H. Ammari
K. Alexopoulos

Numerical Analysis II

ETH Zürich
D-MATH

Problem Sheet 7

Problem 7.1 Error estimate for the trapezium rule method

We consider the trapezium rule method

$$x_{n+1} = x_n + \frac{1}{2}h(f_{n+1} + f_n).$$

for the numerical solution of the initial value problem

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(t, x),$$

where $x_0 = x(0)$ is given, $f_n = f(t_n, x_n)$ and $h = t_{n+1} - t_n$. Let us define the truncation error T_n as

$$T_n := \frac{x(t_{n+1}) - x(t_n)}{h} - \frac{1}{2}\Big(f(t_{n+1}, x(t_{n+1})) + f(t_n, x(t_n))\Big).$$

(7.1a) Show that

$$T_n = -\frac{1}{12}h^2 x'''(\xi_n),$$

for some ξ_n in the interval (t_n, t_{n+1}), where x is the solution of the initial value problem.

HINT: Apply integration by parts to the integral

$$\int_{t_n}^{t_{n+1}} (t - t_{n+1})(t - t_n)x'''(t)dt.$$

Solution: Using integration by parts, we get

$$\begin{aligned}
\int_{t_n}^{t_{n+1}} (t - t_{n+1})(t - t_n)x'''(t)dt &= -\int_{t_n}^{t_{n+1}} (2t - t_n - t_{n+1})x''(t)dt \\
&= -(2t - t_n - t_{n+1})x'(t)\Big|_{t_n}^{t_{n+1}} + \int_{t_n}^{t_{n+1}} 2x'(t)dt \\
&= -hx'(t_{n+1}) - hx'(t_n) + 2(x(t_{n+1}) - x(t_n)) \\
&= -h\Big(f(t_{n+1}, x(t_{n+1})) + f(t_n, x(t_n))\Big) + 2(x(t_{n+1}) - x(t_n)) \\
&= 2hT_n.
\end{aligned}$$

By applying the mean value theorem for integrals, there exists $\xi_n \in (t_n, t_{n+1})$ such that

$$\int_{t_n}^{t_{n+1}} (t - t_{n+1})(t - t_n)x'''(t)dt = x'''(\xi_n)\int_{t_n}^{t_{n+1}} (t - t_{n+1})(t - t_n)dt = -\frac{1}{6}h^3 x'''(\xi_n)$$

and the conclusion follows.

(7.1b) Suppose that f satisfies the Lipschitz condition

$$|f(t,x) - f(t,y)| \leq L|x-y|$$

for all real t, x, y, where L is a positive constant independent of t. Suppose also that there exists some constant M such that $|x'''(t)| \leq M$ for all t. Show that the global error $e_n = x(t_n) - x_n$ satisfies the inequality

$$|e_{n+1}| \leq |e_n| + \frac{1}{2}hL(|e_{n+1}| + |e_n|) + \frac{1}{12}h^3 M.$$

Solution: Let us denote

$$\Phi(t_n, x_n) := \frac{1}{2}(f(t_n, x_n) + f(t_{n+1}, x_{n+1}))$$

and

$$\widetilde{\Phi}(t_n, x) := \frac{1}{2}(f(t_n, x(t_n)) + f(t_{n+1}, x(t_{n+1})))$$

Since

$$x_{n+1} - x_n = h\Phi(t_n, x_n),$$

we have that

$$\begin{aligned}
e_n &= x(t_n) - x_n \\
&= x(t_n) - \left(x_0 + \sum_{k=0}^{n-1}(x_{k+1} - x_k)\right) \\
&= x(t_n) - (x_0 + h\Phi(t_0, x_0) + h\Phi(t_1, x_1) + ... + h\Phi(t_{n-1}, x_{n-1})).
\end{aligned}$$

Thus,

$$\begin{aligned}
e_{n+1} - e_n &= x(t_{n+1}) - x(t_n) - h\Phi(t_n, x_n) \\
&= h\left(\frac{x(t_{n+1}) - x(t_n)}{h} - \widetilde{\Phi}(t_n, x)\right) + h\Big(\widetilde{\Phi}(t_n, x) - \Phi(t_n, x_n)\Big) \\
&= hT_n + h\Big(\widetilde{\Phi}(t_n, x) - \Phi(t_n, x_n)\Big).
\end{aligned}$$

