CG + Gradientenverfahren

<u>Gradientenverfahren am Beispiel Minimierung eines quadratischen Potential</u>

- Minimierung eines Potentials / quadratischen Term eines LGS $\underline{\underline{A}}\,\underline{x} = \underline{y} \Rightarrow f(\underline{x}) = \frac{1}{2}\underline{x}\cdot\underline{\underline{A}}\,\underline{x} - \underline{y}\cdot\underline{x}$
  $\underline{\underline{A}} \in \mathrm{Sym}^+$ (symmetrisch positiv definit) $\longleftrightarrow$ konvexes Potential $f(\underline{x})$
- Minimiere quadrat. Term $\Rightarrow$ Residuum $\underline{r} = -\nabla f(\underline{x}) = \underline{y} - \underline{\underline{A}}\,\underline{x}$ Lösung wenn $\underline{r} = -\nabla f(\underline{x}) = \underline{0}$

  - Iterationsvorschrift:
  - $x_{i+1} = x_i + \alpha r_i = x_i - \alpha \nabla f(x_i)$

  Wie groß darf die Schrittweite α sein?

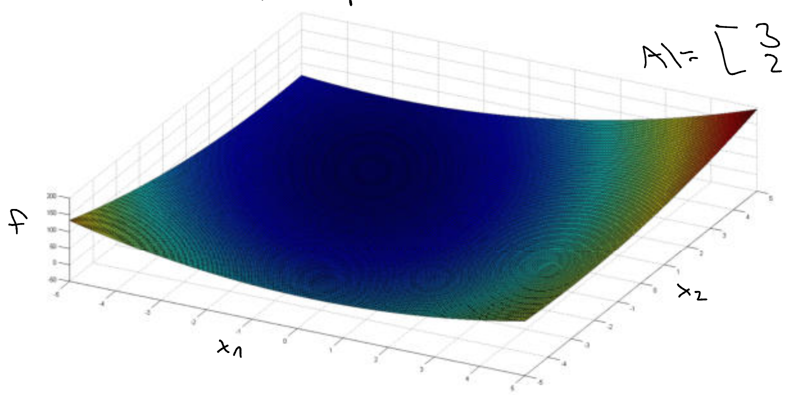
$$\min_\alpha \{f(x_{i+1})\} = \frac{\partial f(x_{i+1})}{\partial x_{i+1}} \cdot \frac{\partial x_{i+1}}{\partial \alpha} = -r_{i+1}\cdot\frac{\partial x_{i+1}}{\partial \alpha} = -r_{i+1}\cdot r_i = 0$$

$\longrightarrow$ $r_{i+1} \perp r_i \longrightarrow$ Orthogonalität der Abstiegsrichtungen

$$-r_{i+1} r_i = \left(\underline{\underline{A}}(x_i + \alpha r_i) - \underline{y}\right)\cdot r_i = 0$$

$$\longrightarrow \quad \alpha = \frac{(\underline{y} - \underline{\underline{A}} x_i)\cdot r_i}{r_i \cdot \underline{\underline{A}} r_i} = \frac{r_i \cdot r_i}{r_i \cdot \underline{\underline{A}} r_i}$$

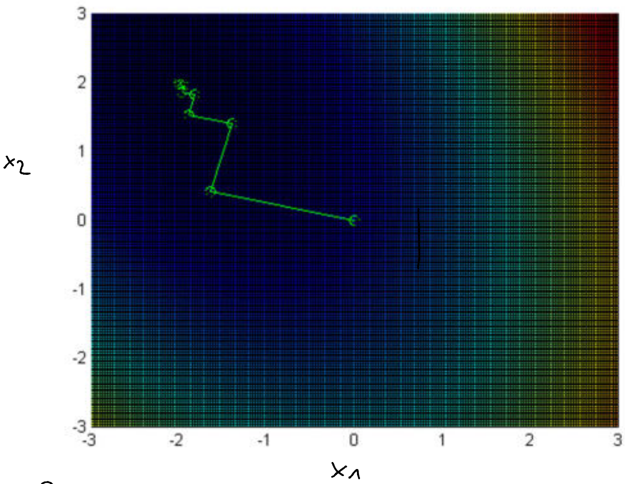
Contourplot



$A1 = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$ $y1 = \begin{bmatrix} 2 \\ -8 \end{bmatrix}$

$x_0 = 10$

Schritte Gradientenverfahren $\Rightarrow$ Suchrichtungen werden mehrmals durchlaufen

