Classical Numerical Analysis, Chapter 19

Abner J. Salgado and Steven M. Wise

asalgad1@utk.edu swise1@.utk.edu
University of Tennessee

Chapter 19
Runge–Kutta Methods

A Second Look at Taylor's Method

Assume that $\mathbf{u} \in C^1\left([0,T];\mathbb{R}^d\right)$ is a classical solution to the initial value problem

$$\mathbf{u}'(t) = \mathbf{f}(t, \mathbf{u}(t)), \quad \mathbf{u}(0) = \mathbf{u}_0. \tag{1}$$

Let $K \in \mathbb{N}$, $\tau = \frac{T}{K}$, and $t_k = \tau k$ for $k = 0, \ldots, K$. As before, we will produce a sequence of approximations, $\{\mathbf{w}^k\}_{k=0}^K$, such that $\mathbf{u}(t_k) \approx \mathbf{w}^k$.

Runge–Kutta methods emerge from having a second look at Taylor's method, where the slope approximation function is given by

$$\mathbf{G}_{TM}(t, s, \mathbf{v}_1) = \mathbf{f}(t, \mathbf{v}_1) + \frac{s}{2}\left[\partial_t \mathbf{f}(t, \mathbf{v}_1) + D_{\mathrm{u}}\mathbf{f}(t, \mathbf{v}_1)\mathbf{f}(t, \mathbf{v}_1)\right].$$

As we saw earlier, this method is convergent with rate $p = 2$. It has one major drawback. Namely, it requires knowledge not only of the slope function $\mathbf{f}$ but also of its partial derivatives. In practice, these functions may not be available, or they may be very difficult to compute.

Simple Two Stage Methods

Theorem (RK2: Autonomous Case)

Let $T > 0$ be given. Consider the general two-stage explicit RK method, defined by

$$\xi^k = w^k + a\tau f(w^k), \quad w^{k+1} = w^k + \tau\Big[b_1 f(w^k) + b_2 f(\xi^k)\Big],$$

for approximating the solution to the scalar autonomous IVP

$$u'(t) = f(u(t)), \quad t \in [0, T], \quad u(0) = u_0.$$

Assume that $f \in \mathcal{F}^2(S)$ and, therefore, $u \in C^3([0, T])$. If the coefficients satisfy $b_1 + b_2 = 1$, $b_1, b_2 \geq 0$, and $ab_2 = \frac{1}{2}$, then the method is consistent of order $p = 2$ and the method is convergent to second order.

Proof.

By Taylor's Thm, for some ζ between $u(t-\tau)$ and $u(t-\tau) + a\tau f(u(t-\tau))$,

$$\begin{aligned} f\left(u(t-\tau) + a\tau f(u(t-\tau))\right) &= f(u(t-\tau)) + a\tau f(u(t-\tau)) f'(u(t-\tau)) \\ &\quad + \frac{1}{2}\left(a\tau f(u(t-\tau))\right)^2 f''(\zeta). \end{aligned}$$

Proof Cont.

Upon setting $b_1 + b_2 = 1$ and $ab_2 = \frac{1}{2}$, the LTE satisfies

$$
\begin{aligned}
\mathcal{E}[u](t,\tau) &= \frac{u(t) - u(t-\tau)}{\tau} \\
&\quad - \left[b_1 f(u(t-\tau)) - b_2 f\left(u(t-\tau) + a\tau f(u(t-\tau)) \right) \right] \\
&= \frac{u(t) - u(t-\tau)}{\tau} - b_1 f(u(t-\tau)) - b_2 f(u(t-\tau)) \\
&\quad - \tau a b_2 f(u(t-\tau)) f'(u(t-\tau)) - \tau^2 \frac{a^2 b_2}{2} f^2(u(t-\tau)) f''(\zeta) \\
&= \frac{u(t) - u(t-\tau)}{\tau} - f(u(t-\tau)) \\
&\quad - \frac{\tau}{2} f(u(t-\tau))\, f'(u(t-\tau)) - \tau^2 \frac{a^2 b_2}{2} f^2(u(t-\tau)) f''(\zeta).
\end{aligned}
$$

Proof Cont.

