LANA 2016 Demo 16-1

NA7.7 Modellen $y(t) = c_1 + c_2 e^{-c_3 t}\sin(c_4 t)$ ska anpassas till datapunkter (t_i, y_i), $i = 1, \dots, m$. Ange residual, Jacobian, minstakvadratproblemet, samt Gauss-Newtons metod för lösningen.

Lösning Beteckna $c := (c_1, c_2, c_3, c_4)$. Residualvektorn $r(c)$ har komponenter $r_i(c) := c_1 + c_2 e^{-c_3 t_i}\sin(c_4 t_i) - y_i$, dess Jacobian är matrisen $J_r(c) = \left(\frac{\partial r_i}{\partial c_j}\right)$, med

$$\frac{\partial r_i}{\partial c_1} = 1, \quad \frac{\partial r_i}{\partial c_2} = e^{-c_3 t_i}\sin(c_4 t_i), \quad \frac{\partial r_i}{\partial c_3} = -t_i c_2 e^{-c_3 t_i}\sin(c_4 t_i),$$

$$\frac{\partial r_i}{\partial c_4} = t_i c_2 e^{-c_3 t_i}\cos(c_4 t_i).$$

Minstakvadratproblemet är att minimera $f(c) := \|r(c)\|_2^2$.

Gauss-Newtons metod för detta problem fås genom att beräkna $c^{(k+1)}$ som lösningen till det linjära minstakvadratproblem som fås när vi approximerar $r(c) \approx r(c^{(k)}) + J_r(c^{(k)})(c - c^{(k)})$, alltså $c^{(k+1)}$ minimerar $f_k(c) := \|r(c^{(k)}) + J_r(c^{(k)})(c - c^{(k)})\|_2^2$.

Via normalekvationerna får vi alltså

$$\begin{aligned} c^{(k+1)} &= \left(J_r(c^{(k)})^T J_r(c^{(k)})\right)^{-1}\left(J_r(c^{(k)})^T J_r(c^{(k)})\, c^{(k)} - J_r(c^{(k)})^T r(c^{(k)})\right) \\ &= c^{(k)} - \left(J_r(c^{(k)})^T J_r(c^{(k)})\right)^{-1} J_r(c^{(k)})^T r(c^{(k)}). \end{aligned}$$

LANA 2016 Demo 16-2

NA 7.10 Låt $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ med $f(x) := \frac{1}{2}(x_1^2 - x_2)^2 + \frac{1}{2}(1 - x_1)^2$.

(a) Var antar f sitt minimum?

Lösning Vi ser att $f(x) \geq 0$ med $f(x) = 0 \Longleftrightarrow x_1 = x_2 = 1$,
så minimum i punkten $x^* = (1, 1)$.

(b) Gör en iteration med Newtons metod från punkten $x^{(0)} = (\overset{2,2}{\cancel{0,0}})$.

Lösning Newtons metod för att bestämma kritiska/stationära punkter, alltså $\nabla f(x) = 0$, lyder

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

där $\nabla^2 f(x) := \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)$ är Hessianen av f. Här har vi

$$\nabla^2 f(x) = \begin{bmatrix} 6x_1^2 - 2x_2 - 1 & -2x_1 \\ -2x_1 & 1 \end{bmatrix}, \quad \nabla f(x) = \begin{bmatrix} 2x_1^3 - 2x_1 x_2 - x_1 + 1 \\ x_2 - x_1^2 \end{bmatrix}.$$

Vi får (se Matlabexempel): $x^{(1)} = x^{(0)} - \nabla^2 f(x^{(0)})^{-1} \nabla f(x^{(0)}) = \frac{1}{3}\begin{bmatrix} 7 \\ 16 \end{bmatrix}$.

(c) På vilket sätt är $x^{(1)}$ bättre än $x^{(0)}$?

Svar $f(x^{(1)}) \approx 0.89 < 2.5 = f(x^{(0)})$, så målfunktionens värde har minskat.

(d) På vilket sätt är $x^{(1)}$ sämre än $x^{(0)}$?

Svar $\|x^{(1)} - x^*\|_2 > \|x^{(0)} - x^*\|_2$, så $x^{(1)}$ är längre bort från x^*.

