5-3-2016

Principles and Analysis of Approximation Techniques

Evan Smith
Boise State University, evansmith246@u.boisestate.edu

Author:
Evan Smith

Boise State University

Principles and Analysis of Approximation Techniques

Senior Thesis

Math 401

Under the supervision of

Dr. Barbara Zubik-Kowal

Boise State University

May 3, 2016

1 Introduction

Many real life problems can be quantified using smooth functions. An example of a real-life model that has a smooth solution is the following model of a chemical process [1]

$$\begin{cases} \dfrac{dy_1}{dt} = -k_1 y_1(t) + k_2 y_2(t) y_3(t), \\ \dfrac{dy_2}{dt} = k_1 y_1(t) - k_2 y_2(t) y_3(t) - k_3 y_2^2(t), \\ \dfrac{dy_3}{dt} = k_3 y_2^2(t), \end{cases}$$

modeling the concentrations of three species $y_1(t)$, $y_2(t)$, $y_3(t)$. The model parameters k_1, k_2, k_3 are positive and represent reaction rates. This model is an example of a system of nonlinear ordinary differential equations (ODEs). For more information on this model and solution techniques see [1].

An example of a partial differential equation (PDE) that has a smooth solution is the Black-Scholes PDE presented in [5]. The Black-Scholes PDE is a financial risk model for investments based on market volatility and interest rates, and it has the following form

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - q) S \frac{\partial V}{\partial S} - rV = 0,$$

for all $S > 0$ and $0 < t < T$ [5]. Also, note that V depends on both S and t, and the specific solution is $V(S, t)$ [5]. For more information on the Black-Scholes PDE and other equations used for financial applications, see [5].

An example of an integro-differential equation that has a smooth solution is a convolution type logistic growth model presented in [3]. This equation models delay effects resulting from the evolution of past populations to determine the dynamics of the present state of the populations, and has the following form

$$\frac{d}{dt} v(t) = r v(t) \left(1 - \frac{1}{k} \int_{-\infty}^{t} \omega(t - s) v(s) ds \right),$$

where r and k are positive constants, $v(t)$ is the population at time t, and $\omega(t)$ is a weight function that determines the degree of emphasis that should be placed on the size of past populations. For more information about this equation and other equations used in biological applications, see [3].

Although exact solutions to many real life problems cannot, in general, be expressed in closed form, it is possible to compute their approximations on finite discrete subsets of their domains. Such a domain that has been reduced to a corresponding finite subset is said to have undergone discretization, see e.g. the monograph by Cheney [2]. Discretization is a preliminary step to approximate unknown solutions as explained in the next chapter.

2 Discrete sets in approximate solutions

In this chapter, we follow the ideas introduced by Cheney [2] and describe the following general process for approximating unknown solutions.

- **Modeling problems on the continuum**
  Start with a problem that needs a solution. Let us call the problem $\mathbf{P}$ and its solution u. Note that u is often a continuous function. The goal is to find u, or find an approximation of u.

- **Discretization**
  Next, replace the problem's domain $\mathbf{D}$ by a discrete subset of $\mathbf{D}$. Let us call that discrete subset $\mathbf{D}_h$, where h is a number that is preferably close to zero so that the spacing between the values in the discrete set $\mathbf{D}_h$ is small. Once this is done, replace the problem $\mathbf{P}$ by $\mathbf{P}_h$ defined on the discrete domain $\mathbf{D}_h$. In this way, the problem has been converted from a problem that is defined on a continuum to a problem defined on a discrete subset of that continuum. For example, approximations to a solution $u = u(t)$ defined for $t \in \mathbf{D} = [a, b]$, where $a < b$, can be computed on a finite discrete subset $\mathbf{D}_h = \{t_i = a + ih \ : \ i = 0, 1, \ldots, N\}$, where $h = \dfrac{b-a}{N}$. The goal of the problem $\mathbf{P}$ is to determine $u(t)$, for $t \in \mathbf{D}$, while the goal of the problem $\mathbf{P}_h$ is to compute approximations to $u(t_i)$, for $t_i \in \mathbf{D}_h$. The construction of $\mathbf{P}_h$ and $\mathbf{D}_h$ depends on $\mathbf{P}$ and $\mathbf{D}$ and the choice of h.

- **Solving discrete systems**
  Now, solve the problem $\mathbf{P}_h$. The solution to $\mathbf{P}_h$ is a discrete function, say v_h, defined on the discrete domain $\mathbf{D}_h$. We refer to v_h as approximations to the exact solution u. The approximations v_h are functions that depend on the choice of the discrete domain $\mathbf{D}_h$ and are constructed by following the procedure described in $\mathbf{P}_h$.

- **Computing continuous extensions**
  Next, using some form of interpolation, see e.g. Suli and Mayers [6] (for example, cubic spline interpolation), determine a function $\overline{v}_h$ that is defined on the domain $\mathbf{D}$ with values that are equal to the values of v_h on the domain $\mathbf{D}_h$.

- **Error analysis**
  $\overline{v}_h$ is known as an approximate solution to the original problem $\mathbf{P}$. Various error estimating techniques can be used to validate that $\overline{v}_h$ is an approximate solution and to address whether or not $\overline{v}_h$ converges to the solution of problem $\mathbf{P}$ as the value of h approaches zero.

3 Approximate solutions to differential equations

In this chapter, we consider the boundary value problem

$$(1)\qquad \begin{cases} u''(t) + a(t)u'(t) + b(t)u(t) = c(t), & 0 < t < 1, \\ u(0) = 0, \qquad u(1) = 0, \end{cases}$$

and explore the above ideas in more detail.

The following lemmas will be needed to solve the above problem numerically.

Lemma 1. *(Cheney [2]). If an $n \times n$ matrix $A = [a_{ij}]_{i,j=1}^{n}$ is diagonally dominant, that is,*

$$|a_{ii}| - \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| > 0, \quad \text{for all } i = 1, \ldots, n,$$

then it is nonsingular, and

$$\|A^{-1}\|_\infty \le \max_{i=1,\ldots,n} \left\{ |a_{ii}| - \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right\}^{-1},$$

where $\| \cdot \|_\infty$ is the infinity norm.

Proof. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be arbitrary such that $\|x\|_\infty \neq 0$ (that is, x is a nonzero vector). Define $y = Ax$ and choose $i \in \{1, \ldots, n\}$ such that $|x_i| = \|x\|_\infty$. Since the ith component of the vector y is given by

$$y_i = \sum_{j=1}^{n} a_{ij}x_j = a_{ii}x_i + \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}x_j,$$

we get

$$a_{ii}x_i = y_i - \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}x_j.$$

Therefore, from the triangle inequality, we get

$$|a_{ii}x_i| \le |y_i| + |\sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij}x_j|.$$

From this and the choice of i, we get

$$\begin{aligned}|a_{ii}| \cdot \|x\|_\infty = |a_{ii}| \cdot |x_i| &\le |y_i| + \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}x_j| \\ = |y_i| + \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \cdot |x_j| &\le |y_i| + \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \cdot \|x\|_\infty \\ = |y_i| + \|x\|_\infty \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|.\end{aligned}$$

Therefore,

$$\|x\|_\infty \Bigg(|a_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Bigg) \le |y_i|.$$

Since $|y_i| \le \|y\|_\infty$, we get

$$\|x\|_\infty \Bigg(|a_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Bigg) \le \|y\|_\infty.$$

Since $\|x\|_\infty > 0$ and A is diagonally dominant, we conclude that $\|y\|_\infty \neq 0$. Thus, $y \neq 0$, which proves that A is nonsingular. Now, we can solve the original equation for x to get $x = A^{-1}y$, which we can use in the final expression that was derived above to obtain

$$\|A^{-1}y\|_\infty = \|x\|_\infty \le \|y\|_\infty \Bigg(|a_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Bigg)^{-1}$$

and dividing by $\|y\|_\infty$, we get

$$\|A^{-1}\|_\infty \le \Bigg(|a_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Bigg)^{-1}.$$

We now take the maximum over i on both sides of the above inequality and obtain

$$\|A^{-1}\|_\infty \le \max_{i=1,\dots,n} \Bigg\{ |a_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}| \Bigg\}^{-1},$$

which finishes the proof of the lemma. ■

The next lemma determines the errors of finite difference operators applied to first and second order derivatives. We will apply the operators to replace the derivatives in the differential equation in the boundary value problem (1).

