NUMERICAL STABILITY;
IMPLICIT METHODS

When solving the initial value problem

$$
\begin{aligned}
Y'(x) &= f(x, Y(x)), && x_0 \le x \le b \\
Y(x_0) &= Y_0
\end{aligned}
$$

we know that small changes in the initial data Y_0 will result in small changes in the solution of the differential equation. More precisely, consider the perturbed problem

$$
\begin{aligned}
Y_\varepsilon'(x) &= f(x, Y_\varepsilon(x)), && x_0 \le x \le b \\
Y_\varepsilon(x_0) &= Y_0 + \varepsilon
\end{aligned}
$$

Then assuming $f(x, z)$ and $\partial f(x, z)/\partial z$ are continuous for $x_0 \le x \le b$, $-\infty < z < \infty$, we have

$$
\max_{x_0 \le x \le b} |Y_\varepsilon(x) - Y(x)| \le c\,|\varepsilon|
$$

for some constant $c > 0$. We would like our numerical methods to have a similar property.

Consider the Euler method

$$
\begin{aligned}
y_{n+1} &= y_n + hf(x_n, y_n), && n = 0, 1, \ldots \\
y_0 &= Y_0
\end{aligned}
$$

and then consider the perturbed problem

$$
\begin{aligned}
y_{n+1}^{\varepsilon} &= y_n^{\varepsilon} + hf(x_n, y_n^{\varepsilon}), && n = 0, 1, \ldots \\
y_0^{\varepsilon} &= Y_0 + \varepsilon
\end{aligned}
$$

We can show the following.

$$
\max_{x_0 \leq x_n \leq b} |y_n^{\varepsilon} - y_n| \leq \widehat{c}\,|\varepsilon|
$$

for some constant $\widehat{c} > 0$ and for all sufficiently small values of the stepsize h.

This implies that Euler's method is stable, and in the same manner as was true for the original differential equation problem.

The general idea of *stability* for a numerical method is essentially that given above for Eulers's method.

There is a general theory for numerical methods for solving the initial value problem

$$Y'(x) = f(x, Y(x)), \quad x_0 \leq x \leq b$$
$$Y(x_0) = Y_0$$

If the *truncation error* in a numerical method has order 2 or greater, then the numerical method is stable **if and only if** it is a convergent numerical method.

Note that Euler's method has a truncation error of order 2.

The numerical methods studied in this chapter are both stable and convergent.

All of these general results on stability and convergence are valid if the stepsize h is *sufficiently small*. What is meant by this?

Rather than studying this for a general problem, we restrict our interest to the *model problem*

$$Y'(x) = \lambda Y(x), \quad x \geq 0, \quad Y(0) = 1$$

The analysis for this problem is generally applicable to the more general differential equation problem.

The true solution is $Y(x) = e^{\lambda x}$. When $\lambda < 0$ or λ is complex with $\text{Real}(\lambda) < 0$, we have $Y(x) \to 0$ as $x \to \infty$. We would like the same to be true for the numerical solution of the model problem.

We begin by studying Euler's method applied to the model problem.

EULER'S METHOD
AND THE MODEL PROBLEM

Apply Euler's method to the model problem.

$$\begin{aligned} y_{n+1} &= y_n + h\lambda y_n \\ &= (1 + h\lambda)\, y_n, \quad n = 0, 1, \dots \end{aligned}$$

with $y_0 = 1$. By induction,

$$y_n = (1 + h\lambda)^n, \quad n \geq 0$$

The numerical solution $y_n \to 0$ as $n \to \infty$ if and only if

$$|1 + h\lambda| < 1$$

In the case λ is real and negative, this is equivalent to

$$-2 < h\lambda < 0$$

If $|\lambda|$ is quite large, then h must be correspondingly small. Usually the truncation error does not require such a small value of h; it is needed only for stability.

NUMERICAL EXAMPLE: Consider the problem

$$Y'(x) = \lambda Y(x) + (1 - \lambda)\cos(x) - (1 + \lambda)\sin(x)$$

with $Y(0) = 1$. The true solution is $Y(x) = \sin(x) + \cos(x)$. We use Euler's method and give numerical results for varying λ and h are given. The truncation error $T_{n+1} = \frac{h^2}{2}Y''(\xi_n)$ does not depend on λ, but the numerical solution does.

