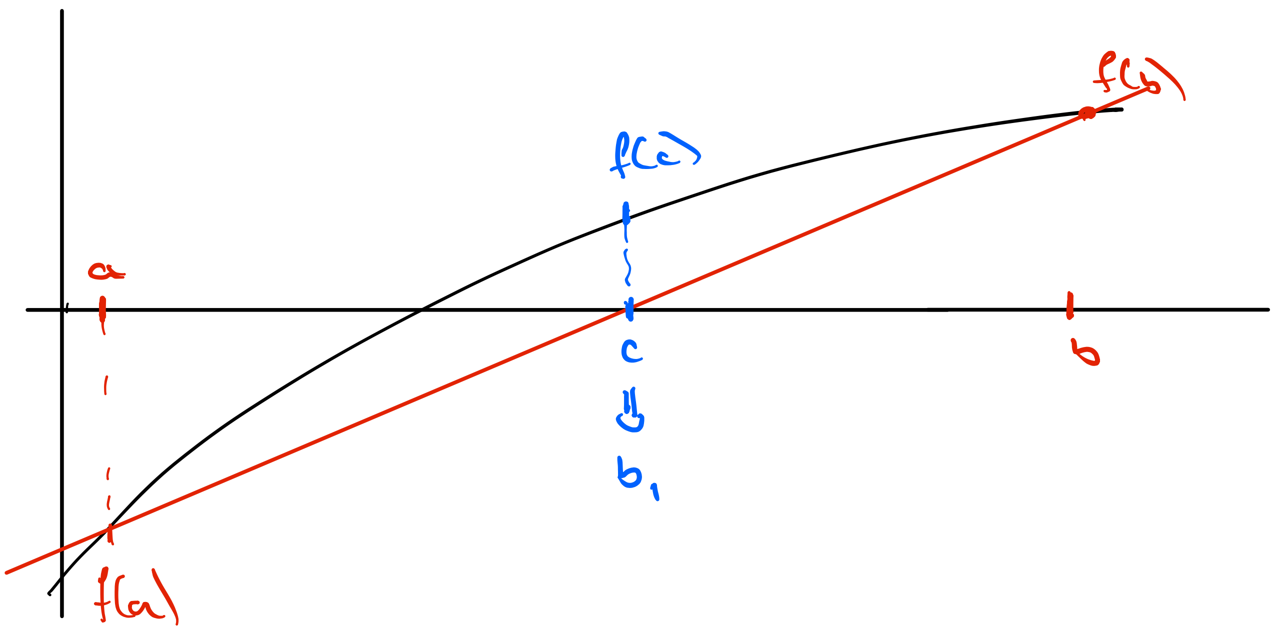
# Method of false position

```
zacetni podatki:  a, b
for k = 2, 3 ...
  c = b - f(b)(b - a)/(f(b) - f(a))
  if f(a)f(c) < 0
      b = c
  else
      a = c
  if je izpolnjen tolerancni pogoj, stop
end
```

c ⇒ secant method (for end points)

$$c = b - f(b)\left[\frac{b-a}{f(b)-f(a)}\right]$$



# Fixed point approach

$$\boxed{f(x)=0 \iff g(x)=x}$$

**Alg**: Choose $x_0$

$x_{k+1} = g(x_k)$ until convergence

**Thm**: If $g$ is continuously differentiable on $I=[a,b]$ & $g(I)\subseteq I$ (closed under $g$) with

$$\sup_{x\in I} |g'(x)| = m < 1$$

① $g(x)=x$ has a unique solution $\xi$ in $I$

② $\forall x_0 \in I$ the sequence $x_n = g(x_{n-1})$ converges to $\xi$ w/

$$|x_{n+1}-\xi| \le \min\left\{ \overset{①}{m^{n+1}|x_0-\xi|},\ \overset{②}{\frac{m}{m-1}|x_{n+1}-x_n|},\ \underset{③}{\frac{m^{n+1}}{1-m}|x_1-x_0|} \right\}$$

**Pf**:

$$|x_n-\xi| = |g(x_{n-1})-g(\xi)| \underset{\substack{\text{for some}\\ \zeta\in(\xi,x_{n-1})}}{=} |g'(\zeta)|\cdot|x_{n-1}-\xi| \le m|x_{n-1}-\xi|$$

