Numerical Solution of Differential Equations I

Endre Süli

Mathematical Institute University of Oxford 2020

Lecture 3

General one-step methods

Definition

A **one-step method** is a function Ψ that takes the triplet

$$(\xi, \eta; h) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0}$$

and a function f, and computes an approximation

$$\Psi(\xi, \eta; h, f) \in \mathbb{R}$$
 of $y(\xi + h)$,

where $y(\xi + h)$ is the value at $x = \xi + h$ of the solution y to the initial-value problem

$$y'(x) = f(x, y(x)), \qquad y(\xi) = \eta.$$



General one-step methods

Definition

A **one-step method** is a function Ψ that takes the triplet

$$(\xi, \eta; h) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}_{>0}$$

and a function f, and computes an approximation

$$\Psi(\xi, \eta; h, f) \in \mathbb{R}$$
 of $y(\xi + h)$,

where $y(\xi + h)$ is the value at $x = \xi + h$ of the solution y to the initial-value problem

$$y'(x) = f(x, y(x)), \qquad y(\xi) = \eta.$$

Remark. The step size (mesh size) h may need to be assumed to be sufficiently small for Ψ to be well-defined.



In the case of the implicit Euler method

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

the function Ψ is defined implicitly, by

$$\Psi(\xi,\eta;h,f)=\eta+hf(\xi+h,\Psi(\xi,\eta;h,f)).$$

In the case of the implicit Euler method

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

the function Ψ is defined implicitly, by

$$\Psi(\xi,\eta;h,f) = \eta + hf(\xi + h, \Psi(\xi,\eta;h,f)).$$

Let f satisfy the Lipschitz condition with Lipschitz constant L.



In the case of the implicit Euler method

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

the function Ψ is defined implicitly, by

$$\Psi(\xi,\eta;h,f)=\eta+hf(\xi+h,\Psi(\xi,\eta;h,f)).$$

Let f satisfy the Lipschitz condition with Lipschitz constant L.

The Contraction Mapping Theorem implies that, given a pair $(\xi, \eta) \in \mathbb{R}^2$ and $h \in (0, 1/L)$, there exists a unique $\Psi(\xi, \eta; h, f)$ in \mathbb{R} satisfying this implicit relationship.

Therefore, for such a "sufficiently small" h, the function Ψ associated with the implicit Euler method is well-defined.



In the case of a general explicit one-step method

$$\Psi(\xi,\eta;h,f)=\eta+h\,\Phi(\xi,\eta;h,f),$$

where $\Phi(\xi, \eta; h, f)$ can be explicitly computed without solving implicit equations in terms of ξ , η , h, and f.

In the case of a general explicit one-step method

$$\Psi(\xi,\eta;h,f)=\eta+h\,\Phi(\xi,\eta;h,f),$$

where $\Phi(\xi, \eta; h, f)$ can be explicitly computed without solving implicit equations in terms of ξ , η , h, and f.

Example

For the explicit Euler method

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

we have that

$$\Psi(\xi, \eta; h, f) = \eta + hf(\xi, \eta).$$

In the case of a general explicit one-step method

$$\Psi(\xi,\eta;h,f)=\eta+h\,\Phi(\xi,\eta;h,f),$$

where $\Phi(\xi, \eta; h, f)$ can be explicitly computed without solving implicit equations in terms of ξ , η , h, and f.

Example

For the explicit Euler method

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

we have that

$$\Psi(\xi,\eta;h,f)=\eta+hf(\xi,\eta).$$

Remark. In what follows, we shall not indicate the dependence of $\Phi(\xi, \eta; h, f)$ on f, and will write $\Phi(\xi, \eta; h)$ instead for simplicity.



General explicit one-step method

A general explicit one-step method is of the form:

$$y_{n+1} = y_n + h \Phi(x_n, y_n; h), \quad n = 0, \dots, N-1, \quad y_0 = y(x_0) [= given],$$
 where $\Phi(\cdot, \cdot; \cdot)$ is a continuous function of its variables.

General explicit one-step method

A general explicit one-step method is of the form:

$$y_{n+1} = y_n + h \Phi(x_n, y_n; h), \quad n = 0, \dots, N-1, \quad y_0 = y(x_0) [= given],$$

where $\Phi(\cdot,\cdot;\cdot)$ is a continuous function of its variables.

