Optimalizace bez omezení (unconstraint)

Nederivační (ad hoc) metody

- Jednoduché metody
- Nelder-Meadova (simplexová) metoda

Derivační metody

- *První derivace*
  - Metoda největšího spádu + další spádové metody
  - Metoda konjugovaných gradientů
- *Druhá derivace*
  - Newton-Raphsonova metoda
  - Quasi-Newtonova metoda

Metoda největšího spádu
-obecně II

Algoritmus:

- zvolíme výchozí bod $x^{(0)}$
- k-tá iterace:

  bod $x^{(k+1)}$ vypočítáme z bodu $x^{(k)}$ pomocí vztahu:

  $x^{(k+1)} = x^{(k)} - \alpha . g^{(k)}$, kde:

$-g^{(k)}$	zjednodušený zápis $-\nabla f(x^{(k)})$, určuje **směr přesunu** z bodu $x^{(k)}$
α	koeficient, popisující **délku daného přesunu**

Metoda největšího spádu

- volba α v metodě největšího spádu

Z bodu $x^{(k)}$ se přesunujeme po polopřímce:

$x(\alpha) = x^{(k)} + \alpha.x^{(k)}$, kde $\alpha > 0$

Hodnotu funkce f na této polopřímce popisuje funkce $\phi(\alpha)$: $\phi(\alpha) = f(x(\alpha))$

Je zřejmé, že musíme zvolit takové α^{OK}, aby platilo:

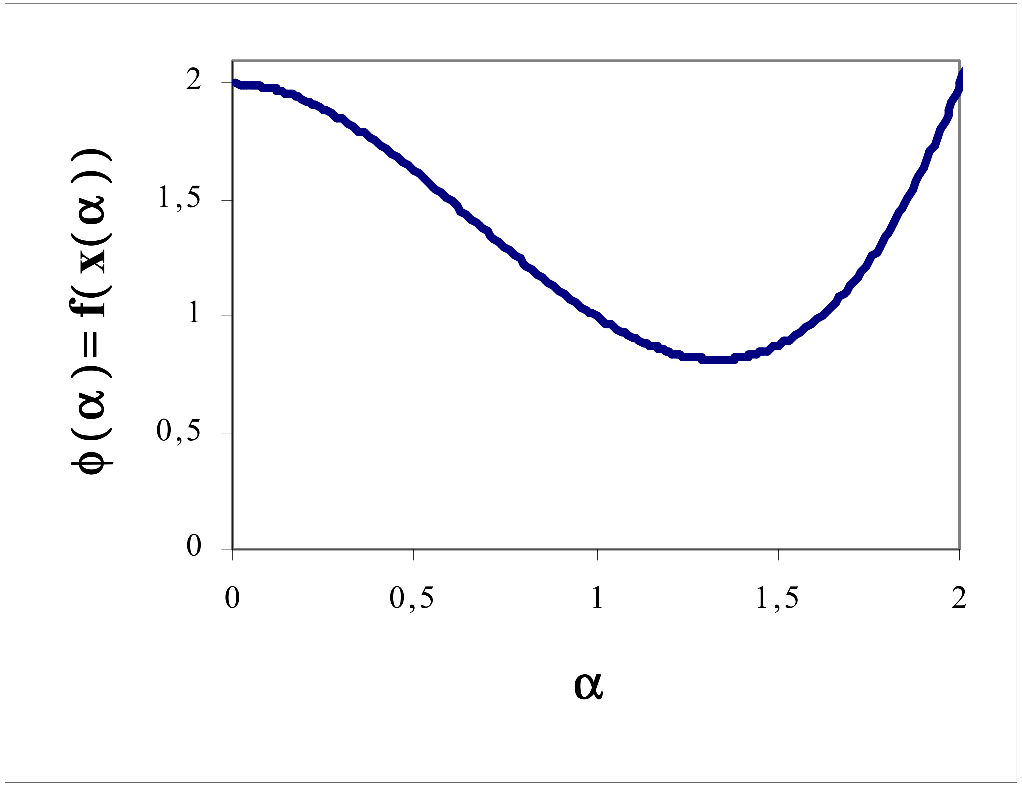
$f(x^{(k)}) > f(x^{(k+1)})$, kde $x^{(k+1)} = x(\alpha^{OK})$ pro dostatečný počet iterací.

Poznámka: „Dostatečný počet“ = dostačuje k tomu, aby metoda konvergovala k minimu ($\lim_{k \to \infty} f(x^{(k)}) = 0$).

Metoda největšího spádu

- volba a v metodě největšího spádu

Funkce $\phi(\alpha)$ má následující tvar:



Metoda největšího spádu

- volba a v metodě největšího spádu

Metoda největšího spádu volí pro každý krok stejnou hodnotu α.

Konkrétně velmi malou hodnotu α.

Poznámka: Hodnoty a musí být dostatečně, aby metoda konvergovala.

Metoda největšího spádu
zhodnocení

Výhody:

- Implementačně jednoduché
- Nízká prostorová složitost

Nevýhody:

- Velmi pomalá konvergence (speciálně v oblastech malého spádu => nízkých hodnot gradientu).
- Chyby, způsobené zaokrouhlením. Mohou vést i k tomu, že se výpočet vůbec nedostane rozumně blízko k minimu. Ale při (ideální) přesné aritmetice metoda konverguje vždy k nějakému lokálnímu minimu.

