

# *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

# *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  $g^{(1)}, g^{(2)}, \dots, g^{(k)}, g^{(k+1)}$ .)

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:

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

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:

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

$$\text{kde: } f(\mathbf{x}^{(k)} + \alpha^{(k)'} \cdot \mathbf{s}^{(k)}) = f(\mathbf{x}^{(k+1)})$$

2)  $\alpha^{(k)}$  by se měla co nejvíce blížit  $\alpha^{(k)*}$ ,

kde  $\alpha^{(k)*}$  je minimum funkce  $f(\mathbf{x}^{(k)} + \alpha^{(k)} \cdot \mathbf{s}^{(k)})$ .

# *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:

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

Kde  $\beta^{(k)}$  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  $\mathbf{G}$ . V tomto případě platí:

$$\mathbf{y}^{(k)} = \mathbf{G} \cdot \boldsymbol{\delta}^{(k)} \quad [2.3]$$

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

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

$$\mathbf{s}^{(k+1)\text{T}} \cdot \mathbf{G} \cdot \boldsymbol{\delta}^{(k)} = 0 \quad [2.4]$$

Z [2.3] a [2.4] vyplývá:

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

# *Metody konjugovaných gradientů*

## *- algoritmus IV*

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

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

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

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

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

# *Metody konjugovaných gradientů*

## *- algoritmus V*

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

$$\begin{aligned} \mathbf{s}^{(k)T} \left( \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)} \right) &= -\mathbf{s}^{(k)T} \cdot \mathbf{g}^{(k)} = \left[ \mathbf{g}^{(k)} - \beta^{(k-1)} \cdot \mathbf{s}^{(k-1)} \right]^T \cdot \mathbf{g}^{(k)} = \\ &= \mathbf{g}^{(k)} \cdot \mathbf{g}^{(k)} = \left\| \mathbf{g}^{(k)} \right\|_2^2 \end{aligned} \quad [2.7]$$



# *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:

$$\beta_{PR}^{(k)} = \frac{\mathbf{g}^{(k+1)T} (\mathbf{g}^{(k+1)} - \mathbf{g}^{(k)})}{\|\mathbf{g}^{(k)}\|_2^2} \quad [2.8]$$

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

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

# *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*

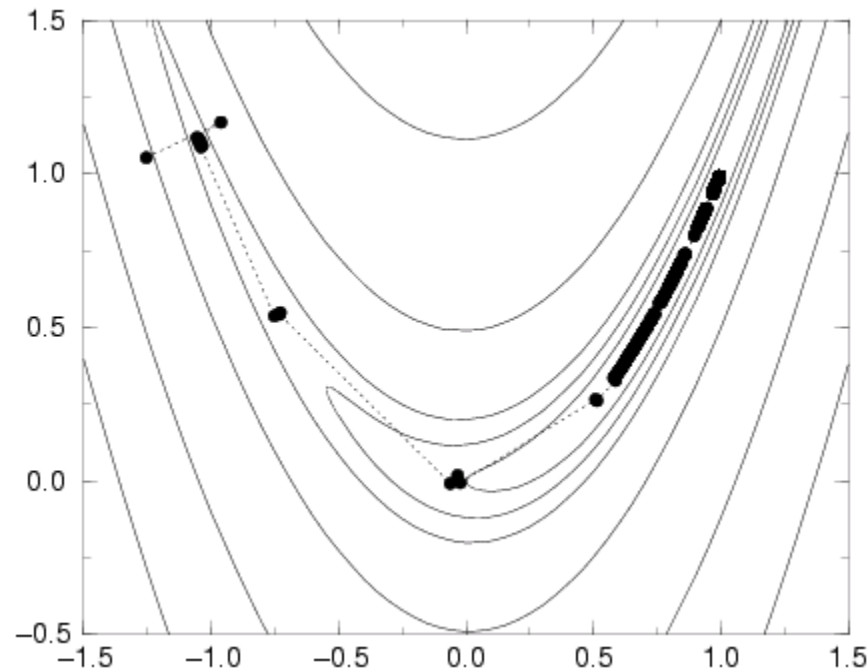
Vhodná i v oblastech poblíž Porovnání metody největšího spádu a metody konjugovaných gradientů (Rosenbrockova funkce):

	<b>Daleko od minima</b> (gradient < 1 kcal.A <sup>-2</sup> )		<b>Poblíž minima</b> (gradient < 0,1 kcal.A <sup>-2</sup> )	
	CPU	Počet iter.	CPU	Počet iter.
<b>Nejv. spád</b>	67	98	1405	1893
<b>Konj. grad.</b>	149	213	257	367