Example

For the improved Euler method

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

we have that

$$\Phi(\xi,\eta;h) = \frac{1}{2} \left[f(\xi,\eta) + f(\xi+h,\eta+hf(\xi,\eta)) \right].$$

In order to assess the accuracy of a one-step method we define the **global error**, e_n , by

$$e_n := y(x_n) - y_n.$$

In order to assess the accuracy of a one-step method we define the **global error**, e_n , by

$$e_n := y(x_n) - y_n.$$

We define the **consistency error** (also called truncation error), T_n , of the method by

$$T_n := \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h). \tag{1}$$

In order to assess the accuracy of a one-step method we define the **global error**, e_n , by

$$e_n := y(x_n) - y_n.$$

We define the **consistency error** (also called truncation error), T_n , of the method by

$$T_n := \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h). \tag{1}$$

Remark. Note that, by definition, the numerical solution satisfies:

$$0 = \frac{y_{n+1} - y_n}{h} - \Phi(x_n, y_n; h). \tag{2}$$

Theorem

Consider a general explicit one-step method where, in addition to being a continuous function of its arguments, Φ is assumed to satisfy a Lipschitz condition with respect to its second argument; i.e., there exists a constant $L_{\Phi}>0$ such that, for $0\leq h\leq h_0$ and for the same region R as in Picard's Theorem (cf. Lecture 1),

$$|\Phi(x,y;h) - \Phi(x,z;h)| \le L_{\Phi}|y-z|,$$
 for (x,y) , (x,z) in R.

Theorem

Consider a general explicit one-step method where, in addition to being a continuous function of its arguments, Φ is assumed to satisfy a Lipschitz condition with respect to its second argument; i.e., there exists a constant $L_{\Phi}>0$ such that, for $0 \le h \le h_0$ and for the same region R as in Picard's Theorem (cf. Lecture 1),

$$|\Phi(x,y;h) - \Phi(x,z;h)| \le L_{\Phi}|y-z|,$$
 for (x,y) , (x,z) in R.

Then, assuming that $|y_n - y_0| \le Y_M$, it follows that

$$|e_n| \le e^{L_{\Phi}(x_n - x_0)}|e_0| + \left[\frac{e^{L_{\Phi}(x_n - x_0)} - 1}{L_{\Phi}}\right] T, \quad n = 0, \dots, N,$$

where $T = \max_{0 \le n \le N-1} |T_n|$.

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

PROOF: By subtracting eq. (2) from eq. (1), we have $e_{n+1} = e_n + h[\Phi(x_n,y(x_n);h) - \Phi(x_n,y_n;h)] + hT_n.$ As $(x_n,y(x_n)),(x_n,y_n) \in \mathbb{R}$, the Lipschitz condition implies $|e_{n+1}| < |e_n| + hL_{\Phi}|e_n| + h|T_n|, \qquad n = 0,\dots,N-1.$

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

As $(x_n, y(x_n)), (x_n, y_n) \in \mathbb{R}$, the Lipschitz condition implies

$$|e_{n+1}| \le |e_n| + hL_{\Phi}|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

That is,

$$|e_{n+1}| \leq (1 + hL_{\Phi})|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

As $(x_n, y(x_n)), (x_n, y_n) \in R$, the Lipschitz condition implies

$$|e_{n+1}| \le |e_n| + hL_{\Phi}|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

That is,

$$|e_{n+1}| \leq (1 + hL_{\Phi})|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

Hence, letting $T := \max_{0 \le n \le N-1} |T_n|$, we have that

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

As $(x_n, y(x_n)), (x_n, y_n) \in \mathbb{R}$, the Lipschitz condition implies

$$|e_{n+1}| \le |e_n| + hL_{\Phi}|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

That is,

$$|e_{n+1}| \leq (1 + hL_{\Phi})|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

Hence, letting $T := \max_{0 \le n \le N-1} |T_n|$, we have that

$$|e_1| \leq (1+hL_{\Phi})|e_0|+hT$$

$$|e_2| \leq (1+hL_{\Phi})^2|e_0|+h[1+(1+hL_{\Phi})]T$$

$$|e_3| \leq (1+hL_{\Phi})^3|e_0|+h[1+(1+hL_{\Phi})+(1+hL_{\Phi})^2]T,$$

etc.

$$|e_n| \leq (1 + hL_{\Phi})^n |e_0| + [(1 + hL_{\Phi})^n - 1]T/L_{\Phi}.$$

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

As $(x_n, y(x_n)), (x_n, y_n) \in R$, the Lipschitz condition implies

$$|e_{n+1}| \le |e_n| + hL_{\Phi}|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

