Variable Step Size Methods

These notes introduce a family of procedures that decide by themselves what step size to use. In all of these procedures the user specifies an acceptable error rate and the procedure attempts to adjust the step size so that each step introduces error at no more than that rate.

Suppose that we wish to generate an approximation to the initial value problem

$$y' = f(t, y), \qquad y(t_0) = y_0$$

for some range of t's and we want the error introduced per unit increase of t to be no more than about ε. Suppose further that we have already produced the approximate solution as far as t_n. The rough strategy is as follows. We do the step from t_n to $t_n + h$ twice using two different algorithms, giving two different answers, that we call A_1 and A_2. The two algorithms are chosen so that

(1) we can use $A_1 - A_2$ to compute an approximate local truncation error and

(2) for efficiency, the two algorithms use almost the same evaluations of f.

In the event that the local truncation error, divided by h, (i.e. the error per unit increase of t) is smaller than ε, we set $t_{n+1} = t_n + h$, accept A_2 (or better still, A_2 minus the computed approximate error in A_2) as the approximate value for $y(t_{n+1})$ and move on to the next step. Otherwise we pick, using what we have learned from $A_1 - A_2$, a new trial step size h and start over again at t_n.

Now for the details. We start with a very simple minded procedure.

Euler and Euler–2step

Denote by $\phi(t)$ the exact solution to $y' = f(t, y)$ that satisfies the initial condition $\phi(t_n) = y_n$. If we apply one step of Euler with step size h, giving

$$A_1 = y_n + hf(t_n, y_n)$$

we know that

$$A_1 = \phi(t_n + h) + Kh^2 + O(h^3)$$

The problem of course is that we don't know what the error is, even approximately, because we don't know what the constant K is. But we can determine K simply by redoing the step from t_n to $t_n + h$ using a judiciously chosen second algorithm. There are a number of different second algorithms that will work. We call the one we use Euler–2step. One step of Euler–2step with step size h just consists of doing two steps of Euler of size $h/2$:

$$A_2 = y_n + \tfrac{h}{2} f(t_n, y_n) + \tfrac{h}{2} f\Big(t_n + \tfrac{h}{2}, y_n + \tfrac{h}{2} f(t_n, y_n)\Big)$$

Here, the first half–step took us from y_n to $y_{\text{mid}} = y_n + \frac{h}{2} f(t_n, y_n)$ and the second half–step took us from y_{mid} to $y_{\text{mid}} + \frac{h}{2} f\big(t_n + \frac{h}{2}, y_{\text{mid}}\big)$. The local truncation error introduced in the first half–step is $K(h/2)^2 + O(h^3)$. That for the second half–step is $K(h/2)^2 + O(h^3)$ with the same K, though a different $O(h^3)$.[1] All together

$$A_2 = \phi(t_n + h) + \tfrac{1}{2} K h^2 + O(h^3)$$

The difference is[2]

$$\begin{aligned} A_1 - A_2 &= \phi(t_n + h) + K h^2 + O(h^3) - \phi(t_n + h) - \tfrac{1}{2} K h^2 - O(h^3) \\ &= \tfrac{1}{2} K h^2 + O(h^3) \end{aligned}$$

So if we do one step of both Euler and Euler–2step, we can estimate

$$\tfrac{1}{2} K h^2 = A_1 - A_2 + O(h^3)$$

We now know that in the step just completed Euler–2step introduced an error of about $\frac{1}{2} K h^2 \approx A_1 - A_2$. That is, the current error rate is about $r = \frac{|A_1 - A_2|}{h} \approx \frac{1}{2} K h$ per unit increase of t. If $r > \varepsilon$, we reject A_2 and repeat the current step with a new trial step size h' chosen so that $\frac{1}{2}|K| h' \approx \frac{r}{h} h' < \varepsilon$. To give ourselves a small safety margin, we could use

$$h' = .9 \, \frac{\varepsilon}{r} \, h$$

---

[1] Because the two half–steps start at values of t only $h/2$ apart, it should not be surprising that we can use the same value of K in both. In case you don't believe me, I have included a derivation of the local truncation error for Euler–2step at the end of these notes. You are not responsible for it.

