Norwegian University of Science and Technology Department of Mathematical Sciences	TMA4180 Optimisation I Spring 2018 **Solutions to exercise set 6**

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 $r = (r_1, r_2, r_3) \colon \mathbb{R}^2 \to \mathbb{R}^3$.

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

Solution: The gradient and Hessian of f equal

$$\nabla f(x, y) = J^\top r = \begin{bmatrix} 1 & 1 & y \\ 1 & -1 & x \end{bmatrix} \begin{bmatrix} x + y - 1 \\ x - y \\ xy - 2 \end{bmatrix} = \begin{bmatrix} 2(x - y) + xy^2 - 1 \\ 2(y - x) + yx^2 - 1 \end{bmatrix}$$

and

$$\begin{aligned} \nabla^2 f(x, y) &= J^\top J + r_1 \nabla^2 r_1 + r_2 \nabla^2 r_2 + r_3 \nabla^2 r_3 \\ &= \begin{bmatrix} 2 + y^2 & xy \\ xy & 2 + x^2 \end{bmatrix} + 0 + 0 + r_3 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 + y^2 & 2(xy - 1) \\ 2(xy - 1) & 2 + x^2 \end{bmatrix}. \end{aligned}$$

Since, for example, $\nabla^2 f(-1, 1)$ has eigenvalues -1 and 7, it follows that f is non-convex. However, f does have a unique minimiser: it is smooth and coercive, and thus we infer that there is a global minimiser, which must also be a stationary point. Coercivity can be seen this way: if $f \le C^2/2$ for some $C > 0$, then $r_1^2 \le C$ and $r_2^2 \le C$. If we put $z_1 = x + y$ and $z_2 = x - y$, then these inequalities imply $|z_1| \le C + 1$ and $|z_2| \le C$. As a consequence we have the

boundedness of $|x| = |(z_1 + z_2)/2| \leq C + 1/2$ and $|y| = |(z_1 - z_2)/2| \leq C + 1/2$. Therefore, if either $|x| \to \infty$ or $|y| \to \infty$, then also $f(x, y) \to \infty$.
The stationary point for f must satisfy the equations

$$xy(x + y) = 2 \qquad \text{and} \qquad xy(x - y) = 4(x - y),$$

which can be seen by adding and subtracting the equations in the system $\nabla f = 0$. If $x \neq y$, then $xy = 4$ from the second equation, so that $y = \frac{1}{2} - x$ from the first. But as $4 = xy = x(\frac{1}{2} - x)$ has complex solutions in x, we reject this case. Therefore $x = y$, which gives solutions $x = y = 1$ from the first equation. Thus the function has only one stationary point, and since the minimum exists and must satisfy the optimality conditions, this is the point of global minimum.

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

Solution: Remember first that any matrix of the form $J^\top J$ is symmetric positive semi-definite (SPSD), which follows from

$$v^\top (J^\top J)v = (Jv)^\top (Jv) = \|Jv\|^2 \geq 0.$$

Moreover, SPSD matrices are characterised by having nonnegative eigenvalues, while a matrix is symmetric positive definite (SPD) if and onlfy if it has strictly positive eigenvalues.
Computing $\det J^\top J = 2(x^2 + y^2 + 2) > 0$, we see that $J^\top J$ is invertible. In particular, all eigenvalues are nonzero, and hence, strictly positive (being nonnegative). Therefore, $J^\top J$ is positive definite.

c) Show that the Gauß–Newton method with Wolfe line search for the minimisation of f converges for all initial values (x_0, y_0) to the unique solution of the nonlinear least squares problems.

Solution: We show that $J(x, y)$ satisfies the "full-rank condition"

$$\|J(x, y)v\| \geq \gamma \|v\|$$

for all $(x, y) \in \mathbb{R}^2$, where $\gamma > 0$ is a constant. Theorem 10.1 in N&W then implies that the Gauß–Newton method with Wolfe line search converges for all initial values.
Now,

$$\begin{aligned}\|J(x, y)v\|^2 &= (v_1 + v_2)^2 + (v_1 - v_2)^2 + (yv_1 + xv_2)^2 \\ &\geq 2\left(v_1^2 + v_2^2\right) = 2\|v\|^2,\end{aligned}$$

and so we may put $\gamma = \sqrt{2}$ to get the desired inequality.

d) 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: With $(x_0, y_0) = (0, 0)$, we have

$$J^\top J = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad J^\top r = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Solving the linear system $J^\top J p = -J^\top r$ gives $p = (1/2, 1/2)$, so that

$$(x_1, y_1) = (x_0, y_0) + p = (1/2, 1/2).$$

2 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 $m(p) = f(x_0) + g^\top p + \frac{1}{2}p^\top Bp$, where $g = \nabla f(x_0)$, $B = \nabla^2 f(x_0)$, and the trust region radius $\Delta = 1$.

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

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

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

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

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

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 $x_1 = x_0 + p_0 = p_0$.

3 Let

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

put $x_0 = (1,1)$, and define the model function $m(p) = f(x_0) + g^\top p + \frac{1}{2}p^\top Bp$ 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 $\Delta = 5/6$.

Solution: 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.)

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

Solution: 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

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

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

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

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.

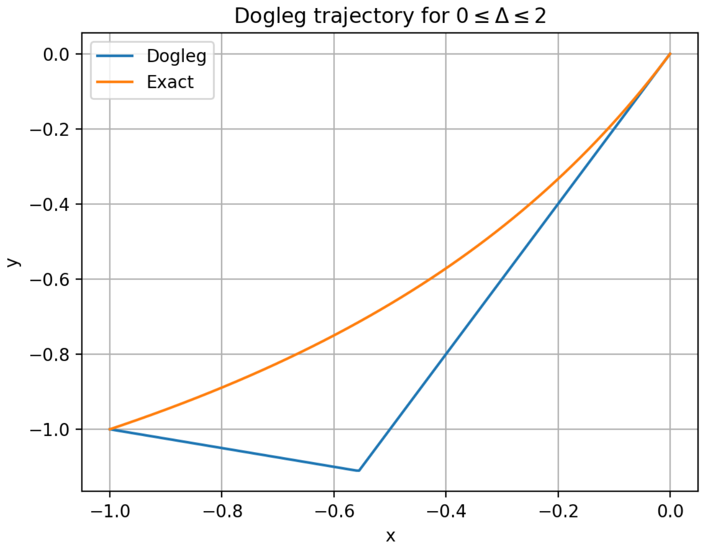

Figure 1: Comparion of the dogleg trajetory vs the exact solutions to the trust region problem.

See Figure 1 which plots both the dogleg trajectory and the exact solution to the trust-region problem for $0 \leq \Delta \leq 2$. The plot has been obtained using mostly symbolic computations from `sympy`; however we also used numerical root finding to solve the Trust Region subproblem exactly, since this is faster. See the source code on the wiki page.