Linjärisering. Jacobimatris. Kedjeregeln. Newtons metod. [1]

Introduction

In this lecture we first write partial derivatives in matrix form. This is very convenient if we have many variables and when we compute partial derivatives in MATLAB. The matrix form of the formulas involving partial derivatives are also easier to remember.

We then formulate Newton's method for systems of equations. The fixed point iteration (and hence also Newton's method) works equally well for systems of equations. For example,

$$\begin{aligned} x_2\big(1-x_1^2\big) &= 0,\\ 2-x_1x_2 &= 0, \end{aligned}$$

is a system of two equations in two unknowns. See Problem 1.6 below. If we define two functions

$$\begin{aligned} f_1(x_1,x_2) &= x_2\big(1-x_1^2\big),\\ f_2(x_1,x_2) &= 2-x_1x_2, \end{aligned}$$

then the equations may be written

$$\begin{aligned} f_1(x_1,x_2) &= 0,\\ f_2(x_1,x_2) &= 0. \end{aligned}$$

With $f=(f_1,f_2)$, $x=(x_1,x_2)$, and $0=(0,0)$, we note that $f:\mathbf{R}^2\to\mathbf{R}^2$ and we can write the equations in the compact form

$$f(x)=0.$$

In this lecture we will see how Newton's method can be applied to such systems of equations.

Note that the bisection algorithm can only be used for a single equation, but not for a system of several equations. This is because it relies on the fact the the graph of a Lipschitz continuous function $f:\mathbf{R}\to\mathbf{R}$ must pass the value zero if it is positive in one point and negative in another point. This has no counterpart for functions $f:\mathbf{R}^2\to\mathbf{R}^2$.

Before we discuss Newton's method we need to define derivatives of such functions, namely, two functions of two variables, and more generally several functions of several variables.

1.1 Function of one variable, $f:\mathbf{R}\to\mathbf{R}$

A function $f:\mathbf{R}\to\mathbf{R}$ of one variable is differentiable at a if the following limit exists:

$$f'(a)=\lim_{x\to a}\frac{f(x)-f(a)}{x-a}.$$

We write this in an equivalent form:

$$\lim_{x\to a}\frac{f(x)-f(a)-f'(a)(x-a)}{x-a}=0.$$

Therefore we can say that a function $f:\mathbf{R}\to\mathbf{R}$ of one variable is differentiable at a if there is a number $m(a)$, such that

$$\lim_{x\to a}\frac{f(x)-f(a)-m(a)(x-a)}{x-a}=0. \tag{1}$$

[1] 2012–03–22, Stig Larsson, Matematiska vetenskaper, Chalmers tekniska högskola

Of course, $m(a)$ is the derivative of f at a:

$$m(a) = f'(a) = Df(a) = \frac{df}{dx}(a).$$

This formulation will be useful when we define the derivative of a function of two variables later. We also obtain the linearization formula

$$f(x) = f(a) + f'(a)(x-a) + E_f(x,a), \tag{2}$$

where the linearization error E_f is smaller than the second term on the right side when x is close to a. (See Theorem 11 in Adams 4.9.)

It is convenient to use the abbreviation $h = x - a$, so that $x = a + h$ and (1) becomes

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - f'(a)h}{h} = 0, \tag{3}$$

and (2) becomes

$$f(x) = f(a+h) = f(a) + f'(a)h + E_f(x,a), \tag{4}$$

where $E_f(x,a)/h \to 0$ as $h \to 0$ according to (3). Note that the first term on the right side, $f(a)$, is constant with respect to x. The second term,

$$f'(a)h = f'(a)(x-a), \tag{5}$$

is a linear function of the increment $h = x - a$. These two terms form the *linearization of f at a*,

$$L(x) = f(a) + f'(a)(x-a). \tag{6}$$

(See Adams 4.9.) The straight line $y = L(x)$ is the tangent to the curve $y = f(x)$ at a.