Metoda největšího spádu
- příklad

Rosenbrockova funkce:

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

Gradient Rosenbrockovy funkce:

$$\nabla f(x_1, x_2) = \left(-400x_1(x_2 - x_1^2) + 2(1 - x_1),\ 200(x_2 - x_1^2)\right)$$

Výchozí bod:

$x_0 = (-2, 2)$

Parametry:

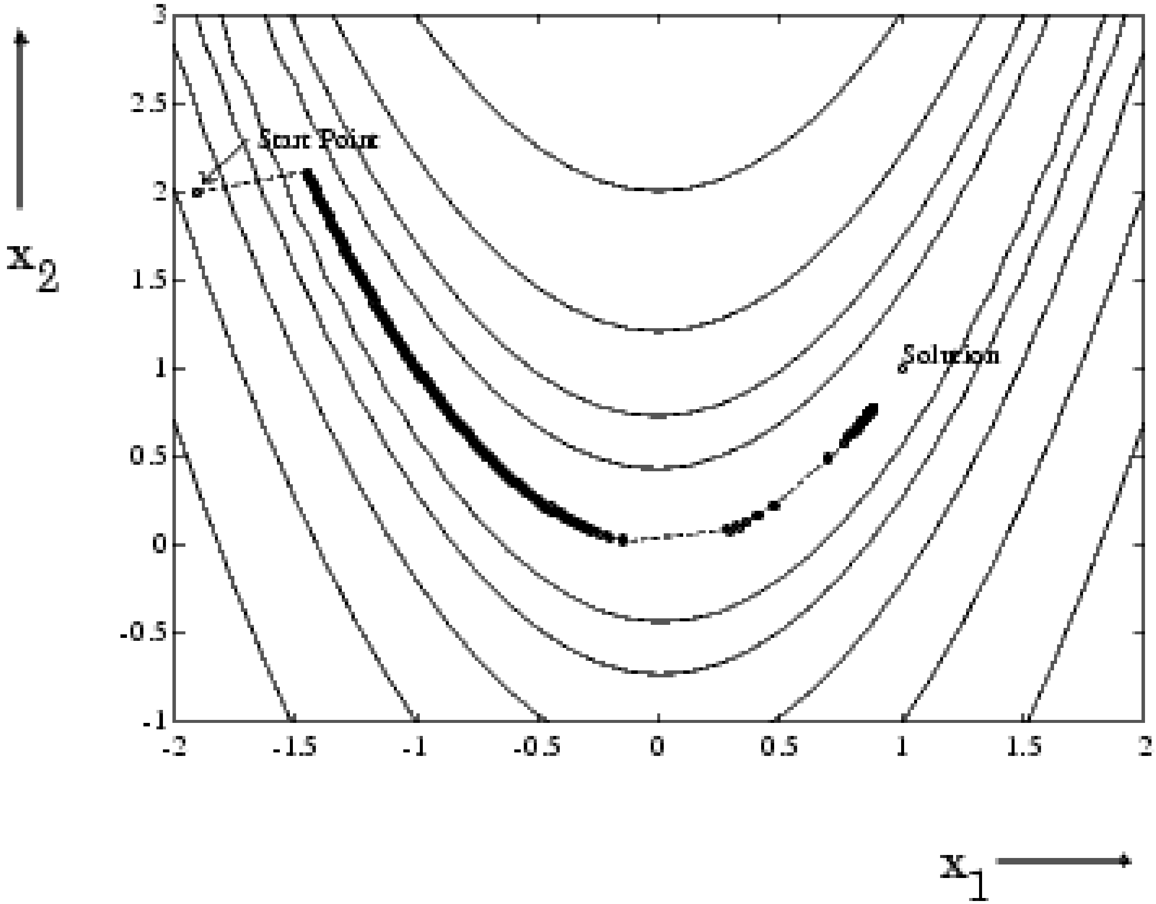
$\alpha = 0{,}001$

Výpočet první iterace - dobrovolník u tabule :-).

Další iterace - příklad v Excelu.

Metoda největšího spádu
- příklad II

Ukázka konvergence metody největšího spádu pro Rosenbrockovu funkci:



Spádové metody
- obecně

Jsou založeny na stejném principu jako metoda největšího spádu:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha.\mathbf{s}^{(k)}, \text{ kde:}$$

$\mathbf{s}^{(k)}$	je **směr přesunu** z bodu $\mathbf{x}^{(k)}$,
	nejčastěji jako směr volíme -$\mathbf{g}^{(k)}$
α	koeficient, popisující **délku daného přesunu**

Využívají sofistikovanější metody k určení koeficientu α.

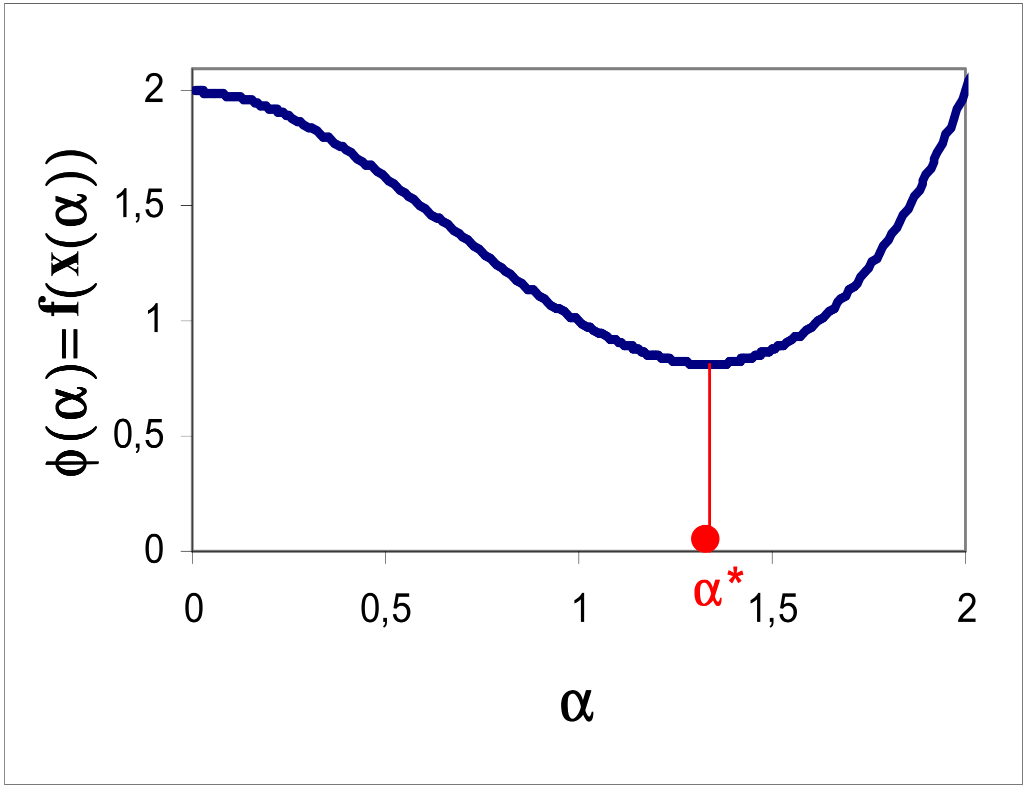
Hodnota koeficientu α je různá pro každou iteraci.

