Introduction to the Numerical Solution of IVP for ODE

Consider the IVP: DE \( x' = f(t, x) \), IC \( x(a) = x_a \). For simplicity, we will assume here that \( x(t) \in \mathbb{R}^n \) (so \( F = \mathbb{R} \)), and that \( f(t, x) \) is continuous in \( t, x \) and uniformly Lipschitz in \( x \) (with Lipschitz constant \( L \)) on \([a, b] \times \mathbb{R}^n \). So we have global existence and uniqueness for the IVP on \([a, b] \).

Moreover, the solution of the IVP \( x' = f(t, x) \), \( x(a) = x_a \) depends continuously on the initial values \( x_a \in \mathbb{R}^n \). This IVP is an example of a well-posed problem: for each choice of the “data” (here, the initial values \( x_a \)), we have:

1. **Existence.** There exists a solution of the IVP on \([a, b] \).

2. **Uniqueness.** The solution, for each given \( x_a \), is unique.

3. **Continuous Dependence.** The solution depends continuously on the data.

Here, e.g., the map \( x_a \mapsto x(t, x_a) \) is continuous from \( \mathbb{R}^n \) into \( (C([a, b]), \| \cdot \|_\infty) \). A well-posed problem is a reasonable problem to approximate numerically.

**Grid Functions**

Choose a mesh width \( h \) (with \( 0 < h \leq b - a \)), and let \( N = \left\lfloor \frac{b-a}{h} \right\rfloor \) (greatest integer \( \leq (b-a)/h \)). Let \( t_i = a + ih \) \((i = 0, 1, \ldots, N)\) be the grid points in \( t \) (note: \( t_0 = a \)), and let \( x_i \) denote the approximation to \( x(t_i) \). Note that \( t_i \) and \( x_i \) depend on \( h \), but we will usually suppress this dependence in our notation.

**Explicit One-Step Methods**

Form of method: start with \( x_0 \) (presumably \( x_0 \approx x_a \)). Recursively compute \( x_1, \ldots, x_N \) by

\[
x_{i+1} = x_i + h\psi(h, t_i, x_i), \quad i = 0, \ldots, N - 1.
\]

Here, \( \psi(h, t, x) \) is a function defined for \( 0 \leq h \leq b-a \), \( a \leq t \leq b \), \( x \in \mathbb{R}^n \), and \( \psi \) is associated with the given function \( f(t, x) \).

**Examples.**

*Euler’s Method.*

\[
x_{i+1} = x_i + hf(t_i, x_i)
\]

Here, \( \psi(h, t, x) = f(t, x) \).
Taylor Methods.

Let $p$ be an integer $\geq 1$. To see how the Taylor Method of order $p$ is constructed, consider the Taylor expansion of a $C^{p+1}$ solution $x(t)$ of $x' = f(t, x)$:

$$x(t + h) = x(t) + hx'(t) + \cdots + \frac{h^p}{p!}x^{(p)}(t) + O(h^{p+1})$$

where the remainder term is $O(h^{p+1})$ by Taylor’s Theorem with remainder. In the approximation, we will neglect the remainder term, and use the DE $x' = f(t, x)$ to replace $x'(t), x''(t), \ldots$ by expressions involving $f$ and its derivatives:

$$x'(t) = f(t, x(t))$$

$$x''(t) = \frac{d}{dt}(f(t, x(t))) = \left. D_t f \right|_{(t, x(t))} + \left. D_x f \right|_{(t, x(t))} \frac{dx}{dt}$$

(for $n = 1$, this is $f_t + f_x f$).

For higher derivatives, inductively differentiate the expression for the previous derivative, and replace any occurrence of $\frac{dx}{dt}$ by $f(t, x(t))$. These expansions lead us to define the Taylor methods of order $p$:

$$p = 1: \quad x_{i+1} = x_i + hf(t_i, x_i) \quad \text{(Euler’s method, } \psi(h, t, x) = f(t, x))$$

$$p = 2: \quad x_{i+1} = x_i + hf(t_i, x_i) + \frac{h^2}{2} \left. (D_t f + (D_x f)f) \right|_{(t_i, x_i)}$$

For the case $p = 2$, we have

$$\psi(h, t, x) = T_2(h, t, x) \equiv \left( f + \frac{h}{2} (D_t f + (D_x f)f) \right) \bigg|_{(t, x)}$$

We will use the notation $T_p(h, t, x)$ to denote the $\psi(h, t, x)$ function for the Taylor method of order $p$.

**Remark.** Taylor methods of order $\geq 2$ are rarely used computationally. They require derivatives of $f$ to be programmed and evaluated. They are, however, of theoretical interest in determining the order of a method.

**Remark.** A “one-step method” is actually an association of a function $\psi(h, t, x)$ (defined for $0 \leq h \leq b - a, a \leq t \leq b, x \in \mathbb{R}^n$) to each function $f(t, x)$ (which is continuous in $t, x$ and Lipschitz in $x$ on $[a, b] \times \mathbb{R}^n$). We study “methods” looking at one function $f$ at a time. Many methods (e.g., Taylor methods of order $p \geq 2$) require more smoothness of $f$, either for their
For “higher-order” methods, this smoothness is essential in getting the error to be higher order in \( h \). We will assume from here on (usually tacitly) that \( f \) is sufficiently smooth when needed.

Examples.

**Modified Euler’s Method**

\[ x_{i+1} = x_i + hf \left( t_i + \frac{h}{2}, x_i + \frac{h}{2}f(t_i, x_i) \right) \]

(so \( \psi(h, t, x) = f \left( t + \frac{h}{2}, x + \frac{h}{2}f(t, x) \right) \)).

Here \( \psi(h, t, x) \) tries to approximate

\[ x' \left( t + \frac{h}{2} \right) \approx f \left( t + \frac{h}{2}, x \left( t + \frac{h}{2} \right) \right), \]

using the Euler approximation to \( x \left( t + \frac{h}{2} \right) \) (\( \approx x(t) + \frac{h}{2}f(t, x(t)) \)).

**Improved Euler’s Method** (or Heun’s Method)

\[ x_{i+1} = x_i + \frac{h}{2} \left( f(t_i, x_i) + f(t_{i+1}, x_i + hf(t_i, x_i)) \right) \]

(so \( \psi(h, t, x) = \frac{1}{2} (f(t, x) + f(t+h, x + hf(t, x))) \)).

