Medical Imaging

Prof. Dr. Tobias Knopp

23. November 2022

Institut für Biomedizinische Bildgebung

Herleitung - CG-Verfahren

Herleitung - CG-Verfahren

Wir betrachten das lineare Gleichungssystem

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \tag{1}$$

$$\boldsymbol{x} = (x_j)_{j=0,\dots,N-1}, \quad \boldsymbol{b} = (b_j)_{j=0,\dots,N-1}, \quad \boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^N, \quad \boldsymbol{A} = (a_{i,j})_{i,j=0,\dots,N-1} \in \mathbb{R}^{N\times N}$$

für eine symmetrisch positiv definite Matrix $\boldsymbol{A}$. Eine Matrix $\boldsymbol{A}$ ist positiv definit falls $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0 \ \forall \boldsymbol{x} \neq \boldsymbol{0}$ gilt. Das lineare Gleichungssystem (1) ist eindeutig lösbar, da jede symmetrisch positiv definite Matrix regulär ist. Wir leiten das CG-Verfahren für reelle Vektoren und Matrizen her.

Herleitung - CG-Verfahren

Die Grundidee des CG-Verfahrens ist folgende. Um das lineare Gleichungssystem (1) zu lösen, suchen wir das Minimum einer Funktion F. Der Minimierer von F ist dann gerade die Lösung des linearen Gleichungssystems. Wir betrachten folgende Funktion $F : \mathbb{R}^N \longrightarrow \mathbb{R}$ mit

$$F(\boldsymbol{x}) := \frac{1}{2}\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle - \langle \boldsymbol{b}, \boldsymbol{x}\rangle.$$

Die Funktion $F(\boldsymbol{x})$ leiten wir nach den Komponenten x_i ab und erhalten

$$\begin{aligned} \frac{\partial F}{\partial x_i} &= \frac{\partial}{\partial x_i}\left(\frac{1}{2}\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}\rangle - \langle \boldsymbol{b}, \boldsymbol{x}\rangle\right) \\ &= \frac{\partial}{\partial x_i}\left(\frac{1}{2}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1} a_{j,k}x_j x_k - \sum_{k=0}^{N-1} b_k x_k\right) \\ &= \sum_{k=0}^{N-1} a_{i,k}x_k - b_i x_i. \end{aligned}$$

Herleitung - CG-Verfahren

Daher ist der Gradient von $F(\boldsymbol{x})$ durch

$$\nabla F(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$$

gegeben. Falls $\boldsymbol{x} = \boldsymbol{x}^*$ die Lösung des linearen Gleichungssystems (1) ist, dann ist der Gradient $\nabla F(\boldsymbol{x}^*) = 0$. Somit ist $F(\boldsymbol{x}^*)$ ein Extremum. Wir betrachten die Hessematrix von $F(\boldsymbol{x})$

$$(\mathcal{H}F(\boldsymbol{x}))_{i,j} \quad = \quad \frac{\partial}{\partial x_j} \sum_{k=0}^{N-1} a_{i,k} x_k - b_i x_i = a_{i,j}.$$

Es ist also die Hessematrix $\mathcal{H}F(\boldsymbol{x}) = \boldsymbol{A}$. Daraus folgt nach Voraussetzung, dass die Hessematrix von $F(\boldsymbol{x})$ symmetrisch positiv definit ist. Somit ist $F(\boldsymbol{x}^*)$ ein globales Minimum.

Um nun die Lösung von (1) zu berechnen, suchen wir den Minimierer von $F(\boldsymbol{x})$. Den Minimierer werden wir iterativ bestimmen, indem wir von einem Startvektor $\boldsymbol{x}^{(0)}$ ausgehen und geeignete Abstiegsvektoren $\boldsymbol{v}_k$ nutzen

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k \quad \text{mit} \quad \alpha_k \in \mathbb{R}. \tag{2}$$

Das Residuum $\boldsymbol{r}_k$ sei durch

$$\boldsymbol{r}_k := \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k \tag{3}$$

definiert.