That is,

$$|e_{n+1}| \leq (1 + hL_{\Phi})|e_n| + h|T_n|, \qquad n = 0, \dots, N-1.$$

Hence, letting $T := \max_{0 \le n \le N-1} |T_n|$, we have that

$$|e_1| \leq (1+hL_{\Phi})|e_0|+hT$$

$$|e_2| \leq (1+hL_{\Phi})^2|e_0|+h[1+(1+hL_{\Phi})]T$$

$$|e_3| \leq (1+hL_{\Phi})^3|e_0|+h[1+(1+hL_{\Phi})+(1+hL_{\Phi})^2]T,$$

etc.

$$|e_n| \leq (1+hL_{\Phi})^n|e_0|+[(1+hL_{\Phi})^n-1]T/L_{\Phi}.$$

As $1 + hL_{\Phi} \leq \exp(hL_{\Phi})$, we obtain the stated bound. \diamond



Consider the initial-value problem $y' = \tan^{-1} y$, $y(0) = y_0$, and suppose that this is solved by the explicit Euler method.

Consider the initial-value problem $y' = \tan^{-1} y$, $y(0) = y_0$, and suppose that this is solved by the explicit Euler method.

The aim of the exercise is to quantify the size of the associated global error; thus, we need to find L_{Φ} and T.

Consider the initial-value problem $y' = \tan^{-1} y$, $y(0) = y_0$, and suppose that this is solved by the explicit Euler method.

The aim of the exercise is to quantify the size of the associated global error; thus, we need to find L_{Φ} and T.

Here $f(x, y) = \tan^{-1} y$, so by the Mean-Value Theorem

$$|f(x,y)-f(x,z)|=\left|\frac{\partial f}{\partial y}(x,\eta)(y-z)\right|,$$

where η lies between y and z. In our case

$$\left|\frac{\partial f}{\partial y}\right| = |(1+y^2)^{-1}| \le 1,$$

and therefore $L_{\Phi} = 1$.



As $T = \max_{0 \le n \le N-1} |T_n|$, and for Euler's method $T_n = \frac{1}{2}hy''(\xi_n)$, where $\xi_n \in (x_n, x_{n+1})$ (cf. Lecture 2), we need to obtain a bound on |y''| (without actually solving the initial-value problem).

As $T = \max_{0 \le n \le N-1} |T_n|$, and for Euler's method $T_n = \frac{1}{2}hy''(\xi_n)$, where $\xi_n \in (x_n, x_{n+1})$ (cf. Lecture 2), we need to obtain a bound on |y''| (without actually solving the initial-value problem). By differentiating both sides of the differential equation w.r.t. x:

$$y'' = \frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1}y) = (1+y^2)^{-1}\frac{\mathrm{d}y}{\mathrm{d}x} = (1+y^2)^{-1}\tan^{-1}y.$$

As $T = \max_{0 \le n \le N-1} |T_n|$, and for Euler's method $T_n = \frac{1}{2}hy''(\xi_n)$, where $\xi_n \in (x_n, x_{n+1})$ (cf. Lecture 2), we need to obtain a bound on |y''| (without actually solving the initial-value problem). By differentiating both sides of the differential equation w.r.t. x:

$$y'' = \frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1}y) = (1+y^2)^{-1}\frac{\mathrm{d}y}{\mathrm{d}x} = (1+y^2)^{-1}\tan^{-1}y.$$

Thus $|y''(x)| \le \frac{1}{2}\pi$, whereby $T = \frac{1}{4}\pi h$. Inserting this and $L_{\Phi} = 1$ into the error bound from the last theorem (note that $x_0 = 0$):

$$|e_n| \le e^{x_n} |e_0| + \frac{1}{4} \pi (e^{x_n} - 1) h, \quad n = 0, \dots, N.$$

As $T = \max_{0 \le n \le N-1} |T_n|$, and for Euler's method $T_n = \frac{1}{2}hy''(\xi_n)$, where $\xi_n \in (x_n, x_{n+1})$ (cf. Lecture 2), we need to obtain a bound on |y''| (without actually solving the initial-value problem). By differentiating both sides of the differential equation w.r.t. x:

$$y'' = \frac{\mathrm{d}}{\mathrm{d}x}(\tan^{-1}y) = (1+y^2)^{-1}\frac{\mathrm{d}y}{\mathrm{d}x} = (1+y^2)^{-1}\tan^{-1}y.$$

