Prednáška 2

Optimalizačné metódy pre funkcie
1-premennej

Študujme **reálnu funkciu** 1-premennej

$$f : R \to R$$

Našou úlohou bude riešiť nasledujúcu optimalizačnú úlohu

$$x_{opt} = \arg\min \, f(x)$$

Táto úloha môže byť riešená mnohými metódami, v tejto prednáške si ukážeme niektoré z týchto metód

1. Newtonova (a Raphsonova) metóda

Pôvodná verzia tejto metódy bola navrhnutá pre riešenie rovnice $f(x)=0$, t.j. hľadanie koreňov tejto rovnice. Bude použitá pre riešenie rovnice $f'(x)=0$.

(a) **Jednoduché odvodenie Newtonovej metódy**. Nech riešenie $f(x)=0$ je vyjadrené v tvare $x=x_o+\delta$, dosadením do pôvodnej rovnice dostaneme $f(x_o+\delta)=0$. Aplikovaním Taylorovho rozvoja dostaneme

$$f(x_0)+f'(x_0)\delta+\ldots=0$$

Riešením tejto rovnice dostaneme približný výraz pre "výchylku" δ

$$\delta\approx-\frac{f(x_o)}{f'(x_o)}$$

Potom riešenie x vyjadríme takto

$$x\approx x_o-\frac{f(x_o)}{f'(x_o)}$$

Túto formulu môžeme interpretovať tak, že jej použitím zostrojíme nové riešenie x_1, toto riešenie budeme považovať za vstupné, čiže pomocou neho zostrojíme nové riešenie x_2, atď. Túto skutočnosť vyjadríme pomocou rekurentnej formule

$$\boxed{x_{k+1}=x_k-\frac{f(x_k)}{f'(x_k)}\quad(\text{pre } k=0,1,2,\ldots)}$$

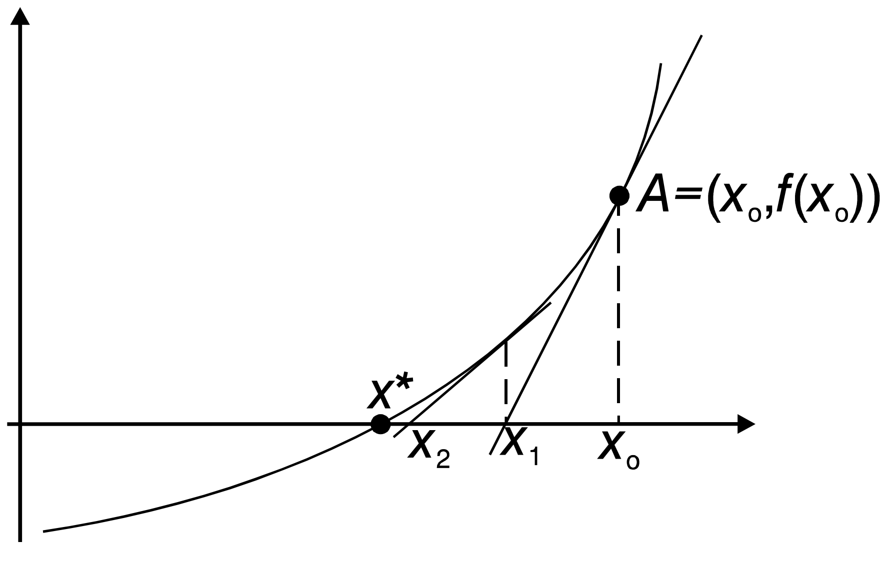
Opakovaným riešením tejto rovnice dostaneme postupnosť bodov

$$\left\{x_k\right\}_{k=1}^{\infty}$$

Budeme hovoriť, že Newtonova metóda **konverguje**, ak postupnosť je konvergentná a jej limita je riešením rovnice $f(x)=0$

$$\lim_{k\to\infty} x_k = x^*, \quad \text{kde } f\left(x^*\right)=0$$

(b) **Geometrická interpretácia Newtonovej metódy**. Newtonová metóda má jednoduchú geometrickú interpretáciu pomocou postupnosti dotyčníc ku grafu funkcie $f(x)$



