Numerical Analysis
(MATH 5610)
Homework 2

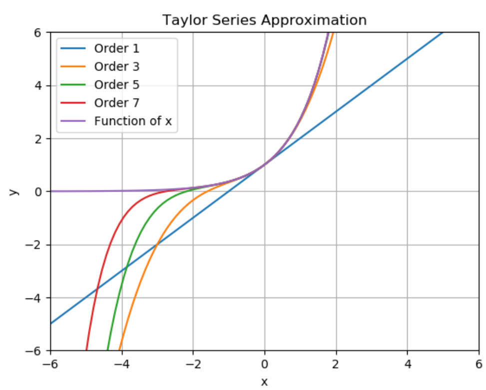
Jordan Saethre

September 26, 2019

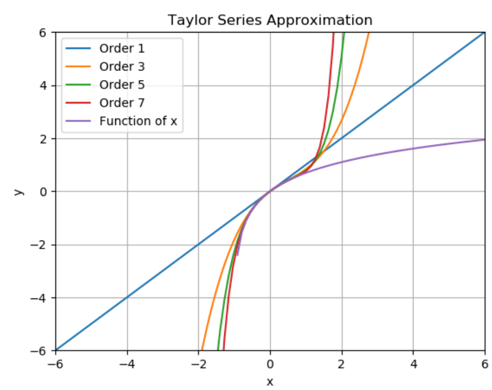
1 Taylor Series

Let $f(x) = e^x$ and $g(x) = \ln(x+1)$, and let p_n andq_n be the Taylor polynomials of degree n for f and g, respectively about $x_0 = 0$. Plot the graphs of f, g, p_n, and q_n, for some small values of n, and comment on your results. Discuss in particular how well f and g are approximated by their Taylor polynomials. Explain your observations in terms of a suitable expression for the error in the approximation.

1.1 Plots of Approximations



Graph of $f(x) = e^x$ and its approximations for some small values of n



Graph of $g(x) = \ln(x+1)$ and its approximations for some small values of n

The approximations for both $f(x)$ and $g(x)$ are best in small neighborhoods around $x = 0$ and get better as n increases. The error in the approximation is given by

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$$

Where c in this expression is between x and a. For our examples we approximated about $x = 0$. The approximation error in the interval $[-\epsilon, \epsilon]$ is bounded, that is

$$\frac{f^{(n+1)}(-\epsilon)}{(n+1)!} \leq error_f \leq \frac{f^{(n+1)}(\epsilon)}{(n+1)!}$$

Similarly for $g(x)$:

$$\frac{g^{(n+1)}(-\epsilon)}{(n+1)!} \leq error_g \leq \frac{g^{(n+1)}(\epsilon)}{(n+1)!}$$

1.2 Code for Approximation Plots

```
import sympy as sy
import numpy as np
import math
from sympy.functions import ln, exp
import matplotlib.pyplot as plt

# Taylor approximation about x_0
def taylor_poly(function, x_0, n, x = sy.Symbol('x')):
    i = 0
    p = 0
    while i <= n:
        p = p + (function.diff(x, i).subs(x, x_0))/
                (math.factorial(i))*(x - x_0)**i
        i = i + 1
    return p

def plot(f, x_0 = 0, n = 7, x_lims = [-6, 6], y_lims = [-6, 6], n_pt = 100,
                                                        x = sy.Symbol('x')):
    x_1 = np.linspace(x_lims[0], x_lims[1], n_pt)

    # Approximate order 1 to n
    for j in range(1, n + 1, 2):
        fn = taylor_poly(f, x_0, j)
        t_lambda = sy.lambdify(x, fn, "numpy")
        print('Taylor expansion at n=' + str(j), fn)
        plt.plot(x_1, t_lambda(x_1), label = 'Order '+ str(j))

    # Plot the function and approximations
    f_lambda = sy.lambdify(x, f, "numpy")
    plt.plot(x_1, f_lambda(x_1), label = 'Function of x')
    plt.xlim(x_lims)
    plt.ylim(y_lims)
    plt.xlabel('x')
    plt.ylabel('y')
    plt.legend()
    plt.grid(True)
    plt.title('Taylor Series Approximation')
    plt.show()

# Define the variable and the function to approximate
x = sy.Symbol('x')
f = ln(1 + x)
plot(f)

x = sy.Symbol('x')
g = exp(x)
plot(g)
```

2 Positive Definite Matrices

Show that every principal submatrix of a positive definite matrix is positive definite.

