Exercise 9

March 12, 2024

Problem 1.

Let

$$f(x) = x_1^4 + 2x_2^4 + x_1x_2 + x_1 - x_2 + 2.$$

Starting at the point $x_0 = (0, 0)$ compute explicitly one step for the trust region method with the model function

$$m(p) = f(x_0) + \langle g, p\rangle + \frac{1}{2}\langle p, Bp\rangle$$

where $g = \nabla f(x_0)$, $B = \nabla^2 f(x_0)$, and the trust region radius is $\Delta = 1$.

Solution.

We invoke Theorem 4.1 in Nocedal & Wright, which says that p_0 is a global minimiser to the trust-region subproblem

$$\min_{\|p\| \le \Delta} m(p),$$

with $\Delta = 1$, if and only if there exists a $\lambda \ge 0$ such that

$$(B + \lambda \operatorname{Id})p_0 = -g, \tag{1}$$
$$\lambda(\Delta - \|p_0\|) = 0, \text{ and} \tag{2}$$
$$B + \lambda \operatorname{Id} \quad \text{is positive semi-definite.} \tag{3}$$

Routine calculations yield that

$$g = \nabla f(x_0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \text{and} \qquad B = \nabla^2 f(x_0) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Since B has eigenvalues ± 1, we must have $\lambda \ge 1$ in order to guarantee the positive semi-definiteness of the matrix $B + \lambda \operatorname{Id}$. As a result, from complementarity condition (2) we must have $\|p_0\| = 1$, so p_0 lies on the trust-region boundary.

The solution of (1) equals

$$p_0 = \frac{1}{1 - \lambda} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

provided $\lambda \neq 1$ (there is no solution for $\lambda = 1$), and from the conditions $\|p_0\| = 1$ and $\lambda > 1$, we thus end up with

$$\lambda = 1 + \sqrt{2}, \qquad \text{and} \qquad p_0 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Next step is therefore $x_1 = x_0 + p_0 = p_0$.

Problem 2.

Let

$$f(x) = \frac{1}{2}x_1^2 + x_2^2,$$

let $x_0 = (1, 1)$, and define the model function $m(p) = f(x_0) + \langle g, p\rangle + \frac{1}{2}\langle p, Bp\rangle$ with $g = \nabla f(x_0)$ and $B = \nabla^2 f(x_0)$.

a) Compute explicitly the next step p in the trust region method using values of $\Delta = 2$ and $\Delta = 5/6$.

b) Compute for all $\Delta > 0$ the next step in the dogleg method.

Solution.

Part a)

Note first that

$$g = \nabla f(x_0) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \text{and} \qquad B = \nabla^2 f(x_0) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},$$

and that the unconstrained minimizer of m equals $p_0^{\mathrm{B}} = -B^{-1}g = -(1, 1)$. When $\Delta = 2$, this point is feasible—indeed, $\|p_0^{\mathrm{B}}\| = \sqrt{2} < 2$—and hence, we compute the next step with $p_0 = p_0^{\mathrm{B}}$ as $x_1 = x_0 + p_0 = (0, 0)$, which turns out to be the global minimizer of f.

If, however, $\Delta = 5/6$, then (1) from Theorem 4.1 in N&W implies that

$$p_0 = -\begin{bmatrix} 1/(1+\lambda) \\ 2/(2+\lambda) \end{bmatrix}$$

for some $\lambda \geq 0$. We cannot have $\lambda = 0$, because then $p_0 = p_0^{\mathrm{B}}$, which is infeasible. Thus $\lambda > 0$ and $\|p_0\| = \Delta = 5/6$ by complementarity condition (2). Written out and simplifying, the latter equation becomes

$$\begin{aligned} 0 &= 25\lambda^4 + 150\lambda^3 + 145\lambda^2 - 132\lambda - 188 \\ &= (\lambda - 1)(25\lambda^3 + 175\lambda^2 + 300\lambda + 188). \end{aligned}$$

Since the second factor in the last expression is positive for all $\lambda \geq 0$, we infer that $\lambda = 1$ is the only possibility. This gives

$$p_0 = (-1/2, -2/3) \qquad \text{and} \qquad x_1 = x_0 + p_0 = (1/2, 1/3).$$

(Note that condition (3) is automatically satisfied because B is positive definite.)

Part b)

If $\Delta \geq 2$, the full step $p_0 = p_0^{\mathrm{B}}$ is feasible, yielding $x_1 = x_0 + p_0 = (0, 0)$.

Next, the steepest descent step equals

$$p_0^{\mathrm{U}} = -\frac{g^\top g}{g^\top B g} g = -\begin{bmatrix} 5/9 \\ 10/9 \end{bmatrix}$$

and satisfies $\|p_0^{\mathrm{U}}\| = 5\sqrt{5}/9 \approx 1.24$. If $\Delta \leq \|p_0^{\mathrm{U}}\|$, the dogleg method chooses p_0 to lie on the "steepest descent trajectory", scaled to lie on the boundary of the trust-region, so that

$$p_0 = \frac{\Delta}{\|p_0^{\mathrm{U}}\|} p_0^{\mathrm{U}} = -\frac{\Delta}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

This yields a new step $x_1 = \left(1 - \frac{\Delta}{\sqrt{5}}, 1 - \frac{2\Delta}{\sqrt{5}}\right)$. Observe that for $\Delta = 5/6$, this gives $x_1 \approx (0.63, 0.25)$, which is not too far from the optimal x_1 found in the previous problem.