Example 1. Let $f(x) = x^2$. Then $f'(x) = 2x$ and the linearization at $a = 3$ is

$$L(x) = 9 + 6(x-3).$$

Numerical computation of the derivative

In a previous lecture, `BM 6.1. Numerisk beräkning av derivata`, and a previous computer exercise, `Numerisk derivata`, we learnt how to compute the derivative numerically. We quickly repeat it here. If we divide (4) by h, then we get

$$\frac{f(a+h) - f(a)}{h} = f'(a) + E_f(x,a)/h. \tag{7}$$

Here the remainder $E_f(x,a)/h \to 0$ when $h \to 0$. This suggests that we can approximate the derivative by the difference quotient

$$f'(a) \approx \frac{f(a+h) - f(a)}{h}. \tag{8}$$

A better approximation is obtained by the symmetric difference quotient:

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h}. \tag{9}$$

The difference quotients in (8) and (9) are of the form ”small number divided by small number”. If this is computed with round-off error on a computer, then the total error will be large if the step h is very small. Therefore we must choose the step “moderately small” here, see `BM 6.1. Numerisk beräkning av derivata`. It can be shown that in Matlab a good choice for (8) is $h = 10^{-8}$ and for (9) $h = 10^{-5}$.

1.2 Function of two variables, $f : \mathbf{R}^2 \to \mathbf{R}$

Let $f(x_1, x_2)$ be a function of two variables, i.e., $f : \mathbf{R}^2 \to \mathbf{R}$. We want to imitate the formula in (3). We write $x = (x_1, x_2)$ and $f(x) = f(x_1, x_2)$ and introduce the increment vector $h = (h_1, h_2) = (x_1 - a_1, x_2 - a_2)$ and its length $|h| = \sqrt{h_1^2 + h_2^2}$.

We now say that function f is differentiable at $a = (a_1, a_2)$, if there are numbers $m_1(a)$, $m_2(a)$, such that

$$\lim_{|h|\to 0} \frac{f(a+h) - f(a) - m_1(a)h_1 - m_2(a)h_2}{|h|} = 0. \tag{10}$$

The corresponding linearization formula is

$$f(x) = f(a+h) = f(a) + m_1(a)h_1 + m_2(a)h_2 + E_f(x, a), \tag{11}$$

where the linearization error E_f is smaller than the second and third terms on the right side, more precisely, $E_f(x, a)/|h| \to 0$ as $|h| \to 0$.

If we take $h = (h_1, 0)$, then we get

$$f(x_1, a_2) = f(a_1 + h_1, a_2) = f(a) + m_1(a)h_1 + E_f(x, a).$$

By comparison with (4) we see that this means that $m_1(a)$ is the derivative of the one-variable function $\hat{f}(x_1) = f(x_1, a_2)$, obtained from f by keeping $x_2 = a_2$ fixed. By taking $h = (0, h_2)$ we see in a similar way that $m_2(a)$ is the derivative of the one-variable function, which is obtained from f by keeping $x_1 = a_1$ fixed. The numbers $m_1(a)$, $m_2(a)$ are called the *partial derivatives* of f at a and we denote them by

$$m_1(a) = f_1'(a) = f_{x_1}'(a) = \frac{\partial f}{\partial x_1}(a), \quad m_2(a) = f_2'(a) = f_{x_2}'(a) = \frac{\partial f}{\partial x_2}(a). \tag{12}$$

Now (11) may be written

$$f(x) = f(a+h) = f(a) + f_{x_1}'(a)h_1 + f_{x_2}'(a)h_2 + E_f(x, a), \quad h = x - a. \tag{13}$$

It is convenient to write this formula by means of matrix notation. Let

$$a = \begin{bmatrix} a_1, a_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

We say that a is a row matrix of type 1×2 (one by two) and that b is a column matrix of type 2×1 (two by one). Their product is defined by

$$ab = \begin{bmatrix} a_1, a_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1 b_1 + a_2 b_2.$$

The result is a matrix of type 1×1 (one real number), according to the rule: $(1 \times 2)(2 \times 1) = 1 \times 1$.

Going back to (13) we define

$$f'(a) = Df(a) = \begin{bmatrix} f_{x_1}'(a) & f_{x_2}'(a) \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}.$$

The row matrix

$$f'(a) = Df(a) = \begin{bmatrix} f_{x_1}'(a) & f_{x_2}'(a) \end{bmatrix}$$

is called the derivative (or Jacobi matrix) of f at a. It is also called the gradient vector of f at a and written $\nabla f(a)$. Then (13) may be written

$$\begin{aligned} f(x) = f(a+h) &= f(a) + \begin{bmatrix} f_{x_1}'(a) & f_{x_2}'(a) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + E_f(x, a) \\ &= f(a) + f'(a)h + E_f(x, a), \quad h = x - a. \end{aligned} \tag{14}$$

Note that the first term on the right side, $f(a)$, is constant with respect to x. The second term,

$$f'(a)h = f'(a)(x-a), \tag{15}$$

is a linear function of the increment $h = x - a$. These terms are called the *linearization of f at a*,

$$L(x) = f(a) + f'(a)(x-a). \tag{16}$$

The plane $x_3 = L(x_1, x_2)$ is the tangent plane at $(a_1, a_2, f(a_1, a_2))$ to the surface $x_3 = f(x_1, x_2)$.