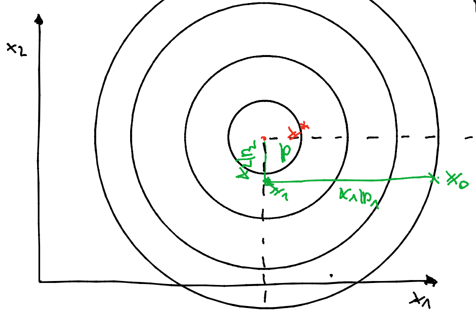


- <u>CG-Verfahren:</u>
  - Lösen eines Gleichungssystems der Form $\underline{\underline{A}}\,\underline{x} = \underline{y}$ $\quad \underline{\underline{A}} \in \mathrm{Sym}^+,\ \lambda_i > 0$
  - Besonders effizient für dünnbesetzte Matrizen / kleine Konditionszahlen $\Rightarrow$ Vorkonditionieren
  - Idee: Minimierung der skalaren quadratischen Form $f(\underline{x}) = \frac{1}{2}\underline{x}\cdot\underline{\underline{A}}\,\underline{x} - \underline{y}\cdot\underline{x} = 0$
- Problem lösbar? $\Rightarrow$ Konvexität der skalaren Funktion $f \longleftrightarrow \underline{\underline{A}} \in$ positiv definit Voraussetzung
  $\nabla^2 f(\underline{x}) = \frac{1}{2}(\underline{\underline{A}}^T + \underline{\underline{A}})$; $\underline{\underline{A}}$ muss positiv definit sein (für 1D-Problem positive Krümmung
  $\Rightarrow$ eindeutiges Minimum)
  $\nabla f(\underline{x}) = \frac{1}{2}(\underline{\underline{A}}^T\underline{x} + \underline{\underline{A}}\,\underline{x}) - \underline{y} = \underline{0}$ nur symmetrischer Teil von $\underline{\underline{A}}$ berücksichtigt $\rightarrow \underline{\underline{A}} \in \mathrm{Sym}$

- Analog zum Gradientenverfahren ergibt sich die Lösung als Linearkombination
  $x_k = x_0 + \sum_{i=1}^{k} \alpha_i p_i$; p_i bilden eine Basis eines über $\underline{A}$ definierten Unterraumes
$\Rightarrow$ Jede beliebige Lösung kann durch Linearkombination der k-Suchrichtungen dargestellt werden

- Vorüberlegung: Welche Eigenschaften können an die Suchrichtungen p_i gestellt werden?
- Für $\underline{A} = \underline{1}$ entspricht $f(\underline{x})$ einem Kreis (2D-Problem)
- Ansatz: Wähle Suchrichtungen, so dass sie den Basisvektoren e_1, e_2 entsprechen $\Rightarrow \underline{x}^* = \alpha_1 e_1 + \alpha_2 e_2 = \alpha_1 p_1 + \alpha_2 p_2$
$\Rightarrow$ für optimal gewähltes α_i sind max. 2 Iterationen bis zur Lösung notwendig

Es gilt $\underline{p}_1 \cdot \underline{p}_2 = 0 \overset{\underline{A}=\underline{1}}{=} \underline{p}_1 \cdot \underline{A}\,\underline{p}_2$



- Was gilt für die Suchrichtungen wenn $\underline{A} \neq \underline{1}$?

$\Rightarrow$ Mache Suchrichtungen A-orthogonal / konjugiert $\Rightarrow$ k = aktueller Rechenschritt

$$\boxed{\underline{p}_i \cdot \underline{A}\,\underline{p}_j = \delta_{ij}\, \underline{p}_i \cdot \underline{A}\,\underline{p}_j} \quad i,j = 1..k \quad (1)$$

Suchrichtungen der anderen Rechenschritte sind orthogonal zur Abbildung von $\underline{A}$ auf $\underline{p}_j \longrightarrow \underline{p}_i$ ist $\underline{A}$-orthogonal / konjugiert

- Für $\underline{A} = \underline{1}$ gilt das Standardskalarprodukt
- Für $\underline{A} \neq \underline{1}$ wird die Streckung / Stauchung / Drehung von $\underline{p}$ durch die Abbildung von $\underline{A}\,\underline{p}_j$ für das Skalarprodukt mitberücksichtigt
  $\Rightarrow$ Skalarprodukt bzgl. eines $\overset{\underline{A}}{\text{nicht}}$-euklidischen Raumes
- $\underline{A} \in sym \longrightarrow \underline{A} = \lambda_i\, \underline{u}_i \underline{u}_i^T \longrightarrow$ Abbildung von $\underline{p}_j$ auf Projektoren (Basis von $\underline{A}$), Streckung, Stauchung von $\underline{p}_j$ über Eigenwerte λ_i
- Konvergenz in $\leq n$ Schritten n = Dimension des Problems

