OPTIMIZATION & NONLIN EQS (MA 784) — HW 2

Unconstrained minimization; first order methods.

INCOMPLETE HW — I will add approx. 1 more problem, but these are posted so that you can get started. **Due on February 19 (Thursday), by the start of the class.** Please turn in your solutions to me, on paper. (*Hints are available if needed, but give it a serious try by yourself first.*)

1. (Justifying the stopping criterion.) It is unreasonable to expect that an iterative numerical optimization method will exactly hit a stationary point, where $\nabla f(\mathbf{x}^*) = 0$ holds precisely. Instead, we may stop when $\|\nabla f(\mathbf{x})\| < \varepsilon$ for some sufficiently small $\varepsilon > 0$ and hope that we are close to a stationary point. In this exercise, we explore whether we can back this intuition up with math.

Throughout, we assume that the function $f : \mathbb{R}^d \to \mathbb{R}$ we wish to minimize is differentiable everywhere on $\mathbb{R}^d$, and that we have a sequence of points $\mathbf{x}_1, \mathbf{x}_2, \ldots$ from $\mathbb{R}^d$ for which

$$\lim_{k\to\infty} \|\nabla f(\mathbf{x}_k)\| = 0,$$

but nothing else (until stated otherwise).

(a) Show by a *simple, univariate* example that without additional assumptions, we cannot even say that the sequence $(\mathbf{x}_k)$ is convergent or that a stationary point even exists! (One example can take care of both of these.)

(b) Suggest a simple heuristic to detect if this is the case (so we can stop the algorithm with a warning to user). Explain in one sentence what happens in your example from the previous part if your new stopping criterion is applied.

(c) From now on, let us suppose (in addition to our initial assumptions) that the sequence $(\mathbf{x}_k)$ converges to a point $\mathbf{x}^*$ and also that f has a stationary point. It is tempting to expect that the limit point $\mathbf{x}^*$ is a stationary point, that is, $\nabla f(\mathbf{x}^*) = 0$, but that is also false! Your task is the following: find a *univariate* function ϕ and a sequence of real numbers $(x_k)_{k=1,2,\ldots}$ that satisfy each of the following conditions:

- $\phi(x_k)$ strictly monotonically decreases to 0,
- $\phi'(x_k)$ strictly monotonically decreases to 0,
- $\phi'(0) = 1$.

(The first two conditions simulate what we might see while running a reasonable descent method for minimization, which will wrongly conclude that we have approached a stationary point.)

(d) Now let us also assume that f is convex, in addition to all the other assumptions we have made. Show that in this case, we can finally conclude that the limit point $\mathbf{x}^*$ is a global minimizer. (You may use all the properties of convex functions we mentioned in class, whether we have proven them or not.)

2. (Misspecified parameters.) We have seen that an underestimated Lipschitz constant for the gradient could lead to non-convergence but is easily detected and mitigated by increasing the estimated L. This exercise shows what happens if we instead overestimate the strong convexity constant.

Let $f(x) = x^2/10$. This is an L-smooth and ℓ-strongly convex function with suitable parameters (ℓ, L).

(a) What are the best possible values for ℓ and L? Use these parameters to determine a constant stepsize $\tau^k = \tau$ that guarantees convergence. Determine, for $M = \{3, \ldots, 8\}$, how many iterations until we reach $f(x^k) < 10^{-M}$ from $x^0 = 1$? Does the rate of convergence appear linear? (You may either make a quick-and-dirty implementation of the gradient method for this simple univariate example or just derive analytically a recursion for $f(x^k)$ and implement that to compute the errors. Either way, include your code and the output in your answer!)

(b) The diminishing step size sequence $\tau_k = 1/(k+1)$ would also guarantee convergence, but it's only a good step size (guaranteeing linear convergence) for the right values of (L, ℓ). It actually does not work here.

Use your previous code to determine the number of iterations until we reach $f(x^k) < 10^{-1}$ from $x^0 = 1$ with this step size sequence. (*Hint:* If your code is correct, you do not want to try and run this nearer to completion–see also the next problem.)

(c) *Extra credit:* use analytic techniques to find a reasonably sharp bound on $f(x^k)$ for general k (as a closed-form formula) when using the step size in the previous part, and show $f(x^k) > 0.001$ even for $k = 10^9$.)