Woods Hole

Methods of Computational Neuroscience

Differential Equations and Linear Algebra

Lecture Notes

1. Classification of differential equations

Differential equations, and their solution methods, are classified by both:

1. the degree of the highest derivative (i.e., the **order**)
2. whether or not the equation is **linear** or **nonlinear** —
   if terms with the unknown are only to the first power, the equation is linear

Examples:

$$\frac{dx}{dt} + x = t\,, \quad \Rightarrow \quad \text{First order linear}$$

$$\frac{dx}{dt} + x^3 = t\,, \quad \Rightarrow \quad \text{First order nonlinear}$$

$$m\frac{d^2y}{dt^2} + b\left(\frac{dy}{dt}\right)^2 + ky = 0\,, \quad \Rightarrow \quad \text{Second order nonlinear}$$

$$\frac{d^2\vartheta}{dt^2} + \frac{g}{\ell}\sin\vartheta = 0\,. \quad \Rightarrow \quad \text{Second order nonlinear}$$

It is possible to have more than one unknown, and that the equations for these unknowns are linked together. This leads to **systems** of equations, such as

$$\frac{dx}{dt} = ax - bxy\,,$$

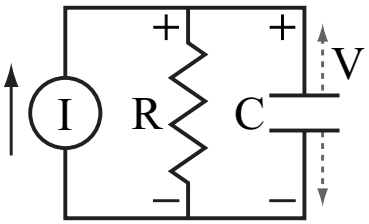
$$\frac{dy}{dt} = -cy + dxy\,.$$

Finally, we make the distinction between **ordinary** and **partial** differential equations. The latter arise when there is more than one independent variable (such as a spatial and a time variable) and *partial derivatives* are therefore required to describe some situation. One example of this is the diffusion equation,

$$\lambda^2 \frac{\partial^2 v}{\partial x^2} = \tau_m \frac{\partial v}{\partial t}\,,$$

which (among other things) describes the spread of voltage in a passive dendrite.

2. A simple RC circuit



Voltage V (units: volts) and current I (units:amps)

Resistor:

$$V_R = I_R R$$

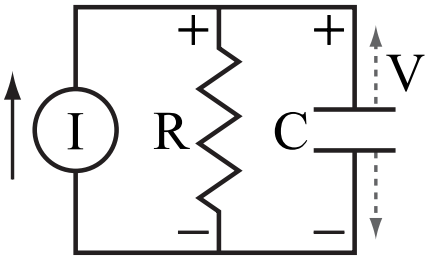
Capacitor:

$$I_C = C\frac{dV_C}{dt}$$

Kirchoff's laws of circuits:

1. voltage drops around any closed loop must always add to zero
2. current can be neither created nor destroyed

Differential equation & equivalent circuit



$$I = \frac{V}{R} + C\frac{dV}{dt} \quad \text{with} \quad V(0) = 0\,,$$

or

$$\frac{dV}{dt} + \frac{V}{RC} = \frac{I}{C} \quad \text{with} \quad V(0) = 0\,. \qquad (1)$$

Choose values in the ballpark of those for a small cell:
$R = 10^9\,\Omega$, $C = 1\,\text{pF}$, and $I = 100\,\text{pA}$

Steady-state solution (when $dV/dt = 0$):
$V = IR = 100\,\text{pA} \times 10^9\,\Omega = 100\,\text{mV}$

Initial slope:
At $t = 0$, $dV/dt = I/C = (100\,\text{pA})/(1\,\text{pF}) = 100\,\text{mV/ms}$

Time to steady state is (roughly) $RI/(I/C) = RC = 1\,\text{ms}$

3. Direction fields and solution curves

One way to understand the behavior of such equations is to generate solutions graphically. Equation (1) is a special case of the more general equation

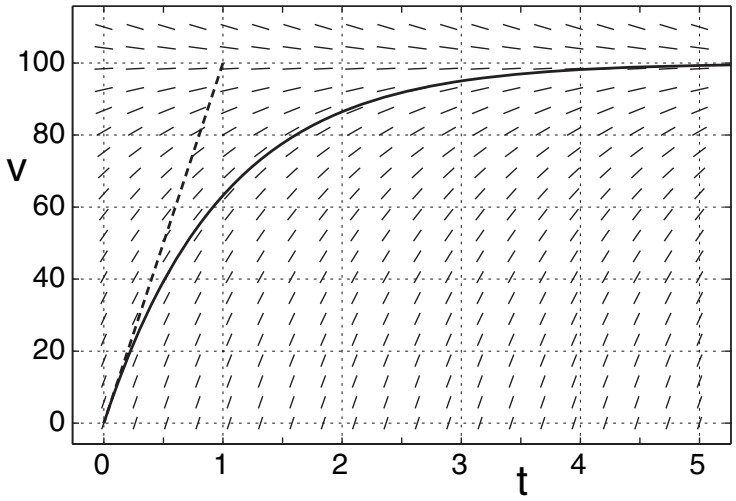
$$\frac{dv}{dt} = f(v, t)\,. \qquad (2)$$

Given an arbitrary point in the (t, v) plane, this allows one to find the **slope of the tangent** dv/dt at that point. Any solution going through that point must have the same slope.

In practice, one can draw the tangents throughout the (t, v) plane. The graphical picture one constructs is then known as a **slope or direction field**. One nice program that does this is `dfield`, a direction field program written in Matlab by John Polking of Rice University. `dfield` also plots solution curves; initial conditions can be entered merely by clicking on the appropriate point)

Here's how the direction field looks for

$$\frac{dV}{dt} = 100 - V \quad \text{with} \quad V(0) = 0\,,$$



Note that the estimate of 1 ms as the time for the voltage to reach 100 mV is not right, because this was based on the slope at V=0 (dashed line), and the slope decreases as V increases.

4. Separable equations

The simplest method of obtaining an explicit solution is when the equation is said to be **separable**, which occurs when all of the dependent variables can be separated from the independent variables so that each of them only appears on one side of the equation:

$$\frac{dV}{dt} = 100 - V \quad \Rightarrow \quad \frac{dV}{100 - V} = dt\,.$$

Both sides of this equation can now be integrated,

$$\int \frac{dV}{100 - V} = \int dt\,, \quad \Rightarrow \quad -\ln|100 - V| = t + K\,,$$

where K is a constant of integration. Some algebra then leads to

$$\ln|100 - V| = -t - K\,, \quad \Rightarrow \quad |100 - V| = e^{-t-K} = Ae^{-t}\,,$$

$$\Rightarrow \quad 100 - V = Ae^{-t}\,, \quad \Rightarrow \quad V = 100 - Ae^{-t}\,.$$

where $A = e^{-K}$. (Can you explain why the absolute value signs were dropped?)

The last result gives V as a function of t and is the solution we were seeking. Because of the arbitrary constant A appearing in the solution, however, this is known as a **general solution**. To find the **particular solution** that satisfies the **initial condition** $V(0) = 0$, we solve for the A that makes this true. We find $A = 100$, and therefore

$$V(t) = 100(1 - e^{-t}) .$$

Note that if the initial condition were $V(0) = 100$, then we would have $V(t) = 100$ for all time, i.e., the solution is just a constant, $dV/dt = 0$. Constant solutions of a differential equation (the value(s) of V that make $dV/dt = 0$) are known as **steady states** or **critical points**.