NA 7.14 (a) Bestäm alla kritiska punkter till $f(x) := 2x_1^3 - 3x_1^2 - 6x_1x_2(x_1 - x_2 - 1)$.

Lösning Kritiska punkter uppfyller $\nabla f(x) = 0$. Vi har

$$\nabla f(x) = 6\begin{bmatrix}(x_1 - x_2)(x_1 - x_2 - 1)\\ x_1(2x_2 - x_1 + 1)\end{bmatrix},$$

så första komponenten är 0 då $x_1 = x_2$ eller $x_1 = x_2 + 1$,
och andra komponenten är 0 då $x_1 = 0$ eller $2x_2 = x_1 - 1$.

Vi kan alltså ha $x_1 = x_2 = 0$, eller $x_1 = x_2$ och $2x_2 = x_1 - 1$ $\Rightarrow x_1 = x_2 = -1$, eller $x_1 = x_2 + 1 = 0 \Rightarrow x_1 = 0,\ x_2 = -1$, eller $x_1 = x_2 + 1$ och $2x_2 = x_1 - 1 \Rightarrow x_1 = 1,\ x_2 = 0$.

Punkterna är alltså $(0,0)$, $(-1,-1)$, $(0,-1)$, och $(1,0)$.

(b) Karaktärisera punkterna som maximum, minimum eller sadelpunkt.

Lösning Punkterna kan karaktäriseras via Hessianen $\nabla^2 f(x)$:

Om alla egenvärden till $\nabla^2 f(x)$ i en kritisk punkt är...
- ... positiva (positivt definit) så är punkten ett minimum,
- ... negativa (negativt definit) så är punkten ett maximum,
- ... både positiva och negativa (indefinit) så är punkten en sadelpunkt.

Här har vi

$$\nabla^2 f(x) = 6\begin{bmatrix}2(x_1 - x_2) - 1 & -(2(x_1 - x_2) - 1)\\ -(2(x_1 - x_2) - 1) & 2x_1\end{bmatrix}.$$

I $x = (0,0)$ har vi $\nabla^2 f(x) = \begin{bmatrix}-6 & 6\\ 6 & 0\end{bmatrix}$ med egenvärden $\lambda = -3 \pm 3\sqrt{3}$,

i $x = (-1,-1)$ har vi $\nabla^2 f(x) = \begin{bmatrix}-6 & 6\\ 6 & -12\end{bmatrix}$ med egenvärden $-9 \pm 3\sqrt{5}$,

i $x = (0,-1)$ har vi $\nabla^2 f(x) = \begin{bmatrix}6 & -6\\ -6 & 0\end{bmatrix}$ med egenvärden $3 \pm 3\sqrt{3}$,

i $x = (1,0)$ har vi $\nabla^2 f(x) = \begin{bmatrix}6 & -6\\ -6 & 12\end{bmatrix}$ med egenvärden $9 \pm 3\sqrt{5}$

så $(0,0)$ är en sadelpunkt, $(-1,-1)$ ett maximum, $(0,-1)$ en sadelpunkt, och $(1,0)$ ett minimum.

LANA 2016 Demo 16-4

Kommentar angående derivering av funktionen $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ med $f(x) = x^T A x + b^T x$ (relevant för NA 7.11):

Det gäller att $\nabla f(x)^T e_i = \dfrac{\partial f}{\partial x_i}(x) = \lim\limits_{h \to 0} \dfrac{f(x+he_i) - f(x)}{h}$

$$= \lim_{h\to 0} \frac{1}{h}\left((x+he_i)^T A (x+he_i) + b^T(x+he_i) - x^T A x - b^T x \right)$$

$$= \lim_{h\to 0} \frac{1}{h}\left(h e_i^T A x + h x^T A e_i + h^2 e_i^T A e_i + h b^T e_i \right)$$

$$= \lim_{h\to 0} \left(e_i^T A x + x^T A e_i + b^T e_i + h e_i^T A e_i \right)$$

$$= e_i^T A x + x^T A e_i + b^T e_i = (Ax)^T e_i + (A^T x)^T e_i + b^T e_i$$

$$= (Ax + A^T x + b)^T e_i \quad , i = 1, \ldots, n$$

så $\nabla f(x) = Ax + A^T x + b.$