[2] Recall that every time the symbol $O(h^3)$ is used it can stand for a different function that is bounded by some constant times h^3 for small h. Thus $O(h^3) - O(h^3)$ need not be zero, but is $O(h^3)$.

If $\left|A_1 - A_2\right|/h < \varepsilon$ we could accept A_2 as an approximate value for $y(t_{n+1} = t_n + h)$ and move on to the next step. But $\phi(t_n + h) = A_2 - \frac{1}{2}Kh^2 + O(h^3) = 2A_2 - A_1 + O(h^3)$, so we do better by setting

$$y_{n+1} = 2A_2 - A_1$$

For the next step, we would repeat the whole process, starting with a trial step size $h' = .9\,\frac{\varepsilon}{r}\,h$.

As a concrete example, suppose that our problem is

$$y(0) = e^{-2},\ y' = 8(1-2t)y,\ \varepsilon = .1$$

and that we have gotten as far as

$$t_7 = .33,\ y_7 = .75,\ \text{trial } h = .094$$

Then, using E to denote the estimated local truncation error in A_2 and r the corresponding error rate

$$\begin{aligned}
f(t_7, y_7) &= 8(1 - 2 \times .33).75 = 2.04 \\
A_1 &= y_7 + hf(t_7, y_7) = .75 + .094 \times 2.04 = .942 \\
y_{\text{mid}} &= y_7 + \tfrac{h}{2} f(t_7, y_7) = .75 + \tfrac{.094}{2} \times 2.04 = .846 \\
f\big(t_7 + \tfrac{h}{2}, y_{\text{mid}}\big) &= 8\big(1 - 2\big(.33 + \tfrac{.094}{2}\big)\big).846 = 1.66 \\
A_2 &= y_7 + \tfrac{h}{2} f(t_7, y_7) + \tfrac{h}{2} f(t_7 + \tfrac{h}{2}, y_{\text{mid}}) = .75 + \tfrac{.094}{2} 2.04 + \tfrac{.094}{2} 1.66 = .924 \\
E &= A_1 - A_2 = .942 - .924 = .018 \\
r &= \tfrac{|E|}{h} = \tfrac{.018}{.094} = .19
\end{aligned}$$

Since $r = .19 > \varepsilon = .1$, the current step size is unacceptable and we have to recompute with step size $h' = .9\frac{\varepsilon}{r}h = .9\frac{.1}{.19}.094 = .045$ to give

$$\begin{aligned}
f(t_7, y_7) &= 8(1 - 2 \times .33).75 = 2.04 \\
A_1 &= y_7 + hf(t_7, y_7) = .75 + .045 \times 2.04 = .842 \\
y_{\text{mid}} &= y_7 + \tfrac{h}{2} f(t_7, y_7) = .75 + \tfrac{.045}{2} \times 2.04 = .796 \\
f\big(t_7 + \tfrac{h}{2}, y_{\text{mid}}\big) &= 8\big(1 - 2\big(.33 + \tfrac{.045}{2}\big)\big).796 = 1.88
\end{aligned}$$

$$A_2 = y_7 + \tfrac{h}{2} f(t_7, y_7) + \tfrac{h}{2} f(t_7 + \tfrac{h}{2}, y_{\text{mid}}) = .75 + \tfrac{.045}{2} 2.04 + \tfrac{.045}{2} 1.88 = .838$$

