MA3227 Numerical Analysis II

Lecture 6: Runge-Kutta Methods

Simon Etter

National University of Singapore

Semester II, AY 2020/2021

Runge-Kutta Methods

Problem statement

Given

$$f : \mathbb{R}^n \to \mathbb{R}^n, \qquad y_0 \in \mathbb{R}^n \qquad \text{and} \qquad T > 0,$$

determine $y : [0, T) \to \mathbb{R}^n$ such that

$$y(0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = f\big(y(t)\big) \quad \text{for all } t \in [0, T).$$

$\dot{y}$ is a shorthand notation for $\dot{y} = \frac{dy}{dt}$.

Terminology: Ordinary differential equations (ODEs)

Problems of the above form are known as *ordinary differential equations*.

ODEs are typically used to model time-evolution phenomena. For this reason, y_0 is called the *initial condition*, and T is called the *final time*.

Outlook

The following slides will illustrate the above problem statement by discussing several example ODEs.

Runge-Kutta Methods

Example 1

Consider the problem of finding $y : [0, \infty) \to \mathbb{R}$ such that

$$y(0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = \lambda\, y(t)$$

for some given $y_0, \lambda \in \mathbb{R}$.
The solution to this problem is given by

$$y(t) = y_0 \,\exp(\lambda t)$$

because this function satisfies

$$y(0) = y_0 \,\exp(\lambda 0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = y_0 \,\exp(\lambda t)\,\lambda = \lambda\, y(t).$$

Runge-Kutta Methods

Example 2

Consider the problem of finding $y : [0, \frac{1}{y_0}) \to \mathbb{R}$ such that

$$y(0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = y(t)^2$$

for some given $y_0 \in \mathbb{R}$.

The solution to this problem is given by

$$y(t) = \frac{y_0}{1 - y_0 t}$$

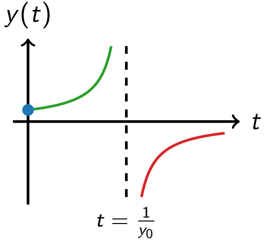
because this function satisfies

$$y(0) = \frac{y_0}{1 - y_0\, 0} = y_0 \qquad \text{and} \qquad \dot{y}(t) = \frac{y_0^2}{(1 - y_0 t)^2} = y(t)^2.$$

Example 2 (continued)

Note that the above solution diverges as t approaches $\frac{1}{y_0}$,

$$y(t) = \frac{y_0}{1 - y_0 t} \qquad \longleftrightarrow$$



Example 2 is hence different from Example 1 in that the above $y(t)$ diverges already after a finite time $\frac{1}{y_0}$ while the solution in Example 1, namely

$$y(t) = y_0 \exp(\lambda t),$$

is finite for all $t \in [0, \infty)$.

Runge-Kutta Methods

Example 3

Consider the problem of finding $x : [0, \infty) \to \mathbb{R}$ such that

$$x(0) = 1, \qquad \dot{x}(0) = 0 \qquad \text{and} \qquad \ddot{x} = -x.$$

The solution to this equation is given by

$$x(t) = \cos(t)$$

since this function satisfies

$$x(0) = \cos(0) = 1, \quad \dot{x}(0) = -\sin(0) = 0$$

and

$$\ddot{x}(t) = \tfrac{d^2}{dt^2}\cos(t) = -\tfrac{d}{dt}\sin(t) = -\cos(t) = -x(t).$$

Runge-Kutta Methods

Example 3 (continued)

The ODE $\ddot{x} = -x$ is not of the form $\dot{y} = f(y)$, but it can be reduced to this form by setting

$$y = \begin{pmatrix} x \\ \dot{x} \end{pmatrix} \qquad \text{and} \qquad f(y) = \begin{pmatrix} y_2 \\ -y_1 \end{pmatrix}$$

since then

$$\dot{y} = \begin{pmatrix} \dot{x} \\ \ddot{x} \end{pmatrix} = \begin{pmatrix} y_2 \\ -y_1 \end{pmatrix} = f(y).$$

This trick can be generalised to reduce ODEs with arbitrarily high derivatives to a system of ODEs involving only first-order derivatives.

Runge-Kutta Methods

Real-world example: Newton's law of motion

A famous real-world example of an ODE is Newton's law of motion

$$m\,\ddot{x}(t) = F\bigl(x(t)\bigr).$$

This equation relates the acceleration $\ddot{x}$ of a point particle of mass m to the forces $F\bigl(x(x)\bigr)$ acting on the particle at the current position $x(t)$.

