2. Přednáška: Line search

Úloha:

$$\min_{x \in \mathbb{R}^n} f(x),$$

kde $x \in \mathbb{R}^n$, $n \geq 1$ a $f : \mathbb{R}^n \to \mathbb{R}$ je dvakrát spojitě diferencovatelná.

Iterační algoritmy:

- Začínám v x_0 a vytvářím posloupnost iterací $\{x_k\}_{k=1}^{\infty}$ tak, aby minimum bylo aproximováno s dostatečnou přesností, ideálně končila v přesném řešení.
- K posunutí z x_k do x_{k+1}, aby $f(x_{k+1}) < f(x_k)$ mohu použít informace o $x_0, \ldots, x_{k-1}$ a jejich funkčních hodnotách.
- 2 základní strategie, jak se posunout z x_k do x_{k+1}: Line search, trust region.

Line search algoritmus:

Dáno x_0, interval (a, b), ε atd. a pro $k = 0, 1, 2, \ldots$:

1) najdi směr $p_k \in \mathbb{R}^n$ - spádový směr,
2) najdi α_k tak, aby $x_k + \alpha_k p_k$ bylo "lepší" než x_k (ideálně minimální - exaktní line search, dostatečně malé - aproximace),
3) polož $x_{k+1} = x_k + \alpha_k p_k$.

Priklad3.m

ad 1) viz další přednášky,
ad 2) hledání minima jednodimenzionální funkce $\phi(\alpha) = f(x_k + \alpha p_k)$, tj. řeším úlohu:

$$\min_{\alpha > 0} \phi(\alpha).$$

Jednodimenzionální minimalizace:

Definice: Unimodální funkce
Funkce $\phi : \mathbb{R} \to \mathbb{R}$ *je unimodální na intervalu* $\langle a, b \rangle$, *jestliže existuje* $c \in \langle a, b \rangle$ *takové, že* ϕ *je klesající na* $\langle a, c \rangle$ *a rostoucí na* $\langle c, b \rangle$.

Věta: Pozice minima
Nechť ϕ *je unimodální v* $\langle a, b \rangle$ *a nabývá svého minima v* $\alpha^* \in \langle a, b \rangle$. *Pak pro libovolné* x *a* y, *kde* $a \leq x < y \leq b$ *platí následující tvrzení:*

- *Je-li* $\phi(x) < \phi(y)$, *pak* $\alpha^* < y$.
- *Je-li* $\phi(x) > \phi(y)$, *pak* $\alpha^* > x$.
- *Je-li* $\phi(x) = \phi(y)$, *pak* $x < \alpha^* < y$.

Algoritmus:

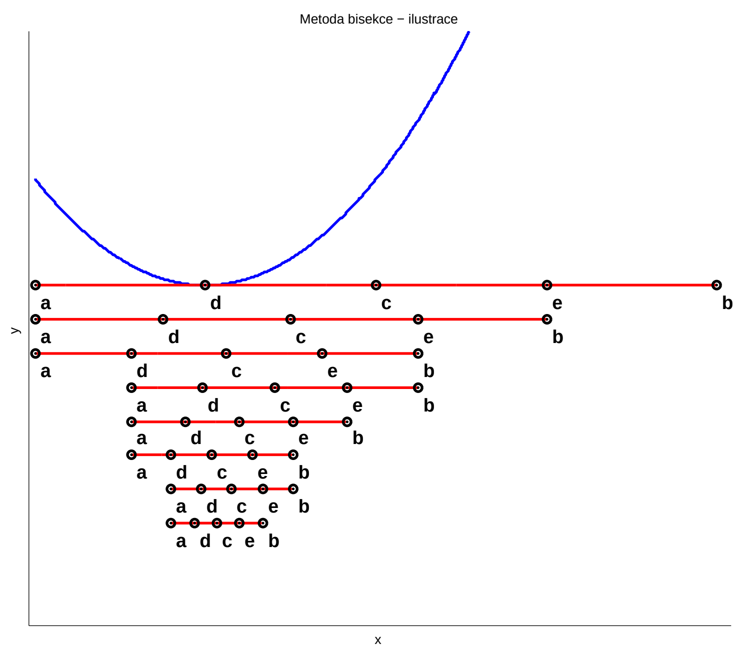
Za předpokladu unimodální funkce ϕ a pomocí výše uvedené věty generuji posloupnost zmenšujících se intervalů obsahujících řešení α^* (intervaly neurčitosti):

$$\langle a, b \rangle = I_0 \supset I_1 \supset \ldots I_{k-1} \supset I_k = \langle a_k, b_k \rangle,$$

$\Delta_k \equiv b_k - a_k$ značí délku k-tého intervalu. Požaduji co nejmenší vyhodnocování funkce ϕ, co největší redukci intervalů, jednoduchost.