Thus $|y''(x)| \le \frac{1}{2}\pi$, whereby $T = \frac{1}{4}\pi h$. Inserting this and $L_{\Phi} = 1$ into the error bound from the last theorem (note that $x_0 = 0$):

$$|e_n| \le e^{x_n} |e_0| + \frac{1}{4} \pi (e^{x_n} - 1) h, \quad n = 0, \dots, N.$$

In particular if we assume that no error has been committed initially (i.e. $e_0 = 0$), we have that

$$|e_n| \leq \frac{1}{4}\pi \left(e^{x_n} - 1\right)h, \quad n = 0, \dots, N.$$

Thus, given a positive tolerance TOL we can ensure that the error between the (unknown) analytical solution and its numerical approximation over an interval $[x_0, X_M]$ (in this example $x_0 = 0$) does not exceed TOL, by choosing a positive step size h such that

$$h \leq \frac{4}{\pi} (e^{X_M} - 1)^{-1} \text{ TOL}.$$

Thus, given a positive tolerance TOL we can ensure that the error between the (unknown) analytical solution and its numerical approximation over an interval $[x_0, X_M]$ (in this example $x_0 = 0$) does not exceed TOL, by choosing a positive step size h such that

$$h \leq \frac{4}{\pi} (e^{X_M} - 1)^{-1} \text{ TOL.}$$

For such h we shall have $|y(x_n) - y_n| = |e_n| \le \text{TOL}$ for each n = 0, ..., N, as required. \diamond

Consistency

Definition

A general one-step method is said to be **consistent** with the ODE y' = f(x, y) if the associated consistency error T_n is such that for any $\varepsilon > 0$ there exists a positive $h(\varepsilon)$ for which $|T_n| < \varepsilon$ for $0 < h < h(\varepsilon)$ and any pair of points $(x_n, y(x_n))$, $(x_{n+1}, y(x_{n+1}))$ on any solution curve contained in R (cf. Lecture 1).

As we have assumed that the function $\Phi(\cdot,\cdot;\cdot)$ is continuous, and also y' is a continuous function of x on $[x_0,X_M]$, it follows that

$$\lim_{h \to 0, x_n \to x \in [x_0, X_M]} T_n = y'(x) - \Phi(x, y(x); 0)$$

$$= f(x, y(x)) - \Phi(x, y(x); 0).$$

As we have assumed that the function $\Phi(\cdot,\cdot;\cdot)$ is continuous, and also y' is a continuous function of x on $[x_0,X_M]$, it follows that

$$\lim_{h \to 0, x_n \to x \in [x_0, X_M]} T_n = y'(x) - \Phi(x, y(x); 0)$$

$$= f(x, y(x)) - \Phi(x, y(x); 0).$$

So, a general explicit one-step method is consistent if, and only if,

$$\Phi(x,y;0) \equiv f(x,y).$$

As we have assumed that the function $\Phi(\cdot,\cdot;\cdot)$ is continuous, and also y' is a continuous function of x on $[x_0,X_M]$, it follows that

$$\lim_{h \to 0, x_n \to x \in [x_0, X_M]} T_n = y'(x) - \Phi(x, y(x); 0)$$

$$= f(x, y(x)) - \Phi(x, y(x); 0).$$

So, a general explicit one-step method is consistent if, and only if,

$$\Phi(x,y;0) \equiv f(x,y).$$

Now we are ready to state a convergence theorem for the general one-step method.

Convergence

Theorem

Suppose that the solution of the initial-value problem y' = f(x, y), $y(x_0) = y_0$ lies in R (cf. Lecture 1) as does its approximation generated from $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$ when $h \le h_0$.

Convergence

Theorem

Suppose that the solution of the initial-value problem y'=f(x,y), $y(x_0)=y_0$ lies in R (cf. Lecture 1) as does its approximation generated from $y_{n+1}=y_n+h\Phi(x_n,y_n;h)$ when $h\leq h_0$. Let $\Phi(\cdot,\cdot;\cdot)$ be uniformly continuous on $R\times[0,h_0]$ and satisfy the consistency condition $\Phi(x,y;0)=f(x,y)$ and the Lipschitz condition

$$|\Phi(x,y;h) - \Phi(x,z;h)| \le L_{\Phi}|y-z|$$
 on $\mathbb{R} \times [0,h_0]$.