Here again \( \psi(h, t, x) \) tries to approximate

\[ x' \left( t + \frac{h}{2} \right) \approx \frac{1}{2} (x'(t) + x'(t+h)). \]

Or \( \psi(h, t, x) \) can be viewed as an approximation to the trapezoid rule applied to

\[ \frac{1}{h} (x(t+h) - x(t)) = \frac{1}{h} \int_t^{t+h} x' \approx \frac{1}{2} x'(t) + \frac{1}{2} x'(t+h). \]

Modified Euler and Improved Euler are examples of 2\(^{nd}\) order two-stage Runge-Kutta methods. Notice that no derivatives of \( f \) need be evaluated, but \( f \) needs to be evaluated twice in each step (from \( x_i \) to \( x_{i+1} \)).

Before stating the convergence theorem, we introduce the concept of *accuracy*.

**Local Truncation Error**

Let \( x_{i+1} = x_i + h\psi(h, t_i, x_i) \) be a one-step method, and let \( x(t) \) be a solution of the DE \( x' = f(t, x) \). The *local truncation error* (LTE) for \( x(t) \) is defined to be

\[ l(h, t) \equiv x(t+h) - (x(t) + h\psi(h, t, x(t))), \]
that is, the local truncation error is the amount by which the true solution of the DE fails to satisfy the numerical scheme. \( l(h, t) \) is defined for those \((h, t)\) for which \(0 < h \leq b - a\) and \(a \leq t \leq b - h\).

Define
\[
\tau(h, t) = \frac{l(h, t)}{h}
\]
and set \( \tau_i(h) = \tau(h, t_i) \). Also, set
\[
\tau(h) = \max_{a \leq t \leq b - h} |\tau(h, t)| \quad \text{for } 0 < h \leq b - a.
\]

Note that \( l(h, t_i) = x(t_{i+1}) - (x(t_i) + h\psi(h, t_i, x(t_i))) \), explicitly showing the dependence of \( l \) on \( h, t_i \), and \( x(t) \).

**Definition.** A one-step method is called formally accurate of order \( p \) (for a positive integer \( p \)) if for any solution \( x(t) \) of the DE \( x' = f(t, x) \) which is \( C^{p+1} \), we have \( l(h, t) = O(h^{p+1}) \).

**Definition.** A one-step method is called consistent if \( \psi(0, t, x) = f(t, x) \). Consistency is essentially minimal accuracy:

**Proposition.** A one-step method
\[
x_{i+1} = x_i + h\psi(h, t_i, x_i),
\]
where \( \psi(h, t, x) \) is continuous for \(0 \leq h \leq h_0, a \leq t \leq b, x \in \mathbb{R}^n \) for some \( h_0 \in (0, b - a) \), is consistent with the DE \( x' = f(t, x) \) if and only if \( \tau(h) \to 0 \) as \( h \to 0^+ \).

**Proof.** Suppose the method is consistent. Fix a solution \( x(t) \). For \( 0 < h \leq h_0 \), let
\[
Z(h) = \max_{a \leq s, t \leq b, |s-t| \leq h} |\psi(0, s, x(s)) - \psi(h, t, x(t))|.
\]

By uniform continuity, \( Z(h) \to 0 \) as \( h \to 0^+ \). Now
\[
l(h, t) = x(t+h) - x(t) - h\psi(h, t, x(t))
= \int_t^{t+h} [x'(s) - \psi(h, t, x(t))] \, ds
= \int_t^{t+h} [f(s, x(s)) - \psi(h, t, x(t))] \, ds
= \int_t^{t+h} [\psi(0, s, x(s)) - \psi(h, t, x(t))] \, ds,
\]
so \( |l(h, t)| \leq hZ(h) \). Therefore \( \tau(h) \leq Z(h) \to 0 \).

Conversely, suppose \( \tau(h) \to 0 \). For any \( t \in [a, b) \) and any \( h \in (0, b - t] \),
\[
\frac{x(t+h) - x(t)}{h} = \psi(h, t, x(t)) + \tau(h, t).
\]
Taking the limit as \( h \downarrow 0 \) gives \( f(t, x(t)) = x'(t) = \psi(0, t, x(t)) \).

□
Introduction to the Numerical Solution of IVP for ODE

Convergence Theorem for One-Step Methods

Theorem. Suppose $f(t, x)$ is continuous in $t, x$ and uniformly Lipschitz in $x$ on $[a, b] \times \mathbb{R}^n$. Let $x(t)$ be the solution of the IVP $x' = f(t, x)$, $x(a) = x_a$ on $[a, b]$. Suppose that the function $\psi(h, t, x)$ in the one step method satisfies the following two conditions:

1. (Stability) $\psi(h, t, x)$ is continuous in $h, t, x$ and uniformly Lipschitz in $x$ (with Lipschitz constant $K$) on $0 \leq h \leq h_0$, $a \leq t \leq b$, $x \in \mathbb{R}^n$ for some $h_0 > 0$ with $h_0 \leq b - a$, and

2. (Consistency) $\psi(0, t, x) = f(t, x)$.

Let $e_i(h) = x(t_i(h)) - x_i(h)$, where $x_i$ is obtained from the one-step method $x_{i+1} = x_i + h\psi(h, t_i, x_i)$. (Note that $e_0(h) = x_a - x_0(h)$ is the error in the initial value $x_0(h)$.) Then

$$|e_i(h)| \leq e^{K(t_i(h) - a)}|e_0(h)| + \tau(h) \left( \frac{e^{K(t_i(h) - a)} - 1}{K} \right),$$

so

$$|e_i(h)| \leq e^{K(b-a)}|e_0(h)| + \frac{e^{K(b-a)} - 1}{K} \tau(h).$$

Moreover, $\tau(h) \to 0$ as $h \to 0$. Therefore, if $e_0(h) \to 0$ as $h \to 0$, then

$$\max_{0 \leq i \leq \frac{b-a}{h}} |e_i(h)| \to 0 \quad \text{as} \quad h \to 0,$$

that is, the approximations converge uniformly on the grid to the solution.