MVT

①

$$\Rightarrow |x_n-\xi| \le m^n|x_0-\xi|$$

$$|x_{n+k+1}-x_{n+k}| = |g(x_{n+k})-g(x_{n-k-1})| \le m|x_{n+k}-x_{n+k-1}| \le m^k|x_{n+1}-x_n|$$

(2,3)

$$\Rightarrow |x_{n+1}-\xi| \le |x_{n+2}-x_{n+1}| + |x_{n+3}-x_{n+2}| + \cdots \le$$

$$\le (m+m^2+\ldots)|x_{n+1}-x_n| = \overset{②}{\frac{m}{1-m}|x_{n+1}-x_n|}$$

$$\le \frac{m^{n+1}}{1-m}(x_1-x_0) \quad ③$$

**Thm** (Speed of convergence)

In the neighborhood of $\xi$ assuming it is $p$-differentiable:

$$g'(\xi) = g''(\xi) = \dots = g^{(p-1)}(\xi) = 0$$
$$g^{(p)}(\xi) \neq 0$$

then order of convergence is $p$

**Pf**: Write Taylor series (since all derivatives to $p$ are zero

$$x_{n+1} = g(x_n) = \xi + \frac{1}{p!} g^{(p)}(\underset{\text{MVT}}{\xi})(x_n - \xi)^p$$

$$\frac{|x_{n+1} - \xi|}{|x_n - \xi|^p} = \frac{1}{p!} |g^{(p)}(\xi)|$$

fzero: MATLAB function

python: root_scalar, brentq

Julia: Roots.jl find_zero

# Systems of nonlinear equations

$$f_1(x_1, \dots x_n) = 0$$
$$f_2(x_1 \dots x_n) = 0$$
$$\vdots$$
$$f_n(x_1 \dots x_n) = 0$$

$$\underline{f} := (f_1 \dots f_n) : \mathbb{R}^n \to \mathbb{R}^n$$

We write $\underline{f}(\underline{x}) = 0$

Newton (tangent method)
Jacobi (fixed point)

---

## Newton iteration

$$\underline{x}^{(r+1)} = \underline{x}^{(r)} - J_{\underline{f}}(\underline{x}^{(r)})^{-1} \underline{f}(\underline{x}^{(r)})$$

$$J_{\underline{f}}(\underline{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_n} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} (\underline{x})$$

Jacobi matrix

in practice we don't compute the inverse but solve the system

$$J_{\underline{f}}(\underline{x}^{(r)})\,\Delta\underline{x}^{(r)} = -\underline{f}(\underline{x}^{(r)})$$

linear system

$$\underline{x}^{(r+1)} = \underline{x}^{(r)} + \Delta\underline{x}^{(r)}$$

## Derivation

① Write function via Taylor series

$$f_i(\underline{x}+\Delta\underline{x}) = f_i(\underline{x}) + \sum_{k=1}^{n}\frac{\partial f_i}{\partial x_k}(\underline{x})\,\Delta x_k + \dots$$

② Drop higher order terms & set

$$f_i(\underline{x}+\Delta\underline{x})=0$$

③ We get the above system

## Jacobi iteration

① $\underline{f}(x)=0 \iff g(\underline{x})=\underline{x}$

$$g: \mathbb{R}^n \to \mathbb{R}^n$$

② Choose $\underline{x}^{(0)} \in \mathbb{R}^n$ (initial guess)

③ $\underline{x}^{(r+1)} = g(\underline{x}^{(r)})$

1...n

**Thm** (Jacobi iteration)

Given $g: \mathbb{R}^n \to \mathbb{R}^n$ on $\Omega \subset \mathbb{R}^n$

① $g(\Omega) \subset \Omega$ (closed on the set)

② $\|g(\underline{x}) - g(\underline{y})\| \le m\|\underline{x}-\underline{y}\|$ $\forall \underline{x},\underline{y} \in \Omega$ for some $m \in [0,1]$ (contracting map)

Then $\underline{g}(\underline{x}) = \underline{x}$ has a unique solution $\xi$ on $\Omega$ & $\underline{x}^{(r+1)}$ converges to $\xi$ for any $\underline{x}^{(0)}$

Also $\| \underline{x}^{(r+1)} - \xi \| \leq \frac{m}{1-m} \| \underline{x}^{(1)} - \underline{x}^{(0)} \|$

<u>Thm</u> ($2^{nd}$ convergence theorem for Jacobi iteration)

