EXERCISE 7 – Advanced Numerical Analysis, Spring 2015

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Prof. D. Kressner
M. Steinlechner

1 ► Quiz

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and bounded from below.

(a) Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. Then for every norm $0 < \lambda_{min}(A)\|\mathbf{x}\|^2 \leq \mathbf{x}^\mathsf{T} A\mathbf{x} \leq \lambda_{max}(A)\|\mathbf{x}\|^2$. **true** ☐ **false** ☐

(b) A point $\mathbf{x} \in \mathbb{R}^n$ is a critical point of f, if and only if it holds $\nabla f(\mathbf{x})^\mathsf{T}\mathbf{h} \leq 0$ for all $\mathbf{h} \in \mathbb{R}^n$. **true** ☐ **false** ☐

(c) If $\mathbf{p}$ is not a descent direction, then there exists no $\alpha_0 > 0$ such that $f(\mathbf{x} + \alpha\mathbf{p}) < f(\mathbf{x})$ for all $\alpha \in (0, \alpha_0]$. **true** ☐ **false** ☐

(d) For the function $\mathbf{x} \mapsto \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\mathbf{x}$ the steepest descent algorithm with *exact* line search converges in one step. **true** ☐ **false** ☐

(e) Let the sequences $(\mathbf{x}_k)$, $(\mathbf{p}_k)$, and (α_k) be generated by a line search method using descent directions and the Armijo condition. Then it holds: $f(\mathbf{x}_{k+1}) - f(\mathbf{x}_k) \to 0 \quad \Rightarrow \quad \nabla f(\mathbf{x}_k)^\mathsf{T}\mathbf{p}_k/\|\mathbf{p}_k\| \to 0.$ **true** ☐ **false** ☐

(f) For a given starting point $\mathbf{x}_0$, let the level set of $\{\mathbf{x}\colon f(\mathbf{x}) \leq f(\mathbf{x}_0)\}$ be bounded. The sequence $f(\mathbf{x}_k)$ produced by a descent algorithm converges. **true** ☐ **false** ☐

For any two clusterpoints $\mathbf{x}^*$ and $\mathbf{x}^{**}$ of the sequence $(\mathbf{x}_k)$ it holds $f(\mathbf{x}^*) = f(\mathbf{x}^{**})$. **true** ☐ **false** ☐

(g) Let the level set again be bounded, and assume $\mathbf{x}_0$ is not a critical point. Then the steepest descent algorithm (with Armijo line search) produces at least one subsequence which converges to a local minimum. **true** ☐ **false** ☐

(h) A bounded sequence $(\mathbf{x}_k)$ generated by a line search method, satisfying the Wolfe conditions and the angle condition $\cos\theta_k \geq \delta > 0$ (see page 51), has a convergent subsequence $(\mathbf{x}_{k_l})$ such that $\|\nabla f(\mathbf{x}_{k_{l+1}})\| \leq \frac{1}{2}\|\nabla f(\mathbf{x}_{k_l})\|$. **true** ☐ **false** ☐

For the whole sequence it at holds $\|\nabla f(\mathbf{x}_{k+1})\| \leq \|\nabla f(\mathbf{x}_k)\|$. **true** ☐ **false** ☐

2 ► Curvature condition for Quasi-Newton methods

In a Quasi-Newton method, after the $(k+1)$-th iteration, a symmetric positive definite matrix B_{k+1} is sought to satisfy $B_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k) = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$. A necessary condition for such a matrix to exist is

$$(\mathbf{x}_{k+1} - \mathbf{x}_k)^T(\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)) > 0. \tag{$\star$}$$

(a) Proof that for a strictly convex function ($\star$) always holds.

(b) Show that the *strong Wolfe curvature condition*

$$|\nabla f(\mathbf{x}_k + \alpha_k\mathbf{p}_k)^T\mathbf{p}_k| \leq -c_2\nabla f(\mathbf{x}_k)^T\mathbf{p}_k$$

with $c_2 \in (0, 1)$ implies ($\star$) for $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{p}_k$.

3 ▶ Steepest descent vs. Newton vs. BFGS

(a) The *Rosenbrock function*

$$f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

serves as an example of functions which are difficult to optimize using steepest descent.

Calculate the gradient $\nabla f(\mathbf{x})$ and Hessian $\nabla^2 f(\mathbf{x})$. Prove that $\mathbf{x}^* = (1, 1)$ is the only local minimizer and $\nabla^2 f(\mathbf{x}^*)$ is positive definite.

(b) A Matlab function on the homepage visualizes the minimization of f using **steepest descent**, **Newton's method** and **BFGS**, all with Armijo backtracking. However, you have to provide these algorithms, as well as the function handles for f, ∇f, and H, to run it. The syntax has to be:

$$\begin{aligned}
&\texttt{X} = \texttt{steepdesc}(\texttt{f}, \texttt{df}, \texttt{x0}, \texttt{c1}, \texttt{alpha0}, \texttt{beta}, \texttt{tol}) \\
&\texttt{X} = \texttt{newton}(\texttt{f}, \texttt{df}, \texttt{ddf}, \texttt{x0}, \texttt{c1}, \texttt{alpha0}, \texttt{beta}, \texttt{tol}) \\
&\texttt{X} = \texttt{BFGS}(\texttt{f}, \texttt{df}, \texttt{x0}, \texttt{B0}, \texttt{c1}, \texttt{alpha0}, \texttt{beta}, \texttt{tol})
\end{aligned}$$

Here `f`, `df`, `ddf` are the function handles for f, ∇f, and $\nabla^2 f$; `x0` is the starting value, and `c1`, `alpha0`, `beta` are the parameters of the Armijo backtracking. The iteration should stop when $\|\nabla f(\mathbf{x}_n)\| \le \texttt{tol}$. The output matrix `X` contains all generated iterates as columns.

For the BFGS algorithm, use the identity $B_0 = I$ as the initial guess for the approximate Hessian at $\mathbf{x}_0$. To improve upon this bad starting guess, do the following:

After having calulated $\mathbf{x}_1$ using B_0, but before updating B_0 to B_1 by the BFGS formula, B_0 is replaced by

$$B_0 \leftarrow B_0 \frac{\mathbf{y}_0^T \mathbf{s}_0}{\mathbf{y}_0^T \mathbf{y}_0},$$

where $\mathbf{s}_0 = \mathbf{x}_1 - \mathbf{x}_0$ and $\mathbf{y}_0 = \nabla f(\mathbf{x}_1) - \nabla f(\mathbf{x}_0)$. Then continue normally for all following B_i.

(c) Use first $\mathbf{x}_0 = (1.2, 1.2)^T$, and then the more difficult starting point $\mathbf{x}_0 = (-1.2, 1)^T$. For both choices plot the error of both methods to the exact value against the iteration number on a semilogarithmic scale.