

Das Verfahren der konjugierten Gradienten

Michael Bauer

11. November 2013

Inhaltsverzeichnis

1.1	Motivation	2
1.2	Definition (A-orthogonal)	2
1.2.1	Bemerkung:	2
1.3	Satz	2
1.3.1	Beweis:	2
1.3.2	Bemerkung:	3
1.4	Lemma	3
1.4.1	Beweis:	3
1.4.2	Bemerkungen:	3
1.5	Algorithmus	3
1.5.1	Bemerkung:	4
1.6	Satz (Verallgemeinerung des Startvektors)	4
1.6.1	Beweis:	4
1.7	Lemma	4
1.7.1	Beweis:	4
1.7.2	Anmerkung:	5
1.8	Lemma	5
1.8.1	Beweis:	5
1.8.2	Bemerkung:	5
1.9	Lemma	5
1.9.1	Beweis:	5
1.9.2	Bemerkung:	6
1.10	Lemma (Zusammenhang zu Krylovräumen)	6
1.10.1	Beweis:	6
1.11	Satz (Bestimmung einer A-orthogonalen Basis)	6
1.11.1	Beweis:	6
1.12	Lemma	6
1.12.1	Beweis:	7
1.12.2	Bemerkung:	7
1.13	Folgerung	7
1.13.1	Bemerkung:	7
1.14	Algorithmus	7
1.15	Rechenaufwand	8
1.16	Literatur:	8

Das Verfahren der konjugierten Gradienten

1.1 Motivation

Löse ein Gleichungssystem $Ax = b$, wobei $A \in \mathbb{R}^n$ s.p.d., $x, b \in \mathbb{R}^n$ und n sehr groß.

1.2 Definition (A-orthogonal)

Sei A eine symmetrische, nicht singuläre Matrix. Zwei Vektoren $x, y \in \mathbb{R}^n$ heißen **konjugiert** oder **A-orthogonal**, wenn $x^T Ay = 0$ ist.

1.2.1 Bemerkung:

- Es definiert $\langle x, y \rangle_A = x^T Ay$ ein Skalarprodukt auf dem \mathbb{R}^n für A s.p.d.
- Wir nennen $\|x\|_A := \sqrt{\langle x, x \rangle_A}$ die Energie-Norm.

1.3 Satz

Sei $A \in \mathbb{R}^{n \times n}$ s.p.d. und

$$f(x) := \frac{1}{2}x^T Ax - b^T x, \quad (1)$$

wobei $b, x \in \mathbb{R}^n$. Dann gilt:

f hat ein eindeutig bestimmtes Minimum und

$$Ax^* = b \iff f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \quad (2)$$

1.3.1 Beweis:

1. Eindeutigkeit: per Widerspruch

Sei \hat{x} ein weiteres Minimum von f . Dann ist $\nabla f(\hat{x}) = A\hat{x} - b = 0 \Rightarrow A\hat{x} = b$.
 $\Rightarrow Ax = b$ hat zwei Lösungen x^* und \hat{x} . Widerspruch, da A eine quadratische Matrix und $\det(A) \neq 0 \Rightarrow$ das GLS hat eine eindeutige Lösung.

2. \Rightarrow : Sei x^* die eind. Lsg. von $Ax = b$. Dann kann man $f(x)$ auch folgendermaßen schreiben:

$$f(x) = \frac{1}{2}(x - x^*)^T A(x - x^*) - c \text{ mit } c = \frac{1}{2}(x^*)^T Ax^*$$

Da $y^T Ay > 0 \forall y \neq 0$ und c konstant ist, da es nicht von x abhängt, folgt

$$f(x) = \underbrace{\frac{1}{2}(x - x^*)^T A(x - x^*)}_{\geq 0} - c$$

ist genau dann minimal, wenn $x = x^*$.

3. \Leftarrow : Sei $f(x^*)$ das Minimum von $f(x)$, dann gilt

$$\nabla f(x^*) = Ax^* - b = 0 \Rightarrow Ax^* = b$$

$\Rightarrow x^*$ löst $Ax = b \Rightarrow$ Beh.

1.3.2 Bemerkung:

Wir wollen die Funktion $f(x)$ in jedem Iterationsschritt minimieren. Dann jedoch in einem entsprechenden Unterraum U_k des \mathbb{R}^n , den wir im weiteren Verlauf kennen lernen werden.