Algorithmus (CG-Verfahren Herleitung)

- Ansatz für Lösung: $\underline{x}_k = \underline{x}_0 + \sum_{i=1}^{k} \alpha_i \underline{p}_i$
  Residuum $\underline{r}_k = \underline{y} - \underline{A}\,\underline{x}_k = -\nabla f(\underline{x})$
- gesucht Lösung: $\underline{x}^*$; Startwert: $\underline{x}^0$
- Ansatz für Abstiegsrichtung $\underline{p}_{k+1}$: $\underline{p}_{k+1} = \underline{r}_k + \sum_{j=1}^{k} \beta_{j+1} \underline{p}_j$
- Bestimmung optimale Schrittweite α_i, $i = 1..k$; $\hat{=}$ beliebigem Iterationsschritt

Minimiere $f(\underline{x}_k) \longrightarrow \min_{\alpha_i} \left\{ f(\underline{x}_k) = \frac{1}{2} \underline{x}_k \cdot \underline{A}\,\underline{x}_k - \underline{y} \cdot \underline{x}_k \right\}$

$$\Rightarrow \frac{\partial f(\underline{x}_k)}{\partial \underline{x}_k} \cdot \frac{\partial \underline{x}_k}{\partial \alpha_i} = \nabla f(\underline{x}_k) \cdot \underline{p}_i = -\underline{r}_k \cdot \underline{p}_i = 0$$

$$\Rightarrow \left(\underline{A}\left(\underline{x}_0 + \sum_i \alpha_i \underline{p}_i\right) - \underline{y} \right) \cdot \underline{p}_i \overset{(1)}{=} \underbrace{\left(\underline{A}\,\underline{x}_0 - \underline{y} \right)}_{-\underline{r}_0} \cdot \underline{p}_i + \underline{A}\, \alpha_i \underline{p}_i \cdot \underline{p}_i = 0$$

$$\longrightarrow \boxed{\alpha_i = \frac{\underline{r}_0 \cdot \underline{p}_i}{\underline{p}_i \cdot \underline{A}\,\underline{p}_i}} \quad (2)$$

Nützliche Beziehungen zwischen $\underline{r}$ und $\underline{p}$:

- Orthogonalität Residuum Abstiegsrichtung: $i = 1..k$

$$\underline{p}_i \cdot \underline{r}_k = \underline{p}_i \cdot \left(\underline{y} - \underline{A}\left(\underline{x}_0 + \sum_{j=1}^{k} \alpha_j \underline{p}_j\right) \right)$$

· Orthogonalität Residuum Abstiegsrichtung: i = 1..k

$$\underline{p}_i \cdot \underline{r}_k = \underline{p}_i \cdot \left(\underline{y} - \underline{\underline{A}}\left(\underline{x}_0 + \sum_{j=1}^{k} \alpha_j \underline{p}_j\right)\right)$$

$$\overset{①}{=} \underline{p}_i \, \underline{r}_0 - \underline{\underline{A}} \, \alpha_i \underline{p}_i \cdot \underline{p}_i = \underline{p}_i \cdot \underline{r}_0 - \underline{\underline{A}} \, \alpha_i \underline{p}_i \cdot \underline{p}_i$$

$$\overset{②}{=} \underline{p}_i \cdot \underline{r}_0 - \frac{\underline{r}_0 \cdot \underline{p}_i}{\underline{p}_i \cdot \underline{\underline{A}} \, \underline{p}_i} \underline{p}_i \cdot \underline{\underline{A}} \, \underline{p}_i = 0 \qquad \boxed{\mathbb{p}_i \cdot \mathbb{r}_k = 0 \quad i = 1..k} \; ③$$

→ Abstiegsrichtungen werden nur 1 mal durchlaufen

· Orthogonalität der Residuen: i = 1..k, j = 1..k

$$\underline{p}_{i+1} = \underline{r}_i + \sum_i \beta_{i+1} \underline{p}_i \rightarrow \underline{r}_i = \underline{p}_{i+1} - \sum \beta_{i+1} \underline{p}_i$$

$$\underline{r}_i \cdot \underline{r}_j = \underline{p}_{i+1} \cdot \underline{r}_j - \beta_{i+1} \underline{p}_i \cdot \underline{r}_j \underset{i<j}{\overset{③}{=}} 0 \qquad \boxed{\mathbb{r}_i \cdot \mathbb{r}_j = \delta_{ij} \, \mathbb{r}_i \cdot \mathbb{r}_j} \; ④$$

$$\cdot \; \underline{r}_i \cdot \underline{\underline{A}} \, \underline{p}_j = (\underline{p}_{i+1} - \beta_{i+1} \underline{p}_i) \cdot \underline{\underline{A}} \, \underline{p}_j \underset{j<i}{\overset{①}{=}} 0$$

$$\boxed{\mathbb{r}_i^T \, \mathbb{A} \, \mathbb{p}_j = 0 \quad i = 1,..,k \quad j<i} \; ⑤$$