Lemma 2. *(Cheney [2]). Suppose* $f \in C^{(4)}([a,b],\mathbb{R})$, $t \in (a,b)$, *and* $h > 0$ *is such that* $t-h,\ t+h \in (a,b)$. *Then there exist* $\xi,\ \eta \in (a,b)$ *such that*

$$f'(t) = \frac{f(t+h)-f(t-h)}{2h} - \frac{h^2}{6}f'''(\eta). \tag{2}$$

$$f''(t) = \frac{f(t+h)-2f(t)+f(t-h)}{h^2} - \frac{h^2}{12}f^{(4)}(\xi). \tag{3}$$

Proof. We apply Taylor's Theorem to $f(t+h)$ and $f(t-h)$ and obtain

$$f(t+h) = f(t) + \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) + \ldots + \frac{h^n}{n!}f^{(n)}(\xi_{1,n}), \tag{4}$$

$$f(t-h) = f(t) - \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) - \ldots \pm \frac{h^n}{n!}f^{(n)}(\xi_{2,n}), \tag{5}$$

where $\xi_{1,n} \in (t, t+h)$ and $\xi_{2,n} \in (t-h, t)$. To show (3), we consider (4) and (5) with $n = 4$ and obtain

$$\begin{aligned} f(t+h) &= f(t) + \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) + \frac{h^3}{3!}f'''(t) + \frac{h^4}{4!}f^{(4)}(\xi_{1,4}), \\ f(t-h) &= f(t) - \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) - \frac{h^3}{3!}f'''(t) + \frac{h^4}{4!}f^{(4)}(\xi_{2,4}). \end{aligned}$$

To obtain an approximation for $f''(t)$, we add the two equations together as follows:

$$\begin{aligned} f(t+h)+f(t-h) &= \left(f(t) + \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) + \frac{h^3}{3!}f'''(t) + \frac{h^4}{4!}f^{(4)}(\xi_{1,4})\right) \\ &\quad + \left(f(t) - \frac{h}{1!}f'(t) + \frac{h^2}{2!}f''(t) - \frac{h^3}{3!}f'''(t) + \frac{h^4}{4!}f^{(4)}(\xi_{2,4})\right) \\ &= 2f(t) + h^2 f''(t) + \frac{h^4}{24}\left(f^{(4)}(\xi_{1,4}) + f^{(4)}(\xi_{2,4})\right). \end{aligned}$$

Since $f \in C^{(4)}([a,b],\mathbb{R})$, there exists $\xi \in (a,b)$ such that $f^{(4)}(\xi) = \frac{1}{2}\left[f^{(4)}(\xi_{1,4}) + f^{(4)}(\xi_{2,4})\right]$. Therefore,

$$h^2 f''(t) = f(t+h) - 2f(t) + f(t-h) - \frac{h^4}{12}f^{(4)}(\xi),$$

which shows (3).

Similarly, to obtain an approximation for $f'(t)$ we apply (4) and (5) with $n = 3$ and subtract the two equations together as follows:

$$\begin{aligned} f(t+h) - f(t-h) &= \left(f(t) + \frac{h}{1!} f'(t) + \frac{h^2}{2!} f''(t) + \frac{h^3}{3!} f'''(\eta_1) \right) \\ &\quad - \left(f(t) - \frac{h}{1!} f'(t) + \frac{h^2}{2!} f''(t) - \frac{h^3}{3!} f'''(\eta_2) \right) \\ &= 2h f'(t) + \frac{h^3}{6} \left(f'''(\eta_1) + f'''(\eta_2) \right), \end{aligned}$$

where $\eta_1 = (t, t+h)$ and $\eta_2 \in (t-h, t)$.

Since $f \in C^{(4)}([a,b], \mathbb{R})$, there exists $\eta \in (a,b)$ such that $f'''(\eta) = \frac{1}{2}\left[f'''(\eta_1) + f'''(\eta_2)\right]$. Therefore,

$$2h f'(t) = f(t+h) - f(t-h) - \frac{h^3}{3} f'''(\eta),$$

which shows (2). ∎

Note that the terms involving ξ and η are the error terms.

We assume that the exact solution u to problem (1) is 4-times continuously differentiable and apply Lemma 1 and Lemma 2 to derive an error bound for a numerical solution to (1). To compute the numerical solution, we discretize the problem using a step size of $h = \frac{1}{n+1} > 0$, where n is a positive integer. For each h, we define the grid points $t_i = ih$, $i = 0, 1, \ldots, n+1$, and approximations $v_i \approx u(t_i)$ such that

$$\begin{cases} \dfrac{v_{i+1} - 2v_i + v_{i-1}}{h^2} + a_i \dfrac{v_{i+1} - v_{i-1}}{2h} + b_i v_i = c_i, \quad i = 1, \ldots, n, \\ v_0 = v_{n+1} = 0, \end{cases} \tag{6}$$

where $a_i = a(t_i)$, $b_i = b(t_i)$, and $c_i = c(t_i)$.

Note that the ith equation in (6) can be written in the following form

$$v_{i-1}\left(\frac{1}{h^2} - \frac{1}{2h} a_i\right) + v_i\left(b_i - \frac{2}{h^2}\right) + v_{i+1}\left(\frac{1}{h^2} + \frac{1}{2h} a_i\right) = c_i. \tag{7}$$

Therefore, (7) can be written in the matrix form

$$A_h v = c_h, \tag{8}$$

where $v = (v_1, \ldots, v_n)^T$, $A_h = \left[A_{ij}^{(h)}\right]_{i,j=1}^{n}$,

$$A_{1j}^{(h)} = \begin{cases} b(t_1) - \dfrac{2}{h^2}, \quad j = 1, \\ \dfrac{1}{h^2} + \dfrac{a(t_1)}{2h}, \quad j = 2, \\ 0, \quad j = 3, \ldots, n, \end{cases} \qquad A_{nj}^{(h)} = \begin{cases} b(t_n) - \dfrac{2}{h^2}, \quad j = n, \\ \dfrac{1}{h^2} - \dfrac{a(t_n)}{2h}, \quad j = n-1, \\ 0, \quad j = 1, \ldots, n-2, \end{cases}$$

$$A_{ij}^{(h)} = \begin{cases} b(t_i) - \dfrac{2}{h^2}, \quad j = i, \\ \dfrac{1}{h^2} - \dfrac{a(t_i)}{2h}, \quad j = i-1, \\ \dfrac{1}{h^2} + \dfrac{a(t_i)}{2h}, \quad j = i+1, \\ 0, \quad j \neq i, i \pm 1, \end{cases}$$

$i = 2, \ldots, n-1$, and $c_h = (c(t_1), \ldots, c(t_n))^T$. The error bound for $|u(t_i) - v_i|$ is presented in the following theorem, found also in [2].

