} = gBe^{i(\omega t-\phi)}$$

Thus

$$x_p = \mathrm{Re}\,(z_p) = gB\cos(\omega t - \phi)$$

—the amplitude of the sinusoidal system response is g times that of the input signal, and lags behind the input signal by ϕ radians.

To make this more explicit, let's use the natural frequency $\omega_n = \sqrt{k/m}$. Then

$$p(i\omega) = m(i\omega)^2 + bi\omega + m\omega_n^2 = m(\omega_n^2 - \omega^2) + bi\omega\,,$$

so

$$W(i\omega) = \frac{bi\omega}{m(\omega_n^2 - \omega^2) + bi\omega}\,.$$

Thus the gain $g(\omega) = |W(i\omega)|$ and the phase lag $\phi = -\mathrm{Arg}(W(i\omega))$ are determined as the polar coordinates of the complex function of ω given by $W(i\omega)$. As ω varies, $W(i\omega)$ traces out a curve in the complex plane, shown by invoking the `[Nyquist plot]` in the applet. To understand this curve, divide numerator and denominator in the expression for $W(i\omega)$ by $bi\omega$, and rearrange:

$$W(i\omega) = \left(1 - \frac{i}{b/m}\,\frac{\omega_n^2 - \omega^2}{\omega}\right)^{-1}\,.$$

As ω goes from 0 to ∞, $(\omega_n^2 - \omega^2)/\omega$ goes from $+\infty$ to $-\infty$, so the expression inside the brackets follows the vertical straight line in the complex plane with real part 1, moving upwards. As z follows this line, $1/z$ follows a circle of radius $1/2$ and center $1/2$, traversed clockwise (exercise!). It crosses the real axis when $\omega = \omega_n$.

This circle is the "Nyquist plot." It shows that the gain starts small, grows to a maximum value of 1 exactly when $\omega = \omega_n$ (in contrast to the spring-driven situation, where the resonant peak is not exactly at ω_n and can be either very large or non-existent depending on the strength of the damping), and then falls back to zero. Near resonance occurs at $\omega_r = \omega_n$.

The Nyquist plot also shows that $-\phi = \mathrm{Arg}(W(i\omega))$ moves from near $\pi/2$ when ω is small, through 0 when $\omega = \omega_n$, to near $-\pi/2$ when ω is large.

And it shows that these two effects are linked to each other. Thus a narrow resonant peak corresponds to a rapid sweep across the far edge

of the circle, which in turn corresponds to an abrupt phase transition from $-\phi$ near $\pi/2$ to $-\phi$ near $-\pi/2$.

14.3. **Second order frequency response using damping ratio.** As explained in Section 13, it is useful to write a second order system with sinusoidal driving term as

$$\ddot{x} + 2\zeta\omega_n\dot{x} + \omega_n^2 x = a\cos(\omega t)\,. \tag{7}$$

The constant ω_n is the "natural frequency" of the system and ζ is the "damping ratio." In this abstract situation, we regard the full right hand side, $a\cos(\omega t)$, as the input signal, and x as the system response.

The best path to the solution of (7) is to view it as the real part of the complex equation

$$\ddot{z} + 2\zeta\omega_n\dot{z} + \omega_n^2 z = ae^{i\omega t}. \tag{8}$$

The Exponential Response Formula of Section 10 tells us that unless $\zeta = 0$ and $\omega = \omega_n$ (in which case the equation exhibits resonance, and has no periodic solutions), this has the particular solution

$$z_p = a\,\frac{e^{i\omega t}}{p(i\omega)} \tag{9}$$

where $p(s) = s^2 + 2\zeta\omega_n s + \omega_n^2$ is the characteristic polynomial of the system. In Section 10 we wrote $W(s) = 1/p(s)$, so this solution can be written

$$z_p = aW(i\omega)e^{i\omega t}\,.$$

The complex valued function of ω given by $W(i\omega)$ is the **complex gain**. We will see now how, for fixed ω, this function contains exactly what is needed to write down a sinusoidal solution to (7).

As in Section 10 we can go directly to the expression in terms of amplitude and phase lag for the particular solution to (7) given by the real part of z_p as follows. Write the polar expression (as in Section 6) for the complex gain as

$$W(i\omega) = \frac{1}{p(i\omega)} = ge^{-i\phi}\,. \tag{10}$$

so that

$$g(\omega) = |W(i\omega)|\,, \quad \phi(\omega) = \mathrm{Arg}(W(i\omega))$$

Then

$$z_p = ag\,e^{i(\omega t-\phi)}, \qquad x_p = ag\cos(\omega t - \phi),$$

The particular solution x_p is the *only* periodic solution to (7), and, assuming $\zeta > 0$, any other solution differs from it by a transient. This solution is therefore the most important one; it is the "steady state"

solution. It is sinusoidal, and hence determined by just a few parameters: its circular frequency, which is the circular frequency of the input signal; its amplitude, which is g times the amplitude a of the input signal; and its phase lag ϕ relative to the input signal.

We want to understand how g and ϕ depend upon the driving frequency ω. The gain is given by

$$g(\omega) = \frac{1}{|p(i\omega)|} = \frac{1}{\sqrt{(\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2}}. \tag{11}$$

Figure 8 shows the graphs of gain against the circular frequency of the signal for $\omega_n = 1$ and several values of the damping ratio ζ (namely $\zeta = 1/(4\sqrt{2}), 1/4, 1/(2\sqrt{2}), 1/2, 1/\sqrt{2}, 1, \sqrt{2}, 2.$) As you can see, the gain may achieve a maximum. This occurs when the square of the denominator in (11) is minimal, and we can discover where this is by differentiating with respect to ω and setting the result equal to zero:

$$\frac{d}{d\omega}\left((\omega_n^2 - \omega^2)^2 + 4\zeta^2\omega_n^2\omega^2\right) = -2(\omega_n^2 - \omega^2)2\omega + 8\zeta^2\omega_n^2\omega, \tag{12}$$

and this becomes zero when ω equals the **resonant frequency**

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2}. \tag{13}$$

When $\zeta = 0$ the gain becomes infinite at $\omega = \omega_n$: this is **true resonance**. As ζ increases from zero, the maximal gain of the system occurs at smaller and smaller frequencies, till when $\zeta = 1/\sqrt{2}$ the maximum occurs at $\omega = 0$. For still larger values of ζ, the only maximum in the gain curve occurs at $\omega = 0$. When ω takes on a value at which the gain is a local maximum we have **practical resonance**.

We also have the phase lag to consider: the periodic solution to (7) is

$$x_p = ga\cos(\omega t - \phi).$$

Returning to (10), ϕ is given by the argument of the complex number

$$p(i\omega) = (\omega_n^2 - \omega^2) + 2i\zeta\omega_n\omega\,.$$

This is the angle counterclockwise from the positive x axis of the ray through the point $(\omega_n^2 - \omega^2, 2\zeta\omega_n\omega)$. Since ζ and ω are nonnegative, this point is always in the upper half plane, and $0 \le \phi \le \pi$. The phase response graphs for $\omega_n = 1$ and several values of ζ are shown in the second figure.

When $\omega = 0$, there is no phase lag, and when ω is small, ϕ is approximately $2\zeta\omega/\omega_n$. $\phi = \pi/2$ when $\omega = \omega_n$, independent of the damping

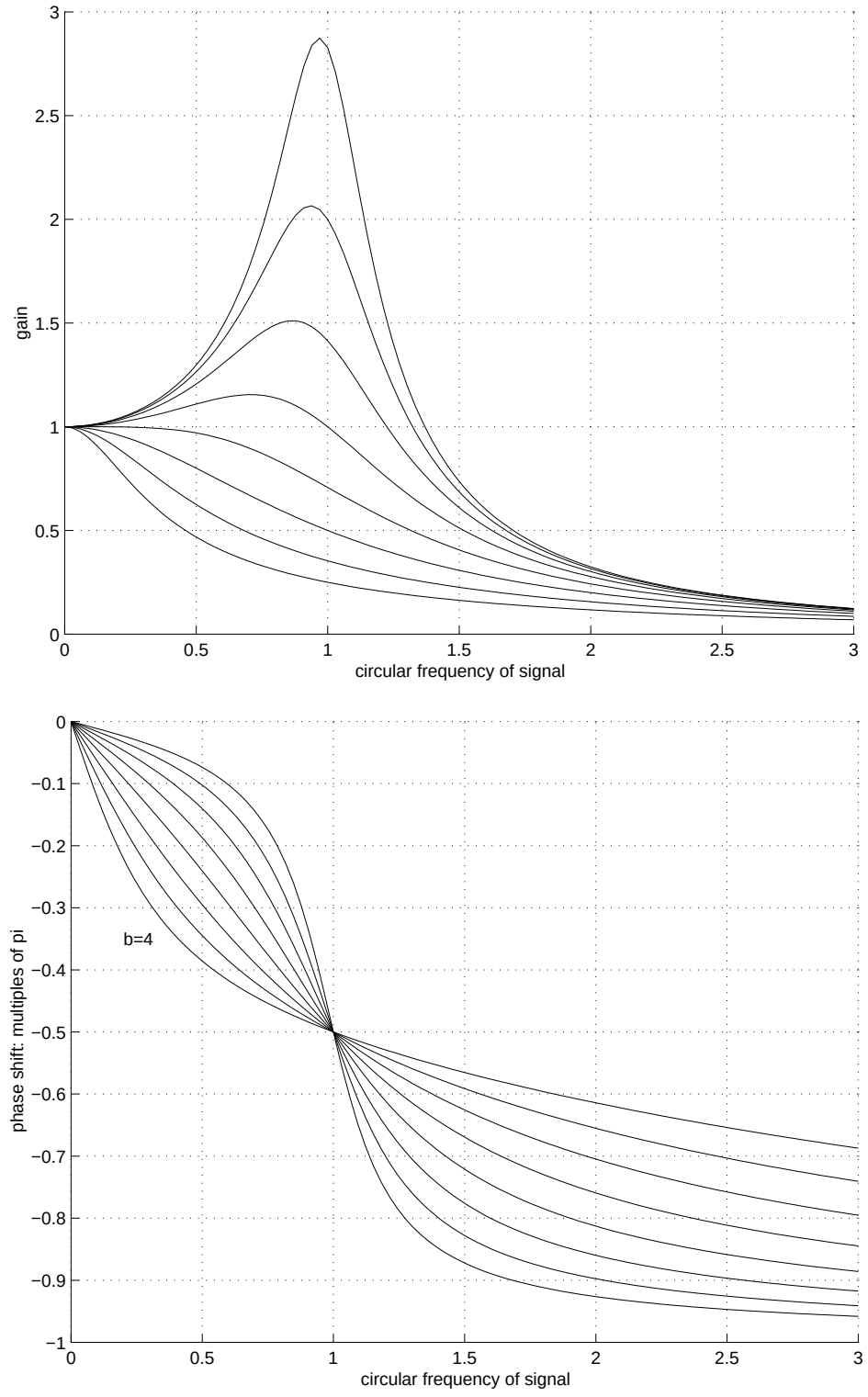


FIGURE 8. Second order amplitude response curves

rato ζ: when the signal is tuned to the natural frequency of the system, the phase lag is $\pi/2$, which is to say that the time lag is one-quarter of a period. As ω gets large, the phase lag tends towards π: strange as it may seem, the sign of the system response tends to be opposite to the sign of the signal.

Engineers typically have to deal with a very wide range of frequencies. In order to accommodate this, and to show the behavior of the frequency response more clearly, they tend to plot $\log_{10}|1/p(i\omega)|$ and the argument of $1/p(i\omega)$ against $\log_{10}\omega$. These are the so-called **Bode plots**.

The expression $1/p(i\omega)$, as a complex-valued function of ω, contains complete information about the system response to periodic input signals. If you let ω run from $-\infty$ to ∞ you get a curve in the complex plane called the **Nyquist plot**. In cases that concern us we may restrict attention to the portion parametrized by $\omega > 0$. For one thing, the characteristic polynomial $p(s)$ has real coefficients, which means that $p(-i\omega) = p(\overline{i\omega}) = \overline{p(i\omega)}$ and so $1/p(-i\omega)$ is the complex conjugate of $1/p(i\omega)$. The curve parametrized by $\omega < 0$ is thus the reflection of the curve parametrized by $\omega > 0$ across the real axis.

15. The Wronskian

We know that a general second order homogeneous linear ODE,

$$y'' + p(x)y' + q(x)y = 0, \tag{1}$$

has a pair of independent solutions; and that if y_1, y_2 is any pair of independent solutions then the general solution is

$$y = c_1 y_1 + c_2 y_2. \tag{2}$$

Suppose we wish to solve the initial value problem with

$$y(x_0) = a\,, \quad y'(x_0) = b.$$

To solve for the constants c_1 and c_2 in (2), we get one equation by substituting $x = x_0$ into this expression. We get another equation by first differentiating (2) and then setting $x = x_0$ and using the value $y'(x_0) = b$. We end up with the system of linear equations

$$y_1(x_0)c_1 + y_2(x_0)c_2 = a\,, \quad y_1'(x_0)c_1 + y_2'(x_0)c_2 = b \tag{3}$$

When you solve for the coefficients of these equations, you get ("Cramer's rule")

$$c_1 = \frac{y_2'(x_0)a - y_2(x_0)b}{W(x_0)}\,, \quad c_2 = \frac{-y_1'(x_0)a + y_1(x_0)b}{W(x_0)}$$

where $W(x_0)$ is the value at x_0 of the **Wronskian** function

$$W(x) = y_1y_2' - y_2y_1' = \det \begin{bmatrix} y_1 & y_2 \\ y_1' & y_2' \end{bmatrix} \tag{4}$$

determined by the pair of solutions y_1, y_2.

You generally wouldn't want to use these formulas for the coefficients; it's better to compute them directly from (3) in the particular case you are looking at. But this calculation does draw attention to the Wronskian function. We can find a linear combination of y_1 and y_2 which solves the IVP for any given choice of initial conditions exactly when $W(x_0) \neq 0$.

On the other hand it's a theorem that one can solve the initial value problem at *any* x value using a linear combination of any linearly independent pair of solutions. A little thought then leads to the following conclusion:

Theorem. Let y_1, y_2 be solutions of (1) and let W be the Wronskian formed from y_1, y_2. Either W is the zero function and one solution is a multiple of the other, or the Wronskian is nowhere zero, neither solution

is a multiple of the other, and any solution is a linear combination of y_1, y_2.

For example, if we compute the Wronskian of the pair of solutions $\{\cos x, \sin x\}$ of $y'' + y = 0$, we get the constant function 1, while the Wronskian of $\{\cos x, 2\cos x\}$ is the constant function 0. One can show (as most ODE textbooks do) that if W is the Wronskian of *some* linearly independent pair of solutions, then the Wronskian of *any* pair of solutions is a constant multiple of W. (That multiple is zero if the new pair happens to be linearly dependent.)

Many references, including Edwards and Penney, encourage the impression that computing the Wronskian of a pair of functions is a good way to check whether or not they are linearly independent. This is silly. Two functions are linearly dependent if one is a multiple of the other; otherwise they are linearly independent. This is always easy to see by inspection.

Nevertheless the Wronskian can teach us important things. To illustrate one, let's consider an example of a second order linear homogeneous system with *nonconstant* coefficient: the **Airy equation**

$$y'' + xy = 0. \tag{5}$$

At least for $x > 0$, this is like the harmonic oscillator $y'' + \omega_n^2 y = 0$, except that the natural circular frequency ω_n keeps increasing with x: the x sits in the position where we expect to see ω_n^2, so near to a given value of x we expect solutions to behave like $\cos(\sqrt{x}x)$ and $\sin(\sqrt{x}x)$. I emphasize that these functions are *not* solutions to (5), but they give us a hint of what to expect. In fact the normalized pair (see Section 9) of solutions to (5), the "Airy cosine and sine functions," have graphs as illustrated in Figure 9

One of the features this picture has in common with the graphs of cosine and sine is the following fact, which we state as a theorem.

Theorem. Let $\{y_1, y_2\}$ be any linearly independent pair of solutions of the second order linear ODE (1), and suppose that x_0 and x_1 are numbers such that $x_0 \neq x_1$ and $y_1(x_0) = 0 = y_1(x_1)$. Then y_2 becomes zero somewhere between x_0 and x_1.

This fact, that zeros of independent solutions interleave, is thus a completely general feature of second order linear equations. It doesn't depend upon the solutions being normalized, and it doesn't depend upon the coefficients being constant.

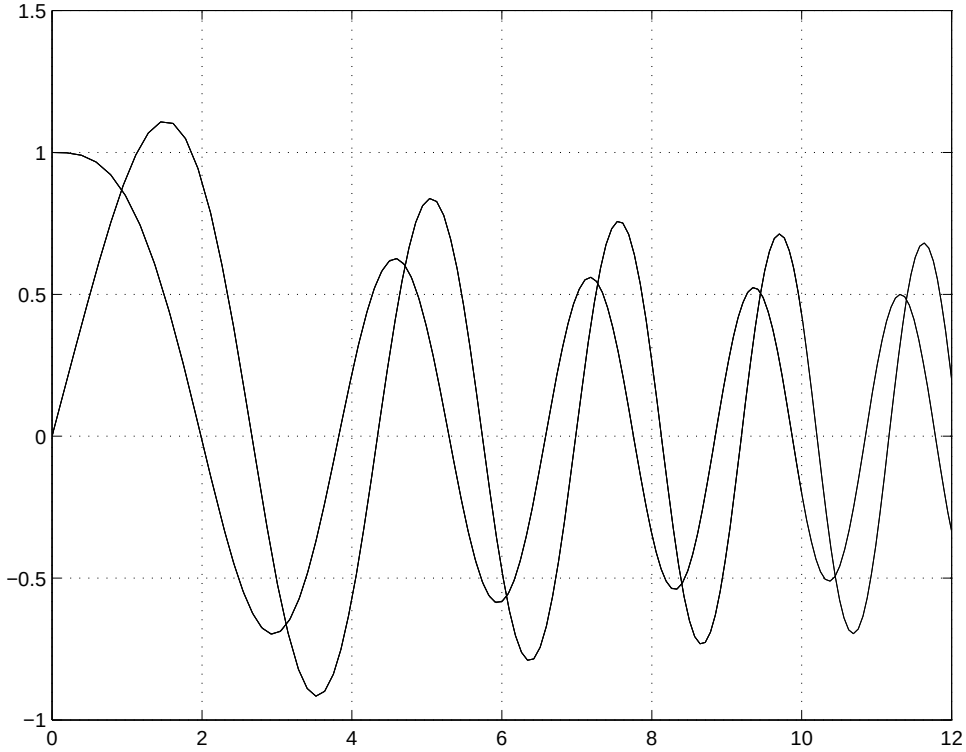


FIGURE 9. Airy cosine and sine

You can see why this must be true using the Wronskian. We might as well assume that y_1 is not zero anywhere between x_0 and x_1. Since the two solutions are independent the associated Wronskian is nowhere zero, and thus has the same sign everywhere. Suppose first that the sign is positive. Then $y_1 y_2' > y_1' y_2$ everywhere. At x_0 this says that $y_1'(x_0)$ and $y_2(x_0)$ have opposite signs, since $y_1(x_0) = 0$. Similarly, $y_1'(x_1)$ and $y_2(x_1)$ have opposite signs. But $y_1'(x_0)$ and $y_1'(x_1)$ must have opposite signs as well, since x_0 and x_1 are neighboring zeros of y_1. (These derivatives can't be zero, since if they were both terms in the definition of the Wronskian would be zero, but $W(x_0)$ and $W(x_1)$ are nonzero.) It follows that $y_2(x_0)$ and $y_2(x_1)$ have opposite signs, and so y_2 must vanish somewhere in between. The argument is very similar if the sign of the Wronskian is negative.

16. More on Fourier series

The Mathlet `Fourier Coefficients` displays many of the effects described in this section.

16.1. **Symmetry and Fourier series.** A function $g(t)$ is **even** if $g(t) = g(-t)$, and **odd** if $g(t) = -g(-t)$.

Fact: *Any* function $f(t)$ is a sum of an even function and an odd function, and this can be done in only one way.

The *even part* of $f(t)$ is

$$f_+(t) = \frac{f(t) + f(-t)}{2}.$$

and the *odd part* is

$$f_-(t) = \frac{f(t) - f(-t)}{2}.$$

It's easy to check that $f_+(t)$ is even, $f_-(t)$ is odd, and that

$$f(t) = f_+(t) + f_-(t).$$

We can apply this to a periodic function. We know that any periodic function $f(t)$, with period 2π, say, has a Fourier expansion of the form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(nt) + b_n \sin(nt)).$$

If $f(t)$ is even then all the b_n's vanish and the Fourier series is simply

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(nt).$$

If $f(t)$ is odd then all the a_n's vanish and the Fourier series is

$$\sum_{n=1}^{\infty} b_n \sin(nt).$$

Most of the time one is faced with a function which is either even or odd. If $f(t)$ is neither even nor odd, we can still compute its Fourier series by computing the Fourier series for $f_+(t)$ and $f_-(t)$ separately and adding the results.

16.2. **Symmetry about other points.** More general symmetries are often present and useful. A function may exhibit symmetry about any fixed value of t, say $t = a$. We say that $f(t)$ is *even about* a if $f(a+t) = f(a-t)$ for all t. It is *odd about* a if $f(a+t) = -f(a-t)$. $f(t)$ is even about a if it behaves the same as you move away from a whether to the left or the right; $f(t)$ is odd about a if its values to the right of a are the negatives of its values to the left. The usual notions of even and odd refer to $a = 0$.

