03 CG-Verfahren

Nichtlineare Optimierung
WS 2020/21

A-konjugierte Vektoren

A-orthogonal

$d_i^T A d_j = 0$ für $i \neq j$

Es sei A spd. A-konjugierte Vektoren $d_j \neq 0$
$d_0, d_1, \ldots, d_k$ sind linear unabhängig.

$$\sum_{j=0}^{k} \alpha_j d_j = 0$$

$$\Rightarrow \underbrace{d_i^T A}_{} \sum_{j=0}^{k} \alpha_j d_j = \sum_{j=0}^{k} \alpha_j \underbrace{d_i^T A d_j}_{= 0 \text{ für } i \neq j} = 0$$

$$\Rightarrow \alpha_i \underbrace{d_i^T A d_i}_{= \|d_i\|_A^2 > 0} = 0$$

$\Rightarrow \alpha_i = 0$ für $0 \leq i \leq k$

Quizfrage

Grad.verfahren

Welche besondere Eigenschaft haben Verfahren zur Minimierung von $\frac{1}{2}x^T A x - b^T x + c$, die mit A-konjugierten Richtungen arbeiten und in jeder Iteration die exakte eindimensionale Minimierung ausführen?

$\longrightarrow$ Umfrage

- A Sie konvergieren in einem Schritt.
- B Sie konvergieren in dim A Schritten.
- C x_k minimiert die Zielfunktion über $x_0 + \operatorname{span}\{d_0, d_1, \ldots, d_{k-1}\}$
- D Sie konvergieren in zwei Schritten.

lin. Hülle

Das CG-Verfahren

$r = Ax - b$
$= \nabla \phi(x)$
$M^{-1} r = \nabla_M \phi(x)$

Verwende die Richtungen

- $d_0 := -M^{-1} r_0$ wie im Gradientenverf.
- $d_k := -M^{-1} r_k + \beta_k d_{k-1}$ für $k \geq 1$

Korrektur für A-Konjugiertheit

$d_{k-1}^T A d_k = - d_{k-1}^T A M^{-1} r_k + \beta_k d_{k-1}^T A d_{k-1} \stackrel{!}{=} 0$

$\Rightarrow \beta_k = \frac{d_{k-1}^T A M^{-1} r_k}{d_{k-1}^T A d_{k-1}} = \dots = \frac{\|r_k\|_{M^{-1}}^2}{\|r_{k-1}\|_{M^{-1}}^2}$

$\|r_k\|_{M^{-1}}^2$ wird sowieso im Verfahren benutzt

Das CG-Verfahren

1: Setze $k := 0$
2: Setze $r_0 := A x_0 - b$
3: Setze $d_0 := -M^{-1} r_0$
4: Setze $\delta_0 := -r_0^T d_0$ // $\delta_0 = \|\nabla_M \phi(x_0)\|_M^2$
5: **while** Abbruchkriterium nicht erfüllt **do**
6: Setze $q_k := A d_k$
7: Setze $\alpha_k := \delta_k / (d_k^T q_k)$
8: Setze $x_{k+1} := x_k + \alpha_k d_k$
9: Setze $r_{k+1} := r_k + \alpha_k q_k$
10: Setze $d_{k+1} := -M^{-1} r_{k+1}$
11: Setze $\delta_{k+1} := -r_{k+1}^T d_{k+1}$ // $\delta_{k+1} = \|\nabla_M \phi(x_{k+1})\|_M^2$
12: Setze $\beta_{k+1} := \frac{\delta_{k+1}}{\delta_k}$
13: Setze $d_{k+1} := d_{k+1} + \beta_{k+1} d_k$
14: Setze $k := k + 1$
15: **end while**
16: **return** x_k

Mitrechnen von Funktionswerten

Bemerkung 4.5

$r_0 = Ax_0 - b$

vermeide zusätzliche Matrix-Vektor-Multipl. mit x_k

$$\phi(x_0) = \frac{1}{2} x_0^T A x_0 - b^T x_0 + c$$

$$= \frac{1}{2} x_0^T r_0 - \frac{1}{2} x_0^T b + c$$

$$= \frac{1}{2} x_0^T (r_0 - b) + c \quad \leftarrow \text{allgemeiner: für } x_k, r_k$$

$$\phi(x_{k+1}) - \phi(x_k) \overset{(4.8)}{=} -\frac{1}{2} \frac{(d_k^T r_k)^2}{d_k^T A d_k}$$

$$= -\frac{1}{2} (d_k^T r_k) \frac{d_k^T r_k}{d_k^T A d_k}$$

$$= -\frac{1}{2} \delta_k \alpha_k$$

Quizfrage

Welche Bausteine muss der Anwender zur Verfügung stellen, um das CG-Verfahren anwenden zu können?

$\longrightarrow$ Umfrage

3 A die Matrizen A und M

o B die Matrizen A und M^{-1}

o C Matrix-Vektor-Produkte mit A und M

9 D Matrix-Vektor-Produkte mit A und M^{-1}

M oder Matrix-Vektor-Produkte mit M werden nicht benötigt

Quizfrage

Welche Eigenschaft hat das CG-Verfahren?

solange nicht $x_k = x^*$

$\longrightarrow$ Umfrage

- A $\phi(x_k)$ fällt streng monoton
- C benötigt ein Matrix-Vektor-Produkt mit A und eines mit M^{-1} pro Iteration
- B die erste Iterierte x_1 ist identisch zu der des Gradientenverfahrens
- D $\|x_k - x_0\|_M$ ist streng monoton wachsend

Lemma 4.20

Berechnung über (4.35)

Quizfrage

Welche Rolle spielt der Vorkonditionierer M?

→ Umfrage

hoffentlich !

1 A legt das M-Innenprodukt im Raum $\mathbb{R}^n$ der Optimierungsvariablen fest

3 C legt das M^{-1}-Innenprodukt im Raum $\mathbb{R}^n$ der Residuen fest

4 B verbessert die Konditionszahl $\frac{\lambda_{\max}(A)}{\lambda_{\min}(A)}$ zu $\frac{\lambda_{\max}(A;M)}{\lambda_{\min}(A;M)}$

2 D bestimmt die affinen Unterräume, in denen ϕ sukzessive minimiert wird

$x_{k+1} = x_k + \alpha_k d_k = \dots = x_0 + \sum_{j=0}^{k} \alpha_j d_j$

$\hookrightarrow \in \operatorname{span}\{d_0, \dots, d_k\} = M^{-1} \operatorname{span}\{r_0, (AM^{-1}) r_0, \dots, (AM^{-1})^k r_0\}$

Demonstration CG-Verfahren

`test_cg_quadratic.m`

Zeit für Ihre Fragen

Was sind Ihre Fragen zu den Themen der Woche?

$\rightarrow$ Benutzen Sie den Chat.

Fragen und Antworten 1

Warum wird beim Training von neuronalen Netzen oft ein (stochastisches) Gradientenverfahren verwendet, wenn doch das (nichtlineare) CG-Verfahren bessere Eigenschaften hat?

Die Eigenschaften, die wir bisher verglichen haben, beziehen sich auf das Gradienten- und CG-Verfahren für quadratische Aufgaben. Bei NN treten aber nichtlineare Aufgaben auf. Hier ist der Unterschied zwischen Gradienten- und nichtlinearem CG-Verfahren nicht immer so gravierend wie bei quadratischen Funktionen. Außerdem zeigt sich, dass bei NN eine schnelle Konvergenz zu einem lokalen Min. sich oft negativ auf die Fähigkeit auswirkt, vom Trainingsdatensatz zu generalisieren.

Fragen und Antworten 2