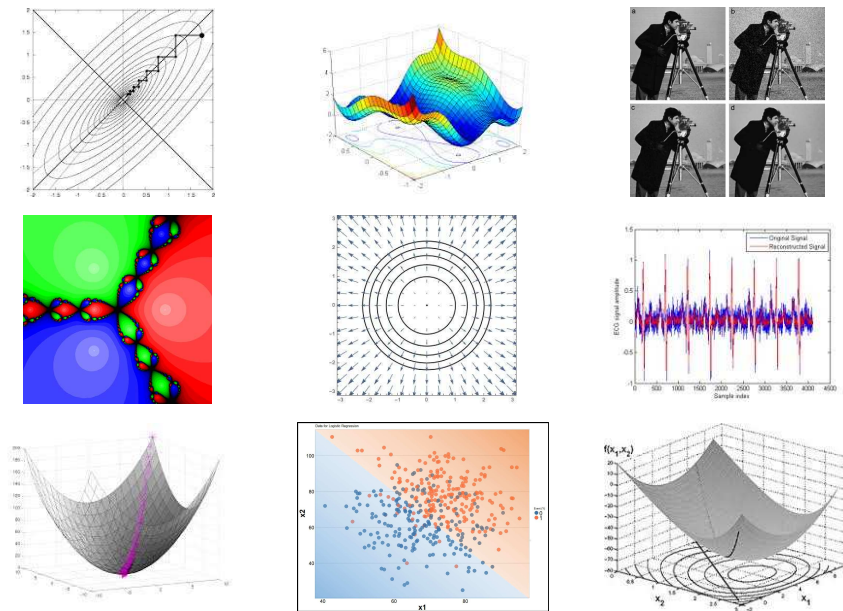
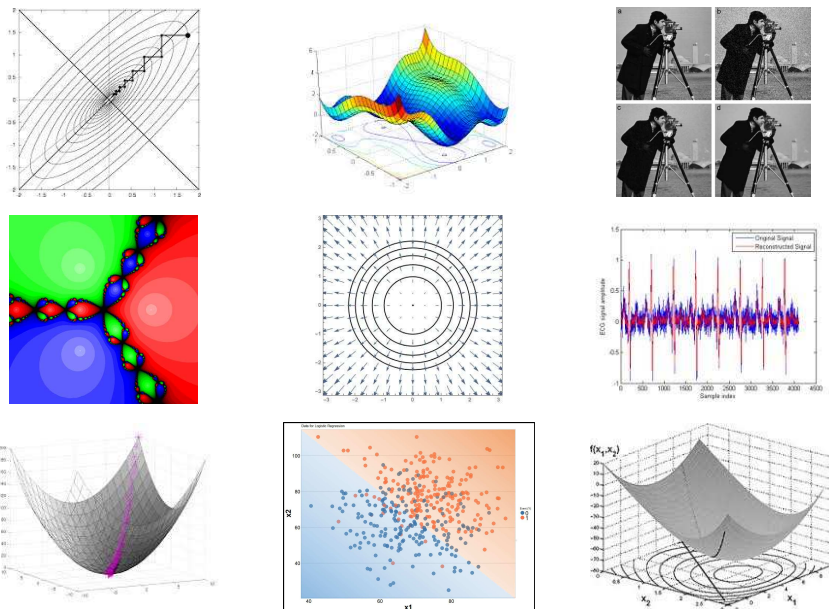




# Metódy voľnej optimalizácie - letný semester 2018

M. Trnovská, KAMŠ, FMFI UK



# Metódy minimalizácie funkcie $n$ premenných - úvod

## Úloha na voľný extrém:

$$\min_{x \in \mathbb{R}^n} f(x) \quad (U1)$$

## Klasifikácia úloh:

- Z hľadiska počtu premenných
  - jednorozmerné
  - viacrozmerné
- Z hľadiska hladkosti funkcie  $f(x)$ 
  - hladké
  - nehladké

## Pripomenieme:

- Úloha (U1) sa zvyčajne rieši nejakým **iteračným algoritmom**, ktorý generuje postupnosť bodov

$$x^0, x^1, x^2, \dots \in \mathbb{R}^n, \quad f(x^k) \rightarrow f(\hat{x}) = \hat{f}, \text{ pre } k \rightarrow \infty$$