On the other hand, using Taylor's Theorem, the exact solution must satisfy

$$\begin{aligned} u(t) &= u(t-\tau) + \tau u'(t-\tau) + \frac{\tau^2}{2} u''(t-\tau) + \frac{\tau^3}{6} u'''(\sigma) \\ &= u(t-\tau) + \tau f(u(t-\tau)) + \frac{\tau^2}{2} f'(u(t-\tau)) f(u(t-\tau)) + \frac{\tau^3}{6} u'''(\sigma) \end{aligned}$$

for some $\sigma \in (t-\tau, t)$. Comparing the expansions,

$$\mathcal{E}[u](t,\tau) = \frac{\tau^2}{6} u'''(\sigma) - \tau^2 \frac{a^2 b_2}{2} f^2(u(t-\tau)) f''(\zeta)$$

provided that $f \in C^2((-\infty, \infty))$, $u \in C^3([0, T])$, $b_1 + b_2 = 1$, and $ab_2 = \frac{1}{2}$. There is some $C > 0$ such that

$$|\mathcal{E}[u](t,\tau)| \le C\tau^2$$

for any $\tau \in (0, T]$ and $t \in [\tau, T]$. The proof for this is subtle and relies on the fact that u is bounded over $[0, T]$.

Proof Cont.

We will only show convergence in the case that $a = 1/2$, $b_1 = 0$. With this simplification, the method reads

$$w^{k+1} = w^k + \tau f\left(w^k + \frac{\tau}{2} f(w^k)\right).$$

The exact solution satisfies

$$u(t_{k+1}) = u(t_k) + \tau f\left(u(t_k) + \frac{\tau}{2} f(u(t_k))\right) + \tau \mathcal{E}^{k+1}[u].$$

Therefore,

$$e^{k+1} = e^k + \tau f\left(u(t_k) + \frac{\tau}{2} f(u(t_k))\right) - \tau f\left(w^k + \frac{\tau}{2} f(w^k)\right) + \tau \mathcal{E}^{k+1}[u].$$

Proof Cont.

Taking absolute values and using the triangle inequality and the Lipschitz continuity of the slope function f, we have

$$\begin{aligned}
|e^{k+1}| &\leq |e^k| + \tau L \left| e^k + \frac{\tau}{2}(f(u(t_k)) - f(w^k)) \right| + \tau \left| \mathcal{E}^{k+1}[u] \right| \\
&\leq (1+\tau L)|e^k| + \frac{\tau^2 L^2}{2}|e^k| + C\tau^3 \\
&= \left(1 + \tau L + \frac{\tau^2 L^2}{2} \right) |e^k| + C\tau^3 .
\end{aligned}$$

Using the discrete Grönwall inequality,

$$|e^k| \leq \frac{C\tau^2}{L + \frac{\tau L^2}{2}} \left[\left(1 + \tau L + \frac{\tau^2 L^2}{2} \right)^k - 1 \right].$$

Proof Cont.

Now, since $\tau L > 0$,

$$1 + \tau L + \frac{\tau^2 L^2}{2} < e^{\tau L};$$

therefore, for any $m = 1, \ldots, K$,

$$(1 + \tau L)^m < e^{m\tau L} \leq e^{K\tau L} = e^{TL},$$

where we used that $K\tau = T$. It follows that, for all $k = 0, \ldots, K$,

$$|e^k| \leq \frac{C}{L}\left[e^{TL} - 1\right]\tau^2.$$

□

General Definition and Basic Properties

Definition (RK)

*Let $r \in \mathbb{N}$. A general r-**stage Runge–Kutta method** (RK method) is a recursive algorithm for generating an approximation $\left\{\mathbf{w}^k\right\}_{k=0}^{K}$ to the solution of (1), via $\mathbf{w}^0 = \mathbf{u}_0$ and, for $k = 0, \dots, K-1$,*

$$\boldsymbol{\xi}_i = \mathbf{w}^k + \tau \sum_{j=1}^{r} a_{i,j} \mathbf{f}(t_k + c_j \tau, \boldsymbol{\xi}_j), \quad i = 1, \dots, r, \tag{2}$$

