Math 252 — Spring 2000
Supplement on Euler's Method

Introduction.

The textbook seems overly enthusiastic about Euler's method. These notes aim to present a more realistic treatment of the value of the method and its relation to other numerical methods for solving differential equations.

First, although Euler's method can be performed on a simple calculator, it cannot be considered well suited for hand computation. The repetitive steps require that the whole program be stored by the computing device to guarantee that they will be performed consistently and correctly. If machine computation is to be used, other methods are available to deliver high accuracy quickly with only a small increase in the complexity of the program. Note that our analytic methods aim to find *general solutions* that contain a parameter, allowing initial value problems to be solved by identifying the value of the parameter that is consistent with the initial data, but numerical methods need the initial condition to give a characterization of a unique function which the method attempts to calculate.

The main value of Euler's method is that it is easy to analyze, and this analysis can be used to prove a form of the *existence and uniqueness theorem* that would be good enough for the purposes of this course. The textbook does not take full advantage of this, although it does describe how one estimates the error in Euler's method in Section 7.1. However, that discussion appears late in the book and the key ideas may be lost in the technicalities needed to give a complete proof.

No proofs are given to the theorems stated in the text since the usual proofs use special methods to get a strong result from a weaker hypothesis. If you require the right side of the differential equation to be continuously differentiable, then you can get an error estimate that bounds the work required to get within a given distance of a solution by Euler's method. This estimate will show that Euler's approximations converge to a solution as the step size goes to zero.

Although we refer to the problem we are studying as an "initial value problem", and usually specify $y(0)$, our method will find a solution for $t < 0$ as well as for $t > 0$.

Examples.

You should use *Maple* to produce the direction fields of these examples. In order to have the necessary tools available, begin a new worksheet with the instruction

```
with(DEtools):
```

Then, you can get the illustrations of these examples using the `dfieldplot` function (see *Maple Help* for a full list of options). In particular,

```
dfieldplot(diff(y(t),t)=(1+t*y(t))/(2+y(t)ˆ2),y(t),t=-2..2,y=-2..2);#A
dfieldplot(diff(y(t),t)=tˆ2+yˆ2,y(t),t=-2..2,y=-2..2);#B
dfieldplot(diff(y(t),t)=2*(t+sqrt(tˆ2-y(t))),y(t),t=-2..2,y=-1..4);# C
```

Examples A and B were constructed to avoid the patterns of equations that can be solved in closed form, yet they will be seen to have a unique solution through each point of the plane. The solutions to A can be extended to be defined for all t, but all solutions of B are unbounded over a bounded interval of t.

Example C is very different. The right side of the equation is only defined for $y \le t^2$, but we can give an exact description of the solution through each such point. First, check that $y = t^2$ is a solution. Then, show that the portion of the line $y = 2ct - c^2$ where $t < c$ is a solution. If you start at some point (t_0, y_0) strictly below $y = t^2$, the solution can be found by solving $y_0 = 2ct_0 - c^2$ for c and selecting the solution that is greater than t_0. There is a unique solution that follows this line until it becomes tangent to the parabola at (c, c^2), to the right of the starting point, and then it follows the parabola. Thus, for each point (a, a^2) on the parabola, the solutions through that point are all the curves that we just described starting from points (t_0, y_0) with $t_0 < a$ and $2at_0 - a^2 < y_0 < a^2$ (i.e., points to the left of the given point between the parabola and the tangent line at the given point).

What is a numerical solution?

The true solution of the equation gives y as a function of t. If there is an exact solution, this is achieved by giving a formula that we know how to evaluate. Such a formula is useful if it can be written *briefly*. This allows the behavior of the function to be illustrated by a graph as well as allowing easy computation of y for arbitrary t to fairly high accuracy. To get similar performance from a numerical method, an interpolation formula will be used to give a similar computation of the function anywhere in its domain from a list of its values at a finite number of points. In Euler's method, the values of t will be $t_k = t_0 + kh$ for some small *step size* h and integers k. We want to allow k to be either positive or negative, but the computation will typically use only positive k. To get the points for negative k, it is customary to change the sign of h and repeat the method of solution. To get the value at other points, linear interpolation between the closest t_k can be used. To fix notation, let $y_k = y(t_k)$ be the value of the solution of the initial value problem at t_k, and let v_k be the approximation to y_k computed by Euler's method. Although the y_k are not known, we shall see that we know enough to produce an upper bound on $|y_k - v_k|$.