Proof. Hold $h > 0$ fixed, and ignore rounding error. Subtracting

$$x_{i+1} = x_i + h\psi(h, t_i, x_i)$$

from

$$x(t_{i+1}) = x(t_i) + h\psi(h, t_i, x(t_i)) + h\tau_i,$$

gives

$$|e_{i+1}| \leq |e_i| + h|\psi(h, t_i, x(t_i)) - \psi(h, t_i, x_i)| + h|\tau_i|$$

$$\quad \leq |e_i| + hK|e_i| + h\tau(h)$$

$$\quad = (1 + hK)|e_i| + h\tau(h).$$

So

$$|e_1| \leq (1 + hK)|e_0| + h\tau(h),$$

and

$$|e_2| \leq (1 + hK)|e_1| + h\tau(h)$$

$$\quad \leq (1 + hK)^2|e_0| + h\tau(h)(1 + (1 + hK)).$$

By induction,

$$|e_i| \leq (1 + hK)^i|e_0| + h\tau(h)(1 + (1 + hK) + (1 + hK)^2 + \cdots + (1 + hK)^{i-1})$$

$$\quad = (1 + hK)^i|e_0| + h\tau(h) \frac{(1 + hK)^i - 1}{(1 + hK) - 1}$$

$$\quad = (1 + hK)^i|e_0| + \tau(h) \frac{(1 + hK)^i - 1}{K}.$$
Since \((1 + hK)^{\frac{1}{i}} \uparrow e^K\) as \(h \to 0^+\) (for \(K > 0\)), and \(i = \frac{i_n - a}{h}\), we have
\[
(1 + hK)^{i} = (1 + hK)^{\frac{i_n - a}{h}} \leq e^{K(t_n - a)}.
\]
Thus
\[
|e_i| \leq e^{K(t_n - a)}|e_0| + \tau(h)\frac{e^{K(t_n - a)} - 1}{K}.
\]
The preceding proposition shows \(\tau(h) \to 0\), and the theorem follows. \(\square\)

If \(f\) is sufficiently smooth, then we know that \(x(t) \in C^{p+1}\). The theorem thus implies that if a one-step method is accurate of order \(p\) and stable [i.e. \(\psi\) is Lipschitz in \(x\)], then for sufficiently smooth \(f\),
\[
l(h, t) = O(h^{p+1}) \quad \text{and thus} \quad \tau(h) = O(h^p).
\]
If, in addition, \(e_0(h) = O(h^p)\), then
\[
\max_i |e_i(h)| = O(h^p),
\]
i.e. we have \(p^{th}\) order convergence of the numerical approximations to the solution.

**Example.** The “Taylor method of order \(p\)” is accurate of order \(p\). If \(f \in C^p\), then \(x \in C^{p+1}\), and
\[
l(h, t) = x(t + h) - \left( x(t) + h x'(t) + \cdots + \frac{h^p}{p!} x^{(p)}(t) \right) = \frac{1}{p!} \int_t^{t+h} (t + h - s)^p x^{(p+1)}(s) ds.
\]
So
\[
|l(h, t)| \leq M_{p+1} \frac{h^{p+1}}{(p+1)!} \quad \text{where} \quad M_{p+1} = \max_{a \leq t \leq b} |x^{(p+1)}(t)|.
\]

**Fact.** A one-step method \(x_{i+1} = x_i + h \psi(h, t, x_i)\) is accurate of order \(p\) if and only if
\[
\psi(h, t, x) = T_p(h, t, x) + O(h^p),
\]
where \(T_p\) is the “\(\psi\)” for the Taylor method of order \(p\).

**Proof.** Since
\[
x(t + h) - x(t) = h T_p(h, t, x(t)) + O(h^{p+1}),
\]
we have for any given one-step method that
\[
l(h, t) = x(t + h) - x(t) - h \psi(h, t, x(t))
= h T_p(h, t, x(t)) + O(h^{p+1}) - h \psi(h, t, x(t))
= h (T_p(h, t, x(t)) - \psi(h, t, x(t))) + O(h^{p+1}).
\]
So \(l(h, t) = O(h^{p+1})\) iff \(h(T_p - \psi) = O(h^{p+1})\) iff \(\psi = T_p + O(h^p)\). \(\square\)

**Remark.** The controlled growth of the effect of the local truncation error (LTE) from previous steps in the proof of the convergence theorem (a consequence of the Lipschitz continuity of \(\psi\) in \(x\)) is called *stability*. The theorem states:

\[
\text{Stability} + \text{ Consistency (minimal accuracy)} \quad \Rightarrow \quad \text{Convergence}.
\]
In fact, here, the converse is also true.
Explicit Runge-Kutta methods

One of the problems with Taylor methods is the need to evaluate higher derivatives of \( f \). Runge-Kutta (RK) methods replace this with the much more reasonable need to evaluate \( f \) more than once to go from \( x_i \) to \( x_{i+1} \). An \( m \)-stage (explicit) RK method is of the form

\[
x_{i+1} = x_i + h \psi(h, t_i, x_i),
\]

with

\[
\psi(h, t, x) = \sum_{j=1}^{m} a_j k_j(h, t, x),
\]

where \( a_1, \ldots, a_m \) are given constants,

\[
k_1(h, t, x) = f(t, x)
\]

and for \( 2 \leq j \leq m \),

\[
k_j(h, t, x) = f(t + \alpha_j h, x + h \sum_{r=1}^{j-1} \beta_{jr} k_r(h, t, x))
\]

with \( \alpha_2, \ldots, \alpha_m \) and \( \beta_{jr} \) (\( 1 \leq r < j \leq m \)) given constants. We usually choose \( 0 < \alpha_j \leq 1 \), and for accuracy reasons we choose

\[
(*) \quad \alpha_j = \sum_{r=1}^{j-1} \beta_{jr} \quad \text{ (} 2 \leq j \leq m \text{)}.
\]

**Example.** \( m = 2 \)

\[
x_{i+1} = x_i + h(a_1 k_1(h, t_i, x_i) + a_2 k_2(h, t_i, x_i))
\]

where

\[
k_1(h, t, x) = f(t_i, x_i)
\]

\[
k_2(h, t, x) = f(t_i + \alpha_2 h, x_i + h \beta_{21} k_1(h, t_i, x_i)).
\]

For simplicity, write \( \alpha \) for \( \alpha_2 \) and \( \beta \) for \( \beta_2 \). Expanding in \( h \),

\[
k_2(h, t, x) = f(t + \alpha h, x + h \beta f(t, x))
\]

\[
= f(t, x) + \alpha h D_t f(t, x) + (D_x f(t, x))(h \beta f(t, x)) + O(h^2)
\]

\[
= [f + h(\alpha D_t f + \beta (D_x f)f)](t, x) + O(h^2).
\]

So

\[
\psi(h, t, x) = (a_1 + a_2)f + h(a_2 \alpha D_t f + a_2 \beta (D_x f)f) + O(h^2).
\]

Recalling that

\[
T_2 = f + \frac{h}{2}(D_t f + (D_x f)f),
\]
and that the method is accurate of order two if and only if

$$\psi = T_2 + O(h^2),$$

we obtain the following necessary and sufficient conditions on a two-stage (explicit) RK method to be accurate of order two:

$$a_1 + a_2 = 1, \quad a_2\alpha = \frac{1}{2}, \quad \text{and} \quad a_2\beta = \frac{1}{2}.$$  

We require $\alpha = \beta$ as in (*) (we now see why this condition needs to be imposed), whereupon these conditions become:

$$a_1 + a_2 = 1, \quad a_2\alpha = \frac{1}{2}.$$  

Therefore, there is a one-parameter family (e.g., parameterized by $\alpha$) of 2nd order, two-stage ($m = 2$) explicit RK methods.

