Gauss-Newtons metod

Betrakta det olinjära minsta-kvadrat-problemet

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|g(x)\|^2$$

där $g: \mathbb{R}^n \to \mathbb{R}^m$, dvs $g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$.

Typiskt är $m > n$ och $g(x) \neq 0$ gäller för alla $x \in \mathbb{R}^n$. Problemet kan tolkas som att finna en minsta-kvadrat-lösning till (det inkonsistenta) olinjära ekvationssystemet $g(x) = 0$. Ett specialfall utgörs av multipel linjär regression, i vilket fall g är en linjär funktion.

Gauss-Newton är en iterativ metod för generella olinjära minsta-kvadrat-problem. Givet en iterationspunkt x^k approximeras funktionen g mha en första ordningens Taylor-utveckling i x^k:

$$\tilde{g}(x, x^k) = g(x^k) + \nabla g(x^k)^T (x - x^k)$$

där $\nabla g(x^k)^T \in \mathbb{R}^{n \times m}$ är Jacobianen (eller funktionalmatrisen) för funktionen g, dvs

$\nabla g(x^k) = (\nabla g_1(x^k), \ldots, \nabla g_m(x^k))$. Nästa iterationspunkt, x^{k+1}, fås som lösningen till det approximativa problemet

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2} \|\tilde{g}(x, x^k)\|^2.$$

Observera att eftersom $\tilde{g}$ är linjär så är detta ett linjärt minsta-kvadrat-problem [och det kan tolkas som att finna en minsta-kvadrat-lösning till (det inkonsistenta) linjära ekvationssystemet $\tilde{g}(x, x^k) = 0$].

Lösningen x^{k+1} ges av $\nabla\left(\tfrac{1}{2}\|\tilde{g}(x,x^k)\|^2\right) = 0$.
Men $\|\tilde{g}(x,x^k)\|^2 = \tilde{g}(x,x^k)^T g(x^k) =$

$$= [g(x^k) + \nabla g(x^k)^T(x-x^k)]^T [g(x^k) + \nabla g(x^k)^T(x-x^k)] =$$

$$= g(x^k)^T g(x^k) + 2[\nabla g(x^k) g(x^k)]^T (x - x^k) +$$

$$+ (x-x^k)^T \nabla g(x^k) \nabla g(x^k)^T (x-x^k)$$

$$\Rightarrow \nabla\left(\tfrac{1}{2}\|\tilde{g}(x,x^k)\|^2\right) = \nabla g(x^k) g(x^k) + \nabla g(x^k) \nabla g(x^k)^T (x-x^k)$$

Om $\nabla g(x^k) \nabla g(x^k)^T$ antas vara icke-singulär så fås alltså

$$x^{k+1} = x^k - (\nabla g(x^k) \nabla g(x^k)^T)^{-1} \nabla g(x^k) g(x^k).$$

<u>Specialfall</u>: Betrakta systemet $Ax = b$, där $A \in \mathbb{R}^{m \times n}$ med $m > n$. Antag att A har full rang (dvs rang$(A) = n$) och att $b \neq 0$. Då saknar

systemet $Ax=b$ lösning, dvs $g(x)=Ax-b\neq 0$ för alla $x\in\mathbb{R}^n$ och $\min_{x\in\mathbb{R}^n}\frac{1}{2}\|g(x)\|^2=\min_{x\in\mathbb{R}^n}\frac{1}{2}\|Ax-b\|^2>0$.

Då fås $\nabla g(x)=A^T$ och det gäller för varje $\bar{x}\in\mathbb{R}^n$ (tex $\bar{x}=0$) att $\tilde{g}(x,\bar{x})=g(x)$, samt att uttrycket ovan ger $x=\bar{x}-(A^TA)^{-1}A^T(A\bar{x}-b)=$ $=(A^TA)^{-1}A^Tb$, vilket är det välbekanta uttrycket för minsta-kvadrat-lösningen till det inkonsistena systemet $Ax=b$.

Egenskaper hos <u>riktningen</u> $x^{k+1}-x^k=$ $=-(\nabla g(x^k)\nabla g(x^k)^T)^{-1}\nabla g(x^k)g(x^k)$? Låt $s\in\mathbb{R}^n$. Då är $s^T\nabla g(x^k)\nabla g(x^k)^Ts=(\nabla g(x^k)^Ts)^T(\nabla g(x^k)^Ts)=$ $=\|\nabla g(x^k)^Ts\|^2\geq 0$. Alltså är matrisen $\nabla g(x^k)\nabla(x^k)^T$ alltid positivt semidefinit, och positivt definit om den är icke-singulär. Vidare är $\nabla g(x^k)g(x^k)=\nabla\left(\frac{1}{2}\|g(x)\|^2\right)\Big|_{x=x^k}$.

