Instructor: Emre Mengi

Math 409/509 (Spring 2011)

Study Guide for Homework 3
This homework concerns the topics listed below.

- Modified Newton's method; in particular
  - modifications based on identity shifts exploiting Gersgorin's Theorem and Frobenius norm
  - modifications based on spectral decomposition (Gill&Wright 3.5.3 and 3.5.4)
- Quasi-Newton methods
  - SR-1 (Nocedal& Wright 8.2)
  - DFP (Nocedal& Wright 8.1)
  - BFGS (Nocedal& Wright 8.1)
- Wolfe conditions (Nocedal& Wright 3.1)
- Convergence of line-search methods (Nocedal& Wright 3.2)

Homework 3 (due on April 13th, Wednesday by 14:00)
Question **7** requires computations in Matlab. Attach the print-outs of the m-files that you implemented, Matlab outputs and plots.

1. Consider the 3×3 matrix

$$A = \begin{bmatrix} -1 & 0.45 & -0.4 \\ 0.45 & -1 & 0.45 \\ -0.4 & 0.45 & -1 \end{bmatrix}$$

(a) Find the Gersgorin intervals containing the eigenvalues of A.

(b) Show that all eigenvalues of A are negative using Gersgorin's theorem. Such a matrix is called *negative definite* and satisfies the property $p^T Ap < 0$ for all nonzero p.

(c) Find a lower bound η for the smallest eigenvalue of A using Gersgorin's theorem. Using the command `eig` in matlab compute the smallest eigenvalue of A and compare it with the lower bound η.

(d) Given a function $f : \mathbb{R}^3 \to \mathbb{R}$ and a point $\bar{x} \in \mathbb{R}^3$ such that $\nabla^2 f(\bar{x}) = A$ and $\nabla f(\bar{x}) = [-1 \;\; -1 \;\; -1]^T$. Find the Newton search direction p_n at $\bar{x}$ for f. Is p_n a descent direction? Find the modified Newton search direction p_m which is computed by replacing $\nabla^2 f(\bar{x})$ with $\nabla^2 f + (-\eta + 0.1)I$ (η is the lower bound from **(c)**) in the Newton iteration. Is p_m a descent direction? You can solve the linear systems to compute p_n and p_m and check whether they are descent directions in Matlab.

2. Consider the matrix

$$A = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}.$$

(a) Find the orthogonal eigenvalue decomposition of $A = V\Lambda V^T$ where Λ is a diagonal matrix and V is an orthogonal matrix, *i.e.* $V^T V = I$.

(b) Let λ_1, λ_2 be the eigenvalues of A. Find a matrix $\tilde{A}$ with the same eigenvectors as A, but with eigenvalues $\tilde{\lambda}_i = |\lambda_i|$, $i = 1, 2$.

(c) (10 points) Consider a function $f : \mathbf{R}^2 \to \mathbf{R}$ and a point $\bar{x} \in \mathbf{R}^2$ such that $\nabla^2 f(\bar{x}) = A$ and $\nabla f(\bar{x}) = [1 \quad -1]^T$. Compute the modified Newton search direction p_m for f at $\bar{x}$ by replacing the Hessian A with $\tilde{A}$. Is p_m a descent direction?

3. This question concerns the application of the quasi-Newton methods to minimize the quadratic function

$$f(x_1, x_2) = x_1^2 + x_2^2$$

starting from $x_0 = (1, -1)$ and with the initial inverse Hessian guess

$$H_0 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

For each method below apply one iteration of the method writing down all the details, in particular the search direction, step-length, new estimate for the minimizer and the updated inverse Hessian approximation. Use exact line search to compute the step-length. Indicate whether the updated Hessians is positive definite or not by referring to one of the standard theorems mentioned in class (and without computing eigenvalues). Recall the notation

$$s_k = x_{k+1} - x_k \quad \text{and} \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

(a) DFP with the update rule

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

(b) BFGS with the update rule

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

4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Given also two estimates $x_k, x_{k+1} \in \mathbb{R}^n$ for a local minimizer of f and the corresponding gradient vectors $\nabla f(x_k), \nabla f(x_{k+1}) \in \mathbb{R}^n$. Consider the DFP update rule given by (1) where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Show that if $H_k \succ 0$ and $s_k^T y_k > 0$, then $H_{k+1} \succ 0$.