Spádové metody
- obecně II

Podmínka pro ideální hodnotu (α^*) koeficientu α:

funkce $\phi(\alpha) = f(x(\alpha))$ má v α^* minimum [1]

Poznámka: Jedná se o nejmenší hodnotu α, v níž má $\phi(\alpha)$ minimum. Navíc samozřejmě platí $\alpha > 0$.



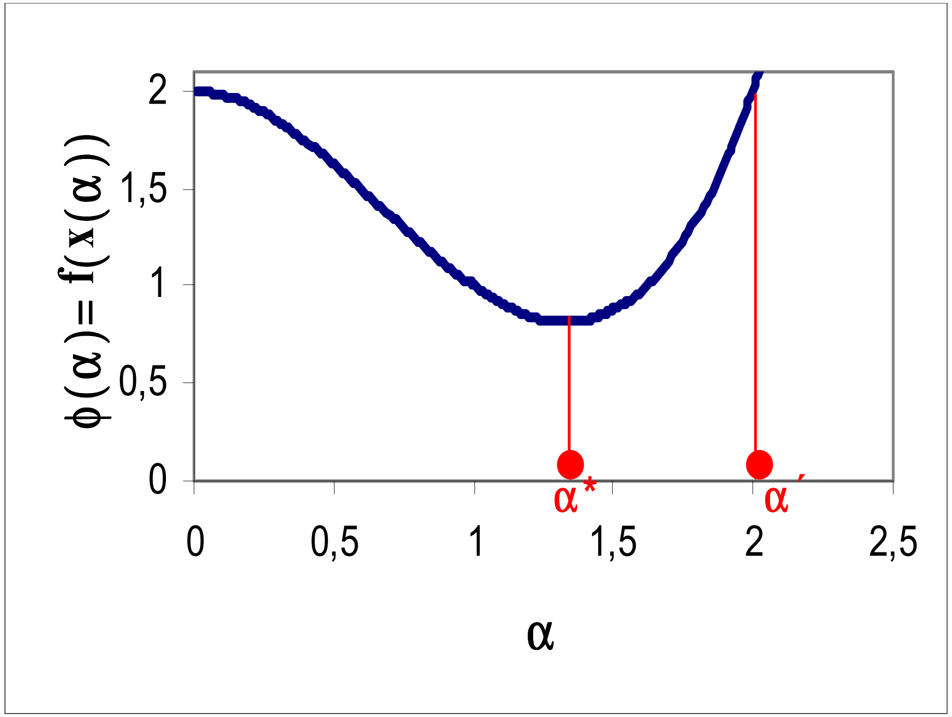
Tuto podmínku nelze využít k volbě koeficientu α.

Potřebujeme totiž určit hodnotu α pro danou iteraci v konečném a pokud možno velmi malém počtu kroků.

Spádové metody

- obecně III

Předpokládejme, že funkce f má průběh naznačený na obrázku.



Pak existuje $\alpha > 0$ tak, že:

$f(x^{(k)} + \alpha s^{(k)}) = f(x^{(k+1)})$

Nejmenší takové α označujeme α'.

=> Nutné podmínky pro koeficient α:

$$0 < \alpha < \alpha' \qquad [2]$$

Spádové metody
- obecně IV

=> Při volbě koeficientu a musíme dodržet podmínky [2] a co nejvíce se přiblížit podmínce [1].

Metody pro nalezení α:

- Goldsteinovy podmínky
- Wolfe-Powellovy podmínky

Značení:

$f(x^{(k+1)}) = f(x^{(k)} + \alpha . s^{(k)})$ budeme značit $\phi(\alpha)$

$f(x^{(k)})$ budeme značit $\phi(0)$

Spádové metody
- Goldsteinovy podmínky I

1. Goldsteinova podmínka zaručuje, že α nebude zvoleno příliš blízko α':

$$\phi(\alpha) \leq \phi(0) + \rho.\alpha.\phi'(0)$$

Parametr ρ je zde pevně zvolené číslo z intervalu (0, ½).

I když α' neexistuje {tj. $\phi(\alpha) < \phi(0)$ pro všechna $\alpha > 0$}, je 1GP schopna omezit volbu $\alpha^{(k)}$. (Pokud je tedy funkce f zdola omezená.)