Exercise: Show for the general case

$$\frac{dV}{dt} + \frac{V}{RC} = \frac{I}{C} \quad \text{with} \quad V(0) = 0$$

that the solution is

$$V(t) = IR\left(1 - e^{-t/RC}\right) .$$

5. First-order linear equations

If in the previous example the applied current is not constant,

$$\frac{dV}{dt} + \frac{V}{RC} = \frac{I(t)}{C} , \tag{3}$$

then the equation is no longer separable. We can still solve the equation, however. First of all, note that if $I(t) \equiv 0$, then the solution is (it's separable)

$$V(t) = V_0\, e^{-t/RC} .$$

This suggests trying to solve the more general problem by using the transformation

$$V(t) = W(t)\, e^{-t/RC} .$$

This particular method is known as **variation of parameters**. We expect the right-hand-side in the equation for $W(t)$ to be proportional to $I(t)$, since when $I(t) \equiv 0$ we must get the solution $W(t) = V_0$. Doing the algebra, we get

$$\frac{dW}{dt} = e^{t/RC}\,\frac{I(t)}{C} .$$

Therefore

$$W(t) = V_0 + \int_0^t e^{\tau/RC}\,\frac{I(\tau)}{C}\,d\tau$$

and

$$V(t) = V_0 e^{-t/RC} + \int_0^t e^{-(t-\tau)/RC}\,\frac{I(\tau)}{C}\,d\tau\,.$$

In general, this method will work on any **first-order, linear** equation, i.e.,

$$\frac{dy}{dt} + p(t)y = f(t)\,.$$

Shortcut: Multiple the equation by the **integrating factor** $e^{\int p(t)\,dt}$; the equation then becomes an exact derivative:

$$e^{\int p(t)\,dt}\,\frac{dy}{dt} + p(t)e^{\int p(t)\,dt}\,y = \frac{d}{dt}\left(e^{\int p(t)\,dt}\,y\right) = e^{\int p(t)\,dt}\,f(t)\,.$$

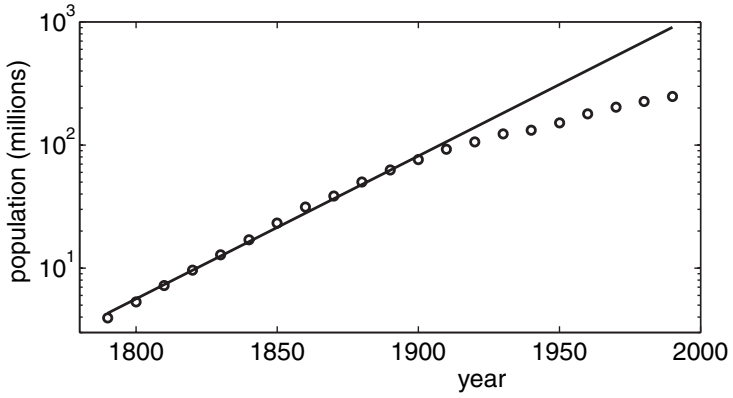
Exercise: Find the solution of Eq. (3) if $I(t) = I_0 + I_1 \sin\omega t$.

6. Exponential and logistic population growth

The simplest model of population growth is a growth rate proportional to the current population:

$$\frac{dP}{dt} = k_0 P \quad \Rightarrow \quad P(t) = P_0 e^{k_0 t}\,.$$

Note that $\ln P = k_0 t + \ln P_0$ is a *linear* function. Example: fit to U.S. population growth between 1790 to 1910 shows exponential growth, but growth after 1910 is slower.



Logistic growth:

$$\frac{dP}{dt} = kP(M-P)$$

"Effective" growth rate: $k(M-P)$. Net growth rate decreases as population increases (due to either real birth rate declining or death rate increasing. This equation is separable; the analytic solution is (**exercise**)

$$P(t) = \frac{P_0 M}{P_0 + (M-P_0)e^{-kMt}}$$

For fitting, note that relative growth rate

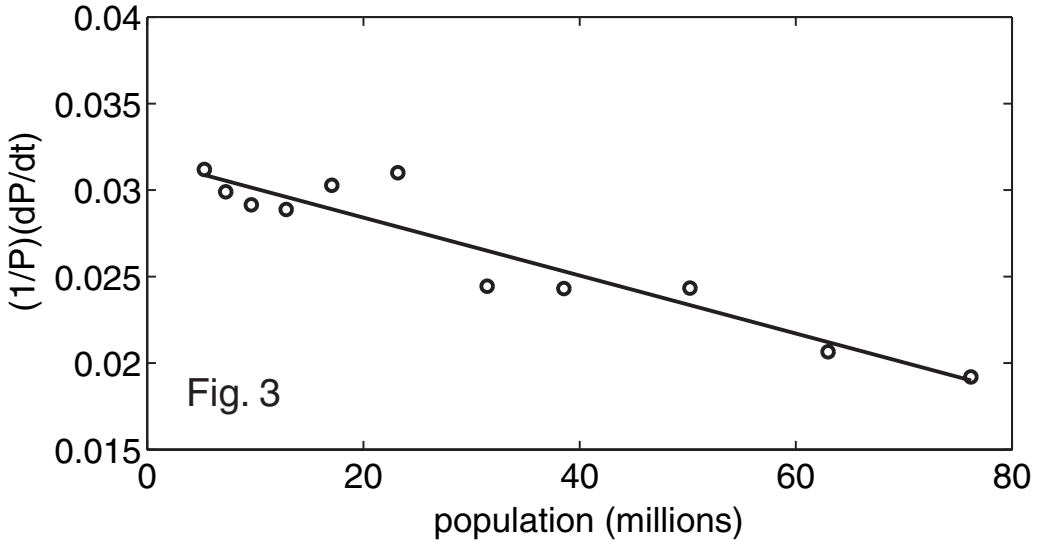
$$\frac{1}{P}\frac{dP}{dt} = kM - kP$$

is a linear function of P. So if we compute the derivative approximately, using

$$\frac{dP}{dt}(t_k) = \frac{P(t_{k+1}) - P(t_{k-1})}{t_{k+1} - t_{k-1}}$$

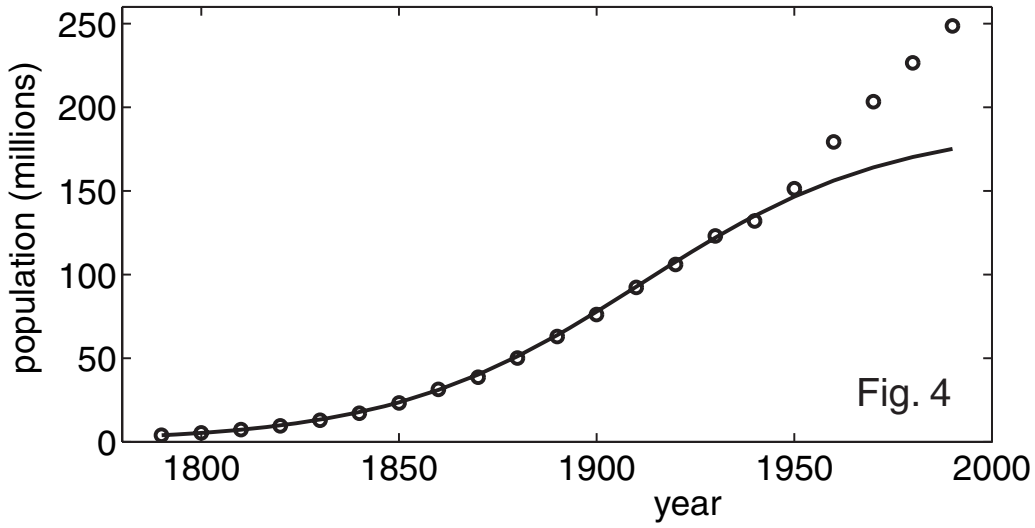
and plot $(1/P)dP/dt$ versus P, we can determine k and M.

Relative growth rate



U.S. population data, 1790 to 1910 — circles; best linear fit — (solid line).