Suppose $f(t)$ is periodic of period 2π, and is even (about 0). $f(t)$ is then entirely determined by its values for t between 0 and π. When we focus attention on this range of values, $f(t)$ may have some *further* symmetry with respect to the midpoint $\pi/2$: it may be even about $\pi/2$ or odd about $\pi/2$, or it may be neither. For example, $\cos(nt)$ is even about $\pi/2$ exactly when n is even, and odd about $\pi/2$ exactly when n is odd. It follows that if $f(t)$ is even and even about $\pi/2$ then its Fourier series involves only *even* cosines:

$$f(t) = \frac{a_0}{2} + \sum_{n \text{ even}} a_n \cos(nt).$$

If $f(t)$ is even about 0 but odd about $\pi/2$ then its Fourier series involves only odd cosines:

$$f(t) = \sum_{n \text{ odd}} a_n \cos(nt).$$

Similarly, the odd function $\sin(nt)$ is even about $\pi/2$ exactly when n is odd, and odd about $\pi/2$ exactly when n is even. Thus if $f(t)$ is odd about 0 but even about $\pi/2$, its Fourier series involves only odd sines:

$$f(t) = \sum_{n \text{ odd}} b_n \sin(nt).$$

If it is odd about both 0 and $\pi/2$, its Fourier series involves only even sines:

$$f(t) = \sum_{n \text{ even}} a_n \sin(nt).$$

16.3. **The Gibbs effect.** The Fourier series for the odd function of period 2π with

$$F(x) = \frac{\pi - x}{2} \quad \text{for } 0 < x < \pi$$

is

$$F(x) = \sum_{k=1}^{\infty} \frac{\sin(kx)}{k}\,.$$

In Figure 10 we show the partial sum

$$F_n(x) = \sum_{k=1}^{n} \frac{\sin(kx)}{k}$$

with $n = 20$ and in Figure 11 we show it with $n = 100$. The horizontal lines of height $\pm\pi/2$ are also drawn.

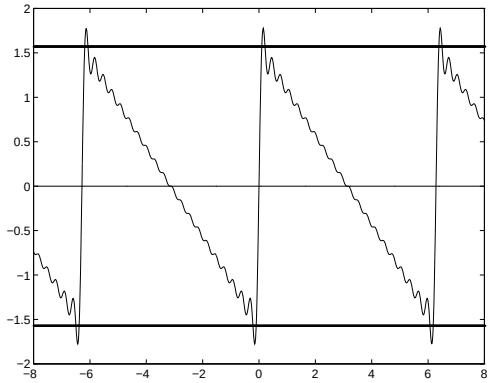


FIGURE 10. Fourier sum through $\sin(20x)$

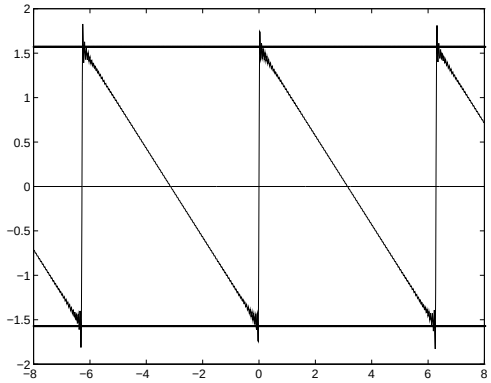


FIGURE 11. Fourier sum through $\sin(100x)$

Notice the "overshoot" near the discontinuities. If you graph $F_n(t)$ for $n = 1000$ or $n = 10^6$, you will get a similar picture. The spike near $x = 0$ will move in closer to $x = 0$, but won't get any shorter. This is the "Gibbs phenomenon." We have $F(0+) = \pi/2$, but it seems that for any n the partial sum F_n overshoots this value by a factor of 18% or so.

A little experimentation with Matlab shows that the spike in $F_n(x)$ occurs at around $x = x_0/n$ for some value of x_0 independent of n. It

turns out that we can compute the limiting value of $F_n(x_0/n)$ for *any* x_0:

Claim. For any x_0,

$$\lim_{n\to\infty} F_n\left(\frac{x_0}{n}\right) = \int_0^{x_0} \frac{\sin t}{t}\, dt.$$

To see this, rewrite the sum as

$$F_n\left(\frac{x_0}{n}\right) = \sum_{k=1}^{n} \frac{\sin(kx_0/n)}{kx_0/n} \cdot \frac{x_0}{n}.$$

Using the notation

$$f(t) = \frac{\sin t}{t}$$

this is

$$F_n\left(\frac{x_0}{n}\right) = \sum_{k=1}^{n} f\left(\frac{kx_0}{n}\right) \cdot \frac{x_0}{n}$$

You will recognize the right hand side as a Riemann sum for the function $f(t)$, between $t = 0$ and $t = x_0$. In the limit we get the integral, and this proves the claim.

To find the largest overshoot, we should look for the maximal value of $\displaystyle\int_0^{x_0} \frac{\sin t}{t}\, dt$. Figure 12 shows a graph of $\dfrac{\sin t}{t}$:

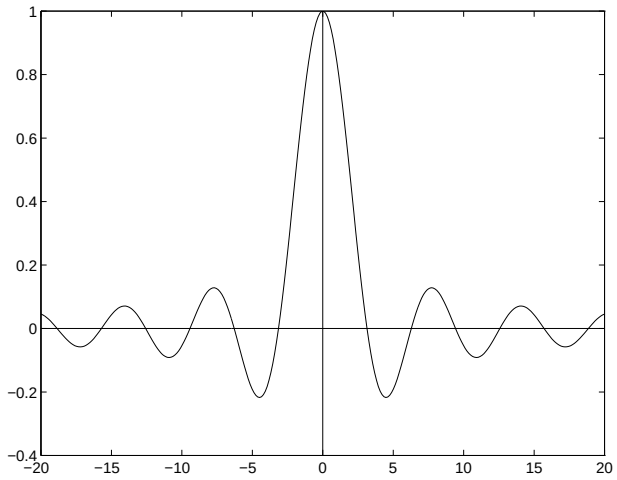


FIGURE 12. $\dfrac{\sin t}{t}$

The integral hits a maximum when $x_0 = \pi$, and the later humps are smaller so it never regains this size again. We now know that

$$\lim_{n\to\infty} F_n\left(\frac{\pi}{n}\right) = \int_0^\pi \frac{\sin t}{t}\,dt.$$

The actual value of this definite integral can be estimated in various ways. For example, the power series for $\sin t$ is

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots.$$

Dividing by t and integrating term by term,

$$\int_0^{x_0} \frac{\sin t}{t}\,dt = x_0 - \frac{x_0^3}{3\cdot 3!} + \frac{x_0^5}{5\cdot 5!} - \cdots.$$

Take $x_0 = \pi$. Pull out a factor of $\pi/2$, to compare with $F(0+) = \pi/2$:

$$\int_0^\pi \frac{\sin t}{t}\,dt = \frac{\pi}{2}\cdot G,$$

where

$$G = 2\left(1 - \frac{\pi^2}{3\cdot 3!} + \frac{\pi^4}{5\cdot 5!} - \cdots\right).$$

The sum converges quickly and gives

$$G = 1.17897974447216727\ldots.$$

We have found, on the graphs of the Fourier partial sums, a sequence of points which converges to the observed overshoot:

$$\left(\frac{\pi}{n}, F_n\left(\frac{\pi}{n}\right)\right) \to \left(0, (1.1789\ldots)\cdot\frac{\pi}{2}\right),$$

that is, about 18% too large. As a proportion of the *gap* between $F(0-) = -\pi/2$ and $F(0+) = +\pi/2$, this is $(G-1)/2 = 0.0894\ldots$ or about 9%. It can be shown that this is the highest overshoot.

The Gibbs overshoot occurs at every discontinuity of a piecewise continuous periodic function $F(x)$. Suppose that $F(x)$ is discontinuous at $x = a$. The overshoot comes to the same 9% of the gap, $F(a+) - F(a-)$, in every case.

Compare this effect to the basic convergence theorem for Fourier series:

Theorem. If $F(x)$ is piecewise continuous and periodic, then for any fixed number a the Fourier series evaluated at $x = a$ converges to $F(a)$ if $F(x)$ is continuous at a, and to the average $\dfrac{F(a+) + F(a-)}{2}$ in general.

The Gibbs effect does not conflict with this, because the point at which the overshoot occurs moves (it gets closer to the point of discontinuity) as n increases.

The Gibbs effect was first noticed by a British mathematician named Wilbraham in 1848, but then forgotten about till it was observed in the output of a computational machine built by the physicist A. A. Michelson (known mainly for the Michelson-Morey experiment, which proved that light moved at the same speed in every direction, despite the motion of the earth through the ether). Michelson wrote to J. Willard Gibbs, the best American physical mathematician of his age and Professor of Mathematics at Yale, who quickly wrote a paper explaining the effect.

16.4. **Fourier distance.** One can usefully regard the Fourier coefficients of a function $f(t)$ as the "coordinates" of $f(t)$ with respect to a certain coordinate system.

Imagine a vector $\mathbf{v}$ in 3-space. We can compute its x coordinate in the following way: move along the x axis till you get to the point *closest to* $\mathbf{v}$. The value of x you find yourself at *is* the x-coordinate of the vector $\mathbf{v}$.

Similarly, move about in the (x, y) plane till you get to the point which is closest to $\mathbf{v}$. This point is the orthogonal projection of $\mathbf{v}$ into the (x, y) plane, and its coordinates are the x and y coordinates of $\mathbf{v}$.

Just so, one way to think of the component $a_n \cos(nt)$ in the Fourier series for $f(t)$ is this: it is the multiple of $\cos(nt)$ which is "closest" to $f(t)$.

The "distance" between functions intended here is hinted at by the Pythagorean theorem. To find the distance between two points in Euclidean space, we take the square root of the sum of squares of differences of the coordinates. When we are dealing with functions (say on the interval between $-\pi$ and π), the analogue is

$$\boxed{\operatorname{dist}(f(t), g(t)) = \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (f(t) - g(t))^2 \, dt \right)^{1/2}}. \tag{1}$$

This number is the **root mean square distance** between $f(t)$ and $g(t)$. The fraction $1/2\pi$ is inserted so that $\operatorname{dist}(1, 0) = 1$ (rather than $\sqrt{2\pi}$) and the calculations on p. 560 of Edwards and Penney show that

for $n > 0$

$$\text{dist}(\cos(nt), 0) = \frac{1}{\sqrt{2}}, \qquad \text{dist}(\sin(nt), 0) = \frac{1}{\sqrt{2}}.$$

The root mean square distance between $f(t)$ and the zero function is called the *norm* of $f(t)$, and is a kind of mean amplitude. The norm of the periodic system response is recorded as "RMS" in the Mathlets `Harmonic Frequency Response` and `Harmonic Frequency Response II`.

One may then try to approximate a function $f(t)$ by a linear combination of $\cos(nt)$'s and $\sin(nt)$'s, by adjusting the coefficients so as to minimize the "distance" from the finite Fourier sum and the function $f(t)$. The Fourier coefficients give the best possible multiples.

Here is an amazing fact. Choose coefficients a_n and b_n *randomly* to produce a function $g(t)$. Then vary one of them, say a_7, and watch the distance between $f(t)$ and this varying function $g(t)$. This distance achieves a minimum precisely when a_7 equals the coefficient of $\cos(7t)$ in the Fourier series for $f(t)$. This effect is *entirely independent* of the other coefficients you have used. You can fix up one at a time, ignoring all the others till later. You can adjust the coefficients to progressively minimize the distance to $f(t)$ in any order, and you will never have to go back and fix up your earlier work. It turns out that this is a reflection of the "orthogonality" of the $\cos(nt)$'s and $\sin(nt)$'s, expressed in the fact, presented on p. 560 of Edwards and Penney, that the integrals of products of distinct sines and cosines are always zero.

16.5. **Complex Fourier series.** With all the sines and cosines in the Fourier series, there must be a complex exponential expression for it. There is, and it looks like this:

$$\boxed{f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}} \tag{2}$$

The power and convenience we have come to appreciate in the complex exponential is at work here too, making computations much easier.

To obtain an integral expression for one of these coefficients, say c_m, the first step is to multiply the expression (2) by e^{-imt} and integrate:

$$\int_{-\pi}^{\pi} f(t)e^{-imt}dt = \sum_{n=-\infty}^{\infty} c_n \int_{-\pi}^{\pi} e^{i(n-m)t}dt \tag{3}$$

Now

$$\int_{-\pi}^{\pi} e^{i(n-m)t}dt = \begin{cases} 2\pi & \text{if } m = n \\ \left.\dfrac{e^{i(n-m)t}}{i(n-m)}\right|_{-\pi}^{\pi} = 0 & \text{if } m \neq n, \end{cases}.$$

The top case holds because then the integrand is the constant function 1. The second case follows from $e^{i(n-m)\pi} = (-1)^{n-m} = e^{i(n-m)(-\pi)}$. Thus only one term in (3) is nonzero, and we conclude that

$$\boxed{c_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-imt}\,dt} \tag{4}$$

This works perfectly well even if $f(t)$ is complex valued. When $f(t)$ is in fact real valued, so that $\overline{f(t)} = f(t)$, (4) implies first that c_0 is real; it's the average value of $f(t)$, that is, in the older notation for Fourier coefficients, $c_0 = a_0/2$. Also, $c_{-n} = \overline{c_n}$ because

$$c_{-n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-i(-n)t}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(t)}e^{int}dt = \overline{c_n}.$$

Since also $e^{-int} = \overline{e^{int}}$, the nth and $(-n)$th terms in the sum (2) are conjugate to each other. We will group them together. The numbers will come out nicely if we choose to write

$$c_n = (a_n - ib_n)/2 \tag{5}$$

with a_n and b_n real. Then $c_{-n} = (a_n + ib_n)/2$, and we compute that

$$c_n e^{int} + c_{-n}e^{-int} = 2\mathrm{Re}\,(c_n e^{int}) = a_n\cos(nt) + b_n\sin(nt).$$

(I told you the numbers would work out well, didn't I?) The series (2) then becomes the usual series

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n\cos(nt) + b_n\sin(nt)\right).$$

Moreover, taking real and imaginary parts of the integral (4) (and continuing to assume $f(t)$ is real valued) we get the usual formulas

$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(nt)dt, \qquad b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(nt)dt.$$

16.6. **Harmonic response.** One of the main uses of Fourier series is to express periodic system responses to general periodic signals. For example, if we drive an undamped spring with a plunger at the end of the spring, the equation is given by

$$m\ddot{x} + kx = kf(t)$$

where $f(t)$ is the position of the plunger, and the x coordinate is arranged so that $x = 0$ when the spring is relaxed and $f(t) = 0$. The natural frequency of the spring/mass system is $\omega = \sqrt{k/m}$, and dividing the equation through by m gives

$$\ddot{x} + \omega^2 x = \omega^2 f(t)\,. \tag{6}$$

This equation is illustrated in the Mathlet `Harmonic Frequency Response`.

An example is given by taking for $f(t)$ the squarewave $\mathrm{sq}(t)$, the function which is periodic of period 2π and such that

$$\mathrm{sq}(t) = \begin{cases} 1 & \text{for} \quad 0 < t < \pi \\ -1 & \text{for} \quad -\pi < t < 0 \end{cases}$$

Its Fourier series is

$$\mathrm{sq}(t) = \frac{4}{\pi}\left(\sin(t) + \frac{\sin(3t)}{3} + \frac{\sin(5t)}{5} + \cdots\right)\,. \tag{7}$$

The periodic system response to the term in the Fourier series for $\omega^2\mathrm{sq}(t)$

$$\frac{4\omega^2}{\pi n}\sin(nt)$$

(where n is an odd integer) is, by the Exponential Reponse Formula (10.10),

$$\frac{4\omega^2}{\pi n}\cdot\frac{\sin(nt)}{\omega^2 - n^2}\,.$$

Thus the periodic system response to $f(t) = \mathrm{sq}(t)$ is given by the Fourier series

$$x_p(t) = \frac{4\omega^2}{\pi}\left(\frac{\sin t}{\omega^2 - 1} + \frac{\sin(3t)}{3(\omega^2 - 9)} + \cdots\right) \tag{8}$$

as long as ω isn't one of the frequencies of the Fourier modes of the signal, i.e. the odd integers.

This expression explains important general features of the periodic solution. When the natural frequency of the system, ω, is near to one of the frequencies present in the Fourier series for the signal (odd integers in this example), the periodic system response x_p is dominated by the

near resonant response to that mode in the signal. When ω is slightly larger than $(2k+1)$ the system response is in phase; as ω decreases though the value $(2k+1)$, the system passes through resonance with the signal (and when $\omega = 2k+1$ there is *no* periodic solution), and comes out on the other side in anti-phase.

In this example, and many others, however, the *same* solution can be obtained quite easily using standard methods of linear ODEs, using some simple features of the solution. These features can be seen directly from the equation, but from our present perspective it's easier to see them from (8). They are:

$$x_p(0) = 0\,, \qquad x_p(\pi) = 0\,.$$

I claim that as long as ω isn't an integer, (6) has just one solution with these properties. That solution is given as a Fourier series by (8), but we can write it out differently using our older methods.

In the interval $[0, \pi]$, the equation is simply

$$\ddot{x} + \omega^2 x = \omega^2\,.$$

We know very well how to solve this! A particular solution is given by a constant, namely 1, and the general solution of the homogeneous equation is given by $a\cos(\omega t) + b\sin(\omega t)$. So

$$x_p = 1 + a\cos(\omega t) + b\sin(\omega t)$$

for some constants a, b.

Substituting $t = 0$ gives $a = -1$, so

$$x_p = 1 - \cos(\omega t) + b\sin(\omega t)\,, \qquad 0 < t < \pi. \tag{9}$$

Substituting $t = \pi$ gives the value for b, depending upon ω:

$$b = \frac{\cos(\pi\omega) - 1}{\sin(\pi\omega)}\,.$$

In the interval $[-\pi, 0]$, the complete signal is $-\omega^2$, so exactly the same calculation gives the negative of the function just written down. Therefore the solution x_p is the *odd* function of period 2π extending

$$x_p = 1 - \cos(\omega t) + \left(\frac{\cos(\pi\omega) - 1}{\sin(\pi\omega)}\right)\sin(\omega t)\,, \qquad 0 < t < \pi\,. \tag{10}$$

The Fourier series of this function is given by (8), but I for one would never have guessed that the expression (8) summed up to such a simple function.

Let's finish up our analysis of this example by thinking about the situation in which the natural frequency ω equals the circular frequency of one of the potential Fourier components of the signal—i.e., an integer, in this case.

In case ω is an even integer, the expression for b is indeterminate since both numerator and denominator are zero. However, in this case the function $x_p = 1 - \cos(\omega t)$ already satisfies $x_p(\pi) = 0$, so we can (and must!) take $b = 0$. Thus x_p is the odd extension of $1 - \cos(\omega t)$. In this case, however, notice that this is not the only periodic solution; indeed, in this case *all* solutions are periodic, since the general solution is (writing $\omega = 2k$)

$$x_p + c_1 \cos(2kt) + c_2 \sin(2kt)$$

and all these are periodic of period 2π.

In case ω is an odd integer, $\omega = 2k + 1$, there are no periodic solutions; the system is in resonance with the Fourier mode $\sin((2k + 1)t)$ present in the signal. We can't solve for the constant b; the zero in its denominator is not canceled by a zero in its numerator. It is not hard to write down a particular solution in this case too, using the Resonance Exponential Response Formula, Section 12.

We have used the undamped harmonic oscillator for this example, but the same methods work in the presence of damping. In that case it is much easier to use the complex form of the Fourier series (Section 16.5 below) since the denominator in the Exponential Response Formula is no longer real.

17. Impulses and generalized functions

In calculus you learn how to model processes using functions. Functions have their limitations, though, and, at least as they are treated in calculus, they are not convenient for modeling some important processes and events, especially those involving sudden changes. In this section we explain how the function concept can be extended to a wider class of objects which conveniently model such processes.

17.1. **From bank accounts to the delta function.** Recall from Section 2 the model of the savings account, beginning with the "difference equation"

$$x(t + \Delta t) = x(t) + I(t)x(t)\Delta t + q(t)\Delta t \tag{1}$$

and continuing to the differential equation

$$\dot{x} - I(t)x = q(t)\,. \tag{2}$$

Here $I(t)$ is the interest rate at time t and $q(t)$ is the rate of contribution. Here $q(t)$ is measured in dollars per year, and, being a mathematician, I have taken the limit and replaced frequent small payments (say \$1 each day) by a continual payment (at a constant *rate* of $q(t) = 365$ dollars per year).

In fact, banks do not behave like mathematicians and take this limit. They use the difference equation (1) and compute intensively. Solving the ODE (2) is much simpler, and leads to good agreement with the discrete calculation. This continuous approximation is critical to the use of differential equations in modeling. The world is in fact discrete, made up of atoms (and smaller discrete structures), but in order to understand this seething digital mass we make continuous—even differentiable—approximations.

On the other hand, there are some processes which cannot be conveniently accommodated by this paradigm. For example, while I continue to save at the rate of \$1/day, my wealthy aunt decided to give me \$1000 as a birthday present, and I deposited this into the bank in one lump sum. How can we model this process?