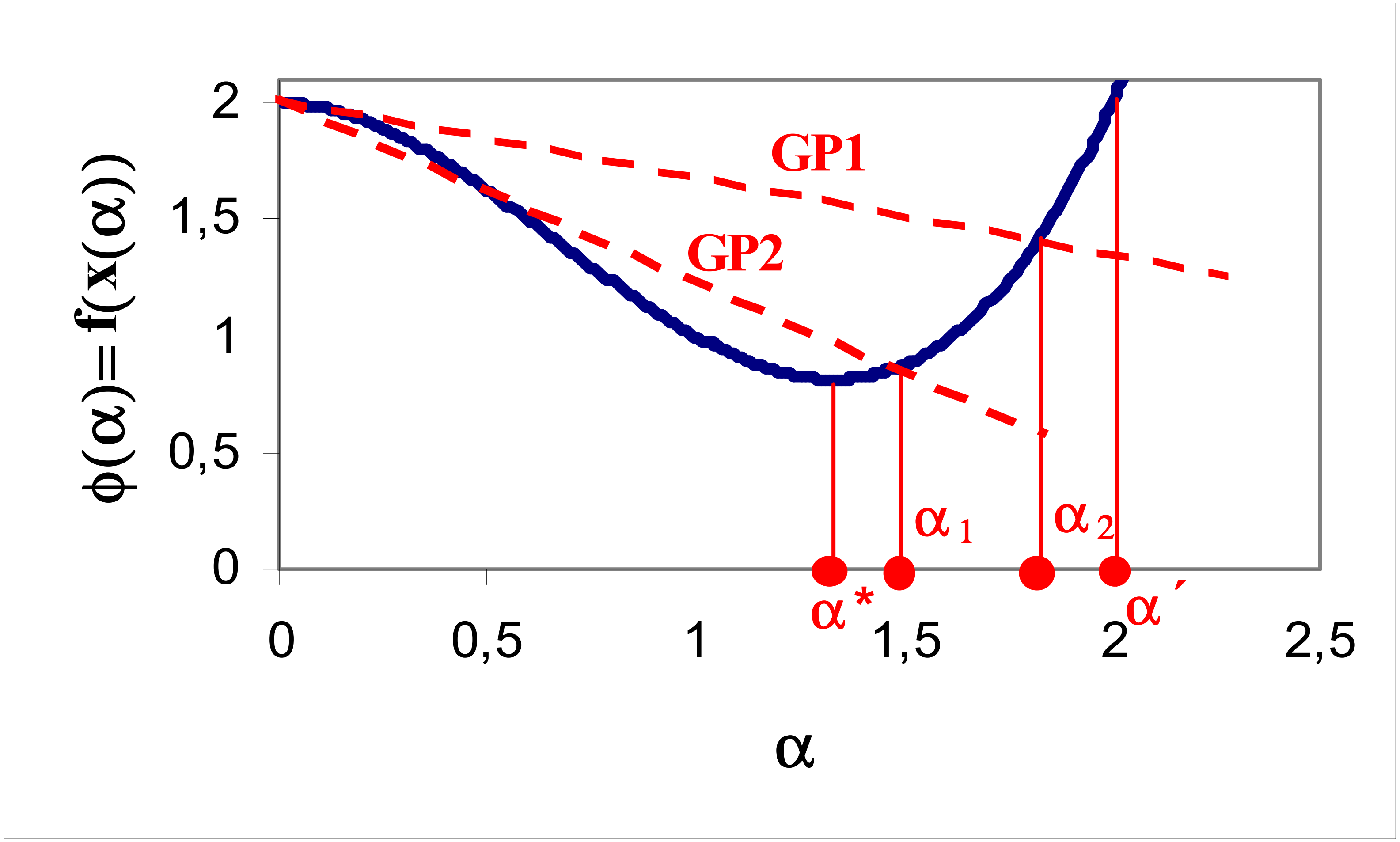
Spádové metody
- Goldsteinovy podmínky II

2. Goldsteinova podmínka zaručuje, že α nebude zvoleno příliš blízko 0:

$$\phi(\alpha) \geq \phi(0) + (1-\rho).\alpha.\phi'(0)$$

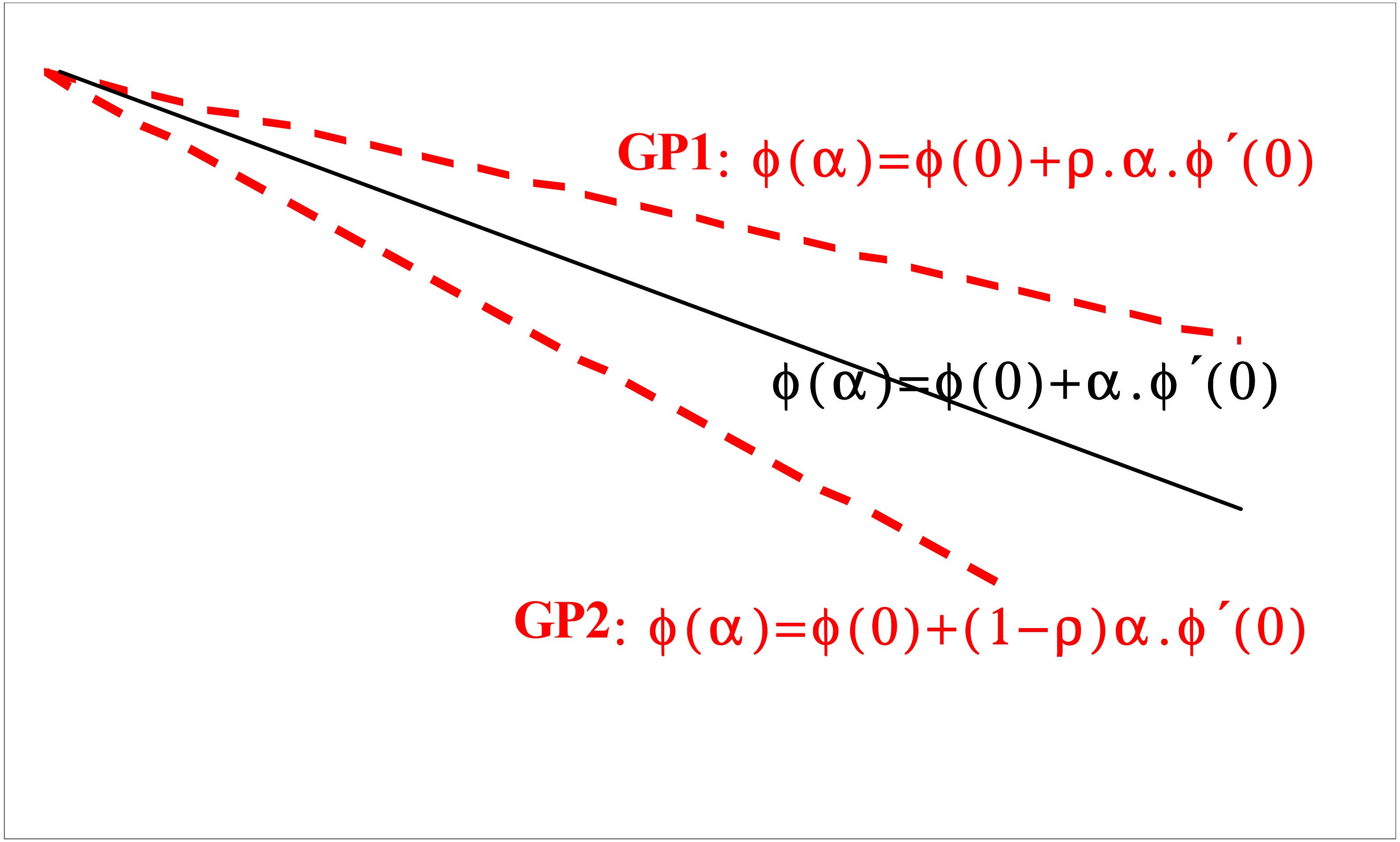
Pravé strany Goldsteinových podmínek určují dvě přímky se zápornou směrnicí. Hodnoty α_1 a α_2, které přísluší průsečíkům těchto přímek s $\phi(\alpha)$, určují interval vhodných hodnot α.