Best logistic fit

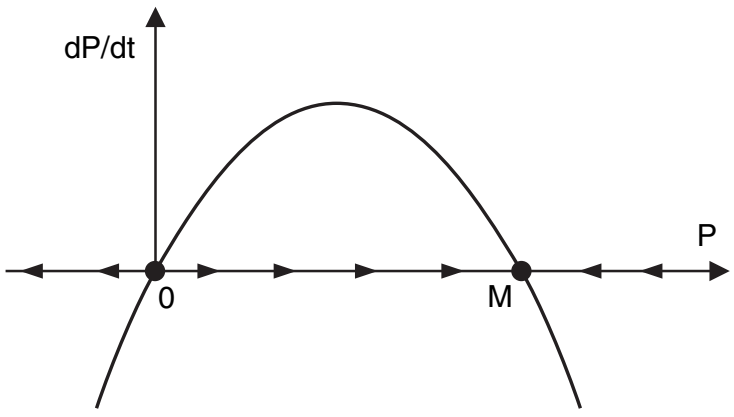


Logistic fit to U.S. data, 1790 to 1910 — solid line; actual population data, 1790 to 1990 — circles

7. Qualitative behavior of solutions; critical points and stability

Sometimes it's impossible find an explicit analytic solution. Nevertheless, much can be learned about the qualitative behavior of solutions from the differential equation itself.

Using logistic growth as an example, one thing that can be done is to plot the derivative dP/dt versus the population P. For the logistic equation, this looks like



Actually, we really don't even need to plot dP/dt; we can just indicate the size of dP/dt by the length of the arrows drawn along the P axis:

This version is called a **phase line**. The points $P = 0$ and $P = M$ are special: at these values of P, $dP/dt = 0$. These points are called **steady-states** or **critical points** of the differential equation.

From either graph, it's clear that $dP/dt > 0$ for $0 < P < M$, while $dP/dt < 0$ for $P > M$ and $P < 0$. This means that P is increasing in the range $0 < P < M$, while it is decreasing for $P > M$. Thus, all positive solutions tend to the final steady-state $P = M$ (negative population values don't really make sense). This asymptotic behavior can also be seen from the exact solution. The population value M is known as the **carrying capacity** of the system, and is the maximum sustainable population that this environment can support.

The critical point $P = 0$ is said to be **unstable** since small perturbations away from this critical point cause the solution to move farther away from it. (If one is right at the critical point, of course, one stays there forever; any deviations from this point are strongly amplified, however.) Similarly, the critical point at $P = 1$ is said to be **stable**, since any perturbations about this point are rapidly undone; deviations from a stable critical point are strongly damped.

Suppose we have a more general problem,

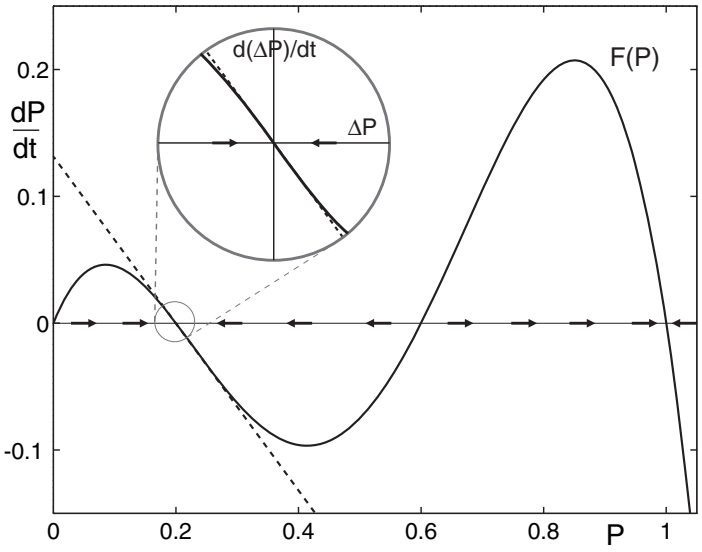
$$\frac{dP}{dt} = F(P)\,.$$

Finding the critical points and looking at the behavior near them is a good tool for understanding how the solutions of a differential equation behave. Determining critical points is relatively easy: one just solves for the zeros of $F(P)$; such solutions P_c will have $dP_c/dt = 0$.

Stability is also relatively easy. If we let $P = P_c + \Delta P$ near the critical point, where $\Delta P \ll P_c$, then since by definition $F(P_c) = 0$,

$$F(P) \approx F(P_c) + F'(P_c)\Delta P = F'(P_c)\Delta P\,.$$

This is just the tangent line approximation to $F(P)$ near P_c.



Near P_c, then, we are left with

$$\frac{d}{dt}(\Delta P) = F'(P_c)(\Delta P)\,. \tag{4}$$

This equation is linear, and is separable. The solution is

$$\Delta P(t) = \Delta P(0)e^{F'(P_c)t}\,.$$

If $F'(P_c) > 0$ the exponential grows as t increases, and therefore the critical point $P = P_c$ is unstable; if $F'(P_c) < 0$, then the exponential decays and the critical point is stable.

This process is called **linearizing** around a critical point and is also known as determining the **linearized stability**.

Exercise: Determine the critical points and their stability for the differential equation

$$\frac{dP}{dt} = -10P(P - 0.2)(P - 0.6)(P - 1.0)\,.$$

8. Introduction to systems of differential equations

Now consider the system of differential equations

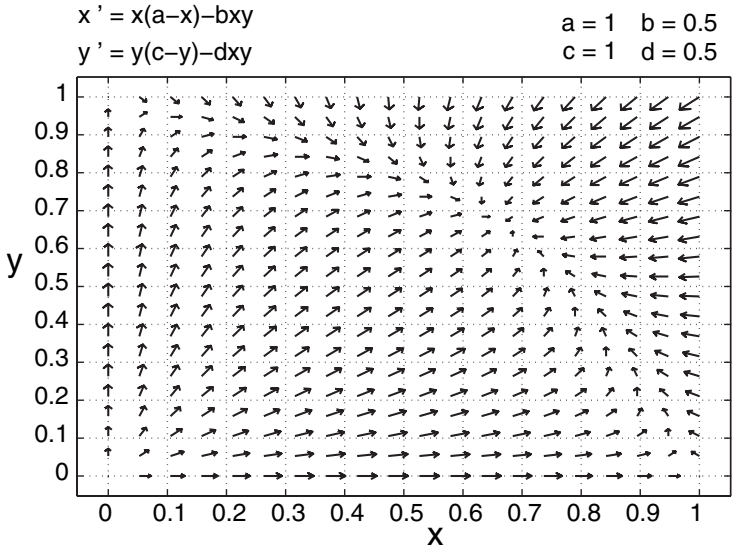
$$\frac{dx}{dt} = x(a-x) - bxy \tag{5a}$$

$$\frac{dy}{dt} = y(c-y) - dxy\,, \tag{5b}$$

where a, b, c and d are positive constants. Think of this as two logistic populations that compete; the more y's there are, the harder it is for x to increase its population, and vice-versus.

In this case, we generalize the phase line — we can use the x axis for indicating how fast the x population is changing, and the y axis for indicating how fast the y population is changing. We combine the x and y arrows into a single vector, $(dx/dt, dy/dt)$ that simultaneously indicates the rate at which both populations are changing.

When we do this, we get what is known as a **phase plane:**



By following the arrows in the phase plane, we can get a fairly good idea about the behavior of solutions.

Looking at the phase plane, we also see locations where the length of the arrows are zero. At these locations, $dx/dt = 0$ and $dy/dt = 0$, i.e., the solutions are constant. These again are **steady states** or **critical points**.