Proof. A matrix is positive definite if it is an $n \times n$ symmetric and $(\vec{x})^T A\vec{x} > 0, \quad \forall\ \vec{x} \neq \vec{0}$. Let $I \subset \{1, 2, ..., n\}$. A principal submatrix is defined by striking out all rows and columns that are not in the chosen set I. Let the principal submatrix for I be A_I. Choose a vector $\vec{x}$ that has 0 for every element whose index is not in I, by definition $(\vec{x})^T A\vec{x}$ must also be positive since it is for any $\vec{x}$. The result of this computation equates to $(\vec{x})^T A_I\vec{x}$ which is therefore also positive definite. $\square$

3 The UL factorization

Show how to compute the factorization $A = UL$ where U is upper triangular with 1s along the diagonal and L is lower triangular. Show how this relates to a way of solving $Ax = b$ by transforming the system into an equivalent system with a lower triangular matrix. (In other words, show that what we did for the LU factorization also works for a UL factorization.) Note: For the purposes of this exercise you may assume that no pivoting is required. This is of course unrealistic but pivoting would only distract from the point of this exercise (which is that conceptually there is no difference between an LU and a UL factorization).

Write A as $A = UL$ where U is the upper triangular matrix, that is $l_{ii} = 1$ $i = 1, 2, ..., n$ and $j < i$ implies $a_{ij} = 0$. L is the lower triangular matrix, that is $l_{ii} = 1$ $i = 1, 2, ..., n$ and $j > i$ implies $a_{ij} = 0$.

We can solve $Ax = b$ by factoring A into UL and solve the equivalent system $Ax = ULx = b$ in two steps:

1. Solve $Uz = b$ (Backward substitution)
2. Solve $Lx = z$ (Forward substitution)

Consider a 4×4 system. The letter x denotes and entry in the working matrix. The symbol $\otimes$ denotes an entry that is final and will not be changed further. The letter m denotes the multipliers, stored in the upper part of the working matrix.

We have the following:

$$\begin{bmatrix} \otimes & x & x & x \\ \otimes & x & x & x \\ \otimes & x & x & x \\ \otimes & x & x & x \end{bmatrix} \longrightarrow \begin{bmatrix} \otimes & m_{12} & m_{13} & m_{14} \\ \otimes & \otimes & x & x \\ \otimes & \otimes & x & x \\ \otimes & \otimes & x & x \end{bmatrix} \longrightarrow \begin{bmatrix} \otimes & m_{12} & m_{13} & m_{14} \\ \otimes & \otimes & m_{23} & m_{24} \\ \otimes & \otimes & \otimes & x \\ \otimes & \otimes & \otimes & x \end{bmatrix} \longrightarrow \begin{bmatrix} \otimes & m_{12} & m_{13} & m_{14} \\ \otimes & \otimes & m_{23} & m_{24} \\ \otimes & \otimes & \otimes & m_{34} \\ \otimes & \otimes & \otimes & \otimes \end{bmatrix}$$

Denote row i of a matrix A by $r_i(A)$ and consider in particular the second row of our working matrix. We have:

$$r_2(L) = r_2(A) - m_{23}r_3(L) - m_{24}r_4(L)$$

Rewrite this as:

$$r_2(A) = r_2(L) + m_{23}r_3(L) + m_{24}r_4(L)$$

Which is exactly what you get in the matrix multiplication $A = UL$

$$\begin{bmatrix} r_1(A) \\ r_2(A) \\ r_3(A) \\ r_4(A) \end{bmatrix} = \begin{bmatrix} 1 & m_{12} & m_{13} & m_{14} \\ 0 & 1 & m_{23} & m_{24} \\ 0 & 0 & 1 & m_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1(L) \\ r_2(L) \\ r_3(L) \\ r_4(L) \end{bmatrix}$$

This applies to all the rows of this 4×4 system.

4 Least Squares Approximation of Functions

Find a linear function such that it minimizes $\int_0^1 (e^x - l(x))^2 dx$.

