Norwegian University of Science and Technology
Department of Mathematical Sciences

Solutions to exercise set 5

1 Observe first that exact line search implies that $\nabla f_{k+1}^\top p_k = 0$ (and $\nabla f_{k+1}^\top s_k = 0$ because $s_k = \alpha_k p_k$). Indeed, minimising f at the current iterate x_k in the direction p_k, that is, finding an optimal step length α_k satisfying

$$\alpha_k \in \arg\min_{\alpha>0} f(x_k + \alpha p_k),$$

means that α_k is a stationary point of $\phi\colon \alpha \mapsto f(x_k + \alpha p_k)$. Differentiating ϕ yields

$$0 = \phi'(\alpha_k) = \nabla f(x_k + \alpha_k p_k)^\top p_k = \nabla f_{k+1}^\top p_k,$$

as desired.

Note next that both this variant of the BFGS method and the Hestenes–Stiefel method iterate on the form

$$x_{k+1} = x_k + \alpha_k p_k.$$

Therefore, assuming exact line search and $p_0 = -\nabla f_0$, it suffices to show that the search directions for the two methods coincide. With

$$s_k = x_{k+1} - x_k = \alpha_k p_k \quad \text{and} \quad y_k = \nabla f_{k+1} - \nabla f_k,$$

we calculate search directions in the BFGS variant as

$$\begin{aligned}
p_{k+1} &= -H_{k+1}\nabla f_{k+1} \\
&= -\left(\mathrm{Id} - \frac{s_k y_k^\top}{y_k^\top s_k}\right)\left(\nabla f_{k+1} - \frac{y_k}{y_k^\top s_k}(s_k^\top \nabla f_{k+1})\right) + \frac{s_k}{y_k^\top s_k}(s_k^\top \nabla f_{k+1}) \\
&= -\left(\mathrm{Id} - \frac{s_k y_k^\top}{y_k^\top s_k}\right)\nabla f_{k+1} \\
&= -\nabla f_{k+1} + \frac{\nabla f_{k+1}^\top y_k}{y_k^\top s_k} s_k \\
&= -\nabla f_{k+1} + \frac{\nabla f_{k+1}^\top y_k}{y_k^\top p_k} p_k \\
&= -\nabla f_{k+1} + \frac{\nabla f_{k+1}^\top (\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^\top p_k} p_k \\
&= -\nabla f_{k+1} + \beta_{k+1} p_k.
\end{aligned}$$

Since β_{k+1} equals that of the Hestenes–Stiefel method, we are done.

2 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, any $\lambda \geq 1$ makes $B + \lambda\,\mathrm{Id}$ positive semi-definite. In particular, we must have $\|p_0\| = 1$ from complementarity condition (2), 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 **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 $p_0^{\mathrm{B}} = -B^{-1}g = -(1,1)$ (why?). When $\Delta = 2$, this direction 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) 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

$$\Delta^2 = \|p(\tau)\|^2 = \|p_0^{\mathrm{U}}\|^2 + 2\tau \left(p_0^{\mathrm{B}} - p_0^{\mathrm{U}}\right)^\top p_0^{\mathrm{U}} + \tau^2 \|p_0^{\mathrm{B}} - p_0^{\mathrm{U}}\|^2.$$

Solving this quadratic equation with respect to τ gives

$$\tau = -\frac{1}{17}\left(10 + \sqrt{1377\Delta^2 - 2025}\right),$$

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.

4 **a)** 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 $\|(x,y)\| \to \infty$, then either $|x| \to \infty$ or $|y| \to \infty$. In either case, it is impossible for all three of r_1, r_2, and r_3 to stay bounded. As such, $f(x,y) \to \infty$.

By adding and equating (since both equal 0) the two components of ∇f, we find that stationary points must satisfy

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

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 $x = y = \pm 1$ from the first equation. Evaluating $f(1,1) = 1$ and $f(-1,-1) = 5$, we conclude that $(x^*, y^*) = (1,1)$ is the unique minimiser.

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

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

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.

Another way to argue is to show that the inequality in ($\star$) is strict for all nonzero v unless $v \in \ker J$. By the rank–nullity theorem,

$$\dim \ker J = 2 - \operatorname{rank} J = 0.$$

Thus $Jv = 0$ only if $v = 0$, and $J^\top J$ is SPD.

c) 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,

$$\begin{aligned} \|J(x,y)v\|^2 &= (v_1 + v_2)^2 + (v_1 - v_2)^2 + (ys_1 + xs_2)^2 \\ &\geq 2\left(v_1^2 + v_2^2\right) = 2\|v\|^2, \end{aligned}$$

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

d) 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

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