· CG-Verfahren Algorithmus: (k = aktuelle Iteration)

$$\underline{x}_{k+1} = \underline{x}_k + \alpha_{k+1} \underline{p}_{k+1} \text{ mit } \underline{p}_{k+1} = \underline{r}_k + \sum_{j=1}^{k} \beta_{j+1} \underline{p}_j$$

Unbekannt: $\underline{p}_{k+1} \,/\, \beta_{k+1}, \; \alpha_{k+1}$

· Forderung aufgrund Konjugiertheit / A-Orthogonalität: $\underline{p}_{k+1} \cdot \underline{\underline{A}} \, \underline{p}_j = 0 \quad j \leq k$

$$\underline{p}_{k+1} \cdot \underline{\underline{A}} \, \underline{p}_j = \underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_j + \sum_{i=1}^{k} \beta_{i+1} \underline{p}_i \cdot \underline{\underline{A}} \, \underline{p}_j = \underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_j + \beta_{i+1} \underline{p}_i \cdot \underline{\underline{A}} \, \underline{p}_j \overset{!}{=} 0$$

$$\rightarrow \begin{cases} j = i < k & \text{da } \underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_j \overset{⑤}{=} 0 \quad \beta_{i+1} \overset{①}{=} 0 \\ j = i = k & \underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_k + \beta_{k+1} \underline{p}_k \cdot \underline{\underline{A}} \, \underline{p}_k = 0 \end{cases} \rightarrow \boxed{\beta_{k+1} = \frac{-\underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_k}{\underline{p}_k \cdot \underline{\underline{A}} \, \underline{p}_k}}$$

$$\Rightarrow \boxed{\underline{p}_{k+1} = \underline{r}_k + \beta_{k+1} \underline{p}_k}$$

· Berechnung α_{k+1}:

$$\alpha_{k+1} = \frac{\underline{r}_0 \cdot \underline{p}_{k+1}}{\underline{p}_{k+1} \cdot \underline{\underline{A}} \, \underline{p}_{k+1}} \qquad \underline{r}_0 \cdot \underline{p}_{k+1} = \left(\underline{y} - \underline{\underline{A}}(\underline{x}_k - \underline{x}_k - \underline{x}_0)\right) \cdot \underline{p}_{k+1} = \left(\underline{r}_k - \underline{\underline{A}}\left(\sum_{i=1}^{k} \alpha_i \underline{p}_i\right)\right) \cdot \underline{p}_{k+1}$$

$$\overset{①}{=} \underline{r}_k \cdot \underline{p}_{k+1} = \underline{r}_k \cdot \left(\underline{r}_k + \sum_{j=1}^{k} \beta_{j+1} \underline{p}_j\right) \overset{③}{=} \underline{r}_k \cdot \underline{r}_k$$

$$\boxed{\alpha_{k+1} = \frac{\underline{r}_k \cdot \underline{r}_k}{\underline{p}_{k+1} \cdot \underline{\underline{A}} \, \underline{p}_{k+1}}}$$

· "Vereinfachung" $\beta_{k+1} = \dfrac{-\underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_k}{\underline{p}_k \cdot \underline{\underline{A}} \, \underline{p}_k}$

$$\underline{r}_k = \underline{r}_{k-1} - \alpha_k \underline{\underline{A}} \, \underline{p}_k \rightarrow \underline{\underline{A}} \, \underline{p}_k = \frac{-\underline{r}_k + \underline{r}_{k-1}}{\alpha_k} \qquad \boxed{\underline{r}_k \cdot \underline{\underline{A}} \, \underline{p}_k = -\frac{1}{\alpha_k} \underline{r}_k \cdot \underline{r}_k} \; ⑥$$

