AMATH/MATH 516 Problem Set 4

FIRST PROGRAMMING ASSIGNMENT

(1) Apply the steepest descent algorithm with backtracking to minimize the Rosenbrock function (see website for Matlab files for the Rosenbrock function). In this algorithm choose the steepest descent search direction in the following three ways:

(a) $d^k := -\nabla f(x^k)/\left\|\nabla f(x^k)\right\|$.

(b) $d^k := -\nabla f(x^k)/\theta_k$, where $\theta_k := (s^k)^T y^k/(s^k)^T s^k$ whenever $(s^k)^T y^k > 0$ and $\theta_k := \left\|\nabla f(x^k)\right\|$ otherwise, with $s^k := x^k - x^{k-1}$ and $y^k := \nabla f(x^k) - \nabla f(x^{k-1})$ (take $\theta_0 := \left\|\nabla f(x^k)\right\|$).

(c) $d^k := -\nabla f(x^k)/\theta_k$, where $\theta_k := (y^k)^T y^k/(y^k)^T s^k$ whenever $(s^k)^T y^k > 0$ and $\theta_k := \left\|\nabla f(x^k)\right\|$ otherwise, with $s^k := x^k - x^{k-1}$ and $y^k := \nabla f(x^k) - \nabla f(x^{k-1})$ (take $\theta_0 := \left\|\nabla f(x^k)\right\|$).

(2) Apply the steepest descent algorithm with weak Wolfe line search to minimize the Rosenbrock function (see website for Matlab files for the Rosenbrock function). In this algorithm choose the steepest descent search direction in the following three ways:

(a) $d^k := -\nabla f(x^k)/\left\|\nabla f(x^k)\right\|$.

(b) $d^k := -\nabla f(x^k)/\theta_k$, where $\theta_k := (s^k)^T y^k/(s^k)^T s^k$ whenever $(s^k)^T y^k > 0$ and $\theta_k := \left\|\nabla f(x^k)\right\|$ otherwise, with $s^k := x^k - x^{k-1}$ and $y^k := \nabla f(x^k) - \nabla f(x^{k-1})$ (take $\theta_0 := \left\|\nabla f(x^k)\right\|$).

(c) $d^k := -\nabla f(x^k)/\theta_k$, where $\theta_k := (y^k)^T y^k/(y^k)^T s^k$ whenever $(s^k)^T y^k > 0$ and $\theta_k := \left\|\nabla f(x^k)\right\|$ otherwise, with $s^k := x^k - x^{k-1}$ and $y^k := \nabla f(x^k) - \nabla f(x^{k-1})$ (take $\theta_0 := \left\|\nabla f(x^k)\right\|$).

(3) Consider the nonlinear least squares problem

$$\min_{\mathbb{R}^7} f(x) := \frac{1}{2}\left\|g(x)\right\|_2^2$$

where matlab function files for the function $g : \mathbb{R}^7 \to \mathbb{R}^8$, its Jacobian, and its Hessians are available on the course website.

(a) Solve this problem using steepest descent with backtracking (or weak Wolfe) line search Choose your own iteration parameters and stopping criteria. However, if you reduce the norm of the gradient to less than 10^{-12} then you have done well and your procedure should terminate.

(b) Solve this problem with the Levenberg-Marquardt method for choosing the search direction and a backtracking line search to choosing the step-length. In the Levenberg-Marquardt method, the search directions d^k are chosen to solve the subproblems

$$\min_{d^k\in\mathbb{R}^n} \frac{1}{2}\left\|g(x^k) + g'(x^k)d\right\|_2^2 + \frac{\alpha_k}{2}\left\|d\right\|_2^2 .$$

Show that these search directions are directions of strict descent for f at x_k as long as $\nabla f(x_k) \neq 0$. Solve these subproblems with the conjugate gradient algorithm. The α_k's used to adjust for the magnitude of the d^k's. For example, one could take

$$\alpha_0 := 10 + \frac{1}{2}\left\|g(x^0)\right\|_2^2$$

and

$$\alpha_{k+1} = \begin{cases} \alpha_k + 10\ , & \text{if } \lambda_k < 10^{-3}, \\ \alpha_k\ , & \text{if } 10^{-3} \leq \lambda_k \leq 1, \\ \alpha_k/2\ , & \text{if } \lambda_k > 1. \end{cases} ,$$

where λ_k is the step length. Is the role of the α_k's significant and, if so, why? Is this a reasonable rule for updating the α_k's and if so why ? Can you suggest a better alternative? If you can, then use that method in your program and justify its use.

Initialize the iterations at $x^0 = 0$. Choose your own line search parameters and termination criteria. Again, if you are able to attain $\|\nabla f(x)\| \leq 10^{-12}$, then you have done very well.

(4) Newton's method for minimizing a function $f : \mathbb{R}^n \to \mathbb{R}$ is just Newton's method applied to the equation $\nabla f(x) = 0$. Thus, the Newton direction ignores the underlying minimization problem. For this reason the Newton direction, if it exists, may not be a direction of descent for the function, or even if it is a direction of descent, it may require an excessive number of steps in a line search procedure to induce descent. Even so, the full Newton step may still be a very good choice of direction even if it may occasionally increase the value of the objective function. For this reason a number of authors have propose *non–monotone* line search procedures in order to increase the likelihood of accepting a Newton (or Newton like) step even if it may occasionally increase the value of the objective. In this problem, we consider one such proposal.

