Optimalizace

7. Volné lokální extrémy

Tom Werner, Tom Kroupa

FEL ČVUT

Analytické podmínky na volné lokální extrémy

Zopakujme: $\mathbf{x} \in \mathbb{R}^n$ je lokální minimum/maximum funkce $f : \mathbb{R}^n \to \mathbb{R}$, jestliže $\mathbf{x}$ je (globální) minimum/maximum funkce f na kouli $B_\epsilon(\mathbf{x})$ pro nějaké $\epsilon > 0$.

Analytické podmínky na volné lokální extrémy

Zopakujme: $\mathbf{x} \in \mathbb{R}^n$ je lokální minimum/maximum funkce $f\colon \mathbb{R}^n \to \mathbb{R}$, jestliže $\mathbf{x}$ je (globální) minimum/maximum funkce f na kouli $B_\epsilon(\mathbf{x})$ pro nějaké $\epsilon > 0$.

Věta (nutná podmínka prvního řádu)

Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je v bodě $\mathbf{x} \in \mathbb{R}^n$ (totálně) diferencovatelná. Jestliže $\mathbf{x}$ je lokální extrém funkce f (na $\mathbb{R}^n$), pak $\nabla f(\mathbf{x}) = f'(\mathbf{x})^T = \mathbf{0}$.

Bod $\mathbf{x} \in \mathbb{R}^n$, ve kterém $f'(\mathbf{x}) = \mathbf{0}$, se nazývá **stacionární bod** funkce f.

Analytické podmínky na volné lokální extrémy

Zopakujme: $\mathbf{x} \in \mathbb{R}^n$ je lokální minimum/maximum funkce $f : \mathbb{R}^n \to \mathbb{R}$, jestliže $\mathbf{x}$ je (globální) minimum/maximum funkce f na kouli $B_\epsilon(\mathbf{x})$ pro nějaké $\epsilon > 0$.

Věta (nutná podmínka prvního řádu)

Nechť funkce $f : \mathbb{R}^n \to \mathbb{R}$ je v bodě $\mathbf{x} \in \mathbb{R}^n$ (totálně) diferencovatelná.
Jestliže $\mathbf{x}$ je lokální extrém funkce f (na $\mathbb{R}^n$), pak $\nabla f(\mathbf{x}) = f'(\mathbf{x})^T = \mathbf{0}$.

Bod $\mathbf{x} \in \mathbb{R}^n$, ve kterém $f'(\mathbf{x}) = \mathbf{0}$, se nazývá **stacionární bod** funkce f.

Věta (podmínky druhého řádu)

Nechť funkce $f : \mathbb{R}^n \to \mathbb{R}$ je v bodě $\mathbf{x} \in \mathbb{R}^n$ dvakrát diferencovatelná.

- Jestliže $\mathbf{x}$ je lokální minimum [maximum] funkce f,
  pak $f'(\mathbf{x}) = \mathbf{0}$ a Hessova matice $f''(\mathbf{x})$ je positivně [negativně] semidefinitní.
- Jestliže $f'(\mathbf{x}) = \mathbf{0}$ a Hessova matice $f''(\mathbf{x})$ je positivně [negativně] definitní,
  pak $\mathbf{x}$ je ostré lokální minimum [maximum] funkce f.

Z věty plyne, že když $f''(\mathbf{x})$ je indefinitní, pak $\mathbf{x}$ není lokální extrém.
Bod $\mathbf{x}$, ve kterém $f'(\mathbf{x}) = \mathbf{0}$ a $f''(\mathbf{x})$ je indefinitní, se nazývá **sedlový bod** funkce f.

Iterační metody na volné lokální extrémy

Rychlost konvergence iteračních algoritmů

Nechť posloupnost bodů $\mathbf{x}_k \in \mathbb{R}^n$ konverguje k bodu $\mathbf{x}^*$.

Pak posloupnost čísel $r_k = \|\mathbf{x}_k - \mathbf{x}^*\| \geq 0$ konverguje k nule.

Definice

Pokud existuje limita

$$\lim_{k\to\infty} \frac{r_{k+1}}{r_k} = \rho$$

řekneme, že posloupnost $\mathbf{x}_k$ (příp. r_k) konverguje

- **sublineárně**, pokud $\rho = 1$,
- **lineárně**, pokud $0 < \rho < 1$,
- **superlineárně**, pokud $\rho = 0$.

Příklady:

- Posloupnost $r_k = \frac{1}{k} = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right)$ konverguje sublineárně.
- Posloupnost $r_k = 2^{-k} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots\right)$ konverguje lineárně.
- Posloupnost $r_k = 2^{-2^k} = \left(\frac{1}{4}, \frac{1}{16}, \frac{1}{256}, \ldots\right)$ konverguje superlineárně.