- Algoritmus končí v  $\varepsilon$ -presnom riešení, t. j. keď

$$f(x^k) - \hat{f} < \varepsilon \text{ alebo } \|\nabla f(x)\| < \varepsilon.$$

## Klasifikácia metód riešenia

- Od dobrého algoritmu sa spravidla očakáva, že v každej iterácii generuje lepšiu aproximáciu minima ako bola predošlá:

$$f(x^{k+1}) < f(x^k), \quad (k = 0, 1, 2, \dots).$$

Z tohto hľadiska sa metódy delia na

- spádové
- nespádové

- **Z hľadiska použitia informácie rozlišujeme metódy**
  - autonómne (jednokrokové, jednoduché)
  - neautonómne (viackrokové, metódy s pamäťou)

## Klasifikácia metód riešenia

Iteračná formula sa zvyčajne zapisuje v tvare aditívnej korekcie predošlej iterácie, t. j. ako

$$x^{k+1} = x^k + p^k, \quad (k = 0, 1, 2, \dots),$$

alebo ako

$$x^{k+1} = x^k + \lambda_k s^k, \quad (k = 0, 1, 2, \dots),$$

kde  $p^k, s^k$  sú nenulové  $n$ -rozmerné vektory a  $\lambda_k$  je nenulový skalár, pričom

$$\lambda_k s^k = p^k = x^{k+1} - x^k$$

je **vektor korekcie**, ktorý sa určuje podľa schémy:

- I. **bod-smer-krok** (metódy s reguláciou dĺžky kroku)
- II. **bod-krok-smer** (metódy s ohraničeným krokom)

## Klasifikácia metód riešenia

- **Z hľadiska voľby dĺžky kroku:**

- metódy s optimálnou dĺžkou kroku

$$\lambda_k = \arg \min_{\lambda} \varphi(\lambda) = f(x^k + \lambda s_k)$$

- metódy s približne optimálnou dĺžkou kroku
- metódy s konštantnou dĺžkou kroku - kontrola spádovosti

- **Z hľadiska voľby smeru:**

- náhodné smery - stochastické algoritmy
- cyklicky sa volí systém  $n$  lineárne nezávislých smerov
- smer  $s_k$  závisí od  $x^k$  a od minimalizovanej funkcie  $f$ , resp jej gradientu  $\nabla f(x^k)$

## Klasifikácia metód riešenia

- Z hľadiska “rádu” použitej informácie:
  - metódy nultého rádu - univerzálne, pomalá konvergencia
  - metódy prvého rádu - využívajú gradient účelovej funkcie
  - metódy druhého rádu - využívajú aj Hessovu maticu účelovej funkcie, napr. Newtonova metóda



## Metódy s reguláciou dĺžky kroku

### Modelový algoritmus riešenia úlohy (U1)

|                    |                                                                                                                                                                                                                                                                        |
|--------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Vstup:</b>      | Účelová funkcia $f(x)$ úlohy (U1),<br>Štartovací bod iteračného procesu $x^0$ ,<br>Tolerančná konštanta $\varepsilon > 0$<br>Nastavenie počítadla iterácií: $k = 0$                                                                                                    |
| <b>Algoritmus:</b> | 1) Testovanie presnosti. Ak $\ \nabla f(x^k)\  < \varepsilon$ , koniec;<br>2) Výpočet spádového smeru $s^k$ ;<br>3) Výpočet dĺžky kroku $\lambda_k$ ;<br>4) Výpočet novej aproximácie $x^{k+1} = x^k + \lambda_k s^k$ ;<br>5) Opakovanie cyklu. $k = k + 1$ choď na 1) |

## Metódy s reguláciou dĺžky kroku

### Spádové smery

**Definícia:** Majme funkciu  $f(x)$  definovanú na  $\mathbb{R}^n$  a bod  $x^k \in \mathbb{R}^n$ . Smer  $s^k \in \mathbb{R}^n$  nazývame **spádový smer** funkcie  $f(x)$  v bode  $x^k$ , ak existuje kladné číslo  $\bar{\lambda}$  tak, že pre všetky  $0 < \lambda < \bar{\lambda}$  platí

$$f(x^k + \lambda s^k) < f(x^k).$$

**Lema:** Nech funkcia  $f(x)$  definovaná na  $\mathbb{R}^n$  má spojité prvé parciálne derivácie. Potom smer  $s^k \in \mathbb{R}^n$  je spádový v bode  $x^k \in \mathbb{R}^n$ , ak platí

$$\nabla f(x^k)^T s^k < 0.$$

**Dôkaz:** ...

## Metódy s reguláciou dĺžky kroku

Predpokladáme, že funkcia  $f(x)$  má spojité prvé, resp. druhé parciálne derivácie.