Spádové metody
- Goldsteinovy podmínky III



Spádové metody
- Goldsteinovy podmínky IV

Zdůvodnění Goldsteinových podmínek:



Spádové metody
- Goldsteinovy podmínky V

Věta: Nechť funkce f je spojitě diferenciovatelná a nechť její gradient $\mathbf{g} = \nabla f$ je lipschitzovsky spojitý na R^n.

Je-li $\{\mathbf{x}^{(k)}\}$ posloupnost generována spádovou metodou a volba a vyhovuje Goldsteinovým podmínkám, pak platí:

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^{(k-1)}) \geq -\rho.\mathbf{g}^{(k)}.\delta^{(k)}$$

kde $\delta^{(k)} = \alpha^{(k)}\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$

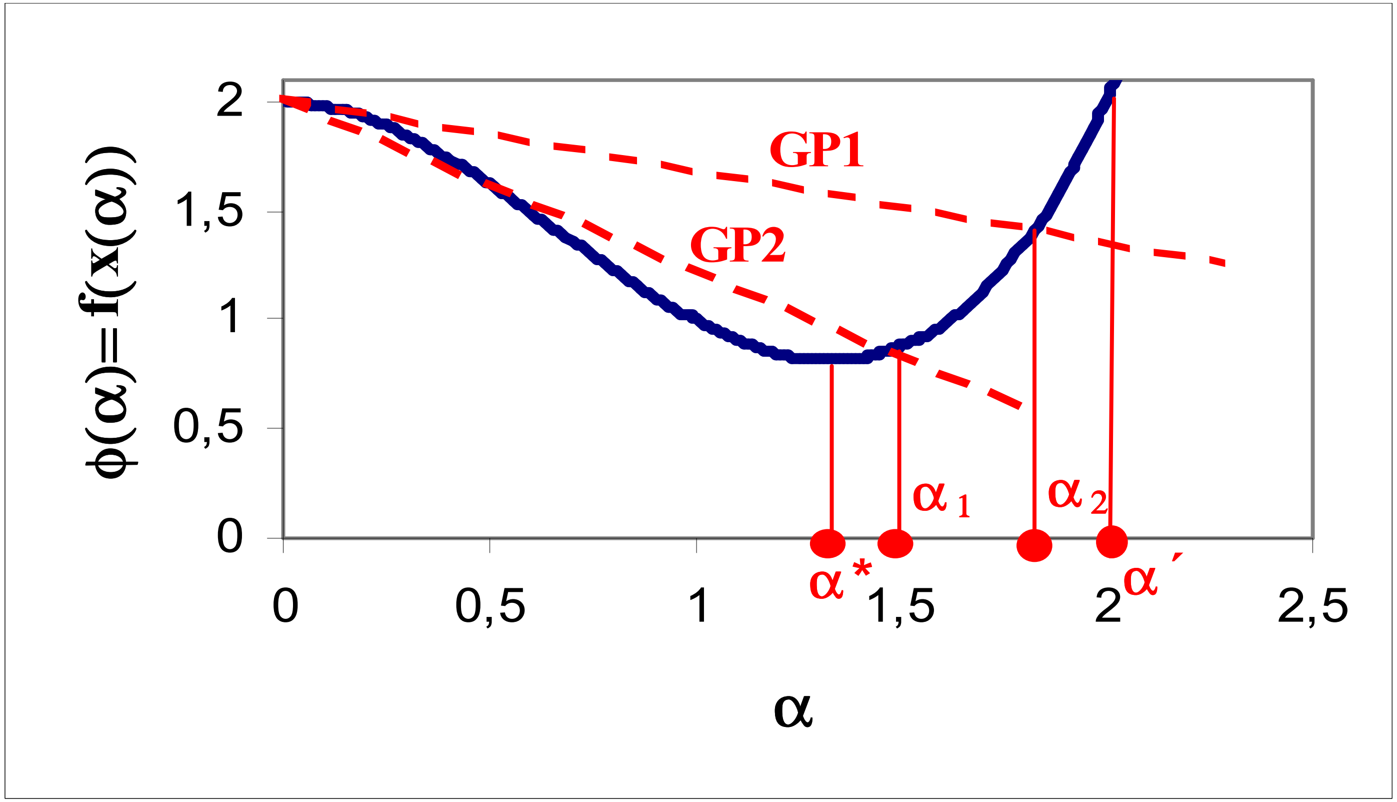
=> Metoda vždy konverguje k minimu funkce f (pro vhodné ρ).

Poznámka: Volba koeficientu $\rho < ½$ zaručuje konvergenci metody pro kvadratické funkce.

Spádové metody
- Goldsteinovy podmínky VI

Nevýhoda Goldsteinových podmínek:

V intervalu mezi α_1 a α_2 se nemusí nacházet minimum funkce $\phi(\alpha)$.



Spádové metody

- Wolfe-Powellovy podmínky

Místo GP2 se testuje sklon funkce $\phi(\alpha)$ v bodě α.