Sestupné metody, názvosloví

Hledáme lokální minimum diferencovatelné funkce $f : \mathbb{R}^n \to \mathbb{R}$.

- Zvolíme **počáteční odhad** $\mathbf{x}_0 \in \mathbb{R}^n$
- Tvoříme posloupnost $\mathbf{x}_1, \mathbf{x}_2, \dots$ pomocí iterace

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{v}_k$$

  kde $\mathbf{v}_k$ je **směr hledání** a $\alpha_k > 0$ **délka kroku** v k-té iteraci.

Sestupné metody, názvosloví

Hledáme lokální minimum diferencovatelné funkce $f : \mathbb{R}^n \to \mathbb{R}$.

- Zvolíme **počáteční odhad** $\mathbf{x}_0 \in \mathbb{R}^n$
- Tvoříme posloupnost $\mathbf{x}_1, \mathbf{x}_2, \ldots$ pomocí iterace

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{v}_k$$

  kde $\mathbf{v}_k$ je **směr hledání** a $\alpha_k > 0$ **délka kroku** v k-té iteraci.

Volba směrů $\mathbf{v}_k$ odlišuje různé metody:

- Uvažujeme jen **sestupné metody**, splňující $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ pro každé k.
- $\mathbf{v}_k$ je **sestupný směr** jestliže

$$f_{\mathbf{v}_k}(\mathbf{x}_k) = f'(\mathbf{x}_k)\mathbf{v}_k < 0$$

  Tvrzení: Je-li $\mathbf{v}_k$ sestupný, pak existuje $\alpha_k > 0$ takové, že $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$.
- $\mathbf{v}_k$ může záviset na hodnotách a derivacích funkce f v bodech $\mathbf{x}_k, \mathbf{x}_{k-1}, \ldots$.
  Často závisí jen na $\mathbf{x}_k$.
  Řád metody je nejvyšší řád použitých derivací.

Volba délky kroku α_k

- **Optimální délka kroku (exact line search)**: α_k je globální minimum funkce

$$\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{v}_k)$$

  na intervalu $(0, \infty)$.

Volba délky kroku α_k

- **Optimální délka kroku (exact line search)**: α_k je globální minimum funkce

  $$\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{v}_k)$$

  na intervalu $(0, \infty)$.

- **Přibližně optimální délka kroku (approximate/inexact line search)**:
  Hledáme nějaké $\alpha_k > 0$ tak, že $\varphi(\alpha_k) < \varphi(0)$.

- **Optimální délka kroku (exact line search)**: α_k je globální minimum funkce

$$\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{v}_k)$$

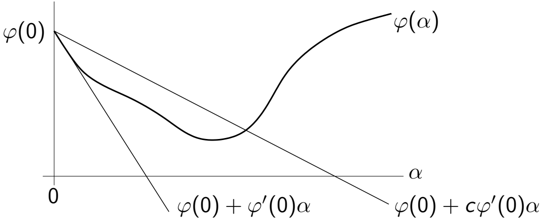
  na intervalu $(0, \infty)$.

- **Přibližně optimální délka kroku (approximate/inexact line search)**: Hledáme nějaké $\alpha_k > 0$ tak, že $\varphi(\alpha_k) < \varphi(0)$.

  Backtracking line search: Začneme od nějakého $\alpha > 0$ a zmenšujeme ho (násobením konstantou), dokud nesplňuje **Armijo-Goldsteinovo pravidlo**

$$\varphi(\alpha) \leq \varphi(0) + c\varphi'(0)\alpha$$

  kde $c \in (0, 1)$ je zvolená konstanta.



- **Pevná posloupnost délek kroku** $(\alpha_k)_{k=0}^{\infty}$. Musí splňovat např.

$$\lim_{k\to\infty} \alpha_k = 0, \qquad \sum_{k=1}^{\infty} \alpha_k = \infty.$$

- **Konstantní délka kroku:** $\alpha_k = \alpha > 0$ pro každé k

Tyto dvě volby negarantují monotónní pokles funkce f.

Gradientní metoda

Gradientní metoda

Iterace: $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k f'(\mathbf{x}_k)^T$

Gradientní metoda

Iterace: $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k f'(\mathbf{x}_k)^T$

- Směr $\mathbf{v}_k = -f'(\mathbf{x}_k)^T = -\nabla f(\mathbf{x}_k)$ je vždy sestupný:

$$f'(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)f'(\mathbf{x}_k)^T = -\|\nabla f(\mathbf{x}_k)\|^2 < 0$$

- Robustní: za velmi obecných předpokladů konverguje, a to obvykle **lineárně**.