λ	x	Error: $h = 0.5$	Error: $h = 0.1$	Error: $h = 0.01$
-1	1	$-2.46\text{E}-1$	$-4.32\text{E}-2$	$-4.22\text{E}-3$
	3	$-2.66\text{E}-2$	$-6.78\text{E}-3$	$-7.22\text{E}-4$
	5	$2.72\text{E}-1$	$4.91\text{E}-2$	$4.81\text{E}-3$
-10	1	$3.98\text{E}-1$	$-6.99\text{E}-3$	$-6.99\text{E}-4$
	3	$1.11\text{E}+2$	$3.86\text{E}-3$	$3.64\text{E}-4$
	5	$2.83\text{E}+4$	$3.78\text{E}-3$	$3.97\text{E}-4$
-50	1	$3.26\text{E}+0$	$1.06\text{E}+3$	$-1.39\text{E}-4$
	3	$1.08\text{E}+6$	$1.17\text{E}+15$	$8.25\text{E}-5$
	5	$3.59\text{E}+11$	$1.28\text{E}+27$	$7.00\text{E}-5$

ABSOLUTE STABILITY

The set of values of $h\lambda$ for which the numerical solution $y_n \rightarrow 0$ as $n \rightarrow \infty$ is called the *region of absolute stability* of the numerical method. We allow λ to be complex, restricting it with Real$(\lambda) < 0$.

With Euler's method, this region is the set of all complex numbers $z = h\lambda$ for which

$$|1 + z| < 1$$

or equivalently,

$$|z - (-1)| < 1$$

This is a circle of radius one in the complex plane, centered at the complex number $-1 + 0 \cdot i$.

If a numerical method has no restrictions on λ in order to have $y_n \rightarrow 0$ as $n \rightarrow \infty$, we say the numerical method is *A-stable*.

THE BACKWARD EULER METHOD

Expand the function $Y(x)$ as a linear Taylor polynomial about x_{n+1}:

$$Y(x) = Y(x_{n+1}) + (x - x_{n+1}) Y'(x_{n+1})$$
$$+ \frac{1}{2} (x - x_{n+1})^2 Y''(\zeta_n)$$

with ζ_n between x and x_{n+1}. Let $x = x_n$, solve for $Y(x_{n+1})$, and use the differential equation to evaluate $Y'(x_{n+1})$:

$$Y(x_{n+1}) = Y(x_n) + hY'(x_{n+1}) - \frac{1}{2}h^2 Y''(\zeta_n)$$
$$= Y(x_n) + hf(x_{n+1}, Y(x_{n+1})) - \frac{1}{2}h^2 Y''(\zeta_n)$$

with ζ_n between x_n and x_{n+1}. The *backward Euler method* is obtained by dropping the truncation error:

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}), \qquad n = 0, 1, \ldots$$
$$y_0 = Y_0$$

The truncation is essentially of the same size as for Euler's method, but of opposite sign.

Apply the backward Euler method to the model problem

$$Y'(x) = \lambda Y(x), \quad x \geq 0, \quad Y(0) = 1$$

This yields

$$
\begin{aligned}
y_{n+1} &= y_n + h\lambda y_{n+1} \\
y_{n+1} &= \frac{1}{1 - h\lambda} y_n, \quad n = 0, 1, \ldots
\end{aligned}
$$

with $y_0 = 1$. By induction,

$$y_n = \left(\frac{1}{1 - h\lambda} \right)^n, \quad n = 0, 1, \ldots$$

We want to know when $y_n \to \infty$ as $n \to \infty$. This will be true if

$$\left| \frac{1}{1 - h\lambda} \right| < 1$$

The hypothesis that $\lambda < 0$ or $\text{Real}(\lambda) < 0$ is sufficient to show this is true, regardless of the size of the stepsize h. Thus the backward Euler method is an A-stable numerical method.

