NTNU

Norwegian University of Science and Technology

TMA4125 Matematikk 4N

Numerical Solution of Nonlinear Equations

Ronny Bergmann

Department of Mathematical Sciences, NTNU.

February 1, 2022

Introduction.

We know that quadratic equations of the form

$$ax^2 + bx + c = 0$$

have the roots (or solutions) $r^{\pm} = \dfrac{-b \pm \sqrt{d^2 - 4ac}}{2a}$

More generally, for a given function f we want to consider the equation

$$f(x) = 0. \tag{1}$$

A solution x^* of (1) is called a root to the equation.

Challenge. In many applications, we encounter equations for which we do not have a simple solution formula as for quadratic functions.
In fact, an analytical solution might not even exist!
$\Rightarrow$ Develop numerical techniques to solve (1).

Scalar and systems of equations

We consider scalar equations first, i. e. $f : \mathbb{R} \to \mathbb{R}$, or with just one equation and variable, for example

$$x^3 + x^2 - 3x = 3.$$

Later we will also consider systems of equations, for example

$$\begin{aligned} x\mathrm{e}^y &= 1, \\ -x^2 + y &= 1. \end{aligned}$$

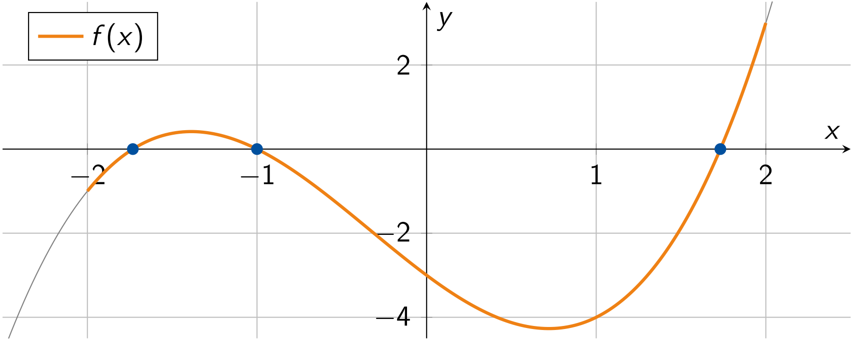
We can write this also as a functions $f : \mathbb{R}^n \to \mathbb{R}^n$, which in this example would be $n = 2$ and

$$f(x, y) = \begin{pmatrix} x\mathrm{e}^y - 1 \\ -x^2 + y - 1 \end{pmatrix} \quad \text{and we want to solve} \quad f(x, y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Scalar equations

Let $f : [a, b] \to \mathbb{R}$ be a continuous function on some interval $[a, b]$.

Goal. Find a zero or root of f.
Let's get an intuition by plotting the function (try yourself in Python).
For example $f(x) = x^3 + x^2 - 3x - 3$ from last slide looks like



$\Rightarrow$ three real roots on $[-2, 2]$. Writing f as $f(x) = (x+1)(x^2-3)$
$\Rightarrow x_1^* = -1$, $x_{2,3}^* = \pm\sqrt{3}$.

Existence and uniqueness of solutions

Theorem. Let $f \in C[a,b]$ be given.

1. if $f(a)$ and $f(b)$ are of different sign, then there exists at lesat one $r \in (a,b)$ such that $f(r) = 0$
   Trick. the condition of different sign is $f(a)f(b) < 0$ in short.
2. The solution is unique if $f \in C^1[a,b]$ and $f'(x) < 0$ or $f'(x) > 0$ for all $x \in (a,b)$.

Bisection method

First part of the theorem already provides a very intuitive algorithm: Divide $[a, b]$ into two parts $[a, c], [c, b]$, check in which half the zero is and continue there.

Algorithm.