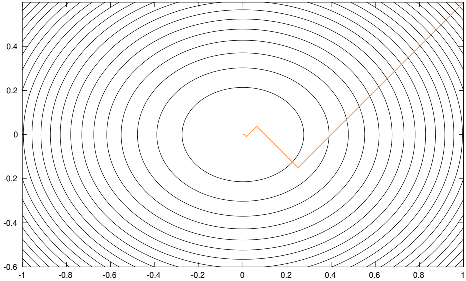
Gradientní metoda

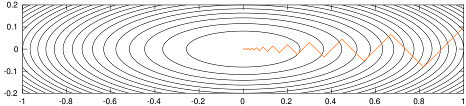
Iterace: $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k f'(\mathbf{x}_k)^T$

- Směr $\mathbf{v}_k = -f'(\mathbf{x}_k)^T = -\nabla f(\mathbf{x}_k)$ je vždy sestupný:

$$f'(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)f'(\mathbf{x}_k)^T = -\|\nabla f(\mathbf{x}_k)\|^2 < 0$$

- Robustní: za velmi obecných předpokladů konverguje, a to obvykle **lineárně**.
- Pomalá konvergence je-li funkce okolo minima protažená (cik-cak chování).





Závislost gradientní metody na lineární transformaci souřadnic

GD v souřadnicích $\mathbf{y} = \mathbf{A}\mathbf{x}$:

Závislost gradientní metody na lineární transformaci souřadnic

GD v souřadnicích $\mathbf{y} = \mathbf{A}\mathbf{x}$:

- Označme

$$g(\mathbf{y}) = f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}), \qquad \text{tedy} \quad g'(\mathbf{y}) = f'(\mathbf{A}^{-1}\mathbf{y})\mathbf{A}^{-1} = f'(\mathbf{x})\mathbf{A}^{-1}$$

- Iterace GD v nových souřadnicích:

$$\mathbf{y}_{k+1} = \mathbf{y}_k - \alpha_k\, g'(\mathbf{y}_k)^T$$

Dosadíme:

$$\begin{aligned} \mathbf{A}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k - \alpha_k\, \mathbf{A}^{-T} f'(\mathbf{x}_k)^T \\ \mathbf{x}_{k+1} &= \mathbf{x}_k - \alpha_k\, (\mathbf{A}^T\mathbf{A})^{-1} f'(\mathbf{x}_k)^T \end{aligned}$$

Závislost gradientní metody na lineární transformaci souřadnic

GD v souřadnicích $\mathbf{y} = \mathbf{A}\mathbf{x}$:

- Označme

$$g(\mathbf{y}) = f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}), \qquad \text{tedy} \quad g'(\mathbf{y}) = f'(\mathbf{A}^{-1}\mathbf{y})\mathbf{A}^{-1} = f'(\mathbf{x})\mathbf{A}^{-1}$$

- Iterace GD v nových souřadnicích:

$$\mathbf{y}_{k+1} = \mathbf{y}_k - \alpha_k \, g'(\mathbf{y}_k)^T$$

Dosadíme:

$$\mathbf{A}\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k - \alpha_k \, \mathbf{A}^{-T} f'(\mathbf{x}_k)^T$$
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \, (\mathbf{A}^T\mathbf{A})^{-1} f'(\mathbf{x}_k)^T$$

Zobecnění GD:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \, \mathbf{C}_k^{-1} f'(\mathbf{x}_k)^T$$

- Je-li $\mathbf{C}_k$ je positivně definitní, směr $\mathbf{v}_k = -\mathbf{C}_k^{-1} f'(\mathbf{x}_k)^T$ bude sestupný.
- Mnoho metod je možno vidět jako zobecněnou GD pro různé volby $\mathbf{C}_k$.
- **Preconditioning:** 'levná' volba $\mathbf{C}_k$ zrychlující konvergenci.

Newtonova metoda

Newtonova metoda na řešení soustavy rovnic

Dáno diferencovatelné zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$, řešíme soustavu rovnic $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Newtonova metoda na řešení soustavy rovnic

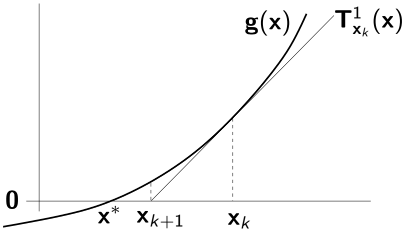
Dáno diferencovatelné zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$, řešíme soustavu rovnic $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Iterace: Pro známé $\mathbf{x}_k$ hledáme $\mathbf{x}_{k+1}$ splňující

$$\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1}) = \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0}$$

To je lineární soustava s řešením

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k)$$



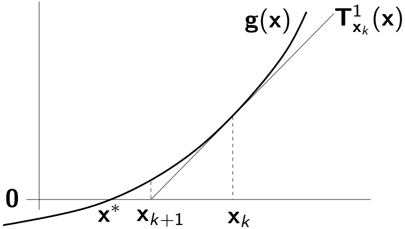
Dáno diferencovatelné zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$, řešíme soustavu rovnic $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Iterace: Pro známé $\mathbf{x}_k$ hledáme $\mathbf{x}_{k+1}$ splňující

$$\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1}) = \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k) = \mathbf{0}$$

To je lineární soustava s řešením

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k)$$



- Konverguje, jestliže počáteční odhad $\mathbf{x}_0$ je ‘dost’ blízko nějakého kořene.
- Obvykle konverguje **superlineárně**.