Theorem 1. Suppose a, b, $c \in C([0,1], \mathbb{R})$ and $b(t) < 0$, for all $t \in [0,1]$. Moreover, suppose that $n \in \mathbb{N}$ and $h = \frac{1}{n+1}$ is such that

$$h|a(t)| \leq 2, \tag{9}$$

for all $t \in [0,1]$, and u is the solution to the boundary value problem (1).

Then

$$\max_{i=1,\ldots,n} |u(t_i) - v_i| \leq Bh^2,$$

where $B > 0$ is a constant which does not depend on n or h.

Proof. Writing the differential equation in (1) at $t = t_i$, we get

$$u''(t_i) + a(t_i)u'(t_i) + b(t_i)u(t_i) = c(t_i),$$

for $i = 1, 2, \ldots, n$. From Lemma 2, we get

$$\begin{aligned} u'(t_i) &= \frac{u(t_i + h) - u(t_i - h)}{2h} - \frac{h^2}{6}u'''(\eta_i), \\ u''(t_i) &= \frac{u(t_i + h) - 2u(t_i) + u(t_i - h)}{h^2} - \frac{h^2}{12}u^{(4)}(\xi_i), \end{aligned}$$

where $\eta_i, \xi_i \in (0,1)$. Therefore,

$$\begin{aligned} u'(t_i) &= \frac{u(t_{i+1}) - u(t_{i-1})}{2h} - \frac{h^2}{6}u'''(\eta_i), \\ u''(t_i) &= \frac{u(t_{i+1}) - 2u(t_i) + u(t_{i-1})}{h^2} - \frac{h^2}{12}u^{(4)}(\xi_i), \end{aligned}$$

and

$$\begin{aligned} &\frac{u(t_{i+1}) - 2u(t_i) + u(t_{i-1})}{h^2} + a(t_i)\frac{u(t_{i+1}) - u(t_{i-1})}{2h} + b(t_i)u(t_i) \\ &= c(t_i) + \frac{h^2}{12}u^{(4)}(\xi_i) + a(t_i)\frac{h^2}{6}u'''(\eta_i), \end{aligned}$$

for $i = 1, 2, \ldots, n$.

Let $u_i = u(t_i)$, for $i = 0, 1, 2, \ldots, n, n+1$, and $d_i = \dfrac{h^2}{12} u^{(4)}(\xi_i) + a_i \dfrac{h^2}{6} u'''(\eta_i)$, for $i = 1, 2, \ldots, n$. Then,

$$\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} + a_i \frac{u_{i+1} - u_{i-1}}{2h} + b_i u_i = c_i + d_i,$$

for $i = 1, 2, \ldots, n$.

We now use the notation $e_i = u_i - v_i$. From this and system (6), we get

$$\frac{e_{i+1} - 2e_i + e_{i-1}}{h^2} + a_i \frac{e_{i+1} - e_{i-1}}{2h} + b_i e_i = d_i,$$

for $i = 1, 2, \ldots, n$, and $e_0 = e_{n+1} = 0$. Then,

$$e_{i-1} \left(\frac{1}{h^2} - \frac{1}{2h} a_i \right) + e_i \left(b_i - \frac{2}{h^2} \right) + e_{i+1} \left(\frac{1}{h^2} + \frac{1}{2h} a_i \right) = d_i.$$

Therefore,

$$A_h e = d,$$

where $e = (e_i, \ldots, e_n)^T$ and $d = (d_i, \ldots, d_n)^T$.

We now verify whether A_h satisfies Lemma 1. We want to show that

$$|A_{ii}^{(h)}| - \sum_{\substack{j=1 \\ j \neq i}}^{n} |A_{ij}^{(h)}| > 0, \tag{10}$$

for $i = 1, \ldots, n$. For $i = 1$, we get

$$|A_{11}^{(h)}| - \sum_{j=2}^{n} |A_{1j}^{(h)}| = |A_{11}^{(h)}| - |A_{12}^{(h)}| = \left| b_1 - \frac{2}{h^2} \right| - \left| \frac{1}{h^2} + \frac{a_1}{2h} \right|.$$

Since $b_1 < 0$, $b_1 - \dfrac{2}{h^2} < 0$ and $\left| b_1 - \dfrac{2}{h^2} \right| = -b_1 + \dfrac{2}{h^2}$. From inequality (9), we get $h|a_1| \le 2$ and $\dfrac{|a_1|}{h} \le \dfrac{2}{h^2}$. Therefore, $\pm \dfrac{a_1}{2h} \le \dfrac{|a_1|}{2h} \le \dfrac{1}{h^2}$ and $0 \le \dfrac{1}{h^2} \pm \dfrac{a_1}{2h}$.

From this, we get

$$|A_{11}^{(h)}| - \sum_{j=2}^{n} |A_{1j}^{(h)}| = -b_1 + \frac{2}{h^2} - \frac{1}{h^2} - \frac{a_1}{2h} = -b_1 + \frac{1}{h^2} - \frac{a_1}{2h} \ge -b_1 > 0,$$

and (10) is proved for $i = 1$.

For $i = n$, we get

$$|A_{n\,n}^{(h)}| - \sum_{j=1}^{n-1} |A_{n\,j}^{(h)}| = |A_{n\,n}^{(h)}| - |A_{n\,n-1}^{(h)}| = \left| b_n - \frac{2}{h^2} \right| - \left| \frac{1}{h^2} - \frac{a_n}{2h} \right| = -b_n + \frac{2}{h^2} - \left(\frac{1}{h^2} - \frac{a_n}{2h} \right)$$

$$= -b_n + \frac{1}{h^2} + \frac{a_n}{2h} \ge -b_n > 0,$$

and we proved (10) for $i = n$.

For $i = 2, 3, \ldots, n-1$, we get

$$|A_{ii}^{(h)}| - \sum_{\substack{j=1 \\ j\neq i}}^{n} |A_{ij}^{(h)}| = |A_{ii}^{(h)}| - |A_{i\,i-1}^{(h)}| - |A_{i\,i+1}^{(h)}| = \left| b_i - \frac{2}{h^2} \right| - \left| \frac{1}{h^2} - \frac{a_i}{2h} \right| - \left| \frac{1}{h^2} + \frac{a_i}{2h} \right|$$
$$= -b_i + \frac{2}{h^2} - \left(\frac{1}{h^2} - \frac{a_i}{2h} \right) - \left(\frac{1}{h^2} + \frac{a_i}{2h} \right) = -b_i > 0,$$

and (10) is proved for $i = 2, 3, \ldots, n-1$. Therefore, A_h is diagonally dominant, and by Lemma 1, A_h is invertible and we conclude that

$$e = A_h^{-1} d.$$

From this and Lemma 1, we get

$$\|e\|_\infty \leq \|A_h^{-1}\|_\infty \|d\|_\infty \leq \max_{i=1,\ldots,n} \left\{ |A_{ii}^{(h)}| - \sum_{\substack{j=1 \\ j\neq i}}^{n} |A_{ij}^{(h)}| \right\}^{-1} \|d\|_\infty \leq \max_{i=1,\ldots,n} \left(\frac{1}{-b_i} \right) \|d\|_\infty.$$