We will deal with these two terms separately. We have, from the Lipschitz condition of f, that

$$\begin{aligned}
&|\widetilde{\Phi}(t_n, x) - \Phi(t_n, x_n)| \\
&\quad = \left|\frac{1}{2}(f(t_n, x(t_n)) + f(t_{n+1}, x(t_{n+1}))) - \frac{1}{2}(f(t_n, x_n) + f(t_{n+1}, x_{n+1}))\right| \\
&\quad \leq \frac{1}{2}\big(|f(t_n, x(t_n)) - f(t_n, x_n)| + |f(t_{n+1}, x(t_{n+1})) - f(t_{n+1}, x_{n+1})|\big) \\
&\quad \leq \frac{L}{2}(|e_n| + |e_{n+1}|).
\end{aligned}$$

From (7.1a) and the assumption that $|x'''(t)| \leq M$, we get

$$|T_n| \leq \frac{1}{12}h^2 M.$$

Therefore, the conclusion follows.

(7.1c) For a uniform step h satisfying $hL < 2$ deduce that, if $x_0 = x(t_0)$, then

$$|e_n| \leq \frac{h^2 M}{12L}\left[\left(\frac{1+\frac{1}{2}hL}{1-\frac{1}{2}hL}\right)^n - 1\right].$$

Solution: Since $1 - \frac{1}{2}hL > 0$, rearranging the result from (7.1b) gives us that

$$|e_{n+1}| \leq \frac{1+\frac{1}{2}hL}{1-\frac{1}{2}hL}|e_n| + \frac{1}{1-\frac{1}{2}hL}\frac{1}{12}h^3 M.$$

Recall, also, that $e_0 = 0$.

In Assignment 5, we saw that, if $a_{n+1} \leq r a_n + b$, then

$$a_n \leq r^n a_0 + b\frac{r^n - 1}{r - 1}.$$

By applying this inequality to e_n, we obtain

$$|e_n| \leq \frac{1}{1-\frac{1}{2}hL}\frac{1}{12}h^3 M\left(\left(\frac{1+\frac{1}{2}hL}{1-\frac{1}{2}hL}\right)^n - 1\right)\left(\frac{1+\frac{1}{2}hL}{1-\frac{1}{2}hL} - 1\right)^{-1}. \tag{7.1.1}$$

Then, since

$$\left(\frac{1+\frac{1}{2}hL}{1-\frac{1}{2}hL} - 1\right)^{-1} = \frac{1-\frac{1}{2}hL}{hL},$$

the conclusion follows.

Problem 7.2 Truncation Error

Consider using a one-step method for the numerical solution of the initial value problem $x' = f(t, x)$, $x(t_0) = x_0$, $f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$. The method is given by

$$x_{n+1} = x_n + \frac{1}{2}h(k_1 + k_2),$$

where

$$k_1 = f(t_n, x_n), \quad \text{and} \quad k_2 = f(t_n + h, x_n + hk_1).$$

Show that the method is consistent and has truncation error

$$T_n = \frac{1}{6}h^2\left(f_x(f_t + f_x f) - \frac{1}{2}(f_{tt} + 2f_{tx}f + f_{xx}f^2)\right) + O(h^3)$$

Solution:

Firstly, since

$$\begin{aligned}\Phi(t, x, \Delta t) :&= \frac{1}{2}(k_1(t, x) + k_2(t, x)) \\ &= \frac{1}{2}(f(t, x) + f(t + \Delta t, x + \Delta t f(t, x)))\end{aligned}$$

it holds that

$$\Phi(t,x,0) = \frac{1}{2}(f(t,x) + f(t,x)) = f(t,x),$$

which means it's consistent.

The truncation error for this one-step method is given by

$$T_k(\Delta t) = \frac{1}{\Delta t}[x(t_{k+1}) - x(t_k)] - \frac{1}{2}(f(t_k, x(t_k)) + f(t_k + h, x(t_k) + hk_1)),$$

where we use the shorthand $k_1 = k_1(t_k, x(t_k))$.