Let $\underline{g} : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at the fixed point $\xi$ & let $\| J_g(\xi) \| \leq 1$ then there exists a closed neighborhood $\Omega \subseteq \mathbb{R}^n$ of $\xi$ so that $\underline{x}^{(r+1)} \to \xi$

<u>Note</u>: the condition is different

# Quasi-Newton methods

Broyden

- Newton: every step requires $O(n^2)$ partial derivatives & $O(n^3)$ operations for solving the linear system

- Broyden method: Like secant method avoids computing differentials

Let $B_r$ approx. $J_f(\underline{x}^{(r)})$

① Solve $B_r \Delta x^{(r)} = -f(x^{(r)})$

② $\underline{x}^{(r+1)} = x^{(r)} + \Delta x^{(r)}$

③ Determine $B_{r+1}$

Broyden: Satisfies

$$B_{r+1}(x^{(r+1)} - x^{(r)}) = f(x^{(r+1)}) - f(x^{(r)})$$

& minimizes $\|B_{r+1} - B_r\|_2$

So we are looking for

$$\Delta B_r = B_{r+1} - B_r \quad w/$$

$$\min \|\Delta B_r\|_2$$

$$\text{s.t. } \Delta B_r \Delta x^{(r)} = f(x^{(r+1)})$$

Claim: Jacobian matrix approx in Broyden method.

$$B_{r+1} = B_r + \frac{f(x^{(r+1)}) (\Delta x^{(r)})^T}{\|(\Delta x^{(r)})\|_2^2}$$

Pf: ① $B_r \Delta x^{(r)} = -f(x^{(r)})$

② $B_{r+1} \Delta x^{(r)} = f(x^{(r+1)}) - f(x^{(r)})$

Combine ① & ②

(0.1) $(B_{r+1} - B_r) \Delta x^{(r)} = f(x^{(r+1)})$

so $$B_{r+1} - B_r = \frac{f(x^{(r+1)}) (\Delta x^{(r)})^T}{\|\Delta x^{(r)}\|_2^2}$$

$\Rightarrow$ every matrix satisfying 0.1 $\Rightarrow$ must have at least this norm. $\sim$ $\dfrac{\|f(x^{(r+1)})\|_2}{\|\Delta x^{(r)}\|_2}$

# Variational methods

If we have a good initial approximation
⇝ we can look for a local minimum

**Claim**

$f(x) = 0$

$g : \mathbb{R}^n \to \mathbb{R}$ $\quad g(x) = \|f(x)\|^2 = \sum_i f_i^2(x)$

**Q.** if $g$ is $C^2$ how can we find extremal points?

## Finding local extremal points

Minimum of a function can be found iteratively in some direction $v_r$

$$x^{(r+1)} = x^{(r)} + \lambda_r v_r$$

where $\lambda_r$ is some real number

$$g(x^{(r+1)}) < g(x^{(r)})$$

General descent: any direction not orthogonal to $\nabla g$

Fastest descent: $-\nabla g = v_r$

Coordinate descent: $e_1, e_2, e_3 \ldots e_n$

For solving step size

$$q(\lambda) = g(\underline{x}^{(r)} + \lambda v_r)$$

① gradient descent $q'(\lambda) = 0$

② tangent descent find intersection tangent to $y = q(\lambda)$ at $\lambda = 0$ w/ x-axis

③ Quadratic parabolic descent

Use tangent to determine $\alpha$

Fit a parabola to $(0, q(0))$, $(\alpha/2, q(\alpha/2))$, $(\alpha, q(\alpha))$ & take the minimum

Using methods for finding roots for finding extremal points

<u>Claim:</u> Local extremal points of $g(x)$ solve the system

$$\nabla g(x) = \left( \frac{\partial g(x)}{\partial x_1}, \ \dots \ \frac{\partial g(x)}{\partial x_n} \right) = 0$$

Type of extremal point - depends on Hessian matrix

$$H_g(x) = \begin{bmatrix} \frac{\partial^2 g(x)}{\partial x_1^2} & \dots & \frac{\partial^2 g(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g(x)}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 g(x)}{\partial x_n^2} \end{bmatrix}$$

if $H_g(x)$ is positive definite $\Rightarrow$ it is a local minimum
negative definite $\Rightarrow$ maximum