Since $-b(t) > 0$, for all $t \in [0,1]$, and b is continuous, there exists a positive number $\delta > 0$ such that $-b(t) > \delta > 0$, for all $t \in [0,1]$. Therefore,

$$\frac{1}{-b(t)} < \frac{1}{\delta}, \qquad \frac{1}{-b_i} < \frac{1}{\delta}, \qquad \max_{i=1,\ldots,n} \left(\frac{1}{-b_i} \right) < \frac{1}{\delta}$$

and

$$\|e\|_\infty \leq \frac{1}{\delta} \|d\|_\infty \leq \frac{c}{\delta} h^2,$$

where c is a positive constant such that

$$c \geq \frac{1}{12} \max_{t\in[0,1]} \left| u^{(4)}(t) \right| + \frac{1}{6} \max_{t\in[0,1]} |a(t)| \max_{t\in[0,1]} |u'''(t)| \,.$$

Therefore,

$$\|e\|_\infty = \max_{i=1,\ldots,n} |u(t_i) - v_i| \leq \frac{c}{\delta} h^2,$$

and the assertion of the theorem is proved with $B = \dfrac{c}{\delta}$.

■

In the next chapter, we present an example and numerical experiments for (1).

4 Numerical experiments

In this chapter, we consider equation (1) with $a(t) = -t$ and $b(t) = -1$. We verify whether the assumptions of Theorem 1 are satisfied, apply the numerical scheme based on A_h, and solve the problem with decreasing step sizes h.

We demonstrate the algorithm defined by (8) with a numerical example. The example given below is a two point boundary value problem posed on $[0, 1]$. Using (8), we compute the approximate solutions v_i, present their errors, and illustrate their convergence to the exact solution u by plotting v_i against t_i, for $i = 0, 1, \dots, n, n+1$, with decreasing $h = 1/(1+n)$.

Example 1. The two point boundary-value problem discussed above is defined as follows:

$$\begin{cases} u''(t) - tu'(t) - u(t) = c(t), & 0 < t < 1, \\ u(0) = 0 \qquad u(1) = 0, \end{cases} \tag{11}$$

where $c(t) = 24(2 + t - t^2)\cos(12t) + 2(145 - 146t)\sin(12t)$.

We begin by applying (7) to transform this problem into a discrete problem. Note that $a(t) = -t$ and $h|a(t)| = ht < h$, for $t \in [0, 1]$. Therefore, condition (9) is satisfied for all step sizes $h \leq 2$, for example, for $h = 1/(1 + n)$ and $n \in \mathbb{N}$. Moreover, $b(t) = -1 < 0$, for $t \in [0, 1]$, and $a, b, c \in C([0, 1], \mathbb{R})$. Therefore, all assumptions of Theorem 1 are satisfied and we conclude that the approximations v_i converge to $u(t_i)$ as $h \to 0$ (the convergence is of order 2).

In order to illustrate the behavior of the approximate solutions v_i for decreasing step sizes h, we use (7), which becomes,

$$v_{i-1}\left(\frac{1}{h^2} + \frac{1}{2h}t_i\right) + v_i\left(-\frac{2}{h^2} - 1\right) + v_{i+1}\left(\frac{1}{h^2} - \frac{1}{2h}t_i\right) = c(t_i),$$

for $i = 1, \dots, n$. Note that this tridiagonal system has to be completed by using the boundary conditions in (11). Then, $v_0 = v_{n+1} = 0$, see Figure 1.

We use MATLAB to solve this tridiagonal system with

$$h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{10}, \frac{1}{20}, \frac{1}{50}, \frac{1}{100}, \frac{1}{10000}$$

(for 2, 4, 8, 10, 20, 50, 100, 10000 steps) in the interval $[0, 1]$. This is accomplished by first building this tridiagonal matrix A_h and solving system (7). Below are the plots of the approximate solutions $v = (v_0, v_1, \dots, v_n, v_{n+1})$ for the step sizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$, and $\frac{1}{10000}$.

Figure 1 shows that all approximate solutions for the four step sizes $h = \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{10000}$ satisfy the boundary conditions. However, the first two approximate solutions are computed with step sizes that are too large $\left(h = \frac{1}{2}\text{ and }\frac{1}{4}\right)$, demonstrating that h needs to be reduced in order to observe the convergence of the approximations. The numerical solution computed with $h = \frac{1}{8}$ is closer to the numerical solution computed with $h = \frac{1}{10000}$ than the ones computed with $h = \frac{1}{2}, \frac{1}{4}$, but it is still seen that the step size h has to be smaller than $\frac{1}{8}$ for good accuracy. The curves obtained from $h = \frac{1}{10}$ and $\frac{1}{20}$ are closer to the black curve obtained from $h = \frac{1}{10000}$, but they are still visibly different than the black curve. The numerical solutions computed with $\frac{1}{50}$ and $\frac{1}{100}$ are both presented by visually the same black curve as the curve obtained by taking $h = \frac{1}{10000}$ demonstrating that a step size of $\frac{1}{50}$ is sufficient.

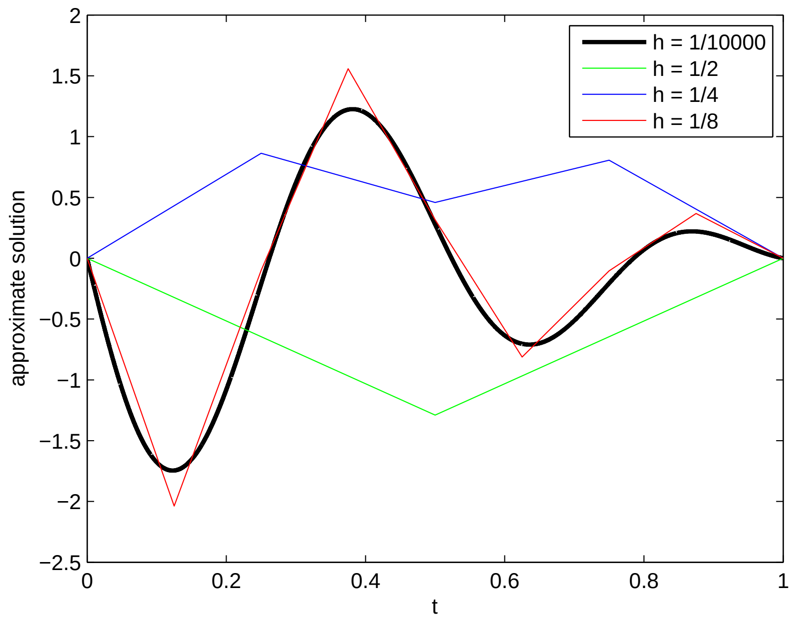

Figure 1: Approximate solutions v_i versus t_i.

Table 1 shows how accurate the approximations are to the true solution at $t = \frac{1}{2}$. The second column lists the approximations v_i to $u(1/2)$ computed with the corresponding step size h from the first column. The index i is determined from $0.5 = t_i = ih$. The third column presents the errors $|v_i - \tilde{v}|$, where $\tilde{v}$ is an approximation to $u(1/2)$ computed by taking $h = \frac{1}{10000}$.