**Examples.**

(1) Setting $\alpha = \frac{1}{2}$ gives $a_2 = 1$, $a_1 = 0$, which is the Modified Euler method.

(2) Choosing $\alpha = 1$ gives $a_2 = \frac{1}{2}$, $a_1 = \frac{1}{2}$, which is the Improved Euler method, or Heun’s method.

**Remark.** Note that an $m$-stage explicit RK method requires $m$ function evaluations (i.e., evaluations of $f$) in each step ($x_i$ to $x_{i+1}$).

**Attainable Orders of Accuracy for Explicit RK methods**

<table>
<thead>
<tr>
<th># of stages ($m$)</th>
<th>highest order attainable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

Explicit RK methods are always stable: $\psi$ inherits its Lipschitz continuity from $f$.

**Example.** Let $L$ be the Lipschitz constant for $f$, and suppose $h \leq h_0$ (for some $h_0 \leq b - a$).

$$x_{i+1} = x_i + hf(t_i + \frac{h}{2}, x_i + \frac{h}{2}f(t_i, x_i))$$

$$\psi(t, h, x) = f\left(t + \frac{h}{2}, x + \frac{h}{2}f(t, x)\right)$$
So
\[
\left| \psi(h, t, x) - \psi(h, t, y) \right| \leq L \left| \left( x + \frac{h}{2} f(t, x) \right) - \left( y + \frac{h}{2} f(t, y) \right) \right|
\]
\[
\leq L |x - y| + \frac{h}{2} L |f(t, x) - f(t, y)|
\]
\[
\leq L |x - y| + \frac{h}{2} L^2 |x - y|
\]
\[
\leq K |x - y|
\]
where \( K = L + \frac{h}{2} L^2 \) is thus the Lipschitz constant for \( \psi \).

**Example.** A popular 4th order four-stage RK method is
\[
x_{i+1} = x_i + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4)
\]
where
\[
k_1 = f(t_i, x_i)
k_2 = f \left( t_i + \frac{h}{2}, x_i + \frac{h}{2} k_1 \right)
k_3 = f \left( t_i + \frac{h}{2}, x_i + \frac{h}{2} k_2 \right)
k_4 = f \left( t_i + h, x_i + hk_3 \right).
\]
The same argument as above shows this method is stable.

**Remark.** RK methods require multiple function evaluations per step (going from \( x_i \) to \( x_{i+1} \)). One-step methods discard information from previous steps (e.g., \( x_{i-1} \) is not used to get \( x_{i+1} \) — except in its influence on \( x_i \)). We will next study a class of multi-step methods. But first, we consider linear difference equations.

**Linear Difference Equations (Constant Coefficients)**

In this discussion, \( x_i \) will be a (scalar) sequence defined for \( i \geq 0 \). Consider the linear difference equation (\( k \)-step)
\[
(\text{LDE}) \quad x_{i+k} + \alpha_{k-1} x_{i+k-1} + \cdots + \alpha_0 x_i = b_i \quad (i \geq 0).
\]
If \( b_i \equiv 0 \), the linear difference equation (LDE) is said to be homogeneous, in which case we will refer to it as (\( lh \)). If \( b_i \neq 0 \) for some \( i \geq 0 \), the linear difference equation (LDE) is said to be inhomogeneous, in which case we refer to it as (\( li \)).

**Initial Value Problem (IVP):** Given \( x_i \) for \( i = 0, \ldots, k-1 \), determine \( x_i \) satisfying (LDE) for \( i \geq 0 \).

**Theorem.** There exists a unique solution of (IVP) for (\( lh \)) or (\( li \)).

**Proof.** An obvious induction on \( i \). The equation for \( i = 0 \) determines \( x_k \), etc. \( \square \)

**Theorem.** The solution set of (\( lh \)) is a \( k \)-dimensional vector space (a subspace of the set of all sequences \( \{ x_i \}_{i \geq 0} \)).
**Proof Sketch.** Choosing
\[
\begin{bmatrix}
x_0 \\
x_1 \\
\vdots \\
x_{k-1}
\end{bmatrix} = e_j \in \mathbb{R}^k
\]
for \(j = 1, 2, \ldots, k\) and then solving (lh) gives a basis of the solution space of (lh).
\(\square\)

Define the *characteristic polynomial* of (lh) to be
\[
p(r) = r^k + \alpha_{k-1}r^{k-1} + \cdots + \alpha_0.
\]
Let us assume that \(\alpha_0 \neq 0\). (If \(\alpha_0 = 0\), (LDE) isn’t really a \(k\)-step difference equation since we can shift indices and treat it as a \(\tilde{k}\)-step difference equation for a \(\tilde{k} < k\), namely \(\tilde{k} = k - \nu\), where \(\nu\) is the smallest index with \(\alpha_\nu \neq 0\).) Let \(r_1, \ldots, r_s\) be the distinct zeroes of \(p\), with multiplicities \(m_1, \ldots, m_s\). Note that each \(r_j \neq 0\) since \(\alpha_0 \neq 0\), and \(m_1 + \cdots + m_s = k\). Then a basis of solutions of (lh) is:
\[
\left\{ \left\{ i^l r_j^i \right\}_{i=0}^{\infty} : 1 \leq j \leq s, \ 0 \leq l \leq m_j - 1 \right\}.
\]

