# Poznámky k metodě konjugovaných směrů a gradientním metodám

- Máme najít minimum funkce více proměnných. Postupujeme tak, že v daném bodu vždy zvolíme vhodný směr, jednodimenzionální minimalizací nalezneme minimum funkce v tomto směru, v minimu zvolíme nový směr, atd., dokud nedosáhneme minima s požadovanou přesností.

- Odvození definice konjugovaných směrů:

  - Předpokládejme že minimum funkce $f$ je v bodu $\vec{x}$ a my jej začínáme hledat z bodu $\vec{P}$ v jeho okolí. Zavedeme-li vektor $\vec{\delta x} = \vec{x} - \vec{P}$ se složkami $\delta x_i = x_i - P_i$, je Taylorův rozvoj funkce

    $$f\left(\vec{x}\right) = f\big(\vec{P} + \vec{\delta x}\big) = f\big(\vec{P}\big) + \sum_i \delta x_i \frac{\partial f(\vec{P})}{\partial x_i} + \frac{1}{2} \sum_{i,j} \delta x_i\, \delta x_j \frac{\partial^2 f(\vec{P})}{\partial x_i\, \partial x_j} + \ldots.$$

    Pro přehlednost zavedeme vektor $\vec{b}$ a matici $\mathbf{A}$ tak, že

    $$b_i = -\frac{\partial f(\vec{P})}{\partial x_i}, \qquad\qquad A_{ij} = \frac{\partial^2 f(\vec{P})}{\partial x_i\, \partial x_j}$$

    a tedy při zanedbání členů od 3. řádu dále

    $$f\left(\vec{x}\right) \approx f(\vec{P}) - \vec{b}^T \,\vec{\delta x} + \frac{1}{2}\, \vec{\delta x}^T \mathbf{A}\, \vec{\delta x}. \tag{1}$$

    V případě rovnosti říkáme funkci $f(\vec{x})$ *kvadratická forma.*

    * Příklad: Pro jednoduchou funkci

      $$f\left(\vec{x}\right) = f\left(x_1, x_2\right) = x_1^2 + x_2^2 - x_1\, x_2$$

      máme

      $$\frac{\partial f}{\partial x_1} = 2\, x_1 - x_2, \qquad \frac{\partial f}{\partial x_2} = 2\, x_2 - x_1,$$
      $$\frac{\partial^2 f}{\partial x_1^2} = 2, \qquad \frac{\partial^2 f}{\partial x_2^2} = 2,$$
      $$\frac{\partial^2 f}{\partial x_1\, \partial x_2} = -1, \qquad \frac{\partial^2 f}{\partial x_2\, \partial x_1} = -1$$

      a tedy

      $$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} p_2 - 2\, p_1 \\ p_1 - 2\, p_2 \end{pmatrix}, \qquad \text{kde } \vec{P} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}.$$

      Protože třetí a vyšší derivace jsou nulové, jedná se o kvadratickou formu.

  - Gradient funkce $f(x_1, \ldots, x_n)$ je definován jako

    $$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}.$$

- Lze snadno ověřit, že gradient kvadratické formy

$$f(\vec{x}) = f(\vec{P}) - \vec{b}^T \vec{x} + \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x}$$

  v bodu $\vec{x}$ je

$$\nabla f(\vec{x}) = \mathbf{A} \vec{x} - \vec{b}$$

  a při posunu o vektor $\vec{v}$ se gradient změní o

$$\delta(\nabla f)_{\vec{v}} = \nabla f(\vec{x} + \vec{v}) - \nabla f(\vec{x}) = \Big(\mathbf{A}(\vec{x} + \vec{v}) - \vec{b}\Big) - \Big(\mathbf{A} \vec{x} - \vec{b}\Big) = \mathbf{A} \vec{v}.$$

- Předpokládejme že jsme do bodu $\vec{x}$ dorazili při hledání minima ve směru $\vec{u}$. Pak v tomto bodě je gradient funkce $f$ na směr $\vec{u}$ kolmý, tedy

$$\nabla f \perp \vec{u} \qquad \text{neboli} \qquad \vec{u}^T \nabla f = 0.$$

  (Kdyby tomu tak nebylo, měl by gradient nenulovou složku ve směru $\vec{u}$ a bod $\vec{x}$ by tedy nebyl v tomto směru minimem.)
- Dále budeme pokračovat ve směru $\vec{v}$. Ten je výhodné volit tak, aby minimum ve směru $\vec{u}$ zůstalo zachováno, neboli aby

$$0 \stackrel{!}{=} \vec{u}^T \delta(\nabla f)_{\vec{v}} = \vec{u}^T \mathbf{A} \vec{v}.$$

- Pokud toto platí, t.j. pokud $\vec{u}^T \mathbf{A} \vec{v} = 0$, řekneme že **směry** $\vec{u}$ a $\vec{v}$ **jsou konjugované**.