One way is the following: solve the ODE (2) with $q = 365$, reflecting the compounding of interest in my account, subject to an appropriate initial condition, say $x(0) = x_0$. Then, at the moment I plan to deposit the gift, say at $t = t_1 > 0$, stop this process and compute my balance at the instant of the big deposit, $x(t_1)$. Write x_1 for this number. Then start the ODE up again with a new initial condition, $x(t_1) = x_1 + 1000$,

and solve anew to get $x(t)$ for $t > t_1$. Another way to think of this is to imagine that a separate fund is set up with the \$1000 gift, and allowed to grow at the same interest rate. The two perspectives are equivalent, by superposition. We get

$$x(t) = \begin{cases} -q/I + (x_0 + q/I)e^{It} & \text{if } 0 < t < t_1 \\ \\ -q/I + (x_0 + q/I)e^{It} + 1000e^{I(t-t_1)} & \text{if } t > t_1 \end{cases}$$

The number $t - t_1$ is the amount of time the gift has been in the bank at time t.

I'd like to introduce some notation here. My bank balance seems to have *two* values at $t = t_1$: x_1, and x_1+1000. There is notation to handle this. One writes $x(t_1-)$ for the balance at $t = t_1$ as estimated from knowledge of the balance from times before the gift; mathematically,

$$x(t_1-) = \lim_{t \uparrow t_1} x(t).$$

Similarly, $x(t_1+)$ is the balance at $t = t_1$ from the perspective of later times; mathematically,

$$x(t_1+) = \lim_{t \downarrow t_1} x(t).$$

The actual value we assign as $x(t_1)$ is unimportant and can be left undeclared.

The fact is that the approach we just used to deal with this windfall situation is often the preferred strategy. Still, it would be convenient to find some way to *incorporate* a one-time essentially instantaneous gift into the *rate* $q(t)$. Then there's the withdrawal I made when I bought the BMW, too—so the machinery should accommodate a series of such events, at different times, of various magnitudes, and of either sign.

A good way to understand a rate is by considering the cumulative total. The cumulative total contribution to the account, from time zero up to time t, is

$$Q(t) = \int_0^t q(\tau) d\tau,$$

so in our case (before the gift) $Q(t) = 365t$. (Note that I had to come up with a new symbol for the time variable inside the integral.) Then the rate is given (by the fundamental theorem of calculus) as the derivative of the cumulative total: $q = Q'(t)$. When I incorporate my aunt's gift, the cumulative total jumps by \$1000 at time t_1: so it is given by

$$Q(t) = \begin{cases} 365t & \text{for } t < t_1 \\ 365t + 1000 & \text{for } t > t_1 \end{cases}$$

(and it doesn't much matter what we declare $Q(t_1)$ to be.)

This is a perfectly good function, but it fails to be differentiable (in the usual sense) at $t = t_1$, so the corresponding rate, $q(t) = Q'(t)$, has problems at $t = t_1$.

We can approximate such a rate in the following way. Imagine that the gift isn't deposited all at once but instead over a short period of time, say h, starting at time t_1. Thus the new rate function has a graph which is horizontal with value 365 until $t = t_1$, then it jumps to a value of $365 + 1000/h$, and then, at $t = t_1 + h$, it falls back to the constant value 365. We can try to take a limit here, letting $h \to 0$, but the result is not a function in the usual sense of the word.

Nevertheless, all these considerations indicate that there may be a mathematical concept a little more general than the function concept which serves to model a *rate which produces a jump in the cumulative total*. This is indeed the case: they are called *generalized functions*, and they are heavily studied by mathematicians and used by engineers and scientists. They form a convenient language.

17.2. **The delta function.** The most basic rate of this sort is the one which produces a "unit step" cumulative total:

$$u(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t > 0 \end{cases}$$

This $u(t)$ is an important function, and it is sometimes called the **Heaviside function**. It doesn't much matter what we declare $u(0)$ to be, and we'll just leave it unspecified; but we do know $u(0-) = 0$ and $u(0+) = 1$. The corresponding rate is the **Dirac delta function**,

$$\delta(t) = u'(0).$$

This object $\delta(t)$ behaves like an ordinary function, in fact like the constant function with value 0, except at $t = 0$, where it can be thought of as taking on such a large value that the area under the graph is 1. The delta function is also called the **unit impulse function**.

Using this notation, my rate of contribution, including my aunt's gift, is

$$q(t) = 365 + 1000\,\delta(t - t_1).$$

Just as for (ordinary) functions, subtracting t_1 inside shifts the graph right by t_1 units; the spike occurs at $t = t_1$ rather than at $t = 0$.

We can also use the delta function to model discrete monthly contributions accurately, without replacing them with a continuous contribution. If the rate is \$365 per year, and I contribute at the start of each month (or more precisely at $t = 0, 1/12, 2/12, \ldots$) starting at $t = 0$, then the rate is

$$q(t) = \frac{365}{12}\left(\delta(t) + \delta(t - \frac{1}{12}) + \delta(t - \frac{2}{12}) + \cdots\right).$$

When these shifted and scaled delta functions are added to "ordinary" functions you get a "generalized function." I'll describe a little part of the theory of generalized functions. The next few paragraphs will sound technical. I hope they don't obscure the simplicity of the idea of generalized functions as a model for abrupt changes.

I will use the following extensions of a definition from Edwards and Penney (p. 268): To prepare for it let me call a collection of real numbers $a_1, a_2, \ldots$, **sparse** if for any $r > 0$ there are only finitely many of k such that $|a_k| < r$. So any finite collection of numbers is sparse; the collection of whole numbers is sparse; but the collection of numbers $1, 1/2, 1/3, \ldots$, is not sparse. Sparse sets don't bunch up. The empty set is sparse.

When I describe a function (on an interval) I typically won't insist on knowing its values for *all* points in the interval. I'll allow a sparse collection of points at which the value is undefined. We already saw this in the definition of $u(t)$ above.

A function $f(t)$ (on an interval) is **piecewise continuous** if (1) it is continuous everywhere (in its interval of definition) except at a sparse collection of points; and (2) for every a, both $f(a+)$ and $f(a-)$ exist.

A function $f(t)$ is **piecewise differentiable** if (1) it is piecewise continuous, (2) it is differentiable everywhere except at a sparse collection of points, and its derivative is piecewise continuous.

We now want to extend this by including delta functions. A **generalized function** is a piecewise continuous function $f_r(t)$ plus a linear combination of delta functions,

$$f_s(t) = \sum b_k \delta(t - a_k), \tag{3}$$

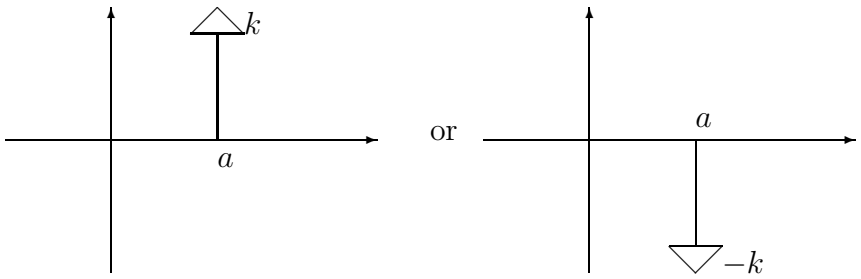
where the a_k's form a sparse set.

Write $f(t)$ for the sum:

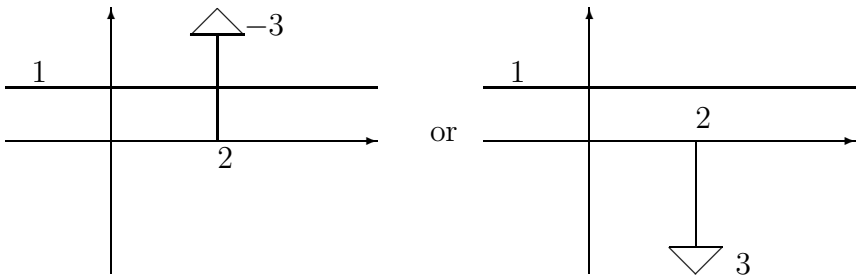
$$f(t) = f_r(t) + f_s(t).$$

$f_r(t)$ is the "regular part" of $f(t)$, and $f_s(t)$ is the "singular part." We define $f(a-)$ to be $f_r(a-)$ and $f(a+)$ to be $f_r(a+)$. Since the actual value of $f(t)$ at $t = a$ is not used to compute these limits, this is a good definition even if a is one of the a_k's.

We will use a "harpoon" to denote a delta function in a graph. The harpoon should be thought to be very high. This notation by itself does not include the information of the area under the graph. To deal with this we will decorate the barb of the harpoon representing $k\delta(t-a)$ with the number k. k may be negative, in which case the harpoon might better be thought of as extending downward. We will denote the same function, $k\delta(t-a)$ equally by a downward harpoon decorated with $-k$:



For example, $1 - 3\delta(t-2)$ can be denoted by either of the following graphs.



A harpoon with $k = 0$ is the same thing as no harpoon at all: $0\delta(t-a) = 0$. We'll call the term $b_k\delta(t-a_k)$ occurring in $f_s(t)$ **the singularity of** $f(t)$ **at** $t = a_k$. If a is not among the a_k's (or if $a = a_k$ but $b_k = 0$) then there is no singularity in $f(t)$ at $t = a$.

17.3. **Integrating generalized functions.** Generalized functions are set up so they can be integrated. We know what the integral of a delta function should be, since we are to think of it as the derivative of the unit step function:

$$\int_b^c \delta(t-a)\,dt = u(c-a) - u(b-a).$$

If $b < a < c$, this is 1. If a is not between b and c, this is 0. If $a = b$ or $a = c$ then this integral involves the expression $u(0)$, which is best thought of as undefined. We can however define

$$\int_{b-}^{c+} f(t)\,dt = \lim_{b'\uparrow b}\lim_{c'\downarrow c}\int_{b'}^{c'} f(t)\,dt,$$

and this gives a well defined result when $f(t) = \delta(t-a)$: Assuming $b \le c$,

$$\int_{b-}^{c+} \delta(t-a)\,dt = 1 \quad \text{if} \quad b \le a \le c,$$

and zero otherwise. In particular,

$$\int_{a-}^{a+} \delta(t-a)\,dt = 1.$$

Now if $f(t)$ is any generalized function, we can define the integral

$$\int_{b-}^{c+} f(t)\,dt$$

by integrating the regular part $f(t)$ in the usual way, and adding the sum of the b_k's over k for which $b \le a_k \le c$ (using the notation of (3)).

The multiple of the delta function that occurs at $t = a$ in a generalized function can be expressed as

$$b = \int_{a-}^{a+} f(t)\,dt.$$

17.4. **The generalized derivative.** Generalized functions let us make sense of the derivative of a function which is merely piecewise differentiable.

For example, we began by saying that the "derivative" of the piecewise differentiable function $u(t-a)$ is the generalized function $\delta(t-a)$. This understanding lets us define the **generalized derivative** of any piecewise continuously differentiable function $f(t)$. It is a generalized function. Its regular part, $f_r'(t)$, is the usual derivative of $f(t)$ (which

is defined except where the graph of $f(t)$ has breaks or corners), and its singular part is given by the sum of terms

$$(f(a+) - f(a-))\delta(t-a),$$

summed over the values a of t where the graph of $f(t)$ has breaks. Each shifted and scaled δ function records the instantaneous velocity needed to accomplish a sudden jump in the value of $f(t)$. When the graph of $f(t)$ has a corner at $t = a$, the graph of $f'(t)$ has a jump at $t = a$ and isn't defined at $t = a$ itself; this is a discontinuity in the piecewise continuous function $f'_r(t)$.

With this definition, the "fundamental theorem of calculus"

$$\int_{b-}^{c+} f'(t)\,dt = f(c+) - f(b-)$$

holds for generalized functions.

For further material on this approach to generalized functions the reader may consult the article "Initial conditions, generalized functions, and the Laplace transform," IEEE Control Systems Magazine 27 (2007) 22–35, by Kent Lundberg, David Trumper, and Haynes Miller. A version is available at `http://www-math.mit.edu/~hrm/papers/lmt.pdf`.

18. Impulse and step responses

In real life, we often do not know the parameters of a system (e.g. the spring constant, the mass, and the damping constant, in a spring-mass-dashpot system) in advance. We may not even know the order of the system—there may be many interconnected springs (or diodes). (We will, however, suppose that all the systems we consider are linear and time independent, LTI.) Instead, we often learn about a system by watching how it responds to various input signals.

The simpler the signal, the clearer we should expect the signature of the system parameters to be, and the easier it should be to predict how the system will respond to other more complicated signals. To simplify things we will always begin the system from "rest."

In section we will study the response of a system from rest initial conditions to two standard and very simple signals: the unit impulse $\delta(t)$ and the unit step function $u(t)$.

The theory of the convolution integral, Section 19, gives a method of determining the response of a system to *any* input signal, given its unit impulse response.

18.1. **Impulse response.** In engineering one often tries to understand a system by studying its responses to known signals. Suppose for definiteness that the system is given by a first order left hand side $\dot{x}+p(t)x$. (The right hand side $q(t)$, isn't part of the "system"; it is the "input signal.") The variable x will be called the "system response," and in solving the ODE we are calculating that response. The analysis proceeds by starting "at rest," by which is meant $x(t) = 0$ for t less than the moment at which the signals occur. One then feeds the system various signals and watches the system response. In a certain sense the simplest signal it can receive is a delta function concentrated at some time t_0: $\delta(t - t_0)$. This signal is entirely concentrated at a single instant of time, but it has an effect nevertheless. In the case of a first order system, we have seen what that effect is, by thinking about what happens when I contribute a windfall to my bank account: for $t < t_0$, $x(t) = 0$; and for $t > t_0$, $x(t)$ is the solution to $\dot{x} + p(t)x = 0$ subject to the initial condition $x(t_0) = 1$. (Thus $x(t_0-) = 0$ and $x(t_0+) = 1$.) If $p(t) = a$ is constant, for example, this amounts to

$$x(t) = \begin{cases} 0 & \text{if } t < t_0 \\ e^{-a(t-t_0)} & \text{if } t > t_0. \end{cases}$$

This system response depends upon t_0, but if the system is LTI, as it is in this example, its dependence is very simple: The response to a unit impulse at $t = 0$ is called the **weight function** or **unit impulse response** of the system, or written $w(t)$. If the system is given by $\dot{x} + ax$, the weight function is given by

$$w(t) = \begin{cases} 0 & \text{for } t < 0 \\ e^{-at} & \text{for } t > 0. \end{cases}$$

In terms of it, the response to a unit impulse at any time t_0 is

$$x(t) = w(t - t_0).$$

18.2. **Impulses in second order equations.** The word "impulse" comes from the interpretation of the delta function as a component of the driving term $q(t)$ in a second order system:

$$m\ddot{x} + b\dot{x} + cx = q(t). \tag{1}$$

In the mechanical interpretation of this equation, $q(t)$ is regarded as an external force acting on a spring-mass-dashpot system. Force affects acceleration, so the cumulative total of force, that is the time integral, affects velocity. If we have a very large force exerted over a very small time, the acceleration becomes very large for a short time, and the velocity increases sharply. In the limit we have an **impulse**, also known as a good swift kick. If $q(t) = a\delta(t - t_0)$, the system response is that the velocity $\dot{x}$ increases abruptly at $t = t_0$ by the quantity a/m. This produces a corner in the graph of x as a function of t, but not a break; the position does not change abruptly.

Thus the system response, $w(t)$, to a unit impulse at $t = 0$ is given for $t < 0$ by $w(t) = 0$, and for $t > 0$ by the solution to (1) subject to the initial condition $x(0) = 0$, $\dot{x}(0) = 1/m$.

For example, if the system is governed by the homogeneous LTI equation $\ddot{x} + 2\dot{x} + 5x = 0$, an independent set of real solutions is $\{e^{-t}\cos(2t),\, e^{-t}\sin(2t)\}$, and the solution to the initial value problem with $x(0) = 0$, $\dot{x}(0) = 1$, is $(1/2)e^{-t}\sin(2t)$. Thus

$$w(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ (1/2)e^{-t}\sin(2t) & \text{for} \quad t > 0. \end{cases}$$

This is illustrated in Figure 13. Note the aspect in this display: the vertical has been inflated by a factor of more than 10. In fact the slope $\dot{w}(0+)$ is 1.

The unit impulse response needs to be defined in two parts; it's zero for $t < 0$. This is a characteristic of *causal* systems: the impulse at

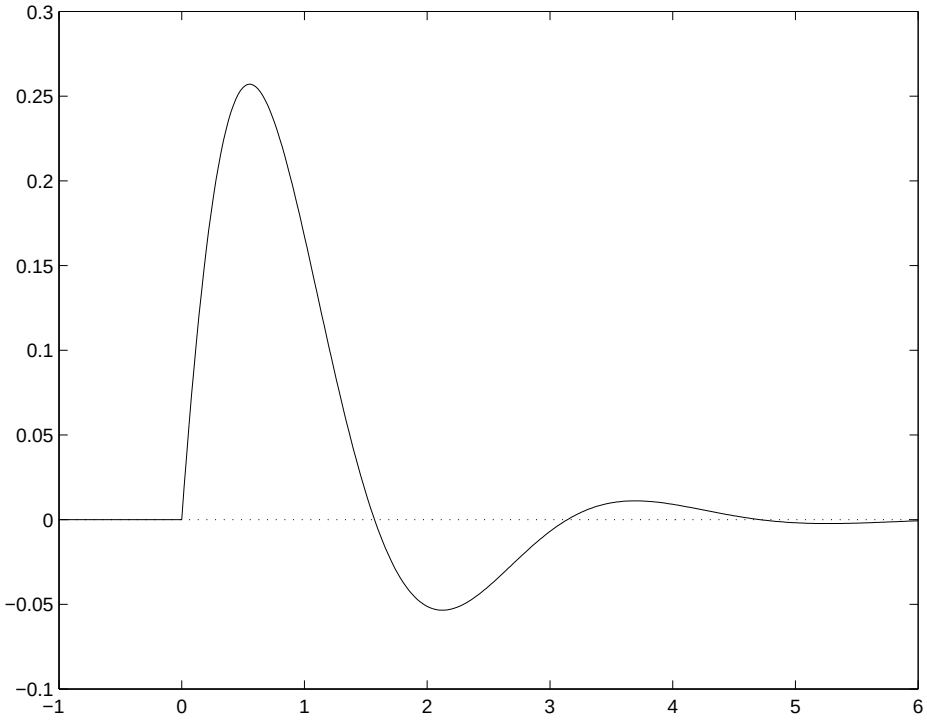


FIGURE 13. The weight function for $\ddot{x} + 2\dot{x} + 5x$

$t = 0$ has no effect on the system when $t < 0$. In a causal system the unit impulse response is always zero for negative time.

18.3. **Singularity matching.** Differentiation increases the order of singularity of a function. For example, the "ramp" function

$$\text{ramp}(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ t & \text{for} \quad t > 0. \end{cases}$$

is not differentiable at $t = 0$ but it is continuous. Its derivative is the step function $u(t)$, which is not continuous at $t = 0$ but it is a genuine function; its singular part is zero. But *its* derivative is the delta function. (This can be made to continue; one can define an even more singular type of generalized function, of which $\delta'(t)$, often called a *doublet*, is an example, but we will not enter into this here.)

Suppose a function satisfies an ODE, say

$$m\ddot{x} + b\dot{x} + cx = q(t),$$

in which $q(t)$ may have a singular part. Whatever singularities x may have get accentuated by the process of differentiation, so the most singular part of $q(t)$ must match up with the most singular part of $m\ddot{x}$. This then forces x to be not too very singular; otherwise its second derivative would be more singular than $q(t)$.

To be more precise, if $q(t)$ is a generalized function in our sense, then its singular part must occur as the singular part of $m\ddot{x}$. The result is that $\dot{x}$ does not have a singular part, but does have discontinuities at the locations at which $q(t)$ has delta components. Similarly, x is continuous, but has jumps in its derivative at those locations. This makes physical sense: a second order system response to a generalized function is continuous but shows sudden jumps in velocity where the signal exhibits impulses.

This analysis is quantitative. If for example $q(t) = 3\delta(t) + 6t$, $m\ddot{x}$ has singular part $3\delta(t)$, so $\ddot{x}$ has singular part $(3/m)\delta(t)$. Thus $\dot{x}$ is continuous except at $t = 0$ where it has a jump in value of $3/m$; and x is differentiable except at $t = 0$, where its derivative jumps by $3/m$ in value.

In a first order system, say $m\dot{x} + kx = q(t)$, the singular part of $m\dot{x}$ is the singular part of $q(t)$, so x is continuous except at those places. If for example $q(t) = 3\delta(t) + 6t$, $\dot{x}$ has singular part $(3/m)\delta(t)$, so x jumps in value by $3/m$ at $t = 0$.

This line of reasoning is called "singularity matching."

18.4. **Step response.** This is the response of a system at rest to a constant input signal being turned on at $t = 0$. I will write $w_1(t)$ for this system response. If the system is represented by the LTI operator $p(D)$, then $w_1(t)$ is the solution to $p(D)x = u(t)$ with rest initial conditions, where $u(t)$ is the unit step function.