**Example.** Fibonacci Sequence:
\[
F_{i+2} - F_{i+1} - F_i = 0, \quad F_0 = 0, \quad F_1 = 1.
\]
The associated characteristic polynomial \(r^2 - r - 1 = 0\) has roots
\[
r_\pm = \frac{1 \pm \sqrt{5}}{2} \quad (r_+ \approx 1.6, r_- \approx -0.6).
\]
The general solution of (lh) is
\[
F_i = C_+ \left( \frac{1 + \sqrt{5}}{2} \right)^i + C_- \left( \frac{1 - \sqrt{5}}{2} \right)^i.
\]
The initial conditions \(F_0 = 0\) and \(F_1 = 1\) imply that \(C_+ = \frac{1}{\sqrt{5}}\) and \(C_- = -\frac{1}{\sqrt{5}}\). Hence
\[
F_i = \frac{1}{\sqrt{5}} \left( \left( \frac{1 + \sqrt{5}}{2} \right)^i - \left( \frac{1 - \sqrt{5}}{2} \right)^i \right).
\]
Since \(|r_-| < 1\), we have
\[
\left( \frac{1 - \sqrt{5}}{2} \right)^i \to 0 \text{ as } i \to \infty.
\]
Hence, the Fibonacci sequence behaves asymptotically like the sequence \(\frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^i\).
Remark. If \( \alpha_0 = \alpha_1 = \cdots = \alpha_{\nu-1} = 0 \) and \( \alpha_{\nu} \neq 0 \) (i.e., 0 is a root of multiplicity \( \nu \)), then \( x_0, x_1, \ldots, x_{\nu-1} \) are completely independent of \( x_i \) for \( i \geq \nu \). So \( x_{i+k} + \cdots + \alpha_{\nu}x_{i+\nu} = b_i \) for \( i \geq 0 \) with \( x_i \) given for \( i \geq \nu \) behaves like a \((k - \nu)\)-step difference equation.

Remark. Define \( \tilde{x}_i = \begin{bmatrix} x_i \\ x_{i+1} \\ \vdots \\ x_{i+k-1} \end{bmatrix} \).

\[
\tilde{x}_{i+1} = A\tilde{x}_i, \quad i \geq 0,
\]

and \( \tilde{x}_0 = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k-1} \end{bmatrix} \) is given by the I.C. So \((lh)\) is equivalent to the one-step vector difference equation

\[
\tilde{x}_i = A\tilde{x}_{i+1}, \quad i \geq 0,
\]

whose solution is \( \tilde{x}_i = A^i\tilde{x}_0 \). The characteristic polynomial of \((lh)\) is the characteristic polynomial of \( A \). Because \( A \) is a companion matrix, each distinct eigenvalue has only one Jordan block. If \( A = PJP^{-1} \) is the Jordan decomposition of \( A \) (\( J \) in Jordan form, \( P \) invertible), then

\[
\tilde{x}_i = PJ_iP^{-1}\tilde{x}_0.
\]

Let \( J_j \) be the \( m_j \times m_j \) block corresponding to \( r_j \) (for \( 1 \leq j \leq s \)), so \( J_j = r_jI + Z_j \), where \( Z_j \) denotes the \( m_j \times m_j \) shift matrix:

\[
Z_j = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}.
\]

Then

\[
J_j^i = (r_jI + Z_j)^i = \sum_{l=0}^{i} \binom{i}{l} r_j^{i-l}Z_j^l.
\]

Since \( \binom{i}{l} \) is a polynomial in \( i \) of degree \( l \) and \( Z_j^{m_j} = 0 \), we see entries in \( \tilde{x}_i \) of the form \((\text{constant}) \cdot r_j^l Z_j^l \) for \( 0 \leq l \leq m_j - 1 \).

Remark. \((li)\) becomes

\[
\tilde{x}_{i+1} = A\tilde{x}_i + \tilde{b}_i, \quad i \geq 0,
\]
where $\tilde{b}_i = [0, \ldots, 0, b_i]^T$. This leads to a discrete version of Duhamel’s principle (exercise).

**Remark.** All solutions \( \{x_i\}_{i \geq 0} \) of \((lh)\) stay bounded (i.e., are elements of \(l^\infty\))

\[
\iff \text{the matrix } A \text{ is power bounded (i.e., } \exists M \text{ so that } \|A^i\| \leq M \text{ for all } i \geq 0) \]

\[
\iff \text{the Jordan blocks } J_1, \ldots, J_s \text{ are all power bounded}
\]

\[
\iff \begin{cases} 
\text{(a) each } |r_j| \leq 1 \\
\text{and (b) for any } j \text{ with } m_j > 1 \text{ (multiple roots), } |r_j| < 1
\end{cases}
\]

If (a) and (b) are satisfied, then the map $\tilde{x}_0 \mapsto \{x_i\}_{i \geq 0}$ is a bounded linear operator from $\mathbb{R}^k$ (or $\mathbb{C}^k$) into $l^\infty$ (exercise).

### Linear Multistep Methods (LMM)

A LMM is a method of the form

\[
\sum_{j=0}^{k} \alpha_j x_{i+j} = h \sum_{j=0}^{k} \beta_j f_{i+j}, \quad i \geq 0
\]

for the approximate solution of an ODE IVP

\[
x' = f(t, x), \quad x(a) = x_a.
\]

Here we want to approximate the solution $x(t)$ of this IVP for $a \leq t \leq b$ at the points $t_i = a + ih$ (where $h$ is the time step), $0 \leq i \leq \frac{b-a}{h}$. The term $x_i$ denotes the approximation to $x(t_i)$. We have set $f_{i+j} = f(t_{i+j}, x_{i+j})$. We normalize the coefficients so that $\alpha_0 = 1$.

The above is called a $k$-step LMM (if at least one of the coefficients $\alpha_0$ and $\beta_0$ is non-zero).

The above equation is similar to a difference equation in that one is solving for $x_{i+k}$ from $x_i, x_{i+1}, \ldots, x_{i+k-1}$. We assume as usual that $f$ is continuous in $(t, x)$ and uniformly Lipschitz in $x$. For simplicity of notation, we will assume that $x(t)$ is real and scalar; the analysis that follows applies to $x(t) \in \mathbb{R}^n$ or $x(t) \in \mathbb{C}^n$ (viewed as $\mathbb{R}^{2n}$ for differentiability) with minor adjustments.