Runge-Kutta Methods

ODEs vs PDEs

ODEs are similar to PDEs in the sense that the problem is to find a function $y(t)$ given an equation in terms of $y(t)$ and its derivatives.

Formally, the difference between ODEs and PDEs is the following.

- In an ODE, the unknown $y(t)$ depends on a single scalar variable and hence the derivatives are *ordinary* derivatives.
- In a PDE, the unknown $u(x_1, \ldots, x_d)$ depends on several variables and hence the derivatives are *partial* derivatives.

The terms "ODE" and "PDE" are hardly ever used in this way, however.

In modern terminology, the defining property of an ODE is that fixed values for $y(t)$ and its derivatives are specified at a single point t_0. For this reason, ODEs are also called *initial value problems*.

The defining property of a PDE is that fixed values of $u(x)$ and its derivatives are specified at two or more points $x \in \partial\Omega$. For this reason, PDEs are also called *boundary value problems*.

Example: ODEs vs PDEs

The one-dimensional Poisson equation $-u''(x) = f(x)$ is an ODE in the formal sense because there is only a single independent variable x.

However, this equation is usually called a PDE because it is almost always paired with boundary conditions rather than initial conditions and hence it is much closer in spirit to e.g. the higher-dimensional Poisson equation $-\Delta u = f$ than to Newton's law of motion $m\ddot{x} = F(x)$.

Runge-Kutta Methods

Outlook

Our main goal in this lecture is of course to develop numerical algorithms for evaluating the ODE map

$$\big(\, f(y),\, y_0,\, T \,\big) \quad \mapsto \quad y(t) \quad \text{such that} \quad y(0) = y_0, \quad \dot{y}(t) = f(y(t)).$$

However, before doing so it is advisable to first study the conditions under which this map is well defined, i.e. the conditions which guarantee that the above problem has precisely one solution.

It turns out that these conidtions are fairly simple: all we need is that $f(y)$ is Lipschitz continuous. The following slides will explain further.

Def: (Global) Lipschitz continuity

A function $f : D \to \mathbb{R}^n$ with $D \subset \mathbb{R}^n$ is called *(globally) Lipschitz continuous with Lipschitz constant* $L > 0$ if for all $y_1, y_2 \in D$ we have

$$\left\| f(y_1) - f(y_2) \right\| \;\leq\; L \, \|y_1 - y_2\|.$$

A function which is Lipschitz continuous with some unspecified Lipschitz constant $L > 0$ is called simply *Lipschitz continuous*.

Picard-Lindelöf theorem, global version

Assume $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous. Then, there exists a unique function $y : [0, \infty) \to \mathbb{R}^n$ such that

$$y(0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = f\big(y(t)\big) \quad \text{for all } t \in [0, \infty).$$

Proof. Beyond the scope of this module.

Runge-Kutta Methods

The Picard-Lindelöf theorem indicates that understanding Lipschitz continuity is important for understanding ODEs.

The following result provides a convenient tool for establishing Lipschitz continuity of a given function $f(y)$.

Thm: Global Lipschitz continuity and differentiability

Assume $D \subset \mathbb{R}^n$ is convex and $f : D \to \mathbb{R}^n$ is differentiable everywhere in D. Then, $f(y)$ is Lipschitz continuous if $\|\nabla f\|$ is bounded.

Proof (not examinable). Immediate corollary of the result on the next slide (which I will skip in class).

Example

Recall from Example 1 on slide 3 the ODE

$$y(0) = y_0, \qquad \dot{y} = \lambda y.$$

Since $f'(y) = \lambda$ is bounded for all y, this function is globally Lipschitz continuous and hence the solution $y(t) = y_0 \exp(\lambda t)$ exists for all $t \geq 0$.