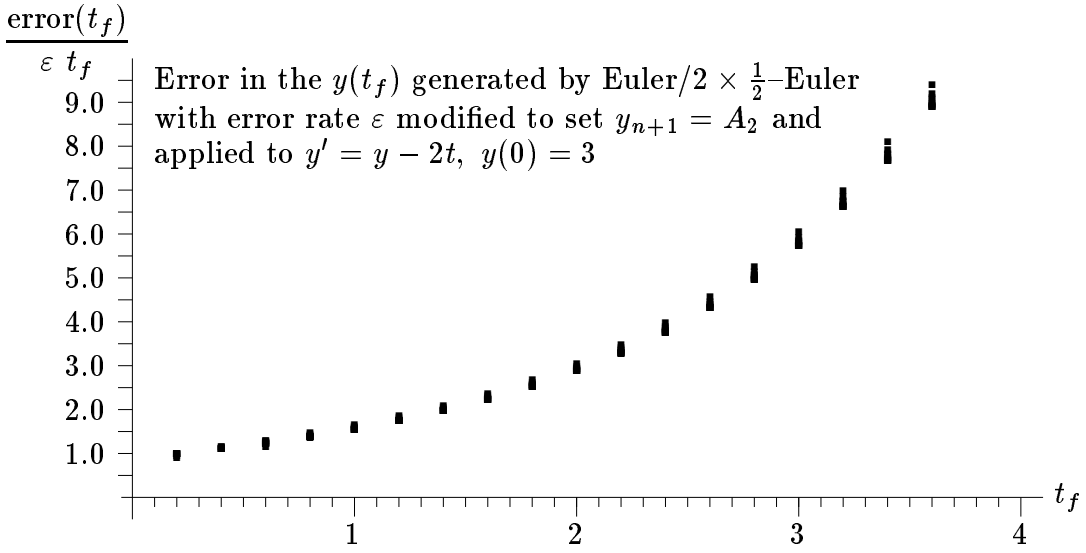
$$E = A_1 - A_2 = .842 - .838 = .004$$

$$r = \tfrac{|E|}{h} = \tfrac{.004}{.045} = .09$$

This time $r = .09 < \varepsilon = .1$, is acceptable so we set $t_8 = .33 + .045 = .375$ and $y_8 = A_2 - E = .838 - .004 = .834$. The trial step size from t_8 to t_9 is $h' = .9 \, \tfrac{.1}{.09} \, .045 = .045$.

Error Behaviour

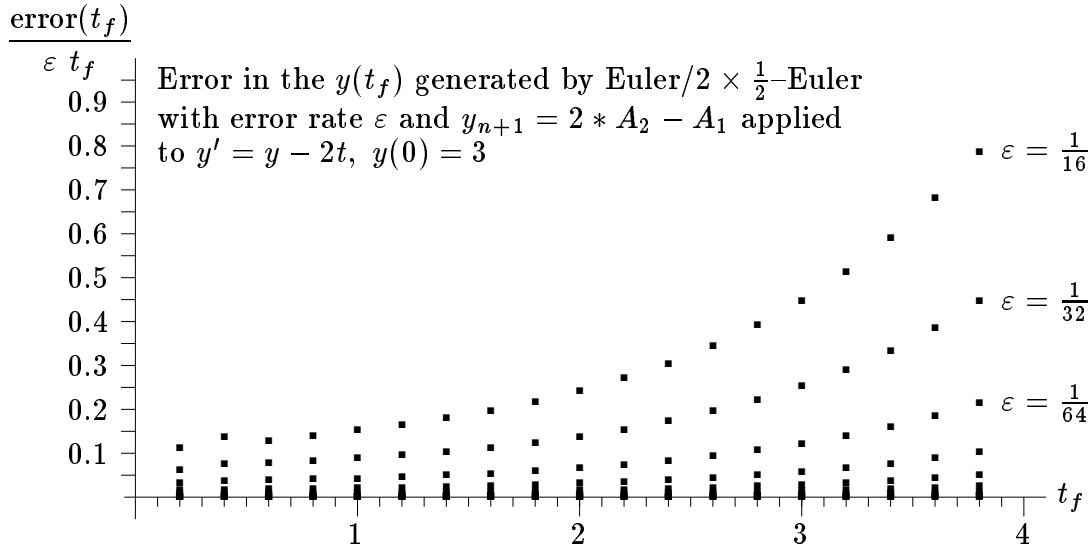
We have been referring loosely to ε as the desired rate for introduction of error, by our variable step size method, as t advances. If the rate of increase of error were exactly ε, then at final time t_f the error would be exactly $\varepsilon(t_f - t_0)$. But our algorithm actually chooses the step size h for each step so that the estimated local truncation error in A_2 for that step is about εh. We have seen that, once some local truncation error has been introduced, its contribution to the global truncation error can grow exponentially with t_f. Here are the results of an experiment that illustrate this effect. In the experiment, the method of the last section is modified by setting y_{n+1} to A_2 (the quantity whose local truncation error is estimated) rather than $A_2 - E$. This modified method is applied to $y' = t - 2y, \; y(0) = 3$ for $\varepsilon = \tfrac{1}{16}, \tfrac{1}{32}, \cdots$ (ten different values) and for $t_f = 0.2, \; 0.4, \; \cdots, \; 3.8$. Here is a plot of the



