Exercise #9

March 14, 2023

Problem 1.

In this exercise, we study the Gauß–Newton method for solving the least-squares problem corresponding to the (overdetermined and inconsistent) system of equations

$$\begin{aligned} x + y &= 1, \\ x - y &= 0, \\ xy &= 2. \end{aligned}$$

To that end, we define

$$\begin{aligned} r_1(x, y) &:= x + y - 1, \\ r_2(x, y) &:= x - y, \\ r_3(x, y) &:= xy - 2, \end{aligned}$$

and

$$f(x, y) := \frac{1}{2} \sum_{j=1}^{3} r_j(x, y)^2.$$

We denote moreover by $J = J(x, y)$ the Jacobian of r.

a) Show that the function f is non-convex, but that it has a unique minimiser (x^*, y^*).

b) Show that the matrix $J^T J$ required in the Gauß–Newton method is positive definite for all x, y.

c) Perform one step of the Gauß–Newton method (without line search) for the solution of this least-squares problem. Use the initial value $(x_0, y_0) = (0, 0)$.

Solution.

Have already discussed in the class!

Problem 2.

Let

$$f(x) = x_1^4 + 2x_2^4 + x_1 x_2 + x_1 - x_2 + 2.$$

Starting at the point $x_0 = (0, 0)$ compute explicitly one step for the trust region method with the model function $m(p) = f(x_0) + g^T p + \frac{1}{2} p^T B p$, where $g = \nabla f(x_0)$, $B = \nabla^2 f(x_0)$, and the trust region radius $\Delta = 1$.

Solution.

Have already discussed in the class!

Problem 3. (Problem 4.1 in N&W)

Let

$$f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

At $x = (0, -1)$ draw the contour lines of the quadratic model

$$\min_p m(p) = f(x) + \langle \nabla f(x), p \rangle + \frac{1}{2} \langle p, Bp \rangle \qquad \text{subject to } \|p\| \leq \Delta, \tag{1}$$

assuming that B is the Hessian of f. Draw the family of solutions of (1) as the trust region radius varies from $\Delta = 0$ to $\Delta = 2$. Repeat this at $x = (0, 0.5)$.

Solution.

The gradient and Hessian of the objective function $f(x) = 10(x_2 - x_1^2)^2 + (1 - x_1)^2$ are

$$\nabla f(x) = \begin{pmatrix} -40(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 20(x_2 - x_1^2), \end{pmatrix}$$

$$\text{and } \nabla^2 f(x) = \begin{pmatrix} 40(3x_1^2 - x_2) + 2 & -40x_1 \\ -40x_1 & 20 \end{pmatrix}, \text{ respectively.}$$

We see that f has only one minimum, that is $x^* = (1, 1)^T$. For $x_k = (0, -1)^T$, we have that

$$f_k = 11, \nabla f_k = \begin{pmatrix} -2 \\ -20 \end{pmatrix} \text{ and } \nabla^2 f_k = \begin{pmatrix} 42 & 0 \\ 0 & 20 \end{pmatrix}.$$

Hence,

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla^2 f_k p = 11 - 2p_1 - 20p_2 + 21p_1^2 + 10p_2^2,$$

is a strictly convex quadratic function with minimizer $p_k^B = -\nabla^2 f_k^{-1} \nabla f_k = (\frac{1}{21}, 1)^T$. We can rewrite $m_k(p)$ to

$$m_k(p) = 21(p_1 - \frac{1}{21})^2 + 10(p_2 - 1)^2 + \frac{20}{21}.$$

Thus the contours of $m_k(p)$ are ellipses. We can get the solution of

$$\min_{\|p\| \leq \Delta} m_k(p) \tag{2}$$

as

$$\begin{cases} \|P\| = \Delta, & \|p_k^B\| > \Delta, \\ p = p_k^B, & \text{otherwise.} \end{cases}$$

For $x_k = (0, 0.5)^T$, we have that

$$f_k = \frac{7}{2}, \ \nabla f_k = \begin{pmatrix} -2 \\ 10 \end{pmatrix}, \ \nabla^2 f_k = \begin{pmatrix} -18 & 0 \\ 0 & 20 \end{pmatrix}.$$

Hence,

$$m_k(p) = \frac{7}{2} - 2p_1 + 10p_2 - 9p_1^2 + 10p_2^2 = -9(p_1 + \frac{1}{9})^2 + 10(p_2 + \frac{1}{2})^2 + \frac{10}{9}$$

has no global maximum or minimum, but a saddle point at $(-\frac{1}{9}, -\frac{1}{2})^T$. Since we have no minimum in the interior of the trust-region, the minimizer p_k of (2) will also here satisfy $\|p_k\| = \Delta$. Observe that the contours of $m_k(p)$ will be hyperbolas. Contour plots of $m_k(x)$, the family of solutions of (2) for $\Delta \in (0, 2]$ and trust region radii for the two different x_k are shown in Figure 1.

