Exercise #8

March 07, 2023

Problem 1.

The secant method for the solution of one-dimensional optimisation problems is given by the iteration

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k).$$

Show that this method coincides with both the BFGS and the DFP Quasi-Newton methods without line search.

Solution.

In the one-dimensional case, we no longer deal with matrices and vectors in the BFGS and DFP methods. Therefore, H_k, p_k, y_k and s_k are scalars, and all the gradients become standard derivatives. In addition, we are considering methods without line search, e.g. with step lengths $\alpha_k = 1$. For the BFGS method, we thereby get

$$x_{k+1} = x_k - H_k^{\text{BFGS}} f'(x_k), \tag{1}$$

where (equation 6.17 in N&W book)

$$H_{k+1}^{\text{BFGS}} = \left(1 - \frac{s_k y_k}{s_k y_k}\right) H_k^{\text{BFGS}} \left(1 - \frac{y_k s_k}{s_k y_k}\right) + \frac{s_k^2}{s_k y_k} = \frac{s_k}{y_k}.$$

With $s_k = x_{k+1} - x_k$ and $y_k = f'(x_{k+1}) - f'(x_k)$, we find that

$$H_k^{\text{BFGS}} = \frac{s_{k-1}}{y_{k-1}} = \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})},$$

and by inserting it into (1), we get

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k).$$

The same reasoning goes for the DFP method; we have

$$x_{k+1} = x_k - H_k^{\text{DFP}} f'(x_k),$$

where (equation 6.15 in N$W book)

$$H_{k+1}^{\text{DFP}} = H_k^{\text{DFP}} - \frac{H_k^{\text{DFP}} y_k y_k H_k^{\text{DFP}}}{y_k H_k^{\text{DFP}} y_k} + \frac{s_k^2}{s_k y_k} = \frac{s_k}{y_k} = \frac{x_{k+1} - x_k}{f'(x_{k+1}) - f'(x_k)},$$

and hence

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k).$$

Problem 2. (See N&W, Problem 6.1.a)

Consider a line search method $x_{k+1} = x_k + \alpha_k p_k$ with search direction

$$p_k = -B_k^{-1} \nabla f(x_k),$$

where $B_k \in \mathbb{R}^{d \times d}$ is a positive definite, symmetric matrix. Denote $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Show that the curvature condition (see N&W, eq. (6.7))

$$\langle s_k, y_k \rangle > 0$$

holds for all step lengths $\alpha_k > 0$, if f is a strictly convex function and $\nabla f(x_k) \neq 0$.

Solution.

The strict convexity of the function f implies

$$f(x_{k+1}) - f(x_k) > \langle \nabla f(x_k), x_{k+1} - x_k \rangle. \tag{2}$$

By interchanging x_{k+1} and x_k in (2), we get

$$f(x_k) - f(x_{k+1}) > \langle \nabla f(x_{k+1}), x_k - x_{k+1} \rangle. \tag{3}$$

By adding equations 2 and 3, we obtain

$$\langle x_{k+1} - x_k, \nabla f(x_{k+1}) - \nabla f(x_k) \rangle > 0,$$

which is

$$\langle s_k, y_k \rangle > 0.$$

Problem 3.

One possibility for lowering the memory requirements of the BFGS-method is to reset the matrix B_k (or its inverse H_k) to the identity matrix after each j-th step for some fixed number j.[1] For $j = 1$ this leads (with the notation of the lecture and Nocedal & Wright, Chapter 6) to the update

$$H_{k+1} = \left(\mathrm{Id} - \frac{s_k y_k^T}{y_k^T s_k}\right)\left(\mathrm{Id} - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}.$$

Assume now that this method is implemented with an exact line search. Show that this yields a non-linear CG-method, where the search directions are defined by

$$p_{k+1} = -\nabla f_{k+1} + \beta_{k+1} p_k$$

with

$$\beta_{k+1} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{(\nabla f_{k+1} - \nabla f_k)^T p_k}$$

(this is the *Hestenes–Stiefel method*, cf. Nocedal & Wright, p. 123).

(Hint: You may need to show in a first step that an exact line search implies that $\nabla f_{k+1}^T p_k = 0 = \nabla f_{k+1}^T s_k$.)

Solution.

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

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

[1] More sophisticated methods are described in Nocedal & Wright, Chapter 7.2.

Let $\phi(\alpha) = f(x_k + \alpha p_k)$. Then we have the optimization problem $\min_{\alpha>0} \phi(\alpha)$ and α_k is a stationary point of it. Therefore, we have

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

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

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

we calculate search directions in the BFGS variant as

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

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

Problem 4.

Implement the BFGS method for the minimisation of the Rosenbrock function.

Note that you will require a Wolfe line search in order to ensure that the matrices stay positive definite and the search directions are actually descent directions.

Solution.

See the implementation on the wiki page.