5. Given a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. Let $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$, where x_k, x_{k+1} are the iterates of a line-search method so that

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

for some positive step-length $\alpha_k \in \mathbb{R}^+$ and search direction $p_k \in \mathbb{R}^n$.

For DFP or BFGS it is essential to choose a step-length so that $s_k^T y_k > 0$, since this condition guarantees that H_{k+1} is positive definite provided H_k is positive definite.

Suppose that the search direction $p_k \in \mathbb{R}^n$ is a descent direction. Furthermore assume that the minimum of $\phi(\alpha) = f(x_k + \alpha p_k)$ is attained at some positive α_* and such an α_* is chosen as the step-length α_k. Show that the condition $s_k^T y_k > 0$ is met.

6. Implement the BFGS algorithm by modifying the routine `Newton_optimize` from the previous homework. It is crucial to use a line search algorithm together with BFGS that ensures the satisfaction of the Wolfe conditions by the step-lengths. Because otherwise, for instance with an Armijo backtracking line search, the updated Hessian approximations do not have to be positive definite and the search directions are not necessarily descent directions. Matlab routines

- `linesearch_wolfe.m`
- `lszoom.m`
- `cubic_interp.m`
- `inside.m`

performing a line-search so that the Wolfe conditions are satisfied are provided on the course webpage. This is an implementation of Algorithm 3.5 (page 60) in Nocedal and Wright by Michael L. Overton. You do not need to understand the details of the line search. Download the line search routines.

In `Newton_optimize` to calculate the step-length the Matlab routine `linesearch.m` is called. Now you need to call `linesearch_wolfe.m`. To be more precise you need to replace the line

```
alpha = linesearch(fname,x,p,f,g);
```

for steepest descent with

```
alpha = linesearch_wolfe(fname,x,p,f,g'*p);
```

for BFGS. (Note that `linesearch_wolfe.m` requires directional derivative `g'*p` instead of the gradient `g` as the last parameter.) You also need to modify the inverse Hessian approximation at the end of each iteration using the BFGS update rule. (Specifically you should first generate the search direction, then the step-length. After refining the estimate x_k and retrieve x_{k+1} for the local minimizer, you can update the inverse Hessian approximation.) Initially set the inverse Hessian approximation equal to the identity matrix.

Recall also the shortest route question (question **12**) in homework 2. Solve the problem, that is find a shortest route, using your BFGS implementation (*i.e.*, repeat Question **12.c** in homework 2 but with BFGS instead of steepest descent). What is the rate of convergence you observe in practice? Compare the running times of the steepest descent and BFGS. You can use the profile utility in Matlab for this purpose. First type `>> profile on` and run your steepest descent or BFGS code. Then type `>> profile report`

7. Given a multivariate function $f : \mathbb{R}^n \to \mathbb{R}$, a point $\bar{x}$ and a search direction $\bar{p}$. Suppose that the graph of the function $\phi : \mathbb{R} \to \mathbb{R}$

$$\phi(\alpha) = f(\bar{x} + \alpha\bar{p})$$

is as provided below. Shade the intervals of step-lengths α on the horizontal axis that satisfy the Wolfe conditions. The Wolfe conditions are comprised of

(i) <u>the sufficient decrease condition</u>

$$f(\bar{x} + \alpha\bar{p} \leq f(\bar{x}) + \mu_1 \alpha \nabla f(\bar{x})^T \bar{p}$$

or equivalently (defining $\ell(\alpha) = f(\bar{x}) + \mu_1 \alpha \nabla f(\bar{x})^T \bar{p}$)