Table 1: Approximations and their errors for decreasing step sizes h.

h	Approximation to $u(1/2)$	Error of the approximation to $u(1/2)$
1/2	-1.2903737	1.5697893
1/4	0.4588711	0.1794557
1/8	0.3170318	0.0376162
1/10	0.3029271	0.0235116
1/20	0.2851098	0.0056942
1/50	0.2803185	0.0009030
1/100	0.2796410	0.0002254

5 Further extensions

Lemma 2 provides approximations for the first and second order derivatives. In this chapter, we prove the following lemma for an approximation to the third order derivative, see Problem 2 in the monograph by Cheney [2], Section 4.1.

Lemma 3. *Suppose $f \in C^{(5)}([a,b],\mathbb{R})$, $x \in (a,b)$, and $h > 0$ is such that $x + ih \in (a,b)$, where $i = \pm 2$. Then,*

$$\left| \frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3} - f'''(x) \right| \leq Ch^2, \tag{12}$$

where C is a positive constant which does not depend on h or x.

Proof. To derive an approximation to $f'''(x)$, we begin by writing out the Taylor expansion for $f(x+2h)$ and $f(x-2h)$

$$\begin{aligned} f(x+2h) &= f(x) + \frac{2h}{1!}f'(x) + \frac{(2h)^2}{2!}f''(x) + \ldots + \frac{(2h)^n}{n!}f^{(n)}(\xi_1), \\ f(x-2h) &= f(x) - \frac{2h}{1!}f'(x) + \frac{(2h)^2}{2!}f''(x) - \ldots \pm \frac{(2h)^n}{n!}f^{(n)}(\xi_2), \end{aligned}$$

where ξ_1, $\xi_2 \in (a,b)$. For $n = 5$, we get

$$\begin{aligned} f(x+2h) &= f(x) + 2hf'(x) + 2h^2f''(x) + \frac{4h^3}{3}f'''(x) + \frac{2h^4}{3}f^{(4)}(x) + \frac{4h^5}{15}f^{(5)}(\zeta_1), \\ f(x-2h) &= f(x) - 2hf'(x) + 2h^2f''(x) - \frac{4h^3}{3}f'''(x) + \frac{2h^4}{3}f^{(4)}(x) - \frac{4h^5}{15}f^{(5)}(\zeta_2), \end{aligned} \tag{13}$$

and, from (4)-(5), we get

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(x) + \frac{h^5}{120}f^{(5)}(\eta_1), \\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f'''(x) + \frac{h^4}{24}f^{(4)}(x) - \frac{h^5}{120}f^{(5)}(\eta_2), \end{aligned} \tag{14}$$

where ζ_1, ζ_2, η_1, $\eta_2 \in (a,b)$. From (13) and (14), we get

$$\begin{aligned} f(x+h) - f(x-h) &= 2hf'(x) + \frac{h^3}{3}f'''(x) + \frac{h^5}{120}\left(f^{(5)}(\eta_1) + f^{(5)}(\eta_2)\right), \\ -2f(x+h) + 2f(x-h) &= -4hf'(x) - \frac{2h^3}{3}f'''(x) - \frac{h^5}{60}\left(f^{(5)}(\eta_1) + f^{(5)}(\eta_2)\right), \end{aligned}$$

and

$$f(x+2h) - f(x-2h) = 4hf'(x) + \frac{8h^3}{3}f'''(x) + \frac{4h^5}{15}\left(f^{(5)}(\zeta_1) + f^{(5)}(\zeta_2)\right).$$

Therefore,

$$f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h) = \frac{6h^3}{3} f'''(x)$$
$$- \frac{h^5}{60}\left(f^{(5)}(\eta_1) + f^{(5)}(\eta_2)\right) + \frac{4h^5}{15}\left(f^{(5)}(\zeta_1) + f^{(5)}(\zeta_2)\right),$$

and

$$\frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3}$$
$$= f'''(x) - \frac{h^2}{60} \cdot \frac{1}{2}\left(f^{(5)}(\eta_1) + f^{(5)}(\eta_2)\right) + \frac{4h^2}{15} \cdot \frac{1}{2}\left(f^{(5)}(\zeta_1) + f^{(5)}(\zeta_2)\right)$$
$$= f'''(x) - \frac{h^2}{60} f^{(5)}(\eta) + \frac{4h^2}{15} f^{(5)}(\zeta),$$

where η, $\zeta \in (a, b)$. Therefore,

$$\left| \frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3} - f'''(x) \right|$$
$$= \frac{h^2}{15} \left| 4f^{(5)}(\zeta) - \frac{1}{4} f^{(5)}(\eta) \right| \le \frac{h^2}{15}\left(4\left|f^{(5)}(\zeta)\right| + \frac{1}{4}\left|f^{(5)}(\eta)\right|\right)$$
$$\le \frac{h^2}{15} \cdot \frac{17}{4} C \le Ch^2,$$

where $C > 0$ is such that

$$\max_{\xi \in [a,b]} |f^{(5)}(\xi)| \le C.$$

Therefore, the proof of Lemma 3 is finished

■

Problem (1) is constrained by boundary conditions $u(0) = 0$ and $u(1) = 0$, and is formulated in terms of an independent variable $t \in [0, 1]$. Below, we discuss how to transform a family of other boundary value problems into (1). The first boundary value problem defines the independent variable t to be in an arbitrary interval $[\alpha, \beta]$.

This boundary value problem is defined as follows:

$$\begin{cases} u''(t) + a(t)u'(t) + b(t)u(t) = c(t), & t \in [\alpha, \beta], \\ u(\alpha) = 0, \quad u(\beta) = 0. \end{cases}$$

By changing the independent variable from t to $s \in [0, 1]$ using the following transformation $t = \beta s + \alpha(1-s)$, this boundary value problem has the following equivalent form on the interval $[0, 1]$:

$$(15) \qquad \begin{cases} \frac{1}{(\beta-\alpha)^2} v''(s) + \frac{a\left(\beta s + \alpha(1-s)\right)}{\beta - \alpha} v'(s) + b\left(\beta s + \alpha(1-s)\right) v(s) \\ = c\left(\beta s + \alpha(1-s)\right), \quad s \in [0, 1], \\ v(0) = 0, \quad v(1) = 0. \end{cases}$$

To show this, we define $v(s) = u(t)$. Then, solving for s in terms of t, we get

$$s = \frac{t - \alpha}{\beta - \alpha}.$$

At the boundaries $t = \alpha$ and $t = \beta$, we get

$$s = \frac{\alpha - \alpha}{\beta - \alpha} = 0, \qquad s = \frac{\beta - \alpha}{\beta - \alpha} = 1$$

and $v(0) = u(\alpha) = 0$ and $v(1) = u(\beta) = 0$. Furthermore, since

$$v(s) = u(t) = u(\beta s + \alpha(1 - s)),$$

by the chain rule, we get

$$\begin{aligned} v'(s) &= \frac{d}{dt} u(\beta s + \alpha(1 - s)) \cdot \frac{d}{ds}[\beta s + \alpha(1 - s)] \\ &= (\beta - \alpha) u'(t). \end{aligned}$$

Using the chain rule again, we get

$$\begin{aligned} v''(s) &= (\beta - \alpha) \frac{d}{dt} u'(\beta s + \alpha(1 - s)) \cdot \frac{d}{ds}[\beta s + \alpha(1 - s)] \\ &= (\beta - \alpha) u''(t) \cdot (\beta - \alpha) \\ &= (\beta - \alpha)^2 u''(t). \end{aligned}$$

Now substituting these results into the initial boundary value problem, we get

$$\begin{aligned} &u''(t) + a(t)u'(t) + b(t)u(t) \\ &= \frac{1}{(\beta - \alpha)^2} v''(s) + \frac{a(t)}{(\beta - \alpha)} v'(s) + b(t)v(s) \\ &= \frac{1}{(\beta - \alpha)^2} v''(s) + \frac{a(\beta s + \alpha(1 - s))}{(\beta - \alpha)} v'(s) + b(\beta s + \alpha(1 - s))v(s) = c(\beta s + \alpha(1 - s)), \end{aligned}$$

which shows (15).

