Norwegian University of Science and Technology
Department of Mathematical Sciences

MA2501 Numeriske Metoder
Olivier Verdier

Training Assignment 2

2012-01-19

This assignment has 5 tasks.

Exercise 1. Write down Newton's method for find solutions to the following equation system:

$$x^2 + (y-8)^2 = 8$$
$$x + .15x^2 + xy = 1$$

$$F(x,y) = (x^2 + (y-8)^2 - 8, x + .15x^2 + y - 2)$$

So the Jacobian is

$$F'(x,y) = \begin{bmatrix} 2x & 2y \\ 1 + .3x + y & x \end{bmatrix}$$

Newton's method is thus

$$2x_n \triangle x + 2y \triangle y = -x_n^2 - (y_n - 8)^2 + 8$$
$$(2 + .3x_n + y_n)\triangle x + x\triangle y = -x_n - .15x_n^2 - x_n y_n + 1$$

Exercise 2. What is the following code doing?

```
x = 10.
for i in range(10):
  y = x**2
```

What is the value of `y` after the code is executed? What do you think of the usefulness of that code?

This code repeats the instruction `y = x**2` ten times. It is not very useful as one iteration would have produced the same result.

Exercise 3. Write down the simplified Newton's algorithm (i.e., when the derivative of the function is only computed at the first point x_0), in one dimension (i.e., we are looking for a root of a function from $\mathbf{R}$ to $\mathbf{R}$). Express it as a fixed point method. What is the condition for this fixed point method to converge? Will this method always converge if the starting point is sufficiently close to a root of f?

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}$$

The fixed point function G is thus

$$G(x) = x - \frac{f(x)}{f'(x_0)}$$

The condition for convergence is that at the fixed point $\bar{x}$ we have $|G'(\bar{x})| < 1$, so

$$\left| 1 - \frac{f'(\bar{x})}{f'(x_0)} \right| < 1$$

Exercise 4. Consider the floating point system with 3 significant decimal digits and 2 decimal exponents (i.e., the possible exponents range from -50 to 49), and without tricks, so we cannot represent the zero.

4.a) Prove that if the first digit is not allowed to be zero, two different set of digits lead to two different numbers.

4.b) Assuming that the first significant digit is not zero, what is

- The smallest possible positive, non-zero number?
- The smallest number strictly greater than one?
- The value of the *machine epsilon*?
- The biggest possible number?

Smallest number is 1.00×10^{-50}. Smallest number greater than one is 1.01, so the machine epsilon is 10^{-2}. The biggest number is 999×10^{49}.

Exercise 5. A 2×2 matrix A and a vector $b \in \mathbf{R}^2$ are fixed. Consider the fixed point problem

$$Ax + b = x.$$

5.a) Find an example of A and b such that there is no solution. What is the condition for this problem to have a solution?

For instance, if $A = I$, then the equation is $x + b = x$, which has a solution only if $b = 0$. So, in general, if $A = I$ and $b \neq 0$, there is no solution.

The general condition is obtained by rewriting the equation as

$$(A - I)x = b.$$

There is a unique solution if and only if $A - I$ is invertible.

5.b) What is the condition for the fixed point iteration

$$x_{n+1} = Ax_n + b$$

to converge to a given fixed point $\bar{x}$?

The solution of the equation $Ax + b = x$ by $\bar{x}$, so we have

$$\bar{x} = A\bar{x} + b. \tag{1}$$

.

The error E_n at step n is

$$E_n \coloneqq x_n - \bar{x}.$$

We have (using (1))

$$E_{n+1} = x_{n+1} - \bar{x} = (Ax_n + b) - (A\bar{x} + b) = A(x_n - \bar{x}) = AE_n.$$

Now, suppose that A is diagonalisable, that is $A = PDP^{-1}$, where D is a diagonal matrix with diagonal $(\lambda_1, \dots, \lambda_d)$. (We can always suppose that A is diagonalisable, because any matrix, *possibly after an arbitrary small perturbation*, is diagonalisable) The relation with the successive errors is now written

$$E_{n+1} = PDP^{-1}E_n.$$

Now, let us denote

$$e_n \coloneqq P^{-1}E_n.$$

We obtain:

$$e_{n+1} = De_n$$

Now, the component i of the error e_n at step n, namely e_n^i, is related to the previous error component by

$$e_{n+1}^i = \lambda_i e_n^i$$

As a result, the i-th component e_n^i converges to zero if and only if $|\lambda_i| < 1$.

Now, the condition for the fixed point iteration to converge is that E_n converges to zero. That is equivalent to the convergence of e_n to zero. That is equivalent to all the components of e_n converging to zero. That is equivalent to

$$|\lambda_i| < 1 \qquad 1 \leq i \leq d$$

If we denote the *spectral radius*

$$\rho(A) := \max_{1 \leq i \leq d} |\lambda_i|,$$

the condition for convergence is

$$\rho(A) < 1.$$