Example 2. Let $f(x) = x_1^2 x_2^5$. Then

$$\frac{\partial f}{\partial x_1}(x) = \frac{\partial f}{\partial x_1}\Big(x_1^2 x_2^5\Big) = 2x_1 x_2^5, \quad \frac{\partial f}{\partial x_2}(x) = \frac{\partial f}{\partial x_2}\Big(x_1^2 x_2^5\Big) = 5x_1^2 x_2^4,$$

so that $f'(x) = \begin{bmatrix} 2x_1 x_2^5 & 5x_1^2 x_2^4 \end{bmatrix}$ and the linearization at $a = (3, 1)$ is

$$L(x) = 9 + \begin{bmatrix} 6 & 45 \end{bmatrix} \begin{bmatrix} x_1 - 3 \\ x_2 - 1 \end{bmatrix}.$$

The tangent plane at $(3, 1, 9)$ to the surface $x_3 = x_1^2 x_2^5$ is

$$x_3 = 9 + 6(x_1 - 3) + 45(x_2 - 1).$$

1.3 Two functions of two variables, $f : \mathbf{R}^2 \to \mathbf{R}^2$

Let $f_1(x_1, x_2)$, $f_2(x_1, x_2)$ be two functions of two variables. We write $x = (x_1, x_2)$ and $f(x) = (f_1(x_1, x_2), f_2(x_1, x_2))$, i.e., $f : \mathbf{R}^2 \to \mathbf{R}^2$. The function f is differentiable at $a = (a_1, a_2)$, if there are numbers $m_{11}(a)$, $m_{12}(a)$, $m_{21}(a)$, $m_{22}(a)$ such that

$$\begin{aligned} f_1(x) = f_1(a+h) = f_1(a) + m_{11}(a)h_1 + m_{12}(a)h_2 + E_{f_1}(x, a), \\ f_2(x) = f_2(a+h) = f_2(a) + m_{21}(a)h_1 + m_{22}(a)h_2 + E_{f_2}(x, a), \end{aligned} \tag{17}$$

where $h = x - a$ and the linearization errors E_{f_j} satisfy $E_{f_j}(x, a)/|h| \to 0$ when $|h| \to 0$. As before $|h| = \sqrt{h_1^2 + h_2^2}$ denotes the norm (length) of the increment vector $h = (h_1, h_2) = (x_1 - a_1, x_2 - a_2)$. From the previous subsection we recognize that the constants $m_{ij}(a)$ are the partial derivatives of the functions f_i at a and we denote them by

$$\begin{aligned} m_{11}(a) = f'_{1,x_1}(a) = \frac{\partial f_1}{\partial x_1}(a), \quad m_{12}(a) = f'_{1,x_2}(a) = \frac{\partial f_1}{\partial x_2}(a), \\ m_{21}(a) = f'_{2,x_1}(a) = \frac{\partial f_2}{\partial x_1}(a), \quad m_{22}(a) = f'_{2,x_2}(a) = \frac{\partial f_2}{\partial x_2}(a). \end{aligned}$$

It is convenient to use matrix notation. Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

We say that A is a matrix of type 2×2 (two by two) and that b is a column matrix of type 2×1 (two by one). Their product is defined by

$$Ab = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_{11}b_1 + a_{12}b_2 \\ a_{21}b_1 + a_{22}b_2 \end{bmatrix}.$$

The result is a matrix of type 2×1 (column matrix), according to the rule: $(2 \times 2)(2 \times 1) = 2 \times 1$.