For the example system, with $a = 1$, $b = 1/2$, $c = 1$ and $d = 1/2$, by solving the differential equations with the derivatives set to zero,

$$x(1-x) - \frac{1}{2}xy = 0$$

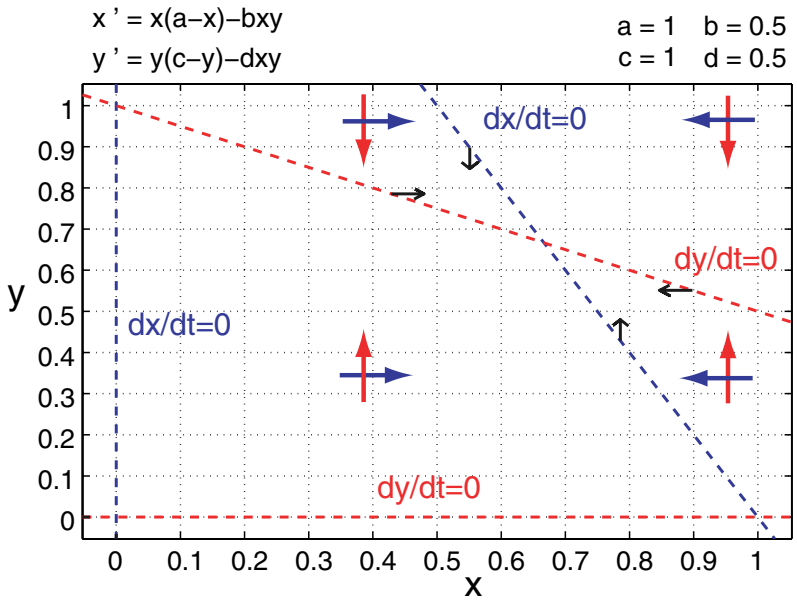
$$y(1-y) - \frac{1}{2}xy = 0\,,$$

we find the critical points $(0,0)$, $(1,0)$, $(0,1)$, and $(2/3, 2/3)$.

As for the phase line, we are also interested in the **stability** of these critical points. From the phase plane, it appears that $(0,0)$, $(1,0)$, and $(0,1)$ will all be unstable, and that $(2/3, 2/3)$ will be stable (the two populations can stably co-exist). We are going to have to develop some additional tools before we can verify this.

Before doing so, however, we note that another graphical way to analyze the behavior in the phase plane is to plot the **nullclines**: the curves in the phase plane along which $dx/dt = 0$ and $dy/dt = 0$. In this case, these curves are

$x = 0$, $y = 0$, $1 - x - y/2 = 0$ and $1 - y - x/2 = 0$:



By definition, dx/dt or dy/dt changes sign when one crosses one of the **nullclines**; this allows a quick graphical determination of the solution behavior.

Exercise: determine the critical points for the system

$$\frac{dx}{dt} = x(a-x) - bxy$$
$$\frac{dy}{dt} = y(c-y) - dxy\,,$$

with $a = 1$, $b = 2$, $c = 1$ and $d = 2$. What happens to the two competing populations in this case?

9. Linearized stability near a critical point

As for the case of the phase line, the way to do stability is to linearize the equations in the vicinity of the critical point. Suppose we have the system of equations

$$\frac{dx}{dt} = F(x,y) \tag{6a}$$
$$\frac{dy}{dt} = G(x,y)\,. \tag{6b}$$

The critical points in this case are just the points x_c and y_c that make the derivatives vanish, i.e., we want

$$F(x_c,y_c) = 0 \quad \text{and} \quad G(x_c,y_c) = 0\,. \tag{7}$$

To find the critical points, one therefore has to solve these two coupled equations for x_c and y_c.

Let's suppose that we've found the critical points, and we want to determine the stability of one of them, say (x_c, y_c). (There may be more than one, of course, but we examine their stability one at a time.) To do so, we look in the vicinity of the critical point by letting

$$x = x_c + \Delta x \quad \text{and} \quad y = y_c + \Delta y\,,$$

where Δx and Δy are both small.

When we then substitute these expressions into the differential equations, we first have, since x_c and y_c are both constant,

$$\frac{dx}{dt} = \frac{d}{dt}(\Delta x) \quad \text{and} \quad \frac{dy}{dt} = \frac{d}{dt}(\Delta y)\,.$$

We also need to simplify the right-hand sides of the equation. For example, we need $F(x, y) = F(x_c + \Delta x, y_c + \Delta y)$. But note that the result for the differential of a function of two variables tells us that

$$\Delta F = F(x_c + \Delta x, y_c + \Delta y) - F(x_c, y_c) \approx F_x(x_c, y_c)\Delta x + F_y(x_c, y_c)\Delta y \,,$$

where

$$F_x \equiv \frac{\partial F}{\partial x} \quad \text{and} \quad F_y \equiv \frac{\partial F}{\partial y} \,.$$

In addition, recall $F(x_c, y_c) = 0$ because (x_c, y_c) is an equilibrium point, so

$$F(x_c + \Delta x, y_c + \Delta y) \approx F_x(x_c, y_c)\Delta x + F_y(x_c, y_c)\Delta y \,.$$

Similarly, for $G(x, y)$ near the critical point we have

$$G(x_c + \Delta x, y_c + \Delta y) \approx G_x(x_c, y_c)\Delta x + G_y(x_c, y_c)\Delta y \,.$$

Thus, finally we obtain the linearized system of equations

$$\frac{d}{dt}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} F_x & F_y \\ G_x & G_y \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \,.$$

(Aside: the matrix of partial derivatives,

$$\mathbf{A} = \begin{bmatrix} F_x & F_y \\ G_x & G_y \end{bmatrix} \,,$$

is called the *Jacobian* of the functions F and G.)

It's important to note that the partial derivatives appearing in linearized system of equations,

$$\frac{d}{dt}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} F_x & F_y \\ G_x & G_y \end{bmatrix}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \mathbf{A}\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \,. \tag{8}$$

are all *constant*, since they are all evaluated at the critical point (x_c, y_c).

Let's get rid of the Δ's by writing $x_1 = \Delta x$ and $x_2 = \Delta y$. In addition, we will write $\mathbf{x}$ for the vector with components x_1 and x_2, i.e.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \,.$$

10. Introduction to the eigenvalue method

As an example, consider the two coupled differential equations

$$\frac{dx_1}{dt} = 4x_1 - 5x_2, \quad \frac{dx_2}{dt} = 2x_1 - 3x_2 \,.$$

If we write

$$\mathbf{A} = \begin{bmatrix} 4 & -5 \\ 2 & -3 \end{bmatrix}$$

then the equations become, in matrix form,

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} \,.$$

How do we find the solution of this equation? A scalar equation, such as

$$\frac{dx}{dt} = ax \,, \quad x = x_0 \text{ at } t = 0 \,,$$

has the solution

$$x(t) = e^{at} x_0 \,.$$

By analogy, we could guess the solution of the vector equation is

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}_0 \,,$$

but we don't know how to interpret the exponential of a matrix.

Therefore we will instead look for solutions of the form

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$$

where λ is a scalar and $\mathbf{v}$ is a constant vector, or equivalently

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = e^{\lambda t} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \,.$$

Note that the **main idea** is to separate the time dependence from the vector nature of the equation: we use a *scalar* time dependent term in combination with a *vector* constant. Since with this choice for $\mathbf{x}$ we have

$$\frac{d\mathbf{x}}{dt} = \lambda e^{\lambda t}\mathbf{x} \,,$$

we get

$$\begin{aligned} \lambda e^{\lambda t}\mathbf{v} &= \mathbf{A}e^{\lambda t}\mathbf{v} \\ \Rightarrow \mathbf{A}\mathbf{v} &= \lambda\mathbf{v}\,. \end{aligned}$$

This is the fundamental equation for the **eigenvalue** λ and the **eigenvector** $\mathbf{v}$. We can also write this as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0\,,$$

where $\mathbf{I}$ is the identity matrix (i.e., ones along the diagonal, and zeros elsewhere), so that $\mathbf{v}$ must lie in the nullspace of the matrix

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{bmatrix}.$$

For most values of λ the equation

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$$

has as its only solution $\mathbf{v} = \mathbf{0}$. It is only for particular values of λ that we will have non-zero solutions, and these are the values of λ in which we are interested (the term 'eigen' is just German for 'special' or 'particular').