Superlineární konvergence Newtonovy metody

Věta

Nechť $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{x}^* \in \mathbb{R}^n$, $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$ a $\mathbf{g}'(\mathbf{x}^*)$ je regulární.
Nechť posloupnost $\mathbf{x}_k$ splňující $\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k)$ konverguje k $\mathbf{x}^*$. Pak

$$\lim_{\mathbf{x}_k \to \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 0.$$

Důkaz: Chceme dokázat

$$\lim_{\mathbf{x}_k \to \infty} \frac{\mathbf{x}_{k+1} - \mathbf{x}^*}{\|\mathbf{x}_k - \mathbf{x}^*\|} = \lim_{\mathbf{x}_k \to \infty} \frac{\mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k) - \mathbf{x}^*}{\|\mathbf{x}_k - \mathbf{x}^*\|} = \mathbf{0}.$$

To plyne z následujícího:

$$\begin{aligned}
\lim_{\mathbf{x} \to \mathbf{x}^*} \frac{\mathbf{x} - \mathbf{g}'(\mathbf{x})^{-1}\mathbf{g}(\mathbf{x}) - \mathbf{x}^*}{\|\mathbf{x} - \mathbf{x}^*\|} &= \lim_{\mathbf{x} \to \mathbf{x}^*} \mathbf{g}'(\mathbf{x})^{-1} \frac{\mathbf{g}'(\mathbf{x})(\mathbf{x} - \mathbf{x}^*) - \mathbf{g}(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}^*\|} \\
&= \mathbf{g}'(\mathbf{x}^*)^{-1} \lim_{\mathbf{x} \to \mathbf{x}^*} \frac{\mathbf{g}'(\mathbf{x})(\mathbf{x} - \mathbf{x}^*) - \mathbf{g}(\mathbf{x})}{\|\mathbf{x} - \mathbf{x}^*\|} \\
&= \mathbf{g}'(\mathbf{x}^*)^{-1} \lim_{\mathbf{x} \to \mathbf{x}^*} \frac{\mathbf{g}(\mathbf{x}^*) - \mathbf{g}(\mathbf{x}) - \mathbf{g}'(\mathbf{x})(\mathbf{x}^* - \mathbf{x})}{\|\mathbf{x}^* - \mathbf{x}\|} = \mathbf{0}
\end{aligned}$$

kde poslední limita je nulová z definice derivace (protože $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$).

Příklad: Babylónská metoda na výpočet druhé odmocniny

Řešíme

$$g(x) = x^2 - a = 0$$

Iterace:

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - \frac{x_k^2 - a}{2x_k} = \frac{1}{2}\left(x_k + \frac{a}{x_k}\right)$$

Příklad: Babylónská metoda na výpočet druhé odmocniny

Řešíme

$$g(x) = x^2 - a = 0$$

Iterace:

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - \frac{x_k^2 - a}{2x_k} = \frac{1}{2}\left(x_k + \frac{a}{x_k}\right)$$

Matlab:

```
a = 4;
x = a;
for i=1:5
  x = (x+a/x)/2;
  fprintf('x=%.12g  g=%.12g\n',x,x^2-a);
end
```

```
x=2.5  g=2.25
x=2.05  g=0.2025
x=2.0006097561  g=0.00243939619274
x=2.00000009292  g=3.71689187872e-07
x=2  g=8.881784197e-15
```

Příklad: Průsečík dvou rovinných křivek

Najdi řešení soustavy dvou rovnic o dvou neznámých:

$$\begin{aligned} (x-1)^2 + y^2 &= 1 \\ x^4 + y^4 &= 1 \end{aligned}$$

Příklad: Průsečík dvou rovinných křivek

Najdi řešení soustavy dvou rovnic o dvou neznámých:

$$(x-1)^2 + y^2 = 1$$
$$x^4 + y^4 = 1$$

Je

$$\mathbf{g}(x,y) = \begin{bmatrix} (x-1)^2 + y^2 - 1 \\ x^4 + y^4 - 1 \end{bmatrix} \qquad \mathbf{g}'(x,y) = \begin{bmatrix} 2(x-1) & 2y \\ 4x^3 & 4y^3 \end{bmatrix}$$