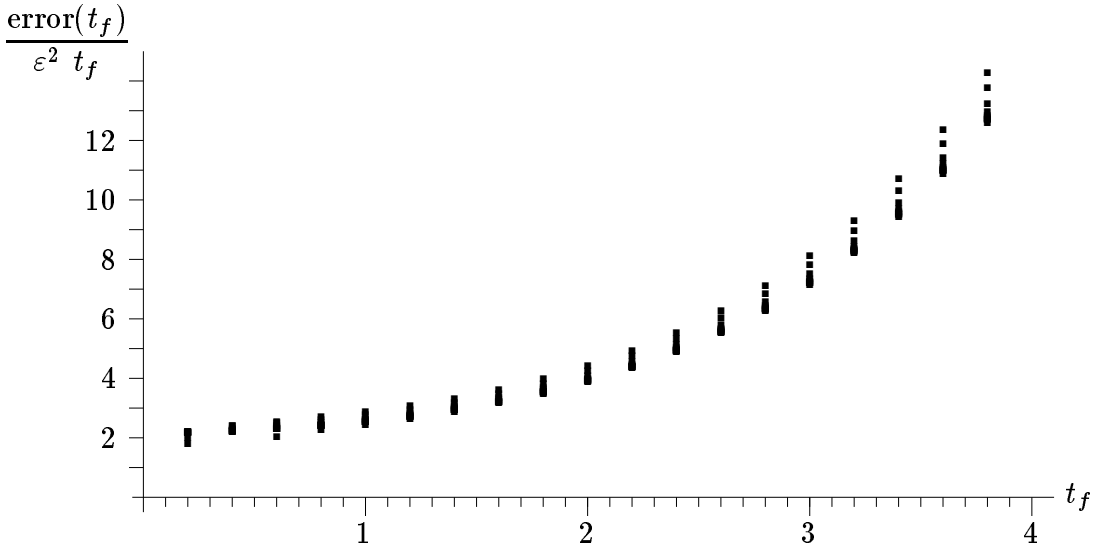
resulting $\frac{\text{actual error at } t=t_f}{\varepsilon t_f}$ against t_f. There is a small square on the graph for each different pair ε, t_f. So for each value of t_f there are ten squares on the line $x = t_f$. This numerical experiment suggests that $\frac{\text{actual error at } t=t_f}{\varepsilon t_f}$ is relatively independent of ε and starts at about

one, as we want, but grows (perhaps exponentially) with t_f.

Here are the results of a second experiment, similar to the first but using the real Euler/Euler–2step algorithm, with y_{n+1} set to $A_2 - E = 2A_2 - A_1$. Now, the step size is chosen based on an $O(h^2)$ local truncation error while the value of y_{n+1} is chosen using an $O(h^3)$ local truncation error. So one would expect the ratio $\frac{\text{actual error at } t=t_f}{\varepsilon t_f}$ to decrease by a factor of two each time ε is decreased by a factor of two. This is indeed what happens in the following data.



Note that the scale of the y–axis has been expanded by a factor of ten. Here is a plot of $\frac{\text{actual error at } t=t_f}{\varepsilon^2 t_f}$ for the same data.