**Example.** (Midpoint rule) (explicit)

\[
x(t_{i+2}) - x(t_i) = \int_{t_i}^{t_{i+2}} x'(s)ds \approx 2hx'(t_{i+1}) = 2hf(t_{i+1}, x(t_{i+1})).
\]

This approximate relationship suggests the LMM

**Midpoint rule:** $x_{i+2} - x_i = 2hf_{i+1}$.

**Example.** (Trapezoid rule) (implicit)

The approximation

\[
x(t_{i+1}) - x(t_i) = \int_{t_i}^{t_{i+1}} x'(s)ds \approx \frac{h}{2} (x'(t_{i+1}) + x'(t_i))
\]
suggests the LMM

\[
\text{Trapezoid rule: } x_{i+1} - x_i = \frac{h}{2} (f_{i+1} + f_i).
\]

**Explicit vs Implicit.**

If \( \beta_k = 0 \), the LMM is called *explicit*: once we know \( x_i, x_{i+1}, \ldots, x_{i+k-1} \), we compute immediately

\[
x_{i+k} = \sum_{j=0}^{k-1} (h\beta_j f_{i+j} - \alpha_j x_{i+j}).
\]

On the other hand, if \( \beta_k \neq 0 \), the LMM is called *implicit*: knowing \( x_i, x_{i+1}, \ldots, x_{i+k-1} \), we need to solve

\[
x_{i+k} = h\beta_k f(t_{i+k}, x_{i+k}) - \sum_{j=0}^{k-1} (\alpha_j x_{i+j} - h\beta_j f_{i+j})
\]

for \( x_{i+k} \).

**Remark.** If \( h \) is sufficiently small, implicit LMM methods also have unique solutions given \( h \) and \( x_0, x_1, \ldots, x_{k-1} \). To see this, let \( L \) be the Lipschitz constant for \( f \). Given \( x_i, \ldots, x_{i+k-1} \), the value for \( x_{i+k} \) is obtained by solving the equation

\[
x_{i+k} = h\beta_k f(t_{i+k}, x_{i+k}) + g_i,
\]

where

\[
g_i = \sum_{j=0}^{k-1} (h\beta_j f_{i+j} - \alpha_j x_{i+j})
\]

is a constant as far as \( x_{i+k} \) is concerned. That is, we are looking for a fixed point of

\[
\Phi(x) = h\beta_k f(t_{i+k}, x) + g_i.
\]

Note that if \( h|\beta_k|L < 1 \), then \( \Phi \) is a contraction:

\[
|\Phi(x) - \Phi(y)| \leq h|\beta_k| |f(t_{i+k}, x) - f(t_{i+k}, y)| \leq h|\beta_k|L|x - y|.
\]

So by the Contraction Mapping Fixed Point Theorem, \( \Phi \) has a unique fixed point. Any initial guess for \( x_{i+k} \) leads to a sequence converging to the fixed point using functional iteration

\[
x_{i+k}^{(l+1)} = h\beta_k f(t_{i+k}, x_{i+k}^{(l)}) + g_i
\]

which is initiated at some initial point \( x_{i+k}^{(0)} \). In practice, one chooses either

1. iterate to convergence, or

2. a fixed number of iterations, using an *explicit* method to get the initial guess \( x_{i+k}^{(0)} \).

This pairing is often called a Predictor-Corrector Method.
Function Evaluations. One FE means evaluating $f$ once.

Explicit LMM: 1 FE per step (after initial start)
Implicit LMM: ? FEs per step if iterate to convergence
usually 2 FE per step for a Predictor-Corrector Method.

Initial Values. To start a $k$-step LMM, we need $x_0, x_1, \ldots, x_{k-1}$. We usually take $x_0 = x_a$, and approximate $x_1, \ldots, x_{k-1}$ using a one-step method (e.g., a Runge-Kutta method).

Local Truncation Error. For a true solution $x(t)$ to $x' = f(t, x)$, define the LTE to be

$$l(h, t) = \sum_{j=0}^{k} \alpha_j x(t + jh) - h \sum_{j=0}^{k} \beta_j x'(t + jh).$$

If $x \in C^{p+1}$, then

$$x(t + jh) = x(t) + jhx'(t) + \cdots + \frac{(jh)^p}{p!} x^{(p)}(t) + O(h^{p+1}) \quad \text{and}$$

$$hx'(t + jh) = hx'(t) + jh^2 x''(t) + \cdots + \frac{j^{p-1} h^p}{(p-1)!} x^{(p)}(t) + O(h^{p+1})$$

and so

$$l(h, t) = C_0 x(t) + C_1 hx'(t) + \cdots + C_p h^p x^{(p)}(t) + O(h^{p+1}),$$

where

$$C_0 = \alpha_0 + \cdots + \alpha_k$$

$$C_1 = (\alpha_1 + 2\alpha_2 + \cdots + k\alpha_k) - (\beta_0 + \cdots + \beta_k)$$

$$\vdots$$

$$C_q = \frac{1}{q!} (\alpha_1 + 2^q \alpha_2 + \cdots + k^q \alpha_k) - \frac{1}{(q-1)!} (\beta_1 + 2^{q-1} \beta_2 + \cdots + k^{q-1} \beta_k).$$

Definition. A LMM is called accurate of order $p$ if $l(h, t) = O(h^{p+1})$ for any solution of $x' = f(t, x)$ which is $C^{p+1}$.

Fact. A LMM is accurate of order at least $p$ iff $C_0 = C_1 = \cdots = C_p = 0$. (It is called accurate of order exactly $p$ if also $C_{p+1} \neq 0$.)

Remarks.

(i) For the LTE of a method to be $o(h)$ for all $f$’s, we must have $C_0 = C_1 = 0$. To see this, for any $f$ which is $C^1$, all solutions $x(t)$ are $C^2$, so

$$l(h, t) = C_0 x(t) + C_1 hx'(t) + O(h^2) \text{ is } o(h) \quad \text{iff} \quad C_0 = C_1 = 0.$$ 

(ii) Note that $C_0, C_1, \ldots$ depend only on $\alpha_0, \ldots, \alpha_k, \beta_0, \ldots, \beta_k$ and not on $f$. So here, “minimal accuracy” is first order.
**Definition.** A LMM is called *consistent* if \( C_0 = C_1 = 0 \) (i.e., at least first-order accurate).

**Remark.** If a LMM is consistent, then any solution \( x(t) \) for any \( f \) (continuous in \( (t,x) \), Lipschitz in \( x \)) has \( l(h,t) = o(h) \). To see this, note that since \( x \in C^1 \),

\[
x(t + jh) = x(t) + jhx'(t) + o(h) \quad \text{and} \quad hx'(t + jh) = hx'(t) + o(h),
\]

so

\[
l(h,t) = C_0 x(t) + C_1 hx'(t) + o(h).
\]

Exercise: Verify that the \( o(h) \) terms converge to 0 uniformly in \( t \) (after dividing by \( h \)) as \( h \to 0 \): use the uniform continuity of \( x'(t) \) on \([a,b] \).