Input a function f and an interval, such that $f(a)f(b) < 0$

1. Set $a^{(0)} = a$, $b^{(0)} = b$
2. For $k = 0, 1, 2, \ldots$
   - $c^{(k)} = \dfrac{a^{(k)} + b^{(k)}}{2}$
   - Set the next interval to

$$[a^{(k+1)}, b^{(k+1)}] = \begin{cases} [a^{(k)}, c^{(k)}] & \text{if } f(a^{(k)})f(c^{(k)}) < 0 \\ [c^{(k)}, b^{(k)}] & \text{if } f(c^{(k)})f(b^{(k)}) < 0. \end{cases}$$

Idea. $c^{(k)}$ approximates the root r, note that $|c^{(k)} - r| \leq \frac{b^{(k)} - a^{(k)}}{2}$

We can stop if $\frac{b^{(k)} - a^{(k)}}{2}$ is small enough or if $f(c^{(k)})$ is zero.

Bisection method – exercises and summary

Exercises.

- Choose appropriate intervals and find the other two roots of f
- Compute the solution(s) of $x^2 + sin(x) - 0.5 = 0$ using the bisection method
- Given the interval $[1.5, 2]$. How many iterations are required to guarantee that the error $|c^{(k)} - r|$ is less than 10^{-4}

Summary.
The bisection method is very robust but not particularly fast.

Fix point iterations

A major class of iteration schemes are the so-called fix point iterations.

General Idea. Given an equation $f(x) = 0$ with root r.
Rewrite the equation

$$f(x) = 0$$

to a a fix point form $x = g(x)$, i. e. construct a function g such that the root x^* of f is a fixed point of g, that is x^* satisfies

$$x^* = g(x^*).$$

Algorithm.

Input Given a function g and a starting value $x^{(0)}$

1. For $k = 0, 1, 2, \ldots$ compute

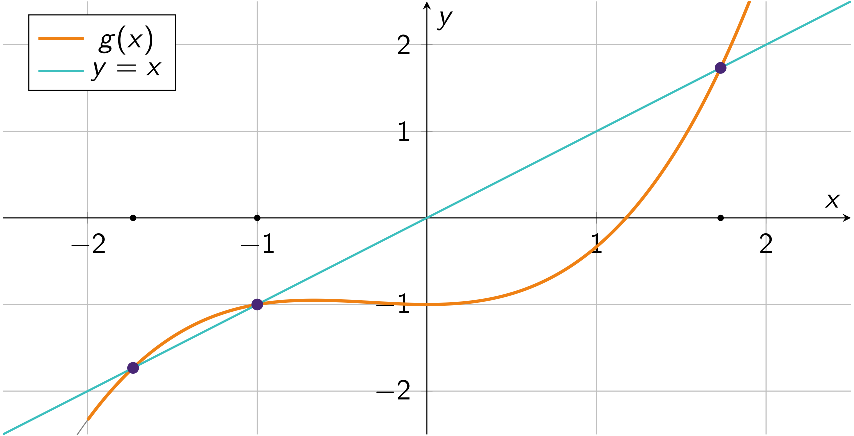
$$x^{(k+1)} = g(x^{(k)}).$$

Example. Fixed point equation

We rewrite

$$f(x) = x^3 + x^2 - 3x - 3 = 0 \quad \text{to} \quad x = \frac{x^3 + x^2 + 3}{3} = g(x).$$

Interpretation: Fixed points are intersections of $g(x)$ with $y = x$.

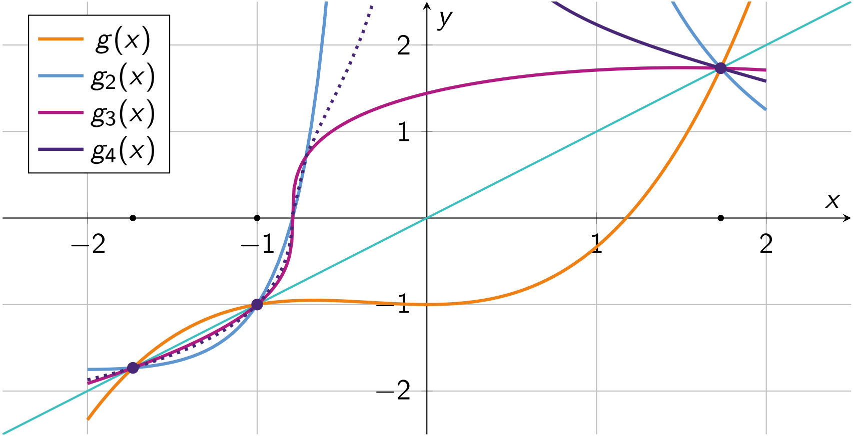


Example. Fixed point equation (cont.)