Solution: Let $l(x) = ax + b$. To minimize we take partial derivatives with respect to both a and b. We set this equal to zero and solve the system of two equations with two unknowns a and b.

$$\int_0^1 (e^x - (ax+b))^2 dx \implies \int_0^1 (e^x - ax - b)^2 dx$$

First, differentiate with respect to a:

$$\frac{\partial}{\partial a}\left(\int_0^1 (e^x - ax - b)^2 dx\right) \implies \int_0^1 \frac{\partial}{\partial a}(e^x - ax - b)^2 dx \implies \int_0^1 2(e^x - ax - b)(-x) dx$$

Set this equal to zero and integrate:

$$\int_0^1 -2xe^x + 2ax^2 + 2bx\ dx = 0 \implies \int_0^1 ax^2 + bx\ dx = \int_0^1 xe^x\ dx \implies \frac{1}{3}ax^3 + \frac{1}{2}bx^2|_0^1 = 1$$

$$\frac{1}{3}a + \frac{1}{2}b = 1 \tag{1}$$

Second, differentiate with respect to b:

$$\frac{\partial}{\partial b}\left(\int_0^1 (e^x - ax - b)^2 dx\right) \implies \int_0^1 \frac{\partial}{\partial b}(e^x - ax - b)^2 \implies \int_0^1 2(e^x - ax - b)(-1)dx$$

Set this equal to zero and integrate:

$$\int_0^1 -2e^x + 2ax + 2b\ dx = 0 \implies \int_0^1 ax + b\ dx = \int_0^1 e^x\ dx \implies \frac{1}{2}ax^2 + bx|_0^1 = e$$

$$\frac{1}{2}a + b = e \tag{2}$$

Now take (1) and (2) and solve the system of equations to get

$$a = 3 - e, \quad b = \frac{3}{2}(e-1)$$

5 An Alternative Approximation Problem

Find a linear function such that it minimizes $\int_0^1 |e^x - l(x)|\ dx$.

Solution: Let $l(x) = ax + b$. To minimize the distance between $ax + b$ and e^x we must first break up the integral into three intervals: $[0, u]$, $[u, v]$, and $[v, 1]$. We note that between u and v our linear equation $ax + b$ will be above e^x. We will then solve the three separate integrals in terms of u, v, a, and b at which point we can take partial derivatives with respect to a and b to result in a system of two equations with two unknowns. Solving this system for u and v we will find the two points that our linear equation $ax + b$ must pass through in order to minimize the distance from it and e^x.

$$\int_0^1 |(e^x - (ax+b)|\ dx \implies \int_0^u e^x - ax - b\ dx + \int_u^v ax + b - e^x\ dx + \int_v^1 e^x - ax - b\ dx$$

$$\implies e^x - \frac{a}{2}x^2 - bx\ |_0^u + \frac{a}{2}x^2 + bx - e^x\ |_u^v + e^x - \frac{a}{2}x^2 - bx\ |_v^1$$

$$\implies 2e^u - au^2 - 2bu - 1 + av^2 + 2bv - 2e^v + e - \frac{a}{2} - b$$

$$\frac{\partial}{\partial a}\left(2e^u - au^2 - 2bu - 1 + av^2 + 2bv - 2e^v + e - \frac{a}{2} - b\right)$$

$$\frac{\partial}{\partial b}\left(2e^u - au^2 - 2bu - 1 + av^2 + 2bv - 2e^v + e - \frac{a}{2} - b\right)$$

$$v^2 - \frac{1}{2} = u^2 \tag{3}$$

$$2v - 2u = 1 \tag{4}$$

Solving the system of equation composed of (3) and (4) you find that $u = \frac{1}{4}$ and $v = \frac{3}{4}$. This provides us with the intersection points of e^x and our linear equation $ax + b$, that is $(\frac{1}{4}, e^{1/4})$ and $(\frac{3}{4}, e^{3/4})$. The line that passes through these two points is

$$y = (2e^{3/4} - 2e^{1/4})x + \left(\frac{-e^{3/4} - 3e^{1/4}}{2}\right)$$

6 Another Alternative Approximation Problem

Find a linear function $l(x)$ such that $\max_{0\leq x\leq 1}|e^x - l(x)|$ is minimized.

Solution: Break up problem into three intervals like we did in question 5 above: $[0, u]$, $[u, v]$, and $[v, 1]$. For the first interval the distance between e^x and $ax + b$ will be maximized when $x = 0$. Similarly for the last interval the distance between e^x and $ax + b$ will be maximized when $x = 1$. This gives us two points to construct the line that minimizes $\max_{0\leq x\leq 1}|e^x - l(x)|$. They are $(0, 1)$ and $(1, e)$ which give us the line $y = (e-1)x + 1$.

7 More on Newton's Method

Construct an example where Newton's method cycles, i.e., $x_{n+2} = x_n$ and $x_n \neq x_{n+1}$ for all n.