Iterace:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 2(x_k-1) & 2y_k \\ 4x_k^3 & 4y_k^3 \end{bmatrix}^{-1} \begin{bmatrix} (x_k-1)^2 + y_k^2 - 1 \\ x_k^4 + y_k^4 - 1 \end{bmatrix}$$

Příklad: Průsečík dvou rovinných křivek

Najdi řešení soustavy dvou rovnic o dvou neznámých:

$$(x-1)^2 + y^2 = 1$$
$$x^4 + y^4 = 1$$

Je

$$\mathbf{g}(x,y) = \begin{bmatrix} (x-1)^2 + y^2 - 1 \\ x^4 + y^4 - 1 \end{bmatrix} \qquad \mathbf{g}'(x,y) = \begin{bmatrix} 2(x-1) & 2y \\ 4x^3 & 4y^3 \end{bmatrix}$$

Iterace:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 2(x_k-1) & 2y_k \\ 4x_k^3 & 4y_k^3 \end{bmatrix}^{-1} \begin{bmatrix} (x_k-1)^2 + y_k^2 - 1 \\ x_k^4 + y_k^4 - 1 \end{bmatrix}$$

Matlab:

```
x = [1;1];
for iter = 1:5
  g = [(x(1)-1)^2+x(2)^2-1; x(1)^4+x(2)^4-1];
  x = x - [2*(x(1)-1) 2*x(2); 4*x'.^3] \ g;
  fprintf('x=(%.12g,%.12g) g=(%.12g,%.12g)\n',x,g);
end

x=(0.75,1)  g=(0.0625,0.31640625)
x=(0.678779069767,0.950944767442)  g=(0.00747883674452,0.0300334591266)
x=(0.671937746776,0.944701508411)  g=(8.57819836066e-05,0.000339082408219)
x=(0.671859761262,0.944629025098)  g=(1.13355711484e-08,4.46057650816e-08)
x=(0.671859751039,0.944629015546)  g=(0,8.881784197e-16)
```

Použití Newtonovy metody na minimalizaci funkce

Hledáme lokální minimum dvakrát diferencovatelné funkce $f : \mathbb{R}^n \to \mathbb{R}$.

Iterace: Řešíme stacionární podmínku $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T = \mathbf{0}$ Newtonovou metodou:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$$

Použití Newtonovy metody na minimalizaci funkce

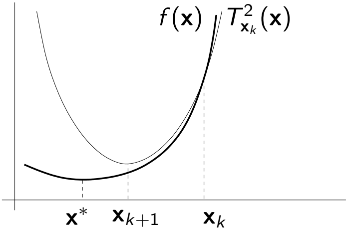
Hledáme lokální minimum dvakrát diferencovatelné funkce $f\colon \mathbb{R}^n \to \mathbb{R}$.

Iterace: Řešíme stacionární podmínku $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T = \mathbf{0}$ Newtonovou metodou:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$$

Jiné odvození: Minimalizujeme Taylorův polynom funkce f stupně 2 v okolí $\mathbf{x}_k$

$$T^2_{\mathbf{x}_k}(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) + f'(\mathbf{x}_k)\,(\mathbf{x}_{k+1} - \mathbf{x}_k) + \tfrac{1}{2}(\mathbf{x}_{k+1} - \mathbf{x}_k)^T f''(\mathbf{x}_k)\,(\mathbf{x}_{k+1} - \mathbf{x}_k)$$



Použití Newtonovy metody na minimalizaci funkce

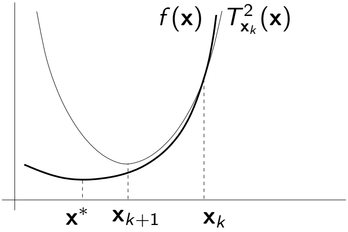
Hledáme lokální minimum dvakrát diferencovatelné funkce $f\colon \mathbb{R}^n \to \mathbb{R}$.

Iterace: Řešíme stacionární podmínku $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T = \mathbf{0}$ Newtonovou metodou:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$$

Jiné odvození: Minimalizujeme Taylorův polynom funkce f stupně 2 v okolí $\mathbf{x}_k$

$$T^2_{\mathbf{x}_k}(\mathbf{x}_{k+1}) = f(\mathbf{x}_k) + f'(\mathbf{x}_k)\,(\mathbf{x}_{k+1} - \mathbf{x}_k) + \tfrac{1}{2}(\mathbf{x}_{k+1} - \mathbf{x}_k)^T f''(\mathbf{x}_k)\,(\mathbf{x}_{k+1} - \mathbf{x}_k)$$



Pro lepší konvergenci můžeme přidat délku kroku:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$$

(Pro $\alpha_k = 1$ se nazývá **čistá** Newtonova metoda.)