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \tau \sum_{j=1}^{r} b_j \mathbf{f}(t_k + c_j \tau, \boldsymbol{\xi}_j). \tag{3}$$

*Here, $a_{i,j} \in \mathbb{R}$ and $b_j, c_j \in [0,1]$ for $i, j = 1, \dots, r$. An RK method is completely determined by its weights $\mathsf{A} = [a_{i,j}]_{i,j=1}^{r} \in \mathbb{R}^{r\times r}$, $\mathbf{b} = [b_i]_{i=1}^{r} \in \mathbb{R}^r$, and $\mathbf{c} = [c_i]_{i=1}^{r} \in \mathbb{R}^r$, which are often expressed in **tableau** form*

$$\begin{array}{c|c} \mathbf{c} & \mathsf{A} \\ \hline & \mathbf{b}^{\intercal} \end{array},$$

*which is commonly referred to as the **Butcher tableau** of the method.*

Definition (RK Cont.)

An RK method, expressed in tableau form as

$$\begin{array}{c|c} \mathbf{c} & A \\ \hline & \mathbf{b}^{\intercal} \end{array},$$

is called **explicit** *(ERK) if and only if $a_{i,j} = 0$ for all $i \leq j$, and is called* **implicit** *(IRK) otherwise. The RK method is called* **diagonally implicit** *(DIRK) if and only if $a_{i,j} = 0$ for all $i < j$.*

Theorem (Properties of Weights)

Assume that $\mathbf{f} \in \mathcal{F}^1(S)$. *Consider the general* r*-stage RK method given by the weights* $\mathsf{A} = [a_{i,j}]_{i,j=1}^r \in \mathbb{R}^{r\times r}$, $\mathbf{b} = [b_i]_{i=1}^r \in [0,1]^r$, *and* $\mathbf{c} = [c_i]_{i=1}^r \in [0,1]^r$. *Let* $\mathbf{1} = [1]_{i=1}^r \in \mathbb{R}^r$.

1. *For the method to be at least first order, it is necessary that*
$$\mathbf{b}^\intercal \mathbf{1} = 1.$$
2. *For the* j*th RK stage* $\boldsymbol{\xi}_j$ *to be at least a first-order approximation of* $\mathbf{u}(t_k + c_j\tau)$, *it is necessary that*
$$\mathsf{A}\mathbf{1} = \mathbf{c}. \tag{4}$$
3. *Suppose that* $\mathbf{f} \in \mathcal{F}^2(S)$ *and* (4) *holds. For the method to be at least second order, it is necessary that*
$$\mathbf{b}^\intercal \mathbf{c} = \frac{1}{2}.$$

Theorem (Properties of Weights (Cont.))

4. *Suppose that* $\mathbf{f} \in \mathcal{F}^3(S)$ *and* (4) *holds. For the method to be at least third order, it is necessary that*

$$\mathbf{b}^\intercal \mathsf{A}\mathbf{c} = \frac{1}{6}.$$

Proof.

We sketch the proof and leave the details to the reader. To prove the result, use the general r-step method to approximate the solution to the linear scalar problem $u'(t) = u(t)$, $u(0) = 1$, whose exact solution is $u(t) = e^t$. At time $t = \tau$, the solution may be expressed as

$$u(\tau) = 1 + \tau + \frac{\tau^2}{2} + \frac{\tau^3}{6} + \frac{\tau^4}{24} e^\eta$$

for some $\eta \in (0, \tau)$.

Proof Cont.

For the RK stages, assume that the matrix $\mathsf{I} - \tau\mathsf{A}$ is invertible — this will always will be the case provided that τ is sufficiently small — and

$$(\mathsf{I} - \tau\mathsf{A})^{-1} = \mathsf{I} + \tau\mathsf{A} + \tau^2\mathsf{A}^2 + \tau^3\mathsf{A}^3 + \cdots .$$

It is possible to show that the vector of stages satisfies

$$\boldsymbol{\xi} = (\mathsf{I} - \tau\mathsf{A})^{-1}\mathbf{1}.$$

Solve explicitly for w^1 and compare the result with the expansion $u(\tau)$ above. □

Theorem (amplification factor I)

Applying an r-stage explicit RK method to approximate the solution of the differential equation $u'(t) = \lambda u(t)$, $u(0) = u_0$, one obtains

$$w^{k+1} = g(\lambda\tau)w^k, \quad g(z) = \sum_{j=0}^{r} \beta_j z^j, \quad w^0 = u_0,$$

where $\beta_j \in \mathbb{R}$, $j = 0, \ldots, r$. If the method is consistent to exactly order r, then

$$g(z) = \sum_{j=0}^{r} \frac{z^j}{j!},$$

i.e., $\beta_j = \frac{1}{j!}$ for $j = 0, \ldots, r$.