Dotyčnica v bode x_o je určená rovnicou

$$\frac{y - f(x_o)}{x - x_o} = f'(x_o)$$

Jednoduchými úpravami dostaneme

$$y = f(x_o) + f'(x_o)(x - x_o)$$

Priesečník x_1 dotyčnice s osou x je určený podmienkou y=0

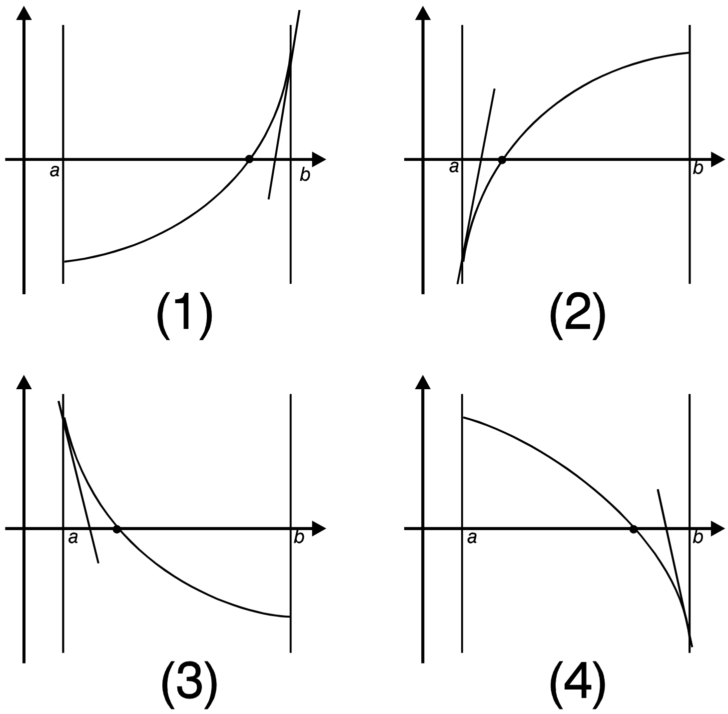
$$x_1 = x_o - \frac{f(x_o)}{f'(x_o)}$$

Táto formula je totožná s rekurentnou formulou odvodenou pomocou Taylorovho rozvoja. Pomocou geometrickej interpretácie ľahko "dokážeme" nasledujúcu vetu o konvergencii Newtonovej metódy.

Veta. Nech funkcia $f(x)$ je definovaná na intervale $<a,b>$, pričom na tomto intervale je 2-krát diferencovateľná a platí $f(a)f(b)<0$. Newtonova metóda je **konvergentná** ak platí jedna z nasledujúcich 4 podmienok

(1) $f(a)<f(b) \wedge f''(x)>0 \wedge x_o=b$, alebo
(2) $f(a)<f(b) \wedge f''(x)<0 \wedge x_o=a$, alebo
(3) $f(a)>f(b) \wedge f''(x)>0 \wedge x_o=a$, alebo
(4) $f(a)>f(b) \wedge f''(x)<0 \wedge x_o=a$.

Jednotlivé podmienky majú túto geometrickú interpretáciu



2. Modifikácia Newtonovej metódy k hľadaniu minima funkcie

Newtonová metóda bude použitá k riešeniu rovnice $f'(x)=0$, t.j. budeme hľadať stacionárne body funkcie $f(x)$. Rekurentná formula má potom tvar

$$\boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad (\text{pre } k = 0,1,2,\ldots)}$$

Za predpokladu, že iteračné riešenie tejto formule konverguje, potom postupnosť bodov

$$\{x_k\}_{k=1}^{\infty}$$

konverguje k stacionárnemu bodu x^*, kde $f'(x^*)=0$. Znamienko druhej derivácie $f''(x^*)$ bude rozhodovať o tom, či sa jedná o minimum alebo o maximum. Iteračné riešenie sa končí ak pre existuje také K, že pre $\forall k \geq K$ platí

$$|x_{k+1} - x_k| < \varepsilon$$

kde ε je dané kladné malé číslo.

Algoritmus Newtonovej metódy k nájdeniu stacionárneho bodu funkcie

```
read(x,ε,kmax);
norm:=∞; k:=0;
while (norm>ε) and (k<kmax) do
begin k:=k+1;
      x':=x-f'(x)/f''(x);
      norm:=abs(x-x');
      x:=x';
end;
write(x,f(x),f'(x),f''(x));
```