The unit step response can be related to the unit impulse response using the following observation: The time invariance of $p(D)$ is equivalent to the fact that as operators

$$p(D)D = Dp(D).$$

We can see this directly:

$$(a_nD^n + \cdots + a_0I)D = a_nD^{n+1} + \cdots + a_0D = D(a_nD^n + \cdots + a_0I)\,.$$

Using this we can differentiate the equation $p(D)w_1 = 1$ to find that $p(D)(Dw_1) = \delta(t)$, with rest initial conditions. That is to say, $\dot{w}_1(t) = w(t)$, or:

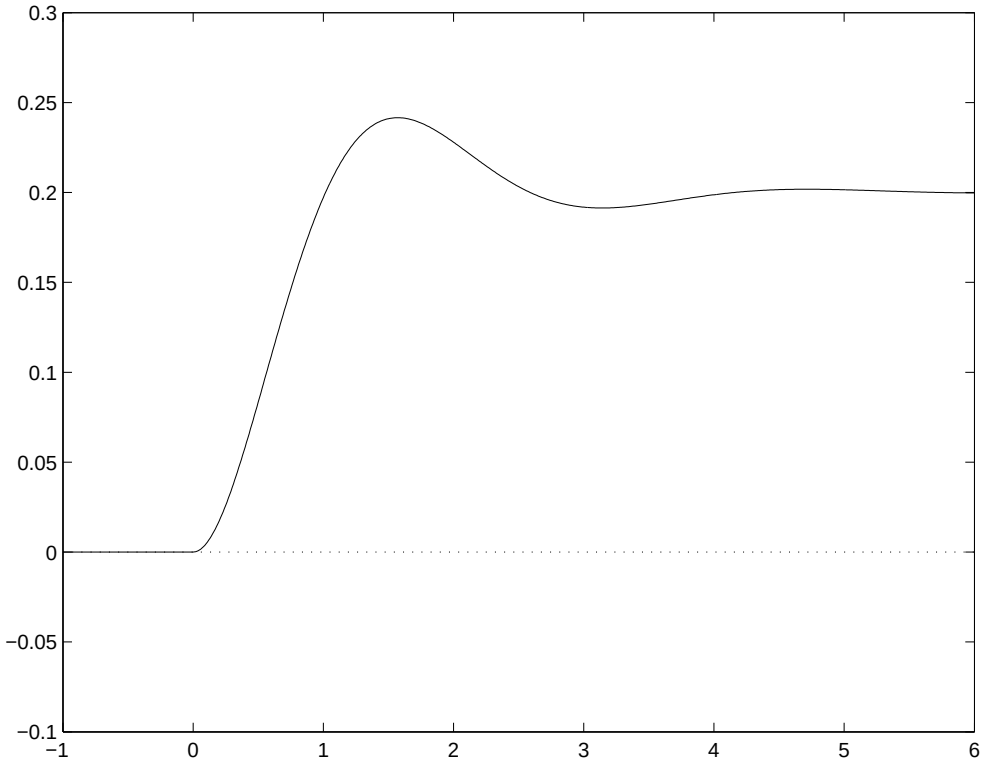


FIGURE 14. The unit step response for $\ddot{x} + 2\dot{x} + 5x$

The derivative of the unit step response is the unit impulse response.

If we return to the system represented by $\ddot{x} + 2\dot{x} + 5x$ considered above, a particular solution to $\ddot{x} + 2\dot{x} + 5x = 1$ is given by $x = 1/5$, so the general solution is $x = (1/5) + e^{-t}(a\cos(2t) + b\sin)2t))$. Setting $x(0) = 0$ and $\dot{x}(0) = 0$ leads to

$$w_1(t) = \begin{cases} 0 & \text{for} \quad t < 0 \\ (1/5) - (e^{-t}/10)(2\cos(2t) + \sin(2t)) & \text{for} \quad t > 0 \end{cases}$$

as illustrated in Figure 14. You can check that the derivative of this function is $w(t)$ as calculated above. In this example the unit impulse response is a simpler function than the unit step response, and this is generally the case.

19. Convolution

19.1. **Superposition of infinitesimals: the convolution integral.** The system response of an LTI system to a general signal can be reconstructed explicitly from the unit impulse response.

To see how this works, start with an LTI system represented by a linear differential operator L with constant coefficients. The system response to a signal $f(t)$ is the solution to $Lx = f(t)$, subject to some specified initial conditions. To make things uniform it is common to specify "rest" initial conditions: $x(t) = 0$ for $t < 0$.

We will approach this general problem by decomposing the signal into small packets. This means we partition time into intervals of length say Δt: $t_0 = 0, t_1 = \Delta t, t_2 = 2\Delta t$, and generally $t_k = k\Delta t$. The kth signal packet is the null signal (i.e. has value zero) except between $t = t_k$ and $t = t_{k+1}$, where it coincides with $f(t)$. Write $f_k(t)$ for the kth packet. Then $f(t)$ is the sum of the $f_k(t)$'s.

Now by superposition the system response (with rest initial conditions) to $f(t)$ is the sum of the system responses to the $f_k(t)$'s separately.

The next step is to estimate the system response to a single packet, say $f_k(t)$. Since $f_k(t)$ is concentrated entirely in a small neighborhood of t_k, it is well approximated as a rate by a multiple of the delta function concentrated at t_k, $\delta(t - t_k)$. The multiple should be chosen so that the cumulative totals match up; that is, it should be the integral under the graph of $f_k(t)$, which is itself well approximated by $f(t_k)\Delta t$. Thus we replace $f_k(t)$ by

$$f(t_k) \cdot \Delta t \cdot \delta(t - t_k).$$

The system response to this signal, a multiple of a shift of the unit impulse, is the same multiple of the same shift of the weight function (= unit impulse response):

$$f(t_k) \cdot \Delta t \cdot w(t - t_k).$$

By superposition, adding up these packet responses over the packets which occur before the given time t gives the system response to the signal $f(t)$ at time t. As $\Delta t \to 0$ this sum approximates an integral taken over time between time zero and time t. Since the symbol t is already in use, we need to use a different symbol for the variable in the integral; let's use the Greek equivalent of t, τ ("tau"). The t_k's get

replaced by τ in the integral, and Δt by $d\tau$:

$$\boxed{x(t) = \int_0^t f(\tau) w(t-\tau)\, d\tau} \tag{1}$$

This is a really wonderful formula. Edwards and Penney call it "Duhamel's principle," but they seem somewhat isolated in this. Perhaps a better name would be the "superposition integral," since it is no more and no less than an integral expression of the principle of superposition. It is commonly called the **convolution integral**. It describes the solution to a general LTI equation $Lx = f(t)$ subject to rest initial conditions, in terms of the unit impulse response $w(t)$. Note that in evaluating this integral τ is always less than t, so we never encounter the part of $w(t)$ where it is zero.

19.2. **Example: the build up of a pollutant in a lake.** Every good formula deserves a particularly illuminating example, and perhaps the following will serve for the convolution integral. It is illustrated by the Mathlet `Convolution: Accumulation`. We have a lake, and a pollutant is being dumped into it, at a certain variable rate $f(t)$. This pollutant degrades over time, exponentially. If the lake begins at time zero with no pollutant, how much is in the lake at time $t > 0$?

The exponential decay is described as follows. If a quantity p of pollutant is dropped into the lake at time τ, then at a later time t it will have been reduced in amount to $pe^{-a(t-\tau)}$. The number a is the decay constant, and $t - \tau$ is the time elapsed. We apply this formula to the small drip of pollutant added between time τ and time $\tau + \Delta\tau$. The quantity is $p = f(\tau)\Delta\tau$ (remember, $f(t)$ is a *rate*; to get a *quantity* you must multiply by time), so at time t the this drip has been reduced to the quantity

$$e^{-a(t-\tau)} f(\tau)\Delta\tau$$

(assuming $t > \tau$; if $t < \tau$, this particular drip contributed zero). Now we add them up, starting at the initial time $\tau = 0$, and get the convolution integral (1), which here is

$$x(t) = \int_0^t f(\tau) e^{-a(t-\tau)}\, d\tau. \tag{2}$$

We found our way straight to the convolution integral, without ever mentioning differential equations. But we can also solve this problem by setting up a differential equation for $x(t)$. The amount of this chemical in the lake at time $t+\Delta t$ is the amount at time t, minus the fraction

that decayed, plus the amount newly added:

$$x(t + \Delta t) = x(t) - ax(t)\Delta t + f(t)\Delta t$$

Forming the limit as $\Delta t \to 0$, we obtain

$$\dot{x} + ax = f(t), \qquad x(0) = 0. \tag{3}$$

We conclude that (2) gives us the solution with rest initial conditions.

An interesting case occurs if $a = 0$. Then the pollutant doesn't decay at all, and so it just builds up in the lake. At time t the total amount in the lake is just the total amount dumped in up to that time, namely

$$\int_0^t f(\tau)\, d\tau\,,$$

which is consistent with (2).

19.3. **Convolution as a "product".** The integral (1) is called the *convolution* of $w(t)$ and $f(t)$, and written using an asterisk:

$$w(t) * f(t) = \int_0^t w(t - \tau) f(\tau)\, d\tau, \qquad t > 0. \tag{4}$$

We have now fulfilled the promise we made at the beginning of Section 18: we can explicitly describe the system response, with rest initial conditions, to any input signal, if we know the system response to just one input signal, the unit impulse:

Theorem. The solution to an LTI equation $Lx = f(t)$, of any order, with rest initial conditions, is given by

$$x(t) = w(t) * f(t),$$

where $w(t)$ is the unit impulse response.

If L is an LTI differential operator, we should thus be able to reconstruct its characteristic polynomial $p(s)$ (so that $L = p(D)$) from its unit impulse response. This is one of the things the Laplace transform does for us; in fact, the Laplace transform of $w(t)$ is the reciprocal of $p(s)$: see Section 21.

The expression (4) can be interpreted formally by a process known as "flip and drag." It is illustrated in the Mathlet `Convolution: Flip and Drag`.

20. Laplace transform technique: coverup

I want to show you some practical tricks which will help you to find the inverse Laplace transform of a rational function. These are refinements on the method of partial fractions which you studied when you learned how to integrate rational functions. Some of this will use complex numbers.

20.1. **Simple case.** First, let's do an easy case:

$$F(s) = \frac{1}{s^2 - 4}.$$

To begin, factor the denominator, and write

$$F(s) = \frac{1}{(s-2)(s+2)} = \frac{a}{s-2} + \frac{b}{s+2} \tag{1}$$

for as yet unknown constants a and b. One way to proceed is to cross multiply and collect terms in the numerator. That is fine but the following is more fun.

To find a, *first multiply through by the corresponding denominator*, $(s-2)$ in this case. You get

$$\frac{1}{s+2} = a + (s-2)(\text{other terms}),$$

in which the "other terms" (namely, $\dfrac{b}{s+2}$) don't have a factor of $(s-2)$ in the denominator. *Then set* $s = 2$:

$$\frac{1}{2+2} = a + (2-2)(\text{other terms}) = a$$

since the second term vanishes. So $a = 1/4$. In exactly the same way, you can multiply (1) through by $s+2$ and then set $s = -2$, to find

$$\frac{1}{-2-2} = b$$

or $b = -1/4$. Thus

$$F(s) = \frac{1/4}{s-2} - \frac{1/4}{s+2}.$$

The tables then show that

$$f(t) = (1/4)(e^{2t} - e^{-2t}).$$

This approach to partial fractions has been called the "cover-up method"; you cover up the denominators of the terms you wish to compute. You can do it without writing anything down; just cover

up the denominators and write down the answer. It works well but does not completely handle the case in which the denominator has a repeated root.

20.2. **Repeated roots.** For example look at

$$F(s) = \frac{s}{s^2 + 2s + 1}.$$

The denominator factors as $(s+1)^2$. We want to find a and b such that

$$F(s) = \frac{s}{(s+1)^2} = \frac{a}{s+1} + \frac{b}{(s+1)^2}.$$

We can begin to use the coverup method: *multiply through by* $(s+1)^2$ *and set* $s = -1$: The left hand side is just -1; the first term vanishes; and the second term is b: so $b = -1$. We can't get a this way, though. One way to find a is to *set* s *to some* other *value*. Any other value will do, and we might as well make our arithmetic as simple as possible. Let's take $s = 0$: then we have

$$0 = \frac{a}{1} + \frac{-1}{1}$$

so $a = 1$:

$$F(s) = \frac{1}{s+1} - \frac{1}{(s+1)^2}.$$

Now the tables show

$$f(t) = e^{-t} - te^{-t}.$$

20.3. **Completing the square.** Suppose

$$F(s) = \frac{1}{s^2 + 2s + 2}.$$

The first part of the method here is to *complete the square in the denominator,* and *rewrite the numerator in the same terms*:

$$\frac{s}{s^2 + 2s + 2} = \frac{a(s+1) + b}{(s+1)^2 + 1}.$$

This works with $a = 1$ and $b = -1$:

$$F(s) = \frac{(s+1) - 1}{(s+1)^2 + 1}.$$

Now the s-shift rule applies, since $F(s)$ is written in terms of $s - a$ (where here $a = -1$). The second part of this method gives you a way to use the s-shift rule without getting too confused. You should *invent*

a new function name—say $G(s)$—and use it to denote the function such that that $F(s) = G(s-a)$. Thus, here,

$$G(s) = \frac{s-1}{s^2+1}.$$

Now the s-shift rule says that if $g(t) \rightsquigarrow G(s)$ then $e^{-t}g(t) \rightsquigarrow G(s+1) = F(s)$, which is to say that $f(t) = e^{-t}g(t)$. The tables give

$$g(t) = \cos t - \sin t$$

so

$$f(t) = e^{-t}(\cos t - \sin t).$$

20.4. **Complex coverup.** Now let's take an example in which the quadratic factor does not occur alone in the denominator: say

$$F(s) = \frac{1}{s^3+s^2-2}.$$

The denominator factors as $s^3+s^2-2 = (s-1)(s^2+2s+2)$. In the example above we learned that the factor s^2+2s+2 should be handled by completing the square and grouping the $(s+1)$ in the numerator:

$$F(s) = \frac{1}{(s-1)((s+1)^2+1)} = \frac{a}{s-1} + \frac{b(s+1)+c}{(s+1)^2+1}.$$

Find a just as before: multiply through by $s-1$ and then set $s=1$, to get $a = 1/5$. To find b and c, *multiply through by the quadratic factor* $(s+1)^2+1$ *and then set* s *equal to a root of that factor.* Having already completed the square, it's easy to find a root: $(s+1)^2 = -1$, so $s+1 = i$ for example, so $s = -1+i$. We get:

$$\frac{1}{(-1+i)-1} = b((-1+i)+1)+c$$

or, rationalizing the denominator,

$$\frac{-2-i}{5} = c+bi$$

Since we want b and c real, we must have $c = -2/5$ and $b = -1/5$:

$$F(s) = \frac{1}{5}\left(\frac{1}{s-1} - \frac{(s+1)+2}{(s+1)^2+1}\right).$$

We're in position to appeal to the s-shift rule, using the tricks described in 20.3, and find

$$f(t) = \frac{1}{5}\left(e^t - e^{-t}(\sin t + 2\cos t)\right).$$

20.5. **Complete partial fractions.** There is another way to deal with quadratic factors: just factor them over the complex numbers and use the coverup method in its original form as in Section 20.1. I don't recommend using this in practice, but it's interesting to see how it works out, and we will use these ideas in Section 22. Using the example

$$F(s) = \frac{1}{s^3 + s^2 - 2}$$

again, we can find *complex* constants a, b, c such that

$$F(s) = \frac{a}{s-1} + \frac{b}{s-(-1+i)} + \frac{c}{s-(-1-i)} \tag{2}$$

Expect that a will be real, and that b and c will be complex conjugates of each other.

Find a just as before; $a = 1/5$. To find b, do the same: multiply through by $s-(-1+i)$ to get

$$\frac{1}{(s-1)(s-(-1-i))} = b + (s-(-1+i))(\text{other terms})$$

and then set $s = -1+i$ to see

$$\frac{1}{(-2+i)(2i)} = b$$

or $b = 1/(-2-4i) = (-1+2i)/10$. The coefficient c can be computed similarly. Alternatively, you can use the fact that the two last terms in (2) must be complex conjugate to each other (in order for the whole expression to come out real) and so discover that $c = \bar{b} = (-1-2i)/10$:

$$F(s) = \frac{1/5}{s-1} + \frac{(-1+2i)/10}{s-(-1+i)} + \frac{(-1-2i)/10}{s-(-1-i)}.$$

The numerators a, b, and c, in this expression are called the **residues** of the poles of $F(s)$; see 22.1 below.

It's perfectly simple to find the inverse Laplace transforms of the terms here:

$$f(t) = \frac{1}{5}\,e^t + \frac{-1+2i}{10}\,e^{(-1+i)t} + \frac{-1-2i}{10}\,e^{(-1-i)t}.$$

The last two terms are complex conjugates of each other, so their sum is twice the real part of each, namely,

$$2\,\frac{e^{-t}}{10}\,\text{Re}\,((-1+2i)(\cos t + i\sin t)) = \frac{e^{-t}}{5}\,(-\cos t - 2\sin t).$$

We wind up with the same function $f(t)$ as before.

List of properties of the Laplace transform

1. $\mathcal{L}$ is linear: $af(t) + bg(t) \rightsquigarrow aF(s) + bG(s)$.

2. $F(s)$ essentially determines $f(t)$.

3. s-shift theorem: If $f(t) \rightsquigarrow F(s)$, then $e^{at}f(t) \rightsquigarrow F(s-a)$.

4. t-shift theorem: If $f(t) \rightsquigarrow F(s)$, then $f_a(t) \rightsquigarrow e^{-as}F(s)$, where

$$f_a(t) = \begin{cases} f(t-a) & \text{if } t > a \\ 0 & \text{if } t < a \end{cases}.$$

5. s-derivative theorem: If $f(t) \rightsquigarrow F(s)$, then $tf(t) \rightsquigarrow -F'(s)$.

6. t-derivative theorem: If $f(t) \rightsquigarrow F(s)$, then $f'(t) \rightsquigarrow sF(s) - f(0+)$ where $f(t)$ is continuous for $t > 0$ and the notation $f'(t)$ indicates the ordinary derivative of $f(t)$.

7. If $f(t) \rightsquigarrow F(s)$ and $g(t) \rightsquigarrow G(s)$, then $f(t) * g(t) \rightsquigarrow F(s)G(s)$.

8. $\delta(t) \rightsquigarrow 1$.

21. The Laplace transform and generalized functions

21.1. **Laplace transform of impulse and step responses.** Laplace transform affords a way to solve LTI IVPs. If the ODE is

$$p(D)x = f(t)\,,$$

application of the Laplace transform results in an equation of the form

$$p(s)X = F(s) + G(s)$$

where $G(s)$ is computed from the initial conditions. Rest initial conditions lead to $G(s) = 0$, so in that case

$$X = W(s)F(s)$$

where $W(s) = 1/p(s)$ is the **transfer function** of the operator.

The very simplest case of this is when $f(t) = \delta(t)$. Then we are speaking of the unit impulse response $w(t)$, and we see that

The Laplace transform of the unit impulse response $w(t)$ is the transfer function $W(s)$.

This is an efficient way to compute the unit impulse response.

The next simplest case is when $f(t) = u(t)$, the unit step function. Its Laplace transform is $1/s$, so the unit step response $w_1(t)$ is the inverse Laplace transform of

$$W_1(s) = \frac{W(s)}{s} = \frac{1}{sp(s)}\,.$$

By way of example, suppose the operator is D^2+2D+2. The transfer function is $W(s) = 1/(s^2 + 2s + 2) = 1/((s + 1)^2 + 1)$. By the s shift rule and the tables,

$$w(t) = u(t)e^{-t}\sin t\,.$$

The Laplace transform of the unit step response is $W_1(s) = 1/s(s^2 + 2s + 2)$, which we can handle using complex cover up: write

$$\frac{1}{s((s+1)^2+1)} = \frac{a}{s} + \frac{b(s+1)+c}{(s+1)^2+1}\,.$$

Multiply through by s and set $s = 0$ to see $a = 1/2$. Then multiply through by $(s+1)^2+1$ and set $s = -1+i$ to see $bi+c = 1/(-1+i) = (-1-i)/2$, or $b = c = -1/2$: so

$$W_1(s) = \frac{1}{2}\left(\frac{1}{s} - \frac{(s+1)+1}{(s+1)^2+1}\right)\,.$$

Thus the unit step response is

$$w_1(t) = \frac{u(t)}{2}(1 - e^{-t}(\cos t + \sin t)) .$$

21.2. **What the Laplace transform doesn't tell us.** What do we mean, in the list of properties at the end of Section 20, when we say that $F(s)$ "essentially determines" $f(t)$?

The Laplace transform is defined by means of an integral. We don't need complete information about a function to determine its integral, so knowing its integral or integrals of products of it with exponentials won't be enough to completely determine it.

For example, if we can integrate a function $g(t)$ then we can also integrate any function which agrees with $g(t)$ except at one value of t, or even except at a finite number of values, and the integral of the new function is the same as the integral of g. Changing a few values doesn't change the "area under the graph."

Thus if $f(t) \rightsquigarrow F(s)$, and $g(t)$ coincides with $f(t)$ except at a few values of t, then also $g(t) \rightsquigarrow F(s)$. We can't hope to recover every value of $f(t)$ from $F(s)$ unless we put some side conditions on $f(t)$, such as requiring that it should be continuous.

Therefore, in working with functions via Laplace transform, when we talk about a function $f(t)$ it is often not meaningful to speak of the value of f at any specific point $t = a$. It does make sense to talk about $f(a-)$ and $f(a+)$, however. Recall that these are defined as

$$f(a-) = \lim_{t \uparrow a} f(t), \qquad f(a+) = \lim_{t \downarrow a} f(t).$$

This means that $f(a-)$ is the limiting value of $f(t)$ as t increases towards a from below, and $f(a+)$ is the limiting value of $f(t)$ as t decreases towards a from above. In both cases, the limit polls infinitely many values of f near a, and isn't changed by altering any finite number of them or by altering $f(a)$ itself; in fact f does not even need to be defined at a for us to be speak of $f(a\pm)$. The best policy is to speak of $f(a)$ only in case both $f(a-)$ and $f(a+)$ are defined and are equal to each other. In this case we can define $f(a)$ to be this common value, and then $f(t)$ is continuous at $t = a$.