Runge-Kutta Methods

Lemma: Lipschitz constants and derivatives (not examinable)

Assume $D \subset \mathbb{R}^n$ is convex and $f : D \to \mathbb{R}^n$ has a bounded derivative. Then,

$$\|f(y_1) - f(y_2)\| \leq L \|y_1 - y_2\| \qquad \text{where} \qquad L = \sup_{y \in D} \|\nabla f(y)\|$$

Proof.

$$\|f(y_1) - f(y_2)\| = \left\| \int_0^1 \frac{d}{dt}\Big(f\left(y_1 + t\left(y_2 - y_1\right)\right)\Big)\, dt \right\|$$

(fundamental theorem of calculus)

$$= \left\| \int_0^1 \nabla f\left(y_1 + t\left(y_2 - y_1\right)\right)\left(y_2 - y_1\right) dt \right\|$$

(chain rule)

$$\leq \int_0^1 \left\|\nabla f\left(y_1 + t\left(y_2 - y_1\right)\right)\right\| \left\|y_2 - y_1\right\| dt$$

($\| \int f(t)\, dt\| \leq \int \|f(t)\|\, dx$ and matrix norm definition)

$$\leq \left(\sup_{y \in D} \|\nabla f(y)\| \right) \left\|y_2 - y_1\right\|.$$

Runge-Kutta Methods

The Picard-Lindelöf theorem on slide 12 assumes that $f(y)$ is globally Lipschitz but in return guarantees existence and uniqueness of the solution for all $t \geq 0$. There is also a version of the Picard-Lindelöf theorem which assumes that $f(y)$ is only *locally Lipschitz continuous* (see below) but in return guarantees existence and uniqueness of the solution only over some potentially finite interval $[0, T)$.

Def: Local Lipschitz continuity

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is called *locally Lipschitz continuous* if for every $y_1 \in \mathbb{R}^n$ there exists a pair $\delta, L > 0$ such that for all $y_2 \in \mathbb{R}^n$ we have

$$\|y_1 - y_2\| \leq \delta \qquad \Longrightarrow \qquad \big\|f(y_1) - f(y_2)\big\| \leq L\,\|y_1 - y_2\|.$$

Picard-Lindelöf theorem, local version

Assume $f : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous. Then, there exists a $T > 0$ and a unique function $y : [0, T) \to \mathbb{R}^n$ such that

$$y(0) = y_0 \qquad \text{and} \qquad \dot{y}(t) = f\big(y(t)\big) \quad \text{for all } t \in [0, T).$$

Proof. Beyond the scope of this module.

Local Lipschitz continuity is again related to differentiability.

Thm: Lipschitz continuity and differentiability
Assume $D \subset \mathbb{R}^n$ is convex and $f : D \to \mathbb{R}^n$ is differentiable everywhere in D. Then, $f(y)$ is locally Lipschitz continuous.

Proof (not examinable). Corollary of the result on slide 14.

Comparing the above against the analogous theorem for global Lipschitz continuity on slide 13, we conclude that for differentiable functions $f(y)$, the difference between local and global Lipschitz continuity is whether $\|\nabla f(y)\|$ is bounded.

Runge-Kutta Methods

Example

Recall from Example 2 on slide 4 the ODE

$$y(0) = y_0, \qquad \dot{y} = y^2.$$

Since $f'(y) = 2y$ exists but is unbounded, $f(y) = y^2$ is locally but not globally Lipschitz continuous and hence solutions may exist only over some finite interval $[0, T]$.

Indeed, we have observed previously that the solution

$$y(t) = \frac{y_0}{1 - y_0 t}$$

exists as a function $y : \mathbb{R} \to \mathbb{R}$ only on $[0, \frac{1}{y_0})$.

Runge-Kutta Methods

Continuity of the ODE map

To approximate the ODE map $\big(f(y), y_0, T\big) \mapsto y(t)$ numerically, we need this map to be not only well defined but also continuous with respect to the initial conditions; otherwise any small perturbation in y_0 (e.g. rounding errors) may lead to arbitrarily large errors in the solution $y(t)$.

Fortunately, it turns out that Lipschitz continuity of $f(y)$ guarantees not existence and uniqueness of solutions but also that these solutions are a Lipschitz continuous function of the initial conditions y_0.

Thm: Lipschitz continuity of the ODE map

Assume $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with Lipschitz constant L, and assume $y_1, y_2 : [0, T) \to \mathbb{R}^n$ are two solutions to the same ODE $\dot{y}_k = f(y_k)$ but with different initial conditions $y_1(0)$ and $y_2(0)$. Then,

$$\|y_1(t) - y_2(t)\| \;\leq\; \exp(Lt)\,\|y_1(0) - y_2(0)\| \qquad \text{for any } t < T.$$

Proof (not examinable). See the following slides (which I will skip in class).

Runge-Kutta Methods

Lipschitz continuity of the ODE map is a consequence of the following auxiliary result.