Using a Taylor expansion, we have that

$$\begin{aligned}
f(t_k + h, x(t_k) + hk_1) &= f(t_k, x(t_k)) + hf_t(t_k, x(t_k)) + hk_1 f_x(t_k, x(t_k)) \\
&\quad + \frac{1}{2}h^2 f_{tt}(t_k, x(t_k)) + h^2 k_1 f_{tx}(t_k, x(t_k)) + \frac{1}{2}h^2 k_1^2 f_{xx}(t_k, x(t_k)) + O((\Delta t)^3),
\end{aligned}$$

where f_t and f_x means the partial derivative with respect to t and x of f, respectively.

We will also use the Taylor expansion

$$f(t, x(t)) = f(t_k, x(t_k)) + Df(t_k, x(t_k))(t - t_k) + D^2 f(t_k, x(t_k))\frac{(t-t_k)^2}{2} + O((\Delta t)^3),$$

as well as the fact that $Df(t_k, x(t_k)) := f_t(t_k, x(t_k)) + f(t_k, x(t_k))f_x(t_k, x(t_k))$, and

$$\begin{aligned}
D^2 f(t_k, x(t_k)) &= D(f_t + ff_x) \\
&= f_{tt} + ff_{tx} + f(f_{xt} + ff_{xx}) + (f_t + ff_x)f_x.
\end{aligned}$$

Put everything together (and using the fundamental theorem of calculus to write T_k as an integral), we have

$$\begin{aligned}
T_k(\Delta t) &= \frac{1}{\Delta t}\left(\int_{t_k}^{t_{k+1}} f(t, x(t)) - \frac{1}{2}(f(t_k, x(t_k)) + f(t_k + h, x(t_k) + hk_1))\mathrm{d}t\right) \\
&= \frac{1}{\Delta t}\left(\int_{t_k}^{t_{k+1}} f(t_k, x(t_k)) + Df(t_k, x(t_k))(t - t_k) + D^2 f(t_k, x(t_k))\frac{(t-t_k)^2}{2} + O((\Delta t)^3)\right. \\
&- \frac{1}{2}f(t_k, x(t_k)) - \frac{1}{2}(f(t_k, x(t_k)) + hf_t(t_k, x(t_k)) + hk_1 f_x(t_k, x(t_k)) \\
&\left. + \frac{1}{2}h^2 f_{tt}(t_k, x(t_k)) + h^2 k_1 f_{tx}(t_k, x(t_k)) + \frac{1}{2}h^2 k_1^2 f_{xx}(t_k, x(t_k)) + O((\Delta t)^3))\mathrm{d}t\right) \\
&= \frac{1}{6}(\Delta t)^2(f_{tt} + 2ff_{tx} + f^2 f_{xx} + f_t f_x + ff_x^2) - \frac{1}{4}(\Delta t)^2(f_{tt} + 2ff_{tx} + f^2 f_{xx})|_{(t_k, x(t_k))} + O((\Delta t)^3) \\
&= \left[-\frac{1}{12}(\Delta t)^2(f_{tt} + 2ff_{tx} + f^2 f_{xx}) + \frac{1}{6}(\Delta t)^2(f_t f_x + ff_x^2)\right]|_{(t_k, x(t_k))} + O((\Delta t)^3).
\end{aligned}$$

Problem 7.3 Roundoff Error Effects

In practical situations, computers always round off real numbers. In numerical methods rounding erros become important when the step size Δt is comparable with the precision of the computations. Thus, if taking rounding error into consideration, the Explicit Euler method will become the following perturbed scheme:

$$x^{k+1} = x^k + \Delta t f(t_k, x^k) + (\Delta t)\mu^k + \rho^k,$$

where μ^k and ρ^k represent the errors in f and in the assembling, respectively. Assume that $|\mu^k| \leq \mu$ and $|\rho^k| \leq \rho$ for all $k = 0, 1, 2, \dots$ and $f \in C^1$. Let $e^k := x(t_k) - x^k$, and try to prove that

$$|e^{k+1}| \leq (1 + \Delta t C)|e^k| + \Delta t \mu + \rho + \sup_{\xi \in [t_k, t_{k+1}]} |Df(\xi)| \frac{1}{2}(\Delta t)^2,$$

and hence

$$|e^k| \leq e^{CT}|e^0| + \frac{\mu e^{CT}}{C} + \frac{\rho e^{CT}}{C\Delta t} + \frac{1}{2C} \sup_{\xi \in [0,T]} |Df(\xi)| e^{CT} \Delta t,$$

where C is the Lipschitz constant for f, and Df denotes the differentiation to f where $f(t, x(t))$ is regarded as a function with single parameter t.

