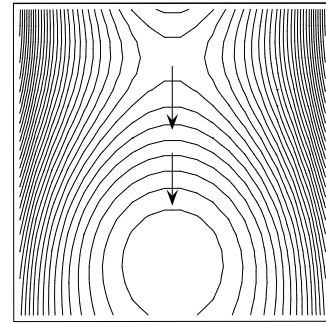
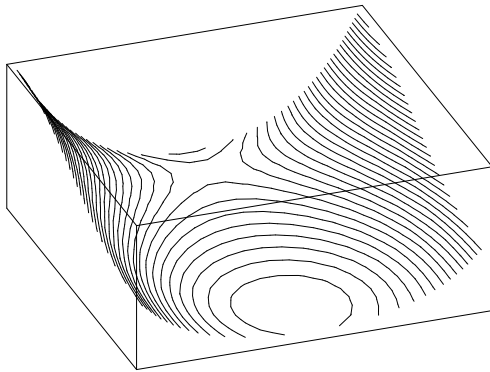


Gradientenverfahren



Mit dem Gradientenverfahren können Minimierungsprobleme bearbeitet werden. Dabei geht man von einem Näherungswert aus und schreitet dann in Richtung des negativen Gradienten, also der Richtung des steilsten Abstiegs, bis man keine Verbesserung mehr erzielt.

Iteration

Startvektor $\vec{x}^{(0)}$

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \vec{d}^{(k)}, \quad \vec{d}^{(k)} = -\nabla f(\vec{x}^{(k)})$$

Verbesserung durch Schrittweitensteuerung

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda^{(k)} \vec{d}^{(k)} \quad \text{mit } \lambda^{(k)} \text{ Minimumstelle von } f(\lambda) = f(\vec{x}^{(k)} + \lambda \vec{d}^{(k)})$$

$$f(x, y) = (x - 3)^2 + 2y^2, \quad \vec{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\nabla f(\vec{x}) = \begin{pmatrix} 2(x - 3) \\ 4y \end{pmatrix}$$

$$\text{Abstiegsrichtung: } \vec{d}^{(0)} = \begin{pmatrix} 4 \\ -4 \end{pmatrix}, \quad \vec{x}^{(0)} + \lambda \vec{d}^{(0)} = \begin{pmatrix} 1 + 4\lambda \\ 1 - 4\lambda \end{pmatrix}, \quad \lambda \text{ statt } \lambda^{(0)}$$

$$f(\lambda) = (1 + 4\lambda - 3)^2 + 2(1 - 4\lambda)^2 \implies \text{Minimum an der Stelle } \lambda = \frac{1}{3}$$

(f vereinfachen, 1. und 2. Ableitung, usw.)

$$\vec{x}^{(1)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 4 \\ -4 \end{pmatrix} = \begin{pmatrix} \frac{7}{3} \\ -\frac{1}{3} \end{pmatrix}$$

$$\text{nächster Iterationsschritt } \vec{d}^{(1)} = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix} \quad \text{usw.}$$

Gradientenverfahren für quadratische Funktionen

Diese Funktionen haben immer die Form

$$f(\vec{x}) = \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x} + \vec{a}^T \vec{x} + \alpha.$$

Es ist dann $\nabla f(\vec{x}) = \mathbf{A} \vec{x} + \vec{a}$.

Mit $\vec{d}^{(k)} = -\nabla f(\vec{x}^{(k)})$

erhalten wir (nach längerer Rechnung, \mathbf{A} wird als symmetrisch vorausgesetzt)

$$\lambda_k = \frac{(\vec{d}^{(k)})^T \cdot \vec{d}^{(k)}}{(\vec{d}^{(k)})^T \cdot \mathbf{A} \cdot \vec{d}^{(k)}}$$

$$f(x, y) = (x - 3)^2 + 2y^2 = \frac{1}{2}(x, y) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (-6, 0) \begin{pmatrix} x \\ y \end{pmatrix} + 9, \quad \vec{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\nabla f(\vec{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -6 \\ 0 \end{pmatrix} = \begin{pmatrix} 2(x - 3) \\ 4y \end{pmatrix}$$

$$\text{Abstiegsrichtung: } \vec{d}^{(0)} = -\nabla f(\vec{x}^{(0)}) = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$$

$$\lambda_0 = \frac{(4, -4) \begin{pmatrix} 4 \\ -4 \end{pmatrix}}{(4, -4) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 4 \\ -4 \end{pmatrix}} = \frac{1}{3}$$