Going back to (17) we define

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix}, \quad f'(a) = Df(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}. \tag{18}$$

The matrix $f'(a) = Df(a)$ is called the derivative (or Jacobi matrix) of f at a. Then (17) may be written

$$\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} f_1(a+h) \\ f_2(a+h) \end{bmatrix} = \begin{bmatrix} f_1(a) \\ f_2(a) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} E_{f_1}(x,a) \\ E_{f_2}(x,a) \end{bmatrix}, \tag{19}$$

or in more compact form

$$f(x) = f(a+h) = f(a) + f'(a)h + E_f(x,a), \quad h = x - a. \tag{20}$$

Note that it is important that f, h are written as column vectors here. Note that $x = (x_1, x_2)$, $a = (a_1, a_2)$ are *points* not vectors, and are therefore sometimes written with ordinary parentheses, ignoring if they are rows or columns. On the other hand h is an "arrow" that goes from the point a to the point x and must be written as a column vector.

Note that the first term on the right side, $f(a)$, is constant with respect to x. The second term,

$$f'(a)h = f'(a)(x-a), \tag{21}$$

is a linear function of the increment $h = x - a$. These terms are called the *linearization of f at a*,

$$L(x) = f(a) + f'(a)(x-a). \tag{22}$$

Example 3. Let $f(x) = \begin{bmatrix} x_1^2 x_2^5 \\ x_2^3 \end{bmatrix}$. Then

$$f'(x) = Df(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) \\ \frac{\partial f_2}{\partial x_1}(x) & \frac{\partial f_2}{\partial x_2}(x) \end{bmatrix} = \begin{bmatrix} 2x_1 x_2^5 & 5x_1^2 x_2^4 \\ 0 & 3x_2^2 \end{bmatrix}$$

and the linearization at $a = (3, 1)$ is

$$L(x) = \begin{bmatrix} 9 \\ 1 \end{bmatrix} + \begin{bmatrix} 6 & 45 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 - 3 \\ x_2 - 1 \end{bmatrix}.$$

1.4 Several functions of several variables, $f : \mathbf{R}^n \to \mathbf{R}^m$

It is now easy to generalize to any number of functions in any number of variables. Let f_i be m functions of n variables x_j, i.e., $f : \mathbf{R}^n \to \mathbf{R}^m$. As in (18) we define

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{bmatrix},$$

$$f(x) = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_m(x_1, \ldots, x_n) \end{bmatrix}, \quad f'(a) = Df(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}.$$

The $m \times n$ matrix $f'(a) = Df(a)$ is called the derivative (or Jacobi matrix) of f at a. In a similar way to (20) we get

$$f(x) = f(a+h) = f(a) + f'(a)h + E_f(x,a), \quad h = x - a. \tag{23}$$

The linearization of f at a is

$$L(x) = f(a) + f'(a)(x-a). \tag{24}$$

Numerical computation of the derivative. In order to compute the j-th column $\dfrac{\partial f}{\partial x_j}(a)$ of the Jacobi matrix, we choose the increment h such that $h_j = \delta$ and $h_i = 0$ for $i \neq j$, i.e.,

$$h = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \delta \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \delta \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \delta e_j, \quad e_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \text{ row number } j.$$

Here the steplength δ is a small positive number and e_j is the j-th standard basis vector. If we use this increment in a symmetric difference quotient, see (9), we get

$$\frac{\partial f}{\partial x_j}(a) \approx \frac{f(a+\delta e_j) - f(a - \delta e_j)}{2\delta}. \tag{25}$$

Note that f is a column so the result is a column: the j^{th} column of the matrix $f'(a)$. Remember that the steplength δ should be small, but not too small.

1.5 The chain rule

(Adams 12.5) It is now very easy to write down the chain rule in its general form. Assume that

1. $g : \mathbf{R}^n \to \mathbf{R}^m$, $f : \mathbf{R}^m \to \mathbf{R}^p$,
2. $b = g(a)$,
3. the derivatives $g'(a)$ and $f'(b)$ exist.

Then the composite function $u = f \circ g : \mathbf{R}^n \to \mathbf{R}^p$, defined by $u(x) = f(g(x))$, is differentiable at a with the derivative

$$u'(a) = f'(b)g'(a), \quad \text{where } b = g(a).$$

This is a product of Jacobi matrices of the type

$$p \times n = (p \times m)(m \times n),$$

more precisely,

$$\begin{bmatrix} \frac{\partial u_1}{\partial x_1}(a) & \dots & \frac{\partial u_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial u_p}{\partial x_1}(a) & \dots & \frac{\partial u_p}{\partial x_n}(a) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(b) & \dots & \frac{\partial f_1}{\partial x_m}(b) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(b) & \dots & \frac{\partial f_p}{\partial x_m}(b) \end{bmatrix} \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(a) & \dots & \frac{\partial g_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1}(a) & \dots & \frac{\partial g_m}{\partial x_n}(a) \end{bmatrix}.$$

