Norwegian University of Science and Technology
Department of Mathematical Sciences

TMA4180 Optimisation I Spring 2018

Exercise set 5

1 This exercise illustrates numerically the relationship between BFGS and CG with exact linesearch applied to convex quadratic problems. As in Exercise 1 from the previous week, let

$$A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 2 \end{bmatrix} \qquad \text{and} \qquad b = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Use BFGS method with $x_0 = 0$ and $H_0 = I$ for solving minimizing the function $f(x) = 0.5x^\top Ax - b^\top x$. Compare the results with those produced by CG (see Theorem 6.4 in N&W).

2 We consider limited memory BFGS (L-BFGS) method, where we store only one pair of vectors s_k, y_k at each iteration, and simply set the initial approximation for the inverse Hessian to be $H_k^0 = I$.

Show that this method with exact linesearch is equivalent with Hestenes–Stiefel non-linear CG method (cf. N&W, p. 123) with exact linesearch.

3 Some analysis of Nedler–Mead method is deferred to the exercises in Nocedal & Wright.

a) Exercise 9.11: show that the average function value at the Nedler–Mead simplex points will decrease over one step if any of the points $\tilde{x}(-1)$, $\tilde{x}(-2)$, $\tilde{x}(-1/2)$, $\tilde{x}(1/2)$ are adopted as a replacement for x_{n+1}.

b) Exercise 9.12: show that if f is a convex function, the shrinkage step in the Nedler–Mead simplex method will not increase the average function value over the simplex. Show that unless $f(x_1) = f(x_2) = \cdots = f(x_{n+1})$, the average value will in fact decrease.

Note that since there is no "sufficient decrease" guarantee in these exercises, the method may stagnate! Indeed, the following example is due to McKinnon (https://doi.org/10.1137/S1052623496303482):

c) Consider

$$f(x,y) = \begin{cases} \theta\phi|x|^\tau + y + y^2, & \text{if } x < 0, \\ \theta x^\tau + y + y^2, & \text{if } x \geq 0, \end{cases}$$

where θ, ϕ, τ are positive constants. For $\tau > 2$ the function f is twice continuously differentiable, strictly convex, and the point $0, 0$ is not a point of its minimum (e.g., the direction $(0, -1)$ is a descent direction at this point). In fact, the minimum of this separable function is atained at $(0, -1/2)$.

Let $\lambda_1 = (1+\sqrt{33})/8$, $\lambda_2 = (1-\sqrt{33})/8$, and consider the simplex with vertices $(0,0)$, $(\lambda_1^n, \lambda_2^n)$, and $(\lambda_1^{n+1}, \lambda_2^{n+1})$ for some integer $n > 0$. It turns out that for carefully select parameters θ, ϕ and τ, Nedler–Mead algorithm for this function performs the "inside contraction" step, resulting in a sequence of simplices with the same structure as the starting simplex, $n \to \infty$. The resulting sequence of simplices "converges" to the non-stationary point $(0,0)$. Therefore, even for strictly convex functions in 2D the method may fail.

Implement Nedler–Mead algorithm, test it on this function to confirm the described behaviour. You can for example use $\tau = 3$, $\theta = 6$, and $\phi = 400$.

$\boxed{4}$ Implement the linesearch Newton-CG algorithm (algorithm 7.1 in the book) and test it on the Rosenbrock function for larger values of n:

$$f(x) = \sum_{i=1}^{n-1} [\alpha(x_{i+1} - x_i^2)^2 + (1 - x_i)^2],$$

where $\alpha > 0$ is a parameter, e.g. $\alpha = 100.0$. The global minimum is attained at $x^* = (1, 1, \ldots, 1)$.