MATH 689: Numerical Optimization

Prof. Wolfgang Bangerth // Blocker 507D // bangerth@math.tamu.edu

Homework assignment 3 – due Tuesday 02/11/2014

Problem 1 (Search directions). The steepest descent method chooses a search direction $p_k = -\nabla f(x_k)$. By Taylor's theorem it is easy to show that this direction is a direction of descent, i.e.: there exists a number $T > 0$ so that if we go in direction p_k no more than T times $\|p_k\|$, our objective function will become smaller (we hope of course that eventually we will land in a minimum). In formulas: for this choice of p_k, there exists $T > 0$ so that

$$f(x_k + \alpha p_k) < f(x_k) \qquad \forall 0 < \alpha < T.$$

Prove that a similar statement (possibly with a different value for T) is also true if we choose as search direction $p_k = -B^{-1}\nabla f(x_k)$ for any given matrix B that is positive definite.

(The importance of this statement is that (i) if we are close to a strict minimum, we know that the Hessian is a positive definite matrix and the statement above with $B = H_k$ guarantees that the Newton direction is a descent direction; (ii) it guarantees that we can *approximate* the Hessian matrix H_k by any matrix B and still get a descent direction as long as B is positive definite.) **(3 points)**

Problem 2 (Radius of convergence for Newton's method). You have seen in class that Newton's method (with step length $\alpha = 1$) can only be proven to converge if we have a starting point x_0 that is close enough to the solution x^*. One reason may be that far enough away from the minimum x^* the function $f(x)$ may not be convex any more. Explain in words for a function $f(x)$ of a single variable $x \in \mathbb{R}^1$ what would go wrong if we started in an area where $f(x)$ is not convex (i.e. where $f''(x) < 0$). In particular, think about how Newton's method choses its search direction. If you want a concrete example to explain things with, use the function

$$f(x) = -\frac{1}{1 + x^2}$$

whose minimum is at $x^* = 0$ but whose Hessian is positive definite only in an interval around the origin. You can illustrate your explanation with numerical results that show what Newton's method does if, for example you start at $x = 0.1, 0.5, 1, 2, 5, \ldots$. **(4 points)**

Problem 3 (Radius of convergence for Newton's method). The case discussed in the previous problem is not the only one where Newton's method may not converge. Consider

$$f(x) = x \ \arctan x - \frac{1}{2}\ln(1 + x^2).$$

This function is convex everywhere since $f''(x) = \frac{1}{1+x^2} > 0$. Yet, Newton's method (with step length $\alpha = 1$) only converges if started within an interval $(-r, r)$ around the minimum $x^* = 0$. Determine numerically the radius of convergence r for this problem. What happens if you start with $x_0 = r$? What if $x_0 > r$?
(5 points)

Problem 4 (Rate of convergence for Newton's method). For the same function as in the previous problem, check that whenever Newton's method converges with step length $\alpha = 1$, that it converges with at least quadratic order. Determine the actual convergence order. To do so start at a point x_0 within $(-r, r)$, calculate x_k for a number of iterations k, and compute the convergence order of the sequence $\|x_k - x^*\|$.

(3 points)

We know that whenever Newton's method with full step length converges, it does so with at least quadratic order. What you see here is different, however. Can you explain? **(2 bonus points)**

A problem for your consideration (Newton's method with line search). For the same function as in the previous problems, implement Newton's method with a line search procedure to determine the step length α_k in each step, i.e. we no longer assume that we can work with $\alpha = 1$. Using your implementation, show the following: (i) the algorithm now converges for all starting points x_0, even those that lie far outside the interval $[-r, r]$; and (ii) when close to the solution, the line search criterion allows a step length of $\alpha_k = 1$ to be chosen (in other words, $\alpha_k = 1$ satisfies the Wolfe conditions), guaranteeing quadratic convergence once we get close to the solution.

Show your results by plotting the sequences x_k and α_k for a starting point $x_0 = 10^6$ (which is far outside the convergence radius for Newton's method with $\alpha = 1$!). **(No points)**

If you have comments on the way I teach – in particular suggestions how I can do things better, if I should do more or less examples, powerpoint slides vs whiteboard, etc – or on other things you would like to critique, feel free to hand those in with your homework as well. I want to make this as good a class as possible, and all comments are certainly much appreciated!