Note to have non-zero solutions for $\mathbf{v}$ the matrix $\mathbf{A} - \lambda\mathbf{I}$ must be singular. If a matrix is singular, its determinant is zero. Therefore, we are looking for the values of λ which make

$$\det(\mathbf{A}-\lambda\mathbf{I}) = 0\,.$$

This is called the **characteristic equation** for the matrix $\mathbf{A}$, and λ will be an eigenvalue if and only if it is satisfied.

In our example

$$\begin{aligned} \det(\mathbf{A}-\lambda\mathbf{I}) &= \begin{vmatrix} 4-\lambda & -5 \\ 2 & -3-\lambda \end{vmatrix} = (4-\lambda)(-3-\lambda)+10 \\ &= \lambda^2 - \lambda - 2 = (\lambda+1)(\lambda-2) = 0 \end{aligned}$$

so $\lambda_1 = -1$ and $\lambda_2 = 2$ are the two eigenvalues. To find the corresponding eigenvectors, we go back to $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = 0$.

When $\lambda = \lambda_1 = -1$ we get

$$\begin{bmatrix} 5 & -5 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \;\Rightarrow\; \begin{bmatrix} 5 & -5 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so v_2 is free (i.e., undetermined) and

$$\mathbf{v} = v_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

We see that the eigenvectors are not uniquely determined; any multiple of $[1\ 1]^T$ will work (any vector in the nullspace of $\mathbf{A} - \lambda\mathbf{I}$ works). Since the multiplier is arbitrary, we are free to choose anything that is convenient. Here we will take

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} .$$

(In mathematical terms, this vector is a basis for the nullspace, which is another way of saying that any scalar multiple of it is also a solution.) It's always good to remember that we can always multiply an eigenvector by an arbitrary constant and it will still be an eigenvector.

Similarly, when $\lambda = \lambda_2 = 2$ we get

$$\begin{bmatrix} 2 & -5 \\ 2 & -5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{v} = v_2 \begin{bmatrix} 5/2 \\ 1 \end{bmatrix} \Rightarrow \mathbf{v}_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix} .$$

Returning to the differential equation, we have therefore found two pure exponential solutions

$$\mathbf{x}_1 = c_1 e^{\lambda_1 t}\mathbf{v}_1 = c_1 e^{-t} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

and

$$\mathbf{x}_2 = c_2 e^{\lambda_2 t}\mathbf{v}_2 = c_2 e^{2t} \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

(the constants c_1 and c_2 are put in merely because any multiples of the eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are also eigenvectors).

These two solutions are special, in that when the two components are in the proper ratios with one another (as shown by the components of the eigenvectors), then the time dependence of both variables in the coupled system becomes identical. For example, when $x_1 = x_2$, the differential equations

$$\frac{dx_1}{dt} = 4x_1 - 5x_2 ,$$

$$\frac{dx_2}{dt} = 2x_1 - 3x_2 ,$$

become

$$\frac{dx_1}{dt} = -x_1\,,$$
$$\frac{dx_2}{dt} = -x_2\,,$$

so that $x_1 = x_2 = c_1 e^{-t}$. Similarly, when $2x_1 = 5x_2$, the differential equations become

$$\frac{dx_1}{dt} = 2x_1\,,$$
$$\frac{dx_2}{dt} = 2x_2\,,$$

so that $x_1/5 = x_2/2 = c_2 e^{2t}$. Looking for eigenvalues and eigenvectors is equivalent to looking for combinations of the equations that produce such simple time dependencies.

What do we do next? Well, because the systems of differential equations is **linear** and **homogeneous** (i.e., there is no forcing function on the right-hand side), the solutions obey the **principle of superposition**: if we have two solutions, we can construct a new one by taking a linear combination of them.

Superposition: if $\mathbf{x}_1$ and $\mathbf{x}_2$ are solutions, then so is $c_1\mathbf{x}_1 + c_2\mathbf{x}_2$ (where c_1 and c_2 are constants). This follows since

$$\frac{d}{dt}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2) = c_1\frac{d\mathbf{x}_1}{dt} + c_2\frac{d\mathbf{x}_2}{dt} = c_1\mathbf{A}\mathbf{x}_1 + c_2\mathbf{A}\mathbf{x}_2$$
$$= \mathbf{A}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2)\,.$$

How do we know we have enough solutions? Well, because we started with two scalar equations, we have two derivatives. That means we expect two integration constants (if we could integrate the two equations directly, we would have directly obtained two added constants of integration). Thus, we expect a solution with two arbitrary constants; we found two solutions using eigenvalues and eigenvectors, so the linear combination of the two should give the full solution.

(Note: it is possible that sometimes gets the same solution twice, rather than

two different solutions. Two such solutions are said to be *linearly dependent*. What we want are two solutions that are different, i.e., *linearly independent*. A basic way to find out of two solutions are linearly independent is to try to apply arbitrary initial conditions. If one can always solve for the constants given any initial condition, then the two solutions are linearly independent.)

Back to the problem at hand, we had two eigenvalues and two eigenvectors. Because the equation is linear and homogeneous any linear combination of the two is also a solution, so that

$$\mathbf{x} = c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2$$

should be the general solution. Again, we have two arbitrary constants, c_1 and c_2, and two initial conditions to satisfy, so the counting is right.

Finally, we must satisfy any initial conditions, if we have them. Suppose at $t = 0$ we want $\mathbf{x}(0) = \mathbf{x}_0 = [8 \;\; 5]^T$ (this is really two conditions, of course)

$$\begin{aligned}
&\Rightarrow\; c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = \mathbf{x}_0 \\
&\Rightarrow\; c_1\begin{bmatrix}1\\1\end{bmatrix} + c_2\begin{bmatrix}5\\2\end{bmatrix} = \begin{bmatrix}8\\5\end{bmatrix} \\
&\Rightarrow\; \begin{bmatrix}1 & 5\\1 & 2\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix} = \begin{bmatrix}8\\5\end{bmatrix} \\
&\Rightarrow\; c_1 = 3 \text{ and } c_2 = 1\,, \\
&\Rightarrow\; \mathbf{x}(t) = 3e^{-t}\begin{bmatrix}1\\1\end{bmatrix} + e^{2t}\begin{bmatrix}5\\2\end{bmatrix}\,,
\end{aligned}$$

or equivalently,

$$\begin{aligned}
x_1(t) &= 3e^{-t} + 5e^{2t}\,, \\
x_2(t) &= 3e^{-t} + 2e^{2t}\,.
\end{aligned}$$

11. A phase plane gallery

Let's look at a generic linear system

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} \qquad \text{where} \qquad \mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

To compute the eigenvalues, we want

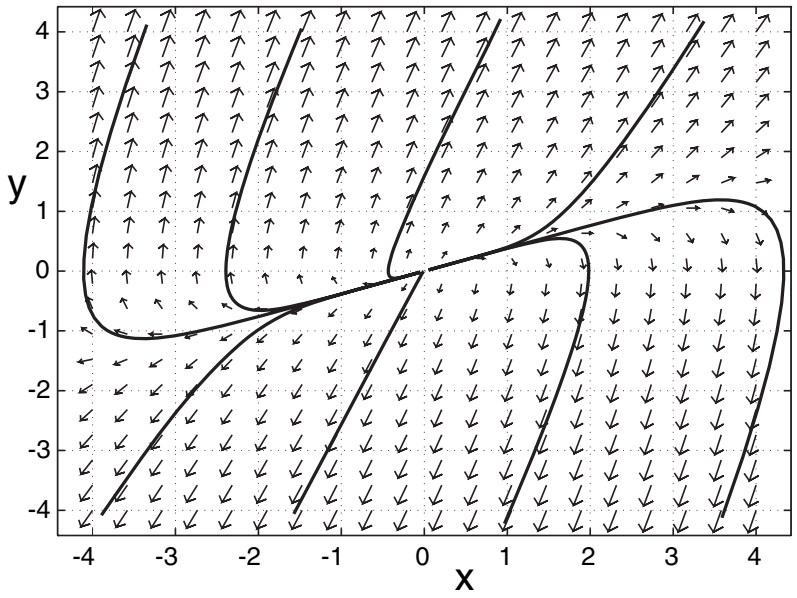
$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} a-\lambda & b \\ c & d-\lambda \end{bmatrix} = \lambda^2 - (a+d)\lambda + (ad-bc) = 0\,.$$