NUMERICAL EXAMPLE

We apply the backward Euler method to the problem

$$Y'(x) = \lambda Y(x) + (1 - \lambda)\cos(x) - (1 + \lambda)\sin(x)$$

with $Y(0) = 1$. The true solution is $Y(x) = \sin(x) + \cos(x)$. We give numerical results with varying λ and with $h = 0.5$. The truncation error $T_{n+1} = -\frac{h^2}{2}Y''(\xi_n)$ does not depend on λ, but the numerical solution does. But in contrast to Euler's method, there are no stability problems with the numerical solution.

x	Error: $\lambda = -1$	Error: $\lambda = -10$	Error: $\lambda = -50$
2	$2.08\text{E} - 1$	$1.97\text{E} - 2$	$3.60\text{E} - 3$
4	$-1.63\text{E} - 1$	$-3.35\text{E} - 2$	$-6.94\text{E} - 3$
6	$-7.04\text{E} - 2$	$8.19\text{E} - 3$	$2.18\text{E} - 3$
8	$2.22\text{E} - 1$	$2.67\text{E} - 2$	$5.13\text{E} - 3$
10	$-1.14\text{E} - 1$	$-3.04\text{E} - 2$	$-6.45\text{E} - 3$

SOLVING THE BACKWARD EULER METHOD

For a general differential equation, we must solve

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}) \qquad (1)$$

for each n. In most cases, this is a rootfinding problem for the equation

$$z = y_n + hf(x_{n+1}, z) \qquad (2)$$

with the root $z = y_{n+1}$. Such numerical methods (1) for solving differential equations are called *implicit methods*.

Methods in which y_{n+1} is given explicitly are called *explicit methods*. Euler's method is an explicit method.

Fixed point iteration is often used to solve (2):

$$y_{n+1}^{(k+1)} = y_n + hf\left(x_{n+1}, y_{n+1}^{(k)}\right), \quad k = 0, 1, \ldots \quad (3)$$

For an initial guess, we use $y_{n+1}^{(0)} = y_n$ or something even better. The iteration error satisfies

$$y_{n+1} - y_{n+1}^{(k+1)} \approx h\frac{\partial f(x_{n+1}, y_{n+1})}{\partial y}\left[y_{n+1} - y_{n+1}^{(k)}\right]$$
$$(4)$$

We have convergence if

$$h\left|\frac{\partial f(x_{n+1}, y_{n+1})}{\partial y}\right| < 1 \quad (5)$$

which is true if h is sufficiently small. If this is too restrictive on h, then another rootfinding method must be used to solve (2).

THE TRAPEZOIDAL METHOD

The backward Euler method is stable, but still is lacking in accuracy. A similar but more accurate numerical method is the *trapezoidal method*:

$$y_{n+1} = y_n + \frac{h}{2}\left[f\left(x_n, y_n\right) + f\left(x_{n+1}, y_{n+1}\right)\right], \quad (6)$$
$$n = 0, 1, \ldots$$

It is derived by applying the simple trapezoidal numerical integration rule to the equation

$$Y(x_{n+1}) = Y(x_n) + \int_{x_n}^{x_{n+1}} f\left(t, Y(t)\right)\ dt$$

obtaining

$$Y(x_{n+1}) = Y(x_n) + \frac{h}{2}\left[f\left(x_n, Y(x_n)\right)\right.$$
$$\left. + f\left(x_{n+1}, Y(x_{n+1})\right)\right] - \frac{h^3}{12}Y'''(\xi_n)$$

The method (6) results from dropping the truncation error term.

As with the backward Euler method, the equation (6) is a nonlinear equation with a root of y_{n+1}. Again, fixed point iteration can be used to solve it:

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

for $j = 0, 1, 2, \ldots$ The iteration will converge if

$$\left| \frac{h}{2} \cdot \frac{\partial f(x_{n+1}, y_{n+1})}{\partial z} \right| < 1$$

provided also that $y_{n+1}^{(0)}$ is a sufficiently good initial guess. Often we use

$$y_{n+1}^{(0)} = y_n + hf(x_n, y_n)$$

or

$$y_{n+1}^{(0)} = y_n + \frac{h}{2}[3f(x_n, y_n) - f(x_{n-1}, y_{n-1})]$$

These latter formulas are called *predictor formulas*. The first is simply Euler's method. The second is an *Adams-Bashforth method of order 2*; and it is an *explicit 2-step method*. Other rootfinding methods are used for more difficult problems.