### - Cauchyovský smer

$$s_k = -\nabla f(x^k)$$

- Nech  $H_k$  je ľubovoľná kladne definitná matica. Smer

$$s_k = -H_k \nabla f(x^k)$$

je spádový. (**Kvázinewtonovské metódy**)

### - Newtonovský smer

$$s^k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k)$$

## Metódy s reguláciou dĺžky kroku

### Dĺžka kroku

#### 1. Optimálna dĺžka kroku

$$\lambda_k = \arg \min_{\lambda} f(x^k + \lambda s_k)$$

- Presadzovala sa do polovice 60. rokov.
- Algoritmy typu združených smerov - konvergencia za  $n$  krokov pre kvadratickú funkciu  $n$  premenných je ukázaná za predpokladu optimálnej dĺžky kroku.
- Používanie optimálneho kroku (pre všeobecnú nekvadratickú funkciu) vyžaduje riešenie optimalizačnej úlohy v každej iterácii.
- Nie vždy zabezpečuje dobrú konvergenciu.

## Metódy s reguláciou dĺžky kroku

### Optimálna dĺžka kroku pre kvadratickú funkciu

Uvažujme kvadratickú funkciu

$$Q(x) = \frac{1}{2}x^T Gx + h^T x$$

s kladne definitnou maticou  $G$  a gradientom

$$\nabla Q(x) = g(x) = Gx + h.$$

Nech  $s_k$  je spádový smer funkcie  $Q(x)$  v bode  $x^k \in \mathbb{R}^n$  a  $g_k = g(x^k)$ .

Potom optimálna dĺžka kroku  $\lambda_k$  v smere  $s^k$  je daná výrazom

$$\lambda_k = -\frac{g_k^T s_k}{s_k^T G s_k}.$$

## Metódy s reguláciou dĺžky kroku Príklad.

Ani Cauchyovský smer  $s_k = -\nabla f(x^k)$  (smer najväčšieho spádu) v kombinácii s optimálnym krokom nemusí dávať dobrú konvergenciu:

Uvažujme minimalizáciu kvadratickej funkcie

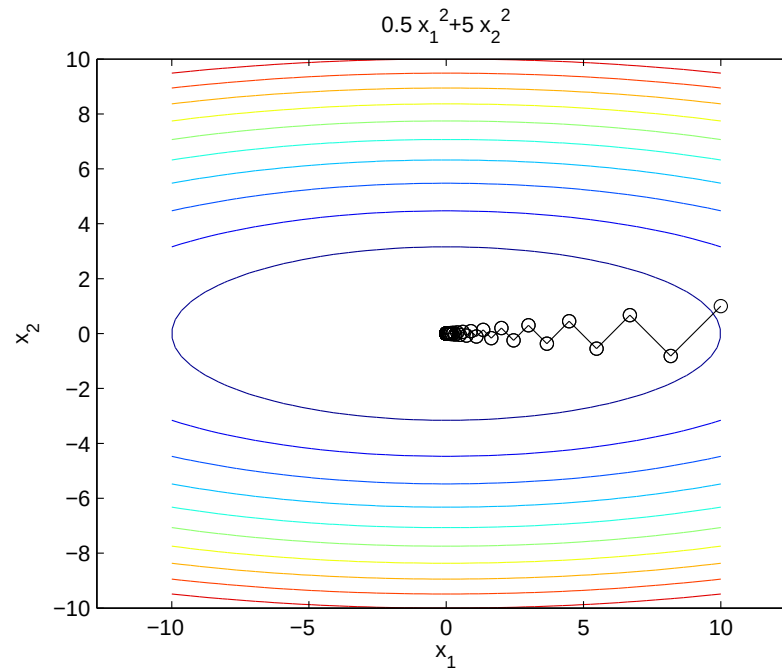
$$Q(x) = \frac{1}{2}(x_1^2 + 10x_2^2)$$

so štartovacím bodom  $x^0 = [10 \ 1]^T$ .