Lemma

Das Minimum von $F(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k)$ *wird bei festen Vektoren* $\boldsymbol{x}_k$ *und* $\boldsymbol{v}_k$ *durch*

$$\alpha_k = \frac{\langle \boldsymbol{r}_k, \boldsymbol{v}_k \rangle}{\langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle} \tag{4}$$

erreicht. Das Residuum $\boldsymbol{r}_k$ *kann rekursiv durch*

$$\boldsymbol{r}_k = \boldsymbol{r}_{k-1} - \alpha_{k-1} \boldsymbol{A}\boldsymbol{v}_{k-1}. \tag{5}$$

berechnet werden.

Beweis: Wir setzen die partiellen Ableitungen

$$
\begin{aligned}
\frac{\partial}{\partial \alpha_k} F(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) &= \frac{\partial}{\partial \alpha_k} \left(\frac{1}{2} \langle \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k, \boldsymbol{A}(\boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k) \rangle - \langle \boldsymbol{b}, \boldsymbol{x}_k + \alpha_k \boldsymbol{v}_k \rangle \right) \\
&= \frac{\partial}{\partial \alpha_k} \left(\frac{1}{2} \Big(\langle \boldsymbol{x}_k, \boldsymbol{A}\boldsymbol{x}_k \rangle + \alpha_k \langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{x}_k \rangle + \alpha_k \langle \boldsymbol{x}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle + \right. \\
&\qquad \left. \alpha_k^2 \langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle \Big) - \langle \boldsymbol{b}, \boldsymbol{x}_k \rangle - \alpha_k \langle \boldsymbol{b}, \boldsymbol{v}_k \rangle \right) \\
&= \langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{x}_k \rangle + \alpha_k \langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle - \langle \boldsymbol{b}, \boldsymbol{v}_k \rangle
\end{aligned}
$$

gleich Null.

Herleitung - CG-Verfahren

Dann lösen wir nach α_k auf, so dass wir

$$\alpha_k = \frac{\langle \boldsymbol{b}, \boldsymbol{v}_k \rangle - \langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{x}_k \rangle}{\langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle}$$

erhalten. Mit (3) folgt die Behauptung (4). Für das Residuum $\boldsymbol{r}_k$ gilt

$$\begin{aligned} \boldsymbol{r}_k &= \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k \\ &\stackrel{(2)}{=} \boldsymbol{b} - \boldsymbol{A}\left(\boldsymbol{x}_{k-1} + \alpha_{k-1}\boldsymbol{v}_{k-1}\right) \\ &= \boldsymbol{r}_{k-1} - \alpha_{k-1}\boldsymbol{A}\boldsymbol{v}_{k-1}, \end{aligned}$$

womit Behauptung (5) folgt. ■

Jetzt stellt sich die Frage welche Abstiegsrichtungen $\boldsymbol{v}_k$ wir wählen. $\boldsymbol{A}$-orthogonale Vektoren $\boldsymbol{v}_k \neq 0$ $(k = 0\ldots, N-1)$ haben sich als günstig erwiesen. Zwei Vektoren $\boldsymbol{c} \in \mathbb{R}^N$ und $\boldsymbol{d} \in \mathbb{R}^N$ heißen $\boldsymbol{A}$-orthogonal falls

$$\langle \boldsymbol{c}, \boldsymbol{A}\boldsymbol{d} \rangle = 0$$

ist. Es gilt nun folgendes Lemma.

Lemma

Seien die Vektoren $\boldsymbol{v}_j$ $(j = 0\ldots, N-1)$ paarweise A-orthogonal, dann ist

$$\langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle = 0 \qquad (k = 1, \ldots, N)\,(j = 0, \ldots, k-1) \tag{6}$$

und es gilt

$$\boldsymbol{x}^{(N)} = \boldsymbol{x}^* = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{7}$$

Beweis: Durch vollständige Induktion über den Index k beweisen wir Gleichung (6).