Proof.

This is a homework problem. □

Theorem (amplification factor II)

Applying an r-stage implicit RK method to approximate the solution of the differential equation $u'(t) = \lambda u(t)$, $u(0) = u_0$, one obtains

$$w^{k+1} = g(\lambda\tau)w^k, \quad g(z) = \frac{p_1(z)}{p_2(z)}, \quad w^0 = u_0,$$

where $p_1, p_2 \in \mathbb{P}_r$ and $p_2 \not\equiv 0$. In particular, g is the rational polynomial

$$g(z) = 1 + z\mathbf{b}^\intercal(\mathsf{I} - z\mathsf{A})^{-1}\mathbf{1} = \frac{\det(\mathsf{I} - z\mathsf{A} + z\mathbf{1}\,\mathbf{b}^\intercal)}{\det(\mathsf{I} - z\mathsf{A})},$$

where $\mathbf{1} = [1]_{i=1}^r \in \mathbb{R}^r$.

Proof.

This is also a homework problem. □

The Importance of Amplification Factors

Remark (amplification factor)

The function g (a polynomial in the ERK case and a rational function in the IRK case) that appears in the to previous Theorems is called the linear amplification factor or just the amplification factor. It will be important in the study of stability of numerical approximations to IVPs.

Remark (LTE)

The consistency error for any explicit RK method is defined in a straightforward way, since the RK stages can be computed explicitly in terms of the approximations. Specifically,

$$\tau\boldsymbol{\mathcal{E}}[\mathbf{u}](t,\tau) = \mathbf{u}(t) - \mathbf{u}(t-\tau) - \tau\sum_{i=1}^{r} b_i \mathbf{f}(t-\tau+c_i\tau, \boldsymbol{\xi}_{e,i}),$$

where $\boldsymbol{\xi}_{e,1} = \mathbf{u}(t-\tau)$ *and, for* $i = 2, \ldots, r$,

$$\boldsymbol{\xi}_{e,i} = \mathbf{u}(t-\tau) + \tau\sum_{j=1}^{i-1} a_{i,j}\mathbf{f}(t-\tau+c_j\tau, \boldsymbol{\xi}_{e,j}).$$

Notice that, in the end, the $\boldsymbol{\xi}_{e,i}$ *can be completely eliminated, which is a key insight. For implicit RK methods, the situation is a bit more complicated, as we shall see.*

Theorem (two-stage RK method)

Suppose that $\mathbf{f} \in \mathcal{F}^2(S)$*, so that* $\mathbf{u} \in C^3\big([0, T]; \mathbb{R}^d\big)$ *is a classical solution to the IVP* (1)*. Consider an explicit two-stage RK method given by the tableau*

$$\begin{array}{c|cc} 0 & 0 & 0 \\ c_2 & a_{2,1} & 0 \\ \hline & b_1 & b_2 \end{array}.$$

The method is consistent to order $p = 2$ *if and only if*

$$b_1 + b_2 = 1, \quad a_{2,1} = c_2, \quad b_2 c_2 = \frac{1}{2}.$$

Proof.

We have already done the autonomous case. The proof for the general, not-necessarily-autonomous case, is in the book. Notice that the necessary conditions from a previous Theorem are satisfied: $\mathbf{b}^\intercal \mathbf{1} = 1$, $\mathsf{A}\mathbf{1} = \mathbf{c}$, $\mathbf{b}^\intercal \mathbf{c} = \frac{1}{2}$. $\square$

Example

The following are second-order explicit methods.