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \lambda^{(0)} \vec{d}^{(0)} = \begin{pmatrix} \frac{7}{3} \\ -\frac{1}{3} \end{pmatrix}$$

$$\text{nächster Iterationsschritt } \vec{d}^{(1)} = -\nabla f(\vec{x}^{(1)}) = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}$$

$$\lambda_1 = \frac{(\frac{4}{3}, \frac{4}{3}) \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}}{(\frac{4}{3}, \frac{4}{3}) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}} = \frac{1}{3}$$

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \lambda_1 \vec{d}^{(1)} = \begin{pmatrix} \frac{25}{9} \\ \frac{1}{9} \end{pmatrix},$$

$$\vec{x}^{(3)} = \vec{x}^{(2)} + \lambda_2 \vec{d}^{(2)} = \begin{pmatrix} \frac{25}{9} \\ \frac{1}{9} \end{pmatrix} + \frac{1}{3} \begin{pmatrix} \frac{4}{9} \\ \frac{4}{9} \end{pmatrix} = \begin{pmatrix} \frac{79}{27} \\ -\frac{1}{27} \end{pmatrix}$$

Methode der konjugierten Gradienten

Eine geschickte Anpassung der Richtungen durch konjugierte Gradienten in jedem Schritt bewirkt, dass das Verfahren nach höchstens n Schritten (\mathbf{A} ist eine $n \times n$ -Matrix, symmetrisch und positiv definit) das Minimum erreicht. Der Start des Verfahrens erfolgt mit dem steilsten Abstieg, während bei allen weiteren Schritten ein Korrekturterm in Abhängigkeit der vorherigen Suchrichtung dazukommt.

$$f(\vec{x}) = \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x} + \vec{a}^T \vec{x} + \alpha \longrightarrow \text{Minimum} \iff \nabla f(\vec{x}) = \mathbf{A} \vec{x} + \vec{a} = \vec{0} \iff \mathbf{A} \vec{x} = \vec{b}$$

Startvektor sei $\vec{x}^{(0)}$, beachte: $\vec{a} = -\vec{b}$.

1. Iteration vorbereiten: $-\nabla f = -(\mathbf{A} \vec{x}^{(0)} + \vec{a}) = \vec{b} - \mathbf{A} \vec{x}^{(0)}$, $\vec{g}^{(0)} = \vec{d}^{(1)} = -\nabla f$

$$\vec{u}^{(1)} = \mathbf{A} \cdot \vec{d}^{(1)} \quad \text{erspart doppelte Berechnung der rechten Seite}$$

$$\lambda_1 = \frac{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}}{(\vec{d}^{(1)})^T \cdot \vec{u}^{(1)}}$$

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \lambda_1 \vec{d}^{(1)}$$

2. Iteration vorbereiten: $\vec{g}^{(1)} = \vec{g}^{(0)} - \lambda_1 \vec{u}^{(1)}$, $\beta_1 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}}$, $\vec{d}^{(2)} = \vec{g}^{(1)} + \beta_1 \vec{d}^{(1)}$

$$\vec{u}^{(2)} = \mathbf{A} \cdot \vec{d}^{(2)}$$

$$\lambda_2 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{d}^{(2)})^T \cdot \vec{u}^{(2)}}$$

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \lambda_2 \vec{d}^{(2)}$$

3. Iteration vorbereiten: $\vec{g}^{(2)} = \vec{g}^{(1)} - \lambda_2 \vec{u}^{(2)}$, $\beta_2 = \frac{(\vec{g}^{(2)})^T \cdot \vec{g}^{(2)}}{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}$, $\vec{d}^{(3)} = \vec{g}^{(2)} + \beta_2 \vec{d}^{(2)}$

$$\vec{u}^{(3)} = \mathbf{A} \cdot \vec{d}^{(3)}$$

$$\lambda_3 = \frac{(\vec{g}^{(2)})^T \cdot \vec{g}^{(2)}}{(\vec{d}^{(3)})^T \cdot \vec{u}^{(3)}}$$