1.4 Lemma

Sei U_k ein k -dimensionaler Teilraum des \mathbb{R}^n ($k \leq n$), und p^0, p^1, \dots, p^{k-1} eine A -orthogonale Basis dieses Teilraums, also $\langle p^i, p^j \rangle_A = 0$ für $i \neq j$. Sei $v \in \mathbb{R}^n$, dann gilt für $u^k \in U_k$:

$$\|u^k - v\|_A = \min_{u \in U_k} \|u - v\|_A \quad (3)$$

genau dann, wenn u^k die A -orthogonale Projektion von v auf $U_k = \text{span}\{p^0, \dots, p^{k-1}\}$ ist. Außerdem hat \mathbf{u}^k die Darstellung

$$P_{U_k, \langle \cdot, \cdot \rangle_A}(v) = \mathbf{u}^k = \sum_{j=0}^{k-1} \frac{\langle v, p^j \rangle_A}{\langle p^j, p^j \rangle_A} p^j \quad (4)$$

1.4.1 Beweis:

Folgt direkt aus dem Satz des Projektionssatzes (z.B. Numerik 1).

1.4.2 Bemerkungen:

- Wie wir in Lemma 1.4. gesehen haben, nutzen wir den Projektionssatz aus, um das Minimum in U_k zu berechnen, also Minimierung von $f(x)$ in U_k .
- Wir wählen im Folgenden den Startvektor $x^0 = 0$.
- Wegen Lemma 1.3 ist die Richtung des steilsten Abstiegs gegeben durch $r^0 = b - Ax^0$, wobei wir den Vektor r Residuum nennen werden.
- Spezifisch für das CG-Verfahren ist die Nutzung des A -Skalarprodukts.

1.5 Algorithmus

Die folgenden Teilschritte definieren die Vorgehensweise zur Erzeugung der Lösung x^* durch Näherungen x^1, x^2, \dots

$U_1 := \text{span}\{r^0\}$, wobei $r^0 = b - Ax^0$
dann gilt für $k = 1, 2, 3, \dots$, falls $r^{k-1} = b - Ax^{k-1} \neq 0$:

CG_a : Bestimme A -orthogonale Basis

$$p^0, \dots, p^{k-1} \text{ von } U_k \quad (5)$$

CG_b : Bestimme $x^k \in U_k$, so dass

$$\|x^k - x^*\|_A = \min_{u \in U_k} \|x - x^*\|_A \quad (6)$$

d.h.

$$x^k = \sum_{j=0}^{k-1} \frac{\langle x^*, p^j \rangle_A}{\langle p^j, p^j \rangle_A} p^j \quad (7)$$

CG_c : Erweiterung des Teilraumes:

$$U_{k+1} := \text{span}\{p^0, \dots, p^{k-1}, r^k\} \text{ wobei } r^k := b - Ax^k \quad (8)$$

1.5.1 Bemerkung:

Wir wählen also die Richtung des steilsten Abstiegs, berechnen das Residuum. Erweitern unseren Teilraum U_k mit diesem und wenden dann ein Orthogonalisierungsverfahren an, um eine A-orthogonale Basis für U_k zu erhalten. Anschließend verwenden wir den Projektionssatz, um das neue x^k zu berechnen. Nach der erneuten Berechnung wiederholt sich dieses Vorgehen bis zum Abbruch des Algorithmus.

1.6 Satz (Verallgemeinerung des Startvektors)

Das Verfahren der konjugierten Gradienten ist unabhängig von der Wahl des Startvektors x^0 .

1.6.1 Beweis:

Zu lösen: $Ax^* = b$.

Sei $x^0 \neq 0$. Definiere für das transformierte System $A\tilde{x} = \tilde{b}$, $\tilde{x} := x^* - x^0$ und $\tilde{b} := b - Ax^0$

$$\implies A\tilde{x} = A(x^* - x^0) = b - Ax^0 = r^0$$

Sei nun: $\tilde{x}^0 = 0$ Startvektor mit Residuum \tilde{r} .

$$\implies \tilde{x}^k = x^k - x^0 \implies x^k = \tilde{x}^k + x^0$$

$$\implies \tilde{r}^k = \tilde{b} - A\tilde{x}^k = b - Ax^0 - A\tilde{x}^k$$

$$= b - A(x^0 - \tilde{x}^k) = b - Ax^k = r^k$$

$$\implies \tilde{r}^k = r^k$$

1.7 Lemma

Sei x^* die Lösung in Gleichung (6). Dann gilt für $y \in U_k$:

$$\langle x^*, y \rangle_A = \langle b, y \rangle \quad (9)$$

1.7.1 Beweis:

Wir nutzen die Eigenschaften des (A-orthogonalen) Skalarproduktes aus:

$$\langle x^*, y \rangle_A \stackrel{\text{Def.1.1}}{=} x^{*T} Ay \stackrel{\text{Symmetrie}}{=} y^T Ax^* = y^T b \stackrel{\text{Symmetrie}}{=} b^T y = \langle b, y \rangle$$

1.7.2 Anmerkung:

Um nun einen numerischen Algorithmus zu entwickeln, werden uns die folgenden Lemmata weiter helfen.