Now, we show how to transform a boundary value problem with nonzero boundary conditions $u(0) = \alpha$ and $u(1) = \beta$.

Let us consider the boundary value problem

$$\begin{cases} u''(t) + a(t)u'(t) + b(t)u(t) = c(t), & t \in [0, 1], \\ u(0) = \alpha, \quad u(1) = \beta, \end{cases} \tag{16}$$

and apply the transformation $v(t) = u(t) - \alpha - (\beta - \alpha)t$. Then, the boundary value problem (16) has the following equivalent form with homogeneous boundary conditions:

$$\begin{cases} v''(t) + a(t)v'(t) + b(t)v(t) = \tilde{c}(t), & t \in [0, 1], \\ v(0) = 0, \quad v(1) = 0, \end{cases}$$

where $\tilde{c}(t) = c(t) + (\alpha - \beta)a(t) + ((\alpha - \beta)t - \alpha)b(t)$.

To show this, note that

$$\begin{aligned} v(0) &= u(0) - \alpha = \alpha - \alpha = 0, \\ v(1) &= u(1) - \alpha + \alpha - \beta = \beta - \beta = 0. \end{aligned}$$

Furthermore,

$$\begin{aligned} v'(t) &= u'(t) - (\beta - \alpha) \\ v''(t) &= u''(t). \end{aligned}$$

Now substituting these results into (16), we get

$$v''(t) + a(t)\left[v'(t) + \beta - \alpha\right] + b(t)\left[v(t) + \alpha + (\beta - \alpha)t\right] = c(t)$$

and

$$v''(t) + a(t)v'(t) + b(t)v(t) = c(t) + (\alpha - \beta)a(t) + ((\alpha - \beta)t - \alpha)b(t).$$

In the next chapter, we consider a family of iterative processes and address the question on whether or not they converge to unique solutions of operator equations.

6 Approximate solutions by iteration

In this chapter, we consider iterative processes written in the following form:

$$x_{n+1} = F(x_n), \tag{17}$$

where $n = 0, 1, 2, \ldots$, $F : [a, b] \to \mathbb{R}$ is such that $F(x) \in [a, b]$, for all $x \in [a, b]$, and $x_0 \in [a, b]$ is arbitrary.

In order to investigate the convergence of the sequence (17) we need to introduce the following definition.

Definition 1. Let the sequence $\{x_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ be given. We say that $\{x_n\}_{n=1}^{\infty}$ converges to x ($\lim_{n\to\infty} x_n = x$) if and only if for any $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n > N$ the inequality $|x_n - x| < \epsilon$ holds.

We define the concept of a Cauchy sequence, below.

Definition 2. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$. Then $\{x_n\}_{n=1}^{\infty}$ is a Cauchy sequence if and only if there exists $N \in \mathbb{N}$ such that $|x_n - x_m| < \epsilon$, for all $m, n \geq N$.

This definition can be found in numerous texts, e.g. [7]. In $\mathbb{R}$, each Cauchy sequence is a convergent sequence, that is, it converges to an element of $\mathbb{R}$.

The following definition of a fixed point can be found in [6].

Definition 3. Suppose $F : [a,b] \to \mathbb{R}$ is continuous and such that $F(x) \in [a,b]$, for all $x \in [a,b]$. Then, $\xi \in [a,b]$ such that $F(\xi) = \xi$ is a fixed point of F.

The following definition can be found in [2].

Definition 4. Suppose $F : [a,b] \to \mathbb{R}$ satisfies the inequality

$$|F(x) - F(y)| \le L|x-y|, \tag{18}$$

for all $x, y \in [a,b]$, where $L \in (0,1)$. Then, F is called a contraction.

The following theorem (see e.g. [6]) provides sufficient conditions for the sequence (17) to converge to a unique fixed point of F.

Theorem 2. Contraction Mapping Theorem. If $F : [a,b] \to [a,b]$ is a contraction, then F has a unique fixed point $\xi \in [a,b]$. The point ξ is the limit of any sequence generated from an arbitrary point $x \in [a,b]$ by iteration (17). That is,

$$\lim_{n\to\infty} F^{(n)}(x) = \xi,$$

where the sequence $\left\{F^{(n)}(x)\right\}_{n=0}^{\infty}$ is defined by

$$F^{(0)}(x) = x, \quad F^{(1)}(x) = F(x), \quad F^{(2)}(x) = F(F(x)), \ldots, \quad F^{(n)}(x) = \overbrace{F(F(\ldots F}^{\text{n-times}}(x))), \ldots.$$

Proof. Let $x = x_0 \in [a,b]$ be arbitrary and $\{x_n\}_{n=0}^{\infty}$ be defined by (17). Then,

$$\begin{aligned}
|x_m - x_{m-1}| &= |F(x_{m-1}) - F(x_{m-2})| \le L|x_{m-1} - x_{m-2}| \\
&= L|F(x_{m-2}) - F(x_{m-3})| \le L^2|x_{m-2} - x_{m-3}| \\
&= L^2|F(x_{m-3}) - F(x_{m-4})| \le L^3|x_{m-3} - x_{m-4}| \le \ldots \\
&\le L^{m-1}|x_1 - x_0|,
\end{aligned}$$

where $m = 1, 2, \ldots$. Moreover, for $k > m$, we get

$$\begin{aligned}
&|x_k - x_m| = |x_k - x_{k-1} + x_{k-1} - x_{k-2} + x_{k-2} - \ldots - x_{m+1} + x_{m+1} - x_m| \\
&\le |x_k - x_{k-1}| + |x_{k-1} - x_{k-2}| + \ldots + |x_{m+1} - x_m| \\
&\le L^{k-1}|x_1 - x_0| + L^{k-2}|x_1 - x_0| + \ldots + L^m|x_1 - x_0| \\
&= |x_1 - x_0| \left(L^{k-1} + L^{k-2} + \ldots + L^m\right) \\
&= |x_1 - x_0| L^m \left(L^{k-1-m} + L^{k-2-m} + \ldots + 1\right) \\
&= |x_1 - x_0| L^m \frac{1 - L^{k-m}}{1 - L} \\
&< |x_1 - x_0| \frac{L^m}{1 - L}.
\end{aligned}$$

Let $\epsilon > 0$. Since $L \in (0,1)$, there exists $N \in \mathbb{N}$ such that

$$|x_1 - x_0| \frac{L^m}{1-L} < \epsilon,$$

for all $m \geq N$. Therefore,

$$|x_k - x_m| < \epsilon,$$

for all $k \geq m \geq N$, and $\{x_n\}_{n=0}^{\infty}$ is a Cauchy sequence in the interval $[a,b]$, thus, because $[a,b]$ is closed, it converges to a certain $\xi \in [a,b]$.

