Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimization Theory Spring 2015

Exercise set 5

The first three of the following exercises are concerned with Quasi-Newton methods. The first exercise asks for an implementation of the method; here it can be interesting to compare the results with those obtained with Newton's method and with a gradient descent method. The second exercise is concerned with the one-dimensional case; here it turns out that Quasi-Newton methods are nothing else than the well-known secant method. The third exercise shows that Quasi-Newton methods and CG methods are closely related. Finally, the fourth exercise deals with non-linear least squares methods and, in particular, the Gauß-Newton method for their solution.

1 Implement the BFGS method for the minimization of the Rosenbrock function.

Note that you will require a Wolfe line search in order to ensure that the matrices stay positive definite. One possible implementation of this line search might roughly look as follows:

1. Test whether the step size 1 satisfies both Wolfe conditions. If yes, use a step size of 1.
2. If $\alpha = 1$ does not work, find values $0 \le \alpha_L < \alpha_R$ such that

$$\begin{aligned} f(x + \alpha_R p) &> f(x) + \alpha_R c_1 \nabla f(x)^T p, \\ f(x + \alpha_L p) &\le f(x) + \alpha_L c_1 \nabla f(x)^T p, \\ \nabla f(x + \alpha_L p)^T p &< f(x) + \alpha_L c_2 \nabla f(x)^T p. \end{aligned}$$

   That is, α_R violates the first Wolfe condition, while α_L violates the second Wolfe condition but not the first one. This can be done as follows:
   - If $\alpha = 1$ violates the first Wolfe condition, choose $\alpha_L = 0$ and $\alpha_R = 1$.
   - Else define $\alpha_L = 1$, choose some $\rho > 1$ and subsequently consider the values ρ^k until either one of them satisfies both Wolfe conditions, in which case we can use it as step size, or violates the first Wolfe condition, in which case we can set $\alpha_R = \rho^k$.
3. If we have not yet found a suitable step size, do the following:
   - Consider $\alpha = (\alpha_L + \alpha_R)/2$.
   - If α satisfies both Wolfe conditions, we are done.
   - If α violates the first Wolfe condition (that is, $f(x+\alpha p) > f(x)+c_1\alpha\nabla f(x)^T p$), replace α_R by α, repeat step 3.
   - If α violates the second Wolfe condition but not the first one, replace α_L by α, repeat step 3.

2 The secant method for the solution of one-dimensional optimization 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.

3 One possibility for lowering the memory requirements of the BFGS-method is to reset the matrix B_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

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

(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 $\nabla f_{k+1}^T p_k = 0 = \nabla f_{k+1}^T s_k$.)

4 Consider the non-linear least squares problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|r(x)\|^2$$

with $r = (r_1, r_2, \ldots, r_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ continuously differentiable. Assume that x^* is the minimizer of this problem and that the Jacobian $J(x^*)$ of the residual function r at x^* has full rank.

a) Show that in a neighbourhood of x^* the Gauß–Newton method for the solution of this problem is well-defined and yields descent directions.

b) Assume in addition that r is twice continuously differentiable and that $r(x^*) = 0$. Show that in this case the Gauß–Newton method without line search converges locally superlinearly to x^*.
(Hint: You might want to use Theorem 3.7 in Nocedal & Wright.)

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