WEEKLY PARTICIPATION 6: ITERATION COMPLEXITIES

Consider the problem

$$\operatorname{argmin}_{x \in \mathbb{R}} f(x),$$

where f is a convex function. We want to solve this optimization problem to ε-suboptimality in the function value, with $\varepsilon = 10^{-3}$. We have several options for optimization algorithms, let us see how many iterations they will each require.

(1) If f is differentiable and β-smooth, how many iterations T of gradient descent suffice[1]?

(2) If f is differentiable and α-strongly convex, how many iterations of gradient descent suffice?

(3) If f is differentiable and has convex condition number κ, how many iterations of gradient descent suffice?

(4) If f is non-differentiable, how many iterations of subgradient descent suffice?

(5) If f is twice-differentiable and we start sufficiently close to the optimizer $x^\star$ that the convergence theory for pure Newton phases apply, how many iterations of Newton's method suffice[2]? Assume the Hessian is L-Lipschitz and $\nabla^2 f(x^\star) \succeq \mu I$.

(6) If f is differentiable and μ-strongly convex and stochastic gradient descent with a constant stepsize α is used, give the tightest constant lower bound you can on how small $\mathbb{E}[f(x_T) - f(x^\star)]$ can become, regardless of how many iterations T are used. Use this to determine an upper bound on the step sizes α that can be used to achieve ε-suboptimality.

---

[1] For this and the other problems, give an expression for T that depends only on the given value of ε and the stated properties of the function. This means, e.g. ignoring the dependence on x_0.

[2] Convince yourself that in the setting where the convergence theory for pure Newton phases applies and f is convex, the minimizer is unique.