

Stat 310, Part II, Optimization. Homework 1.

Problem 1: (computation; Newton's method)

- I. Implement Newton's method ((3.38) in Nocedal and Wright, or from lecture notes).
- II. Apply the method to the following function (Fenton's function ; which you can download from my website, from lecture list area, including with links to the automatic differentiation package discussed in class -- if you wish to go this route. If not you may have to compute the gradient and Hessian by hand. It is also an easy function to code in AMPL if you want to test your work – but this is not required).

$$f(x) = \left\{ 12 + x_1^2 + \frac{1 + x_2^2}{x_1^2} + \frac{x_1^2 x_2^2 + 100}{(x_1 x_2)^4} \right\} \left(\frac{1}{10} \right)$$

- III. Initialize the method at $x^1 = [3, 2]$. Describe what you observe.
- IV. Modify the method in the following two ways; use $x^{k+1} - x^k = -B_k^{-1} \nabla f(x^k)$, where the matrix B_k is, instead of the Newton choice (1) $B_k = \nabla^2 f(x^k)$; one of:

$$(2) \quad B_k = \left(1 + \frac{1}{k^2} \right) I + \nabla^2 f(x^k);$$

$$(3) \quad B_k = \frac{1}{k^2} I + \nabla^2 f(x^k)$$

Start the method from (3,2), again.

- V. Propose a way to estimate the rate of convergence of the 3 methods (Newton + (2),(3)) and carry it out for the experiment at point 1.IV. I suggest finding a way to graphically represent your answers, perhaps use a log-log plot.

Problem 2: (divided differences, derivative calculations). Consider the following divided difference approach (called "central differences") by which to approximate the derivative of the function $f(x) : R \rightarrow R$, $f \in C^4$, (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

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

- 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 ∇f , $f : R^n \rightarrow R$ (by doing the perturbation above in each component) over the forward differences approach that we have discussed in class? Please explain your conclusion.

Problem 3: (global minimum; convexity.). We say that a function $f : R^n \rightarrow R$ is convex if $x, y \in R^n, \alpha \in [0, 1] \Rightarrow f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$. We say that a subset $S \subset R^n$ is convex if $\forall x, y \in S, \alpha \in [0, 1] \Rightarrow \alpha x + (1 - \alpha)y \in S$.

- I. Prove that the set of global minimizers of a *convex* function f (which we assume nonempty) is convex.
- II. Using the consequences of sufficient second-order conditions for optimality discussed in class, prove the following: If a global minimum x^* of a *convex* function f satisfies the strong second-order condition $\nabla_{xx}^2 f(x^*) \succ 0$ (the Hessian is positive definite at x^* ; we assume here that f is twice continuously differentiable) then x^* is the UNIQUE global minimizer.