• Přehled názvosloví metod probíraných v přednášce:

  - **Metoda konjugovaných (sdružených) směrů**
    postupuje ve směrech vzájemně konjugovaných, nemusí to být gradienty. V praxi při ní není třeba explicitně počítat vektor $\vec{b}$ ani matici $\mathbf{A}$, tedy žádné parciální derivace funkce, protože konjugovanost vektorů (směrů) je zajištěna způsobem jejich konstrukce.
  - **Metoda nejprudšího spádu**
    postupuje vždy ve směru opačném ke gradientu v daném bodu.
  - **Metoda konjugovaných (sdružených) gradientů**
    postupuje po směrech vzájemně konjugovaných, jednotlivé směry jsou zkombinovány z gradientů v současném bodu a v bodech předchozích.

• Kromě hledání minima funkce ve více dimenzích lze metodu konjugovaných gradientů použít i pro řešení soustavy lineárních rovnic s pozitivně definitní maticí, a to následovně:

  - Hlavní myšlenkou tohoto postupu je místo vlastního řešení soustavy minimalizovat rozdíl aktuálního řešení oproti přesnému. Díky tomu, že řešená soustava je lineární, se vzorce velmi zjednoduší a tím se tento způsob jejího řešení stává efektivním.
  - Výše jsme ukázali, že je-li $f(\vec{x})$ kvadratická forma se symetrickou pozitivně definitní maticí

$$f(\vec{x}) = \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x} - \vec{b}^T \vec{x}, \qquad \forall \vec{x} \neq \vec{0}, \quad (\mathbf{A}\vec{x}, \vec{x}) > 0,$$

    pak v jejím minimu $\tilde{\vec{x}}$ platí

$$\vec{0} = \nabla f\big(\tilde{\vec{x}}\big) = \mathbf{A} \tilde{\vec{x}} - \vec{b} \qquad \text{neboli} \qquad \mathbf{A} \tilde{\vec{x}} = \vec{b}.$$

    Pro řešení lineární soustavy $\mathbf{A} \vec{x} = \vec{b}$ tedy stačí nalézt minimum příslušné funkce $f(\vec{x})$.

– Začneme jej hledat z bodu (počátečního odhadu) $\vec{x}_0$ a budeme postupovat ve směru záporného gradientu

$$\vec{s}_0 = -\nabla f(\vec{x}_0) = -\mathbf{A}\,\vec{x}_0 + \vec{b}$$

do bodu $\vec{x}_1$

$$\vec{x}_1 = \vec{x}_0 + \lambda\,\vec{s}_0, \tag{2}$$

který je minimem v tomto směru. Jak jsme ukázali výše, v bodu $\vec{x}_1$ bude nový (záporný) gradient funkce kolmý ke směru $\vec{s}_0$ (jinak by to nebylo minimum). Tedy

$$\begin{aligned} 0 \;&=\; \vec{s}_0^{\,T}\Big(-\nabla f(\vec{x}_1)\Big) \;=\; \vec{s}_0^{\,T}\Big(-\mathbf{A}\,\vec{x}_1 + \vec{b}\Big) \;=\; \vec{s}_0^{\,T}\Big(-\mathbf{A}\big(\vec{x}_0 + \lambda\,\vec{s}_0\big) + \vec{b}\Big) = \\ &=\; -\vec{s}_0^{\,T}\,\mathbf{A}\,\vec{x}_0 - \lambda\,\vec{s}_0^{\,T}\,\mathbf{A}\,\vec{s}_0 + \vec{s}_0^{\,T}\,\vec{b} \;=\; \vec{s}_0^{\,T}\Big(-\mathbf{A}\,\vec{x}_0 + \vec{b}\Big) - \lambda\,\vec{s}_0^{\,T}\,\mathbf{A}\,\vec{s}_0 = \\ &=\; \vec{s}_0^{\,T}\,\vec{s}_0 - \lambda\,\vec{s}_0^{\,T}\,\mathbf{A}\,\vec{s}_0, \end{aligned}$$

a tedy parametr $\lambda$ určíme ze vztahu

$$\lambda = \frac{\vec{s}_0^{\,T}\,\vec{s}_0}{\vec{s}_0^{\,T}\,\mathbf{A}\,\vec{s}_0}.$$

Jeho dosazením do (2) získáme nový bod $\vec{x}_1$ a z něj budeme postupovat dále ve směru nového záporného gradientu

$$\vec{s}_1 = -\nabla f(\vec{x}_1) = -\mathbf{A}\,\vec{x}_1 + \vec{b}.$$

– Vidíme že nalezení minima ve směru se zde redukuje na násobení vektoru maticí a skalární součiny, přičemž pro řídké matice je násobení vektoru maticí velmi rychlé (řádu $N$ místo $N^2$). Proto se tento postup zvláště hodí pro řešení lineárních soustav s velkou řídkou maticí. Je to v podstatě také iterační metoda.