Metody:

- **Metoda bisekce:**
  - dělení intervalu (a, b) na poloviny,
  - jednoduché, drahé - hodně vyhodnocování ϕ,
  - více možností,
  - Priklady3_bisekce.m



- **Fibonacciho metoda:**
  - Fibonacciho posloupnost: $F_0 = F_1 = 1$, $F_k = F_{k-1} + F_{k-2}$, $k = 2.\dots, N$

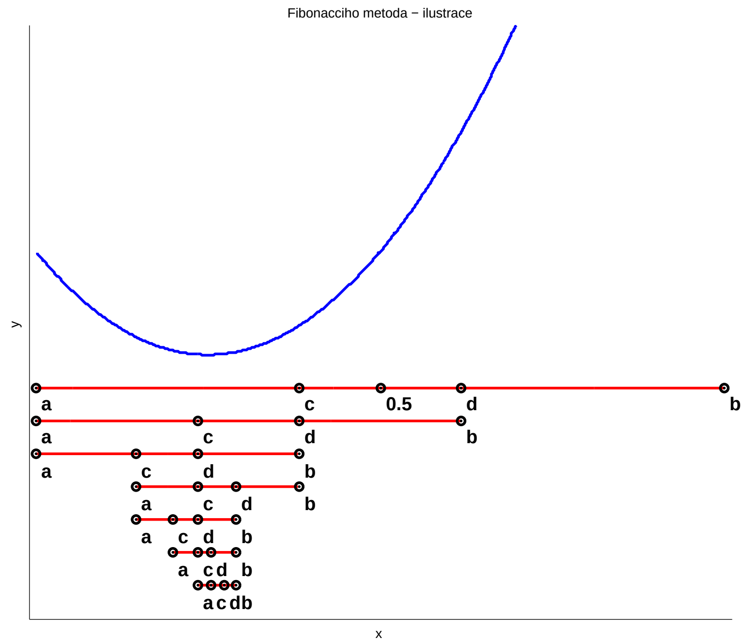
$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \dots$$

  - Priklady3_fibonacci.m
  - optimální, dělení intervalu (a, b) v poměru fibonacciho čísel $\frac{F_{N-1}}{F_N} > \frac{1}{2}$,
  - když chceme zpřesnit, musíme začít od začátku (nová volba N),
  -

$$\Delta_0 = b_0 - a_0,$$
$$c_0 = b_0 - \frac{F_{N-1}}{F_N}\Delta_0,$$
$$d_0 = a_0 + \frac{F_{N-1}}{F_N}\Delta_0,$$
$$c_0 = a_0 + \frac{F_{N-2}}{F_N}\Delta_0,$$
$$\Delta_1 = b_1 - a_1 = \frac{F_{N-1}}{F_N}\Delta_0,$$
$$c_1 = b_1 - \frac{F_{N-2}}{F_N}\Delta_1,$$
$$d_1 = a_1 + \frac{F_{N-2}}{F_N}\Delta_1,$$

  - nový bod je umístěn v daném intervalu symetricky s již existujícím bodem

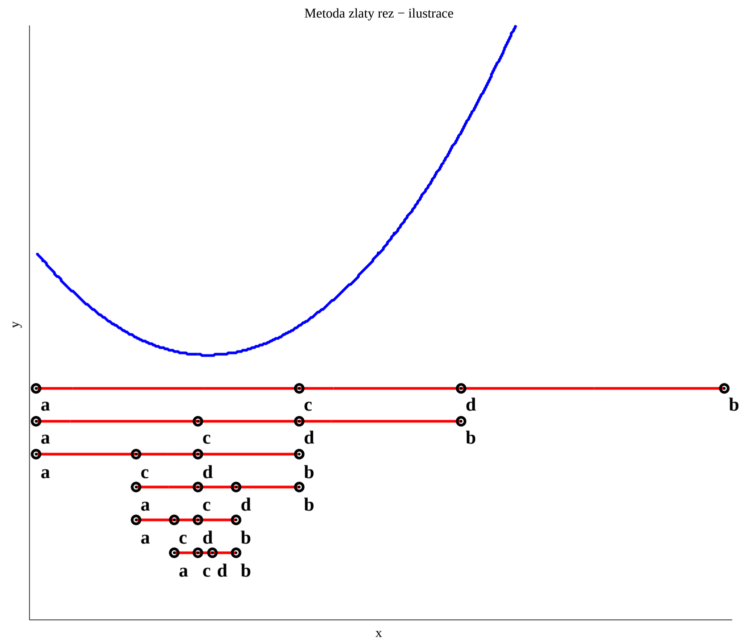
$$\Delta_k = \frac{F_{N-k}}{F_N}\Delta_0.$$