1. Midpoint method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & 0 & 1 \end{array}.$$

2. Heun's method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

3. Ralston's method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{2}{3} & \frac{2}{3} & 0 \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array}.$$

Theorem (three-stage ERK methods)

Suppose that $\mathbf{f} \in \mathcal{F}^3(S)$*, so that* $\mathbf{u} \in C^4\left([0,T];\mathbb{R}^d\right)$ *is a classical solution to the IVP* (1)*. Consider an explicit three-stage RK method given by the tableau*

$$
\begin{array}{c|ccc}
0 & 0 & 0 & 0 \\
c_2 & a_{2,1} & 0 & 0 \\
c_3 & a_{3,1} & a_{3,2} & 0 \\
\hline
 & b_1 & b_2 & b_3
\end{array}.
$$

The method is consistent to order $p = 3$ *if and only if*

$$
b_1 + b_2 + b_3 = 1, \quad b_2 c_2 + b_3 c_3 = \frac{1}{2}, \quad b_2 c_2^2 + b_3 c_3^2 = \frac{1}{3}, \quad b_3 a_{3,2} c_2 = \frac{1}{6}.
$$

Proof.

The proof can be found in the classic book by Butcher (2008), or the book by Iserles (2009). □

Example

1. The following three-stage explicit RK method is consistent to exactly order $p = 3$:
$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & -1 & 2 & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}.$$
This method is called the classical RK method.

2. RK4: The following four-stage explicit RK method is consistent to exactly order $p = 4$:
$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}.$$
For a proof of the consistency, see Butcher (2008).

Theorem

An RK method (implicit or explicit) is of order 3 iff

1. $\sum_{j=1}^{r} b_j = 1$,
2. $\sum_{j=1}^{r} \sum_{k=1}^{r} b_j a_{j,k} = \frac{1}{2}$,
3. $\sum_{j=1}^{r} \sum_{k=1}^{r} \sum_{\ell=1}^{r} b_j a_{j,k} a_{j,\ell} = \frac{1}{3}$,
4. $\sum_{j=1}^{r} \sum_{k=1}^{r} \sum_{\ell=1}^{r} b_j a_{j,k} a_{k,\ell} = \frac{1}{6}$.

Proof.

The proof can be found in Hairer, Wanner, and Norsett (1993). □

Collocation Methods

Definition (RK collocation method)

Let $\mathbf{f} \in C\left(S; \mathbb{R}^d\right)$*. Suppose that the so-called* **collocation points** *satisfy*

$$0 \leq c_1 < c_2 < \cdots < c_r \leq 1.$$

Let $\mathbf{w}^k \in \mathbb{R}^d$ *be given. Assume that* $\mathbf{p}_k \in [\mathbb{P}_r]^d$ *satisfies, if possible,*

$$\mathbf{p}_k(t_k) = \mathbf{w}^k, \quad \mathbf{p}_k'(t_k + c_j\tau) = \mathbf{f}(t_k + c_j\tau, \mathbf{p}_k(t_k + c_j\tau)), \tag{5}$$

for $j = 1, \ldots, r$*. Define* $\mathbf{w}^{k+1} = \mathbf{p}_k(t_{k+1})$*, for* $k = 0, \ldots, K-1$*, with* $\mathbf{w}^0 = \mathbf{u}_0$*. This algorithm for producing the approximation sequence* $\left\{\mathbf{w}^k\right\}_{k=0}^K \subset \mathbb{R}^d$ *is called a* **Runge–Kutta collocation method***.*

Existence

Remark (existence)

The previous definition only makes sense if we can find a vector-valued polynomial

$$\mathbf{p}_k(t) = \sum_{j=0}^{r} \mathbf{a}_j t^j, \quad \mathbf{a}_j \in \mathbb{R}^d, \quad j = 0, \ldots, r$$

that satisfies (5). *If so, we say that the implicit r-stage RK collocation method is well defined. In fact, it may be the case that such a polynomial will not exist or will not be uniquely determined unless $\tau > 0$ is sufficiently small.*

Theorem (collocation)