The uniqueness theorem for the inverse Laplace transform asserts that if f and g have the same Laplace transform, then $f(a-) = g(a-)$ and $f(a+) = g(a+)$ for all a. If $f(t)$ and $g(t)$ are both continuous at a, so that $f(a-) = f(a+) = f(a)$ and $g(a-) = g(a+) = g(a)$, then it follows that $f(a) = g(a)$.

Part of the strength of the theory of the Laplace transform is its ability to deal smoothly with things like the delta function. In fact, we can form the Laplace transform of a generalized function as described in Section 17, assuming that it is of exponential type. The Laplace transform $F(s)$ determines the singular part of $f(t)$: if $F(s) = G(s)$ then $f_s(t) = g_s(t)$.

21.3. **Worrying about $t = 0$.** When we consider the Laplace transform of $f(t)$ in this course, we always make the assumption that $f(t) = 0$ for $t < 0$. Thus

$$f(0-) = 0.$$

What happens at $t = 0$ needs special attention, and the definition of the Laplace transform offered in Edwards and Penney (and most other ODE textbooks) is not consistent with the properties they assert.

They define

$$F(s) = \int_0^\infty e^{-st} f(t)\, dt.$$

Suppose we let $f(t) = \delta(t)$, so that $F(s)$ should be the constant function with value 1. The integrand is $\delta(t)$ again, since $e^{-st}|_{t=0} = 1$. The indefinite integral of $\delta(t)$ is the step function $u(t)$, which does not have a well-defined value at $t = 0$. Thus the value they assign to the the definite integral with lower limit 0 is ambiguous. We want the answer to be 1. This indicates that we should really define the Laplace transform of $f(t)$ as the integral

$$\boxed{F(s) = \int_{0-}^\infty e^{-st} f(t)\, dt.} \tag{1}$$

The integral is defined by taking the limit as the lower limit increases to zero from below. It coincides with the integral with lower limit $-\infty$ since $f(t) = 0$ for $t < 0$:

$$F(s) = \int_{-\infty}^\infty e^{-st} f(t)\, dt.$$

With this definition, $\delta(t) \rightsquigarrow 1$, as desired.

Let's now check some basic properties of the Laplace transform, using this definition.

21.4. **The t-derivative rule.** Integration by parts gives us

$$\int_{0-}^\infty e^{-st} f'(t)\, dt = f(0-) - (-s) \int_{0-}^\infty e^{-st} f(t)\, dt.$$

Since $f(0-) = 0$, we find that if $f(t) \rightsquigarrow F(s)$ then

$$f'(t) \rightsquigarrow sF(s). \tag{2}$$

Here I am using the **generalized derivative** *described in Section 17.* Including the delta terms in the derivative is important even if $f(t)$ is continuous for $t > 0$ because the jump from $f(0-) = 0$ to $f(0+)$ contributes $f(0+)\delta(t)$ to $f'(t)$. Let's assume that $f(t)$ is continuous for $t > 0$. Then $f'(t) = f(0+)\delta(t) + f_r'(t)$, where $f_r'(t)$ is the regular part of the generalized derivative, that is, the ordinary derivative of the function $f(t)$. Substituting this into (2) and using $\delta(t) \rightsquigarrow 1$ and linearity gives us the familiar formula

$$f_r'(t) \rightsquigarrow sF(s) - f(0+). \tag{3}$$

Formula (16) on p. 281 of Edwards and Penney is an awkward formulation of (2).

21.5. **The initial singularity formula.** If $f(t)$ is a generalized function with singular part at zero given by $b\delta(t)$, then

$$\lim_{s \to \infty \cdot 1} F(s) = b.$$

The notation means that we look at values of $F(s)$ for large *real* values of s.

To see this, break $f(t)$ into regular and singular parts. We have a standing assumption that the regular part $f_r(t)$ is of exponential order, and we know (from Edwards and Penney, formula (25) on p. 271 for example) that its Laplace transform dies off as $s \to \infty \cdot 1$.

Each delta function $\delta(t-a)$ in $f(t)$ contributes a term e^{-as} to $F(s)$, and as long as $a > 0$, these all decay to zero as $s \to \infty \cdot 1$ as well. Only $a = 0$ is left, and we know that $b\delta(t) \rightsquigarrow b$. This finishes the proof.

When $b = 0$—that is, when $f(t)$ is nonsingular at $t = 0$—the result is that

$$\lim_{s \to \infty \cdot 1} F(s) = 0.$$

21.6. **The initial value formula.** If $f(t)$ is a piecewise differentiable generalized function, then

$$\lim_{s \to \infty \cdot 1} sF(s) = f(0+).$$

To see this, let $f'(t)$ again denote the generalized derivative. The jump from $f(0-) = 0$ to $f(0+)$ contributes the term $f(0+)\delta(t)$ to $f'(t)$.

The initial value formula results directly from the initial singularity formula applied to $f'(t)$.

For example, if $f(t) = \cos t$ then $F(s) = \dfrac{s}{s^2+1}$, and

$$\lim_{s\to\infty} \frac{s^2}{s^2+1} = 1$$

which also happens to be $\cos(0)$. With $f(t) = \sin t$, on the other hand, $F(s) = \dfrac{1}{s^2+1}$, and $\lim\limits_{s\to\infty} \dfrac{s}{s^2+1} = 0$ in agreement with $\sin 0 = 0$.

21.7. **Initial conditions.** Let's return to the first order bank account model, $\dot{x}+px = q(t)$. When we come to specify initial conditions, at $t = 0$, the following two procedures are clearly equivalent: (1) Fix $x(0+) = x_0$ and proceed; and (2) Fix $x(0-) = 0$—*rest* initial conditions—but at $t = 0$ deposit a lump sum of x_0 dollars. The second can be modeled by altering the signal, replacing the rate of deposit $q(t)$ by $q(t) + x_0\delta(t)$. Thus if you are willing to accept a generalized function for a signal, you can get always away with rest initial conditions.

Let's see how this works out in terms of the Laplace transform. In the first scenario, the t-derivative theorem gives us

$$(sX - x_0) + pX = Q.$$

In the second scenario, the fact that $\delta(t) \rightsquigarrow 1$ gives us

$$sX + pX = Q + x_0,$$

an equivalent expression.

This example points out how one can absorb certain non-rest initial conditions into the signal. The general picture is that the *top* derivative you specify as part of the initial data can be imposed using a delta function.

The mechanical model of the second degree case helps us understand this. We have an equation

$$m\ddot{x} + b\dot{x} + kx = q(t).$$

Suppose the initial *position* is $x(0) = 0$. Again, we have two alternatives: (1) Fix the initial velocity at $\dot{x}(0+) = v_0$ and proceed; or (2) Fix $\dot{x}(0-) = 0$—so, together with $x(0-) = 0$ we have rest initial conditions—and then at $t = 0$ give the system an impulse, by adding to the signal the term $mv_0\delta(t)$. The factor of m is necessary, because we want to use this impulse to jack the velocity up to the value v_0, and the force required to do this will depend upon m.

Exercise 21.7.1. As above, check that this equivalence is consistent with the Laplace transform.

For a higher order example, consider the LTI operator $L = (D^3 + 2D^2 - 2I)$ with transfer function $W(s) = 1/(s^3 + s^2 - 2)$. In the section on the Laplace Transform in the complex plane we computed the corresponding weight function: $w(t) = e^t/5 + (e^{-t}/5)\,(-\cos t - 2\sin t)$ (for $t > 0$). This is a solution to the ODE $Lx = \delta(t)$ with rest initial conditions. This is equivalent (for $t > 0$) to the IVP $Lx = 0$ with initial conditions $x(0) = \dot{x}(0) = 0$ and $\ddot{x}(0) = 1$, and indeed $w(t)$ satisfies all this.

In order to capture initial conditions of lower derivatives using impulses and such in the signal, one must consider not just the delta function but also its derivatives. The desire to do this is a good motivation for extending the notion of generalized functions, but not something we will pursue in this course.

22. The pole diagram and the Laplace transform

When working with the Laplace transform, it is best to think of the variable s in $F(s)$ as ranging over the *complex* numbers. In the first section below we will discuss a way of visualizing at least some aspects of such a function—via the "pole diagram." Next we'll describe what the pole diagram of $F(s)$ tells us—and what it does not tell us—about the original function $f(t)$. In the third section we discuss the properties of the integral defining the Laplace transform, allowing s to be complex. The last section describes the Laplace transform of a periodic function of t, and its pole diagram, linking the Laplace transform to Fourier series.

22.1. **Poles and the pole diagram.** The real power of the Laplace transform is not so much as an algorithm for explicitly computing linear time-invariant system responses as in gaining insight into these responses *without* explicitly computing them. (A further feature of the Laplace transform is that it allows one to analyze systems which are not modeled by ODEs at all, by exactly the same methodology.) To achieve this insight we will have to regard the transform variable s as *complex*, and the transform function $F(s)$ as a complex-valued function of a complex variable.

A simple example is $F(s) = 1/(s-z)$, for a fixed complex number z. We can get some insight into a complex-valued function of a complex variable, such as $1/(s-z)$, by thinking about its absolute value: $|1/(s-z)| = 1/|s-z|$. This is now a *real-valued* function on the complex plane, and its graph is a surface lying over the plane, whose height over a point s is given by the value $|1/(s-z)|$. This is a tent-like surface lying over the complex plane, with elevation given by the reciprocal of the distance to z. It sweeps up to infinity like a hyperbola as s approaches z; it's as if it is being held up at $s = z$ by a tent-pole, and perhaps this is why we say that $1/(s-z)$ "has a pole at $s = z$." Generally, a function of complex numbers has a "pole" at $s = z$ when it becomes infinite there.

$F(s) = 1/(s-z)$ is an example of a **rational function**: a quotient of one polynomial by another. The Laplace transforms of many important functions are rational functions, and we will start by discussing rational functions.

A product of two rational functions is again a rational function. Because you can use a common denominator, a sum of two rational functions is also a rational function. The reciprocal of any rational function except the zero function is again a rational function—exchange

numerator and denominator. In these algebraic respects, the collection of rational functions behaves like the set of rational *numbers*. Also like rational numbers, you can simplify the fraction by cancelling terms in numerator and denominator, till the two don't have any common factors. (In the case of rational numbers, you do have to allow ± 1 as a common factor! In the case of rational functions, you do have to allow nonzero contants as common factors.)

When written in reduced form, the magnitude of $F(s)$ blows up to ∞ as s approaches a root of the denominator. The complex roots of the denominator are the **poles** of $F(s)$.

In case the denominator doesn't have any repeated roots, partial fractions let you write $F(s)$ as

$$F(s) = p(s) + \frac{w_1}{s - z_1} + \cdots + \frac{w_n}{s - z_n} \tag{1}$$

where $p(s)$ is a polynomial, $z_1, \ldots z_n$ are complex constants, and $w_1, \ldots, w_n$ are nonzero complex constants.

For example, the calculation done in Section 20.5 shows that the poles of $F(s) = 1/(s^3 + s^2 - 2)$ are at $s = 1$, $s = -1 + i$, and $s = -1 - i$.

The **pole diagram** of a complex function $F(s)$ is just the complex plane with the poles of $F(s)$ marked on it. Figure 15 shows the pole diagram of the function $F(s) = 1/(s^3 + s^2 - 2)$.

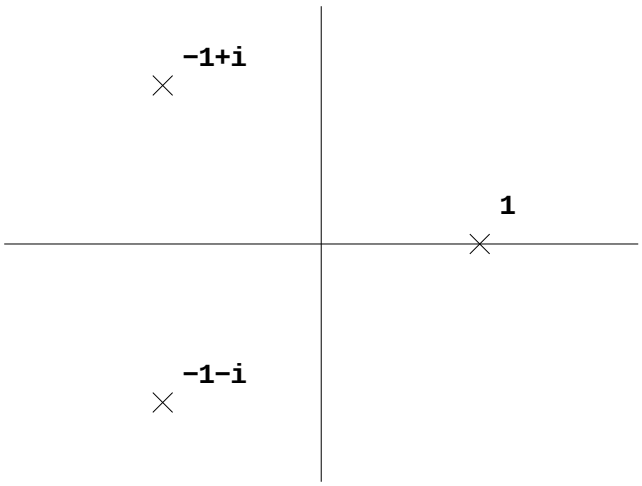


FIGURE 15. Pole diagram for $1/(s^3 + s^2 - 2)$

The constant w_k appearing in (1) is the **residue** of the pole at $s = z_k$. The calculation in 20.5 shows that the residue at $s = 1$ is $1/5$, the

residue at $s = -1 + 2i$ is $(-1 + 2i)/10$, and the residue at $s = -1 - 2i$ is $(-1 - 2i)/10$.

Laplace transforms are not always rational functions. For example, the exponential function occurs: $F(s) = e^{ws}$, for w a complex constant. The exponential function has *no poles*: it takes on well defined complex values for any complex input s.

We can form more elaborate complex functions by taking products—$e^{-s}/(s^3 + s^2 - 2)$, for example. The numerator doesn't contribute any poles. Nor does it kill any poles—it is never zero, so it doesn't cancel any of the roots of the denominator. The pole diagram of this function is the same as the pole diagram of $1/(s^3 + s^2 - 2)$.

A general complex function of the type that occurs as a Laplace transform (the mathematical term is *meromorphic*) does not have a partial fraction decomposition, so we can't use (1) to locate the poles. Poles occur where the value of the function blows up. This can be expressed as follows. Define the **residue** of $F(s)$ at $s = z$ as

$$\text{res}_{s=z} F(s) = \lim_{s \to z} (s - z) F(s). \tag{2}$$

If $F(s)$ does not have a pole at $s = z$, then

$$\text{res}_{s=z} F(s) = 0\,.$$

A complex function is by no means completely specified by its pole diagram. Nevertheless, the pole diagram of $F(s)$ carries a lot of information about $F(s)$, and if $F(s)$ is the Laplace transform of $f(t)$, it tells you a lot of information of a specific type about $f(t)$.

22.2. **The pole diagram of the Laplace transform.**

Summary: The pole diagram of $F(s)$ tells us a lot about *long-term behavior* of $f(t)$. It tells us *nothing* about the near-term behavior.

This is best seen by examples.

Suppose we have just one pole, at $s = 1$. Among the functions with this pole diagram we have:

$$F(s) = \frac{c}{s-1}\,, \quad G(s) = \frac{ce^{-as}}{s-1}\,, \quad H(s) = \frac{c}{s-1} + b\frac{1 - e^{-as}}{s}$$

where $c \neq 0$. (Note that $1 - e^{-as}$ becomes zero when $s = 0$, canceling the zero in the denominator of the second term in $H(s)$.) To be Laplace transforms of real functions we must also assume them all to be real,

and $a \geq 0$. Then these are the Laplace transforms of

$$f(s) = ce^t\,, \quad g(t) = \begin{cases} ce^{t-a} & \text{for } t > a, \\ 0 & \text{for } t < a \end{cases}\,, \quad h(t) = \begin{cases} ce^t & \text{for } t > a, \\ ce^t + b & \text{for } t < a \end{cases}$$

All these functions grow like a multiple of e^t when t is large. You can even say which multiple: it is given by the residue at $s = 1$. (Note that $g(t) = (ce^{-a})e^t$, and the residue of $G(s)$ at $s = 1$ is ce^{-a}.) But their behavior when $t < a$ is all over the map. In fact, the function can be *anything* for $t < a$, for *any* fixed a; as long as it settles down to something close to ce^t for t large, its Laplace transform will have just one pole, at $s = 1$, with residue c.

Now suppose we have two poles, say at $s = a + bi$ and $s = a - bi$. Two functions with this pole diagram are

$$F(s) = \frac{c(s-a)}{(s-a)^2 + b^2}\,, \quad G(s) = \frac{cb}{(s-a)^2 + b^2}.$$

and we can modify these as above to find others. These are the Laplace transform of

$$f(t) = ce^{at}\cos(bt)\,, \quad g(t) = ce^{at}\sin(bt).$$

This reveals that it is the *real part* of the pole that determines the long term *growth* of absolute value; if the function oscillates, this means growth of maxima and minima. The *imaginary part* of the pole determines the *circular frequency of oscillation* for large t. We can't pick out the phase from the pole diagram alone (but the residues do determine the phase). And we can't promise that it will be exactly sinusoidal times exponential, but it will resemble this. And again, the pole diagram of $F(s)$ says *nothing* about $f(t)$ for small t.

Now let's combine several of these, to get a function with several poles. Suppose $F(s)$ has poles at $s = 1$, $s = -1 + i$, and $s = -1 - i$, for example. We should expect that $f(t)$ has a term which grows like e^t (from the pole at $s = 1$), and another term which behaves like $e^{-t}\cos t$ (up to constants and phase shifts). When t is large, the damped oscillation becomes hard to detect as the other term grows exponentially.

We learn that the *rightmost poles dominate*—the ones with *largest real part* have the dominant influence on the long-term behavior of $f(t)$.

The most important consequence relates to the question of *stability*:

If all the poles of $F(s)$ have *negative real part* then $f(t)$ decays exponentially to zero as $t \to \infty$.

If some pole has positive real part, then $|f(t)|$ becomes arbitrarily large for large t.

If there are poles on the imaginary axis, and no poles to the right, then the function $f(t)$ may grow (e.g. $f(t) = t$ has $F(s) = 1/s^2$, which has a pole at $s = 0$), but only "sub-exponentially": for any $a > 0$ there is a constant c such that $|f(t)| < ce^{at}$ for all $t > 0$.

Comment on reality. We have happily taken the Laplace transform of complex valued functions of t: $e^{it} \rightsquigarrow 1/(s - i)$, for example. If $f(t)$ is real, however, then $F(s)$ enjoys a symmetry with respect to complex conjugation:

(3) $$\boxed{\text{If } f(t) \text{ is real-valued then } F(\overline{s}) = \overline{F(s)}.}$$

The pole diagram of a function $F(s)$ such that $F(\overline{s}) = \overline{F(s)}$ is *symmetric about the real axis*: non-real poles occur in complex conjugate pairs. In particular, the pole diagram of the Laplace transform of a real function is symmetric across the real axis.

22.3. **The Laplace transform integral.** In the integral defining the Laplace transform, we really should let s be complex. We are thus integrating a complex-valued function of a real parameter t, $e^{-st}f(t)$, and this is done by integrating the real and imaginary parts separately. It is an improper integral, computed as the limit of $\displaystyle\int_0^T e^{-st}f(t)\,dt$ as $T \to \infty$. (Actually, we will see in Section 21 that it's better to think of the lower limit as "improper" as well, in the sense that we form the integral with lower limit $a < 0$ and then let $a \uparrow 0$.) The textbook assumption that $f(t)$ is of "exponential order" is designed so that if s has large enough real part, the term e^{-st} will be so small (at least for large t) that the product $e^{-st}f(t)$ has an integral which converges as $T \to \infty$. In terms of the pole diagram, we may say that the integral converges when the real part of s is bigger than the real part of any pole in the resulting transform function $F(s)$. The exponential order assumption is designed to guarantee that we won't get poles with arbitrarily large real part.

The region to the right of the rightmost pole is called the **region of convergence**. Engineers abbreviate this and call it the "ROC."

Once the integral has been computed, the expression in terms of s will have meaning for all complex numbers s (though it may have a pole at some).

For example, let's consider the time-function $f(t) = 1$, $t > 0$. Then:

$$F(s) = \int_0^\infty e^{-st} dt = \lim_{T\to\infty} \frac{e^{-st}}{-s}\bigg|_0^T = \frac{1}{-s}\left(\lim_{T\to\infty} e^{-sT} - 1\right).$$

Since $\left|e^{-sT}\right| = e^{-aT}$ if $s = a + bi$, the limit is 0 if $a > 0$ and doesn't exist if $a < 0$. If $a = 0$, $e^{-sT} = \cos(bT) - i\sin(bT)$, which does not have a limit as $T \to \infty$ unless $b = 0$ (which case is not relevant to us since we certainly must have $s \neq 0$). Thus the improper integral converges exactly when $\mathrm{Re}\,(s) > 0$, and gives $F(s) = 1/s$. Despite the fact that the integral definitely diverges for $\mathrm{Re}\,(s) \leq 0$, the expression $1/s$ makes sense for all $s \in \mathbb{C}$ (except for $s = 0$), and it's better to think of the function $F(s)$ as defined everywhere in this way. This process is called **analytic continuation**.

22.4. **Laplace transform and Fourier series.** We now have two ways to study periodic functions $f(t)$. First, we can form the Laplace transform $F(s)$ of $f(t)$ (regarded as defined only for $t > 0$). Since $f(t)$ is periodic, the poles of $F(s)$ lie entirely along the imaginary axis, and the locations of these poles reveal sinusoidal constituents in $f(t)$, in some sense. On the other hand, $f(t)$ has a Fourier series, which explicitly expresses it as a sum of sinusoidal components. What is the relation between these two perspectives?

For example, the standard square wave $\mathrm{sq}(t)$ of period 2π, with value 1 for $0 < t < \pi$ and -1 for $-\pi < t < 0$, restricted to $t > 0$, can be written as

$$\mathrm{sq}(t) = 2(u(t) - u(t-\pi) + u(t-2\pi) - u(t-3\pi) + \cdots) - u(t)$$

By the t-shift formula and $u(t) \rightsquigarrow 1/s$,

$$\mathrm{Sq}(s) = \frac{1}{s}\left(2(1 - e^{-\pi s} + e^{-2\pi s} - \cdots) - 1\right) = \frac{1}{s}\left(\frac{2}{1+e^{-\pi s}} - 1\right)$$

The denominator vanishes when $e^{-\pi s} = -1$, and this happens exactly when $s = ki$ where k is an odd integer. So the poles of $\mathrm{Sq}(s)$ are at 0 and the points ki where k runs through odd integers. $s = 0$ does not occur as a pole, because the expression $\dfrac{2}{1+e^{-\pi s}} - 1$ vanishes when $s = 0$ and cancels the $1/s$.