Exercise. Repeat the experiment with $x^{(0)} = 1.5$ and

$$g_2(x) = \frac{-x^2+3x+3}{x^2}, \quad g_3(x) = \sqrt[3]{3+3x-x^2}, \quad g_4(x) = \sqrt{\frac{3+3x-x^2}{x}}$$

Interpretation: Fixed points are intersections of $g(x)$ with $y = x$.



Existence and uniqueness for fix point iteration

We can adapt the previous Theorem on the equation $f(x) = x - g(x) = 0$

Theorem. Let $g \in C[a, b]$

1. if $a < g(x) < b$ for all $x \in (a, b)$ then g has at least one fixed point x^*
2. if $g \in C^1[a, b]$ and $|g'(x)| < 1$ for all $x \in [a, b]$ then the fixed point x^* is unique.

We write the assumption $a < g(x) < b$ for all $x \in [a, b]$
as $g([a, b]) \subset (a, b)$.

Convergence of fix the point iteration

We denote the error after k iterations as $e^{(k)} = x^* - x^{(k)}$.

The iteration converges if $e^{(k)} \to 0$ as $k \to \infty$.
Under which conditions is this the case?

Here's a trick we will use: For any k we have

$$\begin{aligned} x^{(k+1)} &= g(x^{(k)}), && \text{the iterations} \\ x^* &= g(x^*) && \text{the fixed point.} \end{aligned}$$

Now use the last Theorem from the preliminaries (that generalised Rolle's theorem)

$$|e^{(k+1)}| = |x^* - x^{(k+1)}| = |g(x^*) - g(x^{(k)})| = |g'(\xi_k)| \cdot |x^* - x^{(k)}| = |g'(\xi_k)| \cdot |e_k|$$

where ξ_k is some unknown value between $x^{(k)}$ (known) and x^* (unknown).

The fix point theorem

Theorem. If there is an interval $[a, b]$ such that $g \in C^1[a, b]$, $g([a, b]) \subset (a, b)$ and there exist a positive constant $L < 1$ such that $|g'(x)| \leq L < 1$ for all $x \in [a, b]$, then

- g has a unique fixed point r in (a, b).
- The fixed point iterations $x^{(k+1)} = g(x^{(k)})$ converges towards x^* for all starting values $x^{(0)} \in [a, b]$.

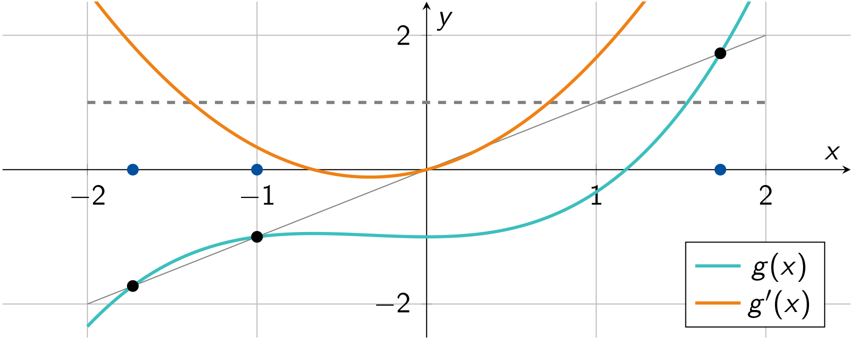
The property $g([a, b]) \subset (a, b)$ guarantees that if $x^{(0)} \in [a, b]$ then $x^{(k)} \in (a, b)$, for $k = 1, 2, \ldots$.
The condition $|g'(x)| \leq L < 1$ guarantees convergence towards the unique fixed point x^*, since

$$|e^{(k+1)}| \leq L|e^{(k)}| \quad \Rightarrow \quad |e^{(k)}| \leq L^k\, |e^{(0)}| \to 0 \quad \text{as } k \to \infty.$$

since $L^k \to 0$ as $k \to \infty$ due to $L < 1$ (also called linear convergence).