3. Numerický výpočet derivácií funkcie

Z definície prvej derivácie môžeme zostrojiť nasledujúce dva približné výrazy pre výpočet prvej derivácie funkcie $f(x)$ v bode x_o

$$f'(x_o) \approx \frac{f(x_o + h) - f(x_o)}{h} + o_+(h)$$

$$f'(x_o) \approx \frac{f(x_o) - f(x_o - h)}{h} + o_-(h)$$

kde h je malé kladné číslo a symbol $o(h)$ znamená, že chyba výpočtu je prvého rádu (vzhľadom k h). Chyba výpočtu o(h) môže byť jednoducho určená pomocou Tazlorovho rozvoja

$$o_+(h) = \frac{1}{2} f''(x_o) h + \frac{1}{6} f'''(x_o) h^2 + \ldots$$

$$o_-(h) = -\frac{1}{2} f''(x_o) h + \frac{1}{6} f'''(x_o) h^2 + \ldots$$

Aritmetickým priemerom dvoch formulí pre prvú deriváciu dostaneme

$$f'(x_o) \approx \frac{f(x_o+h) - f(x_o-h)}{2h} + o(h^2)$$

kde kvadrarická chyba $o(h)$ je určená pomocou tretej derivácie $f(x)$ v bode x_o

$$o(h^2) = \frac{1}{6} f'''(x_o) h^2 + \ldots$$

Takto zostrojený výraz pre numerický výpočet derivácie je vhodnejší, ako predchádzajúci,, pretože jeho presnosť je kvadratická na rozdiel od lineárnej presnosti pôvodného výrazu.

Výraz pre numerický výpočet druhej derivácie môže byť zostrojený ako derivácia prvej derivácie

$$f''(x_o) \approx \frac{f(x_o-h) - 2f(x_o) + f(x_o+h)}{h^2} + o(h^2)$$

4. Metóda kvadratickej interpolácie

Táto metóda patrí medzi tie prístupy k optimalizácii funkcií 1-premennej, ktoré nepožadujú výpočet derivácie, sú založené len na výpočte hodnoty funkcie.
Majme tri body x_1, x_2 a x_3, funkčné hodnoty $f(x)$ v týchto bodoch sú

$$F_1=f(x_1),\ F_2=f(x_2) \text{ a } F_3=f(x_3)$$

Nech

$$\varphi(x) = ax^2 + bx + c$$

je kvadratická parabola, ktorej koeficienty a, b a c sú určené tak, aby jej funkčné hodnoty v bodoch x_1, x_2 a x_3 boli totožné s funkčnými hodnotami funkcie $f(x)$ v týchto bodoch

$$\varphi(x_1) = ax_1^2 + bx_1 + c = F_1$$
$$\varphi(x_2) = ax_2^2 + bx_2 + c = F_2$$
$$\varphi(x_3) = ax_3^2 + bx_3 + c = F_3$$

to znamená, že koeficienty paraboly sú určené systémom lineárnych rovníc, postupným riešením dostaneme

$$a = \frac{\begin{vmatrix} F_1 & x_1 & 1 \\ F_2 & x_2 & 1 \\ F_3 & x_3 & 1 \end{vmatrix}}{\begin{vmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \end{vmatrix}} =$$

$$= \frac{F_1\left(x_2 - x_3\right) + F_2\left(x_3 - x_1\right) + F_3\left(x_1 - x_2\right)}{x_2 x_3\left(x_2 - x_3\right) + x_1 x_3\left(x_3 - x_1\right) + x_1 x_2\left(x_1 - x_2\right)}$$

Ak poznáme hodnotu koeficieta a, potom prvé dve rovnice systému môžeme upraviť do tvaru v ktorom sú neznáme koeficienty b a c

$$bx_1 + c = F_1 - ax_1^2$$

$$bx_2 + c = F_2 - ax_2^2$$

Riešením tohto systému dostaneme koeficient b

$$b = \frac{\begin{vmatrix} F_1 - ax_1^2 & 1 \\ F_2 - ax_2^2 & 1 \end{vmatrix}}{\begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix}} = \frac{F_2 - F_1 - a\left(x_2^2 - x_1^2\right)}{x_2 - x_1}$$

Konečne, poznajúc koeficienta a a b , koefient c je určený prvou rovnicou takto

$$c = F_1 - ax_1^2 - bx_1$$

Týmto spôsobom máme plne určenú parabolickú funkciu $\varphi(x)$, táto funkcia má stacinárny bod určený podmienkou $\varphi'(x)=0$

$$x_4 = -\frac{b}{2a}$$

Charakter tohto stacionárneho bodu je určený znamienkom 2. derivácie

$$\varphi''(x) = 2a$$

To znamená, že ak a>0, potom parabola $\varphi(x)$ má v bode x_4 , ak a<0, potom parabola má v tomto bode maximum.