Since F is a contraction, it is continuous and

$$F(\xi) = F(\lim_{m\to\infty} x_m) = \lim_{m\to\infty} F(x_m) = \lim_{m\to\infty} x_{m+1} = \xi.$$

Therefore, ξ is a fixed point of F.

To show that ξ is a unique fixed point of F, we suppose that $\eta \in [a,b]$ is such that $F(\eta) = \eta$. Then,

$$|\xi - \eta| = |F(\xi) - F(\eta)| \leq L|\xi - \eta|$$

and

$$(1-L)\,|\xi - \eta| \leq 0.$$

If $\xi \neq \eta$, then $|\xi - \eta| > 0$, and from the above inequality $(1-L) \leq 0$, that is, $L \geq 1$, which contradicts the fact that $L \in (0,1)$. Therefore, $\xi = \eta$ and ξ is a unique fixed point of F. Since $x_0 \in [a,b]$ was chosen to be arbitrary, iteration (17) is convergent for any starting point. Note that $F^{(n)}(x) = x_n$, for all $n = 0, 1, \ldots$, and the proof is finished.

■

To generalize Theorem 2, we introduce the following definitions (see e.g. [4]) of a metric space and a complete metric space.

Definition 5. Suppose X is a non empty set. Consider a function $d : X \times X \to [0,\infty)$ such that the following three conditions are satisfied:

(i) $$\begin{cases} d(x,y) \geq 0, \quad \text{for all } x,y \in X, \text{ and} \\ d(x,y) = 0 \text{ if and only if } x = y, \end{cases}$$

(ii) $$d(x,y) = d(y,x), \quad \text{for all } x,y \in X,$$

(iii) $$\underbrace{d(x,y) \leq d(x,z) + d(z,y)}_{\text{triangle inequality}}, \quad \text{for all } x,y,z \in X.$$

Then, d is called a **metric** and the pair (X,d) is called a **metric space**.

The following definition can be found in [7].

Definition 6. If every Cauchy sequence in a metric space X convergences to an element of X, then X is a **complete** metric space.

The following theorem is a generalization of Theorem 2 and was proved by Stefen Banach in 1922, see e.g. [2].

Theorem 3. Contraction Mapping Theorem (Cheney [2]). Suppose X is a complete metric space supplemented by a metric d and $F : X \to X$ is a contraction, that is,

$$d(F(x), F(y)) \leq Ld(x, y), \tag{19}$$

for all $x, y \in X$, with $L \in (0, 1)$. Then, for all $x_0 \in X$, the sequences $x_{m+1} = F(x_m)$, where $m = 0, 1, 2, \ldots$, are such that $\lim_{m\to\infty} x_m = \xi$, where $\xi \in X$ is a unique fixed point of F.

The proof of Theorem 3 can be found in [2].

The Contraction Mapping Theorem is useful in proving that iteratively-obtained approximate solutions to systems of equations (e.g. differential equations or integral equations) converge to their exact solutions.

Theorem 3 can be applied to prove the following theorem on the existence of a unique solution of the differential problem written in the form

$$\begin{cases} u'(t) = f(t, u(t)), \quad t \in [a, b], \\ u(a) = u_0, \end{cases} \tag{20}$$

where $a < b$, $u_0 \in \mathbb{R}$, and the function $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ are given.

To prove the uniqueness of the solution of the initial value problem (20), we introduce the following definition (see e.g. [2]).

Definition 7. If f satisfies the following condition

$$|f(t, u_1) - f(t, u_2)| \leq K|u_1 - u_2|, \tag{21}$$

for all $t \in [a, b]$ and u_1, $u_2 \in \mathbb{R}$, where K is a positive constant, then f is called Lipschitz continuous and (21) is called the Lipschitz condition.

A less general version of the following theorem, Theorem 4, can be found in [2] . Here we prove Theorem 4 for a more general problem, problem (20), than the problem stated in [2], which was restricted to the case $a = 0$. Theorem 4 allows a to be an arbitrary real value.

The proof of Theorem 4 is more general than the proof presented in [2] because of two reasons. Firstly, Theorem 4 allows a to be arbitrary. Secondly, the constant $L \in (0, 1)$ for the contraction condition is arbitrary.

Theorem 4. Assume the function $f : [a, b] \times \mathbb{R} \to \mathbb{R}$ satisfies the following conditions

(i) f is continuous,

(ii) f satisfies Lipschitz condition (21).

Then the initial value problem (20) has exactly one solution in $C([a,b],\mathbb{R})$.

Proof. Let $\gamma > 1$ and $X = C([a,b],\mathbb{R})$ with

$$\|u\| = \max\big\{|u(t)|e^{-\gamma K(t-a)} : t \in [a,b]\big\}, \tag{22}$$

for all $u \in X$. Then X is a Banach space (complete normed space), see e.g. [2]. Note that the initial value problem (20) is equivalent to

$$\begin{aligned}
\int_a^t u'(s)ds &= \int_a^t f(s,u(s))ds,\\
u(t) - u(a) &= \int_a^t f(s,u(s))ds,\\
u(t) &= u(a) + \int_a^t f(s,u(s))ds,\\
u(t) &= u_0 + \int_a^t f(s,u(s))ds,
\end{aligned}$$

and we can define the following operator $F : X \to X$ such that

$$(F(u))(t) = u_0 + \int_a^t f(s,u(s))ds, \tag{23}$$

where $u \in C([a,b],\mathbb{R})$ and $t \in [a,b]$. We now apply Theorem 3 with F defined by (23) and we verify that (19) is satisfied with

$$d(u_1,u_2) = \|u_1 - u_2\|, \tag{24}$$

where $u_1,\ u_2 \in X$.

From (24) and (22), we obtain

$$\begin{aligned}
&d(F(u_1),F(u_2)) = \|F(u_1) - F(u_2)\|\\
&= \max\Big\{\Big|\Big(F(u_1) - F(u_2)\Big)(t)\Big| \exp\big(-\gamma K(t-a)\big) : t \in [a,b]\Big\}\\
&= \max\big\{|F(u_1)(t) - F(u_2)(t)| \exp\big(-\gamma K(t-a)\big) : t \in [a,b]\big\}.
\end{aligned}$$

Then, for arbitrary $t \in [a,b]$, we get

$$\begin{aligned}
|F(u_1)(t) - F(u_2)(t)| &= \left|u_0 + \int_a^t f(s,u_1(s))ds - u_0 - \int_a^t f(s,u_2(s))ds\right|\\
&= \left|\int_a^t f(s,u_1(s)) - f(s,u_2(s))ds\right|.
\end{aligned}$$