Example. The fix point interation (cont. II)

$$g(x) = \frac{x^3 + x^2 - 3}{3}, \qquad g'(x) = \frac{3x^2 + 2x}{3}$$



Exercise. How large can we make $[-0.7, 1.3]$ still having convergence? Check the same conditions for g_2, g_3, g_4.

Newton's method

Since a small $g'(x^*)$ is preferable and we can choose g
$\Rightarrow$ can we choose g such that $g'(x^*) = 0$?

We use $x^{(k)} = x^* - e^{(k)}$ and a Taylor expansion (around x^*)

$$\begin{aligned} e^{(k+1)} &= x^* - x^{(k+1)} = g(x^*) - g(x^{(k)}) = g(x^*) - g(x^* - e^{(k)}) \\ &= -g'(x^*)e^{(k)} + \frac{1}{2}g''(\xi_k)\left(e^{(k)}\right)^2 \end{aligned}$$

Let's choose $g'(x^*) = 0$ and assume that we have a constant M such that $|g''(x)|/2 \leq M$. Then

$$|e^{(k+1)}| \leq M|e^{(k)}|^2$$

Also known as quadratic convergence.

How to find our favourite fix point equation.

Question. Given an equation

$$f(x) = 0$$

with an unknown solution x^*.
Can we find a g with fixed point x^* and $g'(x^*) = 0$?
Idea. Note that for any $h(x)$ we have due to $f(x^*) = 0$ that

$$g(x) = x - h(x)f(x)$$

has a fixed point x^*. The derivate reads

$$g(x) = 1 - h'(x)f(x) - h(x)f'(x)$$

and at the fixed point

$$g'(x^*) = 1 - h(x^*)f'(x^*).$$

Choose $h(x) = 1/f'(x)$ we achieve $g'(x^*) = 0$.

Newton's method. Algorithm

Algorithm.

Input a function f, its derivative f', and a start point $x^{(0)}$.

1. For $k = 0, 1, 2, \ldots$ compute

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}$$

Error analysis

We constructed the method to give quadratic convergence

$$|e^{(k+1)}| \leq M\left|e^{(k)}\right|^2,$$

where $e^{(k)} = x^* - x^{(k)}$.
But under which conditions can we say something about the size of M?

Let's compare the Taylor series around x^* with Newton's method

$$0 = f(x^{(k)}) + f'(x^{(k)})(x^* - x^{(k)}) + \frac{1}{2}f''(\xi_k)(x^* - x^{(k)})^2 \quad \text{(Taylor series)}$$
$$0 = f(x^{(k)}) + f'(x^{(k)})(x^{(k+1)} - x^{(k)}), \quad \text{(Newton's method)}$$

where ξ_k is between x^* and $x^{(k)}$. Subtracting both yields

$$f'(x^{(k)})(x^* - x^{(k+1)}) + \frac{1}{2}f''(\xi_k)(x^* - x^{(k)})^2 = 0 \quad \Rightarrow \quad e^{(k+1)} = -\frac{1}{2}\frac{f''(\xi_k)}{f'(x^{(k)})}\left(e^{(k)}\right)^2$$

$\Rightarrow$ we need $f'(x^{(k)}) \neq 0$, $f \in C^2[a, b]$ and $x^{(0)}$ sufficiently close to x^*.

Convergence of Newton's method

Theorem. Assume that the function f has a root x^*, and let $I_\delta = [x^* - \delta, x^* + \delta]$ for some $\delta > 0$.
Assume further that

- $f \in C^2(I_\delta)$.
- There is a $M > 0$ such that $\left|\dfrac{f''(y)}{f'(x)}\right| \leq 2M, \quad$ for all $x, y \in I_\delta$.