Convergence

Theorem

Suppose that the solution of the initial-value problem y'=f(x,y), $y(x_0)=y_0$ lies in R (cf. Lecture 1) as does its approximation generated from $y_{n+1}=y_n+h\Phi(x_n,y_n;h)$ when $h\leq h_0$. Let $\Phi(\cdot,\cdot;\cdot)$ be uniformly continuous on $R\times[0,h_0]$ and satisfy the consistency condition $\Phi(x,y;0)=f(x,y)$ and the Lipschitz condition

$$|\Phi(x,y;h) - \Phi(x,z;h)| \le L_{\Phi}|y-z|$$
 on $\mathbb{R} \times [0,h_0]$.

Then, if successive approximation sequences (y_n) , generated for $x_n = x_0 + nh$, n = 1, 2, ..., N, are obtained from this one-step method with successively smaller values of h, each less than h_0 , we have convergence of the numerical solution to the solution of the initial-value problem in the sense that

$$|y(x)-y_n| \to 0$$
 as $h \to 0$, $x_n \to x \in [x_0, X_M]$.



Proof: We need to show that

$$|y(x)-y_n| \to 0$$
 as $h \to 0$, $x_n \to x \in [x_0, X_M]$.

Proof: We need to show that

$$|y(x)-y_n|\to 0$$
 as $h\to 0$, $x_n\to x\in [x_0,X_M]$.

However, by the triangle inequality,

$$|y(x) - y_n| \le |y(x) - y(x_n)| + |y(x_n) - y_n|.$$

The first term on the r.h.s. side tends to 0 as $x_n \to x \in [x_0, X_M]$ thanks to the continuity of y, guaranteed by Picard's theorem.

Proof: We need to show that

$$|y(x)-y_n| \to 0$$
 as $h \to 0$, $x_n \to x \in [x_0, X_M]$.

However, by the triangle inequality,

$$|y(x) - y_n| \le |y(x) - y(x_n)| + |y(x_n) - y_n|.$$

The first term on the r.h.s. side tends to 0 as $x_n \to x \in [x_0, X_M]$ thanks to the continuity of y, guaranteed by Picard's theorem.

The second term on the r.h.s. side tends to 0 by the previous theorem, thanks to the assumed consistency of the method, which implies that

$$\lim_{h\to 0,\,x_n\to x\in[x_0,X_M]}T_n=0,$$

and because $e_0 = y(x_0) - y_0 = 0$. \diamond

Order of accuracy

We saw earlier that for Euler's method the absolute value of the consistency error T_n is bounded above by a constant multiple of the step size h, that is

$$|T_n| \le Kh$$
 for $0 < h \le h_0$,

where K is a positive constant, independent of h.

Order of accuracy

We saw earlier that for Euler's method the absolute value of the consistency error T_n is bounded above by a constant multiple of the step size h, that is

$$|T_n| \le Kh$$
 for $0 < h \le h_0$,

where K is a positive constant, independent of h.

However there are other one-step methods (a class of which, called Runge–Kutta methods, will be considered in the next lecture) for which we can do better.

Order of accuracy

We saw earlier that for Euler's method the absolute value of the consistency error T_n is bounded above by a constant multiple of the step size h, that is

$$|T_n| \le Kh$$
 for $0 < h \le h_0$,

where K is a positive constant, independent of h.

However there are other one-step methods (a class of which, called Runge–Kutta methods, will be considered in the next lecture) for which we can do better.

To quantify the asymptotic rate of decay of the consistency error as the step size h tends to zero, we introduce the following definition.

Definition

The one-step method $y_{n+1} = y_n + h \Phi(x_n, y_n; h)$ is said to have **order of accuracy** p, if p is the largest positive integer such that, for any sufficiently smooth solution curve (x, y(x)) in R of the initial-value problem y' = f(x, y), $y(x_0) = y_0$, there exist constants K and h_0 such that

$$|T_n| \le Kh^p$$
 for $0 < h \le h_0$

for any pair of points $(x_n, y(x_n))$, $(x_{n+1}, y(x_{n+1}))$ on the solution curve.

Definition

The one-step method $y_{n+1} = y_n + h \Phi(x_n, y_n; h)$ is said to have **order of accuracy** p, if p is the largest positive integer such that, for any sufficiently smooth solution curve (x, y(x)) in R of the initial-value problem y' = f(x, y), $y(x_0) = y_0$, there exist constants K and h_0 such that

$$|T_n| \le Kh^p$$
 for $0 < h \le h_0$

for any pair of points $(x_n, y(x_n))$, $(x_{n+1}, y(x_{n+1}))$ on the solution curve.

Next, we shall focus on a family of explicit one-step methods which have p-th order of accuracy, $p \ge 1$: explicit Runge-Kutta methods.