The solution of this quadratic equation is

$$\lambda = \frac{a+d}{2} \pm \frac{1}{2}\sqrt{(a+d)^2 - 4(ad-bc)} = \frac{a+d}{2} \pm \frac{1}{2}\sqrt{(a-d)^2 + 4bc}\,.$$

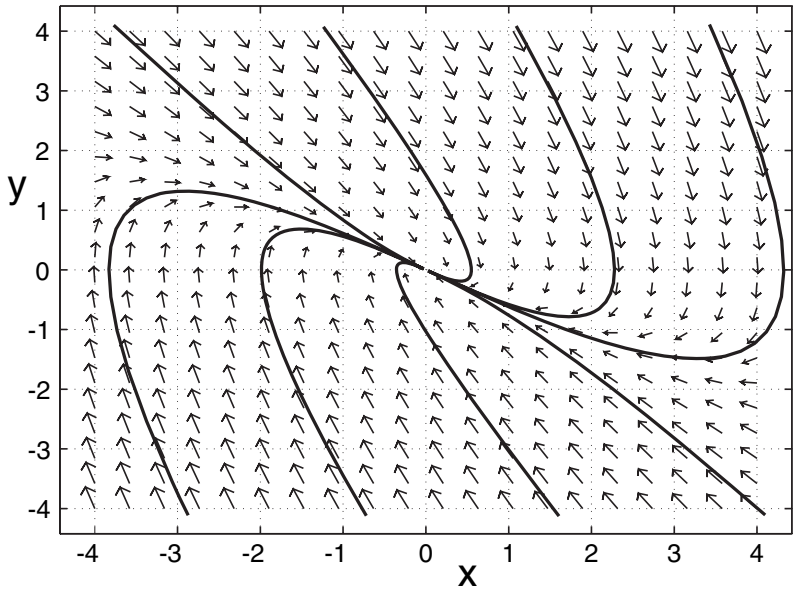
The rest of the details depend upon how the computation for the eigenvectors goes, but the main qualitative behavior of the system, and how its phase plane looks, depends upon the eigenvalues alone. We have only a few cases.

Case 1: both eigenvalues real and positive: unstable node



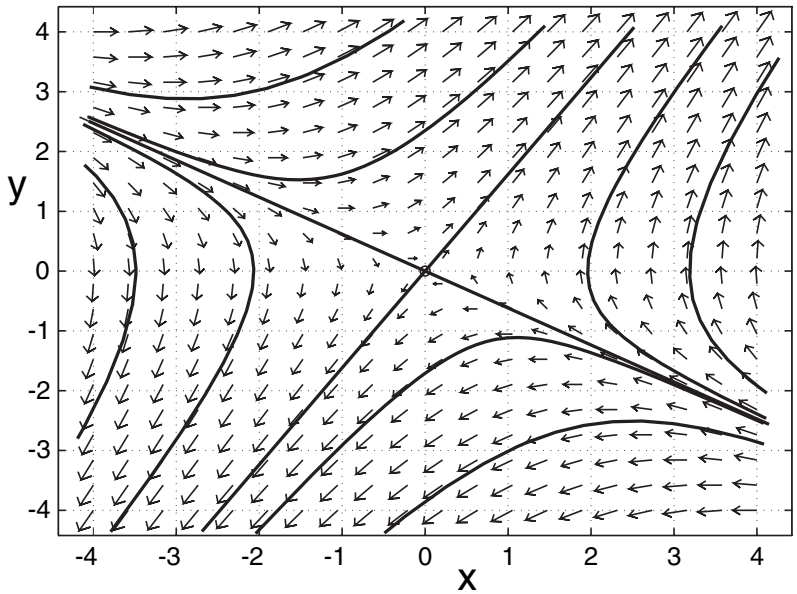
In this case, one eigenvalue is larger than the other, so it usually ends up dominating the solution.

Case 2: both eigenvalues real and negative: stable node



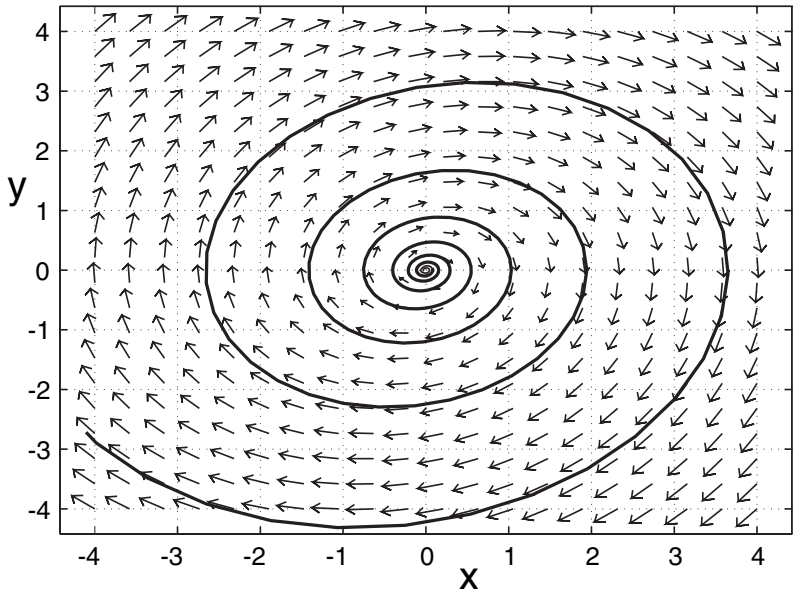
In this case, the eigenvalue that is least negative ends up being the dominant behavior (slowest decay).

Case 3: one positive and one negative eigenvalue: saddle point



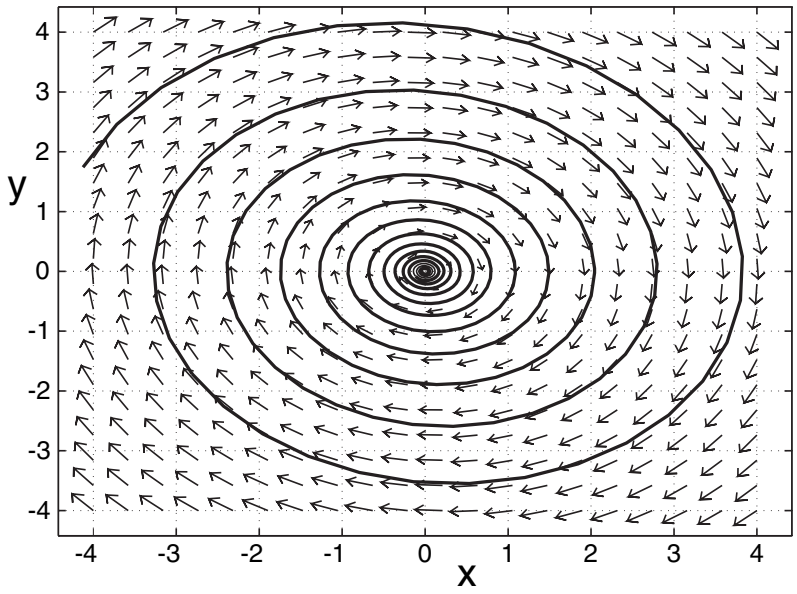
A saddle point is unstable (usually this is not included in the name, however).