Euler's method will use v_k to find v_{k+1}. The initial condition gives y_0 and is is reasonable to take $v_0 = y_0$. Mathematical induction shows that such a process determines all v_k for positive integers k.

Using the equation to find the second derivative.

Although our goal is to prove that equations have solutions, we begin by assuming that we have a solution and seek to learn more about it. Thus, we suppose that we have a solution $y(t)$ to an equation

$$\frac{dy}{dt} = f(t, y) \tag{1}$$

and look for additional properties of the solution. In particular, we can differentiate both sides of (1) with respect to t to obtain (via the chain rule for functions of two variables)

$$\begin{aligned} \frac{d^2y}{dt^2} &= f_t(t, y) + f_y(t, y) \cdot \frac{dy}{dt} \\ &= f_t(t, y) + f_y(t, y) \cdot f(t, y) \end{aligned}$$

where the subscripts on f indicate partial derivatives. Thus, if $f(t, y)$ is differentiable, the solution $y(t)$ will have a second derivative. This can be continued to find higher derivatives of $y(t)$ as long as the partial derivatives of f exist, but the expressions become messy very quickly. Fortunately, no more than d^2y/dt^2 is needed for the analysis of Euler's method.

For examples A and B, these expressions exist everywhere. For example C, existence of this quantity requires $y < t^2$: there is a square root of $t^2 - y$ in the denominator, so we need to be sure both that the square root exists and that it is not zero. Interestingly, in this case, the expression for d^2y/dt^2 simplifies to zero. This tells us that any solutions must lie along straight lines as long as they remain below $y = t^2$.

In general, the expression for the second derivative will depend on both t and y, but in any bounded region where this expression is continuous, we can compute a bound on the second derivative of any function

$y(t)$ satisfying the equation, while the graph of the solution lies in this region. Specifying bounds on y means that we announce that we will lose interest in a solution as soon as it gets too far from the initial value. Without this restriction, we would only be sure of our solution on an interval around the initial value of t that was so small that the solution could not reach the top or bottom of our graphing window. This is the only reason for the mysterious ϵ in the statements in the textbook.

An important consequence of this follows from Taylor's formula:

$$y(t) = y(t_k) + y'(t_k)(t - t_k) + \frac{1}{2}y''(\tau)(t - t_k)^2$$

where τ is some number between t and t_k. The first two terms on the right give the equation of the tangent line to the solution curve at $t = t_k$. Although the last term contains much that is not known, the assumption that the solution lies in our given region means that $y''(\tau)$ is given by $f_t(\tau, \eta) + f_y(\tau, \eta) \cdot f(\tau, \eta)$ at some point (τ, η) in our bounded region. Any bound on this expression translates into a proof that the tangent is close to the curve when $t - t_k$ is small. Confining attention to a bounded rectangle in the (t, y) plane gives a uniform bound $\left|y''(\tau)\right| \leq M$ (assuming that the partial derivatives of $f(t, y)$ are continuous). Thus, for Euler's method with a step size of h, a single step introduces an error that is a bounded multiple of h^2.

Nearby Solutions.

After we have been using Euler's method for a while, the current point is no longer on the solution through the starting point. If this is not to cause too much trouble, the tangent lines at the true point (t_k, y_k) and the calculated point (t_k, v_k) should have roughly the same direction. Fortunately, a quantitative version of this condition can be expressed in terms of things that can be estimated. One needs only the mean value theorem to show that the difference of the slopes, $f(t_k, y_k) - f(t_k, v_k)$ is $f_y(t_k, \eta) \cdot (y_k - v_k)$ with η between v_k and y_k. Again, as long as both the true solution and the approximate solution stay in the given region, we have a quantity L such that $\left|f_y(t, y)\right| \leq L$. For #$C$, $f_y(t, y)$ is unbounded if $y - t^2$ is small, and the uniqueness theorem fails for points on this parabola.

