Lecture 19: 2.7 Numerical Approximation: Euler's method. Most differential equations of the form

$$\frac{dy}{dt} = f(t,y), \qquad y(t_0) = y_0 \tag{2.7.1}$$

can not be solved analytically. Only in special cases like the linear case or the separable case can we obtain an explicit formula for the solution in terms of integrals. Still we know from the existence theorem that there is a solution for some time.
We have already seen that we can get some information about how the solution looks like from the direction field: Given a graph of the direction field we can try to plot a solution curve that is always tangential to the direction field. However, this method is not very exact since the direction field is only plotted at a few point and the curve we try to plot might not pass close to these points. There is however, a more quantitative **numerical** version of this method, called **Euler's method** or the **tangent line method**:

Let us consider how we might approximate a solution curve $y = \phi(t)$ of (2.7.1), near $t = t_0$. We know that the solution curve passes through the point (t_0, y_0) in the $t-y$ plane and, from (2.7.1) we also know the slop at this point $f(t_0, y_0)$. Thus we can write down an equation for the tangent line to the solution curve at (t_0, y_0):

$$y = y_0 + f(t_0, y_0)(t - t_0) \tag{2.7.2}$$

The tangent line is a good approximation to the solution curve on a short time interval. Thus if t_1 is close enough to t_0 we can approximate $\phi(t_1)$ by

$$y_1 = y_0 + f(t_0, y_0)(t_1 - t_0) \tag{2.7.3}$$

To proceed further we can repeat the process. Unfortunately, we do not know the value $\phi(t_1)$ of the solution at time t_1. The best we can do is to use the approximation y_1 instead. Thus we construct a line through (t_1, y_1) with slope $f(t_1, y_1)$:

$$y = y_1 + f(t_1, y_1)(t - t_1) \tag{2.7.3}$$

and we approximate $\phi(t_2)$ at a nearby point t_2 by

$$y_2 = y_1 + f(t_1, y_1)(t_2 - t_1). \tag{2.7.4}$$

Continuing in this manor we define

$$y_{n+1} = y_n + f(t_n, y_n)(t_{n+1} - t_n)$$

Finally, if we assume that we always take the same small step h in the time direction:

$$y_{n+1} = y_n + f(t_n, y_n)h, \qquad t_{n+1} = t_n + h \tag{2.7.5}$$

The values y_n are approximations for $\phi(t_n)$, the value of the true solution at the times t_n. We can then approximate the solution curve $\phi(t)$ by the **polygonal curve** consisting of the line segments between the points (t_n, y_n) and (t_{n+1}, y_{n+1}), for $n = 0, \ldots$. This polygonal curve is not exactly the solution curve but the hope is that it will converge to it as the time step size $h \to 0$.

Ex Use Euler's method to approximately find the value $y(1)$ of the solution of

$$\frac{dy}{dt} = y, \qquad y(0) = 1$$

Use the step size $h = 1/m$, and determine how the error decreases with h (the true solution is $e^1 \sim 2.7183$)
Sol The general formula is that

$$y_{n+1} = y_n + y_n h = y_n(1+h), \qquad t_{n+1} = t_n + h, \qquad t_0 = 0, \quad y_0 = 1 \tag{2.7.6}$$

and hence

$$y_n = (1+h)^n, \qquad t_n = nh \tag{2.7.7}$$

If we pick the step size $h = 1/m$ then

$$y_m = (1 + 1/m)^m, \qquad t_m = m \cdot 1/m = 1$$

We know that

$$(1 + 1/m)^m \to e, \qquad m \to \infty$$

so we expect it to be a good approximation for small step size. If we put $m = 2^k$, $k = 6, 7, 8$ we get

$$(1 + 1/64)^{64} \sim 2.6973.., \quad (1 + 1/128)^{128} \sim 2.7077..., \quad (1 + 1/256)^{256} \sim 2.71230..$$

The error is hence in the three cases

$$0.024..., \qquad 0.010..., \qquad 0.005...$$

One can actually check that the error is linear in h i.e. proportional to h, in fact since $1/256 \sim 4/1000 = 0.0025$ we see that the error is approximately $2h$. It is actually true in general for Euler's method that the error is linear in h.

Note that in general it is not as easy as above to calculate the numerical approximation and (2.7.6) does not simplify to something as simple as (2.7.7) but one has to calculate each step individually. The example above was just meant to illustrate that the method works.

2.8 The existence and uniqueness theorem and Picard's iteration scheme. As pointed out earlier, most differential equations of the form

(2.8.1) $$\frac{dy}{dt} = f(t, y), \qquad y(t_0) = y_0$$

can not be solved analytically. Only in special cases like the linear case or the separable case can we obtain an explicit formula for the solution in terms of integrals. In general we can however say that there is a local solution in some time interval $t_0 - h < t < t_0 + h$. To prove this one iteratively constructs a sequence of function and show that the sequence converges to a solution. This is called **Picard's iteration** or **method of successive approximation**. Let us first rewrite (2.8.1) as an integral equation by integrating it:

(2.8.1) $$y(t) = \int_{t_0}^{t} f(s, y(s))\, ds + y_0$$

It looks like we solved the problem but the unknown function y is in the integral in the right hand side so we can not calculate it exactly without the knowledge of $y(t)$. Therefore we make a successive approximation, starting with y_0 and defining

(2.8.2) $$y_{n+1} = \int_{t_0}^{t} f(s, y_n(s))\, ds + y_0, \qquad n \geq 0.$$

The hope is that the functions $y_n(t)$ will converge to a function $y(t)$ that is a solution of (2.8.1).