- **Metoda zlatého řezu:**
  - optimální,
  - zafixujeme zkracování intervalu na jedné hodnotě:

$$\xi = \lim_{N \to \infty} \frac{F_{N-1}}{F_N},$$

    platí $\xi = \frac{\sqrt{5}-1}{2}$ - zlatý řez,
  - $\Delta_k = \xi^k \Delta_0$,
  - Priklady3_zlatyrez.m



- **Newtonova metoda:**
  - Nechť $\phi(x)$ je hladká funkce, unimodální na intervalu (a, b).
  - Aproximujeme $\phi(x)$ pomocí kvadratické funkce $q(x)$, která se bude shodovat v následujících parametrech:

$$\phi(x_k) = q(x_k), \quad \phi'(x_k) = q'(x_k), \quad \phi''(x_k) = q''(x_k)$$

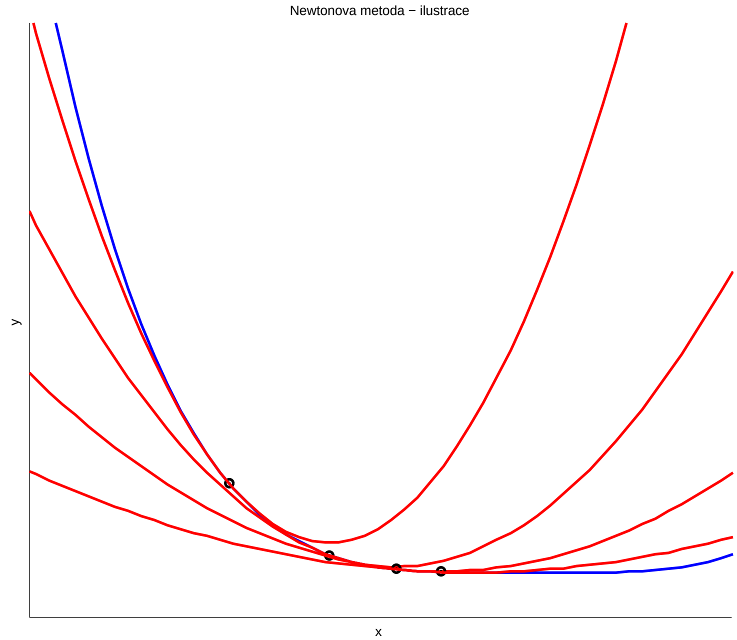
    $\Rightarrow$ Taylorův polynom 2. řádu

$$\phi(x) \approx q(x) = \phi(x_k) + \phi'(x_k)(x - x_k) + \frac{\phi''(x_k)}{2}(x - x_k)^2$$

  - x_{k+1} je bod minima funkce $q(x)$.
  - Nutná podmínka extrému:

$$q'(x) = \phi'(x_k) + \phi''(x_k)(x - x_k) = 0,$$

    Stacionární bod: $x = x_k - \frac{\phi'(x_k)}{\phi''(x_k)}$.
  - Protože jde o kvadratickou funkci, je stacionární bod i lokálním minimem (zároveň i globálním minimem)
    $\Rightarrow$ iterační předpis $x_{k+1} = x_k - \frac{\phi'(x_k)}{\phi''(x_k)}$.
  - Priklady3_newton.m



- **Interpolace pomocí kvadratické funkce:**
  - trojice bodů $[x_1, \phi(x_1)]$, $[x_2, \phi(x_2)]$, $[x_3, \phi(x_3)]$
  - interpolace kvadratickou funkcí (např. Lagrangeův interpolační polynom), nepotřebuji derivace:

$$\phi(x) \approx q(x) = \phi(x_1)\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + \phi(x_2)\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + \phi(x_3)\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}$$

  - x_4 je bodem minima funkce $q(x)$.
  - Vyberu novou trojici bodů porovnáním funkčních hodnot $\phi(x_1)$, $\phi(x_2)$, $\phi(x_3)$, $\phi(x_4) \Rightarrow (x_2, x_4, x_3)$ nebo (x_1, x_2, x_4).