I want to show the following:

$$x_0 = x_0, \quad x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \quad x_2 = \left(x_0 - \frac{f(x_0)}{f'(x_0)}\right) - \frac{f\left(x_0 - \frac{f(x_0)}{f'(x_0)}\right)}{f'\left(x_0 - \frac{f(x_0)}{f'(x_0)}\right)} = x_0$$

Let $x_0 = 0$ and try $f(x) = x^3 - 2x + 2$. We have $f'(x) = 3x^2 - 2$. $f(0) = 2$ and $f'(0) = -2$.

$$x_1 = 0 - \frac{(0)^3 - 2(0) - 2}{3(0)^2 - 2} = 0 - \frac{2}{-2} = 1$$

$$x_2 = 1 - \frac{(1)^3 - 2(1) - 2}{3(1)^2 - 2} = 1 - \frac{1}{1} = 0 = x_0$$

8 Division without Division

Suppose you have a computer or calculator that has no built-in division. Come up with a fixed point iteration that converges to $\frac{1}{r}$ for any given non-zero number r, and that only uses addition, subtraction, and multiplication. Hint: Write down an equation satisfied by $\frac{1}{r}$, apply Newton's method to that equation, and then modify Newton's method so that it doesn't use division. Your resulting method should converge of order 2.

Let $x = \frac{1}{r}$ and $f(x) = rx - 1$. If we apply Newton's Method to this equation we have:

$$g(x) = x - \frac{f(x)}{f'(x)} = x - \frac{rx - 1}{r} = x - f(x) \cdot x$$

This iteration will converge of order 2.

9 A Common Matrix

The following $n \times n$ matrix occurs frequently in the solution of second order ordinary differential equations.

$$A = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \cdots & \cdots & \cdots & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

a) Show that $A^{-1} = [x_{ij}]$ where

$$x_{ij} = \begin{cases} \frac{(n+1-j)i}{n+1} & i \leq j \\ \\ \frac{(n+1-i)j}{n+1} & i > j \end{cases}$$

$$AA^{-1} = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \cdots & \cdots & \cdots & -1 \\ & & & & -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} \frac{n}{n+1} & \frac{n-1}{n+1} & \cdots & \cdots & \cdots & \frac{1}{n+1} \\ \frac{n-1}{n+1} & \frac{2(n-1)}{n+1} & \frac{2(n-2)}{n+1} & \cdots & \cdots & \frac{2}{n+1} \\ \cdots & \frac{2(n-2)}{n+1} & \cdots & \cdots & \cdots & \cdots \\ \frac{2}{n+1} & \cdots & \cdots & \cdots & \cdots & \frac{n-1}{n+1} \\ \frac{1}{n+1} & \cdots & \cdots & \cdots & \frac{n-1}{n+1} & \frac{n}{n+1} \end{bmatrix} = I$$

b) Show that $A = LU$ where $L = [l_{ij}]$ is a unit lower triangular matrix and $U = [u_{ij}]$ is an upper triangular matrix and

$$l_{ii} = 1, \quad l_{i,i-1} = -\frac{i-1}{i}, \quad u_{ii} = \frac{i+1}{i}, \quad \text{and} \quad u_{i,i+1} = -1$$

$$A = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \cdots & \cdots & \cdots & -1 \\ & & & & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & & & & & \\ -\frac{1}{2} & 1 & & & & \\ -\frac{2}{3} & -\frac{2}{3} & 1 & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ \frac{1-n}{n} & \frac{1-n}{n} & \frac{1-n}{n} & \frac{1-n}{n} & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 & -1 & -1 & \cdots & -1 \\ & \frac{3}{2} & -1 & -1 & \cdots & -1 \\ & & \frac{4}{3} & -1 & \cdots & -1 \\ & & & \cdots & \cdots & -1 \\ & & & & & \frac{n+1}{n} \end{bmatrix} = LU$$

Clearly the result of the matrix multiplication will be A.

c) Discuss the differences between solving $Ax = b$ by multiplying with the inverse, and by applying backward and forward substitution to the LU factorization.

Computing A^{-1} is computationally more expensive than decomposing A into upper and lower matrices. If you want to solve $Ax = b$ most efficiently you will use the LU Decomposition.

10 Newton's Method

What do you think of the idea of applying Newton's Method to the square linear system $Ax = b$?

Since $Ax = b$ is a linear system Newton's Method will converge with order 1.

11 A Cubically Convergent Method

Consider the iteration:

$$x_{k+1} = g(x_k) \quad \text{where} \quad g(x) = x - \frac{f(x)}{f'(x)} - \frac{1}{2}\frac{(f(x))^2 f''(x)}{(f'(x))^3}$$

We assume f is sufficiently often differentiable and $f'(x) \neq 0$. Suppose that $f(\alpha) = 0$ and $g(\alpha) = \alpha$. Show that $g'(\alpha) = g''(\alpha) = 0$ and thus the fixed point method will converge of order at least 3 if we start sufficiently close to α.