From this, the triangle inequality, and Lipschitz condition (21), we get

$$|F(u_1)(t) - F(u_2)(t)| \le \int_a^t \big|f(s, u_1(s)) - f(s, u_2(s))\big| ds$$

$$\le \int_a^t K\big|u_1(s) - u_2(s)\big| ds.$$

We now want to apply the definition of the norm $\|\cdot\|$, and continue with

$$|F(u_1)(t) - F(u_2)(t)| \le \int_a^t K\big|u_1(s) - u_2(s)\big| \exp\big(-\gamma K(s-a)\big) \exp\big(\gamma K(s-a)\big) ds$$

$$= \int_a^t K\big|(u_1 - u_2)(s)\big| \exp\big(-\gamma K(s-a)\big) \exp\big(\gamma K(s-a)\big) ds$$

$$\le \int_a^t K\|u_1 - u_2\| \exp\big(\gamma K(s-a)\big) ds.$$

Since $\|u_1 - u_2\|$ does not depend on the variable s of integration, we get

$$\begin{aligned} |F(u_1)(t) - F(u_2)(t)| &\le K\|u_1 - u_2\| \int_a^t \exp\big(\gamma K(s-a)\big) ds \\ &= K\|u_1 - u_2\| \frac{1}{\gamma K}\Big(\exp\big(\gamma K(t-a)\big) - 1\Big) \\ &< \|u_1 - u_2\| \frac{1}{\gamma} \exp\big(\gamma K(t-a)\big). \end{aligned}$$

Therefore,

$$\big|\big(F(u_1) - F(u_2)\big)(t)\big| \le \frac{1}{\gamma}\|u_1 - u_2\| \exp\big(\gamma K(t-a)\big)$$

and

$$\big|\big(F(u_1) - F(u_2)\big)(t)\big| \exp\big(-\gamma K(t-a)\big) \le \frac{1}{\gamma}\|u_1 - u_2\|.$$

We now take the maximum over the interval $[a, b]$ on both sides of the above inequality, and get

$$\max\Big\{\Big|\Big(F(u_1) - F(u_2)\Big)(t)\Big| \exp\big(-\gamma K(t-a)\big) : t \in [a, b]\Big\} \le \frac{1}{\gamma}\|u_1 - u_2\|$$

(note that the right hand side of the above inequality does not depend on t).

Therefore, from the definition of the norm $\|\cdot\|$, we get

$$\|F(u_1) - F(u_2)\| \le \frac{1}{\gamma}\|u_1 - u_2\|,$$

and from (24), we get

$$d(F(u_1), F(u_2)) \leq \frac{1}{\gamma} d(u_1, u_2).$$

Since $u_1, u_2 \in X$ were chosen in an arbitrary way and $0 < \dfrac{1}{\gamma} < 1$, we can conclude that F is a contraction on X with $L = \dfrac{1}{\gamma}$.

Therefore, all assumptions of Theorem 3 are satisfied, and we conclude that (23) has exactly one fixed point in X, which implies that (20) has exactly one solution in X.

■

The example below is an initial value problem that we consider to present application of Theorem 4.

Example 2. Suppose $f(t,x) = \sin(t^2 x) - \sin(t^3) + 1$, $a = 1$, $b = 2$, and $u_0 = 1$. Then, the differential problem (20) is written in the form

$$\begin{cases} u'(t) = \sin\big(t^2 u(t)\big) - \sin(t^3) + 1, & t \in [1,2], \\ u(1) = 1. \end{cases} \tag{25}$$

We now verify whether the assumptions of Theorem 4 are satisfied. Since $f : [1,2] \times \mathbb{R} \to \mathbb{R}$ is continuous, assumption (i) of Theorem 4 is satisfied. Let $t \in [1,2]$, x_1, $x_2 \in \mathbb{R}$ be arbitrary. Then, by the Mean Value Theorem, we get

$$\begin{aligned} f(t,x_1) - f(t,x_2) &= \sin(t^2 x_1) - \sin(t^3) + 1 - \sin(t^2 x_2) + \sin(t^3) - 1 \\ &= \sin(t^2 x_1) - \sin(t^2 x_2) = \cos(\xi)(t^2 x_1 - t^2 x_2), \end{aligned}$$

where ξ is between $t^2 x_1$ and $t^2 x_2$. Since $|\cos(\xi)| \leq 1$, we get

$$|f(t,x_1) - f(t,x_2)| = |\cos(\xi)| \cdot |t^2 x_1 - t^2 x_2| \leq |t^2 x_1 - t^2 x_2| = t^2 |x_1 - x_2| \leq 4|x_1 - x_2|,$$

which shows that (21) is satisfied with $K = 4$, and f is Lipschitz continuous with respect to the second argument. Therefore, by Theorem 4, the initial value problem (25) has exactly one solution.

Example 3. Let $f(t,x) = t\cos^2(tx) - t$, $a = -\frac{1}{2}$, $b = \frac{1}{2}$, and $u_0 = 0$. Then, the initial value problem (25) is written in the form

$$\begin{cases} u'(t) = t\cos^2\big(tu(t)\big) - t, & t \in \left[-\dfrac{1}{2}, \dfrac{1}{2}\right], \\ u\left(-\dfrac{1}{2}\right) = 0. \end{cases} \tag{26}$$

To see whether the assumptions of Theorem 4 are satisfied, we take arbitrary x_1, $x_2 \in \mathbb{R}$, and $t \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ and obtain

$$\begin{aligned} f(t,x_1) - f(t,x_2) &= t\cos^2(tx_1) - t - t\cos^2(tx_2) + t \\ &= t\cos^2(tx_1) - t\cos^2(tx_2) = t\big(\cos(tx_1) - \cos(tx_2))(\cos(tx_1) + \cos(tx_2)\big) \\ &= -t\sin(\xi)(tx_1 - tx_2)(\cos(tx_1) + \cos(tx_2)), \end{aligned}$$

where ξ is between tx_1 and tx_2. Therefore,

$$\begin{aligned} |f(t,x_1) - f(t,x_2)| &= t^2|\sin(\xi)| \cdot |x_1 - x_2| \cdot |\cos(tx_1) + \cos(tx_2)| \\ &\le t^2|x_1 - x_2| \cdot \big(|\cos(tx_1)| + |\cos(tx_2)|\big) \\ &\le 2t^2|x_1 - x_2| \le 2\frac{1}{4}|x_1 - x_2| = \frac{1}{2}|x_1 - x_2|, \end{aligned}$$

which shows that $f : \left[-\frac{1}{2}, \frac{1}{2}\right] \times \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous with respect to the second argument with the Lipschitz constant $K = \frac{1}{2}$ and condition (ii) of Theorem 4 is satisfied. Moreover, f is continuous and, since all conditions of Theorem 4 are satisfied, the initial value problem (26) has exactly one solution.

The rate of convergence of the iterative process (17) depends on F (so, on f). For computational techniques of solving problems of the form (20), we refer the reader to [1].

References

[1] Butcher, J. *Numerical methods for ordinary differential equations*, 2nd ed. John Wiley & Sons, Chichester, 2003.

[2] Cheney, W. *Analysis for Applied Mathematics.* Springer, New York, New York, 2001.

[3] Murray, J. *Mathematical Biology I An Introduction*, 3rd ed. Springer, New York, New York, 2002.

[4] Shirali, S., and Vasudeva, H. *Metric Spaces.* Springer, New York, New York, 2006.

[5] Stefanica, D. *A Primer for the Mathematics of Financial Engineering.* FE Press, New York, New York, 2008.

[6] Suli, E., and Mayers, D. *An Introduction to Numerical Analysis.* Cambridge University Press, London, United Kingdom, 2008.

[7] Sutherland, W. *Introduction to Metric and Topological Spaces.* Oxford University Press, London, United Kingdom, 1975.