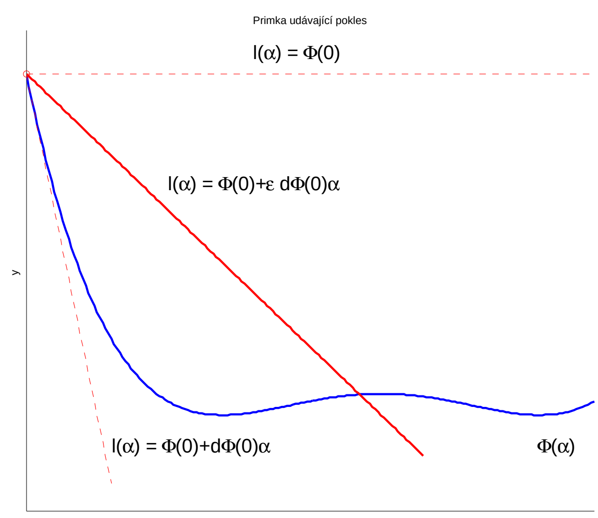
- **Testování dostatečného poklesu:**
  - Dilema přesnost $\times$ výpočetní čas.
  - Použití minimalizační metody (viz výše), ale zastavím hledání řešení, pokud interval neurčitosti splňuje podmínku:

    $$|\alpha^* - \bar{\alpha}| \leq c\bar{\alpha},$$

    kde α^* je přesné řešení, $\bar{\alpha}$ je aktuální řešení, $|\alpha^* - \bar{\alpha}|$ délka intervalu neurčitosti.
  - Spokojím se s dostatečným poklesem.

Přímka udávající pokles:

- x_k, p_k spádový směr $\phi'(0) = \mathrm{grad}^T f(x_k) p_k < 0$.
- Chci zvolit α tak, aby $\phi(\alpha) \equiv f(x_k + \alpha p_k)$ dostatečně pokleslo a α nebylo ani malé ani velké.
- Přímka udávající pokles: pro dané $\varepsilon \in (0,1)$: $l(\alpha) \equiv \phi(0) + \varepsilon\phi'(0)\alpha$.
- Pro všechna α taková, že $\phi(\alpha) < l(\alpha)$ platí $f(x_k + \alpha p_k) = \phi(\alpha) < \phi(0) = f(x_k)$.



Armijův test (podmínka):

- Pokud bude α malé i pokles může být malý.
- Volba tak, aby například dvojnásobné (desetinásobné) α už nesplnilo podmínku $\phi(\alpha) < l(\alpha)$, tj. zvol η ($\eta = 2$) a najdi α tak, aby

$$\phi(\alpha) \leq \phi(0) + \varepsilon\phi'(0)\alpha \tag{1}$$

$$\wedge$$

$$\phi(\eta\alpha) > \phi(0) + \varepsilon\phi'(0)\alpha. \tag{2}$$

- Pseudo-algoritmus:
  - Start: volím ε (v praxi volíme malé, 10^{-4}), η, $\alpha\dots$
  - Pokud je splněna (1), prováděj $\alpha = \eta\alpha$ dokud není splněna (2).
  - Pokud není splněna (1), prováděj $\alpha = \frac{\alpha}{\eta}$, dokud není splněna (1).
- Interval přípustných hodnot α nemusí obsahovat minimum funkce $\phi(\alpha)$.
- Priklad3_pokles.m

Goldsteinův test (podmínka):

- Zvolím $\varepsilon \in \left(0, \frac{1}{2}\right)$ tak, aby
$$\phi(\alpha) \leq \phi(0) + \varepsilon\phi'(0)\alpha. \tag{3}$$
- Aby α nebylo příliš malé, musí ležet nad jinou přímkou, tj.
$$\phi(\alpha) \geq \phi(0) + (1-\varepsilon)\phi'(0)\alpha. \tag{4}$$
- Interval přípustných hodnot α nemusí obsahovat minimum funkce $\phi(\alpha)$.
- Priklad3_goldstein.m

Wolfeho test (podmínka):

- Zvolím $\varepsilon \in \left(0, \frac{1}{2}\right)$ tak, aby
$$\phi(\alpha) \leq \phi(0) + \varepsilon\phi'(0)\alpha. \tag{5}$$
- Dám požadavek na sklon: pokud jsem blízko minima $\phi(\alpha)$ má menší sklon než $\phi(0)$.
- Wolfeho podmínka: (5) a
$$\phi(\alpha)' \geq (1-\varepsilon)\phi'(0). \tag{6}$$
- Silná Wolfeho podmínka: (5) a
$$|\phi(\alpha)'| \leq (1-\varepsilon)|\phi'(0)|. \tag{7}$$
- Nemusím se omezovat jen jedním ε.
- Alternativa: pro $0 < c_1 < c_2 < 1$
$$\phi(\alpha) \leq \phi(0) + c_1\phi'(0)\alpha,$$
$$\phi'(\alpha) \geq c_2\phi'(0),$$
$$|\phi(\alpha)| \leq c_2|\phi'(0)|.$$
- Wolfeho test zaručí jak dostatečný pokles $\phi(\alpha)$, tak dostatečný nárůst $\phi'(\alpha)$ vzhledem k $\alpha = 0$.
- Priklad3_wolfe.m