Case 4: Complex roots, positive real part: unstable spiral



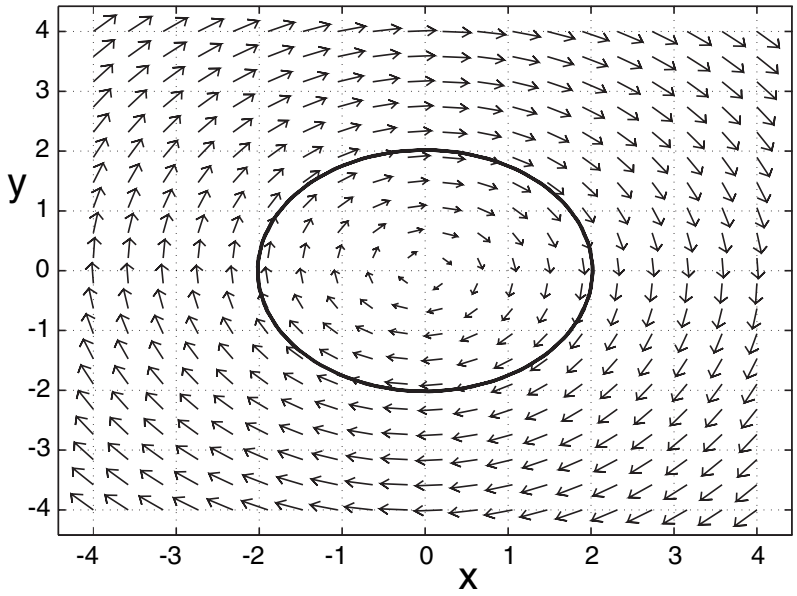
The direction of the spiral is best determined by looking at a couple of specific points in the phase plane (it's harder to do it from eigenvectors).

Case 5: Complex roots, negative real part: stable spiral



Another name for a stable spiral is a spiral sink (unstable: spiral source).

Case 6: Complex roots, zero real part: center



This is sometimes termed neutrally stable. Technically speaking, the stability is indeterminant (determined by terms neglected during the linearization).

Example: Consider again the system of competing logistic populations

$$\begin{aligned}\frac{dx}{dt} &= x(1-x) - \frac{1}{2}xy \\ \frac{dy}{dt} &= y(1-y) - \frac{1}{2}xy\,.\end{aligned}$$

Let's consider the critical point at $(2/3, 2/3)$. The Jacobian associated with the linearization at this point is

$$\begin{bmatrix} 1-2x-\frac{1}{2}y & -\frac{1}{2}x \\ -\frac{1}{2}y & 1-2y-\frac{1}{2}x \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{2}{3} \end{bmatrix}$$

The eigenvalues are

$$\lambda = -\frac{2}{3} \pm \frac{1}{2}\sqrt{\frac{4}{9}} = -\frac{2}{3} \pm \frac{1}{3} \quad \Rightarrow \quad -1 \text{ and } -\frac{1}{3}$$

Since both are negative, this critical point is a stable node.

Exercise: do the rest of the critical points for this case. Also do them for the case where the coupling coefficient is 2 instead of $1/2$.

12. Numerical methods

So, how do the numerical solution programs work? They use various algorithms to simulate numerically an equation or a system of equations. These methods are essential if one can't solve an equation analytically; this is what happens most of the time when one is doing applications. We will look at a few of the simplest methods for getting numerical solutions.

Euler's method

Suppose we have a general first-order differential equation,

$$\frac{dy}{dx} = f(x, y) \quad \text{with} \quad y(x_0) = y_0 \,.$$

This will be assumed to be a scalar equation in what follows, but this is not necessary. Systems of equations (and therefore second-order equations) can be treated in exactly the same way.

Euler's method, or the tangent line method, is the simplest way to get a solution. The idea is that since one knows $y = y_0$ at $x = x_0$, from the equation one knows dy/dx. Since one knows the initial slope, one can then

use a tangent line approximation to get to a nearby point, say $x = x_0 + h$. This gives

$$y(x_0 + h) \approx y(x_0) + hf(x_0, y(x_0)) \,.$$

Now one knows y at a new point, so from the differential equation one knows the slope there, so a new tangent line can be constructed to get to a new point (say $x_0 + 2h$), and so on.

To simplify the notation, define

$$x_{n+1} = x_n + h \quad \text{or} \quad x_n = x_0 + nh \,,$$

and write

$$y_n = y(x_n) \,.$$

Then the general procedure can be written

$$y_{n+1} = y_n + hf(x_n, y_n) \quad \text{with} \quad y_0 \quad \text{given.}$$

Example:

Let's use this with the simple example

$$\frac{dy}{dx} = y, \quad y(0) = 1$$

which has the solution

$$y(x) = e^x$$

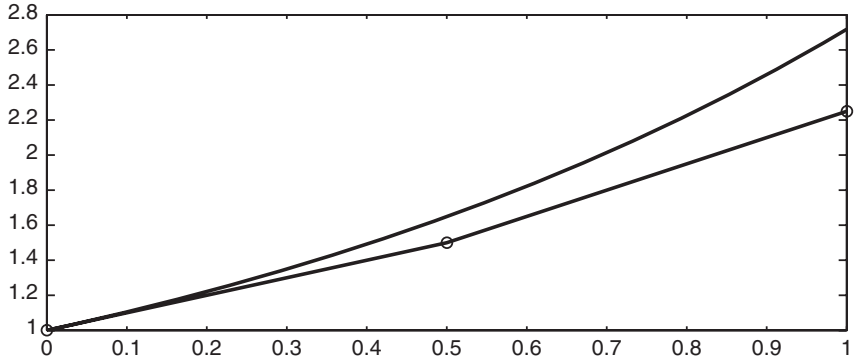
to try to approximate e. If we use a stepsize of $h = 1/2$, we have

$$y_1 = 1 + hy_0 = 1 + 1/2 = 1.5$$

and then

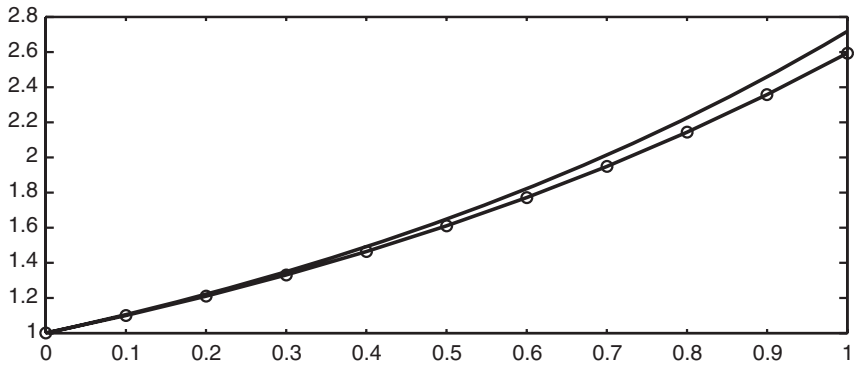
$$y_2 \approx y(1) = y_1 + hy_1 = 1.5 + 0.5 * 1.5 = 2.25$$

This is obviously not a very good approximation to e, as can be seen here:



It's clear from looking at the figure where the error is coming from in the approximation: the true solution curves away from the straight line approximation.

Of course, if we take a smaller step size h, the numerical approximation gets better because the two curves are not allowed to get as far away from one another before the slope is recalculated. For example, if we take $h = 0.1$, the final result after doing the 10 steps is 2.593742460. The entire solution curve is shown here:



Let's try to do this one more time, now taking $h = 0.01$. The final result after doing the 100 steps is now 2.704813834. This is clearly starting to get close

to the exact value, but as you can see a small stepsize is required when one uses Euler's method in order to get a reasonably accurate result. This is one of the disadvantages of Euler's method, and why people like to use "improved" or more accurate methods.