$$\phi(\alpha) \leq \ell(\alpha)$$

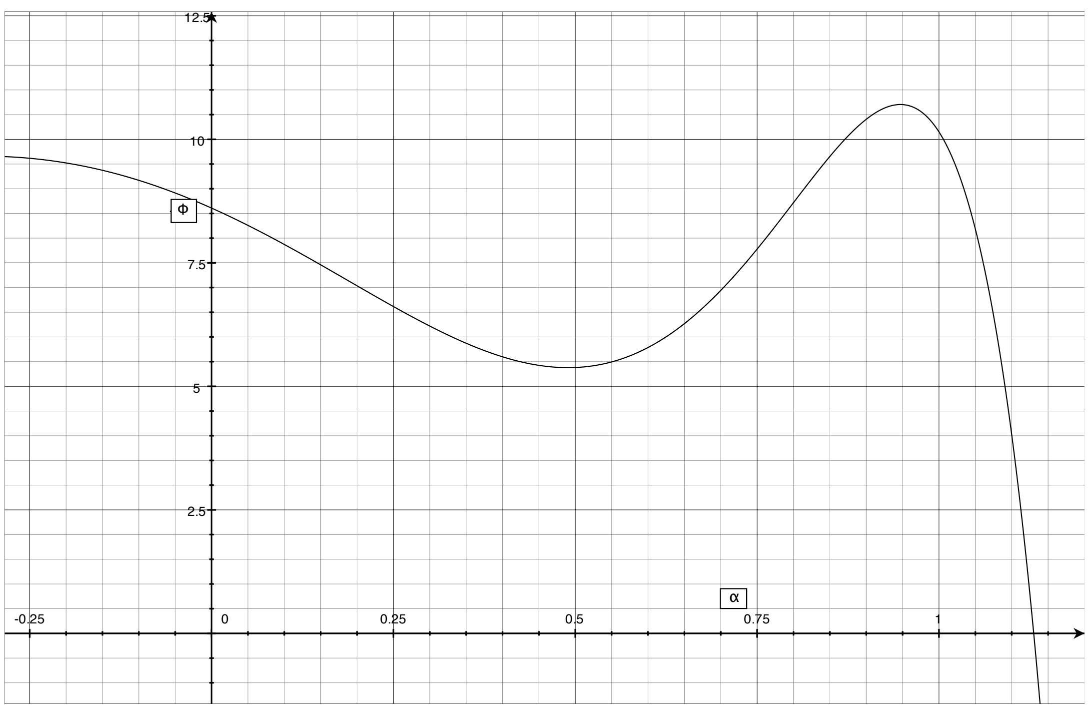
(ii) <u>the sufficient curvature condition</u>

$$\nabla f(\bar{x} + \alpha\bar{p})^T p \geq \mu_2 \nabla f(\bar{x})^T \bar{p}$$

or equivalently

$$\phi'(\alpha) \geq \mu_2 \phi'(0)$$

where the parameters μ_1, μ_2 must satisfy $0 < \mu_1 < \mu_2 < 1$. For this question assume $\mu_1 = 0.1$ and $\mu_2 = 0.9$.



8. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a twice continuously differentiable function (note that this implies f is Lipschitz continously differentiable) that is bounded below.

Which of the following line-search methods (with a line search ensuring Wolfe conditions) are guaranteed to converge to a local minimizer of f globally by Zoutendijk's theorem. Justify your answers.

(a) The quasi-Newton method with the inverse Hessian approximations

$$H_k = \begin{bmatrix} 2 & 0 \\ 0 & 1 + 2^{-k} \end{bmatrix}$$

(b) The quasi-Newton method with the Hessian approximations

$$B_k = \begin{bmatrix} 2 + \frac{k+3}{k+1} & 0 \\ 0 & 1 - \frac{k+1}{k+5} \end{bmatrix}$$

(c) A modified Newton's method based on orthogonal eigenvalue decompositions as follows. Suppose the Hessian matrix has the eigenvalue decomposition $\nabla^2 f(x_k) = V \Lambda V^T$ with the eigenvalues λ_1 and λ_2. The method replaces the Hessian with $\widetilde{\nabla^2 f}(x_k) = V \tilde{\Lambda} V^T$ where

$$\tilde{\Lambda} = \begin{bmatrix} |\lambda_1| + 0.1 & 0 \\ 0 & |\lambda_2| + 0.1 \end{bmatrix}.$$