Vlastnosti

- **Newtonův směr** $\mathbf{v}_k = -f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$ je sestupný, tj.

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -f'(\mathbf{x}_k) f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T < 0,$$

  když Hessián $f''(\mathbf{x}_k)$ je pozitivně definitní (za předpokladu $f'(\mathbf{x}_k) \neq \mathbf{0}$).

Vlastnosti

- **Newtonův směr** $\mathbf{v}_k = -f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$ je sestupný, tj.

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -f'(\mathbf{x}_k) f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T < 0,$$

  když Hessián $f''(\mathbf{x}_k)$ je pozitivně definitní (za předpokladu $f'(\mathbf{x}_k) \neq \mathbf{0}$).

- Soustava

$$f''(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)^T$$

  má symetrickou matici: můžeme ji řešit Choleského rozkladem.

Kvasi-newtonovy metody

- Počítání Hessiánu $f''(\mathbf{x}_k)$ a řešení soustavy $f''(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)^T$ je drahé.
- Myšlenka kvasi-newtonových metod:
  $f''(\mathbf{x}_k)$ aproximujeme pomocí minulých hodnot $f(\mathbf{x}_i)$ a $f'(\mathbf{x}_i)$, $i \leq k$.
- Speciálně pro $n = 1$ proměnnou: metoda sečen (vs. metoda tečen = Newtonova metoda)
- Populární metoda: **Broyden–Fletcher–Goldfarb–Shanno (BFGS)**
  Matlabská funkce `fminunc`.

Nelineární metody nejmenších čtverců

Nelineární úloha nejmenších čtverců

Pro dané diferencovatelné zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ minimalizujeme funkci

$$f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 = \mathbf{g}(\mathbf{x})^T \mathbf{g}(\mathbf{x}) = \sum_{i=1}^{m} g_i(\mathbf{x})^2$$

- Tedy minimalizujeme součet čtverců daných funkcí.
- Úloha lineárních nejmenších čtverců je speciálním případem, kdy

  $$\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

  je afinní.

Gauss-Newtonova metoda

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ ktere minimalizuje funkci

$$\|\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1})\|^2 = \|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2$$

To je lineární úloha nejmenších čtverců s řešením

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$$

Gauss-Newtonova metoda

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ ktere minimalizuje funkci

$$\|\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1})\|^2 = \|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2$$

To je lineární úloha nejmenších čtverců s řešením

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$$

Můžeme přidat délku kroku:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$$

(Pro $\alpha_k = 1$ se nazývá **čistá** Gauss-Newtonova metoda.)

Gauss-Newtonova metoda

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ ktere minimalizuje funkci

$$\|\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1})\|^2 = \|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2$$

To je lineární úloha nejmenších čtverců s řešením

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$$

Můžeme přidat délku kroku:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$$

(Pro $\alpha_k = 1$ se nazývá **čistá** Gauss-Newtonova metoda.)

- Má-li $\mathbf{g}'(\mathbf{x}_k)$ lin. nezávislé sloupce, Gauss-Newtonův směr lze psát jako

$$\begin{aligned}\mathbf{v}_k &= -\mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)\\ &= -(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k)\\ &= -\tfrac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T \qquad (\text{neboť } f'(\mathbf{x}) = 2\mathbf{g}(\mathbf{x})^T\mathbf{g}'(\mathbf{x}))\end{aligned}$$

- Gauss-Newtonův směr je vždy sestupný:

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -\tfrac{1}{2}f'(\mathbf{x}_k)(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T < 0.$$

Rozdíl oproti Newtonově metodě

Minimalizace funkce $f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2$ Newtonovou vs. Gauss-Newtonovou metodou:

- Gauss-Newtonův směr je

$$\mathbf{v}_k = -\tfrac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k))^{-1} f'(\mathbf{x}_k)^T$$

- Newtonův směr je

$$\begin{aligned}\mathbf{v}_k &= -f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T \\ &= -\tfrac{1}{2}\Big(\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k) + \sum_{i=1}^{m} g_i(\mathbf{x}_k) g_i''(\mathbf{x}_k)\Big)^{-1} f'(\mathbf{x}_k)^T\end{aligned}$$

Rozdíl oproti Newtonově metodě

Minimalizace funkce $f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2$ Newtonovou vs. Gauss-Newtonovou metodou:

- Gauss-Newtonův směr je

$$\mathbf{v}_k = -\tfrac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T$$

- Newtonův směr je

$$\begin{aligned}\mathbf{v}_k &= -f''(\mathbf{x}_k)^{-1}f'(\mathbf{x}_k)^T \\ &= -\tfrac{1}{2}\Big(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \sum_{i=1}^{m} g_i(\mathbf{x}_k)g_i''(\mathbf{x}_k)\Big)^{-1}f'(\mathbf{x}_k)^T\end{aligned}$$

Gauss-Newtonova metoda zanedbává členy druhého řádu v Hessiánu $f''(\mathbf{x})$.