In this case, Newton's method converges quadratically,

$$|e^{(k+1)}| \leq M|e^{(k)}|^2$$

for all starting values satisfying $|x^{(0)} - x^*| \leq \min\{1/M, \delta\}$.

Exercises.

- Repeat the example with different $x^{(0)}$ to find the two other roots.
- Verify quadratic convergence numerically (remember the EOC!).
- Solve $x(1 - \cos(x)) = 0$, both by the bisection and Newton's method, $x^{(0)} = 1$. What difference do you observe?

Systems of equations

Instead of one function f with one variable x we now consider

Systems of nonlinear equations.

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0 \\ f_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0 \end{aligned}$$

We can write this in short as

$$\boldsymbol{f}(\boldsymbol{x}) = 0,$$

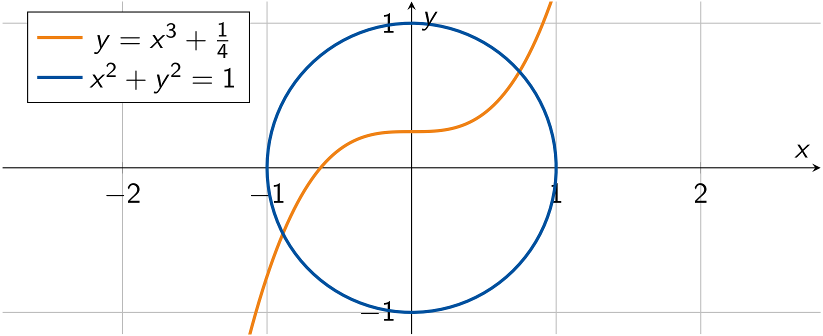
where $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$.

Example.

Example. Consider the two equations

$$x^3 - y + \frac{1}{4} = 0$$
$$x^2 + y^2 - 1 = 0$$

Interpretation.
Rewrite first equation to $y = x^3 + \frac{1}{4} \Rightarrow$ solutions lie on this graph.
Second equation means $1 = x^2 + y^2 \Rightarrow$ point with distance 1 from origin



Towards Newton's method for systems of equations I

Idea. Extend fixed point iterations to systems of equations.
We concentrate on Newton's method and the case $n = 2$:

$$f(x, y) = 0$$
$$g(x, y) = 0$$

to avoid getting lost in indices.

Notation. We denote a solution (root) to these by $\boldsymbol{r} = (x^*, y^*)^{\mathrm{T}}$.
Let $\hat{\boldsymbol{x}} = (\hat{x}, \hat{y})^{\mathrm{T}}$ that approximates $\boldsymbol{r}$.

$\Rightarrow$ Search for a better approximation by linearizing $\boldsymbol{f}(\boldsymbol{x}) = 0$.
In other words: use the multivariate Taylor expansion around $\hat{\boldsymbol{x}}$

$$f(x, y) = f(\hat{x}, \hat{y}) + \frac{\partial f}{\partial x}(\hat{x}, \hat{y})(x - \hat{x}) + \frac{\partial f}{\partial y}(\hat{x}, \hat{y})(y - \hat{y}) + \dots$$
$$g(x, y) = g(\hat{x}, \hat{y}) + \frac{\partial g}{\partial x}(\hat{x}, \hat{y})(x - \hat{x}) + \frac{\partial g}{\partial y}(\hat{x}, \hat{y})(y - \hat{y}) + \dots$$

where we omit higher order terms (in "..."), which are small for $\boldsymbol{x} \approx \hat{\boldsymbol{x}}$.

NTNU

Towards Newton's method for systems of equations II

Idea. Ignore the higher order terms and solve

$$\begin{aligned} f(\hat{x},\hat{y}) + \frac{\partial f}{\partial x}(\hat{x},\hat{y})(x-\hat{x}) + \frac{\partial f}{\partial y}(\hat{x},\hat{y})(y-\hat{y}) &= 0 \\ g(\hat{x},\hat{y}) + \frac{\partial g}{\partial x}(\hat{x},\hat{y})(x-\hat{x}) + \frac{\partial g}{\partial y}(\hat{x},\hat{y})(y-\hat{y}) &= 0 \end{aligned}$$

for x and y as a better approximation (or precise: next iterate).