**Definition.** A \( k \)-step LMM

\[
\sum \alpha_j x_{i+j} = h \sum \beta_j f_{i+j}
\]
is called *convergent* if for each IVP \( x' = f(t,x) \), \( x(a) = x_a \) on \([a,b] \) (\( f \in (C, \text{Lip}) \)) and for any choice of \( x_0(h), \ldots, x_{k-1}(h) \) for which

\[
\lim_{h \to 0} |x(t_i(h)) - x_i(h)| = 0 \quad \text{for} \quad i = 0, \ldots, k - 1,
\]

we have

\[
\lim_{h \to 0} \max_{\{i: a \leq t_i(h) \leq b\}} |x(t_i(h)) - x_i(h)| = 0 .
\]

**Remarks.**

(i) This asks for *uniform* decrease of the error on the grid as \( h \to 0 \).

(ii) By continuity of \( x(t) \), the condition on the initial values is equivalent to \( x_i(h) \to x_a \) for \( i = 0, 1, \ldots, k - 1 \).

**Fact.** If a LMM is convergent, then the zeroes of the (first) characteristic polynomial of the method \( p(r) = \alpha_k r^k + \cdots + \alpha_0 \) satisfy the *Dahlquist root condition*:

(a) all zeroes \( r \) satisfy \( |r| \leq 1 \), and

(b) multiple zeroes \( r \) satisfy \( |r| < 1 \).

**Examples.** Consider the IVP \( x' = 0 \), \( a \leq t \leq b \), \( x(a) = 0 \). So \( f \equiv 0 \). Consider the LMM:

\[
\sum \alpha_j x_{i+j} = 0 .
\]
(1) Let \( r \) be any zero of \( p(r) \). Then the solution with initial conditions

\[
x_i = hr^i \quad \text{for} \quad 0 \leq i \leq k - 1
\]

is

\[
x_i = hr^i \quad \text{for} \quad 0 \leq i \leq \frac{b - a}{h}.
\]

Suppose \( h = \frac{b - a}{m} \) for some \( m \in \mathbb{Z} \). If the LMM is convergent, then

\[
x_m(h) \to x(b) = 0
\]
as \( m \to \infty \). But

\[
x_m(h) = hr^m = \frac{b - a}{m} r^m.
\]

So

\[
|x_m(h) - x(b)| = \frac{b - a}{m} |r^m| \to 0 \quad \text{as} \quad m \to \infty
\]

iff \(|r| \leq 1\).

(2) Similarly if \( r \) is a multiple zero of \( p(r) \), taking \( x_i(h) = hir^i \) for \( 0 \leq i \leq k - 1 \) gives

\[
x_i(h) = hir^i, \quad 0 \leq i \leq \frac{b - a}{h}.
\]

So if \( h = \frac{b - a}{m} \), then

\[
x_m(h) = \frac{b - a}{m} mr^m = (b - a)r^m,
\]

so \( x_m(h) \to 0 \) as \( h \to 0 \) iff \(|r| < 1\).

**Definition.** A LMM is called zero-stable if it satisfies the Dahlquist root condition.

Recall from our discussion of linear difference equations that zero-stability is equivalent to requiring that all solutions of \((lh) \sum_{j=0}^{k} \alpha_j x_{i+j} = 0 \) for \( i \geq 0 \) stay bounded as \( i \to \infty \).

**Remark.** A consistent one-step LMM (i.e., \( k = 1 \)) is always zero-stable. Indeed, consistency implies that \( C_0 = C_1 = 0 \), which in turn implies that \( p(1) = \alpha_0 + \alpha_1 = C_0 = 0 \) and so \( r = 1 \) is the zero of \( p(r) \). Therefore \( \alpha_1 = 1, \alpha_0 = -1 \), so the characteristic polynomial is \( p(r) = r - 1 \), and the LMM is zero-stable.

**Exercise:** Show that if an LMM is convergent, then it is consistent.

**Key Theorem.** [LMM Convergence]

A LMM is convergent if and only if it is zero-stable and consistent. Moreover, for zero-stable methods, we get an error estimate of the form

\[
\max_{a \leq t_i(h) \leq b} |x(t_i(h)) - x_i(h)| \leq K_1 \max_{0 \leq i \leq k-1} |x(t_i(h)) - x_i(h)| + K_2 \max_{i} \frac{|l(h, t_i(h))|}{h} = \text{“growth of error” controlled by zero-stability}
\]
Remark. If $x \in C^{p+1}$ and the LMM is accurate of order $p$, then $|LTE|/h = O(h^p)$. To get $p$th-order convergence (i.e., $\text{LHS} = O(h^p)$), we need

$$x_i(h) = x(t_i(h)) + O(h^p) \quad \text{for} \quad i = 0, \ldots, k - 1.$$ 

This can be done using $k - 1$ steps of a RK method of order $\geq p - 1$.

Lemma. Consider

$$(li) \quad \sum_{j=0}^{k} \alpha_j x_{i+j} = b_i \quad \text{for} \quad i \geq 0 \quad \text{(where} \ \alpha_k = 1),$$

with the initial values $x_0, \ldots, x_{k-1}$ given, and suppose that the characteristic polynomial $p(r) = \sum_{j=0}^{k} \alpha_j r^j$ satisfies the Dahlquist root condition. Then there is an $M > 0$ such that for $i \geq 0$,

$$|x_{i+k}| \leq M \left( \max\{|x_0|, \ldots, |x_{k-1}|\} + \sum_{\nu=0}^{i} |b_\nu| \right).$$

Remark. Recall that the Dahlquist root condition implies that there is an $M > 0$ for which $\|A^i\|_\infty \leq M$ for all $i \geq 0$, where

$$A = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 & 1 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots \\ -\alpha_0 & \cdots & \cdots & -\alpha_k \\ \end{bmatrix}$$

is the companion matrix for $p(r)$, and $\| \cdot \|_\infty$ is the operator norm induced by the vector norm $\| \cdot \|_\infty$. The $M$ in the Lemma can be taken to be the same as this $M$ bounding $\|A^i\|_\infty$.

