Mihai Anitescu, STAT 310, Assigned: 02/21/2012, Due: 02/28/2012

Stat 310, Part II, Optimization. Homework 2.

Problem 1: (computation; line search method, also version of Problem 3.1 in textbook)

Write a program that implements the backtracking (Armijo) line search with modified LDL' factorization ,(first, write a pseudocode of the entire method, this will help in your implementation. Report the pseudocode in your homework). Choose *Bk* to be the exact Hessian. Apply it to solve Fenton's function (from preceding homework) starting at both [3 2] and [3 4]. Experiment with the parameters or by designing your own rules. Report the total number of linear systems solved and the total number of function evaluations.

Problem 2: (computation; "the power of sparse linear algebra").

Apply the algorithm above to solve the problem "cute" which is posted on my website. Be sure the implementations supports sparse linear algebra. Solve versions of the problem of increasing size, initialized at a vector of all ones, to a point which seems reasonable (on my laptop, evaluations of Hessian with INTVAL is very fast up to n=10000, on UNIX, n=1000 should be fast enough). Plot the number of function evaluations and number of linear systems solved as a function of the problem size.

Calling sequences for my function are (for example)
[f,g,H]=cute_wrap(ones(10000,1),2);
[f,g,H]=cute_wrap(x,2);

Problem 3: (divided differences, derivative calculations).

Consider the following divided difference approach (called "central differences") by which to approximate the derivative of the function $f(x): R \to R$, $f \in C^4$, (the function f is 4 times continuously differentiable):

$$\left.\frac{df}{dx}\right|_{x_0} \approx \frac{f(x_0+h)-f(x_0-h)}{2h}.$$

I. Use Taylor series expansions to justify that

$$\left| \left.\frac{df}{dx}\right|_{x_0} - \frac{f(x_0+h)-f(x_0-h)}{2h} \right| = O\left(h^2\right)$$

II. Using the same argument as in class, what is the optimal perturbation amount h with respect to the machine precision $\varepsilon \approx 1e-16$? (that is, the h that produces the minimum value of the error). Is this minimum error smaller than the one for forward differences (the case covered in class?)

III. Would you recommend central differences for approximating $\nabla f, \quad f : R^n \to R$ (by doing the perturbation above in each component) over the forward differences approach that we have discussed in class? Please explain your conclusion.