Lemma: Gronwall's inequality (not examinable)

$$\dot{y}(t) \le \lambda\, y(t) \qquad \Longrightarrow \qquad y(t) \le \exp(\lambda t)\, y(0).$$

Proof (not examinable).
Consider $z(t) = \exp(-\lambda t)\, y(t)$. Then, $z(0) = y(0)$ and

$$\begin{aligned} \dot{z}(t) &= -\lambda\, \exp(-\lambda t)\, y(t) + \exp(-\lambda t)\, \dot{y}(t) \\ &\le -\lambda\, \exp(-\lambda t)\, y(t) + \exp(-\lambda t)\, \lambda\, y(t) = 0; \end{aligned}$$

hence $z(t) \le y(0)$ and thus $y(t) \le y(0)\, \exp(\lambda t)$.

Runge-Kutta Methods

Proof of the theorem on slide 18 (not examinable).

We have for any $y : \mathbb{R} \to \mathbb{R}^n$ that

$$\tfrac{d}{dt}\|y(t)\| = \lim_{\tilde{t}\to t} \tfrac{\|y(\tilde{t})\| - \|y(t)\|}{\tilde{t}-t} \le \lim_{\tilde{t}\to t} \tfrac{\|y(\tilde{t}) - y(t)\|}{\tilde{t}-t} = \|\dot{y}(t)\|.$$

Combining the above with the Lipschitz continuity of $f(y)$, we obtain

$$\begin{aligned}
\tfrac{d}{dt}\|y_2(t) - y_1(t)\| &\le \|\dot{y}_2(t) - \dot{y}_1(t)\| \\
&= \|f(y_2(t)) - f(y_1(t))\| \\
&\le L\,\|y_2(t) - y_1(t)\|.
\end{aligned}$$

The claim then follows by Gronwall's inequality.

Runge-Kutta Methods

Interpretation of the ODE continuity

In the above bound

$$\|y_1(t) - y_2(t)\| \;\le\; \exp(Lt)\, \|y_1(0) - y_2(0)\|,$$

$y_1(t)$ typically represents the exact solution and $y_2(t)$ represents an approximation to $y_1(t)$ resulting from a slightly perturbed initial condition $y_2(0)$. In this context, the above bound is both a blessing and a curse:

- Good: The error at time t is proportional to the error at time 0.
- Bad: The constant of proportionality is $\exp(Lt)$ and hence grows very rapidly once $t > \frac{1}{L}$.

The second point implies that solving ODEs over time spans longer than one over the Lipschitz constant is essentially impossible.

It is the mathematical foundation of phenomena like the butterfly effect, i.e. the claim that the occurrence of a tornado may depend on whether a butterfly flaps its wings.

Runge-Kutta Methods

Solving ODEs using quadrature

We have now established that if $f(y)$ is Lipschitz continuous, then $\dot{y} = f(y)$ has a unique solution and this solution is a Lipschitz continuous function of the initial conditions y_0.

Let us now move on to discuss numerical methods for solving ODEs.

It turns out that solving ODEs is in some sense equivalent to evaluating integrals, namely we have

$$y(0) = y_0, \quad \dot{y} = f(y) \qquad \Longleftrightarrow \qquad y(t) = y_0 + \int_0^t f(y(\tau))\, d\tau.$$

Numerically computing integrals is known as *quadrature*, and it is a problem that we already know how to solve. The following slides will recapitulate the basics.

Runge-Kutta Methods

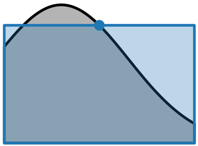
Def: Quadrature rule

A formula of the form

$$\sum_{k=1}^{n} f(x_k)\, w_k \approx \int_a^b f(x)\, dx$$

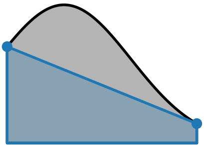
is called a *quadrature rule for* $[a, b]$. The parameters $(x_k \in \mathbb{R})_{k=1}^n$ and $(w_k \in \mathbb{R})_{k=1}^n$ are called *quadrature points* and *quadrature weights*, respectively.