Let $\{c_j\}_{j=1}^r \subset [0,1]$ *be a set of distinct collocation points. Suppose that a unique polynomial* $\mathbf{p}_k \in [\mathbb{P}_r]^d$ *satisfying* (5) *exists. Define*

$$\boldsymbol{\xi}_i = \mathbf{p}_k(t_k + c_i\tau), \quad i = 1, \ldots, r,$$

$$L_j(t) = \prod_{\substack{i=1 \\ i \neq j}}^{r} \frac{(t - c_i)}{(c_j - c_i)}, \quad j = 1, \ldots, r,$$

$$a_{i,j} = \int_0^{c_i} L_j(s)\,\mathrm{d}s, \quad b_j = \int_0^1 L_j(s)\,\mathrm{d}s, \quad i, j = 1, \ldots, r.$$

Then the collocation method of the previous definition is a standard implicit RK method, with the weights $\mathsf{A} = [a_{i,j}]$, $\mathbf{b} = [b_j]$, *and* $\mathbf{c} = [c_j]$, *the last weights being precisely the collocation points.*

Interpolation Facts

The elements

$$L_j(t) = \prod_{\substack{i=1 \\ i \neq j}}^{r} \frac{(t - c_i)}{(c_j - c_i)}, \quad j = 1, \dots, r,$$

are the called the Lagrange basis functions. They are polynomials of degree $r-1$ and satisfy

$$L_j(c_i) = \delta_{i,j}, \quad 1 \leq i, j \leq r.$$

The unique degree $r-1$ polynomial interpolating the points (c_i, y_i), $1 \leq i \leq r$, satisfies

$$q(t) = \sum_{j=1}^{r} L_j(t) y_j.$$

Proof of the Collocation Theorem.

Suppose that $\mathbf{p}_k \in [\mathbb{P}_r]^d$ satisfies (5). Consider the unique Lagrange interpolating polynomial of degree at most $r-1$, $\boldsymbol{\rho} \in [\mathbb{P}_{r-1}]^d$ such that

$$\boldsymbol{\rho}(t_k + c_j\tau) = \mathbf{p}_k'(t_k + c_j\tau) = \mathbf{f}(t_k + c_j\tau, \mathbf{p}_k(t_k + c_j\tau)) = \boldsymbol{\nu}_j, \quad j = 1, \ldots, r.$$

Interpolation theory guarantees that

$$\boldsymbol{\rho}(t) = \sum_{j=1}^{r} L_j\left(\frac{t - t_k}{\tau}\right)\boldsymbol{\nu}_j.$$

Observe that $\mathbf{p}_k' \in [\mathbb{P}_{r-1}]^d$ and, in fact,

$$\mathbf{p}_k'(t_k + c_j\tau) = \boldsymbol{\rho}(t_k + c_j\tau), \quad j = 1, \ldots, r.$$

Therefore, $\mathbf{p}_k' \equiv \boldsymbol{\rho}$, since these polynomials (of degree at most $r-1$) agree at r points. By (5),

$$\mathbf{p}_k'(t) = \sum_{j=1}^{r} L_j\left(\frac{t - t_k}{\tau}\right)\mathbf{f}(t_k + c_j\tau, \mathbf{p}_k(t_k + c_j\tau)).$$

Proof (Cont.)

Integrating the expression

$$\mathbf{p}_k'(t) = \sum_{j=1}^{r} L_j\left(\frac{t-t_k}{\tau}\right)\mathbf{f}(t_k + c_j\tau, \mathbf{p}_k(t_k + c_j\tau)).$$

and using the condition $\mathbf{p}_k(t_k) = \mathbf{w}^k$, we observe that

$$\begin{aligned}\mathbf{p}_k(t) &= \mathbf{w}^k + \int_{t_k}^{t}\sum_{j=1}^{r} L_j\left(\frac{s-t_k}{\tau}\right)\mathbf{f}(t_k + c_j\tau, \mathbf{p}_k(t_k + c_j\tau))\,\mathrm{d}s \\ &= \mathbf{w}^k + \tau\sum_{j=1}^{r}\mathbf{f}(t_k + c_j\tau, \boldsymbol{\xi}_j)\int_0^{\frac{t-t_k}{\tau}} L_j(s)\,\mathrm{d}s.\end{aligned} \tag{6}$$

Proof (Cont.)