1.8 Lemma

Sei x^* die Lösung von Gleichung (6) und x^k die optimale Approximation von x^* in U_k . Dann kann x^k wie folgt berechnet werden:

$$x^k = x^{k-1} + \alpha_{k-1}p^{k-1}, \text{ mit } \alpha_{k-1} := \frac{\langle r^0, p^{k-1} \rangle}{\langle p^{k-1}, Ap^{k-1} \rangle} \quad (10)$$

1.8.1 Beweis:

$$\begin{aligned} x^k &\stackrel{(4)}{=} \sum_{j=0}^{k-1} \frac{\langle x^*, p^j \rangle_A}{\langle p^j, p^j \rangle_A} p^j = \underbrace{\sum_{j=0}^{k-2} \frac{\langle x^*, p^j \rangle_A}{\langle p^j, p^j \rangle_A} p^j}_{=x^{k-1}} + \frac{\langle \overbrace{Ax^*}^{=b-Ax^0=r^0}, p^{k-1} \rangle}{\langle Ap^{k-1}, p^{k-1} \rangle} p^{k-1} = \\ & x^{k-1} + \alpha_{k-1}p^{k-1}, \text{ mit } \alpha_{k-1} = \frac{\langle r^0, p^{k-1} \rangle}{\langle p^{k-1}, Ap^{k-1} \rangle} \end{aligned}$$

1.8.2 Bemerkung:

x^k kann mit wenig Aufwand aus x^{k-1} und p^{k-1} berechnet werden.

1.9 Lemma

Um U_{k+1} zu erhalten, also den Teilraum zu erweitern, muss lediglich das neue Residuum $r^k = b - Ax^k$ berechnet werden. Dieses erhält man durch:

$$r^k = r^{k-1} - \alpha_{k-1}Ap^{k-1} \quad (11)$$

Wobei α_{k-1} wie in (10).

1.9.1 Beweis:

1. Zeige, dass nur das neue Residuum berechnet werden muss:
Da $U_{k+1} = \text{span}\{p^0, \dots, p^{k-1}, r^k\}$ und wir die A-orthogonalen Vektoren p^0, \dots, p^{k-1} bereits bestimmt haben, muss nur noch das Residuum gemäß (8) berechnet werden.
2. Zeige (11) durch Erweiterung von (10):

$$\begin{aligned} x^k &= x^{k-1} + \alpha_{k-1}p^{k-1} \iff Ax^k = Ax^{k-1} + \alpha_{k-1}Ap^{k-1} \\ &\iff b - Ax^k = b - Ax^{k-1} - \alpha_{k-1}Ap^{k-1} \iff r^k = r^{k-1} - \alpha_{k-1}Ap^{k-1} \end{aligned}$$

1.9.2 Bemerkung:

Das α_{k-1} , sowie die Matrix-Vektor-Multiplikation Ap^{k-1} wurden bereits berechnet.

1.10 Lemma (Zusammenhang zu Krylovräumen)

Man kann U_k auch in folgender Form schreiben:

$$U_k := \text{span}\{r^0, r^1, \dots, r^{k-1}\} = \text{span}\{p^0, p^1, \dots, p^{k-1}\} = \text{span}\{r^0, Ar^0, \dots, A^{k-1}r^0\} \quad (12)$$

1.10.1 Beweis:

Per Induktion über k (klar für $k=0$):

Induktionsvoraussetzung:

$$U_k := \text{span}\{r^0, r^1, \dots, r^{k-1}\} = \text{span}\{p^0, p^1, \dots, p^{k-1}\} = \text{span}\{r^0, Ar^0, \dots, A^{k-1}r^0\}$$

$$\implies k \longrightarrow k+1 : U_{k+1}$$

$$r^k = r^{k-1} - \alpha_{k-1}Ap^{k-1}$$

$$p^{k-1} \in U_k = \text{span}\{r^0, \dots, A^{k-1}r^0\}$$

$$\text{da } p^{k-1} = \left(\sum_{i=0}^{k-1} \sigma_i A^i\right)r^0$$

$$\implies Ap^{k-1} = \left(\sum_{i=0}^{k-1} \sigma_i A^{i+1}\right)r^0$$

$$= \sigma_0 Ar^0 + \dots + \sigma_{k-1} A^k r^0$$

$$\implies r_k \in U_{k+1}$$

1.11 Satz (Bestimmung einer A-orthogonalen Basis)

Durch

$$p^{k-1} = r^{k-1} - \sum_{j=0}^{k-2} \frac{\langle r^{k-1}, p^j \rangle_A}{\langle p^j, p^j \rangle_A} p^j \quad (13)$$

wird die A-orthogonale Basis zum Vektor r^{k-1} bestimmt.

1.11.1 Beweis:

Der Beweis folgt direkt aus dem Gram-Schmidt-Orthonormalisierungsverfahren.