Ex Use Picard iteration to find the solution of

$$\frac{dy}{dt} = y, \qquad y(0) = y_0$$

Sol Let $y_0 = 1$ and

$$y_t = 1 + \int_0^t 1\, ds = 1 + t$$

and

$$y_2(t) = 1 + \int_0^t (1 + s)\, ds = 1 + t + \frac{t^2}{2}$$

and so on, in general we obtain:

$$y_3(t) = 1 + \int_0^t \left(1 + s + \frac{s^2}{2}\right) ds = 1 + t + \frac{t^2}{2} + \frac{t^3}{3 \cdot 2}$$

In general we obtain

$$y_n(t) = 1 + \frac{t^2}{2} + \frac{t^3}{3 \cdot 2} + ... + \frac{t^n}{n!}$$

Using the ratio test we can prove that this converges as $n \to \infty$:

$$y(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}$$

which of course is just the Taylor series for the exponential function e^t.

Let

$$T(y)(t) \equiv \int_{t_0}^{t} f(s, y(s))\, ds + y_0$$

Then we want to find y such that $T(y) = y$. We are trying to find y as a limit of an iteration $y_0(t) = y_0$ and for $n \geq 0$; $y_{n+1} = T(y_n)$.

Contractions. A map $T : W \to W$ is called a *contraction*, if for $x, y \in W$:

$$\|T(x) - T(y)\| \le K\|x - y\|, \qquad K < 1 \tag{1}$$

A point $x \in W$ is called a *fixed point* if $T(x) = x$. We have:

Lemma. *Let $T : W_0 \to W_0$ be a contraction of a complete normed space W_0. Then T has a unique fixed point $x \in W_0$. In fact for any $x_0 \in W_0$, $x_k = T^k(x_0) = T \circ \cdots \circ T(x_0)$ (k times) converges to x; $\|x - x_k\| \to 0$, as $k \to \infty$.*

Proof. Using (1) repeatedly we get

$$\|x_{k+1} - x_k\| = \|T(x_k) - T(x_{k-1})\| \le K\|x_k - x_{k-1}\| \le \cdots \le K^k\|x_1 - x_0\| \tag{2}$$

Here $\|x_1 - x_0\| = \|T(x_0) - x_0\| = C$ is a fixed constant. For $m > k$ we write $x_m - x_k = (x_m - x_{m-1}) + (x_{m-1} - x_{m-2}) + ... + (x_{k+1} - x_k)$ and estimate the norm of each term by (2):

$$\|x_m - x_k\| \le \|x_m - x_{m-1}\| + \cdots + \|x_{k+1} - x_k\| \le (K^{m-1} + \cdots + K^{k-1})C$$

This is a geometric sum and since $K < 1$ the infinite sum converges; $\sum_{\ell=k-1}^{m-1} K^\ell \le \sum_{\ell=k-1}^{\infty} K^\ell = K^{k-1}\sum_{n=0}^{\infty} K^n = K^{k-1}/(1-K)$. Hence

$$\|x_m - x_k\| \le \varepsilon(N) = \frac{CK^{N-1}}{1-K}, \qquad \text{if} \quad m, k \ge N,$$

where $\varepsilon(N) \to 0$ as $N \to \infty$, i.e. x_k is a Cauchy sequence.

The uniqueness follows from (1); if $T(x) = x$ and $T(y) = y$ then $\|x - y\| = \|T(x) - T(y)\| \le K\|x - y\|$ and since $K < 1$ it follows that $\|x - y\| = 0$ so $x = y$. ☐

In the application to the differential equation the norm is $\|f\| = \max_{t_0 \le t \le t_0+\delta} |f(t)|$. However, we will show how the idea of contractions works in a simpler case:
Ex. Find an approximation for $\sqrt{2}$. Let

$$g(x) = \frac{x^2 + 2}{2x}$$

Then $\sqrt{2}$ is a fixed point for $g(x)$; $g(\sqrt{2}) = \sqrt{2}$. We claim that it is a contraction of the set $W_0 = \{x; x \ge 1\}$:

$$|g(x) - g(y)| \le \frac{1}{2}|x - y|, \qquad \text{if} \quad x, y \ge 1 \tag{3}$$

and $g(x) \ge 1$ if $x \ge 1$. Therefore, by the above lemma, if we set $x_0 = 1$ and $x_{n+1} = g(x_n)$, for $n \ge 0$ then $x_n \to \sqrt{2}$, as $n \to \infty$. In fact,

$$x_0 = 1, \quad x_1 = 1.5, \quad x_2 = 1.41667..., \quad x_3 = 1.41422..., \cdots$$

To prove (3) we note that $|g'(s)| = |1/2 - 1/s^2| \le 1/2$, if $|s| \ge 1$ and hence

$$|g(x) - g(y)| = \Big| \int_y^x g'(s)\, ds \Big| \le \int_y^x |g'(s)| ds \le \frac{|x - y|}{2}, \quad \text{if} \quad x \ge y \ge 1.$$