$$\vec{x}^{(3)} = \vec{x}^{(2)} + \lambda_3 \vec{d}^{(3)}$$

Methoden der konjugierten Gradienten

$$(k+1)\text{te Iteration vorbereiten: } \vec{g}^{(k)} = \vec{g}^{(k-1)} - \lambda_k \vec{u}^{(k)}, \quad \beta_k = \frac{(\vec{g}^{(k)})^T \cdot \vec{g}^{(k)}}{(\vec{g}^{(k-1)})^T \cdot \vec{g}^{(k-1)}}, \quad \vec{d}^{(k+1)} = \vec{g}^{(k)} + \beta_k \vec{d}^{(k)}$$

$$\vec{u}^{(k+1)} = \mathbf{A} \cdot \vec{d}^{(k+1)}$$

$$\lambda_{k+1} = \frac{(\vec{g}^{(k)})^T \cdot \vec{g}^{(k)}}{(\vec{d}^{(k+1)})^T \cdot \vec{u}^{(k+1)}}$$

$$\vec{x}^{(k+1)} = \vec{x}^{(k)} + \lambda_{k+1} \vec{d}^{(k+1)}$$

Vektoren $\vec{d}^{(0)}, \vec{d}^{(1)}, \dots, \vec{d}^{(n)}$ heißen \mathbf{A} -konjugiert, wenn gilt

$$(\vec{d}^{(i)})^T \cdot \mathbf{A} \cdot \vec{d}^{(j)} = 0 \quad \text{für } i, j \in \{0, 1, \dots, n\}.$$

Die Begriffsbildung ist eine Erweiterung der Orthogonalität, die mit \mathbf{A} als Einheitsmatrix vorliegt. \mathbf{A} -konjugierte Vektoren sind linear unabhängig.

Methode der konjugierten Gradienten Beispiel

$$\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

$$f(\vec{x}) = \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x} + \vec{a}^T \vec{x} + \alpha \rightarrow \text{Minimum} \iff \nabla f(\vec{x}) = \mathbf{A} \vec{x} + \vec{a} = \vec{0} \iff \mathbf{A} \vec{x} = \vec{b}$$

Startvektor sei $\vec{x}^{(0)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, beachte: $\vec{b} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$.

1. Iteration vorbereiten: $-\nabla f = -(\mathbf{A} \vec{x}^{(0)} + \vec{a}) = \vec{b} - \mathbf{A} \vec{x}^{(0)}$, $\vec{g}^{(0)} = \vec{d}^{(1)} = -\nabla f = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$

$$\vec{u}^{(1)} = \mathbf{A} \cdot \vec{d}^{(1)} = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$$

$$\lambda_1 = \frac{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}}{(\vec{d}^{(1)})^T \cdot \vec{u}^{(1)}} = \frac{9}{9} = 1$$

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \lambda_1 \vec{d}^{(1)} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

2. Iteration vorbereiten: $\vec{g}^{(1)} = \vec{g}^{(0)} - \lambda_1 \vec{u}^{(1)} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $\beta_1 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}} = 1$, $\vec{d}^{(2)} = \vec{g}^{(1)} + \beta_1 \vec{d}^{(1)} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}$

$$\vec{u}^{(2)} = \mathbf{A} \cdot \vec{d}^{(2)} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

$$\lambda_2 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{d}^{(2)})^T \cdot \vec{u}^{(2)}} = \frac{1}{2}$$

$$\vec{x}^{(2)} = \vec{x}^{(1)} + \lambda_2 \vec{d}^{(2)} = \begin{pmatrix} \frac{3}{2} \\ -\frac{9}{2} \end{pmatrix}$$

3. Iteration vorbereiten: $\vec{g}^{(2)} = \vec{g}^{(1)} - \lambda_2 \vec{u}^{(2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Ziel erreicht

$$\begin{pmatrix} 3 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{3}{2} \\ -\frac{9}{2} \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

Methode der konjugierten Gradienten 2. Beispiel

$$f(\vec{x}) = \frac{1}{2} \vec{x}^T \mathbf{A} \vec{x} + \vec{a}^T \vec{x} + \alpha \longrightarrow \text{Minimum} \iff \nabla f(\vec{x}) = \mathbf{A} \vec{x} + \vec{a} = \vec{0} \iff \mathbf{A} \vec{x} = -\vec{a}$$

$$f(x, y) = (x - 3)^2 + 2y^2 = \frac{1}{2} (x, y) \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + (-6, 0) \begin{pmatrix} x \\ y \end{pmatrix} + 9$$

Startvektor sei $\vec{x}^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, beachte: $\vec{a} = -\vec{b} = \begin{pmatrix} -6 \\ 0 \end{pmatrix}$.