Proof.

$$g(x) = x - \frac{f(x)}{f'(x)} - \frac{1}{2}\frac{(f(x))^2 f''(x)}{(f'(x))^3}$$

$$\implies g'(x) = 1 - \left(-\frac{f(x)f''(x)}{(f'(x))^2} + \frac{f'(x)}{f'(x)}\right) - \frac{1}{2}\left(\frac{(-3)(f(x))^2(f''(x))^2}{(f'(x))^4} + \frac{(f(x))^2 f'''(x) + 2f(x)f'(x)f''(x)}{(f'(x))^3}\right)$$

$$\implies g'(x) = \frac{3}{2}\frac{(f)^2(f'')^2}{(f')^4} - \frac{(f)^2(f''')^2}{2(f')^3} = f(x)\left(\frac{3}{2}\frac{f(f'')^2}{(f')^4} - \frac{f(f''')^2}{2(f')^3}\right)$$

By assumption $f(\alpha) = 0$ so the above shows that $g'(\alpha) = f(\alpha) \cdot (\text{other stuff}) = 0 \cdot (\text{other stuff}) = 0$.

$$g''(x) = \frac{d}{dx}\left(\frac{3}{2}\frac{(f)^2(f'')^2}{(f')^4} - \frac{(f)^2(f''')^2}{2(f')^3}\right)$$

When you do this computation some terms will cancel and every remaining term will contain an $f(x)$ which when evaluated at α is zero, therefore $g''(\alpha) = 0$.

□

12 Induced Matrix Norms

Show that $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^n |a_{ij}|$.

Proof. Let $\|x\|_1 = \sum_{j=1}^n |x_j|$ and $\|Ax\|_1 = \sum_{i=1}^n \left|\sum_{j=1}^n |a_{ij}|\right|$.

By the triangle inequality $\|Ax\|_1 \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}||x_j|$.

We can rearrange the right hand side of this inequality to look like $\sum_{j=1}^n |x_j| \sum_{i=1}^n |a_{ij}|$.

Now let $\alpha = \max_{j=1,..,n} \sum_{i=1}^n |a_{ij}|$.

Therefore $\|Ax\|_1 \leq \alpha \, \|x|_1$ which implies $\|A\|_1 \leq \alpha$

Let the maximum of $\sum_{i=1}^n |a_{ij}|$ happen when $i = i^*$ and the vector x be the unit vector with a 1 in the i^*th position so that $\|x\|_1 = 1$.

$\|Ax\|_1 = \sum_{i=1}^n \left|\sum_{j=1}^n |a_{ij}|\right| = \sum_{i=1}^n |a_{ii^*}| = \max_{j=1,\ldots,n} \sum_{i=1}^n |a_{ij}|$ ∎

13 Reverse Triangle Inequality

Show that for all vector norms $\|\cdot\|$ and vectors x and y: $\|x-y\| \geq |\|x\| - \|y\||$ and $\|x+y\| \geq |\|x\| - \|y\||$.

Proof. By the regular triangle inequality:

$$\|x - y + y\| \leq \|x - y\| + \|y\|$$

$$\|x\| \leq \|x - y\| + \|y\|$$
$$\|x\| - \|y\| \leq \|x - y\|$$

Similarly for $\|y - x\|$

$$\|x + y - x\| \leq \|x\| + \|y - x\|$$
$$\|y\| \leq \|x\| + \|y - x\|$$
$$\|y\| - \|x\| \leq \|y - x\|$$

If you combine both results and notice that $\|y - x\| = \|x - y\|$ you have:

$$\|x - y\| \geq |\|x\| - \|y\||$$

To show $\|x+y\| \geq |\|x\| - \|y\||$, we follow a similar argument: By the regular triangle inequality:

$$\|x + y - x\| \leq \|x + y\| + \|-x\|$$

$$\|y\| \leq \|x + y\| + \|x\|$$
$$\|y\| - \|x\| \leq \|x + y\|$$

Similarly for $\|y + x\|$

$$\|x + y - y\| \leq \|x + y\| + \|-y\|$$
$$\|x\| \leq \|x + y\| + \|y\|$$
$$\|x\| - \|y\| \leq \|x + y\|$$

If you combine both results and notice that $\|y + x\| = \|x + y\|$ you have:

$$\|x + y\| \geq |\|x\| - \|y\||$$

∎