Konvergence algoritmů typu Line-search:

- Body x_k nalezené Line-search algoritmem obecně nemusí konvergovat k přesnému řešení x^*.
- Konvergenci lze chápat v různém smyslu: $\lim\limits_{k\to\infty} x_k = x^*$ nebo $\lim\limits_{k\to\infty} ||\nabla f(x_k)|| = 0$.
- Označme $\cos\theta_k = \frac{-p_k^T \nabla f_k}{||\nabla f_k|| \, ||p_k||}$, θ_k ... úhel svírající zvolený spádový směr a směr největšího spádu.
- Lipchitzovská spojitost $\nabla f(x)$ na Ω: existuje konstanta $L > 0$ taková, že

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \quad \forall x, y \in \Omega.$$

- Vrstevnicová oblast $\Gamma_{f(x_0)} \equiv \{x : f(x) \leq f(x_0)\}$.

Věta: *Nechť $x_0 \in \mathbb{R}^n$ je počáteční přiblížení a nechť $f \in \mathbf{C}^1(\Omega)$, kde Ω je otevřená množina obsahující vrstevnicovou oblast $\Gamma_{f(x_0)}$. Dále nechť ∇f je Lipschitzovsky spojitý na Ω. Uvažujme algoritmus Line-search, tj. iterace $x_{k+1} = x_k + \alpha_k p_k$, kde p_k je spádový směr a α_k splňuje Wolfeho podmínky. Potom*

$$\sum_{k\geq 0} cos^2(\theta_k)||\nabla f(x_k)||^2 < \infty$$

a tato podmínka implikuje

$$\cos^2(\theta_k)||\nabla f(x_k)||^2 \to 0 \quad pro \quad k \to \infty.$$

Definice: *Řekneme, že konvergentní posloupnost $\{x_k\}$ konverguje k x^* rychlostí řádu p, pokud existují čísla $0 < c < 1$ a $k_0 > 0$ taková, že platí*

$$\frac{||\epsilon_{k+1}||}{||\epsilon_k||^p} \leq c, \quad \forall k \geq k_0.$$

Kde $\epsilon_k \equiv x_k - x^*$.
Pro $p = 1$ mluvíme o lineární rychlosti konvergence,
pro $p = 2$ mluvíme o kvadratické rychlosti konvergence.
Pokud platí

$$\frac{||\epsilon_{k+1}||}{||\epsilon_k||} \leq a_k,$$

kde $\{a_k\}$ je posloupnost nezáporných čísel, $a_k \to 0^+$ pro $k \to \infty$, hovoříme o superlineární konvergenci.

Základní spádové metody

Úloha:

$$\min_{x \in \mathbb{R}^n} f(x),$$

kde $x \in \mathbb{R}^n$, $n \geq 1$ a $f : \mathbb{R}^n \to \mathbb{R}$.

Iterační algoritmy:

- Začínám v x_0 a vytvářím posloupnost iterací $\{x_k\}_{k=0}^{\infty}$, tak, aby minimum bylo aproximováno s dostatečnou přesností, ideálně končila v přesném řešení.
- K posunutí z x_k do x_{k+1}, aby $f(x_{k+1}) < f(x_k)$ mohu použít informace o $x_0, \ldots, x_{k-1}$ a jejich funkčních hodnotách.

Line search algoritmus:

Dáno x_0, interval (a, b), ε atd. a pro $k = 0, 1, 2, \ldots$:

1) najdi spádový směr $p_k \in \mathbb{R}^n$, tj. $\nabla^T f(x_k) p_k < 0$, směr p_k NESMÍ být kolmý k $\nabla f(x_k)$.
2) najdi α_k tak, aby $x_k + \alpha_k p_k$ bylo "lepší" než x_k,
3) polož $x_{k+1} = x_k + \alpha_k p_k$.