On the other hand, the Fourier series for the square wave is

$$\mathrm{sq}(t) = \frac{4}{\pi}\left(\sin(t) + \frac{\sin(3t)}{3} + \frac{\sin(5t)}{5} + \cdots\right).$$

If we express this as a series of complex exponentials, following 16.5, we find that c_k is nonzero for k an odd integer, positive or negative. There must be a relation!

It is easy to see the connection in general, especially if we use the complex form of the Fourier series,

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}.$$

Simply apply the Laplace transform to this expression, using $e^{int} \rightsquigarrow \dfrac{1}{s-in}$:

$$F(s) = \sum_{n=-\infty}^{\infty} \frac{c_n}{s-in}$$

The only possible poles are at the complex numbers $s = in$, and the residue at in is c_n.

If $f(t)$ is periodic of period 2π, the poles of $F(s)$ occur only at points of the form $n\pi i$ for n an integer, and the residue at $s = n\pi i$ is precisely the complex Fourier coefficients c_n of $f(t)$.

23. Amplitude response and the pole diagram

We have seen in Section 10 that the analysis of the system response of an LTI operator to a sinusoidal signal generalizes to the case of a signal of the form $Ae^{at}\cos(\omega t - \phi)$. When $\phi = 0$, one considers an associated complex equation, with input signal given by $Ae^{(a+i\omega)t}$, and applies the Exponential Response Formula.

The Mathlet `Amplitude Response and Pole Diagram` illustrates this situation.

We will consider again the behavior of the spring/mass/dashpot system, in its complex form

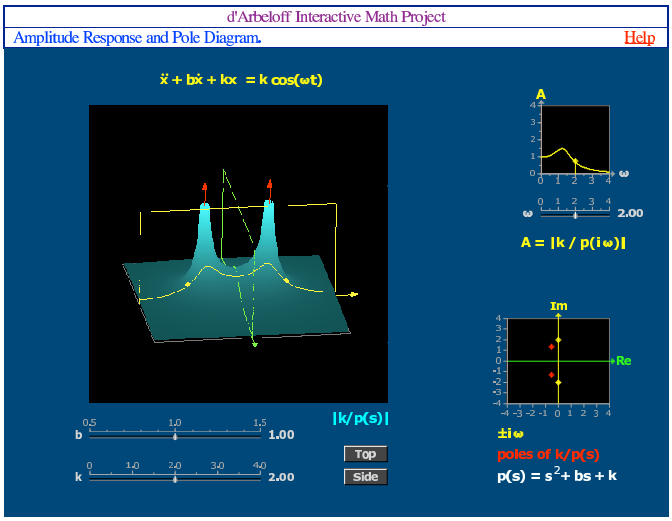
$$m\ddot{z} + b\dot{z} + kz = kAe^{st}$$

where s is a complex constant. The Exponential Response Formula gives the exponential solution

$$z_p = AW(s)e^{st} \tag{1}$$

where $W(s)$ is the "transfer function"

$$W(s) = \frac{k}{p(s)}$$



(Please forgive the missing ω's in this screen capture!)

The input signal gets multiplied by the complex number $W(s)$. This number has a magnitude and an argument. We will focus entirely on

the magnitude. When it is large, the system response is that much larger than the input signal. When it is small, the system response is that much smaller. The real number $|W(s)|$ is called the **gain**.

Now imagine the graph of $|W(s)|$. It is a tent-like surface, suspended over the complex plane. There are places where $|W(s)|$ becomes infinite—at the roots of the characteristic polynomial $ms^2 + bs + k$. These are called **poles** of $W(s)$. The graph flares up to infinite height above the poles of $W(s)$, which may be why they are called poles! The altitude above $s = a+i\omega$ has an interpretation as the "gain" of the system when responding to an input signal of the form $e^{at}\cos(\omega t)$. When s coincides with a pole, the system is in resonance; there is no solution of the form $gAe^{at}\cos(\omega t - \phi)$, but rather one of the form t times that expression.

Near to the poles, the gain is large, and it falls off as s moves away from the poles.

The case of sinusoidal input occurs when s is on the imaginary axis. Imagine wall rising from the imaginary axis. It intersects the graph of $|W(s)|$ in a curve. That curve represents $|W(i\omega)|$ as ω varies over real numbers. This is precisely the "Bode plot," the amplitude frequency response curve.

The `Amplitude Response and Pole Diagram` Mathlet shows this well. (The program chooses $m = 1$.) The left hand window is a 3-D display of the graph of $|W(s)|$. Moving the cursor over that window will reorient the picture. At lower right is the pole diagram of $W(s)$, and above it is the amplitude response curve.

You can see the geometric origin of near resonance: what is happening is that the part of the graph of $|W(s)|$ lying over the imaginary axis moves up over the shoulder of the "volcano" surrounding one of the poles of $W(s)$.

24. The Laplace transform and more general systems

This section gives a hint of how flexible a device the Laplace transform is in engineering applications.

24.1. **Zeros of the Laplace transform: stillness in motion.** The mathematical theory of functions of a complex variable shows that the *zeros* of $F(s)$—the values r of s for which $F(r) = 0$—are just as important to our understanding of it as are the poles. This symmetry is reflected in engineering as well; the location of the zeros of the transfer function has just as much significance as the location of the poles. Instead of recording resonance, they reflect stillness.

We envision the following double spring system: there is an object with mass m_1 suspended by a spring with spring constant k_1. A second object with mass m_2 is suspended from this first object by a second spring with constant k_2. The system is driven by motion of the top of the top spring according to a function $f(t)$. Pick coordinates so that x_1 is the position of the first object and x_2 is the position of the second, both increasing in the downward direction, and such that when $f(t) = x_1 = x_2 = 0$ the springs exert no force.

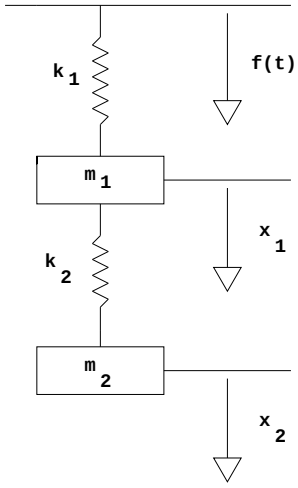


Figure 16. Two spring system

The equations of motion are

$$\text{(1)} \qquad \begin{cases} m_1\ddot{x}_1 &= k_1(f(t) - x_1) - k_2(x_1 - x_2) \\ m_2\ddot{x}_2 &= k_2(x_1 - x_2) \end{cases}$$

This is a *system of second order equations*, and as you can imagine mechanical engineering is full of similar systems.

Suppose that our main interest is in x_1. Let's take Laplace transforms, and assume rest initial conditions.

$$\begin{cases} (m_1 s^2 + (k_1 + k_2))X_1 &= k_2 X_2 + k_1 F \\ (m_2 s^2 + k_2)X_2 &= k_2 X_1. \end{cases}$$

Use the second equation to express X_2 in terms of X_1, and substitute this value into the first equation. Then solve for X_1 to get:

$$X_1(s) = \frac{m_2 s^2 + k_2}{(m_1 s^2 + (k_1 + k_2))(m_2 s^2 + k_2) - k_2^2} \cdot k_1 F(s).$$

The "transfer function" $W(s)$ is then the ratio of the LT of the system response, X_1, and the LT of the input signal, F:

$$W(s) = \frac{k_1(m_2 s^2 + k_2)}{(m_1 s^2 + (k_1 + k_2))(m_2 s^2 + k_2) - k_2^2}.$$

It is still the case that $W(r)$ is the multiple of e^{rt} which occurs as x_1 in a solution to the equations (1) when we take $f(t) = e^{rt}$. Thus the zeros of $W(s)$ at $s = \pm i\sqrt{k_2/m_2}$—the values of s for which $W(s) = 0$—reflect a "neutralizing" circular frequency of $\omega = \sqrt{k_2/m_2}$. If $f(t)$ is sinusoidal of this circular frequency then $x_1 = 0$ is a solution. The suspended weight oscillates with (k_1/k_2) times the amplitude of $f(t)$ and reversed in phase (independent of the masses!), and exactly cancels the impressed force. Check it out!

24.2. **General LTI systems.** The weight function $w(t)$, or its Laplace transform, the transfer function $W(s)$, completely determine the system. The transfer function of an ODE has a very restricted form—it is the reciprocal of a polynomial; but the mechanism for determining the system response makes sense for much more general complex functions $W(t)$, and, correspondingly, much more general "weight functions" $w(t)$: given a very general function $w(t)$, we can define an LTI system by declaring that a signal $f(t)$ results in a system response (with null initial condition, though in fact nontrivial initial conditions can be handled too, by absorbing them into the signal using delta functions) given by the convolution $f(t) * w(t)$. The apparatus of the Laplace transform helps us, too, since we can compute this system response as the inverse Laplace transform of $F(s)W(s)$. This mechanism allows us to represent the *system*, the *signal*, and the *system response*, all three, using *functions* (of t, or of s). Differential operators have vanished from the scene. This flexibility results in a tool of tremendous power.

25. First order systems and second order equations

25.1. **The companion system.** One of the main reasons we study first order systems is that a differential equation of any order may be replaced by an equivalent first order system. Computer ODE solvers use this principle.

To illustrate, suppose we start with a second order homogeneous LTI system,

$$\ddot{x} + b\dot{x} + cx = 0.$$

The way to replace this by a *first order* system is to introduce a new variable, say y, related to x by

$$y = \dot{x}.$$

Now we can replace $\ddot{x}$ by $\dot{y}$ in the original equation, and find a *system*:

$$\begin{cases} \dot{x} &= y \\ \dot{y} &= -cx - by \end{cases} \tag{1}$$

The solution $x(t)$ of the original equation appears as the top entry in the vector-valued solution of this system.

This process works for *any* higher order equation, linear or not, provided we can express the top derivative as a function of the lower ones (and t). An nth order equation gives rise to a first order system in n variables.

The trajectories of this system represent in very explicit form many aspects of the time evolution of the original equation. You no longer have time represented by an axis, but you see the effect of time quite vividly, since the vertical direction, y, records the velocity, $y = \dot{x}$. A stable spiral, for example, reflects damped oscillation. (See the Mathlet `Damped Vibrations` for a clear visualization of this.)

The matrix for the system (1),

$$\begin{bmatrix} 0 & 1 \\ -c & -b \end{bmatrix},$$

is called the *companion matrix*. These matrices constitute quite a wide range of 2×2 matrices, but they do have some special features. For example, if a companion matrix has a repeated eigenvalue then it is necessarily incomplete, since a companion matrix can never be a multiple of the identity matrix.

This association explains an apparent conflict of language: we speak of the characteristic polynomial of a second order equation—in the case

at hand it is $p(s) = s^2 + bs + c$. But we also speak of the characteristic polynomial of a matrix. Luckily (and obviously)

The characteristic polynomial of a second order LTI operator is the same as the characteristic polynomial of the companion matrix.

25.2. **Initial value problems.** Let's do an example to illustrate this process, and see how initial values get handled. We will also use this example to illustrate some useful ideas and tricks about handling linear systems.

Suppose the second order equation is

$$\ddot{x} + 3\dot{x} + 2x = 0. \tag{2}$$

The companion matrix is

$$A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix},$$

so solutions to (2) are the top entries of the solutions to $\dot{\mathbf{x}} = A\mathbf{x}$.

An initial value for (2) gives us values for both x and $\dot{x}$ at some initial time, say $t = 0$. Luckily, this is exactly the data we need for an initial value for the matrix equation $\dot{\mathbf{x}} = A\mathbf{x}$: $\mathbf{x}(0) = \begin{bmatrix} x(0) \\ \dot{x}(0) \end{bmatrix}$.

Let's solve the system first, by finding the exponential e^{At}. The eigenvalues of A are the roots of the characteristic polynomial, namely $\lambda_1 = -1, \lambda_2 = -2$. (From this we know that there are two normal modes, one with an exponential decay like e^{-t}, and the other with a much faster decay like e^{-2t}. The phase portrait is a stable node.)

To find an eigenvector for λ_1, we must find a vector α_1 such that $(A - \lambda_1 I)\alpha_1 = 0$. Now

$$A - (-1)I = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}.$$

A convenient way to find a vector killed by this matrix is to take the entries from one of the rows, reverse their order and one of their signs: so for example $\alpha_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ will do nicely. The other row serves as a check; if it doesn't kill this vector then you have made a mistake somewhere. In this case it does.

Similarly, an eigenvector for λ_2 is $\alpha_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$.

The general solution is thus

$$\mathbf{x} = ae^{-t}\begin{bmatrix} 1 \\ -1 \end{bmatrix} + be^{-2t}\begin{bmatrix} 1 \\ -2 \end{bmatrix}. \tag{3}$$

From this expression we can get a good idea of what the phase portrait looks like. There are two eigendirections, containing straight line solutions. The line through $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ contains solutions decaying like e^{-t}. (Notice that this single line contains *infinitely many* solutions: for any point $\mathbf{x}_0$ on the line there is a solution $\mathbf{x}$ with $\mathbf{x}(0) = \mathbf{x}_0$. If $\mathbf{x}_0$ is the zero vector then this is the constant solution $\mathbf{x} = \mathbf{0}$.) The line through $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ contains solutions decaying like e^{-2t}.

The general solution is a linear combination of these two. Notice that as time grows, the coefficient of $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ varies like the *square* of the coefficient of $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. When time grows large, both coefficients become small, but the second becomes smaller much faster than the first. Thus the trajectory becomes asymptotic to the eigenline through $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. If you envision the node as a spider, the body of the spider is oriented along the eigenline through $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

A fundamental matrix is given by lining up the two normal modes as columns of a matrix:

$$\Phi = \begin{bmatrix} e^{-t} & e^{-2t} \\ -e^{-t} & -2e^{-2t} \end{bmatrix}.$$

Since each column is a solution, any fundamental matrix itself is a solution to $\dot{\mathbf{x}} = A\mathbf{x}$ in the sense that

$$\dot{\Phi} = A\Phi.$$

(Remember, premultiplying Φ by A multiplies the columns of Φ by A separately.)

The exponential matrix is obtained by normalizing Φ, i.e. by post-multiplying Φ by $\Phi(0)^{-1}$ so as to obtain a fundamental matrix which is the identity at $t = 0$. Since

$$\Phi(0) = [\alpha_1 \quad \alpha_2] = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix},$$

$$\Phi(0)^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

and so

$$e^{At} = \Phi(t)\Phi(0)^{-1} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix}.$$

Now any IVP for this ODE is easy to solve: $\mathbf{x} = e^{At}\mathbf{x}(0)$. For example, if $\mathbf{x}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then

$$\mathbf{x} = \begin{bmatrix} 3e^{-t} - 2e^{-2t} \\ -3e^{-t} + 4e^{-2t} \end{bmatrix}.$$

Now let's solve the original second order system, and see how the various elements of the solution match up with the work we just did. The key is always the fact that $y = \dot{x}$: $\mathbf{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$.

As observed, the characteristic polynomial of (2) is the same as that of A, so the eigenvalues of A are the roots, and we have two normal modes: e^{-t} and e^{-2t}. These are the exponential solutions to (2). The general solution is

$$x = ae^{-t} + be^{-2t}.$$

Note that (3) has this as top entry, and its derivative as bottom entry.

To solve general IVPs we would like to find the pair of solutions which is normalized at $t = 0$ as in Section 9. These are solutions x_1 and x_2 such that $\begin{bmatrix} x_1(0) \\ \dot{x}_1(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} x_2(0) \\ \dot{x}_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. This says exactly that we are looking for the columns of the normalized fundamental matrix e^{At}! Thus we can read off x_1 and x_2 from the top row of e^{At}:

$$x_1 = 2e^{-t} - e^{-2t}, \qquad x_2 = e^{-t} - e^{-2t}.$$

The bottom row of e^{At} is of course exactly the derivative of the top row.

The process of finding $\Phi(0)^{-1}$ is precisely the same as the process of finding the numbers a, b, c, d such that $x_1 = ae^{-t} + be^{-2t}$ and $x_2 = ce^{-t} + de^{-2t}$ form a normalized pair of solutions. If A is the companion matrix for a second order homogeneous LTI equation, then the entries in the top row of e^{At} constitute the pair of solutions of the original equation normalized at $t = 0$.

26. Phase portraits in two dimensions

This section presents a very condensed summary of the behavior of two dimensional linear systems, followed by a catalogue of linear phase portraits. A much richer understanding of this gallery can be achieved using the Mathlets `Linear Phase Portraits: Cursor Entry` and `Linear Phase Portraits: Matrix Entry`.

26.1. **Phase portraits and eigenvectors.** It is convenient to represent the solutions of an autonomous system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ (where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$) by means of a *phase portrait*. The x, y plane is called the *phase plane* (because a point in it represents the state or phase of a system). The phase portrait is a representative sampling of trajectories of the system. A *trajectory* is the directed path traced out by a solution. It does *not* include information about the time at which solutions pass through various points (which will depend upon when the clock was set), nor does it display the speed at which the solution passes through the point—only the direction of travel. Still, it conveys essential information about the qualitative behavior of solutions of the system of equations.

The building blocks for the phase portrait of a general system will be the phase portraits of *homogeneous linear constant coefficient* systems: $\dot{\mathbf{x}} = A\mathbf{x}$, where A is a constant square matrix. Notice that this equation *is* autonomous!

The phase portraits of these linear systems display a startling variety of shapes and behavior. We'll want names for them, and the names I'll use differ slightly from the names used in the book and in some other sources.

One thing that can be read off from the phase portrait is the stability properties of the system. A linear autonomous system is *unstable* if most of its solutions tend to infinity with time. (The meaning of "most" will become clearer below.) It is *asymptotically stable* if all of its solutions tend to 0 as t goes to ∞. Finally it is *neutrally stable* if none of its solutions tend to infinity with time but most of them do not tend to zero either. It is an interesting fact that any linear autonomous system exhibits one of these three behaviors.

The *characteristic polynomial* of a square matrix A is defined to be

$$p_A(s) = \det(A - sI).$$

If A is $n \times n$, this polynomial has the following form:

$$p_A(s) = (-s)^n + (\mathrm{tr}A)(-s)^{n-1} + \cdots + (\det A),$$

where the dots represent less familiar combinations of the entries of A. When A is 2×2, say $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, this reads

$$p_A(s) = s^2 - (\mathrm{tr}A)s + (\det A).$$

We remind the reader that in this case, when

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

$$\mathrm{tr}A = a + d\,, \quad \det A = ad - bc.$$

From the eigenvalues we may reconstruct $\mathrm{tr}A$ and $\det A$, since

$$p_A(s) = (s - \lambda_1)(s - \lambda_2) = s^2 - (\lambda_1 + \lambda_2)s + \lambda_1\lambda_2$$

implies

$$\mathrm{tr}A = \lambda_1 + \lambda_2, \qquad \det A = \lambda_1\lambda_2.$$

Thus giving the trace and the determinant is equivalent to giving the pair of eigenvalues.

Recall that the general solution to a system $\dot{\mathbf{x}} = A\mathbf{x}$ is usually of the form $c_1 e^{\lambda_1 t}\alpha_1 + c_2 e^{\lambda_2 t}\alpha_2$, where λ_1, λ_2 are the eigenvalues of the matrix A and α_1, α_2 are corresponding nonzero eigenvectors. The eigenvalues by themselves usually describe most of the gross structure of the phase portrait.

There are two caveats. First, this is not necessarily the case if the eigenvalues coincide. In two dimensions, when the eigenvalues coincide one of two things happens. (1) The *complete* case. Then $A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{bmatrix}$, *every* vector is an eigenvector (for the eigenvalue $\lambda_1 = \lambda_2$), and the general solution is $e^{\lambda_1 t}\alpha$ where α is any vector. (2) The *defective* case. (This covers all the other matrices with repeated eigenvalues, so if you discover your eigenvalues are repeated and you are not diagonal, then you are defective.) Then there is (up to multiple) only one eigenvector, α_1, and the general solution is $\mathbf{x} = e^{\lambda_1 t}(c_1\alpha_1 + c_2(t\alpha_1 + \beta))$, where β is a vector such that $(A - \lambda_1 I)\beta = \alpha_1$. (Such a vector β always exists in this situation, and is unique up to addition of a multiple of α_1.)

The second caveat is that the eigenvalues may be non-real. They will then form a complex conjugate pair. The eigenvectors will also be non-real, and if α_1 is an eigenvector for λ_1 then $\alpha_2 = \overline{\alpha_1}$ is an eigenvector

for $\lambda_2 = \overline{\lambda_1}$. Independent real solutions may be obtained by taking the real and imaginary parts of either $e^{\lambda_1 t}\alpha_1$ or $e^{\lambda_2 t}\alpha_2$. (These two have the same real parts and their imaginary parts differ only in sign.) This will give solutions of the general form e^{at} times a vector whose coordinates are linear combinations of $\cos(\omega t)$ and $\sin(\omega t)$, where the eigenvalues are $a \pm i\omega$.

Each of these caveats represents a failure of the eigenvalues by themselves to determine major aspects of the phase portrait. In the case of repeated eigenvalue, you get a defective node or a star node, depending upon whether you are in the defective case or the complete case. In the case of non-real eigenvalues you know you have a spiral (or a center, if the real part is zero); you know whether it is stable or unstable (look at the sign of the real part of the eigenvalues); but you do *not* know from the eigenvalues alone which way the spiral is spiraling, clockwise or counterclockwise.