# *Metody konjugovaných gradientů*

## *- porovnání II*

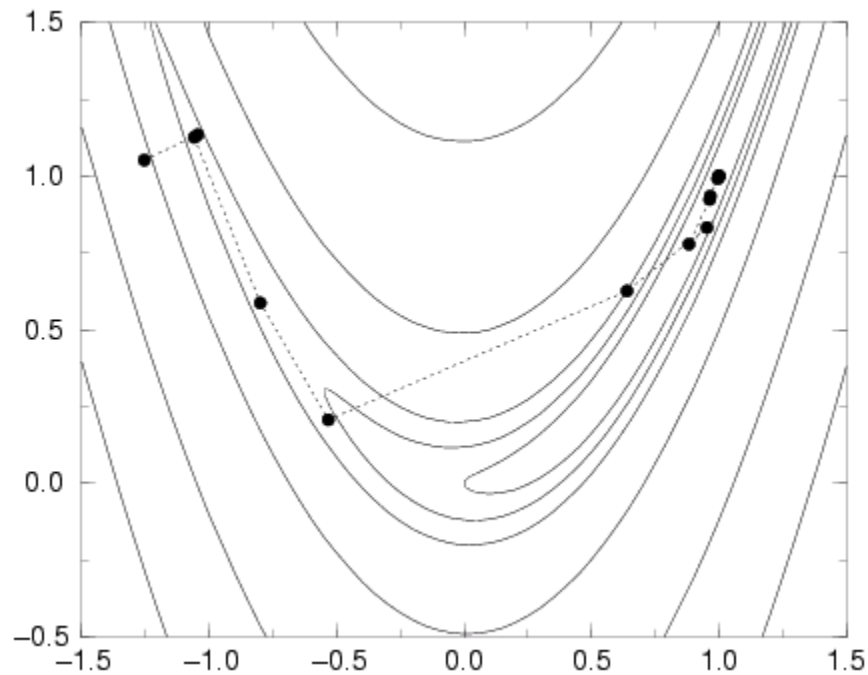
Ukázka konvergence spádové metody pro Rosenbrockovu funkci:



# *Metody konjugovaných gradientů*

## *- porovnání III*

Ukázka konvergence metody konjugovaných gradientů pro Rosenbrockovu funkci:



# *Cvičení*

Proved'te první 2 kroky optimalizace funkce:

$$f(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^2 + 2\mathbf{x}_2^2$$

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:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce:  $s^k = -g^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

Směr v bodě  $x^1$ :  $s^1 = (-2, 0)$

Bod  $x^2$ :  $x^2 = (1, 0) + 0,25 \cdot (-2, 0) = (0,5, 0)$



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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $\alpha = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce:  $s^k = -g^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

Směr v bodě  $x^1$ :  $s^1 = (-2, 0)$

Bod  $x^2$ :  $x^2 = (1, 0) + 0,25 \cdot (-2, 0) = (0,5, 0)$

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

**Beta:**  $\beta_{FR}^{(k)} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}$

Směr funkce:  $s^k = -g^k + \beta^k \cdot s^k$

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

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce:  $s^k = -g^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

Směr v bodě  $x^0$ :  $s^1 = (-2, 0)$

Bod  $x^2$ :  $x^2 = (1, 0) + 0,25 \cdot (-2, 0) = (0,5, 0)$

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

**Beta:**  $\beta_{FR}^{(k)} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}$

Směr funkce:  $s^k = -g^k + \beta^k \cdot s^k$

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

Gradient v bodě  $x^0$ :

Směr v bodě  $x^0$ :

Bod  $x^1$ :

Gradient v bodě  $x^1$ :

**Beta v bodě  $x^1$ :**

**Směr v bodě  $x^1$ :**

**Bod  $x^1$ :**

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce:  $s^k = -g^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

Směr v bodě  $x^0$ :  $s^1 = (-2, 0)$

Bod  $x^1$ :  $x^2 = (1, 0) + 0,25 \cdot (-2, 0) = (0,5, 0)$

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

**Beta:**  $\beta_{FR}^{(k)} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}$

Směr funkce:  $s^k = -g^k + \beta^k \cdot s^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :

**Beta v bodě  $x^1$ :**

**Směr v bodě  $x^1$ :**

**Bod  $x^1$ :**

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

Směr funkce:  $s^k = -g^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

Směr v bodě  $x^0$ :  $s^0 = (-2, 0)$

Bod  $x^1$ :  $x^1 = (1, 0) + 0,25 \cdot (-2, 0) = (0,5, 0)$

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

Výchozí bod:  $x^0 = (2, 1)$

Parametry:  $a = 0.25$

Gradient funkce:  $\nabla f(x_1, x_2) = (2x_1, 4x_2)$

**Beta:**  $\beta_{FR}^{(k)} = \frac{\|g^{(k+1)}\|_2^2}{\|g^{(k)}\|_2^2}$

Směr funkce:  $s^k = -g^k + \beta^k \cdot s^k$

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

Gradient v bodě  $x^0$ :  $g^0 = (4, 4)$

Směr v bodě  $x^0$ :  $s^0 = (-4, -4)$

Bod  $x^1$ :  $x^1 = (2, 1) + 0,25 \cdot (-4, -4) = (1, 0)$

Gradient v bodě  $x^1$ :  $g^1 = (2, 0)$

**Beta v bodě  $x^1$ :**  $b^1 = (2 \cdot 2) / (4 \cdot 4 + 4 \cdot 4) = 4 / 32 = 0,125$

**Směr v bodě  $x^1$ :**  $s^1 = -(2, 0) + 0,125 \cdot (-4, -4) = (-2,5, -0,5)$

**Bod  $x^1$ :**  $x^1 = (1, 0) + 0,25 \cdot (-2,5, -0,5) = (0,375, -0,125)$

## *Domácí úkol*

Proveďte první 3 kroky optimalizace funkce:

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

a) pomocí metody konjugovaných gradientů

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

# *Domácí úkol 2*

Připravte implementaci metody konjugovaných gradientů (např. v Excelu, Pythonu apod.) pro Rosenbrockovu funkci:

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

Gradient Rosenbrockovy funkce:

$$\nabla f(\mathbf{x}) = \left( -400x_1(x_2 - x_1^2) - 2(1 - x_1), \quad 200(x_2 - x_1^2) \right)^T$$

Výchozí bod:

$$\mathbf{x}_0 = (-2, 2)$$

Parametry:

$$\alpha = 0,001$$