Example quadrature rules $\quad (m = \frac{a+b}{2})$



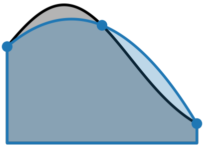
Midpoint rule

$(b - a)\, f(m)$



Trapezoidal rule

$\frac{b-a}{2}\, \big(f(a) + f(b)\big)$



Simpson's rule

$\frac{b-a}{6}\, \big(f(a) + 4\, f(m) + f(b)\big)$

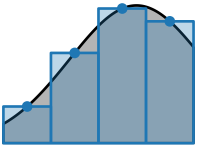
Composite quadrature

Quadrature rules generally become more accurate the larger the number of quadrature points. A simple way to construct quadrature rules with many points is to split the original integral into many small integrals,

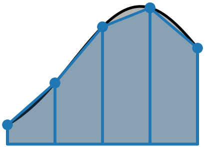
$$\int_a^b f(x)\,dx \;=\; \sum_{k=1}^{n} \int_{c_{k-1}}^{c_k} f(x)\,dx$$

and then apply a simple quadrature rule to each of these small integrals. This trick is known as *composite quadrature*.

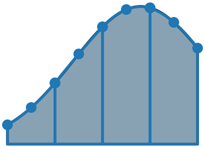
Example composite quadrature rules



Composite midpoint



Composite trapezoidal



Composite Simpson

Runge-Kutta Methods

Solving ODEs using quadrature (continued)

Let us now return to the problem of evaluating the ODE integral

$$y(t) = y_0 + \int_0^t f(y(\tau))\, d\tau.$$

The first challenge that we face when trying to apply quadrature to this integral is that the integration interval $[0, t]$ is of variable rather than fixed width. This circumstance can be remedied by substituting $\tau = xt$, which yields

$$y(t) = y_0 + \int_0^1 f(y(xt))\, t\, dx.$$

Given a quadrature rule $(x, w_k)_{k=1}^s$ for $[0, 1]$, we can hence compute a numerical approximation to $y(t)$ using the formula

$$y(t) \approx y_0 + \sum_{k=1}^{s} f\big(y(x_k t)\big)\, w_k\, t.$$

Runge-Kutta Methods

Solving ODEs using quadrature (continued)

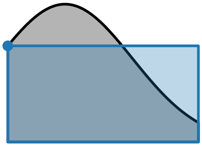
The second challenge is that the highlighted values in

$$y(t) \approx y_0 + \sum_{k=1}^{s} f\big(y(x_k t)\big)\, w_k t$$

are available only if $x_k = 0$.

The only admissible quadrature rule is hence the *left-point rule* given by

$$x_1 = 0, \qquad w_1 = 1 \qquad \longleftrightarrow$$



.

This quadrature rule leads us to the approximation

$$y(t) \approx y(0) + f\big(y(0)\big)\, t$$

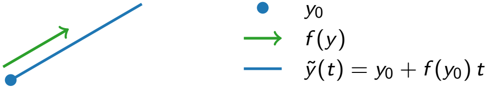
which is known as an *Euler step*.

Runge-Kutta Methods

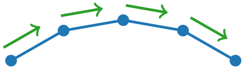
Visual interpretation of Euler's method

The right-hand side $f(y)$ in the ODE $\dot{y} = f(y)$ can be interpreted as the direction in which $y(t)$ should move given its current position.

In this mental model, the Euler step formula corresponds to looking at the direction once and then moving in this direction forever.



It is clear that this procedure will not lead to good approximations. We can obtain better approximations by using the Euler step formula to extrapolate only some small distance into the future and then update the direction $f(y)$.



This procedure is analogous to the composite quadrature idea and known as *Euler's method*.

Runge-Kutta Methods

Def: Euler's method

Approximating the solution to $y(0) = y_0$, $\dot{y} = f(y)$ using

$$\tilde{y}(0) = y_0, \qquad \tilde{y}\big(t_k\big) = \tilde{y}(t_{k-1}) + f\big(\tilde{y}(t_{k-1})\big)\,\big(t_k - t_{k-1}\big)$$

is known as *Euler's method*.

The t_k in this formula refer to a sequence of time points

$$0 = t_0 < t_1 < \ldots < t_{n-1} < t_n = T.$$

Such a sequence is called a *temporal mesh*.

For much of this lecture, I will be using the equispaced temporal mesh

$$\Big(\ t_k = \tfrac{T}{n}\,k\ \Big)_{k=0}^{n},$$

but it will occasionally be useful to also consider more general meshes.

Numerical demonstration

See `euler_step()`, `propagate()` and `example()`.