Konvergensanalys kvasi-Newton metoder

Verktyg: Kantorovichs olikhet

Låt $Q \in \mathbb{R}^{n \times n}$ vara symmetrisk och positivt definit, och låt M och m beteckna dess största respektive minsta egenvärde. Då gäller för varje nollskilt $x \in \mathbb{R}^n$ att

$$\frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \geq \frac{4Mm}{(M+m)^2}.$$

Kommentar: eftersom $M = \max\limits_{\|x\|_2=1} x^T Q x$

och $m = \min\limits_{\|x\|_2=1} x^T Q x$ så fås att

$$\frac{4Mm}{(M+m)^2} = 1 - \frac{(M-m)^2}{(M+m)^2} = 1 - \frac{(\kappa(Q)-1)^2}{(\kappa(Q)+1)^2}$$

där $\kappa(Q) = \dfrac{\max\limits_{\|x\|_2=1} x^T Q x}{\min\limits_{\|x\|_2=1} x^T Q x}$ är konditionstalet för Q.

Problem: $\min f(x) = \frac{1}{2} x^T Q x - b^T x$

Optimum: $x^* = Q^{-1} b \Rightarrow f(x^*) = \frac{1}{2} x^{*T} Q x^* - b^T x^* =$
$= -\frac{1}{2} x^{*T} Q x^* = -\frac{1}{2} b^T Q^{-1} b$ (används nedan)

<u>Metod</u>: $x^{k+1} = x^k - \alpha_k S_k \nabla f(x^k)$, $k = 0, 1, 2, \ldots$,
där varje S_k är symmetrisk och positivt definit och α_k löser $\min\limits_{\alpha \geq 0} f(x^k - \alpha S_k \nabla f(x^k))$.

Obs: $\nabla f(x^k) \neq 0$ och S_k positivt definit
$\Rightarrow -\nabla f(x^k)^T S_k \nabla f(x^k) < 0 \Rightarrow \alpha_k > 0$

Om $S_k = I$, $\forall k \rightarrow$ brantaste-lutnings-metoden.
Om $S_k = \nabla^2 f(x^k)^{-1}$, $\forall k \rightarrow$ Newtons modifierade metod.

Låt $E(x) = f(x) - f(x^*) = \frac{1}{2} x^T Q x - b^T x - \frac{1}{2} x^{*T} Q x^* =$
$= \frac{1}{2} x^T Q x - x^{*T} Q x + \frac{1}{2} x^{*T} Q x^* = \frac{1}{2}(x - x^*)^T Q (x - x^*)$.

Alltså: $\min\limits_{x \in R^n} f(x) \iff \min\limits_{x \in R^n} E(x)$
och $\min\limits_{x \in R^n} E(x) = E(x^*) = 0$.

<u>Sats</u>: Låt M_k och m_k vara det största respektive minsta egenvärdet till $S_k Q$, $k = 0, 1, 2, \ldots$. Då gäller att

$$E(x^{k+1}) \leq \left(\frac{M_k - m_k}{M_k + m_k} \right)^2 E(x^k).$$

<u>Bevis</u>: Låt $g^k = \nabla f(x^k) = Qx^k - b$. Då fås
$f(x^k - \alpha S_k g^k) = \frac{1}{2}(x^k - \alpha S_k g^k)^T Q (x - \alpha S_k g^k) - b^T (x^k - \alpha S_k g^k)$
och $\min\limits_{\alpha \geq 0} f(x^k - \alpha S_k g^k)$ fås för

$$\alpha_k = \frac{g^{kT} S_k g^k}{g^{kT} S_k Q S_k g^k}.$$

(Utnyttja att $S_k^T = S_k$ och att $Qx^k - b = g^k$.)