26.2. **The** $(\mathrm{tr}, \det)$ **plane and structural stability.** We are now confronted with a large collection of autonomous systems, the linear two-dimensional systems $\dot{\mathbf{x}} = A\mathbf{x}$. This collection is parametrized by the four entries in the matrix. We have understood that much of the behavior of such a system is determined by two particular combinations of these four parameters, namely the trace and the determinant.

So we will consider now an entire plane with coordinates (T, D). Whenever we pick a point on this plane, we will be considering the linear autonomous systems whose matrix has trace T and determinant D.

Such a matrix is not well-defined. For given (T, D) there are always infinitely many matrices A with $\mathrm{tr}A = T$ and $\det A = D$. One example is the "associated matrix,"

$$A = \begin{bmatrix} 0 & 1 \\ -D & T \end{bmatrix}.$$

This is a particularly important example, because it represents the system corresponding to the LTI equation $\ddot{x} - T\dot{x} + Dx = 0$, via $y = \dot{x}$. (I'm sorry about the notation here. T and D are just numbers; Dx does not signify the derivative of x.)

The (T, D) plane divides into several parts according to the appearance of the phase portrait of the corresponding matrices. The important regions are as follows.

If $D < 0$, the eigenvalues are real and of opposite sign, and the phase portrait is a saddle (which is always unstable).

If $0 < D < T^2/4$, the eigenvalues are real, distinct, and of the same sign, and the phase portrait is a node, stable if $T < 0$, unstable if $T > 0$.

If $0 < T^2/4 < D$, the eigenvalues are neither real nor purely imaginary, and the phase portrait is a spiral, stable if $T < 0$, unstable if $T > 0$.

These three regions cover the whole of the (T, D) except for the curves separating them from each other, and so are them most commonly encountered and the most important cases. Suppose I have a matrix A with $(\mathrm{tr}A, \det A)$ in one of these regions. If someone kicks my matrix, so that its entries change slightly, I don't have to worry; if the change was small enough, the new matrix will be in the same region and the character of the phase portrait won't have changed very much. This is a feature known as "structural stability."

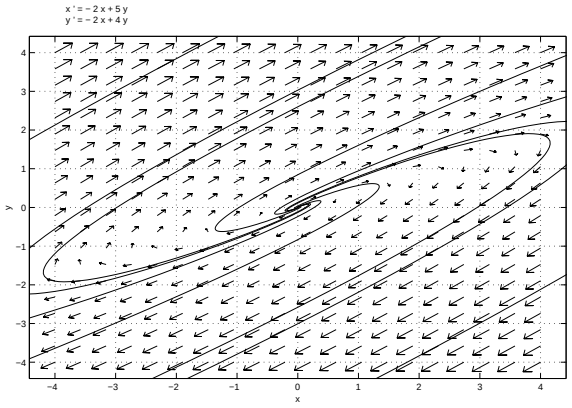
The remainder of the (T, D) plane is populated by matrices exhibiting various other phase portrait types. They are *structurally unstable*, in the sense that arbitrarily small perturbations of their entries can, and almost always will, result in a matrix with phase portrait of a different type. For example, when $0 < D$ and $T = 0$, the eigenvalues are purely imaginary, and the phase portrait is a center. But most perturbations of such a matrix will result in one whose eigenvalues have nonzero real part and hence whose phase portrait is a spiral.

26.3. **The portrait gallery.** Now for the dictionary of phase portraits. In the pictures which accompany these descriptions some elements are necessarily chosen at random. For one thing, most of the time there will be two independent eigenlines (i.e., lines through the origin made up of eigenvectors). Below, if these are real they will be the lines through $\alpha_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\alpha_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. If there is only one eigendirection (this only happens if $\lambda_1 = \lambda_2$ and is then called the "defective case") it will be the line through $\alpha_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. If they are not real, they are conjugate to each other and hence distinct. The question of how they influence the phase portrait is more complex and will not be addressed.

Name: Spiral.

Eigenvalues: Neither real nor purely imaginary: $0 \neq \mathrm{tr}^2/4 < \det$.

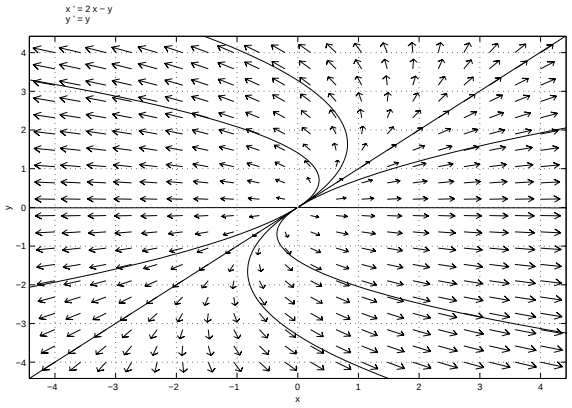
Stability: Stable if $\mathrm{tr} < 0$, unstable if $\mathrm{tr} > 0$.



Name: Node.

Eigenvalues: Real, same sign: $0 < \det < \mathrm{tr}^2/4$.

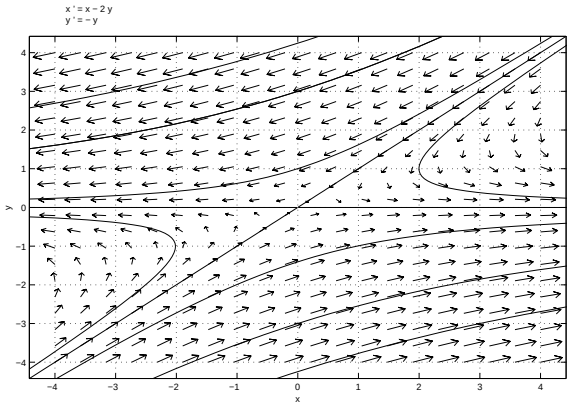
Stability: Stable if $\mathrm{tr} < 0$, unstable if $\mathrm{tr} > 0$.



Name: Saddle.

Eigenvalues: Real, opposite sign: $\det < 0$.

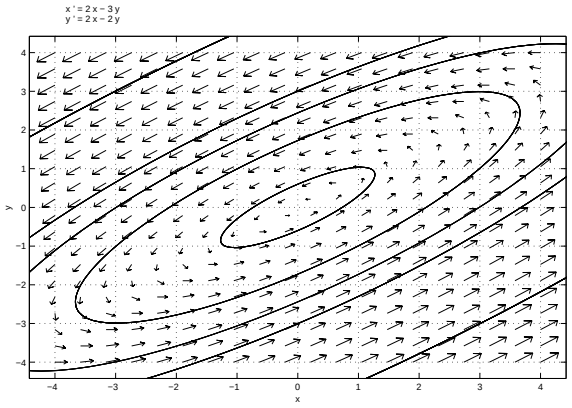
Stability: Unstable.



Name: Center.

Eigenvalues: Purely imaginary, nonzero: $\mathrm{tr} = 0, \det > 0$.

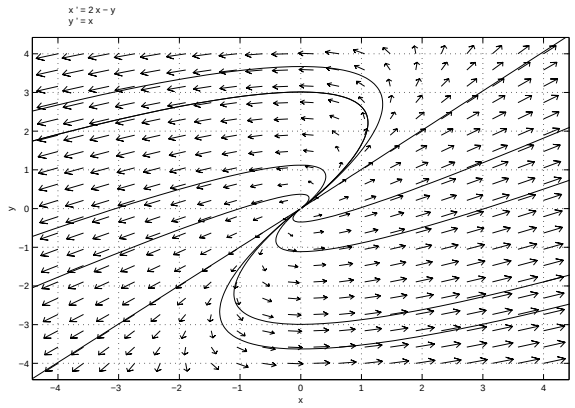
Stability: Neutrally stable.



Name: Defective Node.

Eigenvalues: Repeated (hence real) but nonzero: det $= \mathrm{tr}^2/4 > 0$; defective.

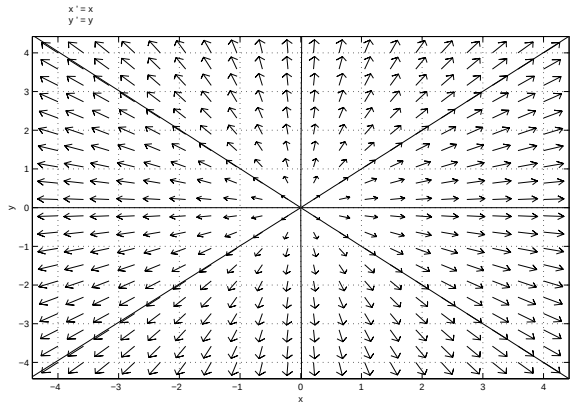
Stability: Stable if $\mathrm{tr} < 0$, unstable if $\mathrm{tr} > 0$.



Name: Star Node.

Eigenvalues: Repeated (hence real) but nonzero; complete: det $= \mathrm{tr}^2/4 > 0$.

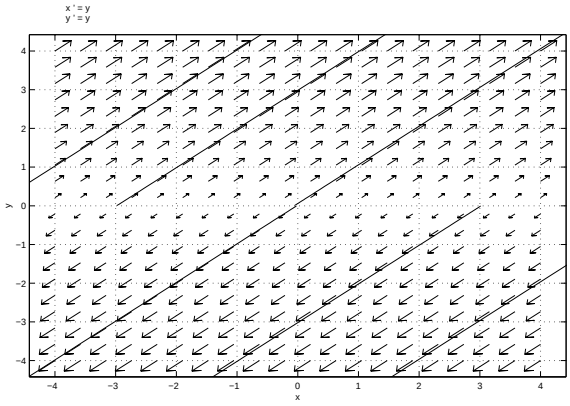
Stability: Stable if $\mathrm{tr} < 0$, unstable if $\mathrm{tr} > 0$.



Name: Degenerate: Comb.

Eigenvalues: One zero (hence both real).

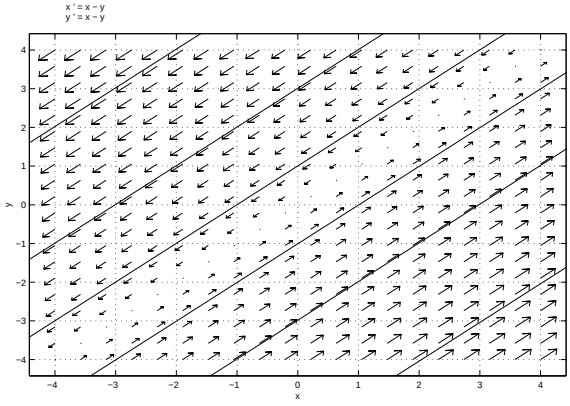
Stability: Stable if $\mathrm{tr} < 0$, unstable if $\mathrm{tr} > 0$.



Name: Degenerate: Parallel Lines.

Eigenvalues: Both zero: $\mathrm{tr} = \det = 0$; defective.

Stability: Unstable.



Name: Degenerate: Everywhere fixed; $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Eigenvalues: Both zero: $\mathrm{tr} = \det = 0$; complete.

Stability: Neutrally stable.

(No picture; every point is a constant solution.)

1. The Kermack-McKendrick equation

The Kermack-McKendrick Equation is an important and simple model for a virus epidemic, which either kills its victims or renders them immune, first considered by W. O. Kermack and A. G. McKendrick in 1927.

Parameter names:

- s is the fraction of the population which is susceptible to infection.
- c is the fraction of the population which is contagious.
- r is the fraction of the population which is removed, either by recovery with immunity or by death.
- β is the transmission rate: the proportionality constant mediating transmission.
- α is the rate of decay of contagiousness: an individual has a $e^{-\alpha\tau}$ chance to still be contagious a time τ after infection; or, alternatively, an individual's level of contagiousness declines according to this exponential decay.

We assume $s + c + r = 1$, so we are supposing that the only way an individual can fail to be susceptible to the desease is either to have it or to have had it. We are thus considering only the initially susceptible population.

The three variables are often written S, I (for "infected"), and R, and this model is often called the SIR model.

Does such an epidemic eventually infect the entire population, or is it somehow self-limiting? Will it take off or sputter out? Can we relate various aspects of the course of this epidemic, such as the peak of the contagious proportion of the population and the proportion of the population which ultimately contracts the disease?

The meaning of β is that $\dot{s} = -\beta sc$.

(a) Explain this: Why should $\dot{s}$ be proportional to both s and c? Why should the proportionality constant be negative?

The meaning of α is that $\dot{c} = \beta sc - \alpha c$.

(b) Explain this: What are the two processes yielding the two terms in $\dot{c}$? Why should one of them be $-\dot{s}$?

Together we have a "system" of equations: two functions of time, and an expression of the derivative of each in terms of the values of both. This is a subject we will take up in earnest at the end of the

course, but we can already get pretty far in analyzing it. The first step is the physics-inspired idea (as in EP §1.8) of eliminating time:

(c) Explain why these two equations imply

$$\frac{dc}{ds} = \gamma s^{-1} - 1$$

for some constant γ. What is γ in terms of α and β? From the expression of γ in terms of the decay rate of infection α, and the transmission rate β, what would you expect small γ to mean about the equation? What would large γ mean?

(d) Solve this equation; you should get

$$c = \gamma \ln |s| - s + a$$

where a is the constant of integration. When c is very small, that is, at the start of the epidemic, s should be near 1. This leads to $a = 1$. We choose to write the result as

$$c = \gamma \ln |s| - (s - 1). \tag{1}$$

Once we know c in terms of s, we can substitute this back into the original equation for $\dot{s}$ to obtain

$$\dot{s} = \beta s(s - 1 - \gamma \ln |s|). \tag{2}$$

This is a subtle variant of the logistic equation, with the advantage that one doesn't have to know the limiting population in advance.

This is an autonomous equation, and we will study its critical points. The effect of this equation depends upon the value of the parameter γ. (From now on we'll assume $s > 0$, as it is in the application at hand; in fact, in our application $0 < s < 1$.) First, draw the graphs of $s - 1$ and of $\ln s$. They are tangent at $(1, 0)$. Multiplying $\ln s$ by the positive constant γ stretches this graph vertically by a factor of $\gamma > 0$. It still meets the graph of $s - 1$ at $s = 1$, but if $\gamma \neq 1$ the graphs meet one additional time, at $s = s_{\text{crit}}$. If $\gamma > 1$, $s_{\text{crit}} > 1$; if $\gamma < 1$, $0 < s_{\text{crit}} < 1$.

The equation $s_{\text{crit}} - 1 = \gamma \ln |s_{\text{crit}}|$ can't be solved analytically for s_{crit} in terms of γ, but we can find γ in terms of s_{crit} (which we will restrict to be positive!):

$$\gamma = \frac{s_{\text{crit}} - 1}{\ln s_{\text{crit}}}.$$

It can be plugged into MATLAB. Let's do that.

MATLAB contains a powerful and accurate ODE solver, called `ode45`. If you fire up MATLAB and type `help ode45` you'll get a screenful of command summaries, from which the following is extracted.

In order to use `ode45` you will have to create and store a file containing a description of the function $F(x,t)$ occuring in the ODE $\dot{x} = F(x,t)$. Here's the file:

```
function sdot=sir(t,s,flag,beta,gamma);

sdot = beta*s*(s-1-gamma*log(s));
```

(Remember, the semicolons suppress the screen printout of the answers.)

Now let's pick some values. Reasonable values are $\gamma = .5$, $\beta = 1$. Take as initial value $s = .999999$ (so only one in a million is contagious). We might watch the evolution of the disease over 75 time units. All this is accomplished by typing:

```
[t,s]=ode45('sir',[0,75],.999999,[],1,.5)
```

When you hit ⟨`enter`⟩, a list of pairs of numbers should stream across the screen. Those are values of t followed by corresponding computed values of s. This command has defined two lists of numbers, t and s, of the same length.

We want to plot s against t. This is easy: type

```
plot(t,s)
```

A window appears with a graph on it. The graph is somewhat deceptive; it probably doesn't extend all the way down to $s = 0$. To make it extend to zero, declare explicitly what you want your axis dimensions to be:

```
axes([0,75,0,1])
```

To line things up with the eye a little better, add a grid, using `grid`. Notice how long it takes for the susceptable population to get enough below 1.0 to be visible on this plot!

Finally, let's add to this the graph of c, the contagious population. First compute it using **(d)** above:

```
c=(.5)*log(s)-s+1
```

Again a list of numbers will stream across the screen. To plot it against t, on the same graph, type `hold on` (which makes future graphing commands plot on the same window instead of a new one) and then `plot(t,c)`. Again, note how long it takes for the contagious population to become noticeable!

(e) From the graph, estimate the time at which the contagious population is a maximum, and what that maximum is. Also, estimate the

fraction of the population which is left susceptible as $t \to \infty$. Print out and hand in the plot you made (by selecting `File/Print` on `Figure No. 1`).

Further comments:

The significance of (1) depends upon the value of γ. The tangent line to $c = \ln|x|$ at $(1, 0)$ is given by $c = s - 1$ and lies above the curve. If $\gamma > 1$, $\gamma \ln|s|$ passes through the same point but more steeply, and intersects $c = s - 1$ at $s = 1$ and again at a point $s > 1$, which is not meaningful in our application. If $\gamma > 1$, $c < 0$ when $0 < s < 1$. What happens here is: nothing, the epidemic never gets started, since the rate of decay of contagiousness is too large relative to the transmission rate of the disease. On the other hand if $\gamma < 1$, the epidemic flares and dies off, with little effect if γ is near 1 and with more devastating effect if γ is small. Small γ reflects a slow decay rate of contagiousness relative to the transmission rate of the desease. The specific values of α and β determine the speed with which the epidemic spreads, but its trajectory in the (s, c) plane depends only on the quotient γ.

The limiting value of s, as $t \to \infty$, can be found by setting $c = 0$ in (1). It satisfies the equation

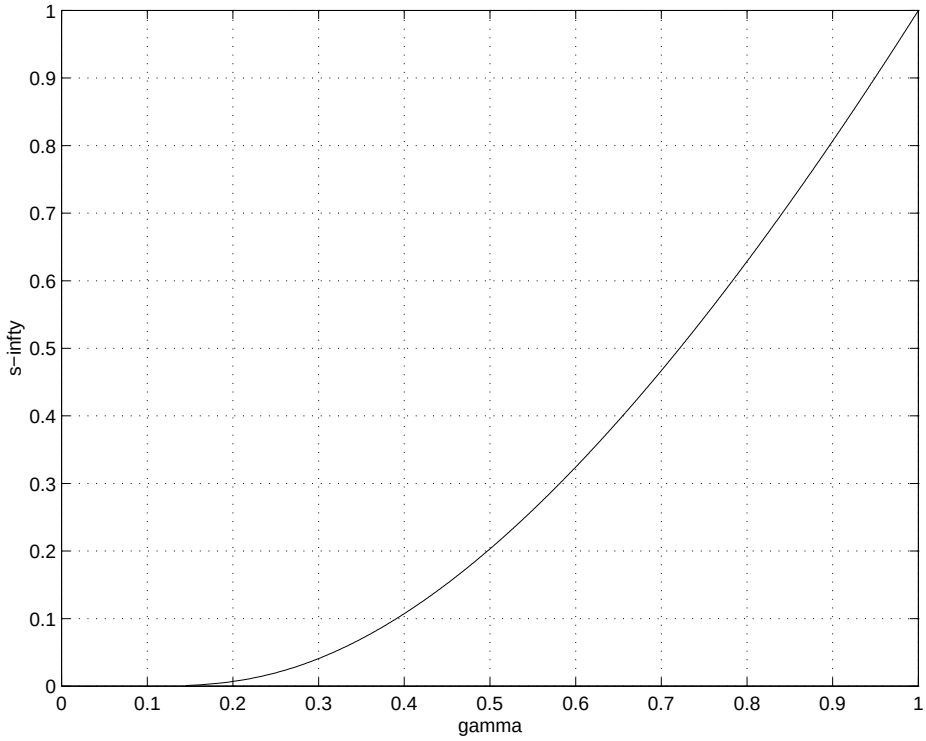
$$\gamma = \frac{s_\infty - 1}{\ln s_\infty} \tag{3}$$

This can't be solved for s_∞ in elementary functions, but it's easy to use Matlab to compute γ in terms of s_∞ and then plot the inverse function. One finds that if $\gamma < .2$ then $s_\infty < .01$: more than 99% of the population is infected. As $\gamma \uparrow 1$, $s_\infty \uparrow 1$: the epidemic affects a very small portion of the population if γ is near 1. In any case, some fraction of the population will always survive.

The differential equation $\dot{c} = \beta sc - \alpha c$ tells us where to find the center of the epidemic: the maximum of c occurs when $s = \gamma$, and thus, by (1), is given by

$$c_{\max} = 1 - \gamma + \gamma \ln \gamma. \tag{4}$$

The invariant γ is thus available to doctors monitoring the disease, assuming they can recognize contagious individuals: they have to watch for the number of contagious individuals to peak. Knowing this peak, they can use (4) to find γ and then (3) to find s_∞—that is, to predict the eventual fraction of the population which will be infected and rendered immune or dead. (Since $c \to 0$, this is $1 - s_\infty$.)


s−infty
gamma
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

1. The Tacoma Narrows Bridge: resonance vs flutter

On July 1, 1940, a bridge spanning the Tacoma Narrows opened to great celebration. It dramatically shortened the trip from Seattle to the Kitsap Peninsula. It was an elegant suspension bridge, a mile long (third longest in the US at the time) but just 39 feet across. Through the summer and early fall, drivers noticed that it tended to oscillate vertically, quite dramatically. It came to be known as "Galloping Gertie." "Motorists crossing the bridge sometimes experienced "roller-coaster like" travel as they watched cars ahead almost disappear vertically from sight, then reappear."[1]

During the first fall storm, on November 7, 1940, with steady winds above 40 mph, the bridge began to exhibit a different behavior. It *twisted*, part of one edge rising while the opposing edge fell, and then the reverse. At 10:00 AM the bridge was closed. The torsional oscillations continued to grow in amplitude, till, at just after 11:00, the central span of the bridge collapsed and fell into the water below. One car and a dog were lost.