Optimalizačná metóda kvadratickej interpolá-cie je založená na tom, že postupne (rekurentne) obnovujeme body x_1, x_2, x_3 a x_4 podľa schémy

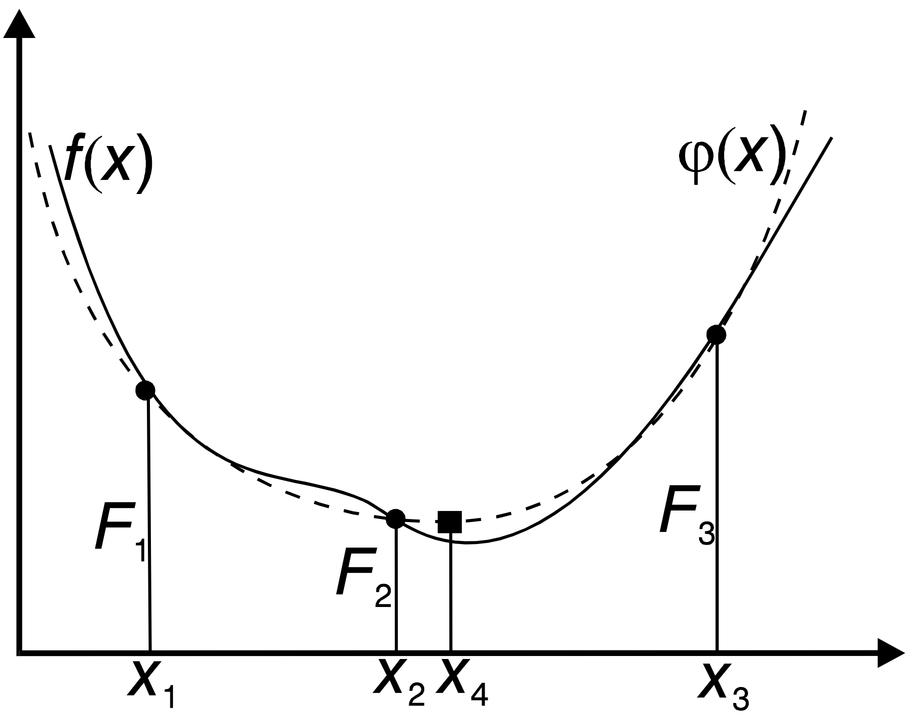
$$x_1 \leftarrow x_2, x_2 \leftarrow x_3, x_3 \leftarrow x_4 \text{ a } x_4 \leftarrow -\frac{b}{2a}$$

Algoritmus je ukončený, ak platí napr.

$$|x_3 - x_4| < \varepsilon$$

kde ε je malé kladné číslo (presnosť).

Geometrická interpretácia optimalizačnej metódy kvadratickej interpolácie



Algoritmus metódy kvadratickej interpolácie k nájdeniu stacionárneho bodu funkcie

```
read(x1,x2,ε,kmax);
x3:=0.5*(x1+x2);
norm:=∞; k:=0;
F1:=f(x1); F2:=f(x2); F3:=f(x3);
while (k<kmax) and (norm>ε) do
begin k:=k+1;
      calculate a and b;
      x4:=-b/(2*a); F4:=f(x4);
      norm:=abs(x3-x4);
      x1:=x2; x2:=x3; x3:=x4;
      F1:=F2; F2:=F3; F3:=F4;
end;
write(x4,F4,a);
```

Quickpropagation Algorithm for Adaptation Process of Feed-Forward Neural Networks

Basic reference:

Scott E. Fahlman: *An Empirical Study of Learning Speed in Back-Propagation Networks*. School of Computer Science, Carnegie Mellon University, Pittsburgh, September 1988. Technical Report CMU-CS-88-162.

Note: In many textbooks on neural networks this algorithm is discussed as a simple alternative to the commonly used steepest-descent algorithm accelerated by momentum term.

The manuscript of this famous paper is still available on anonymous FTP server:

ftp://ftp.funet.fi/pub/sci/neural/
neuroprose/fahlman.quickprop-tr.ps.Z

Simple derivation of quickpropagation algorithm

Two independent derivations are presented:

(1) Simplification of Newton's method
(2) An optimization of quadratic function

1. Simplification of Newton's optimization method

Let is consider a function $f{:}R \to R$, which has the first and second derivatives for all $x \in R$. A solution of optimization problem can be found by an iterative application of the recurrent formula

$$\boxed{x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}}$$

A solution $x_{opt} = \lim_{k\to\infty} x_k$ (postulating that it exists) is a local minimum if $f'(x_{opt})=0$ and $f''(x_{opt})>0$.