More compact we can write

$$\mathrm{f}(\hat{\boldsymbol{x}}) + J(\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}}) = 0,$$

where $J(\boldsymbol{x})$ denotes the Jacobian of $\boldsymbol{f}$ given by

$$J(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f}{\partial x}(x,y) & \frac{\partial f}{\partial y}(x,y) \\ \frac{\partial g}{\partial x}(x,y) & \frac{\partial g}{\partial y}(x,y) \end{pmatrix}$$

Newton's method for systems of equations

Algorithm.

Input A function $\boldsymbol{f}(\boldsymbol{x})$, its Jacobian $J(\boldsymbol{x})$ and a starting value $\boldsymbol{x}^{(0)}$.

1. For $k = 0, 1, 2, \dots$
   - **1.1** Solve the linear system $J(\boldsymbol{x}^{(k)})\Delta^{(k)} = -\boldsymbol{f}(\boldsymbol{x}^{(k)})$
   - **1.2** Set $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \Delta^{(k)}$

Note. This can be generalized to n equations with n unknowns, where the Jacobian reads

$$J(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_1}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}) \\ \frac{\partial f_2}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_2}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_n}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\boldsymbol{x}) \end{pmatrix}$$

NTNU

Example.

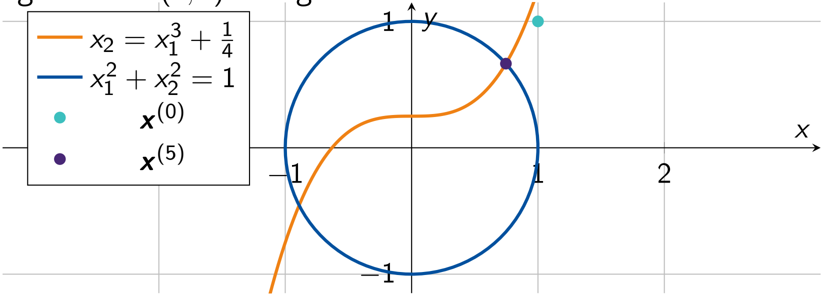
To solve our example from before

$$f(x,y) = x^3 - y + \frac{1}{4} = 0$$
$$g(x,y) = x^2 + y^2 - 1 = 0$$

we need its Jacobian

$$J(x,y) = \begin{pmatrix} 3x^2 & -1 \\ 2x & 2y \end{pmatrix}$$

Starting in $\boldsymbol{x}^{(0)} = (1,1)^{\mathrm{T}}$ we get



Further exercises for you to try

- Search for the solution of the Example in the third quadrant by changing the initial value $\boldsymbol{x}^{(0)}$.
- Apply Newton's method to the system

$$\begin{aligned} xe^{y} &= 1 \\ -x^2 + y &= 1 \end{aligned}$$

using $x^{(0)} = y^{(0)} = 0$.

Error analysis for multivariate Newton's method

Error and convergence analysis for the multivariate case is beyond the scope of this lecture.

Summary. If $\boldsymbol{f}$ is sufficiently differentiable, and there is a solution $\boldsymbol{x}^*$ of the system $\boldsymbol{f}(\boldsymbol{x}) = 0$ with $J(\boldsymbol{x}^*)$ nonsingular, then the Newton iterations will converge quadratically towards $\boldsymbol{x}^*$ for all $\boldsymbol{x}^{(0)}$ sufficiently close to $\boldsymbol{x}^*$.

Outlook. Multivariate Newton's method

- finding solutions in the multivariate case is hard
- especially choosing a good starting point $\boldsymbol{x}^{(0)}$ is
- systems of equations usually have multiple solutions
- how do you find the one you want?
- for large n: Evaluating the Jacobian $J(\boldsymbol{x}^{(k)})$ is expensive
- there exist efficient versions – including slow, but robust algorithms to find $\boldsymbol{x}^{(0)}$.