1.6 Newton's method for $f(x) = 0$

Consider a system of n equations with n unknowns:

$$\begin{aligned} f_1(x_1, \dots, x_n) &= 0, \\ &\vdots \\ f_n(x_1, \dots, x_n) &= 0. \end{aligned}$$

If we define

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad f = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}, \quad 0 = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix},$$

then $f : \mathbf{R}^n \to \mathbf{R}^n$, and we can write our system of equations in the compact form

$$f(x) = 0. \tag{26}$$

Suppose that we have found an approximate solution a. We want to find a better approximation $x = a + h$. Instead of solving (26) directly, which is usually impossible, we solve the *linearized equation at* a:

$$L(a + h) = f(a) + f'(a)h = 0. \tag{27}$$

We must solve for the increment h. Rearranging the terms we get

$$f'(a)h = -f(a). \tag{28}$$

Remember that the Jacobi matrix $f'(a)$ is of type $n \times n$ and the increment h is of type $n \times 1$. Therefore we have to solve a linear system of n equations with n unknowns to get the increment h. It is of the form $Ah = b$ with $A = f'(a)$ and $b = -f(a)$. Then we set $x = a + h$.

In algorithmic form Newton's method can be formulated:

```
while |h|>tol
   evaluate the residual    b=-f(x)
   evaluate the Jacobian    A=f'(x)
   solve the linear system  Ah=b
   update                   x=x+h
end
```

You will implement this algorithm in the studio exercises. You will use the MATLAB command

```
h=A\b
```

to solve the system.

Problems

Problem 1.1. Let

$$a = \begin{bmatrix} 1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Compute the products $\quad ab, \quad Ab, \quad Aa.$

Problem 1.2. Compute the Jacobi matrix $f'(x)$ (also denoted $Df(x)$). Compute the linearization of f at $\bar{x}$.

$$\text{(a)} \quad f(x) = \begin{bmatrix} \sin(x_1) + \cos(x_2) \\ \cos(x_1) + \sin(x_2) \end{bmatrix}, \quad \bar{x} = 0; \qquad \text{(b)} \quad f(x) = \begin{bmatrix} 1 \\ 1 + x_1 \\ 1 + x_1 e^{x_2} \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Problem 1.3. Compute the gradient vector $\nabla f(x)$ (also denoted $f'(x) = Df(x)$). Compute the linearization of f at $\bar{x}$.

$$\text{(a)} \quad f(x) = e^{-x_1} \sin(x_2), \quad \bar{x} = 0; \qquad \text{(b)} \quad f(x) = |x|^2 = x_1^2 + x_2^2 + x_3^2, \quad x \in \mathbf{R}^3, \quad \bar{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Problem 1.4. Here $f : \mathbf{R} \to \mathbf{R}^2$. Compute $f'(t)$. Compute the linearization of f at $\bar{t}$.

$$\text{(a)}\quad f(t) = \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix}, \quad \bar{t} = \pi/2; \qquad \text{(b)}\quad f(t) = \begin{bmatrix} t \\ 1+t^2 \end{bmatrix}, \quad \bar{t} = 0.$$

Problem 1.5. Compute the linearization of f at $\bar{x}$:

$$f(x) = \begin{bmatrix} x_1 - x_1 x_2 \\ -x_2 + x_1 x_2 \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Problem 1.6. (a) Write the system

$$\begin{aligned} u_2\big(1 - u_1^2\big) &= 0, \\ 2 - u_1 u_2 &= 0 \end{aligned}$$

in the form $f(u) = 0$. Find the all the solutions by hand calculation.

(b) Compute the Jacobi matrix $Df(u)$.

(c) Perform the first step of Newton's method for the equation $f(u) = 0$ with initial vector $u^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(d) Solve the equation $f(u)$ with your MATLAB program `newton.m`.

Problem 1.7. (a) Write the system

$$\begin{aligned} u_1\big(1 - u_2\big) &= 0, \\ u_2\big(1 - u_1\big) &= 0, \end{aligned}$$

in the form $f(u) = 0$. Find the all the solutions by hand calculation.

(b) Compute the Jacobi matrix $Df(u)$.