Metoda největšího spádu:

- $f \in \mathbf{C}^1(\mathbb{R}^n)$, směr $p_k = -g_k = -\nabla f(x_k)$.
- Globálně konvergentní, $\theta_k = \pi$, $\forall k$, v praxi pomalá.
- Důležitá z teoretického hlediska.

Kvadratická funkce f:

- Předpokládejme funkci f ve tvaru $f(x) = \frac{1}{2} x^T A x - x^T b$, kde A je symetrická pozitivně definitní matice.
- Gradient: $g_k = A x_k - b$.
- Stacionární bod $A x_k = b$ - soustava lineárních algebraických rovnic.
- Hessova matice $H_k = A$ je pozitivně definitní $\to$ ostře konvexní funkce, tj. stacionární bod je globální minimum.

- Zvolme $f_1(x) = \frac{1}{2}||x - x^*||_A^2$, potom $f(x) = f_1(x) - \frac{1}{2}x^{*T}Ax^* \to f(x)$ a $f_1(x)$ se liší jen o konstantu $\to$ mají stejné lokální minimum, stejný gradient.

$$f_1(x) = \frac{1}{2}||x-x^*||_A^2 = \frac{1}{2}(x-x^*)^TA(x-x^*) = \frac{1}{2}(x^TAx-x^{*T}Ax-x^TAx^*+x^{*T}Ax^*) =$$

$$= \frac{1}{2}x^TAx - x^TAx^* + \frac{1}{2}x^{*T}Ax^* = \frac{1}{2}x^TAx - x^Tb + \frac{1}{2}x^{*T}Ax^*.$$

- Dáno x_0, $x_{k+1} = x_k - \alpha_k g_k$, kde $g_k = Ax_k - b$ a $\alpha_k = \frac{g_k^Tg_k}{g_k^TAg_k}$ (exaktní line-search).

$$\phi(\alpha) = f(x_k-\alpha g_k) = \frac{1}{2}(x_k-\alpha g_k)^TA(x_k-\alpha g_k)-(x_k-\alpha g_k)^Tb = \frac{1}{2}x_k^TAx_k-\frac{1}{2}\alpha g_k^TAx_k-$$

$$-\frac{1}{2}\alpha x_k^TAg_k+\frac{1}{2}\alpha^2g_k^TAg_k-x_k^Tb+\alpha g_k^Tb = \frac{1}{2}x_k^TAx_k-\alpha g_k^TAx_k+\alpha g_k^Tb-x_k^Tb+\frac{1}{2}\alpha^2g_k^TAg_k.$$

$$\frac{d\phi(\alpha)}{d\alpha} = -g_k^TAx_k + g_k^Tb + \alpha g_k^TAg_k = 0$$

$$g_k^T(b - Ax_k) + \alpha g_k^TAg_k = 0$$

$$\alpha g_k^TAg_k = g_k^Tg_k \quad \Rightarrow \quad \alpha = \frac{g_k^Tg_k}{g_k^TAg_k}.$$

- Gradient lze počítat rekurzivně: $g_{k+1} = g_k - \alpha_kAg_k$.

$$g_{k+1} = Ax_{k+1} - b = A(x_k - \alpha_kg_k) - b = Ax_k - \alpha_kAg_k - b = g_k - \alpha_kAg_k.$$

- Priklady4.m

Rychlost konvergence:

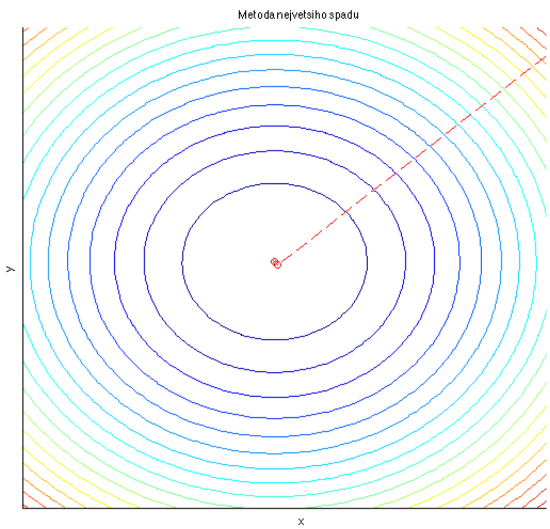
- $||x_{k+1} - x^*||_A^2 = \left(1 - \frac{(g_k^Tg_k)^2}{(g_k^TAg_k)(g_k^TA^{-1}g_k)}\right)||x_k - x^*||_A^2 \ldots$ špatně se interpretuje.

