Broyden-familjen

Rang-två-uppdatering av approximation av invers till Hessian.
Bevarar positiv definithet.

Uppdateringsformel:

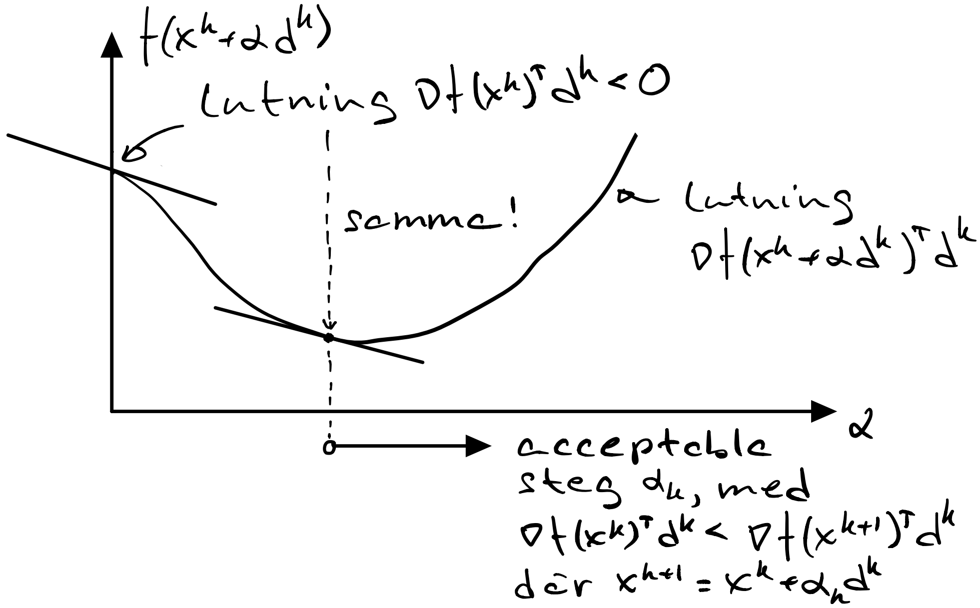
$$D_{k+1} = D_k + \frac{p^k p^{kT}}{p^{kT} q^k} - \frac{D_k q^k q^{kT} D_k}{q^{kT} D_k q^k} + \xi^k \tau^k v^k v^{kT}$$

där $v^k = \dfrac{p^k}{p^{kT} q^k} - \dfrac{D_k q^k}{\tau^k}$, $\tau^k = q^{kT} D_k q^k$ och
ξ^k är en parameter, sådan att $0 \le \xi^k \le 1$.
D_0 väljs symmetrisk och positivt definit, men kan i övrigt väljas godtyckligt.
Uppenbarligen gäller att $D_k^T = D_k \Rightarrow D_{k+1}^T = D_{k+1}$, dvs uppdateringen bevarar symmetrin hos approximationen.

Valet $\xi^k = 0$, $\forall k$, kallas Davidon-Fletcher-Powell (DFP).
Valet $\xi^k = 1$, $\forall k$, kallas Broyden-Fletcher-Goldfarb-Shanno (BFGS).
Klassen av uppdateringsformler som fås för $0 \le \xi^k \le 1$ kallas Broyden-familjen.
DFP är äldst. BFGS anses vara den bästa generella kvasi-Newton metoden.

Rang-två eftersom uppdateringen görs med de två vektorerna p^k och $D_k q^k$.

Sats: Låt $d^k = -D_k \nabla f(x^k)$ och $x^{k+1} = x^k + \alpha_k d^k$, samt antas att D_k är positivt definit. Då gäller att om $\alpha_k \geq 0$ väljs så att x^{k+1} uppfyller att $\nabla f(x^k)^T d^k < \nabla f(x^{k+1})^T d$ så är D_{k+1} positivt definit.



Bevis: Från $\nabla f(x^k)^T d^k < \nabla f(x^{k+1})^T d^k$ följer att $x^{k+1} \neq x^k$, $d^k \neq 0$ och $\_^k = \nabla f(x^{k+1}) - \nabla f(x^k) \neq 0$.
Vidare: D_k positivt definit och $d^k \neq 0 \Rightarrow$
$\nabla -(x^k)^T d = d_k^T D_k d^k < 0 \Rightarrow \alpha_k > 0$.
Från $\nabla f(x^k)^T d^k < f(x^{k+1})^T d^k$ och $\alpha_k > 0$ följer att
$p^{kT} q^{kT} = (x^{k+1} - x^k)^T q^k = \alpha_k d^{kT} \left(\nabla f(x^{k+1}) - \nabla f(x^k)\right) > 0$.
D_k positivt definit och $q^k \neq 0$ ger att $q^{kT} D_k q^k > 0$,

vilket tillsammans med $p^{k^T}q^k \neq 0$ ger att D_{k+1} är väldefinierad (ty ingen nämnare är noll).

Låt $z \in \mathbb{R}^n$ vara noll-skild. Då fås

$$z^T D_{k+1} z = z^T D_k z + \frac{(z^T p^k)^2}{p^{k^T} q^k} - \frac{(q^{k^T} D_k z)^2}{q^{k^T} D_k q^k} + \xi^k \tau^k (v^{k^T} z)^2.$$

Med $a = D_k^{1/2} z$ och $b = D_k^{1/2} q^k$ fås att

$$z^T D_{k+1} z = \underbrace{\frac{\|a\|^2 \cdot \|b\|^2 - (a^T b)^2}{\|b\|^2}}_{\substack{\geq 0 \\ \text{enligt Schwartz} \\ \text{olikhet: } |a^T b| \leq \|a\| \cdot \|b\|}} + \underbrace{\frac{(z^T p^k)^2}{p^{k^T} q^k}}_{\substack{\geq 0 \\ \text{enligt} \\ \text{ovan}}} + \underbrace{\xi^k}_{\geq 0} \underbrace{\tau^k}_{> 0} \underbrace{(v^{k^T} z)^2}_{\geq 0}$$

Alltså gäller $z^T D_{k+1} z \geq 0$. Olikheten $|a^T b| \leq \|a\| \cdot \|b\|$ ger upphov till två fall.

