TMA4180
Solutions to recommended exercises in Chapter 6 of N&W

Exercise 6.1a

We will assume that f is twice continuously differentiable. Recall that a function f is strongly convex if its Hessian, $\nabla^2 f(x)$, is positive definite. Hence, there exists a $\sigma > 0$ s.t.

$$p^T \nabla^2 f(x) p \geq \sigma \|p\|^2,$$

for any p. From Taylor's Theorem (Eq. (2.5) in N&W), we have that

$$\nabla f(x_{k+1}) = \nabla f(x_k) + \int_0^1 \left[\nabla^2 f(x_k + t\alpha_k p_k) \alpha_k p_k \right] \mathrm{d}t.$$

We can now deduce that

$$\begin{aligned} s_k^T y_k &= \alpha_k p_k^T (\nabla f(x_{k+1}) - \nabla f(x_k)) \\ &= \alpha_k^2 \int_0^1 \left[p_k^T \nabla^2 f(x_k + t\alpha_k p_k) p_k \right] \mathrm{d}t \\ &\geq \alpha_k^2 \sigma \|p\|^2 > 0. \end{aligned}$$

Exercise 6.3

We want to show that

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T$$

and

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T}{s_k^T B_k s_k} + \rho_k y_k y_k^T$$

are inverses of each other, i.e., that $H_{k+1} B_{k+1} = I$ for all k. We will show this by induction. Obviously, $H_0 B_0 = I$ by construction. Further, we let $k \geq 1$ and assume that $H_k B_k = I$. We need to show that $H_{k+1} B_{k+1} = I$. For better readability we skip subindex k in the derivation. Thus, we write

$$\begin{aligned} H_{k+1} &= (I - \rho s y^T) H (I - \rho y s^T) + \rho s s^T \\ &= H - \rho s y^T H - \rho H y s^T + \rho^2 s y^T H y s^T + \rho s s^T. \end{aligned}$$

From now on let $\frac{1}{s^T Bs} = \gamma$. We can expand the product $H_{k+1}B_{k+1}$ as follows,

$$\begin{aligned} H_{k+1}B_{k+1} = &\Big[HB - \rho sy^T HB - \rho Hys^T B + \rho^2 sy^T Hys^T B + \rho ss^T B\Big] \\ &+ \Big[-\gamma HBss^T B + \rho\gamma sy^T HBss^T B + \rho\gamma Hys^T Bss^T B \\ &\quad - \rho^2\gamma sy^T Hys^T Bss^T B - \rho\gamma ss^T Bss^T B\Big] \\ &+ \Big[\rho Hyy^T - \rho^2 sy^T Hyy^T - \rho^2 Hys^T yy^T + \rho^3 sy^T Hys^T yy^T + \rho^2 ss^T yy^T\Big]. \end{aligned}$$

By using the induction hypothesis $HB = I$ and the relations $\rho = \frac{1}{y^T s}$ and $\gamma = \frac{1}{s^T Bs}$, one can verify that every term but the first has a negative counterpart and thus vanish. Hence, we are left with $H_{k+1}B_{k+1} = HB = I$.

Exercise 6.12

By the assumptions, it follows that f is strongly convex on $\mathcal{L}$. Thus there exists a unique solution, x^*, to the minimization problem $\min f(x)$ for $x \in \mathcal{L}$. Further, since $\liminf \|\nabla f_k\| = 0$, there exists a subsequence $\nabla f(x_{k_j})$ that converges to 0. If we do a Taylor expansion of $\nabla f(x_{k_j})$ around $\nabla f(x^*)$, we get

$$\nabla f(x_{k_j}) = \nabla f(x^*) + G(x^* + t(x_{k_j} - x^*))(x_{k_j} - x^*)$$

for some $t \in [0, 1]$ and where $G = \nabla^2 f$ is the Hessian of f. We know that $\nabla f(x^*) = 0$, and if we further premultiply with $(x_{k_j} - x^*)^T$ and use the assumptions, we see that

$$(x_{k_j} - x^*)^T \nabla f(x_{k_j}) = (x_{k_j} - x^*)^T G(x^* + t(x_{k_j} - x^*))(x_{k_j} - x^*) \geq m\|x_{k_j} - x^*\|^2.$$

Now, the left-hand-side converges to 0, and hence $\|x_{k_j} - x^*\| \to 0$, i.e., the subsequence x_{k_j} converges to x^*. By continuity of f this implies that $f(x_{j_k}) \to f(x^*)$. In fact, the whole sequence converges, $f(x_k) \to f(x^*)$, since $f(x_k)$ is decreasing (by the sufficient decrease condition).

Next, consider the Taylor expansion of $f(x_k)$ around $f(x^*)$,

$$f(x_k) = f(x^*) + \nabla f(x^*)(x_k - x^*) + \frac{1}{2}(x_k - x^*)^T G(x^* + t(x_k - x^*))(x_k - x^*).$$

By rearranging and using the assumptions, we end up with the estimate

$$f(x_k) - f(x^*) \geq m\|x_k - x^*\|^2.$$

Again, since the left-hand-side converges to 0, $\|x_k - x^*\| \to 0$, and thus $x_k \to x^*$.

Exercise 4.3

The trust-region Newton CG method is implemented in `trustregion_newtonCG.m`, while a function handle for the objective function,

$$f(x) = \sum_{i=1}^{n} \left[(1 - x_{2i-1})^2 + 10(x_{2i} - x_{2i-1}^2)^2 \right]$$

is implemented in `rosenbrock.m`[1]. A useful observation is that the sum is uncoupled. Hence, the first and second order derivatives can be expressed as

$$\begin{aligned}
\frac{\partial f}{\partial x_{2i-1}} &= -2(1 - x_{2i-1}) - 40x_{2i-1}(x_{2i} - x_{2i-1}^2), \\
\frac{\partial f}{\partial x_{2i}} &= 20(x_{2i} - x_{2i-1}^2), \\
\frac{\partial^2 f}{\partial x_{2i-1}\partial x_{2i-1}} &= 2 - 40x_{2i} + 120x_{2i-1}^2, \\
\frac{\partial^2 f}{\partial x_{2i-1}\partial x_{2i}} &= -40x_{2i-1}, \\
\frac{\partial^2 f}{\partial x_{2i}\partial x_{2i}} &= 20, \\
\frac{\partial^2 f}{\partial x_{2i}\partial x_{2i-1}} &= -40x_{2i-1}, \\
\frac{\partial^2 f}{\partial x_i \partial x_j} &= 0, \qquad \text{otherwise.}
\end{aligned}$$

In `lbfgs.m`, the L-BFGS algorithm (p. 179 in N&W) is implemented with the same test problem.

[1]OBS! The n-dimensional Rosenbrock function comes in two different variants, see e.g. Wikipedia: `http://en.wikipedia.org/wiki/Rosenbrock_function`.