Dá sa ukázať, že  $\lambda_k = 2/11$  v každej iterácii a pre postupnosť  $\{x^k\}$  generovanú týmto algoritmom platí

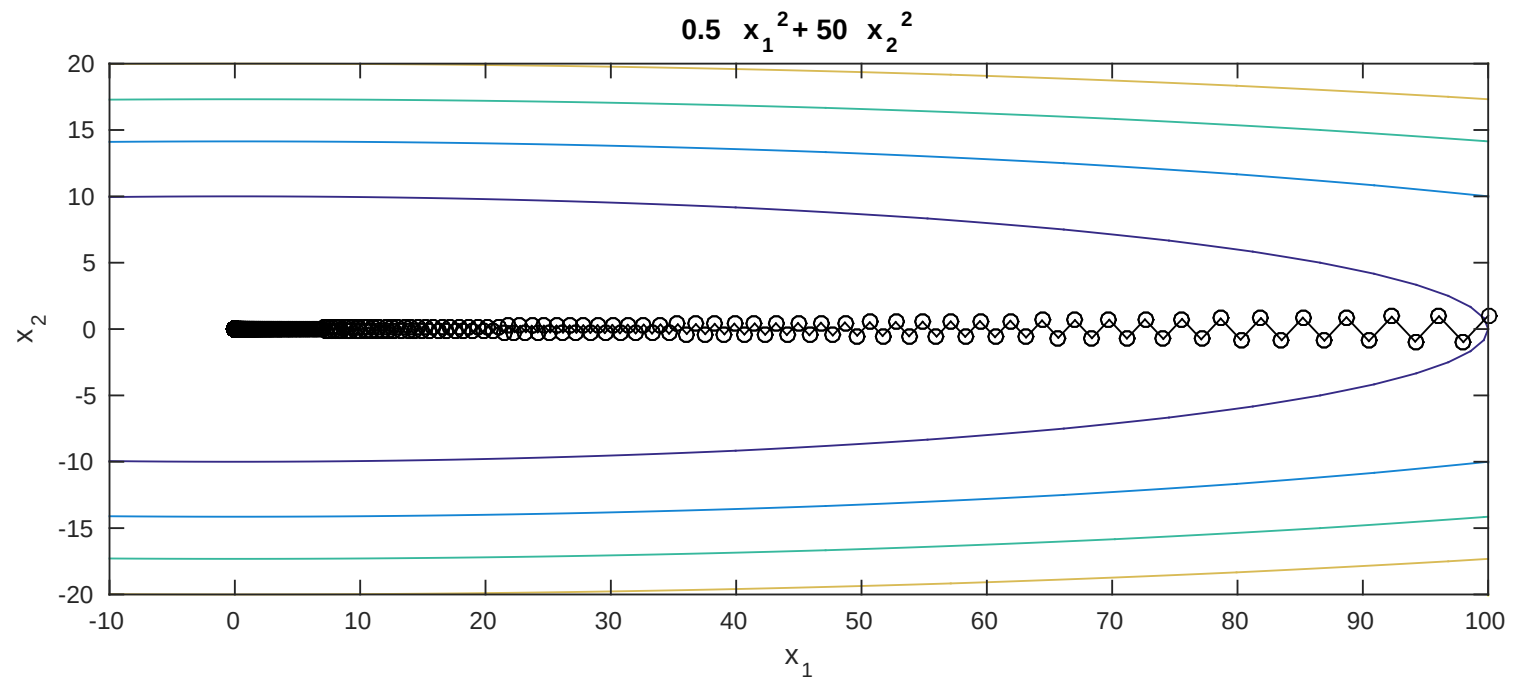
$$x^k = \left(\frac{9}{10}\right)^k \begin{bmatrix} 10 \\ (-1)^k \end{bmatrix}$$

## Metódy s reguláciou dĺžky kroku



Iterácie gradientnej metódy s optimálnym krokom pre funkciu  $Q(x) = \frac{1}{2}(x_1^2 + 10x_2^2)$  so štartovacím bodom  $x^0 = [10 \ 1]^T$ .  
Metóda skonvergovala po 74 iteráciách ( $\varepsilon = 10^{-6}$ ).

## Metódy s reguláciou dĺžky kroku



Iterácie gradientnej metódy s optimálnym krokom pre funkciu  $Q(x) = \frac{1}{2}(x_1^2 + 100x_2^2)$  so štartovacím bodom  $x^0 = [100 \ 1]^T$ .  
Metóda skonvergovala po 743 iteráciách ( $\varepsilon = 10^{-6}$ ).



## Metódy s reguláciou dĺžky kroku

### 2. Približne optimálna dĺžka kroku

Ako zvoliť krok  $\lambda_k$ , aby bola zabezpečená konvergencia metódy?