Fehlberg's Method

Of course, in practice more accurate methods than Euler and Euler–2step are used. Fehlberg's method uses improved Euler and a second more accurate method. Each step involves three calculations of f:

$$f_1 = f(t_n, y_n)$$
$$f_2 = f(t_n + h, y_n + hf_1)$$
$$f_3 = f\left(t_n + \tfrac{h}{2}, y_n + \tfrac{h}{4}[f_1 + f_2]\right)$$

Once these three evaluations have been made, the method generates two approximations for $y(t_n + y)$:

$$A_1 = y_n + \tfrac{h}{2}[f_1 + f_2]$$
$$A_2 = y_n + \tfrac{h}{6}[f_1 + f_2 + 4f_3]$$

The local truncation error in A_1 (which is just the improved Euler's method) is $Kh^3 + O(h^4)$ while that in A_2 is $O(h^4)$. The error in A_1 is

$$E = Kh^3 + O(h^4) = A_1 - A_2 + O(h^4)$$

and our estimate for rate at which error is introduced in A_1 is

$$r = \frac{|A_1 - A_2|}{h} \approx Kh^2$$

If $r > \varepsilon$ we try again with step size h' chosen so that $Kh'^2 \approx \frac{r}{h^2}h'^2 < \varepsilon$. With our traditional safety factor

$$h' = .9\sqrt{\tfrac{\varepsilon}{r}}\,h$$

If $r \leq \varepsilon$ we set $t_{n+1} = t_n + h$ and $y_{n+1} = A_2$ (since A_2 should be considerably more accurate than A_1) and move on to the next step with trial step size $h' = .9\sqrt{\frac{\varepsilon}{r}}\,h$.

Reference E. Fehlberg, NASA Technical Report R315 (1969) and NASA Technical Report R287 (1968).

The Kutta–Merson Process

The Kutta–Merson process uses two variations of the Runge–Kutta method. Each step involves five calculations of f:

$$k_1 = f(t_n, y_n)$$
$$k_2 = f(t_n + \tfrac{1}{3}h, y_n + \tfrac{1}{3}hk_1)$$
$$k_3 = f(t_n + \tfrac{1}{3}h, y_n + \tfrac{1}{6}hk_1 + \tfrac{1}{6}hk_2)$$
$$k_4 = f(t_n + \tfrac{1}{2}h, y_n + \tfrac{1}{8}hk_1 + \tfrac{3}{8}hk_3)$$
$$k_5 = f(t_n + h, y_n + \tfrac{1}{2}hk_1 - \tfrac{3}{2}hk_3 + 2hk_4)$$

Once these five evaluations have been made, the process generates two approximations for $y(t_n + h)$:

$$A_1 = y_n + h\left[\tfrac{1}{2}k_1 - \tfrac{3}{2}k_3 + 2k_4\right]$$
$$A_2 = y_n + h\left[\tfrac{1}{6}k_1 + \tfrac{2}{3}k_4 + \tfrac{1}{6}k_5\right]$$

The error in A_1 is $\frac{1}{120}h^5K + O(h^6)$ while that in A_2 is $\frac{1}{720}h^5K + O(h^6)$ with the same constant K. This unknown constant can be determined, to within an error $O(h)$, by

$$K = \tfrac{720}{5h^5}(A_1 - A_2)$$

and the approximate error in A_2 and its corresponding rate are

$$E = \tfrac{1}{720}h^5K = \tfrac{1}{5}(A_1 - A_2)$$
$$r = \tfrac{1}{720}h^4K = \tfrac{1}{5h}(A_1 - A_2)$$

If $r > \varepsilon$ we try again with step size h' chosen so that $\frac{1}{720}h'^4K \approx \frac{r}{h^4}h'^4 < \varepsilon$. With our traditional safety factor

$$h' = .9\left(\tfrac{\varepsilon}{r}\right)^{1/4} h$$

If $r \leq \varepsilon$ we set $t_{n+1} = t_n + h$ and $y_{n+1} = A_2 - E$ (since E is our estimate of the error in A_2) and move on to the next step with trial step size $h' = .9\left(\frac{\varepsilon}{r}\right)^{1/4} h$.

Reference L. Fox, Numerical Solution of Ordinary and Partial Differential Equations.