1.12 Lemma

Für jedes r^{k-1} und p^j gilt:

$$\langle r^{k-1}, p^j \rangle_A = 0 \quad \text{für } 0 \leq j \leq k-3$$

1.12.1 Beweis:

Sei $k \geq 3$ fest gewählt. Aus $U_1 = \text{span}\{r^0\}$, $U_2 = U_1 \oplus \text{span}\{r^1\} = \text{span}\{r^0, r^1\}$ usw. erhält man

$$U_m = \text{span}\{r^0, r^1, \dots, r^{m-1}\} \quad m = 1, 2, \dots, k \quad (14)$$

Aus der Definition von x^m ergibt sich $x^m - x^* \perp_A U_m$, also $-r^m = A(x^m - x^*) \perp U_m$. Zusammen mit (14) folgt hieraus

$$r^i \perp r^j \quad \text{für } 0 \leq i, j \leq k, i \neq j \quad (15)$$

Da $r^j \neq 0$ für $j \leq k-1$ gilt folgt wegen (15) muss dann $r^j \neq r^{j-1}$ gelten, also auch $x^j \neq x^{j-1}$, $j \leq k-1$. Aus (10) erhält man damit, dass $\alpha_i \neq 0$ für $j \leq k-2$ gilt. Nun gilt für $j \leq k-3$

$$\langle r^{k-1}, p^j \rangle_A = \langle r^{k-1}, Ap^j \rangle \stackrel{(11)}{=} \langle r^{k-1}, \frac{1}{\alpha_j} (r^j - r^{j-1}) \rangle = \frac{1}{\alpha_j} \langle r^{k-1}, r^j \rangle - \frac{1}{\alpha_j} \langle r^{k-1}, r^{j+1} \rangle \stackrel{(15)}{=} 0$$

1.12.2 Bemerkung:

Im Beweis nutzen wir aus, dass laut dem Projektionssatz die $x^m - x^*$ A-orthogonal auf U_m stehen. Durch das Ausnutzen der A-Orthogonalität können wir sehen, dass das negative Residuum senkrecht (nicht A-orthogonal) auf U_m steht. Durch Ausnutzung der Eigenschaften des Residuums und des Iterationsvektors und durch geschicktes Umformen, erhalten wir Lemma 1.12.

1.13 Folgerung

Wegen Lemma 1.11 vereinfacht sich (13) auf

$$p^{k-1} = r^{k-1} - \frac{\langle r^{k-1}, Ap^{k-2} \rangle}{\langle p^{k-2}, Ap^{k-2} \rangle} p^{k-2}$$

1.13.1 Bemerkung:

Somit können wir p^{k-1} einfach aus r^{k-1} und p^{k-2} berechnen.

Durch das Ausnutzen von Lemma 1.12. kann der Rechenaufwand nochmals verbessert werden (ohne Beweis) und man erhält den folgenden Algorithmus.

1.14 Algorithmus

Gegeben: $A \in \mathbb{R}^n$ s.p.d., $b \in \mathbb{R}^n$, Startvektor $x^0 \in \mathbb{R}^n$, $\beta_{-1} := 0$. Berechne $r^0 = b - Ax^0$. Für $k = 1, 2, \dots$, falls $r^{k-1} \neq 0$:

$$p^{k-1} = r^{k-1} + \beta_{k-2} p^{k-2}, \quad \text{wobei } \beta_{k-2} = \frac{\langle r^{k-1}, r^{k-1} \rangle}{\langle r^{k-2}, r^{k-2} \rangle} \quad \text{mit } (k \geq 2), \quad (16a)$$

$$x^k = x^{k-1} + \alpha_{k-1} p^{k-1}, \quad \text{mit } \alpha_{k-1} = \frac{\langle r^{k-1}, r^{k-1} \rangle}{\langle p^{k-1}, Ap^{k-1} \rangle} \quad (16b)$$

$$r^k = r^{k-1} - \alpha_{k-1} Ap^{k-1} \quad (16c)$$

1.15 Rechenaufwand

Im Allgemeinen hat das CG-Verfahren einen Aufwand von $\mathcal{O}(n^2)$ Operationen pro Iterationsschritt. Es müssen zwei Skalarprodukte und eine Matrix-Vektor-Multiplikation berechnet werden. In bestimmten Fällen, z.B. beim Lösen der diskreten Poisson Gleichung, kann sich der Rechenaufwand auf $\mathcal{O}(n)$ pro Iterationsschritt verbessern.

1.16 Literatur:

1. "Numerik für Ingenieure und Naturwissenschaftler",
W.Dahmen & A.Reusken, 2.,korrigierte Auflage, 2008, Seiten 566-572
2. "Finite Elemente", Dietrich Braess,, 5. überarbeitete Auflage, 2013, Seiten 200-203