Vidare fås $E(x^{k+1}) = f(x^{k+1}) - f(x^*) =$

$$= \frac{1}{2}(x^k - \alpha_k S_k g^k)^T Q (x^k - \alpha_k S_k g^k) - b^T(x^k - \alpha_k S_k g_k) - f(x^*) =$$

$$= \frac{1}{2}(x^{kT} Q x^k - 2\alpha_k g^{kT} S_k \underline{Q x^k} + \alpha_k^2 g_k^T S_k Q S_k g^k) -$$

$$- b^T x^k + \alpha_k g^{kT} S_k b - f(x^*) =$$

utnyttja $Qx^k = g^k + b$ och sätt in α_k

$$= \frac{1}{2}\left(x^T Q x^k - \frac{(g^{kT} S_k g^k)^2}{g^{kT} S_k Q S_k g^k}\right) - b^T x^k - f(x^*).$$

Eftersom $E(x^k) = \frac{1}{2} x^{kT} Q x^k - b^T x^k - f(x^*)$ fås

$$E(x^{k+1}) = E(x^k) - \frac{1}{2}\frac{(g^{kT} S_k g^k)^2}{g^{kT} S_k Q S_k g^k}.$$

Vidare är $E(x^k) = f(x^k) - f(x^*) =$

$$= \frac{1}{2} x^{kT} Q x^k - b^T x^k + \frac{1}{2} b^T Q^{-1} b = \frac{1}{2}(Qx^k - b)^T Q^{-1} (Qx^k - b) =$$

$$= \frac{1}{2} g^{kT} Q^{-1} g^k \Rightarrow$$

$$E(x^{k+1}) = E(x^k) - \frac{1}{2}\frac{(g^{kT} S_k g^k)^2}{g^{kT} S_k Q S_k g^k} \cdot \underbrace{\frac{E(x^k)}{\frac{1}{2} g^{kT} Q^{-1} g^k}}_{=1}$$

$$\Rightarrow E(x^{k+1}) = \left(1 - \frac{(g^{kT} S_k g^k)^2}{(g^{kT} S_k Q S_k g^k)(g^{kT} Q^{-1} g^k)}\right) E(x^k).$$

Låt $T_k = S_k^{1/2} Q S_k$ och $p^k = S_k^{1/2} g^k$. Då fås

$$E(x^{k+1}) = \left(1 - \frac{(p^{kT} p^k)^2}{(p^{kT} T_k p^k)(p^{kT} T_k^{-1} p^k)}\right) E(x^k).$$

Men $S_k^{1/2} T_k S_k = S_k^{1/2} (S_k^{1/2} Q S_k^{1/2}) S_k^{-1/2} = S_k Q$, dvs $S_k Q$ är likformig med T_k och har därför samma egenvärden som T_k. Alltså har T_k största och minsta egenvärden M_k respektive m_k.

Kantorovichs olikhet ger då att

$$E(x^{k+1}) \leq \left(1 - \frac{4M_k m_k}{(M_k+m_k)^2}\right) E(x^k)$$

$$\Rightarrow E(x^{k+1}) \leq \left(\frac{M_k - m_k}{M_k + m_k}\right)^2 E(x^k).$$

V.S.V.

Följdsats: Om $\{S_k\}$ väljs så att $m_k \geq \varepsilon > 0$, $\forall k$, och $\{M_k\}$ är begränsad så gäller att $\{x^k\} \to x^*$.

Kommentarer:

Om $S_k \approx Q^{-1}$ så fås $M_k \approx 1$ och $m_k \approx 1$
$\Rightarrow M_k - m_k \approx 0 \Rightarrow E(x^k)$ minskar mycket fort.

För icke-kvadratiska problem ger valet $S_k = \nabla^2 f(x^k)^{-1}$, dvs Newton-riktningen, bästa möjliga konvergenshastighet (i denna metodklass).

Det finns probleminstanser för vilka

$$E(x^{k+1}) \approx \left(\frac{M_k - m_k}{M_k + m_k}\right)^2 E(x^k)$$

gäller i praktiken. (Resultatet i satsen är alltså det starkaste möjliga generellt giltiga.)

Brantaste-lutnings-metoden svarar mot valet $S_k = I, \forall k$. Dess konvergenshastighet avgörs alltså av egenvärdena till Q, och om Q är illa-konditionerad så kan konvergensen alltså bli mycket långsam.