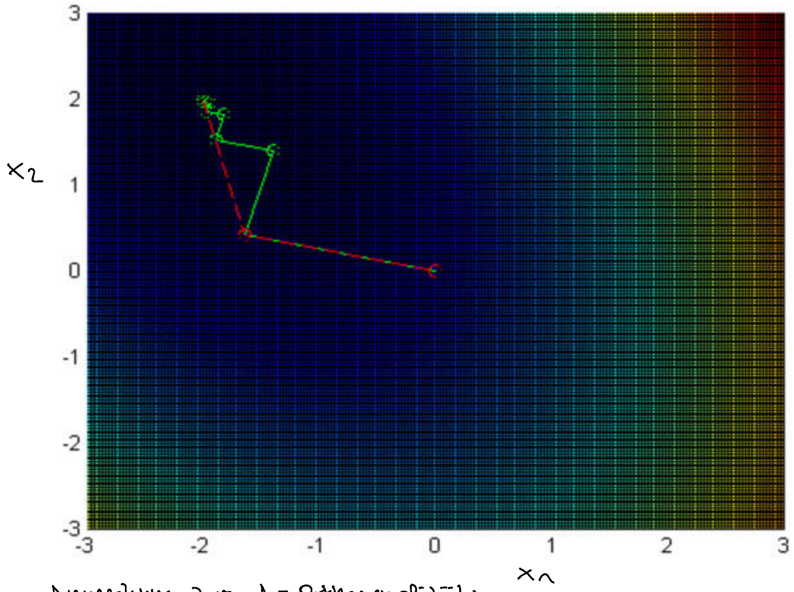
$$\underline{p}_k = \underline{r}_{k-1} + \beta_k \underline{p}_{k-1} \rightarrow \underline{p}_k \cdot \underline{\underline{A}} \, \underline{p}_k = (\underline{r}_{k-1} + \beta_k \underline{p}_{k-1}) \cdot (-\underline{r}_k + \underline{r}_{k-1}) \frac{1}{\alpha_k}$$

$$\overset{②③}{=} \underline{r}_{k-1} \cdot \underline{r}_{k-1} \frac{1}{\alpha_k} \; ⑦$$

$$\boxed{\beta_{k+1} = \frac{\underline{r}_k \cdot \underline{r}_k}{\underline{p}_k \cdot \underline{\underline{A}} \, \underline{p}_k} \overset{⑥⑦}{=} + \frac{\underline{r}_k \cdot \underline{r}_k}{\underline{r}_{k-1} \cdot \underline{r}_{k-1}}}$$

· Vergleich Gradientenverfahren (grün) vs. CG-Verfahren (rot) für obiges 2D-Problem:

– CG-Verfahren konvergiert in 2 Schritten, Gradientenverfahren in ca. 30



– Anmerkung zur A-Orthogonalität:
Würde man den Contourplot so stauchen/strecken, dass die Höhenlinien kreisförmig verlaufen, stünden die Suchrichtungen $\underline{p}_1$ und $\underline{p}_2$ bzgl der e_{x_2}, e_{x_1} Basis senkrecht aufeinander

· Algorithmus CG-Verfahren

$\underline{x}_0 = 0,\ k = 0,\ \underline{r}_0 = \underline{y},\ \underline{p}_0 = \underline{r}_0,\ \beta_1 = 0$

WHILE $\left(\frac{\|\underline{r}_k\|}{\|\underline{y}\|} > TOL \ \&\&\ k < KMAX \right)$

$\underline{p}_{k+1} = \underline{r}_k + \beta_{k+1} \underline{p}_k$

$\alpha_{k+1} = \frac{\underline{r}_k \cdot \underline{r}_k}{\underline{p}_{k+1} \cdot \underline{\underline{A}}\, \underline{p}_{k+1}}$

$\underline{x}_{k+1} = \underline{x}_k + \alpha_{k+1} \underline{p}_{k+1}$

$\underline{r}_{k+1} = \underline{r}_k - \alpha_{k+1} \underline{\underline{A}}\, \underline{p}_{k+1}$

$\beta_{k+1} = \frac{\underline{r}_{k+1} \cdot \underline{r}_{k+1}}{\underline{r}_k \cdot \underline{r}_k}$

$k = k+1$

END

· Algorithmus Gradienten-Verfahren:

$k = 0$

$\underline{x}_0 = \underline{x}_s \quad \underline{x}_s =$ Startwert

$\underline{r}_0 = \underline{y} - \underline{\underline{A}}\, \underline{x}_0$

WHILE $\left(\frac{\|\underline{r}_k\|}{\|\underline{y}\|} > TOL \ \&\&\ k < kmax \right)$

$\alpha_k = \frac{\underline{r}_k \cdot \underline{r}_k}{\underline{r}_k \cdot \underline{\underline{A}}\, \underline{r}_k}$

$\underline{x}_{k+1} = \underline{x}_k + \alpha_k \underline{r}_k$

$\underline{r}_{k+1} = \underline{y} - \underline{\underline{A}}\, \underline{x}_{k+1}$

$k = k+1$

END