Introduce

$$\phi(\Delta t) = \frac{\rho e^{CT}}{C\Delta t} + \frac{1}{2C} \sup_{\xi \in [0,T]} |Df(\xi)| e^{CT} \Delta t,$$

when does ϕ attain its minimum, and therefore what suggestion do you have for the minimal step size Δt? HINT: Use the the arithmetic mean–geometric mean inequality to find a bound for ϕ.

Solution: Since

$$x^{k+1} = x^k + \Delta t f(t_k, x^k) + (\Delta t)\mu^k + \rho^k$$

and

$$x(t_{k+1}) = x(t_k) + \int_{t_k}^{t_{k+1}} f(t, x(t)) \mathrm{d}t,$$

taking the difference between these two equations gives

$$x^{k+1} - x(t_{k+1}) = x^k - x(t_k) + \Delta t(f(t_k, x^k) - f(t_k, x(t_k))) + \Delta t \mu^k + \rho^k - \int_{t_k}^{t_{k+1}} f(t, x(t)) - f(t_k, x(t_k)) \mathrm{d}t.$$

Using the Lipschitz condition on f, we reach

$$|e^{k+1}| \leq (1 + \Delta t C)|e^k| + \Delta t \mu + \rho + \left| \int_{t_k}^{t_{k+1}} f(t, x(t)) - f(t_k, x(t_k)) \mathrm{d}t \right|.$$

Applying the mean value theorem, we find that for each $t \in [t_k, t_{k+1}]$ there exists some $\xi(t) \in [t_k, t]$ such that $f(t, x(t)) - f(t_k, x(t_k)) = Df(\xi(t))(t - t_k)$. Thus, we reach the bound

$$|e^{k+1}| \leq (1 + \Delta t C)|e^k| + \Delta t \mu + \rho + \sup_{\xi \in [t_k, t_{k+1}]} |Df(\xi)| \frac{1}{2}(\Delta t)^2.$$

Applying this bound iteratively, we have that

$$|e^k| \leq (1 + \Delta t C)^k |e^0| + (\Delta t \mu + \rho + \frac{1}{2}(\Delta t)^2 \sup_{\xi \in [0,T]} |Df(\xi)|) \frac{(1 + \Delta t C)^k - 1}{\Delta t C}.$$

Recall that $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Applying this here gives the bound

$$\begin{aligned}
|e^k| &\leq e^{CT}|e^0| + (\mu \Delta t + \rho + \frac{1}{2}(\Delta t)^2 \sup_{\xi \in [0,T]} |Df(\xi)|) \frac{e^{CT}}{C\Delta t} \\
&= e^{CT}|e^0| + \frac{\mu e^{CT}}{C} + \frac{\rho e^{CT}}{C\Delta t} + \frac{1}{2C} \sup_{\xi \in [0,T]} |Df(\xi)| e^{CT} \Delta t
\end{aligned}$$

This is an upper bound for truncation error. We would like to minimise this error bound. Consider only the terms that depend on Δt (this quantity is denoted by ϕ in the question). By the arithmetic mean–geometric mean inequality, we have that

$$\begin{aligned}\phi(\Delta t) = \frac{\rho e^{CT}}{C\Delta t} + \frac{1}{2C} \sup_{\xi\in[0,T]} |Df(\xi)| e^{CT} \Delta t &\geq 2\sqrt{\frac{1}{2C^2}\rho e^{2CT} \sup_{\xi\in[0,T]} |Df(\xi)|} \\ &= \frac{\sqrt{2}e^{CT}}{C}\sqrt{\rho \sup_{\xi\in[0,T]} |Df(\xi)|}\end{aligned}$$

In the above inequality, equality holds if and only if the two quantities are equal. Thus, the minimum of $\phi(\Delta t)$ is attained when $\Delta t = \sqrt{2\rho \sup_{\xi\in[0,T]} |Df(\xi)|^{-1}}$, which is the minimal suggested step size for this algorithm.

Published on 13 April 2022.

To be submitted by 21 April 2022.

Last modified on February 17, 2022