Let p be some positive integer (usually $p = 3$ or $p = 4$). Suppose that $x^0 \in \mathbb{R}^n$ is given and define $x^{-\ell} = x^0$ for $\ell = 1, 2, \ldots, p-1$. Consider the following backtracking procedure for choosing a step length.

Non–Monotone Backtracking: $0 < c < 1$ and $0 < \gamma < 1$
Let $d^k \in \mathbb{R}^n$ be such that $\nabla f(x^k)^T d^k < 0$. Set

$$\begin{array}{rll} \lambda_k = & \max & \gamma^s \\ & \text{subject to} & s = 0, 1, 2, \ldots \\ & & f(x^k + \gamma^s d^k) \le \left[\max\limits_{\ell = 0,1,\ldots,p-1} f(x^{k-\ell}) \right] + c\gamma^s \nabla f(x^k)^T d^k \ . \end{array}$$

Observe that this line search procedure does not guarantee that $x^{k+1} = x^k + \lambda_k d^k$ satisfies $f(x^{k+1}) < f(x^k)$.

(a) Show that the sequence $\{f_k^{\max}\}$ given by

$$f_k^{\max} = \max_{\ell = 0,1,\ldots,p-1} f(x^{k-\ell})$$

for $k = 0, 1, \ldots$ is non-increasing.

(b) Show by induction that the sequence $\{f_k^{\max}\}$ is p–**step strictly** decreasing sequence, i.e. $f_{k+p}^{\max} < f_k^{\max}$ for $k = 0, 1, \ldots$.

(c) For each $k = 0, 1, \ldots$, define $m(k)$ to be the smallest integer in the set $\{k, k-1, \ldots, k-p+1\}$ such that

$$f(x^{m(k)}) = \max_{\ell = 0,1,\ldots,p-1} f(x^{k-\ell}),$$

so that $f_k^{\max} = f(x^{m(k)})$. By paralleling the proof of Theorem 2.1.1, prove the following result.

Theorem Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable with ∇f Lipschitz continuous on the closed convex hull of the set $\{x : f(x) \le f(x^0)\} + \delta\mathbb{B}$ where $\delta > 0$ and $x^0 \in \mathbb{R}^n$ are given. Let $\{x^k\}$ be a sequence of iterates given by the following algorithm:

Non-Monotone Descent Algorithm

Step 1: If $D_k \subset \{d : f'(x^k; d) < 0\}$ is empty, STOP; otherwise choose $d^k \in D_k$.
Step 2: Choose the step length λ_k by the non-monotone backtracking procedure specified above.
Step 3: Set $x^{k+1} = x^k + \lambda_k d^k$, $k = k + 1$, and return to Step 1.

If $\|d^k\| \le \delta$ for all $k = 0, 1, \ldots$, then one of the following must occur:

(i) There is a k_0 such that $D_{k_0} = \emptyset$.
(ii) $f_k^{\max} \downarrow -\infty$.
(iii) $f'(x^{m(k)-1}; d^{m(k)-1}) \to 0$.

(d) Use the Newton search direction with $d^\nu = -\nabla^2 f(x^\nu)^{-1} \nabla f(x^\nu)$ ($H_\nu := \nabla^2 f(x^\nu)$) in conjunction with this non-monotone line search procedure to minimize the Rosenbrock function. Safeguard the Newton search direction by two checks:

(i) If $\|H_k^{-1} \nabla f(x_k)\| \ge 100 \|\nabla f(x_k)\|$, reset $d_k = -\nabla f(x_k)$.
(ii) If $-\nabla f(x_k)^T d_k \le 10^{-4} \|\nabla f(x_k)\| \|d_k\|$, reset $d_k = -\nabla f(x_k)$.

Use the starting point and stopping criteria from the previous problem set. Try the values $p = 3$ and $p = 4$ as suggested above. Compare the outcome of this numerical experiment with those of the previous problem set.

(5) In this problem set we revisit the nonlinear least squares problem of the previous problem set. That is, we are interested solving the nonlinear least squares problem

$$\mathcal{P} \qquad \min_{x \in B} f(x) := \frac{1}{2} \|g(x)\|_2^2$$

where matlab function files for the function $g : \mathbb{R}^7 \to \mathbb{R}^8$, its Jacobian, and its Hessians are available through the course webpage.

Each of the algorithms described below is to be initiated at the point $x^0 = zeros(7, 1)$ with a stopping criteria of $\|\nabla f(x)\| \le 10^{-12}$. As part of your output, you need to include

(i) a graph of the primary stopping criteria, e.g. the norm of the gradient in the unconstrained case
(ii) a graph of the function values
(iii) a graph of the magnitude of the steps taken at each iteration
(iv) a table listing the total number of function, gradient, and Hessian evaluations.

Use a `maxit` in excess of 5000.

(a) Solve the problem using the safe-guarded Newton step of the previous problem and the non-monotone backtracking line search.
(b) Solve the problem using the safe-guarded Newton step of the previous problem and the weak Wolfe line search.
(c) Solve the problem using (inverse) BFGS updating in conjunction with the weak Wolfe line search.