CONVERGENCE: We can show that for all sufficiently small values of h,

$$\max_{x_0 \leq x_n \leq b} |Y(x_n) - y_n| \leq ch^2 \max_{x_0 \leq x \leq b} \left| Y'''(x) \right|$$

The constant c depends on the Lipschitz constant K for $f(x, z)$:

$$K = \max_{\substack{x_0 \leq x \leq b \\ -\infty < z < \infty}} \left| \frac{\partial f(x, z)}{\partial z} \right|$$

Thus it converges more rapidly than either the Euler method or the backward Euler method.

STABILITY: The trapezoidal method is *absolutely stable*. Apply the method to the problem

$$Y'(x) = \lambda Y(x), \quad x \geq 0, \quad Y(0) = 1$$

Then

$$y_{n+1} = y_n + \frac{h}{2}[\lambda y_n + \lambda y_{n+1}]$$

with $y_0 = 1$. Solving for y_{n+1},

$$y_{n+1} = \left(\frac{1 + \frac{h\lambda}{2}}{1 - \frac{h\lambda}{2}}\right) y_n, \quad n \geq 0$$

By induction,

$$y_n = \left(\frac{1 + \frac{h\lambda}{2}}{1 - \frac{h\lambda}{2}}\right)^n, \quad n \geq 0$$

Then for λ real and negative, and also for λ complex with $\text{Real}(\lambda) < 0$, we can use this formula to show $y_n \to 0$ as $n \to \infty$. This shows the trapezoidal method is absolutely stable.

NUMERICAL EXAMPLE: We apply the trapezoidal method to the problem

$$Y'(x) = \lambda Y(x) + (1 - \lambda)\cos(x) - (1 + \lambda)\sin(x)$$

with $Y(0) = 1$. The true solution is $Y(x) = \sin(x) + \cos(x)$. We give numerical results with varying λ and with $h = 0.5$. The truncation error $T_{n+1} = -\frac{h^3}{12}Y'''(\xi_n)$ does not depend on λ, but the numerical solution does. As with the backward Euler method, there are no stability problems with the numerical solution. Moreover, it converges more rapidly than does the backward Euler method.

x	Error: $\lambda = -1$	Error: $\lambda = -10$	Error: $\lambda = -50$
2	$-1.13\mathrm{E} - 2$	$-2.78\mathrm{E} - 3$	$-7.91\mathrm{E} - 4$
4	$-1.43\mathrm{E} - 2$	$-8.91\mathrm{E} - 5$	$-8.91\mathrm{E} - 5$
6	$2.02\mathrm{E} - 2$	$2.77\mathrm{E} - 3$	$4.72\mathrm{E} - 4$
8	$-2.86\mathrm{E} - 3$	$-2.22\mathrm{E} - 3$	$-5.11\mathrm{E} - 4$
10	$-1.79\mathrm{E} - 2$	$-9.23\mathrm{E} - 4$	$-1.56\mathrm{E} - 4$

STIFF PROBLEMS: Linear equations

$$Y'(x) = \lambda Y(x) + g(x)$$

with $\text{Real}(\lambda) < 0$, $|\text{Real}(\lambda)|$ large, are called *stiff differential equations*. For general equations, the role of λ is replaced by

$$\frac{\partial f(x, Y(x))}{\partial z}$$

With many numerical methods, the truncation error for these equations may be satisfactorily small with not too small a value of h. But the large size of $|\text{Real}(\lambda)|$ may force h to be smaller so that λh will be in the stability region of the numerical method.

For stiff differential equations, one must use a numerical method with a large region of absolute stability, or else h must be chosen very small.

The backward Euler method and the trapezoidal method are very desirable because their stability region contains all λh where $\text{Real}(\lambda) < 0$. The backward Euler method and the trapezoidal rule have convergence orders 1 and 2, respectively.

The iteration method for calculating y_{n+1} in

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

must be something like Newton's method, something where the convergence is not affected by the size of λh.