- Proto obvykle konverguje o něco pomaleji než Newtonova metoda.
- Výhoda: nemusíme počítat Hessiány $g_i''(\mathbf{x})$

Příklad: Přibližný průsečík tří křivek

Najdi přibližné řešení soustavy tří rovnic o dvou neznámých:

$$
\begin{aligned}
(x-1)^2 + y^2 &= 1 \\
x^4 + y^4 &= 1 \\
x^2 + (y-1)^2 &= \tfrac{1}{2}
\end{aligned}
$$

Minimalizujeme funkci

$$
\begin{aligned}
f(x,y) &= \mathbf{g}(x,y)^T \mathbf{g}(x,y) \\
&= ((x-1)^2 + y^2 - 1)^2 + (x^4 + y^4 - 1)^2 + (x^2 + (y-1)^2 - \tfrac{1}{2})^2
\end{aligned}
$$

Příklad: Přibližný průsečík tří křivek

Najdi přibližné řešení soustavy tří rovnic o dvou neznámých:

$$\begin{aligned} (x-1)^2 + y^2 &= 1 \\ x^4 + y^4 &= 1 \\ x^2 + (y-1)^2 &= \tfrac{1}{2} \end{aligned}$$

Minimalizujeme funkci

$$\begin{aligned} f(x,y) &= \mathbf{g}(x,y)^T \mathbf{g}(x,y) \\ &= ((x-1)^2 + y^2 - 1)^2 + (x^4 + y^4 - 1)^2 + (x^2 + (y-1)^2 - \tfrac{1}{2})^2 \end{aligned}$$

Je

$$\mathbf{g}(x,y) = \begin{bmatrix} (x-1)^2 + y^2 - 1 \\ x^4 + y^4 - 1 \\ x^2 + (y-1)^2 - \frac{1}{2} \end{bmatrix}, \qquad \mathbf{g}'(x,y) = \begin{bmatrix} 2(x-1) & 2y \\ 4x^3 & 4y^3 \\ 2x & 2(y-1) \end{bmatrix}$$

Iterace čisté Gauss-Newtonovy metody:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 2(x_k-1) & 2y_k \\ 4x_k^3 & 4y_k^3 \\ 2x_k & 2(y_k-1) \end{bmatrix}^+ \begin{bmatrix} (x_k-1)^2 + y_k^2 - 1 \\ x_k^4 + y_k^4 - 1 \\ x_k^2 + (y_k-1)^2 - \frac{1}{2} \end{bmatrix}$$

Matlab:

```
x = [1;1];
for iter = 1:10
  g = [(x(1)-1)^2+x(2)^2-1; x(1)^4+x(2)^4-1; x(1)^2+(x(2)-1)^2-.5];
  x = x - [2*(x(1)-1) 2*x(2); 4*x'.^3; 2*x(1) 2*(x(2)-1)] \ g;
  fprintf('x=(%.12g,%.12g) g=(%.12g,%.12g,%.12g)\n',x,g);
end
```

```
x=(0.75,1) g=(0,1,0.5)
x=(0.696777860013,0.945770115246) g=(0.0625,0.31640625,0.0625)
x=(0.691092552216,0.940578214706) g=(-0.0135752229284,0.0358061117489,-0.0115597333957)
x=(0.691002680826,0.94054818438) g=(-0.0198888107249,0.0107820331379,-0.0188601357041)
x=(0.691002154829,0.940548357781) g=(-0.0198897696029,0.0105634499047,-0.0189807767107)
x=(0.691002152527,0.940548357855) g=(-0.0198891183559,0.0105633328127,-0.0189815242591)
x=(0.691002152516,0.940548357857) g=(-0.0198891167934,0.0105633300224,-0.0189815274491)
x=(0.691002152516,0.940548357857) g=(-0.0198891167821,0.0105633300147,-0.0189815274653)
x=(0.691002152516,0.940548357857) g=(-0.019889116782,0.0105633300146,-0.0189815274654)
x=(0.691002152516,0.940548357857) g=(-0.019889116782,0.0105633300146,-0.0189815274654)
```

Zopakujme:

- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ nabývá minima pro $\mathbf{x} = \mathbf{A}^+\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.
- $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \mu\|\mathbf{x}\|^2$ (kde $\mu > 0$) nabývá minima pro $\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$.
- Matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je regulární pro každou $\mathbf{A}$. Člen $\mu\|\mathbf{x}\|^2$ je regularizátor.