The technical part of the proof constructs an inductive argument to show that this component of the error has a an effect that is bounded independent of the step size. For Euler's method, this means that for any a for which the solution between $t = t_0$ and $t = a$ remains in the given bounded region, $y(a)$ has an error bounded by a fixed multiple of h^2 for each step of size h. The number of steps is proportional to $1/h$, so the total error is proportional to h.

This is good enough to show that the method approximates solutions, but not good for computing solutions to a reasonable accuracy. To increase the accuracy by a single decimal place requires ten times as much computation. In addition, the simple act of adding together a million numbers means that a million round-off errors are accumulated. This requires that higher accuracy must be maintained throughout the computation to keep these errors from affecting the part of the answer that is expected to be accurate.

Existence and Uniqueness.

The error estimate in Euler's method shows that, on any closed and bounded region D where the value of d^2y/dt^2 computed from the equation is continuous, the computed approximations converge to any solution of a given initial value problem. Since a convergent sequence has a unique limit, this means that an initial value can have at most one solution on D.

To prove the existence of solutions, similar methods are applied to estimate the difference between the quantities computed by Euler's method with different step sizes. This shows that Euler's method will converge to something. The final step is to show that any such limiting function must satisfy the differential equation.

Technicalities of compounding.

In addition to the t_k, y_k and v_k, let

$$w_{k+1} = y_k + h \cdot f(t_k, y_k).$$

That is, w_{k+1} is the point that would be computed if we made one step of Euler's method starting from (t_k, y_k). The triangle inequality gives

$$|y_{k+1} - v_{k+1}| \le |y_{k+1} - w_{k+1}| + |w_{k+1} - v_{k+1}|.$$

The comments above show that

$$|y_{k+1} - w_{k+1}| \le \frac{Mh^2}{2},$$

and

$$|w_{k+1} - v_{k+1}| \le |y_k - v_k| \cdot (1 + Lh).$$

Induction on this gives

$$|y_k - v_k| \le \frac{Mh^2}{2} \sum_{j=0}^{k-1} (1 + Lh)^j.$$

The sum is a geometric series, so we have

$$|y_k - v_k| \le \frac{Mh^2}{2} \frac{(1 + Lh)^k - 1}{(1 + Lh) - 1}.$$

The fraction at the end of this expression is a difference quotient of the function x^k, so

$$\frac{(1 + Lh)^k - 1}{(1 + Lh) - 1} = k(1 + \theta Lh)^{k-1}$$

for some θ between 0 and 1. Since $(1 + 1/x)^x$ increases to e as $x \to +\infty$,

$$(1 + \theta Lh)^{k-1} < (1 + Lh)^k < e^{Lhk},$$

and

$$|y_k - v_k| \le \frac{Mh^2k}{2} e^{Lhk}.$$

In this expression, $hk = t_k - t_0$, which is the total horizontal displacement. This depends on the point being computed, but not on the step size. We are assuming a bound of the form $|t_k - t_0| \le H$, so we find

$$|y_k - v_k| \le \frac{MHh}{2} e^{LH}.$$

Now, everything in this bound is constant except for one factor of h.

Instead of comparing the Euler approximation to a true solution, one can compare Euler approximations of different step sizes. A similar result will hold. If the step size is repeatedly cut in half, this error estimate will show that the sequence functions $e_k(t)$ obtained will have allow a function $y(t)$ to be defined as

$$y(t) = \lim_{k \to \infty} e_k(t),$$

and this function will satisfy the given initial value problem.

The only part of this analysis that is specific to Euler's method is the single-step bound of $Mh^2/2$. One factor of h gets multiplied by k to give H. Any local error estimate will suffer the same fate. The Runge-Kutta method, which Maple uses for its numerical solutions, has a single-step bound of the form Rh^5, so its global error bound is RHh^4e^{LH}.