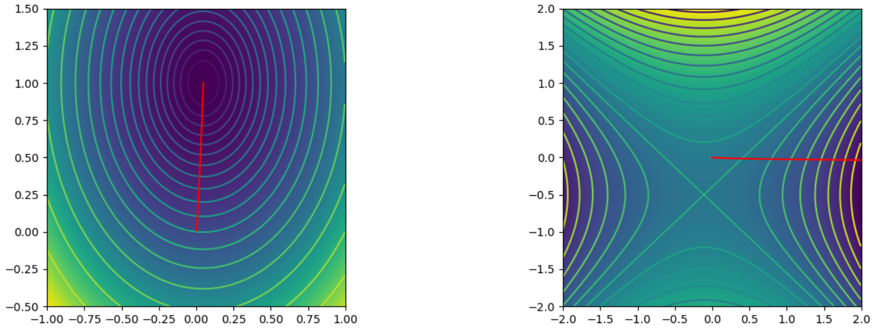

Figure 1: Contour plot of $m_k(p)$ and the family of solutions of (5)(in red)

Problem 4. (Problem 4.5 in N&W)

Let $\phi(\tau) = m_k(\tau p_k^s)$, where $p_k^s = -\frac{\Delta_k}{\|g_k\|} g_k$ and $m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B p$ with $B \in \mathbb{R}^{d\times d}$ symmetric. Show that the minimizer of $\phi(\tau)$ subject to $\|\tau p_k^s\| \leq \Delta_k$ and $\tau \geq 0$ is given as

$$\begin{cases} 1, & \text{if } g_k^T B_k g_k \leq 0, \\ \min\left(\frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}, 1\right), & \text{otherwise.} \end{cases} \tag{3}$$

Solution.

With the definition of $m_k(p)$ and p_k^s, we can write

$$\begin{aligned} \phi(\tau) &= f_k + g_k^T \left(-\frac{\tau \Delta_k g_k}{\|g_k\|}\right) + \frac{1}{2}\left(-\frac{\tau \Delta_k g_k}{\|g_k\|}\right)^T B_k \left(-\frac{\tau \Delta_k g_k}{\|g_k\|}\right) \\ &= f_k - \tau \Delta_k \|g_k\| + \frac{\frac{1}{2}\tau^2 \Delta_k^2 g_k^T B_k g_k}{\|g_k\|^2}. \end{aligned}$$

Furthermore, observe that the constraint $\|\tau p_k^s\| \leq \Delta_k$ is equivalent to $|\tau| \leq 1$. Together with $\tau > 0$, this means that $\tau \in (0, 1]$. First, if $g_k = 0$, then $\phi(\tau)$ is a constant, so $\tau = 1$ will be a minimizer. This is in agreement with (3). Second, if $g_k^T B_k g_k = 0$, then $\phi(\tau)$ is linear and decreasing, so the minimizer is the highest possible value, i.e., $\tau = 1$, which is also in agreement with (3). Lastly, if $g_k^T B_k g_k \neq 0$, then we have a critical point where

$$\phi'(\tau) = -\Delta_k \|g_k\| \frac{\tau \Delta_k^2 g_k^T B_k g_k}{\|g_k\|^2} = 0,$$

that is

$$\tau = \frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}. \tag{4}$$

Now if $g_k^T B_k g_k < 0$ then this is a maximizer, and the minimizer must be at the endpoints of the interval $(0, 1]$. Since we have $\phi(0) > \phi(1)$, the minimizer must be $\tau = 1$. This is in agreement with (3). Otherwise, if $g_k^T B_k g_k > 0$ then (4) is a minimizer. If this value in bigger than 1, then ϕ is decreasing across the interval $(0, 1]$, and thus the minimizer is $\tau = 1$, as in (3). If (4) is less than 1, then (4) is the minimizer. This is captured by (3).