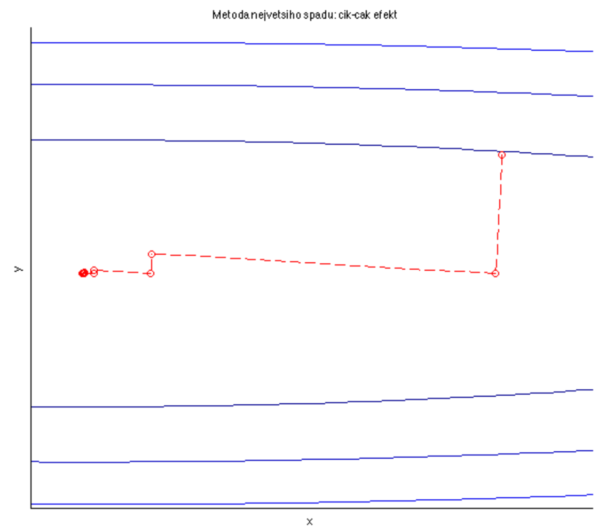
- Důkaz:

$$||x_{k+1} - x^*||_A^2 = (1 - c)||x_k - x^*||_A^2 = ||x_k - x^*||_A^2 - c||x_k - x^*||_A^2$$

$$||x_k - x^*||_A^2 - ||x_{k+1} - x^*||_A^2 = c||x_k - x^*||_A^2$$

1. $x_{k+1} = x_k - \alpha_kg_k$
2. $\alpha_k = \frac{g_k^Tg_k}{g_k^TAg_k}$
3. $g_k = Ax_k - b = A(x_k - x^*)$
4. $||x_k-x^*||_A^2 = ||x_k-A^{-1}b||_A^2 = (x_k-A^{-1}b)^TA(x_k-A^{-1}b) = x_k^TAx_k-b^TA^{-1}Ax_k-x_k^TAA^{-1}b + b^TA^{-1}AA^{-1}b = (Ax_k - b)^TA^{-1}(Ax_k - b) = g_k^TA^{-1}g_k$





$$||x_k - x^*||_A^2 - ||x_{k+1} - x^*||_A^2 = (x_k - x^*)^T A(x_k - x^*) - (x_{k+1} - x^*)^T A(x_{k+1} - x^*) =$$
$$= x_k^T Ax_k - x^{*T}Ax_k - x_k^T Ax^* + x^{*T}Ax^* - x_{k+1}^T Ax_{k+1} + x^{*T}Ax_{k+1} + x_{k+1}^T Ax^* - x^{*T}Ax^* = \dots$$

$$x_{k+1}^T Ax_{k+1} = (x_k - \alpha_k g_k)^T A(x_k - \alpha_k g_k) = x_k^T Ax_k - \alpha_k g_k^T Ax_k - \alpha_k x_k A g_k + \alpha_k^2 g_k^T A g_k$$
$$x^{*T}Ax_{k+1} + x_{k+1}^T Ax^* = 2x^{*T}Ax_{k+1} = 2x^{*T}A(x_k - \alpha_k g_k) = 2x^{*T}Ax_k - 2\alpha_k x^{*T}Ag_k$$
$$-x^{*T}Ax_k - x_k^T Ax^* = -2x^{*T}Ax_k$$

$$\dots = x_k^T Ax_k - 2x^{*T}Ax_k - x_k^T Ax_k + \alpha_k g_k^T Ax_k + \alpha_k x_k^T Ag_k - \alpha_k^2 g_k^T Ag_k + 2x^{*T}Ax_k - 2\alpha_k x^{*T}Ag_k =$$
$$= 2\alpha_k g_k^T A(x_k - x^*) - \alpha_k^2 g_k^T Ag_k = \frac{2g_k^T g_k}{g_k^T Ag_k} g_k^T g_k - \frac{(g_k^T g_k)^2}{(g_k^T Ag_k)^2} \cdot g_k^T Ag_k = \frac{(g_k^T g_k)^2}{(g_k^T Ag_k)} \frac{||x_k - x^*||_A^2}{||x_k - x^*||_A^2} =$$
$$= \frac{(g_k^T g_k)^2}{(g_k^T Ag_k)(g_k^T A^{-1} g_k)} ||x_k - x^*||_A^2$$

- $c \in (0, 1) \rightarrow 1 - c < 1$

Věta: *Nechť je funkce f(x) ostře konvexní (A je symetrická, pozitivně definitní) a aplikujeme metodu největšího spádu, tj.* $x_{k+1} = x_k - \alpha_k g_k$, *kde* $\alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}$, *potom pro chybu platí:*

$$||x_{k+1} - x^*||_A^2 \leq \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}\right)^2 ||x_k - x^*||_A^2,$$

kde $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$, *jsou vlastní čísla matice A.*