<u>Gauss - Newton iteration</u>

Solving overdetermined system

$$f: \mathbb{R}^n \to \mathbb{R}^m \quad f(x) = (0, \dots, 0)$$

$m$ non linear equations in $n$ unknowns
$m > n$

Since generally no solution exists, find solution s.t.

$$\|f(\alpha)\|^2 = \min \{ \|f(x)\|^2 \}$$

# Gauss - Newton

$$\underline{x}^{(r+1)} = \underline{x}^{(r)} - \underbrace{J_r(\underline{x}^{(r)})^+}_{} f(x^{(r)})$$

pseudo inverse
rather than inverse

At each step solve
$x^{(r+1)}$ is the "solution" of

$$J_r(x^{(r)})\,\Delta x^{(r)} = f(x^{(r)})$$

no convergence $\Rightarrow$ even when there is, it is only local convergence

<u>Example</u>:

$(x_i, y_i) \in \mathbb{R}^2 \quad i = 1, \dots, m$

Find a function $f(x, a, b) = ae^{bx}$

$F : \mathbb{R}^2 \to \mathbb{R}^m$

$$F(a,b) = (y_1 - ae^{bx_1}, \dots y_n - ae^{bx_n})$$

$$DF(a,b) = \begin{bmatrix} -e^{bx_1} & ax_1e^{bx_1} \\ \vdots & \vdots \\ -e^{bx_n} & ax_ne^{bx_n} \end{bmatrix}$$

$$\begin{bmatrix} a_{r+1} \\ b_{r+1} \end{bmatrix} = \begin{bmatrix} a_r \\ b_r \end{bmatrix} - DF^+(a_r, b_r)\, F(a_r, b_r)^T$$

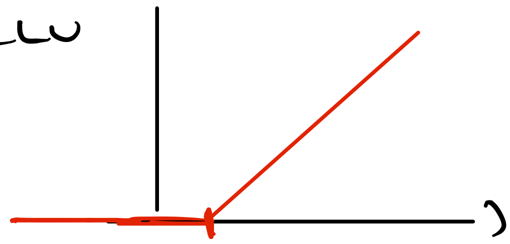
What about non-differentiable functions?

① Automatic differentiation - backprop

② Almost differentiable

a) - max (maxpooling) $\max(\underline{x})$

b) - ReLU




②

b) Just use the fact that we never hit the corner point (always evaluate $f'(\underline{x}_k)$)

a) Say multiple entries are equal

$$y = (x_1, \dots, x_n)$$

$$\nabla y = (0 .. 0 1 ... 0) \quad \text{if unique}$$

otherwise

$$\nabla y = (0 .. 0 \tfrac{1}{k} .. 0 .. \tfrac{1}{k}) \quad \tfrac{1}{k} \text{ for } k \text{ max values}$$

or choose randomly

$\Rightarrow$ practical solution

① Idea (will cover later)

$\Rightarrow$ write derivatives of basic function

$\Rightarrow$ use chain rule to construct derivative

$$\frac{\partial}{\partial x} f(g(x)) = f'(g(x))\, g'(x)$$

forward $\quad x \mapsto (x, 1) \qquad \sin(x) \mapsto (\sin(x), \cos(x))$

$$x^2 \mapsto (x^2, 2x)$$

backward

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial x}$$

Ex $\quad f(x) = (x^2 + 3x) \sin x$

$a = x^2$

$b = 3x$

$c = a + b$

$d = \sin x$

$e = c \cdot d$

$x \mapsto (x, 1)$

$a: \; x^2 \mapsto (x^2, 2x)$

$b: \; b \mapsto (3x, 3)$

$d: \; \sin x \mapsto (\sin x, \cos x)$

$c \mapsto (a + b, a' + b')$

$(x^2 + 3x, 2x + 3)$

$e: \; c \cdot d \mapsto (f(x), c'd + d'c)$

Dual numbers $\Rightarrow$ usually at a point

$$x = a + c\varepsilon \Rightarrow x = a + f'(x)\varepsilon$$

Start at $(y, \varepsilon) \to f(y) = (f(y), f'(y)\varepsilon)$

Multiply 2 dual numbers

$$(a+b\varepsilon)(c+d\varepsilon) = \underbrace{ac} + \underbrace{(ad+bc)}\varepsilon + \underbrace{bd\varepsilon^2}_{\text{set to } 0}$$