Why did this collapse occur? Were the earlier oscillations a warning sign? Many differential equations textbooks announce that this is an example of *resonance*: the gusts of wind just happened to match the natural frequency of the bridge.

The problem with this explanation is that the wind was not gusting—certainly not at anything like the natural frequency of the bridge. This explanation is worthless.

Structural engineers have studied this question in great detail. They had determined already before the bridge collapsed that the vertical oscillation was self-limiting, and not likely to lead to a problem. The torsion oscillation was different. To model it, pick a portion of the bridge far from the support towers. Let $\theta(t)$ denote its angle off of horizontal, as a function of time. The torsional dynamics can be modeled by a second order differential equation of the form

$$\ddot{\theta} + b\dot{\theta} + \omega_n^2\theta = F$$

where ω_n^2 is the natural circular frequency of the torsional oscillation, and b is a damping term. The forcing term F depends upon θ itself, and its derivatives. To a reasonable approximation we can write

$$F = a_0\theta + a_1\dot{\theta}$$

where a_0 and a_0 are functions of the wind velocity v which are determined by the bridge characteristics. The resulting differential equation

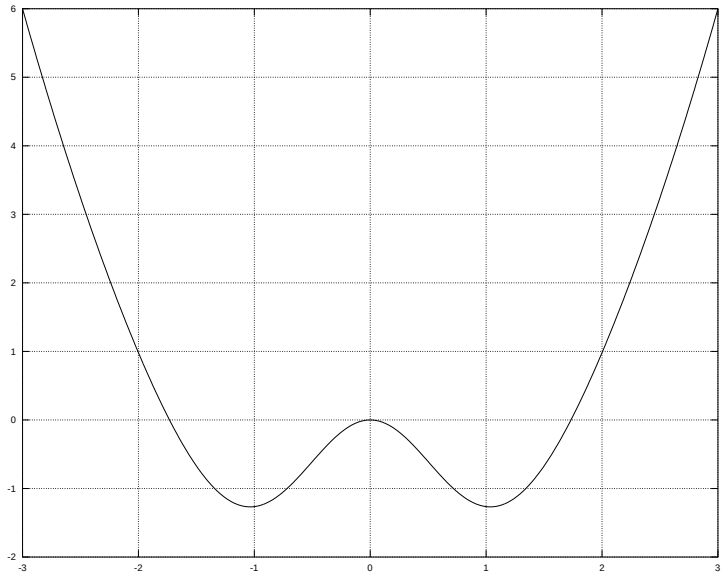


FIGURE 1. Dependence of the forcing term on wind velocity

is analogous to the equation governing the behavior of the mass in a spring/mass/dashpot system which is driven through both the mass and the dashpot—except that that "input signal," which is the position of the forcing plate in the spring/mass/dashpot system, is now the output signal, the angular deflection itself. This is an instance of "self-excitation."

Notice that this equation can be rewritten as

$$\ddot{\theta} + (b - a_1)\dot{\theta} + (\omega_n^2 - a_0)\theta = 0 \tag{1}$$

It turns out that in the case of the Tacoma Narrows bridge the value of a_0 is small relative to ω_n^2; the effect is to slightly alter the effective natural frequency of torsional oscillation. For simplicity we'll just suppose it's negligable and drop it.

The function $a_1(v)$ reflects mainly turbulence effects. The technical term for this effect is *flutter*. The same mechanism makes flags flap and snap in the wind. It turns out that the graph of $a_1(v)$ has the following shape.

When $|v|$ is small, $a_1(v) < 0$: the wind actually increases the damping of the bridge; it becomes *more* stable. When $|v|$ is somewhat larger,

$a_1(v) = 0$, and the wind has no damping effect. When $|v|$ increases still more, it starts to erode the damping of the bridge, till, when v hits a certain critical value, it overwhelms the intrinsic damping of the bridge. The result is *anti-damping*, a negative effective damping constant. For the Tacomah Narrows Bridge, the critical value of velocity was discovered, on November 7, 1940, to be around 40 miles per hour.

Solutions to (1) are linear combinations of the functions e^{rt} where r is a root of the characteristic polynomial $p(s) = s^2 + (b - a_1)s + \omega_n^2$:

$$r = -\frac{b - a_1}{2} \pm \sqrt{\frac{(b - a_1)^2}{4} - \omega_n^2}$$

The movies of the bridge collapse clearly show large oscillations, so in this regime $|b - a_1| < 2\omega_n$, square root is negative, and the roots have nonzero imaginary parts. The real part of each root is $k = (a_1 - b)/2$, and when v is such that $a_1(v) > b$ this is positive. If we write $r = k \pm i\omega$, the general solution is

$$\theta = Ae^{kt} \cos(\omega t - \phi)$$

Its peaks grow in magnitude, exponentially.

This spells disaster. There are compensating influences which slow down the rate of growth of the maxima, but in the end the system will—and did—break down.

References

[1] K. Y. Billah and R. H. Scanlan, Resonance, Tacomah Narrows bridge failure, and undergraduate physics textbooks, Am. J. Phys. 59 (1991) 118–124.

1. Linearization: the phugoid equation as example

"Linearization" is one of the most important and widely used mathematical terms in applications to Science and Engineering. In the context of Differential Equations, the word has two somewhat different meanings.

On the one hand, it may refer to the procedure of analyzing solutions of a nonlinear differential equation near a critical point by studying an approximating linear equation. This is linearizing an *equation*.

On the other hand, it may refer to the process of systematically dropping negligibly small terms in the mathematical expression of the model itself, under the assumption that one is near an equilibrium. The result is that you obtain a linear differential equation directly, without passing through a nonlinear differential equation. This is linearizing a *model*.

A virtue of the second process is that it avoids the need to work out the full nonlinear equation. This may be a challenging problem, often requiring clever changes of coordinates; while, in contrast, it is always quite straightforward to write down the linearization near equilibrium, by using a few general ideas. We will describe some of these ideas in this section.

Most of the time, the linearization contains all the information about the behavior of the system near equilibrium, and we have a pretty complete understanding of how linear systems behave, at least in two dimensions. There aren't too many behaviors possible. The questions to ask are: is the system stable or unstable? If it's stable, is it underdamped (so the solution spirals towards the critical point) or overdamped (so it decays exponentially without oscillation)? If it's underdamped, what is the period of oscillation? In either case, what is the damping ratio?

One textbook example of this process is the analysis of the linear pendulum. In this section we will describe a slightly more complicated example, the "phugoid equation" of airfoil flight.

1.1. **The airplane system near equilibrium.**

If you have ever flown a light aircraft, you know about "dolphining" or "phugoid oscillation." This is precisely the return of the aircraft to the equilibrium state of steady horizontal flight. We'll analyze this effect by linearizing the model near to this equilibrium. To repeat, the questions to ask are: Is this equilibrium stable or unstable? (Experience suggests it's stable!)

Is it overdamped or underdamped? What is the damping ratio? If it's underdamped, what is the period (or, more properly, the quasiperiod)?

There are four forces at work: thrust F, lift L, drag D, and weight $W = mg$. At equilibrium F and D cancel, and L and W cancel. Here's a diagram. In it the airplane is aligned with the thrust vector, since the engines provide a force pointing parallel with the body of the airplane.

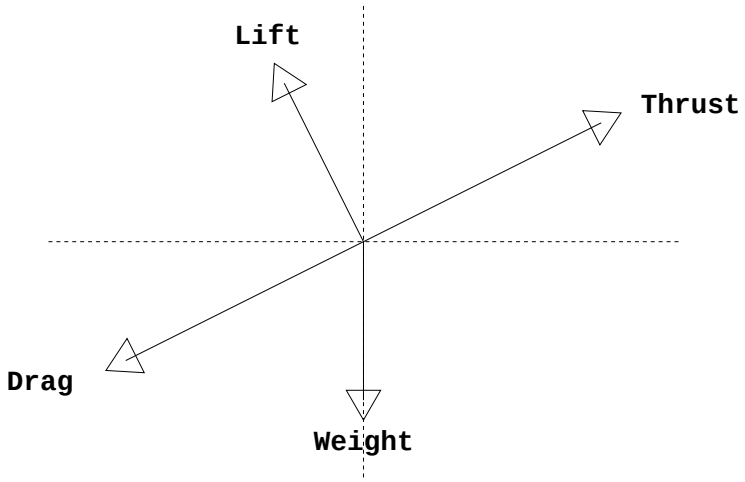


FIGURE 1. Forces on an airfoil

I'll make the following simplifying assumptions: **(1)** the air is still relative to the ground (or, more generally, the ambient air is moving uniformly and we use a coordinate frame moving with the air); **(2)** the weight and the thrust are both constant.

Lift and the drag are more complicated than weight and thrust. They are components of a "frictional" force exerted on the plane by the surrounding air. The drag is, by definition, the component of that force in the direction of the thrust (directed backwards), and the lift is the perpendicular component, directed towards the "up" side of the airfoil.

When we call this force "frictional," what we mean is that it depends upon the velocity of the plane (through the air) and on nothing else.

Friction is a complex process, and it shows up differently in different regimes. Let's first think about friction of a particle moving along the x axis. It is then a force $\phi(v)$ dependent upon $v = \dot{x}$. It always takes the value zero when the velocity is zero and is directed against the direction of motion. The tangent line approximation then lets us approximate $\phi(v)$ by a multiple of v when $|v|$ is small. This is "linear damping," and

it plays a big role in our study of second order LTI systems. When the velocity is relatively large, consideration of the nonlinear dependence of friction on velocity becomes unavoidable. Often, for v in a range of values the frictional force is reasonably well approximated by a power law:

$$\phi(v) = \begin{cases} -c|v|^p & \text{for } v \geq 0 \\ c|v|^p & \text{for } v < 0 \end{cases} \tag{1}$$

where $c > 0$ is a constant. This rather complicated looking expression guarantees that the force acts against the direction of motion. The magnitude is $|\phi(v)| = c|v|^p$.

Often the power involved is $p = 2$, so $\phi(v) = -cv^2$ when $v > 0$. (Since squares are automatically positive we can drop the absolute values and the division into cases in (1).) To analyze motion near a given velocity v_0, the tangent line approximation indicates that we need only study the rate of change of $\phi(v)$ near the velocity v_0, and when $p = 2$ and $v_0 > 0$,

$$\phi'(v_0) = -2cv_0 = \frac{2\phi(v_0)}{v_0}\,. \tag{2}$$

We rewrote the derivative in terms of $\phi(v_0)$ because doing so eliminates the constant c.

Now let's go back to the airfoil. Our last assumption is that near equilibrium velocity v_0, drag and lift depend quadratically on speed. Stated in terms of (2) we have our next assumption: **(3)** the drag $D(v)$ and the lift $L(v)$ are quadratic, so by (2) they satisfy

$$D'(v_0) = \frac{2D(v_0)}{v_0}\,, \qquad L'(v_0) = \frac{2L(v_0)}{v_0}\,.$$

There is an equilibrium velocity at which the forces are in balance: cruising velocity v_0. Our final assumption is that at cruising velocity the pitch of the airplane is small: so **(4)** the horizontal component of lift is small. The effect is that to a good approximation, lift balances weight and thrust balances drag:

$$D(v_0) = F\,, \qquad L(v_0) = mg\,.$$

This lets us rewrite the equations for the derivatives can be rewritten

$$D'(v_0) = \frac{2F}{v_0}\,, \qquad L'(v_0) = \frac{2mg}{v_0}\,. \tag{3}$$

This is all we need to know about the dynamics of airfoil flight.

There are several steps in our analysis of this situation from this point. A preliminary observation is that in the phugoid situation the airplane has no contact with the ground, so **everything is invariant under space translation**. After all, the situation is the same for all altitudes (within a range over which atmospheric conditions and gravity are reasonably constant) and for all geographical locations. The implication is that Newton's Law can be written entirely in terms of velocity and its derivative, acceleration. Newton's Law is a *second order* equation for position, but if the forces involved don't depend upon position it can be rewritten as a *first order* equation for velocity. This reasoning is known as *reduction of order*.

1.2. **Deriving the linearized equation of motion.** The fundamental decision of linearization is this:

Study the situation near the equilibrium we care about, and systematically use the tangent line approximation at that equilibrium to simplify expressions.

The process of replacing a function by its tangent line approximation is referred to as "working to first order."

Let's see how this principle works out in the phugoid situation.

One of the first steps in any mathematical analysis is to identify and give symbols for relevant parameters of the system, and perhaps to set up a well-adapted coordinate system. Here, we are certainly interested in the velocity. We have already introduced v_0 for the equilibrium velocity, which by assumption **(4)** is horizontal. We write the actual velocity as equilibrium plus a correction term: Write

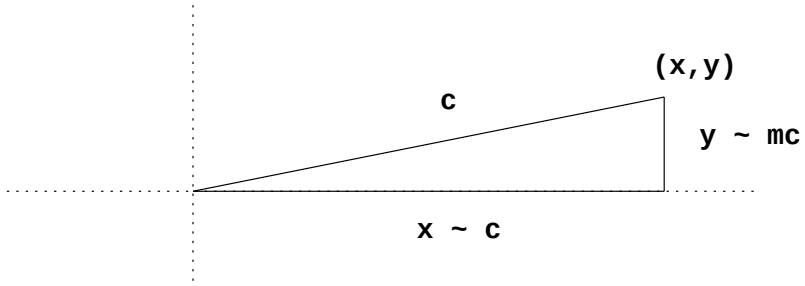
w for the vertical component of velocity, and

$v_0 + u$ for the horizontal component,

and suppose the axes are arranged so that the plane is moving in the direction of the positive x axis. We are assuming that the plane is not too far from equilibrium, so we are assuming that w and u are both small.

We will want to approximate the actual speed in terms of v_0, u, and w. To do this, and for other reasons too, we will use a geometric principle which arises very often in linearization of physical systems.

If a vector makes a small angle with the positive x axis, then to first order its x component is its length and its y component is its length times the slope.



This is geometrically obvious, and equivalent to the facts that $\cos'(0) = 0$ and $\sin'(0) = 1$.

If we take $x = v_0 + u$, $y = w$, and $c = v$, the estimate $x \sim c$ says that the speed is approximately $v_0 + u$; the normal component w makes only a "second order" contribution and we will ignore it.

Now we use the linearization principle again: we plug this estimate of the speed into the tangent line approximation for $D(v)$ and $L(v)$ and use (3) and the values $D(v_0) = F$ and $L(v_0) = mg$ to find

$$D \simeq F + \frac{2F}{v_0}u\,, \quad L \simeq mg + \frac{2mg}{v_0}u\,.$$

Subscript L, W, T, and D by h and v to denote their horizontal and vertical components. Writing down similar triangles, we find (to first order, always—ignoring terms like u^2, uw, and w^2):

$$L_v \simeq L \simeq mg + \frac{2mg}{v_0}u\,, \quad L_h \simeq \frac{w}{v_0}L \simeq \frac{w}{v_0}mg$$

$$W_v = mg\,, \quad W_h = 0\,, \quad T_v = \frac{w}{v_0}F\,, \quad T_h \simeq F$$

$$D_v \simeq \frac{w}{v_0}D \simeq \frac{w}{v_0}F\,, \quad D_h \simeq D \simeq F + \frac{2F}{v_0}u.$$

In words, to first order the vertical components of thrust and drag still cancel and the vertical component of the lift in excess of the weight is given by $(2mg/v_0)u$, so, by Newton's law,

$$m\dot{w} = \frac{2mg}{v_0}u. \tag{4}$$

Also, to first order, the horizontal component of the excess of drag over thrust is $(2F/v_0)u$, and the horizontal component of the lift is $-mg(w/v_0)$: so

$$m\dot{u} = -\frac{2F}{v_0}u - \frac{mg}{v_0}w. \tag{5}$$

We can package these findings in matrix terms:

$$\frac{d}{dt}\begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} -2F/mv_0 & -g/v_0 \\ 2g/v_0 & 0 \end{bmatrix}\begin{bmatrix} u \\ w \end{bmatrix}. \tag{6}$$

and we could go on to use the methods of linear systems to solve it. Instead, though, we will solve the equations (4), (5) by elimination. Differentiating the equation for $\dot{w}$ and substituting the value for $\dot{u}$ from the other equation gives the homogeneous second order constant coefficient linear differential equation

$$\boxed{\ddot{w} + \frac{2F}{mv_0}\dot{w} + \frac{2g^2}{v_0^2}w = 0} \tag{7}$$

1.3. **Implications.** From this (or from the system (6)) we can read off the essential characteristics of motion near equilibrium. We have in (7) a second order homogeneous linear ODE with constant coefficients; it is of the form

$$\ddot{w} + 2\zeta\omega_n\dot{w} + \omega_n^2 w = 0,$$

where ω_n is the *natural circular frequency* and ζ is the *damping ratio* (for which see Section **??**). Comparing coefficients,

$$\omega_n = \frac{\sqrt{2}\,g}{v_0} \quad , \qquad \zeta = \frac{F}{\sqrt{2}\,mg}.$$

We have learned the interesting fact that the period

$$P = \frac{2\pi}{\omega_n} = \frac{\sqrt{2}\,\pi}{g}v_0$$

of phugoid oscillation depends *only on the equilibrium velocity* v_0. In units of meters and seconds, P is about $0.45\,v_0$. The nominal equilibrium speeds v_0 for a Boeing 747 and an F15 are 260 m/sec and 838 m/sec, respectively. The corresponding phugoid periods are about 118 sec and 380 sec.

We have also discovered that the phugoid damping ratio depends *only on the "thrust/weight ratio,"* a standard tabulated index for aircraft. Both ζ and F/mg are dimensionless ratios, and ζ is about $.707(F/mg)$, independent of units. F/mg is about 0.27 for a Boeing 747, and about 0.67 for an F15.

The system is underdamped as long as $\zeta < 1$, i.e. $(F/mg) < \sqrt{2}$. Even an F15 doesn't come close to having a thrust/weight approaching 1.414.

To see a little more detail about these solutions, let's begin by supposing that the damping ratio is negligible. The equation (7) is then simply a harmonic oscillator with angular frequency ω_n, with general solution of the form

$$w = w_0 \cos(\omega_n t - \phi) .$$

Equation (4) then shows that $u = (v_0/2g)\dot{w} = -(v_0/2g)\omega_n w_0 \sin(\omega_n t - \phi)$. But $\omega_n = \sqrt{2}g/v_0$, so this is

$$u = -(w_0/\sqrt{2}) \sin(\omega_n t - \phi) .$$

That is: The vertical amplitude is $\sqrt{2}$ times as great as the horizontal amplitude.

Integrate once more to get the motion in space:

$$x = x_0 + v_0 t + a \cos(\omega_n t - \phi)$$

where $a = v_0 w_0/g$—as a check, note that a does have units of length!—and

$$y = y_0 + \sqrt{2}\, a \, \sin(\omega_n t - \phi) ,$$

for appropriate constants of integration x_0 (which is the value of x at $t = 0$) and y_0 (which is the average altitude). Relative to the frame of equilibrium motion, the plane executes an ellipse whose vertical axis is $\sqrt{2}$ times its horizontal axis, moving counterclockwise. (Remember, the plane is moving to the right.)

Relative to the frame of the ambient air, the plane follows a roughly sinusoidal path. The horizontal deviation u from equilibrium velocity is small and would be hard to detect in the flightpath.

Reintroducing the damping, the plane spirals back to equilibrium.

We can paraphrase the behavior in physics terms like this: Something jars the airplane off of equilibrium; suppose it is hit by a downdraft and the vertical component of its velocity, w, acquires a negative value. This puts us on the leftmost point on the loop. The result is a decrease in altitude, and the loss in potential energy translates to a gain in kinetic energy. The plane speeds up, increasing the lift, which counteracts the negative w. We are now at the bottom of the loop. The excess velocity continues to produce excess lift, which raises the plane past equilibrium (at the rightmost point on the loop). The plane now has $w > 0$, and rises above its original altitude. Kinetic energy is converted to potential energy, the plane slows down, passes through the top of the loop; the lowered speed results in less lift, and the plane returns to where it was just after the downdraft hit (in the frame of equilibrium motion).

A typical severe downdraft has speed on the order of 15 m/sec, so we might take $c = 10$ m/sec. With the 747 flying at 260 m/sec, this results in a vertical amplitude of 265 meters; the F15 flying at 838 m/sec gives a vertical amplitude of 855 meters, which could pose a problem if you are near the ground!

Historical note: The term *phugoid* was coined by F. W. Lanchester in his 1908 book *Aerodonetics* to refer to the equations of airfoil flight. He based this neologism on the Greek $\phi\upsilon\gamma\acute{\eta}$, which does mean flight, but in the sense of the English word *fugitive*, not in the sense of movement through the air. Evidently Greek was not his strong suit.

Question: Assumption **(3)** is the most suspect part of this analysis. Suppose instead of quadratic dependence we assume some other power law, for lift and drag. What is the analogue of (3), and how does this alter our analysis?

18.03 Differential Equations

Spring 2010