Setting $t = t_k + c_i\tau$ in (6), we have

$$\boldsymbol{\xi}_i = \mathbf{w}^k + \tau \sum_{j=1}^{r} a_{i,j}\mathbf{f}(t_k + c_j\tau, \boldsymbol{\xi}_j).$$

Setting $t = t_{k+1}$ in (6), we find

$$\mathbf{w}^{k+1} = \mathbf{w}^k + \tau \sum_{j=1}^{r} b_j\mathbf{f}(t_k + c_j\tau, \boldsymbol{\xi}_j),$$

and the proof is finished. $\square$

Theorem (collocation order)

Suppose that $\{c_j\}_{j=1}^r \subset [0,1]$ is the set of r distinct collocation points that determine the RK collocation method. Define

$$q(t) = \prod_{n=1}^{r}(t - c_n) \in \mathbb{P}_r.$$

If, for some $m \in \{1, \ldots, r\}$,

$$\int_0^1 q(s)p(s)\mathrm{d}s = 0, \quad \forall p \in \mathbb{P}_{m-1},$$

but there is some $\tilde{p} \in \mathbb{P}_m$ such that

$$\int_0^1 q(s)\tilde{p}(s)\mathrm{d}s \neq 0,$$

then the RK collocation method is consistent to exactly order $p = r + m$.

Proof.

The proof is in the textbook. □

Orthogonal Polynomials: The Transformed Legendre Polynomials

Definition (transformed Legendre polynomials)

By $\left\{\tilde{P}_j\right\}_{j\in\mathbb{N}_0}$ *we denote the set of* **transformed Legendre polynomials**, *which have the property that*

$$\int_0^1 \tilde{P}_i(s)\tilde{P}_j(s)\mathrm{d}s = \frac{1}{2j+1}\delta_{i,j}.$$

j	$\tilde{P}_j(t)$
0	1
1	$2t-1$
2	$6t^2-6t+1$
3	$20t^3-30t^2+12t-1$
4	$70t^4-140t^3+90t^2-20t+1$
5	$252t^5-630t^4+560t^3-210t^2+30t-1$

Table: The first six transformed Legendre polynomials.

Corollary (Gauss–Legendre–RK)

Let the collocation points $c_1, \ldots, c_r$ be precisely the zeros of the transformed Legendre polynomial $\tilde{P}_r \in \mathbb{P}_r$. According to a Theorem from Chapter 11, these lie in the open interval $(0, 1)$. Then the corresponding collocation method is consistent to order exactly $p = 2r$.

Proof.

In this case, $q \equiv C_r \tilde{P}_r$, where $0 \neq C_r \in \mathbb{R}$. Since the $\tilde{P}_i$ form an orthogonal basis for $\mathbb{P}_r$, for any $j \in \{0, 1, 2, \ldots, r\}$, we can express

$$t^j = \sum_{m=0}^{j} \beta_{j,m} \tilde{P}_m(t)$$

for some constants $\beta_{j,1}, \ldots, \beta_{j,j}$.

Proof (Cont.)

Therefore,

$$\int_0^1 q(s)s^j \mathrm{d}s = C_r \sum_{m=0}^{j} \beta_{j,m} \int_0^1 \tilde{P}_r(s)\tilde{P}_m(s)\mathrm{d}s = 0,$$

provided that $j \leq r-1$. By the previous theorem, the method is exactly of order $p = r + r = 2r$. $\square$

Definition (Gauss–Legendre–RK method)

The implicit r-stage RK methods constructed as collocation methods whose collocation points are the zeros of the transformed Legendre polynomial $\tilde{P}_r$ are called **Gauss–Legendre–Runge–Kutta methods***.*

Midpoint Rule

Suppose that $r = 1$. The transformed Legendre polynomial of order one is

$$\tilde{P}_1(t) = 2t - 1 \implies c_1 = \frac{1}{2}.$$

The corresponding Gauss–Legendre IRK method is given by

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array}$$

and is of order $2r = 2$. For a scalar autonomous system, $u' = f(u)$, the method can be expressed as

$$w^{k+1} = w^k + \tau f\left(w^k + \frac{\tau}{2}\kappa_1\right), \quad \kappa_1 = f\left(w^k + \frac{\tau}{2}\kappa_1\right). \tag{7}$$