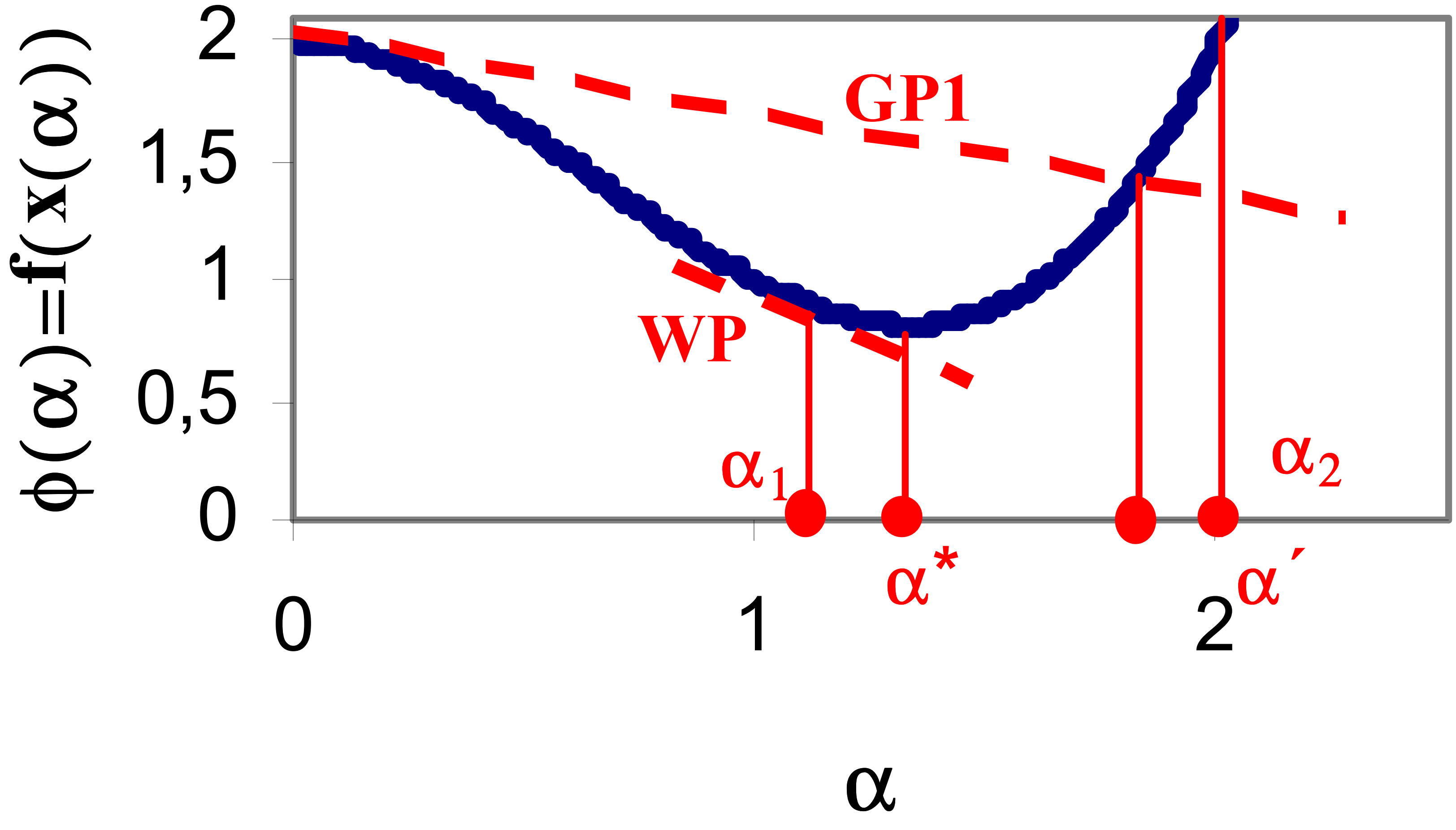
Využívá se tedy následující podmínka:

$$|\phi'(\alpha)| \leq \sigma |\phi'(0)|$$

kde $\sigma \in (\rho, 1)$.

Spádové metody

- Wolfe-Powellovy podmínky II



Metody konjugovaných gradientů
- obecně

= metody sdružených gradientů

= conjugate gradient method

= speciální případ metod sdružených směrů

Základní myšlenka: Pro určení směru přesunu z bodu $x^{(k)}$ do bodu $x^{(k+1)}$ se využívají nejen hodnotu $g^{(k+1)}$, ale rovněž hodnotu $g^{(k)}$. (V obecném případě je možno využít hodnot $g^{(1)}$, $g^{(2)}$, ..., $g^{(k)}$, $g^{(k+1)}$.)

Zdůvodnění: Spojení informací o současném a předchozím sklonu studované funkce umožňuje rychlejší sestup do minima (zlepšení konvergence na plochých oblastech).

Metody konjugovaných gradientů

- algoritmus

Výpočet $\mathbf{x}^{(k+1)}$:

$\mathbf{x}^{(k+1)}$ se určuje pomocí stejného vztahu jako u spádových metod:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)}.\mathbf{s}^{(k)} \qquad [2.1]$$

kde:

$\mathbf{s}^{(k)}$ **směr přesunu** z bodu $\mathbf{x}^{(k)}$

$\alpha^{(k)}$ koeficient, popisující **délku daného přesunu**

Metody konjugovaných gradientů

- algoritmus II

Výpočet $\alpha^{(k)}$:

Analogické jako u spádových metod:

1) Je nutno zvolit koeficient $\alpha^{(k)}$ tak, aby platilo:

$$0 < \alpha^{(k)} < \alpha^{(k)\prime}$$

kde: $f(x^{(k)} + \alpha^{(k)\prime}.s^{(k)}) = f(x^{(k+1)})$

2) $\alpha^{(k)}$ by se měla co nejvíce blížit $\alpha^{(k)*}$, kde $\alpha^{(k)*}$ je minimum funkce $f(x^{(k)} + \alpha^{(k)}.s^{(k)})$.

Metody konjugovaných gradientů

- algoritmus III

Výpočet $\mathbf{s}^{(k+1)}$:

$\mathbf{s}^{(k+1)}$ se vypočte pomocí gradientu $\mathbf{g}^{(k+1)}$ a směru $\mathbf{s}^{(k)}$. (Přičemž $\mathbf{s}^{(k)}$ byl vypočítán pomocí předchozích hodnot gradientů ...). Konkrétně:

$$\mathbf{s}^{(k+1)} = -\mathbf{g}^{(k+1)} + \beta^{(k)}.\mathbf{s}^{(k)} \qquad [2.2]$$

Kde b$^{(k)}$ je koeficient, který určuje míru vlivu směru přesunu v kroku k ($\mathbf{s}^{(k)}$) na směr přesunu v následujícím kroku ($\mathbf{s}^{(k+1)}$).

Výpočet $\mathbf{s}^{(0)}$:

$$\mathbf{s}^{(0)} = -\mathbf{g}^{(0)}$$

Metody konjugovaných gradientů
- algoritmus III

Výpočet $\beta^{(k+1)}$:

Existuje více možností, jak volit číslo $\beta^{(k+1)}$.