Der Induktionsanfang für $k = 1$ ist mit Gleichung (4) und (5) durch

$$\begin{aligned}\langle \boldsymbol{v}_0, \boldsymbol{r}_1 \rangle &= \langle \boldsymbol{v}_0, \boldsymbol{r}_0 - \alpha_0 \boldsymbol{A}\boldsymbol{v}_0 \rangle \\ &= \langle \boldsymbol{v}_0, \boldsymbol{r}_0 \rangle - \frac{\langle \boldsymbol{r}_0, \boldsymbol{v}_0 \rangle}{\langle \boldsymbol{v}_0, \boldsymbol{A}\boldsymbol{v}_0 \rangle} \langle \boldsymbol{v}_0, \boldsymbol{A}\boldsymbol{v}_0 \rangle \ = \ 0.\end{aligned}$$

gegeben. Es gelte nun für einen Index $k \geq 1$

$$\langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle = 0 \qquad (j = 0 \ldots, k-1),$$

dann gilt für das Skalarprodukt mit Gleichung (4) und (5) die Beziehung

$$\begin{aligned}\langle \boldsymbol{v}_j, \boldsymbol{r}_{k+1} \rangle &= \langle \boldsymbol{v}_j, \boldsymbol{r}_k - \alpha_k \boldsymbol{A}\boldsymbol{v}_k \rangle \\ &= \langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle - \alpha_k \langle \boldsymbol{v}_j, \boldsymbol{A}\boldsymbol{v}_k \rangle \\ &= \langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle - \frac{\langle \boldsymbol{v}_k, \boldsymbol{r}_k \rangle}{\langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle} \langle \boldsymbol{v}_j, \boldsymbol{A}\boldsymbol{v}_k \rangle. \qquad (8)\end{aligned}$$

Falls $j \neq k$ ist das Skalarprodukt $\langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle$ wegen der Induktionsvoraussetzung und das Skalarprodukt $\langle \boldsymbol{v}_j, \boldsymbol{A}\boldsymbol{v}_k \rangle$ wegen der $\boldsymbol{A}$-Orthogonalität gleich Null. Für den Fall $j = k$ kürzt sich der der hintere Bruch in Gleichung (8) wodurch $\langle \boldsymbol{v}_j, \boldsymbol{r}_k \rangle - \langle \boldsymbol{v}_k, \boldsymbol{r}_k \rangle = 0$ folgt. Somit ist das Skalarprodukt $\langle \boldsymbol{v}_j, \boldsymbol{r}_{k+1} \rangle$ gleich Null für $(j = 0, \ldots, k)$.

Der zweite Teil des Lemmas 2, die Gleichung (7), folgt aus Gleichung (6), denn es gilt

$$\langle \boldsymbol{v}_j, \boldsymbol{r}_N \rangle = 0 \quad (j = 0, \ldots, N-1).$$

Daraus folgt, dass das Residuum $\boldsymbol{r}_N = \boldsymbol{0}$ sein muss, da die Vektoren $\boldsymbol{v}_j$ paarweise orthogonal und ungleich dem Nullvektor sind. Dies liegt daran, dass es in einem N-dimensionalen Vektor-Raum nur N paarweise orthogonale vom Nullvektor verschiedene Vektoren geben kann. Es gilt also $\boldsymbol{A}\boldsymbol{x}^{(N)} - \boldsymbol{b} = \boldsymbol{0}$. ∎

Herleitung - CG-Verfahren

Nach N Schritten liefert das CG-Verfahren wegen Beziehung (7) die richtige Lösung. Um die paarweise $\boldsymbol{A}$-orthogonalen Vektoren $\boldsymbol{v}_k$ zu erhalten wird das Orthogonalisierungsverfahren nach Gram-Schmidt auf die Basis $\boldsymbol{r}_0, \dots, \boldsymbol{r}_{N-1}$ angewendet. Der Startvektor ist durch

$$\boldsymbol{v}_0 = \boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(0)}$$

gegeben. Die paarweise $\boldsymbol{A}$-orthogonalen Vektoren $\boldsymbol{v}_k$ entstehen nach dem Orthogonalisierungsverfahren durch

$$\boldsymbol{v}_k = \boldsymbol{r}_k - \sum_{j=0}^{k-1} b_{k,j}\boldsymbol{v}_j \quad \text{mit} \quad b_{k,j} = \frac{\langle \boldsymbol{r}_k, \boldsymbol{A}\boldsymbol{v}_j \rangle}{\langle \boldsymbol{v}_j, \boldsymbol{A}\boldsymbol{v}_j \rangle}. \tag{9}$$

Die Summe in (9) hat höchstens einen Summanden ungleich Null. Um dies zu beweisen, benötigen wir folgendes Lemma.