The Local Truncation Error for Euler–2step

Recall that, by definition, the local truncation error for an algorithm is the error generated by a single step of the algorithm, under the assumptions that we start the step with the exact solution and that there is no roundoff error. Denote by $\phi(t)$ the exact solution to

$$y'(t) = f(t,y)$$
$$y(t_n) = y_n$$

In other words, $\phi(t)$ obeys

$$\phi'(t) = f\big(t,\phi(t)\big) \qquad \text{for all } t$$
$$\phi(t_n) = y_n$$

In particular $\phi'(t_n) = f\big(t_n,\phi(t_n)\big) = f(t_n,y_n)$ and

$$\phi''(t_n) = \tfrac{d}{dt} f\big(t,\phi(t)\big)\big|_{t=t_n} = \big[f_t\big(t,\phi(t)\big) + f_y\big(t,\phi(t)\big)\phi'(t)\big]_{t=t_n} = f_t(t_n,y_n) + f_y(t_n,y_n) f(t_n,y_n)$$

By definition, the local truncation error for Euler is

$$E_1(h) = \phi(t_n+h) - y_n - hf\big(t_n,y_n\big)$$

while that for Euler–2step is

$$E_2(h) = \phi(t_n+h) - y_n - \tfrac{h}{2} f(t_n,y_n) - \tfrac{h}{2} f\Big(t_n+\tfrac{h}{2}, y_n + \tfrac{h}{2} f(t_n,y_n)\Big)$$

We just use Taylor's Theorem, $f(h) = f(0) + f'(0)h + \frac{1}{2}f''(0)h^2 + O(h^3)$, to expand both $E_1(h)$ and $E_2(h)$ in powers of h to order h^2. For Euler

$$E_1(h) = \phi(t_n+h) - y_n - hf\big(t_n,y_n\big) \qquad E_1(0) = \phi(t_n) - y_n = 0$$
$$E_1'(h) = \phi'(t_n+h) - f\big(t_n,y_n\big) \qquad E_1'(0) = \phi'(t_n) - f\big(t_n,y_n\big) = 0$$
$$E_1''(h) = \phi''(t_n+h) \qquad E_1''(0) = \phi''(t_n)$$

By Taylor's Theorem, the local truncation error for Euler obeys

$$\boxed{E_1(h) = \tfrac{1}{2}\phi''(t_n)h^2 + O(h^3) = Kh^2 + O(h^3) \qquad \text{with } K = \tfrac{1}{2}\phi''(t_n)}$$

For Euler–2step

$$
\begin{aligned}
E_2(h) &= \phi(t_n+h) - y_n - \tfrac{h}{2} f(t_n,y_n) - \tfrac{h}{2} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right) \\
E_2'(h) &= \phi'(t_n+h) - \tfrac{1}{2} f(t_n,y_n) - \tfrac{1}{2} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right) \\
&\qquad - \tfrac{h}{2}\tfrac{d}{dh} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right) \\
E_2''(h) &= \phi''(t_n+h) - 2\times\tfrac{1}{2}\tfrac{d}{dh} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right) \\
&\qquad - \tfrac{h}{2}\tfrac{d^2}{dh^2} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right)
\end{aligned}
$$

so that

$$
\begin{aligned}
E_2(0) &= \phi(t_n) - y_n = 0 \\
E_2'(0) &= \phi'(t_n) - \tfrac{1}{2} f(t_n,y_n) - \tfrac{1}{2} f(t_n,y_n) = 0 \\
E_2''(0) &= \phi''(t_n) - \tfrac{d}{dh} f\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right)\Big|_{h=0} \\
&= \phi''(t_n) - \tfrac{1}{2} f_t\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right)\Big|_{h=0} \\
&\qquad - \tfrac{1}{2} f(t_n,y_n) f_y\left(t_n+\tfrac{h}{2}, y_n+\tfrac{h}{2}f(t_n,y_n)\right)\Big|_{h=0} \\
&= \phi''(t_n) - f_t(t_n,y_n)\tfrac{1}{2} - f_y(t_n,y_n)\tfrac{1}{2} f(t_n,y_n) \\
&= \tfrac{1}{2}\phi''(t_n)
\end{aligned}
$$

and the local truncation error for Euler–2step obeys

$$
\boxed{E_2(h) = \tfrac{1}{4}\phi''(t_n)h^2 + O(h^3) = \tfrac{1}{2}Kh^2 + O(h^3) \qquad \text{with } K = \tfrac{1}{2}\phi''(t_n)}
$$