<u>Om</u> $\|a\|^2 \cdot \|b\|^2 > |a^T b|$ gäller så fås $z^T D_{k+1} z > 0$.

<u>Om</u> $\|a\|^2 \cdot \|b\|^2 = |a^T b|$ gäller så är $a = \lambda b$ för något $\lambda \in \mathbb{R}$. Men $a = \lambda b \Leftrightarrow z = \lambda q^k$, där $z \neq 0$ och $q^k \neq 0$, varför $\lambda \neq 0$ gäller.

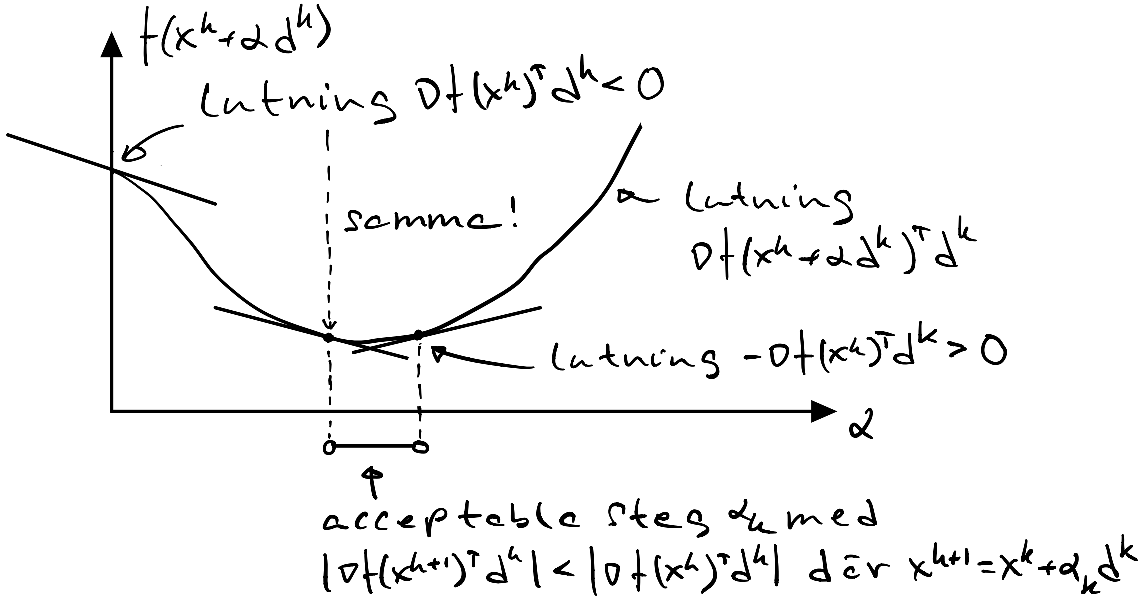
Då fås $z^T p^k = \underbrace{\lambda}_{\neq 0} \underbrace{q^{k^T} p^k}_{\substack{> 0 \\ \text{enligt ovan}}} \neq 0 \Rightarrow z^T D_{k+1} z > 0$

Alltså gäller alltid $z^T D_{k+1} z > 0$, varför D_{k+1} är positivt definit.

V.S.V.

Hur säkerställs att villkoret $\nabla f(x^k)^T d^k < \nabla f(x^{k+1})^T d^k$ uppfylls?

Ett sätt: kräv att $|\nabla f(x^{k+1})^T d^k| < |\nabla f(x^k)^T d^k|$ gäller. Eftersom $\nabla f(x^k)^T d^k < 0$ så gäller då att $\nabla f(x^k)^T d^k < \nabla f(x^{k+1})^T d^k < -\nabla f(x^k)^T d^k$



Speciellt uppfylls detta krav om α_k ges av den exakta linjesökningen $\min_{\alpha \geq 0} f(x^k + \alpha d^k)$, vilken ger $\nabla f(x^{k+1})^T d^k = 0$.

Satsen förutsätter inget om ξ^k. Alltså gäller den speciellt för DFP och BFGS.

Man kan visa att om α_k ges av en exakt linjesökning så gäller för godtycklig målfunktion f att iterationssekvensen $\{x^k\}$ inte beror av sekvensen $\{\xi^k\}$. Valen av ξ^k är alltså av betydelse endast då approximativa linjesökningar används.

Man kan vidare visa att om målfunktionen f är kvadratisk, säg $f(x) = \frac{1}{2}x^T Q x - b^T x$, och exakta linjesökningar används så gäller:

$$\begin{cases} d^k,\ k=0,1,\ldots, \text{ är } Q\text{-konjugerade,} \\ D_n = Q^{-1}, \\ \text{konvergens till optimum fås på} \\ \text{högst } n \text{ iterationer.} \end{cases}$$

Vidare gäller då:

$D_0 = I$ ger konjugerade gradient metoden.

$D_0 \neq I$ ger en prekonditionerad konjugerad gradientmetod med $H = SS^T = D_0$.

Kvasi-Newton metoder:

- kräver inte regelbundna omstarter
- kräver inte exakta linjesökningar
- är minneskrävande om Hessian-invers-approximationen D_k är stor och tät.