Lemma
Seien die Vektoren $\boldsymbol{v}_k$ und $\boldsymbol{r}_k$ nach der Vorschrift (9) konstruiert. Dann gilt

$$\langle \boldsymbol{r}_k, \boldsymbol{v}_k \rangle = \langle \boldsymbol{r}_k, \boldsymbol{r}_k \rangle \tag{10}$$

und

$$\langle \boldsymbol{r}_k, \boldsymbol{r}_l \rangle = 0 \qquad (l = 0, \ldots, k-1). \tag{11}$$

Beweis: Die Gleichung (10) zeigen wir, indem wir die Ansatzgleichung (9) umformen. Durch

$$\boldsymbol{v}_k = \boldsymbol{r}_k - \sum_{j=0}^{k-1} b_{k,j}\boldsymbol{v}_j$$

folgt, indem wir mit $\boldsymbol{r}_k^T$ multiplizieren

$$\boldsymbol{r}_k^T\boldsymbol{v}_k = \boldsymbol{r}_k^T\boldsymbol{r}_k - \sum_{j=0}^{k-1} b_{k,j}\boldsymbol{r}_k^T\boldsymbol{v}_j.$$

Dies ergibt

$$\langle\boldsymbol{r}_k,\boldsymbol{v}_k\rangle = \langle\boldsymbol{r}_k,\boldsymbol{r}_k\rangle - \sum_{j=0}^{k-1} b_{k,j} \underbrace{\langle\boldsymbol{r}_k,\boldsymbol{v}_j\rangle}_{(= \quad 0 \quad \text{wegen (6)})} = \langle\boldsymbol{r}_k,\boldsymbol{r}_k\rangle.$$

Herleitung - CG-Verfahren

Analog beweisen wir Gleichung (11). Durch Umformung der Gleichung (9) ist

$$\boldsymbol{r}_l \quad = \quad \boldsymbol{v}_l - \sum_{j=0}^{l-1} b_{l,j} \boldsymbol{v}_j .$$

Indem wir die Gleichung mit $\boldsymbol{r}_k^T$ multiplizieren, folgt

$$\boldsymbol{r}_k^T \boldsymbol{r}_l \quad = \quad \boldsymbol{r}_k^T \boldsymbol{v}_l - \sum_{j=0}^{k-1} b_{k,j} \boldsymbol{r}_k^T \boldsymbol{v}_j .$$

Nun schreiben wir die Vektor-Multiplikationen als Skalarprodukt, wodurch

$$\langle \boldsymbol{r}_k, \boldsymbol{r}_l \rangle \quad = \quad \underbrace{\langle \boldsymbol{r}_k, \boldsymbol{v}_l \rangle}_{(= \quad 0 \quad \text{wegen (6)})} - \sum_{j=0}^{k-1} b_{k,j} \underbrace{\langle \boldsymbol{r}_k, \boldsymbol{v}_j \rangle}_{(= \quad 0 \quad \text{wegen (6)})} = 0$$

und somit Gleichung (11) folgt. ■

Lemma

Die Vektoren $\boldsymbol{v}_{k+1}$ $(k = 0, \ldots, N-1)$ *lassen sich durch*

$$\boldsymbol{v}_{k+1} = \boldsymbol{r}_{k+1} + \beta_k \boldsymbol{v}_k \quad \textit{mit} \quad \beta_{k-1} = \frac{\langle \boldsymbol{r}_k, \boldsymbol{r}_k \rangle}{\langle \boldsymbol{r}_{k-1}, \boldsymbol{r}_{k-1} \rangle} \tag{12}$$

berechnen. Für die skalaren Werte α_k *gilt*

$$\alpha_k = \frac{\langle \boldsymbol{r}_k, \boldsymbol{r}_k \rangle}{\langle \boldsymbol{v}_k, \boldsymbol{A}\boldsymbol{v}_k \rangle}. \tag{13}$$

Beweis: Aus der Rekursionsgleichung (5) folgt

$$\boldsymbol{A}\boldsymbol{v}_k = \frac{\boldsymbol{r}_k - \boldsymbol{r}_{k+1}}{\alpha_k}.$$