Alltså är $-(\nabla g(x^k)\nabla g(x^k)^T)^{-1}\nabla g(x^k)g(x^k)$ en <u>avtagenderiktning</u> för $\frac{1}{2}\|g(x)\|^2$ i x^k.

Komplikationer:

- $\nabla g(x^k)\nabla g(x^k)^T$ kan vara singulär eller nästan singulär,
- det kan inträffa att $\frac{1}{2}\|g(x^{k+1})\|^2>\frac{1}{2}\|g(x^k)\|^2$ gäller (dvs steget att är för långt).

Vanlig modifierad form:

$$x^{k+1} = x^k - \alpha_k \left(\nabla g(x^k) \nabla g(x^k)^T + \Delta^k\right)^{-1} \nabla g(x^k)\, g(x^k)$$

där Δ^k är en diagonalmatris sådan att $\nabla g(x^k) \nabla g(x^k)^T + \Delta^k$ är (signifikant) positivt definit och α_k tas genom en (approximativ) linjesökning map $\frac{1}{2}\|g(x)\|^2$, dvs

$$\min_{\alpha \geq 0} \frac{1}{2}\left\|g\left(x^k - \alpha \left(\nabla g(x^k) \nabla g(x^k)^T + \Delta^k\right)^{-1} \nabla g(x^k)\, g(x^k)\right)\right\|^2.$$

Då gäller alltid att $\frac{1}{2}\|g(x^{k+1})\|^2 < \frac{1}{2}\|g(x^k)\|^2$. Det vanligaste valet är $\Delta^k = \nu I$, där $\nu > 0$ är tillräckligt stor. Metoden kallas då Levenberg-Marquardt.

Uppdateringen i Gauss-Newtons metod,

$$x^{k+1} = x^k - \left(\nabla g(x^k) \nabla g(x^k)^T\right)^{-1} \nabla g(x^k)\, g(x^k),$$

påminner om Newtons metod. Vilken är relationen? I Newtons metod skulle uppdateringen ges av

$$x^{k+1} = x^k - \nabla^2\left(\frac{1}{2}\|g(x)\|^2\right)^{-1}\Big|_{x=x^k} \nabla\left(\frac{1}{2}\|g(x)\|^2\right)\Big|_{x=x^k}.$$

Enligt ovan är $\nabla\left(\frac{1}{2}\|g(x)\|^2\right) = \nabla g(x)\, g(x)$.

Vidare är $\nabla^2\left(\frac{1}{2}\|g(x)\|^2\right) = \nabla\left(\nabla g(x)\, g(x)\right) =$

$$= \nabla g(x) \nabla g(x)^T + \sum_{i=1}^{m} g_i(x)\, \nabla^2 g_i(x).$$

I Gauss-Newtons metod försummas alltså den andra termen i Hessianen. Notera att den innehåller Hessianer av alla funktionerna g_i och därför kan förväntas vara mer krävande att beräkna än Jacobianen av g.

Gauss-Newton-approximation av $\nabla^2(\frac{1}{2}\|g(x)\|^2)$ är god om funktionen g är måttligt olinjär (ty då är alla $\nabla^2 g_i(x^k) \approx 0$) eller om $g(x^k) \approx 0$ gäller. (Det senare gäller speciellt om $\min_{x \in \mathbb{R}} \frac{1}{2}\|g(x)\|^2 \gtrsim 0$ och x^k är nära-optimal.) Under någon av dessa omständigheter blir alltså Gauss-Newton-steget ett approximativt Newton-steg.

Om Gauss-Newton-approximationen av $\nabla^2(\frac{1}{2}\|g(x)\|^2)$ inte är tillräckligt god för att få bra konvergens så kan termen $\sum_{i=1}^{m} g_i(x) \nabla^2 g_i(x)$ approximeras som i kvasi-Newton-metoder.

Specialfall: Givet att $m=n$ och att det olinjära ekvationssystemet $g(x)=0$ har

Lösning. Det kan då lösas genom att finna ett x sådant att $\|g(x)\| = 0$, vilket kan göras genom att söka $\min_{x \in R^n} \frac{1}{2}\|g(x)\|^2$.

Om $\nabla g(x^k)$ är icke-singulär så fås

$$x^{k+1} = x^k - (\nabla g(x^k)\nabla g(x^k)^T)^{-1}\nabla g(x^k) g(x^k) =$$

$$= x^k - (\nabla g(x^k)^T)^{-1}\nabla g(x^k)^{-1}\nabla g(x^k) g(x^k) =$$

$$= x^k - (\nabla g(x^k)^T)^{-1} g(x^k),$$ vilket är Newtons metod på systemet $g(x) = 0$.