(c) Perform the first step of Newton's method for the equation $f(u) = 0$ with initial vector $u^{(0)} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$.

(d) Solve the equation $f(u)$ with your MATLAB program `newton.m`.

Answers and solutions

1.1. Use MATLAB to check your answers.

$$ab = 5, \quad Ab = \begin{bmatrix} 5 \\ 11 \end{bmatrix}, \quad Aa = \text{not defined}.$$

1.2.

(a)

$$f'(x) = \begin{bmatrix} \cos(x_1) & -\sin(x_2) \\ -\sin(x_1) & \cos(x_2) \end{bmatrix}, \qquad L(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

(b)

$$f'(x) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ e^{x_2} & x_1 e^{x_2} \end{bmatrix}, \qquad L(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) = \begin{bmatrix} 1 \\ 2 \\ 1+e \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ e & e \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}.$$

1.3.

(a)

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} -e^{-x_1}\sin(x_2), & e^{-x_1}\cos(x_2) \end{bmatrix}, \\ L(x) &= f(\bar{x}) + f'(\bar{x})(x - \bar{x}) = 0 + \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_2. \end{aligned}$$

(b)

$$\begin{aligned} \nabla f(x) &= \begin{bmatrix} 2x_1 & 2x_3 & 2x_3 \end{bmatrix}, \\ L(x) &= f(\bar{x}) + f'(\bar{x})(x - \bar{x}) = 3 + \begin{bmatrix} 2 & 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \\ x_3 - 1 \end{bmatrix} = -3 + 2x_1 + 2x_2 + 2x_3. \end{aligned}$$

1.4.

(a)

$$\begin{aligned} f'(t) &= \begin{bmatrix} -\sin(t) \\ \cos(t) \end{bmatrix}, \\ L(t) &= f(\bar{t}) + f'(\bar{t})(t - \bar{t}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 0 \end{bmatrix} (t - \pi/2). \end{aligned}$$

(b)

$$\begin{aligned} f'(t) &= \begin{bmatrix} 1 \\ 2t \end{bmatrix}, \\ L(t) &= f(\bar{t}) + f'(\bar{t})(t - \bar{t}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} t = \begin{bmatrix} t \\ 1 \end{bmatrix}. \end{aligned}$$

1.5.

$$f'(x) = \begin{bmatrix} 1 - x_2 & -x_1 \\ x_2 & -1 + x_2 \end{bmatrix},$$

$$L(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$

1.6. (a) The solutions are given by

$$f(u) = \begin{bmatrix} u_2(1 - u_1^2) \\ 2 - u_1 u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We find two solutions $\bar{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\bar{u} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$.

(b) The Jacobian is

$$Df(u) = \begin{bmatrix} -2u_1 u_2 & 1 - u_1^2 \\ -u_2 & -u_1 \end{bmatrix}.$$

(c) The first step of Newton's method:

evaluate $\quad A = Df(1,1) = \begin{bmatrix} -2 & 0 \\ -1 & -1 \end{bmatrix}$ and $b = -f(1,1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

solve $\quad Ah = b, \quad \begin{bmatrix} -2 & 0 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$$\begin{cases} -2h_1 = 0, \\ -h_1 - h_2 = -1, \end{cases} \quad h = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

update $\quad u^{(1)} = u^{(0)} + h = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \bar{u}$

Bingo! We found one of the solutions.

1.7. (a) The solutions are given by

$$f(u) = \begin{bmatrix} u_1(1 - u_2) \\ u_2(1 - u_1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We find two solutions $\bar{u} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\bar{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(b) The Jacobian is

$$Df(u) = \begin{bmatrix} 1 - u_2 & -u_1 \\ -u_2 & 1 - u_1 \end{bmatrix}.$$

(c) The first step of Newton's method:

evaluate $\quad A = Df(2,2) = \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix}$ and $b = -f(2,2) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

solve $\quad Ah = b, \quad \begin{bmatrix} -1 & -2 \\ -2 & -1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix},$

$$\begin{cases} -h_1 - 2h_2 = 2, \\ -2h_1 - h_2 = 2, \end{cases} \quad h = \begin{bmatrix} -2/3 \\ -2/3 \end{bmatrix}$$

update $\quad u^{(1)} = u^{(0)} + h = \begin{bmatrix} 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -2/3 \\ -2/3 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 4/3 \end{bmatrix}$

Getting closer to one of the solutions $\bar{u}$!

`/stig`