Nyní odvodíme hodnotu $\beta^{(k+1)}$ za předpokladu, že minimalizovaná funkce je kvadratická a má pozitivně definitní Hessovou matici **G**. V tomto případě platí:

$$\mathbf{y}^{(k)} = \mathbf{G}.\delta^{(k)} \qquad [2.3]$$

kde $\mathbf{y}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$ a $\delta^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$.

Protože vektory $\mathbf{s}^{(k+1)}$ a $\mathbf{s}^{(k)}$ mají být sdružené vzhledem ke **G**, musí platit:

$$\mathbf{s}^{(k+1)T}.\mathbf{G}.\delta^{(k)} = 0 \qquad [2.4]$$

Z [2.3] a [2.4] vyplývá:

$$0 = \mathbf{s}^{(k+1)T}.\mathbf{G}.\delta^{(k)} = \mathbf{s}^{(k+1)T}.\mathbf{y}^{(k)}$$

Metody konjugovaných gradientů

- algoritmus IV

$$\mathbf{s}^{(k+1)} = -\mathbf{g}^{(k+1)} + \beta^{(k)}.\mathbf{s}^{(k)} \qquad [2.2]$$

Po transponování rovnice [2.2] a násobení vektorem $\mathbf{y}^{(k)}$ zprava dostáváme:

$$0 = -\mathbf{g}^{(k+1)T}.\mathbf{y}^{(k)} + \beta^{(k)}.\mathbf{s}^{(k)T}.\mathbf{y}^{(k)} \qquad [2.5]$$

Odtud plyne hodnota, kterou navrhli Hestenes a Stiefel v roce 1952:

$$\beta_{HS}^{(k)} = \frac{\boldsymbol{g}^{(k+1)T}(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)})}{\boldsymbol{s}^{(k)T}(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)})} \qquad [2.6]$$

Metody konjugovaných gradientů

- algoritmus V

Pokud je $\mathbf{x}^{(k)}$ v rovnici [2.1] přesně určeným minimem ve směru $\mathbf{s}^{(k)}$, musí platit $\mathbf{s}^{(k)T}.\mathbf{g}^{(k+1)} = 0$ (jinak by se hodnota f ve směru $\mathbf{s}^{(k)}$ dala ještě snížit), obdobně $\mathbf{s}^{(k-1)T}.\mathbf{g}^{(k)} = 0$. Pak je jmenovatel ve vztahu [2.6] roven:

$$\boldsymbol{s}^{(k)T}\left(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)}\right) = -\boldsymbol{s}^{(k)T}\boldsymbol{.}\boldsymbol{g}^{(k)} = \left[\boldsymbol{g}^{(k)} - \beta^{(k-1)}\boldsymbol{.}\boldsymbol{s}^{(k-1)}\right]^T\boldsymbol{.}\boldsymbol{g}^{(k)} =$$

$$= \boldsymbol{g}^{(k)}\boldsymbol{.}\boldsymbol{g}^{(k)} = \left\|\boldsymbol{g}^{(k)}\right\|_2^2 \qquad [2.7]$$

Metody konjugovaných gradientů

- algoritmus VI

Po dosazení [2.7] do [2.6] získáme vyjádření $\beta^{(k)}$, které poprvé publikovali Polak a Ribiere v roce 1969:

$$\beta_{PR}^{(k)} = \frac{\boldsymbol{g}^{(k+1)T}(\boldsymbol{g}^{(k+1)} - \boldsymbol{g}^{(k)})}{\left\|\boldsymbol{g}^{(k)}\right\|_2^2} \qquad [2.8]$$

Obdobně lze ukázat, že $\mathbf{g}^{(k+1)T}.\mathbf{g}^{(k)} = 0$, odtud plyne vztah pro $\beta^{(k)}$ podle Fletchera a Reevese (1963):

$$\beta_{FR}^{(k)} = \frac{\left\|\boldsymbol{g}^{(k+1)}\right\|_2^2}{\left\|\boldsymbol{g}^{(k)}\right\|_2^2}$$

Metody konjugovaných gradientů

- algoritmus VII

Pro kvadratické funkce je $\beta_{HS}^{(k)} = \beta_{PR}^{(k)} = \beta_{FR}^{(k)}$.

To ale neplatí pro obecnější funkce. Při testovacích úlohách dává obvykle nejlepší výsledky varianta Polaka a Ribiera.

Metody konjugovaných gradientů

- zhodnocení

Výhody:

Spolehlivější než spádové metody.

Vhodná i v oblastech poblíž minima

Nevýhody:

Výpočetně náročnější.

Větší prostorová složitost (nutnost ukládat několik n-prvkových vektorů).

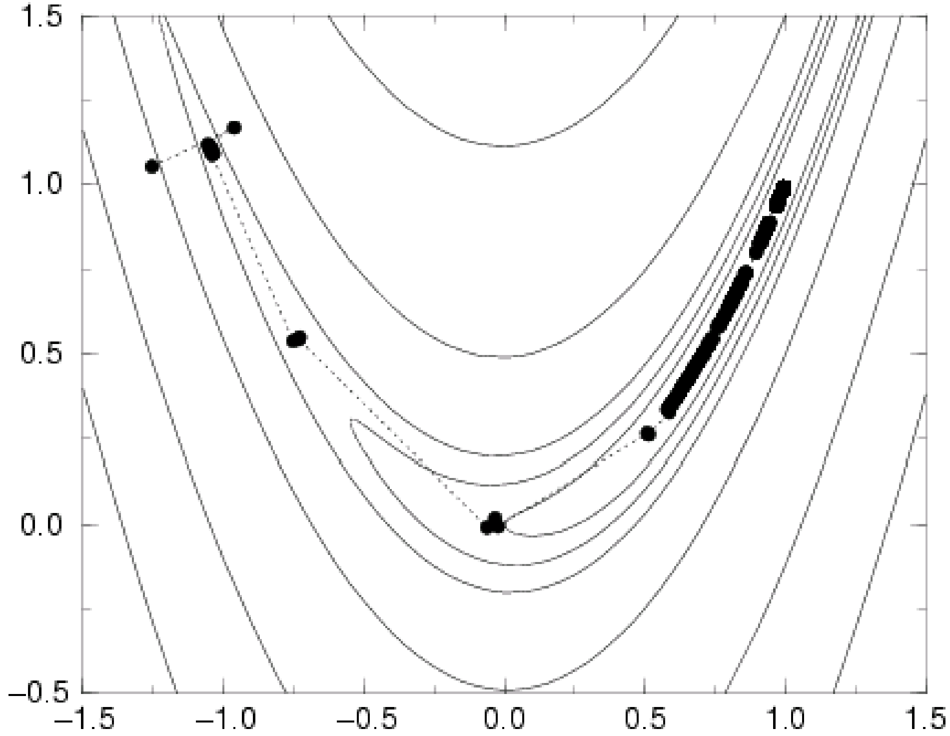
Metody konjugovaných gradientů
- porovnání