It is a simple exercise to show that this is equivalent to the midpoint rule,

$$w^{k+1} = w^k + \tau f\left(\frac{w^{k+1} + w^k}{2}\right). \tag{8}$$

Midpoint Rule

But let us write this another way. Define

$$\tilde{w}^{k+\frac{1}{2}} = \frac{w^{k+1} + w^k}{2}.$$

Then we can express the midpoint rule as

$$\tilde{w}^{k+\frac{1}{2}} = w^k + \frac{\tau}{2} f\left(\tilde{w}^{k+\frac{1}{2}}\right), \quad w^{k+1} = 2\tilde{w}^{k+\frac{1}{2}} - w^k. \tag{9}$$

Still another, equivalent, way of writing this method is as follows:

$$\tilde{w}^{k+\frac{1}{2}} = w^k + \frac{\tau}{2} f\left(\tilde{w}^{k+\frac{1}{2}}\right), \quad w^{k+1} = \tilde{w}^{k+\frac{1}{2}} + \frac{\tau}{2} f\left(\tilde{w}^{k+\frac{1}{2}}\right). \tag{10}$$

Observe that method (9) shows that the midpoint rule is essentially a backward (implicit) Euler method with half the time step size followed by an extrapolation. Method (10) expresses the midpoint rule as a half-step-size backward Euler method followed by a half-step-size forward (explicit) Euler method.

Example

Suppose that $r = 2$. The transformed Legendre polynomial of order two is

$$\tilde{P}_2(t) = 6t^2 - 6t + 1 \implies c_1 = \frac{1}{2} - \frac{\sqrt{3}}{6},\; c_2 = \frac{1}{2} + \frac{\sqrt{3}}{6}.$$

The Gauss–Legendre IRK method is given by

$$\begin{array}{c|cc} \frac{1}{2} - \frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{2} + \frac{\sqrt{3}}{6} & \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

and is of order $2r = 4$.

Example

Suppose that $r = 3$. The transformed Legendre polynomial of order three is

$$\tilde{P}_3(t) = 20t^3 - 30t^2 + 12t - 1 \implies c_1 = \frac{1}{2} - \frac{\sqrt{15}}{10},\ c_2 = \frac{1}{2},\ c_3 = \frac{1}{2} + \frac{\sqrt{15}}{10}.$$

The Gauss–Legendre IRK method is given by

$$\begin{array}{c|ccc}
\frac{1}{2} - \frac{\sqrt{15}}{10} & \frac{5}{36} & \frac{2}{9} - \frac{\sqrt{15}}{15} & \frac{5}{36} - \frac{\sqrt{15}}{30} \\
\frac{1}{2} & \frac{5}{36} + \frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36} - \frac{\sqrt{15}}{24} \\
\frac{1}{2} + \frac{\sqrt{15}}{10} & \frac{5}{36} + \frac{\sqrt{15}}{30} & \frac{2}{9} + \frac{\sqrt{15}}{15} & \frac{5}{36} \\
\hline
 & \frac{5}{18} & \frac{4}{9} & \frac{5}{18}
\end{array}$$

and is of order $2r = 6$.

Example

Not all IRK methods are of collocation type. Consider, for example, the methods given by the tables

$$\begin{array}{c|cc} 0 & \frac{1}{4} & -\frac{1}{4} \\ \frac{2}{3} & \frac{1}{4} & \frac{5}{12} \\ \hline & \frac{1}{4} & \frac{3}{4} \end{array} \qquad \begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array}.$$

For both methods, the necessary conditions for consistency are satisfied. The method on the left, which is consistent to exactly order $p = 3$, is not of collocation type. (How do we know this?) The method on the right is of collocation type. One can check that the collocation points $c_1 = \frac{1}{3}$ and $c_2 = 1$ completely determine the other weights. The method on the right is consistent to exactly order $p = 3$. To see this, observe that

$$\int_0^1 \left(s - \frac{1}{3}\right)(s-1)s^j \mathrm{d}s = 0$$

only for $j = 0$. Thus, invoking the collocation-order theorem, we have $m = 1$ and $p = r + m = 3$.