#### Príklad:

$$f(x) = x^2, \quad x^0 = 2, \quad s^{2k} = -1, \quad s^{2k+1} = 1, \quad \lambda_k = 2 + \frac{3}{2^{k+1}}$$

Použitím iteračnej formule  $x^{k+1} = x^k + \lambda_k s^k$  dostávame:

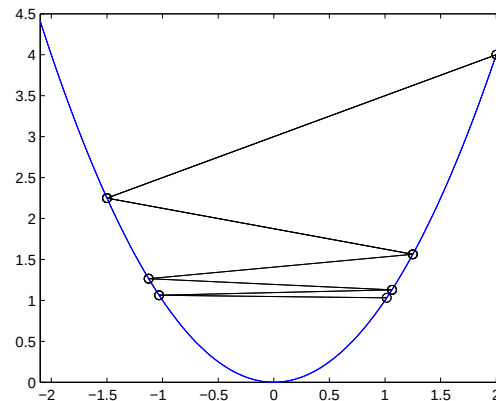
$$\begin{aligned} x^1 &= -1 - 1/2 \\ x^2 &= +1 + 1/4 \\ x^3 &= -1 - 1/8 \\ &\vdots \\ x^k &= (-1)^k \left(1 + \frac{1}{2^k}\right) \end{aligned}$$

## Metódy s reguláciou dĺžky kroku

Postupnosť bodov  $x^k$  má dva hromadné body  $\pm 1$  - teda nekonverguje k minimu, hoci je to spádová metóda:

$$|x^{k+1}| < |x^k| \Rightarrow f(x^{k+1}) < f(x^k)$$

Zvolený krok bol príliš veľký.



## Metódy s reguláciou dĺžky kroku

### Príklad:

$$f(x) = x^2, \quad x^0 = 2, \quad s^k = -1, \quad \lambda_k = \frac{1}{2^{k+1}}$$

Použitím iteračnej formule  $x^{k+1} = x^k + \lambda_k s^k$  dostávame:

$$\begin{aligned} x^1 &= 1 + 1/2 \\ x^2 &= 1 + 1/4 \\ &\vdots \\ x^k &= 1 + \frac{1}{2^k} \end{aligned}$$

Zrejme

$$\lim_{k \rightarrow \infty} x^k = 1, \quad \lim_{k \rightarrow \infty} f(x^k) = 1.$$

Zvolený krok bol príliš malý.

## Metódy s reguláciou dĺžky kroku

### Horná hranica dĺžky kroku (Goldstein):

Zvoľme hodnotu  $0 < \alpha < \frac{1}{2}$ . Nech v bode  $x^k$  je smer  $s^k$  spádový - t.j.  $\nabla f(x^k)^T s^k < 0$ . Potom krok  $\lambda_k$  pokladáme za **dostatočne malý**, ak platí

$$f(x^k + \lambda_k s^k) \leq f(x^k) + \alpha \lambda_k \nabla f(x^k)^T s^k, \quad (G)$$

resp.

$$f_{k+1} \leq f_k + \alpha \nabla f(x^k)^T (x^{k+1} - x^k).$$

## Metódy s reguláciou dĺžky kroku

### Dolná hranica dĺžy kroku (Wolfe-Powell):

Zvoľme hodnotu  $\beta : 0 < \alpha < \beta < 1$ . Nech v bode  $x^k$  je smer  $s^k$  spádový. Potom krok  $\lambda_k$  pokladáme za **dostatočne veľký**, ak platí

$$\nabla f(x^k + \lambda_k s^k)^T s^k \geq \beta \nabla f(x^k)^T s^k, \quad (WP)$$

resp.

$$\nabla f(x^{k+1})^T (x^{k+1} - x^k) \geq \beta \nabla f(x^k)^T (x^{k+1} - x^k).$$

## Metódy s reguláciou dĺžky kroku

### Lema:

Nech funkcia  $f(x)$  definovaná na  $\mathbb{R}^n$  má spojité prvé parciálne derivácie a je na  $\mathbb{R}^n$  zdola ohraničená. Nech  $0 < \alpha < \frac{1}{2}$  a  $\alpha < \beta < 1$ . Nech v bode  $x^k$  je smer  $s^k$  spádový. Potom existuje interval  $[d, h]$ ,  $0 < d < h$  tak, že pre ľubovoľné  $\lambda_k \in [d, h]$  platia súčasne vzťahy (G), (WP).

Pri praktickej realizácii algoritmov stačí za parameter  $\alpha$  zvoliť malú hodnotu, napr.  $10^{-1}$  až  $10^{-5}$ . Parameter  $\beta$  sa spravidla volí z intervalu  $[\alpha, 5\alpha]$ .