The second derivative $f''(x)$ may be approximated by

$$f''(x_k) \approx \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}$$

An introduction of this approximate formula for the evaluation of the second derivative into Newton's recurrent formula gives

$$\boxed{x_{k+1} = x_k + \frac{f'(x_k)}{f'(x_{k-1}) - f'(x_k)}(x_k - x_{k-1})}$$

This recurrent formula can be used in a similar way as Newton's formula for looking for local extremes of functions that are at least one-times differentiable.

2. An optimization of quadratic function

Let us consider a quadratic function

$$f(x) = ax^2 + bx + c$$

where *a,b,c* are constants. Its first derivative is simply determined by

$$f'(x) = 2ax + b$$

For two different points we have prescribed values of first derivatives

$$2ax_1 + b = f'(x_1)$$
$$2ax_2 + b = f'(x_2)$$

This means that constants *a* and *b* are determined by above system of linear equations, we get

$$a = \frac{f'(x_2) - f'(x_1)}{2(x_2 - x_1)}, \quad b = \frac{x_2 f'(x_1) - x_1 f'(x_2)}{x_2 - x_1}$$

A global minimum of the above quadratic function is determined by

$$f'\left(x_{opt}\right)=0 \Rightarrow x_{opt}=-\frac{b}{2a}=\frac{x_1 f'(x_2)-x_2 f'(x_1)}{f'(x_2)-f'(x_1)}$$

After simple algebraic manipulations it can be rewritten in a form as follows

$$x_3=x_{opt}=x_2+\frac{f'(x_2)}{f'(x_1)-f'(x_2)}(x_2-x_1)$$

We see that for quadratic functions the used approach gives immediately after one step correct solutions, for other than quadratic functions this formula can be generalized to the following recurrent form

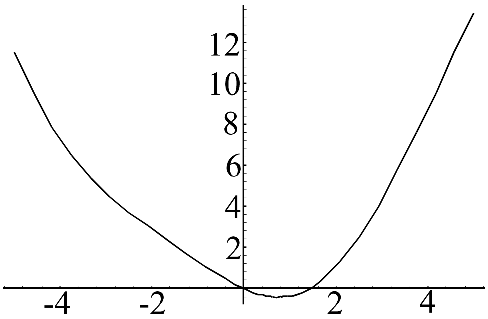
$$\boxed{x_{k+1}=x_k+\frac{f'(x_k)}{f'(x_{k-1})-f'(x_k)}(x_k-x_{k-1})}$$

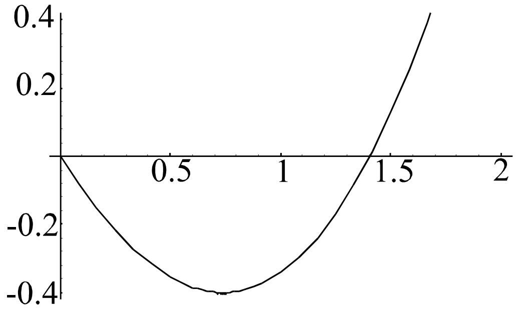
The derived recurrent formula is fully identical to those one derived by a simple modification of the standard Newton's recurrent formula.

Note: The method is finished if a difference between two last solutions is smaller than a prescribed precision ε.

Illustrative example

$$f(x) = \frac{1}{2}x^2 - \sin x, \quad x_{opt} = 0.73908$$





Quickpropagation method:

iteration # 1 x_1= 1.000000 x_2= 2.000000 x_3= 0.765035
iteration # 2 x_1= 2.000000 x_2= 0.765035 x_3= 0.742299
iteration # 3 x_1= 0.765035 x_2= 0.742299 x_3= 0.739103
iteration # 4 x_1= 0.742299 x_2= 0.739103 x_3= 0.739085
iteration # 5 x_1= 0.739103 x_2= 0.739085 x_3= 0.739085

Newton's method:

iteration # 1 x_1= 1.000000 x_2= 0.750364
iteration # 2 x_1= 0.750364 x_2= 0.739113
iteration # 3 x_1= 0.739113 x_2= 0.739085
iteration # 4 x_1= 0.739085 x_2= 0.739085