As you can see from the above, one must take a relatively small step size h in order to make the error made in the numerical solution small. Summarizing:

h	approx.	error	error/h
0.1	2.593742460	0.124539368	1.245394
0.01	2.704813834	0.013466994	1.346699
0.001	2.716923932	0.001357896	1.357896

Roughly speaking, when one reduces the step size by a factor of 10, the error is reduced by a factor of 10 as well. Thus, the error seems to be proportional to the step size h.

One can get an idea of what the error should be by looking at the error made in one step, called the **local truncation error**. Recall

$$y_{n+1} = y_n + hf(x_n, y_n)$$

and we want to compare this with the exact answer when one starts at y_n at $x = x_n$, which we will write as $y(x)$ (now we must distinguish y_n which is the approximation at x_n, with $y(x_n)$, which is the true solution at y_n). Using a Taylor series approximation,

$$y(x_{n+1}) = y(x_n + h) = y(x_n) + hy'(x_n) + \frac{h^2}{2}y''(x_n) + \dots .$$

Note also that using the differential equation

$$y'(x_n) = f(x_n, y(x_n)),$$

and if we want to we can evaluate $y''(x_n)$ as well by differentiating the differential equation implicitly. Taking the difference, we then have

$$d_{n+1} = y_{n+1} - y(x_{n+1}) = -\frac{h^2}{2}y''(x_n) + \dots .$$

Thus, we see that the error per step is of size h^2. To numerically integrate an equation from $x = x_0$ to $x = x_f$ with N steps, we have $h = (x_f - x_0)/N$. If we make an error of size h^2 per step, and have N steps, then the total error

will be of size

$$Nh^2 = \frac{x_f - x_0}{h}\, h^2 = (x_f - x_0)\, h\,.$$

Of course, there should be some constant multiplying this (coming from the $y''(x_n)/2$ multiplying the local error), but even so one can get an idea of the trend of the error from the above formula: for fixed x_0 and x_f the error is proportional to h, as expected.

Improved Euler method

Rather than taking smaller and smaller values of the stepsize h to get more accurate numerical solutions to a differential equation, a better idea is to improve the formula used to generate the numerical solution.

Consider the differential equation

$$y'(x) = f(x, y), \qquad y(x_0) = y_0$$

and let $y(x)$ denote its solution. By integrating the above equation from x_n to x_{n+1} we have

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x))\, dx\,.$$

This can be thought of as just another way to write the differential equation.

We then get various approximate numerical solutions for the differential equation by using various numerical approximations for the integral. For example, if we use the left endpoint to approximate the integral, we get Euler's forward method, or *forward Euler:*

$$y_{n+1} = y_n + hf(x_n, y_n)$$

We can also use the right endpoint, which gives us *backward Euler:*

$$y_{n+1} = y_n + f(x_{n+1}, y_{n+1})$$

Note that this gives an implicit equation for y_{n+1} which must be solved for with some method like a Newton iteration. Such methods are called *implicit* methods, and because one must in general solve a nonlinear equation at each step they generally viewed as difficult to use than *explicit* methods (i.e., ones in which y_{n+1} only appears on the left-hand side of the equation). Nevertheless, implicit methods often have advantages (namely, better stability properties) in comparison to explicit methods.

Continuing the argument, we know that a trapezoidal approximation to an integral is better than just using rectangles. Therefore, a more accurate method

should result if we use the trapezoidal rule, which gives:

$$y_{n+1} = y_n + \frac{f(x_n, y_n) + f(x_{n+1}, y_{n+1})}{2}\, h$$

which is known as the midpoint method. This method is more accurate, but is still implicit. An explicit method can be created by approximating y_{n+1} in the above by the result obtained from the forward Euler method. This gives

$$y_{n+1} = y_n + \frac{f(x_n, y_n) + f(x_{n+1}, y_n + hf(x_n, y_n))}{2}\, h\,.$$

This method is known as the improved Euler formula. Its local truncation error is larger than the midpoint method, but both truncation errors are still of size h^3, which gives a total error in each case proportional to

$$(x_f - x_0)\, h^2$$

(The improved Euler method has a larger error only because the constant multiplying the local truncation error is larger than that for the midpoint method.)

Here are some numerical results for our simple example using the improved Euler method:

h	approx.	error
0.1	2.714080847	0.004200982
0.01	2.718236863	0.000044966
0.001	2.718281376	0.000000453

As you can see, the error goes down by a factor of 100 each time the step size h is reduced by a factor of 10. This is accurate enough that a graph of the numerical solution using 10 steps (a stepsize h of 0.1) is already virtually indistinguishable from the exact solution.

Before continuing, we note that the improved Euler method can also be written as the two-step iteration

$$\begin{aligned} u_{n+1} &= y_n + hf(x_n, y_n)\,, \\ y_{n+1} &= y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_{n+1}, u_{n+1})\right]\,. \end{aligned}$$

In this form the first step is thought as the **predictor** and the second step is thought of as a **corrector**. In addition, the whole process can alternatively be written

$$\begin{aligned} k_1 &= f(x_n, y_n)\,, \\ k_2 &= f(x_{n+1}, y_n + hk_1)\,, \\ y_{n+1} &= y_n + \frac{h}{2}(k_1 + k_2)\,. \end{aligned}$$

Another way of improving Euler's method is merely to use more terms of the Taylor series to construct a one-step approximation that is better than the tangent line. The Runge-Kutta method and its variants is probably used more often than this method, however.

The Runge-Kutta Method

The Runge-Kutta method can be thought of as an extended version of the improved Euler method. The method is:

$$y_{n+1} = y_n + \frac{h}{6}\left[k_1 + 2k_2 + 2k_3 + k_4\right]$$

where

$$\begin{aligned} k_1 &= f(x_n, y_n)\,, \\ k_2 &= f(x_n + h/2, y_n + hk_1/2)\,, \\ k_3 &= f(x_n + h/2, y_n + hk_2/2)\,, \\ k_4 &= f(x_n + h, y_n + hk_3)\,. \end{aligned}$$

The method is somewhat difficult to motivate, and it is even more difficult to determine the truncation error by expanding things using Taylor series because of the large number of terms required. One way to understand the idea, though, is to note that if going from one function evaluation per step (Euler) to two function evaluations per step (implicit Euler) increases the accuracy

of the numerical method from $O(h)$ to $O(h^2)$, then adding two more should increase it to $O(h^4)$, which is indeed the case. Once one gets this idea, one can determine the details by considering four successive function evaluations each of the form $f(x_n + \alpha_i h, y_n + \beta_i k_{i-1})$, where $i = 1 \ldots 4$, and then substituting, expanding, and picking the constants α_i and β_i to make the truncation error as small as possible.

Another way to understand things is to note that in the case where $f(x, y)$ does not depend upon y, the above reduces to

$$y_{n+1} = y_n + \frac{h}{6}[f(x_n) + 4f(x_n + h/2) + f(x_n + h)]$$

which is Simpson's rule. Simpson's rule, of course, has a one-step error of h^5. It therefore is reasonable that the Runge-Kutta method should have a local truncation error proportional to h^5, and thus a total error proportional to $(x_f - x_0)h^4$. (Note that Edwards and Penney motivate the Runge-Kutta method by starting with Simpson's rule, but it is still not clear from their explanation exactly where the particular constants come from.)

The high accuracy of the method and relative ease of evaluation make it a very popular method. For our simple example, here are the results:

h	approx.	error
0.1	2.718279744	2.08E-6
0.01	2.718281828	2.25E-10
0.001	2.718281828	2.26E-14

As can be easily seen, the error drops by a factor of 10,000 each time the step size is reduced by a factor of 10.

Some difficulties with numerical methods

One of the problems of working with numerical approximations to the solutions of differential equations is the fact that computers only do calculations with finite precision. If one does not keep enough digits when doing calculations, the accuracy of the approximation can be lost. In fact, since one makes an error each time an arithmetic operation is done (addition, subtraction, multiplication, etc.), it is possible that taking smaller and smaller step sizes only makes things worse after a point: one makes more and more errors which add up to make the final result worse.