Vhodná i v oblastech poblíž Porovnání metody největšího spádu a metody konjugovaných gradientů (Rosenbrockova funkce):

	Daleko od minima (gradient < 1kcal.A^{-2})		**Poblíž minima** (gradient < 0,1kcal.A^{-2})	
	CPU	Počet iter.	CPU	Počet iter.
Nejv. spád	67	98	1405	1893
Konj. grad.	149	213	257	367

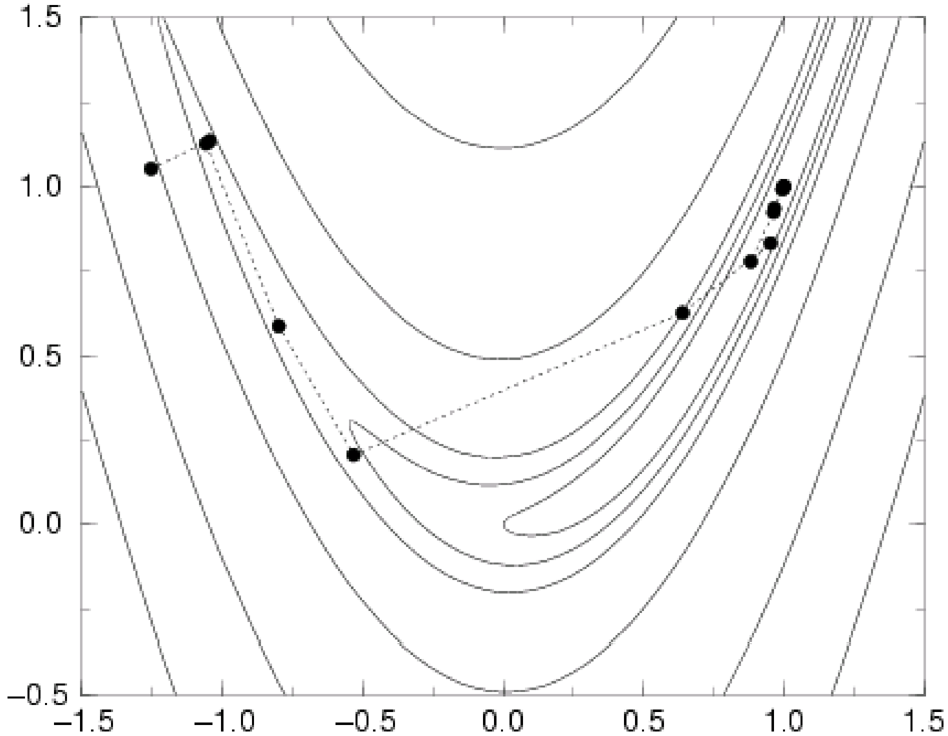
Metody konjugovaných gradientů
- porovnání II

Ukázka konvergence spádové metody pro Rosenbrockovu funkci:



Metody konjugovaných gradientů
- porovnání III

Ukázka konvergence metody konjugovaných gradientů pro Rosenbrockovu funkci:



Cvičení

Proveďte první 2 kroky optimalizace funkce:

$f(x_1, x_2) = x_1^2 + 2x_2^2$

a) pomocí metody největšího spádu

b) pomocí metody konjugovaných gradientů

Poznámka: Využijte $\alpha = 0{,}25$. Počáteční bod je (2,1)

Funkce: $f(x_1, x_2) = x_1^2 + 2x_2^2$
Výchozí bod: $x^0 = (2, 1)$
Parametry: $\alpha = 0.25$
Gradient funkce: $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce: $s^k = -g^k$

Následující bod: $x^{k+1} = x^k + \alpha.s^k$

Gradient v bodě x^0: $g^0 = (4,4)$
Směr v bodě x^0: $s^0 = (-4,-4)$
Bod x^1: $x^1 = (2,1) + 0,25.(-4,-4) = (1,0)$

Gradient v bodě x^1: $g^1 = (2,0)$

Směr v bodě x^0: $s^0 = (-2,0)$
Bod x^1: $x^1 = (1,0) + 0,25.(-2,0) = (0.5,0)$

Funkce: $f(x_1, x_2) = x_1^2 + 2x_2^2$
Výchozí bod: $x^0 = (2, 1)$
Parametry: $\alpha = 0.25$
Gradient funkce: $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Beta: $\boldsymbol{\beta_{FR}^{(k)} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}}$
Směr funkce: $s^k = -g^k$ $\mathbf{+ \beta^k.s^k}$

Následující bod: $x^{k+1} = x^k + \alpha.s^k$

Gradient v bodě x^0: $g^0 = (4,4)$
Směr v bodě x^0: $s^0 = (-4,-4)$
Bod x^1: $x^1 = (2,1) + 0,25.(-4,-4) = (1,0)$

Gradient v bodě x^1: $g^1 = (2,0)$
Beta v bodě x^1: $b^1 = (2.2)/(4.4+4.4) = 4/32 = 0,125$
Směr v bodě x^1: $s^1 = -(2,0)+0.125.(-4,-4) = (-2.5 ,-0.5)$
Bod x^1: $x^1 = (1,0) + 0,25.(-2,5,-0.5) = (0.375,-0.125)$

Příloha

Příklad je řešen v přiloženém excelovém souboru „3uncon_metody_prvni_derivace.xlsx“