Somit gilt für die Koeffizienten

$$\begin{aligned} b_{k,j} &= \frac{\left\langle \boldsymbol{r}_k, \frac{\boldsymbol{r}_j - \boldsymbol{r}_{j+1}}{\alpha_j} \right\rangle}{\left\langle \boldsymbol{v}_j, \frac{\boldsymbol{r}_j - \boldsymbol{r}_{j+1}}{\alpha_j} \right\rangle} \\ &= \frac{\langle \boldsymbol{r}_k, \boldsymbol{r}_j \rangle - \langle \boldsymbol{r}_k, \boldsymbol{r}_{j+1} \rangle}{\langle \boldsymbol{v}_j, \boldsymbol{r}_j \rangle - \langle \boldsymbol{v}_j, \boldsymbol{r}_{j+1} \rangle}. \end{aligned}$$

Für den ersten Fall $l < k - 1$ ist nun

$$b_{k,j} = \frac{\overbrace{\langle \boldsymbol{r}_k, \boldsymbol{r}_j \rangle}^{= \quad 0} - \overbrace{\langle \boldsymbol{r}_k, \boldsymbol{r}_{j+1} \rangle}^{= \quad 0}}{\langle \boldsymbol{v}_j, \boldsymbol{r}_j \rangle - \langle \boldsymbol{v}_j, \boldsymbol{r}_{j+1} \rangle} = 0,$$

für den zweiten Fall $l = k - 1$ ist

$$b_{k,k-1} = \frac{\overbrace{\langle \boldsymbol{r}_k, \boldsymbol{r}_{k-1} \rangle}^{= \quad 0} - \langle \boldsymbol{r}_k, \boldsymbol{r}_k \rangle}{\underbrace{\langle \boldsymbol{v}_{k-1}, \boldsymbol{r}_{k-1} \rangle}_{\langle \boldsymbol{r}_{k-1}, \boldsymbol{r}_{k-1} \rangle} - \underbrace{\langle \boldsymbol{v}_{k-1}, \boldsymbol{r}_k \rangle}_{= \quad 0}} = -\frac{\langle \boldsymbol{r}_k, \boldsymbol{r}_k \rangle}{\langle \boldsymbol{r}_{k-1}, \boldsymbol{r}_{k-1} \rangle}.$$

Wir setzten $\beta_{k-1} = -b_{k,k-1}$ womit die Behauptung (12) folgt. Der zweite Teil des Lemmas (4), die Gleichung (13), folgt direkt mit Gleichung (10). ∎

Herleitung - CG-Verfahren

Jetzt haben wir alle Voraussetzungen für das CG-Verfahren. In der Anwendung liefert das CG-Verfahren oft schon nach wenigen Iterationen gute Approximationen an die Lösung $\boldsymbol{x}^*$. Wir führen daher ein Abbruchkriterium ein. Falls das Residuum $\boldsymbol{r}_k$ in der 2-Norm kleiner als der Abbruchparameter $\varepsilon \in \mathbb{R}$ ist, bricht der Algorithmus ab.

Herleitung - CG-Verfahren

Bemerkung: In dem folgenden Algorithmus wurde noch die Hilfsvariable $\boldsymbol{z}_k$ eingeführt, damit das Produkt $\boldsymbol{A}\boldsymbol{v}_k$ nur einmal pro Iteration aufgerufen werden muss.

Algorithm 1 CG-Verfahren

$\boldsymbol{r}_0 \leftarrow \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$

$\boldsymbol{v}_0 \leftarrow \boldsymbol{r}_0$

for $k = 0, \ldots, N-1$ **do**

  $\boldsymbol{z}_k \leftarrow \boldsymbol{A}\boldsymbol{v}_k$

  $\alpha_k \leftarrow \frac{\boldsymbol{r}_k^{\mathsf{H}}\boldsymbol{r}_k}{\boldsymbol{v}_k^{\mathsf{H}}\boldsymbol{z}_k}$

  $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k\boldsymbol{v}_k$

  $\boldsymbol{r}_{k+1} \leftarrow \boldsymbol{r}_k - \alpha_k\boldsymbol{z}_k$

  $\beta_k \leftarrow \frac{\boldsymbol{r}_{k+1}^{\mathsf{H}}\boldsymbol{r}_{k+1}}{\boldsymbol{r}_k^{\mathsf{H}}\boldsymbol{r}_k}$

  $\boldsymbol{v}_{k+1} \leftarrow \boldsymbol{r}_{k+1} + \beta_k\boldsymbol{d}_k$

end for