1. Iteration vorbereiten: $-\nabla f = -(\mathbf{A} \vec{x}^{(0)} + \vec{a}) = \vec{b} - \mathbf{A} \vec{x}^{(0)}$, $\vec{g}^{(0)} = \vec{d}^{(1)} = -\nabla f = \begin{pmatrix} 4 \\ -4 \end{pmatrix}$

$$\vec{u}^{(1)} = \mathbf{A} \cdot \vec{d}^{(1)} = \begin{pmatrix} 8 \\ -16 \end{pmatrix}$$

$$\lambda_1 = \frac{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}}{(\vec{d}^{(1)})^T \cdot \vec{u}^{(1)}} = \frac{1}{3}$$

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \lambda_1 \vec{d}^{(1)} = \begin{pmatrix} \frac{7}{3} \\ -\frac{1}{3} \end{pmatrix}$$

2. Iteration vorbereiten: $\vec{g}^{(1)} = \vec{g}^{(0)} - \lambda_1 \vec{u}^{(1)} = \begin{pmatrix} \frac{4}{3} \\ \frac{4}{3} \end{pmatrix}$, $\beta_1 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{g}^{(0)})^T \cdot \vec{g}^{(0)}} = \frac{1}{9}$, $\vec{d}^{(2)} = \vec{g}^{(1)} + \beta_1 \vec{d}^{(1)} = \begin{pmatrix} \frac{16}{9} \\ \frac{8}{9} \end{pmatrix}$

$$\vec{u}^{(2)} = \mathbf{A} \cdot \vec{d}^{(2)} = \begin{pmatrix} \frac{32}{9} \\ \frac{32}{9} \end{pmatrix}$$

$$\lambda_2 = \frac{(\vec{g}^{(1)})^T \cdot \vec{g}^{(1)}}{(\vec{d}^{(2)})^T \cdot \vec{u}^{(2)}} = \frac{3}{8}$$

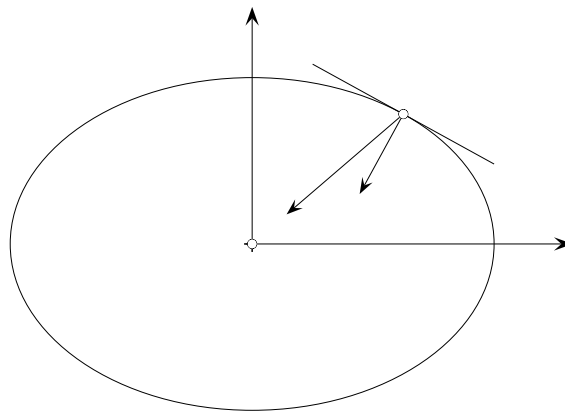
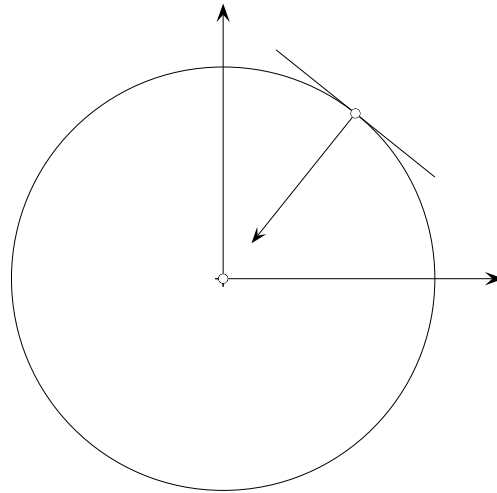
$$\vec{x}^{(2)} = \vec{x}^{(1)} + \lambda_2 \vec{d}^{(2)} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$$

3. Iteration vorbereiten: $\vec{g}^{(2)} = \vec{g}^{(1)} - \lambda_2 \vec{u}^{(2)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Ziel erreicht

$$\begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}$$

Methode der konjugierten Gradienten Idee



\mathbf{A} kann als affine Abbildung gesehen werden. Kreise gehen in Ellipsen über. Für eine optimale Suchrichtung ist das zu berücksichtigen. Sie ist dann nicht mehr orthogonal zur Höhenlinie. Mit \mathbf{A} wird ein Skalarprodukt definiert. Die Orthogonalität bezüglich dieses Skalarprodukts ist der neuen Situation angepasst. Der Algorithmus enthält eine Orthogonalisierung.