Levenberg-Marquardtova metoda

Zopakujme:

- $\|\mathbf{Ax}-\mathbf{b}\|^2$ nabývá minima pro $\mathbf{x}=\mathbf{A}^+\mathbf{b}=(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.
- $\|\mathbf{Ax}-\mathbf{b}\|^2+\mu\|\mathbf{x}\|^2$ (kde $\mu>0$) nabývá minima pro $\mathbf{x}=(\mathbf{A}^T\mathbf{A}+\mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$.
- Matice $\mathbf{A}^T\mathbf{A}+\mu\mathbf{I}$ je regulární pro každou $\mathbf{A}$. Člen $\mu\|\mathbf{x}\|^2$ je regularizátor.

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ které minimalizuje funkci

$$\|\mathbf{g}(\mathbf{x}_k)+\mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1}-\mathbf{x}_k)\|^2+\ \mu_k\|\mathbf{x}_{k+1}-\mathbf{x}_k\|^2.$$

Ta nabývá minimum pro

$$\mathbf{x}_{k+1}=\mathbf{x}_k-(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)+\ \mu_k\mathbf{I})^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k).$$

Levenberg-Marquardtova metoda

Zopakujme:

- $\|\mathbf{Ax} - \mathbf{b}\|^2$ nabývá minima pro $\mathbf{x} = \mathbf{A}^+\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.
- $\|\mathbf{Ax} - \mathbf{b}\|^2 + \mu\|\mathbf{x}\|^2$ (kde $\mu > 0$) nabývá minima pro $\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$.
- Matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je regulární pro každou $\mathbf{A}$. Člen $\mu\|\mathbf{x}\|^2$ je regularizátor.

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ které minimalizuje funkci

$$\|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2 + \mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

Ta nabývá minimum pro

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k).$$

- Pro $\mu_k \approx 0$ se blíží Gauss-Newtonově iteraci.
- Pro $\mu_k \gg 0$ se blíží iteraci gradientní metody:

$$\mathbf{x}_{k+1} \approx \mathbf{x}_k - \frac{1}{\mu_k}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \frac{1}{2\mu_k}f'(\mathbf{x}_k)^T$$

Levenberg-Marquardtova metoda

Zopakujme:

- $\|\mathbf{Ax} - \mathbf{b}\|^2$ nabývá minima pro $\mathbf{x} = \mathbf{A}^+\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$.
- $\|\mathbf{Ax} - \mathbf{b}\|^2 + \mu\|\mathbf{x}\|^2$ (kde $\mu > 0$) nabývá minima pro $\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}$.
- Matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je regulární pro každou $\mathbf{A}$. Člen $\mu\|\mathbf{x}\|^2$ je regularizátor.

Iterace: Pro známé $\mathbf{x}_k$ najdeme $\mathbf{x}_{k+1}$ které minimalizuje funkci

$$\|\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x}_{k+1} - \mathbf{x}_k)\|^2 + \mu_k\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2.$$

Ta nabývá minimum pro

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k).$$

- Pro $\mu_k \approx 0$ se blíží Gauss-Newtonově iteraci.
- Pro $\mu_k \gg 0$ se blíží iteraci gradientní metody:

$$\mathbf{x}_{k+1} \approx \mathbf{x}_k - \frac{1}{\mu_k}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \frac{1}{2\mu_k}f'(\mathbf{x}_k)^T$$

Heuristika pro volbu μ_k (také nahrazuje line search):

- Pokud $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$, iteraci přijmeme a μ_k zmenšíme (např. 2 krát).
- Pokud $f(\mathbf{x}_{k+1}) \geq f(\mathbf{x}_k)$, iteraci odmítneme a μ_k zvětšíme (např. 2 krát).

Přehled probraných metod

úloha	metoda	iterace
$\min_{\mathbf{x}} f(\mathbf{x})$	gradientní	$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, f'(\mathbf{x}_k)^T$
$\mathbf{g}(\mathbf{x}) = \mathbf{0}$	Newton	$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k)$
$\min_{\mathbf{x}} f(\mathbf{x})$	Newton	$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$
$\min_{\mathbf{x}} \underbrace{\lVert\mathbf{g}(\mathbf{x})\rVert^2}_{f(\mathbf{x})}$	Gauss-Newton	$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, \mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k)$ $= \mathbf{x}_k - \alpha_k\, (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}\, \underbrace{\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k)}_{\frac{1}{2}f'(\mathbf{x}_k)^T}$
$\min_{\mathbf{x}} \lVert\mathbf{g}(\mathbf{x})\rVert^2$	Levenberg-Marquardt	$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k)$

Porovnejte s iterací zobecněné gradientní metody:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, \mathbf{C}^{-1} f'(\mathbf{x}_k)^T$$