## Metódy s reguláciou dĺžky kroku

### Backtracking

- metóda hľadania približne optimálnej dĺžky kroku

### Myšlienka:

Redukovanie funkcie  $\varphi(\lambda) = f(x^k + \lambda s_k)$  pozdĺž polpriamky

$$\{x^k + \lambda s_k \mid \lambda \geq 0\},$$

založené na postupnom overovaní Goldsteinovej podmienky.

Efektívna a v praxi obľúbená technika.

## Metódy s reguláciou dĺžky kroku

### Algoritmus Backtracking

|                    |                                                                                                                                                                                                                                                              |
|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Vstup:</b>      | Účelová funkcia $f(x)$ úlohy (U1),<br>Bod $x = x^k$ a spádový smer $s = s_k$ ,<br>Parameter $\alpha \in (0, \frac{1}{2})$ , v praxi spravidla $\alpha \in [0.01, 0.3]$<br>Parameter redukcie $\delta \in (0, 1)$ , v praxi spravidla $\delta \in [0.1, 0.8]$ |
| <b>Algoritmus:</b> | $\lambda = 1;$<br>while $f(x + \lambda s) > f(x) + \alpha \lambda \nabla f(x)^T s$<br>$\lambda = \delta \lambda;$                                                                                                                                            |

Pre dostatočne malé  $\lambda$  máme

$$f(x + \lambda s) \sim f(x) + \lambda \nabla f(x)^T s < f(x) + \alpha \lambda \nabla f(x)^T s,$$

a teda Backtracking algoritmus napokon skončí.



## Metódy s reguláciou dĺžky kroku

### Konvergenčné vety

**Veta:** (*Wolfe, 1971*)

Nech funkcia  $f(x)$  definovaná na  $\mathbb{R}^n$  má spojité prvé parciálne derivácie, je na  $\mathbb{R}^n$  zdola ohraničená a jej gradient  $\nabla f(x)$  je Lipschitzovský. Potom pre ľubovoľné  $x^0 \in \mathbb{R}^n$  postupnosť bodov  $\{x^k\}$  generovaná iteračným predpisom

$$x^{k+1} = x^k + \lambda_k s^k$$

so spádovými smermi  $s^k$  a dĺžkami krokov  $\lambda_k$  spĺňajúcimi vzťahy (G) a (WP) má nasledovnú vlastnosť: buď pre nejaké  $k > 0$  je  $\nabla f(x^k) = 0$  alebo

$$\lim_{k \rightarrow \infty} \frac{\nabla f(x^k)^T s^k}{\|s^k\|_2} = 0.$$

## Metódy s reguláciou dĺžky kroku

**Veta:** (*Dennis-Moré, 1974*)

Nech funkcia  $f(x)$  definovaná na  $\mathbb{R}^n$  má spojité druhé parciálne derivácie a jej Hessova matica je Lipschitzovská. Nech postupnosť bodov  $\{x^k\}$  generovaná iteračným predpisom

$$x^{k+1} = x^k + \lambda_k s^k$$

so spádovými smermi  $s^k$  a dĺžkami krokov  $\lambda_k$  spĺňajúcimi vzťahy (G) a (WP) konverguje k bodu, v ktorom je Hessova matica kladne definitná a platí

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f(x^k) + \nabla^2 f(x^k) s^k\|_2}{\|s^k\|_2} = 0.$$

Potom existuje index  $k_0 > 0$  taký, že pre  $\forall k > k_0$  krok  $\lambda_k = 1$  spĺňa vzťahy (G) a (WP), pričom rýchlosť konvergenzie postupnosti  $\{x^k\}$  k lokálnemu minimu funkcie  $f(x)$  je superlineárna.

## Metódy s reguláciou dĺžky kroku

**Rád konvergence metódy** - miera efektívností algoritmu

Nech  $\{x^k\}$  je postupnosť bodov konvergujúca k optimálnemu riešeniu.

**Lineárna konvergenca:**  $\exists r \in (0, 1) : \frac{\|x^{k+1} - \hat{x}\|}{\|x^k - \hat{x}\|} \leq r, \quad \forall k \geq K$

**Superlineárna konvergenca:**  $\lim_{k \rightarrow \infty} \frac{\|x^{k+1} - \hat{x}\|}{\|x^k - \hat{x}\|} = 0.$

**Kvadratická konvergenca:**  $\exists r \in (0, 1) : \frac{\|x^{k+1} - \hat{x}\|}{\|x^k - \hat{x}\|^2} \leq r, \quad \forall k \geq K$