- Lineární rychlost konvergence.
- Kantorovičova nerovnost: Nechť $A \in \mathbb{R}^{n\times n}$ je symetrická pozitivně definitní matice, potom pro všechna netriviální x platí

$$\frac{(x^Tx)^2}{(x^TAx)(x^TA^{-1}x)} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2}.$$

- $c = \left(\frac{\lambda_n-\lambda_1}{\lambda_n+\lambda_1}\right)^2 = \left(\frac{\kappa-1}{\kappa+1}\right)^2$, $\kappa = \frac{\lambda_n}{\lambda_1}$ je číslo podmíněnosti matice A.
  - $\kappa = \text{cond}(A) \approx 800 \rightarrow$ dobrá podmíněnost matice A
    $c = \left(\frac{799}{801}\right)^2 = 0.995 \rightarrow$ pomalá konvergence
  - čím menší je c, tj. $\lambda_1 \approx \lambda_n$ (například $A = kI \rightarrow \lambda_1 = \cdots = \lambda_n = 1$), potom jsou hladiny podobné kružnicím $\rightarrow$ rychlá konvergence
    čím větší c, potom se hladiny protahují do elips a čím protaženější elipsy, tím horší cik-cak efekt.
- Priklady4.m

Nekvadratická funkce f:

- Podobné použití, problém gradient a line-search.
- V blízkosti minima funkce $f(x)$ se začne blížit kvadratické funkci (Taylorův polynom).
- Konverguje lineárně s faktorem c a λ_1, λ_n jsou vlastní čísla Hessovy matice $H(x_k) = \nabla^2 f(x_k)$.

Newtonova metoda:

- f aproximujeme lokálně kvadratickou funkcí, minimum je další aproximace.
- Taylorův polynom v bodě x_k:

$$f(x) \approx q(x) = f(x_k) + \nabla f(x_k)^T(x-x_k) + \frac{1}{2}(x-x_k)^T H(x_k)(x-x_k).$$

- Stacionární bod:

$$\nabla q(x) = \nabla f(x_k) + H(x_k)(x-x_k) = 0,$$
$$x = x_{k+1} = x_k - H(x_k)^{-1}\nabla f(x_k).$$

- Vyřešit $H(x_k)d_k = -g_k$, $x_{k+1} = x_k + d_k$ - čistá Newtonova metoda.
- V každém kroku se řeší soustava lineárních algebraických rovnic.
- V blízkosti řešení, nebo pokud je $f(x)$ ostře konvexní, potom $H(x_k)$ je pozitivně definitní.

Věta: *Nechť* $f(x) \in \mathbf{C}^2(\mathbb{R}^n)$ *a* $\nabla^2 f(x) = H(x)$ *je Lipschitzovsky spojitá v okolí řešení* x^**. Předpokládejme Newtonovu iterační metodu* $x_{k+1} = x_k + p_k$*, kde* $p_k = -H^{-1}(x_k)g_k$*, potom*

a) pokud je x_0 *blízko řešení* x^**, potom posloupnost Newtonových iterací konverguje k řešení,*

b) rychlost konvergence posloupnosti $\{x_k\}$ *je kvadratická,*

c) posloupnost $\{||\nabla f_k||\}$ *konverguje k nule kvadraticky.*

- Základní Newtonova metoda není vhodná pro praktické výpočty.
- H_k nemusí být symetrická pozitivně definitní a i pokud je, nemusí iterace konvergovat (x_0 je daleko od x^*).

Line search:

- Pokud H_k je symetrická pozitivně definitní a $d_k = -H_k^{-1} g_k$, potom $g_k^T d_k = -g_k^T H_k^{-1} g_k < 0 \to d_k$ je spádový směr.
- $x_{k+1} = x_k - \alpha_k H_k^{-1} g_k$, minimalizace ve směru blízko řešení, tj, $\alpha_k \approx 1$.
- Priklady4_Newton.m

Obecně lze uvažovat třídu algoritmů

$$x_{k+1} = x_k - \alpha_k M_k g_k,$$

kde $M_k \in \mathbb{R}^{n \times n}$, pokud $M_k = I \to$ metoda největšího spádu, pro $M_k = H_k^{-1} \to$ Newtonova metoda. Nejlépe pokud M_k je symetrická pozitivně definitní $\to$ zajistím spádovost směru $p_k = -M_k g_k$.

- $g_k^T p_k < 0 \to -g_k^T M_k g_k < 0$ pokud M_k je SPD