For the remaining case $5\sqrt{5}/9 < \Delta < 2$, we follow the dogleg path

$$p(\tau) = p_0^{\mathrm{U}} + \tau \left(p_0^{\mathrm{B}} - p_0^{\mathrm{U}}\right), \qquad \tau \in (0, 1)$$

until it hits the boundary of the trust-region, that is, when

$$\begin{aligned} \Delta^2 = \|p(\tau)\|^2 &= \|p_0^{\mathrm{U}}\|^2 + 2\tau \left(p_0^{\mathrm{B}} - p_0^{\mathrm{U}}\right)^\top p_0^{\mathrm{U}} + \tau^2 \|p_0^{\mathrm{B}} - p_0^{\mathrm{U}}\|^2 \\ &= \frac{17}{81}\tau^2 + \frac{20}{81}\tau + \frac{125}{81} \end{aligned}$$

Solving this quadratic equation with respect to τ gives

$$\tau = -\frac{10}{17} + \frac{9}{17}\sqrt{17\Delta^2 - 25},$$

where the other solution has been discarded since it results in $\tau < 0$. Next step is therefore $x_1 = x_0 + p(\tau)$, with τ as above.

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

$$\min_p m(p) = f(x) + \langle \nabla f(x), p \rangle + \frac{1}{2}\langle p, Bp \rangle \qquad \text{subject to } \|p\| \leq \Delta, \quad (4)$$

assuming that B is the Hessian of f. Draw the family of solutions of (4) 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

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

and

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

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

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

Hence,

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

which is a strictly convex quadratic function with minimizer

$$p_k^B = -\nabla^2 f_k^{-1} \nabla f_k = \begin{pmatrix} \frac{1}{21} \\ 1 \end{pmatrix}.$$

We can rewrite $m_k(p)$ as

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

Thus the contours of $m_k(p)$ are ellipses centered at $p = (\frac{1}{21}, 1)$. We then get that $\|p_k^B\| = \sqrt{1 + \frac{1}{21^2}} < 2$, so for $\Delta \geq \|p_k^B\|$ we get $p_k = p_k^B$. For $\Delta < \|p_k^B\|$ we can apply (1) from Theorem 4.1 in N&W, to get that the solution is given by $p = -(\nabla^2 f_k + \lambda I)^{-1} \nabla f_k$. This then gives

$$p_{k,\lambda} = -\begin{pmatrix} \frac{1}{42+\lambda} & 0 \\ 0 & \frac{1}{20+\lambda} \end{pmatrix} \begin{pmatrix} -2 \\ -20 \end{pmatrix} = \begin{pmatrix} \frac{2}{42+\lambda} \\ \frac{20}{20+\lambda} \end{pmatrix}.$$

For this to be a valid solution, we require that $B + \lambda I$ be positive definite, which it clearly is. Finally, we need to then find $\lambda > 0$ such that $\|p_{k,\lambda}\| = \Delta$. Now, we have a parameterization of the solution by using $0 \leq \lambda$ as a parameter for the solution curve

$$\left\{ \left(\frac{2}{42+\lambda}, \frac{20}{20+\lambda} \right) : \lambda \in [0, \infty) \right\}.$$

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

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

Hence,

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

Because $\nabla^2 f_k$ is indefinite, the function m_k has no global maximum or minimum, though it has 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 (4) will satisfy $\|p_k\| = \Delta$. Observe that the contours of $m_k(p)$ will be hyperbolas.

To find the solutions, we again use (1) from Theorem 4.1 in N&W, $p = -(\nabla^2 f_k + \lambda I)^{-1} \nabla f_k$. This gives

$$p_{k,\lambda} = -\begin{pmatrix} \frac{1}{\lambda - 18} & 0 \\ 0 & \frac{1}{\lambda + 20} \end{pmatrix} \begin{pmatrix} -2 \\ 10 \end{pmatrix} = \begin{pmatrix} \frac{2}{\lambda - 18} \\ \frac{-10}{20+\lambda}. \end{pmatrix}$$

In order to guarantee that $\nabla^2 f_k + \lambda I$ is positive definite, we require $\lambda > 18$. Finally, we need to pick $\lambda(\Delta)$ so that $\|p_{k,\lambda(\Delta)}\| = \Delta$, and then we have the solution. The curve of valid solutions for $\Delta \in (0, 2]$ is then given by

$$\left\{ \left(\frac{2}{\lambda - 18}, \frac{-10}{20+\lambda} \right) : \lambda \in [\lambda(2), \infty) \right\}.$$

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 minimiser of $\phi(\tau)$ subject to $\|\tau p_k^s\| \leq \Delta_k$ and $\tau \geq 0$ is given as

$$\tau = \begin{cases} 1, & \text{if } \langle g_k, B_k g_k \rangle \leq 0, \\ \min\left\{ \dfrac{\|g_k\|^3}{\Delta_k \langle g_k, B_k g_k \rangle}, 1 \right\}, & \text{otherwise.} \end{cases} \tag{5}$$

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 (5). 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 (5). 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

$$\tau = \frac{\|g_k\|^3}{\Delta_k g_k^T B_k g_k}. \tag{6}$$

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 (5). Otherwise, if $g_k^T B_k g_k > 0$ then (6) 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 (5). If (6) is less than 1, then (6) is the minimizer. This is captured by (5).