Newton-metoder

Problem: $\min_{x \in \mathbb{R}^n} f(x)$ där $f \in C^2$

Givet en iterationspunkt x^k.
Approximera $f(x)$ mha en andra ordningens Taylor-utveckling i x^k:

$$q(x) = f(x^k) + \nabla f(x^k)^T (x - x^k) + \frac{1}{2}(x - x^k)^T \nabla^2 f(x^k)(x - x^k).$$

Antag att $\nabla^2 f(x^k)$ är icke-singulär och sök stationär punkt, $\bar{x}^k$, till approximationen $q(x)$.

$$\nabla q(x) = 0 \Rightarrow \bar{x}^k - x^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$$

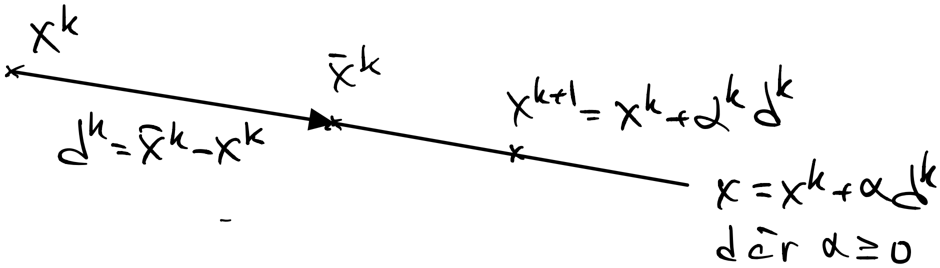
Newtons metod: välj $x^{k+1} = \bar{x}^k \Rightarrow$

$$x^{k+1} = x^k - \nabla^2 f(x^k)^{-1} \nabla f(x^k).$$

[För $f : \mathbb{R} \to \mathbb{R}$ fås $x^{k+1} = x^k - \dfrac{f'(x^k)}{f''(x^k)}$, dvs iterationen i Newton-Raphsons metod på ekvationen $f'(x) = 0$.]

Newtons modifierade metod: definiera sökriktningen $d^k = \bar{x}^k - x^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$, låt α_k lösa linjesökningsproblemet $\min_{\alpha \geq 0} f(x^k + \alpha d^k)$

och definiera $x^{k+1} = x^k + \alpha_k d^k$.



Observera att $d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k)$ kan tolkas som en vridning av $\nabla f(x^k)$ och att $d^k = 0$ om $\nabla f(x^k) = 0$.

Om linjesökningen ersätt med $\alpha_k = 1$, $\forall k \Rightarrow$ Newtons metod.

Om $\bar{x}^k$ ligger nära x^k så är $q(x)$ troligen en god approximation av $f(x)$ i $\bar{x}^k$ och $\alpha_k \approx 1$ kan förväntas.

Riktningen $d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k)$ beräknas vanligen som lösningen till ekvationssystemet $\nabla^2 f(x^k) d = -\nabla f(x^k)$, dvs inversen $\nabla^2 f(x^k)^{-1}$ beräknas inte.

Vad är riktningsderivatan för Newton-riktningen $d^k = -\nabla^2 f(x^k)^{-1}\nabla f(x^k)$?

$$\nabla f(x^k)^T d^k = (-\nabla^2 f(x^k) d^k)^T d^k = -d^{kT}\nabla^2 f(x^k) d^k$$

Alltså gäller (givet att $\nabla f(x^k) \neq 0$):

$\nabla^2 f(x^k)$ positivt definit $\Rightarrow \nabla f(x^k)^T d^k < 0$
$\nabla^2 f(x^k)$ negativt definit $\Rightarrow \nabla f(x^k)^T d^k > 0$
annars: $\nabla f(x^k)^T d^k$ osäkert

Observera: f konvex på hela $\mathbb{R}^n \Leftrightarrow$ $\nabla^2 f(x)$ positivt semidefinit på hela $\mathbb{R}^n$.
Alltså: f konvex på $\mathbb{R}^n$ och $\nabla^2 f(x^k)$ icke-singulär $\Rightarrow d^k = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$ är en avtagenderiktning för f i x^k.

Vad göra om f är icke-konvex och $\nabla f(x^k)^T d^k > 0$ fås (i ett minimerigsproblem)?
Ersätt $\nabla^2 f(x^k)$ med $\nabla^2 f(x^k) + \nu I$, där $\nu > 0$!
Då gäller: λ egenvärde till $\nabla^2 f(x^k)$ $\Rightarrow \lambda + \nu$ egenvärde till $\nabla^2 f(x^k) + \nu I$.
Vidare: $\nu > -$[mest negativa egenvärdet till $\nabla^2 f(x^k)$] $\Rightarrow \nabla^2 f(x^k) + \nu I$ är positivt definit och riktningen $d^k_{NM} = -(\nabla^2 f(x^k) + \nu I)^{-1} \nabla f(x^k)$ är en avtagenderiktning. Kallas Newton-Marquardt. Notera att då ν väljs mycket stor så kommer d_{NM} så mot att bli parallell med $-\nabla f(x^k)$, dvs med

branteste-lutnings-riktningen, som är den brantaste avtagande-riktningen. Newton-Marquardt "vrider" alltså Newton-riktningen mot branteste-lutnings-riktningen, tillräckligt mycket för att en avtagande riktning ska fås.

I praktiken kan beräkningen av $\nabla^2 f(x^k)$ eller $d = -\nabla^2 f(x^k)^{-1} \nabla f(x^k)$ vara alltför krävande (m.a.p. minne eller tid). Om $-\nabla^2 f(x^k)^{-1}$ ersätts med en positivt definit approximation D_k, som ger riktningen $d^k = -D_k \nabla f(x^k)$, så fås en kvasi-Newton metod.