Proof. Let $\tilde{x}_i = [x_i, \ x_{i+1}, \ldots, \ x_{i+k-1}]^T$ and $\tilde{b}_i = [0, \ldots, 0, \ b_i]^T$. Then $\tilde{x}_{i+1} = A\tilde{x}_i + \tilde{b}_i$, so by induction

$$\tilde{x}_{i+1} = A^{i+1}\tilde{x}_0 + \sum_{\nu=0}^{i} A^{i-\nu}\tilde{b}_\nu.$$ 

Thus

$$|x_{i+k}| \leq \|\tilde{x}_{i+1}\|_\infty$$

$$\leq \|A^{i+1}\|_\infty \|\tilde{x}_0\|_\infty + \sum_{\nu=0}^{i} \|A^{i-\nu}\|_\infty \|\tilde{b}_\nu\|_\infty$$

$$\leq M(\|\tilde{x}_0\|_\infty + \sum_{\nu=0}^{i} |b_\nu|).$$
Proof of the LMM Convergence Theorem. The fact that convergence implies zero-stability and consistency has already been discussed. Suppose a LMM is zero-stable and consistent. Let \(x(t)\) be the true solution of the IVP \(x' = f(t, x), x(a) = x_a\) on \([a, b]\), let \(L\) be the Lipschitz constant for \(f\), and set
\[
\beta = \sum_{j=0}^{k} |\beta_j|.
\]

Hold \(h\) fixed, and set 
\[
e_i(h) = x(t_i(h)) - x_i(h),
\]
\[
l_i(h) = l(h, t_i(h)),
\]
where \(I = \{i \geq 0 : i + k \leq \frac{b-a}{h}\}\).

**Step 1.** The first step is to derive a “difference inequality” for \(|e_i|\). This difference inequality is a discrete form of the integral inequality leading to Gronwall’s inequality. For \(i \in I\), we have
\[
\sum_{j=0}^{k} \alpha_j x(t_{i+j}) = h \sum_{j=0}^{k} \beta_j f(t_{i+j}, x(t_{i+j})) + l_i
\]
\[
\sum_{j=0}^{k} \alpha_j x_{i+j} = h \sum_{j=0}^{k} \beta_j f_{i+j}.
\]
Subtraction gives
\[
\sum_{j=0}^{k} \alpha_j e_{i+j} = b_i,
\]
where 
\[
b_i \equiv h \sum_{j=0}^{k} \beta_j (f(t_{i+j}, x(t_{i+j})) - f(t_{i+j}, x_{i+j})) + l_i.
\]
Then 
\[
|b_i| \leq h \sum_{\nu=0}^{i} |\beta_{\nu}| L |e_{i+\nu}| + |l_{\nu}|.
\]
So, by the preceding Lemma with \(x_{i+k}\) replaced by \(e_{i+k}\), we obtain for \(i \in I\)
\[
|e_{i+k}| \leq M \left[ E + \sum_{\nu=0}^{i} |b_{\nu}| \right]
\]
\[
\leq M \left[ E + hL \sum_{\nu=0}^{i} \sum_{j=0}^{k} |\beta_{\nu}| |e_{i+\nu+j}| + \sum_{\nu=0}^{i} |l_{\nu}| \right]
\]
\[
\leq M \left[ E + hL |\beta_k| |e_{i+k}| + hL \beta \sum_{\nu=0}^{i+k-1} |e_{\nu}| + \sum_{\nu=0}^{i} |l_{\nu}| \right] .
\]
From here on, assume $h$ is small enough that

$$MhL|\beta_k| \leq \frac{1}{2}.$$  

(Since $\{h \leq b - a : MhL|\beta_k| \geq \frac{1}{2}\}$ is a compact subset of $(0, b - a]$, the estimate in the Key Theorem is clearly true for those values of $h$.) Moving $MhL|\beta_k||e_{i+k}|$ to the LHS gives

$$|e_{i+k}| \leq hM_1 \sum_{\nu=0}^{i+k-1} |e_{\nu}| + M_2 E + M_3 \lambda/h$$

for $i \in I$, where $M_1 = 2ML\lambda$, $M_2 = 2M$, and $M_3 = 2M(b-a)$. (Note: For explicit methods, $\beta_k = 0$, so the restriction $MhL|\beta_k| \leq \frac{1}{2}$ is unnecessary, and the factors of 2 in $M_1$, $M_2$, $M_3$ can be dropped.)

**Step 2.** We now use a discrete “comparison” argument to bound $|e_i|$. Let $y_i$ be the solution of

$$(*) \quad y_{i+k} = hM_1 \sum_{\nu=0}^{i+k-1} y_{\nu} + (M_2 E + M_3 \lambda/h) \quad \text{for } i \in I,$$

with initial values $y_j = |e_j|$ for $0 \leq j \leq k - 1$. Then clearly by induction $|e_{i+k}| \leq y_{i+k}$ for $i \in I$. Now

$$y_k \leq hM_1 kE + (M_2 E + M_3 \lambda/h) \leq M_4 E + M_3 \lambda/h,$$

where $M_4 = (b - a)M_1 k + M_2$. Subtracting $(*)$ for $i$ from $(*)$ for $i + 1$ gives

$$y_{i+k+1} - y_{i+k} = hM_1 y_{i+k}, \quad \text{and so } y_{i+k+1} = (1 + hM_1)y_{i+k}.$$

Therefore, by induction we obtain for $i \in I$:

$$y_{i+k} = (1 + hM_1)^i y_k \leq (1 + hM_1)^{(b-a)/h} y_k \leq e^{M_1(b-a)/h} y_k \leq K_1 E + K_2 \lambda/h,$$

where $K_1 = e^{M_1(b-a)} M_4$ and $K_2 = e^{M_1(b-a)} M_3$. Thus, for $i \in I$,

$$|e_{i+k}| \leq K_1 E + K_2 \lambda/h;$$

since $K_1 \geq M_4 \geq M_2 \geq M \geq 1$, also $|e_j| \leq E \leq K_1 E + K_2 \lambda/h$ for $0 \leq j \leq k - 1$. Since consistency implies $\lambda = o(h)$, we are done.

**Remarks.**

1. Note that $K_1$ and $K_2$ depend only on $a, b, L, k$, the $\alpha_j$’s and $\beta_j$’s, and $M$.

2. The estimate can be refined — we did not try to get the best constants $K_1$, $K_2$. For example, $e^{M_1(b-a)}$ could be replaced by $e^{M_1(b_i-a)}$ in both $K_1$ and $K_2$, yielding more precise estimates depending on $i$, similar to the estimate for one-step methods.