Note: The purpose of this problem set is to motivate convex optimization, as well as introducing you to common tools used in analyzing the performance or fundamental limits of optimization algorithms. As you will see, optimizing a generic function is hard in a meaningful sense, yet optimizing so-called 'convex' functions is quite easy. Since convex functions are surprisingly ubiquitous in the area of machine learning and statistics, it's important to know these potentially surprising properties.

---

1 Suppose we wish to minimize a continuous function $f : [0, 1] \to \mathbb{R}$ via an algorithm which sequentially queries $f(x_i)$ at points $x_1, x_2, \ldots, x_n \in [0, 1]$ before reporting some $\overline{x} \in [0, 1]$ that is intended to approximate the minimizer $x^\star \in [0, 1]$ with $f(x^\star) = \min_{x \in [0,1]} f(x)$. In the *oracle* model of computation, we measure the number of queries n (the *sample complexity*) to our function needed to ensure that $f(\overline{x}) < f(x^\star) + \epsilon$ for some $\epsilon > 0$ as a way to indicate the complexity of our algorithm. Such an $\overline{x}$ is known as an ϵ-approximate minimizer, and since in this setting we are only given information about the function value (and not, say, the derivative at each x_i as well) this is known as a *zero-th order oracle*. We will prove that any algorithm which returns an ϵ-approximate minimizer of a continuous function $f : [0, 1] \to \mathbb{R}$ (in particular, a polynomial) has infinite sample complexity via a technique known as a *resisting oracle*. Flesh out the following sketch (and in particular, construct q):

Suppose by way of contradiction that we have an algorithm which can find an ϵ-approximate minimizer to any continuous function using at most $n < \infty$ queries to the oracle. In an attempt to 'trick' the algorithm into thinking that an input f is the zero function, suppose our oracle returns $f(x_i) = 0$ for all $x_1, x_2, \ldots, x_n$ and consider the resulting $\overline{x}$. Since the algorithm is deterministic, it must return a fixed $\overline{x}$, yet we can construct a polynomial $q : [0, 1] \to \mathbb{R}$ which satisfies $q(x_1) = \cdots = q(x_n) = q(\overline{x}) = 0$ yet $\min_{x \in [0,1]} q(x) = -\epsilon$. This contradicts the fact that our algorithm supposedly guaranteed that $0 = q(\overline{x}) < q(x^\star) + \epsilon = 0$, and hence the sample complexity of such an algorithm is *at least* n. But since n was an arbitrary natural number, this means that sample complexity is infinite, and one cannot hope to minimize an arbitrary continuous (even analytic) function without querying a dense subset of the domain and effectively determining the function uniquely.

Note that q only needs to *achieve* $q(x) = -\epsilon$ to create a contradiction. Without loss of generality assume that all $x_1, x_2, \ldots, x_n, \overline{x}, x$ are distinct, and use Lagrange interpolation to construct q. ∎

---

2 Prove that any algorithm which returns an ϵ-approximate minimizer of a function $f : [0, 1]^n \to \mathbb{R}$ with $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \le L\|\boldsymbol{x} - \boldsymbol{y}\|_\infty$ requires at least $\left(\left\lfloor \frac{L}{2\epsilon} \right\rfloor\right)^n$ queries in the zero-th order oracle model if $\epsilon < L/2$.

Write $p = \lfloor \frac{L}{2\epsilon} \rfloor$. In light of the solution to Problem 1, we only need to find an L-Lipschitz function $f : [0,1]^n \to \mathbb{R}$ that has $f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) = \ldots = f(\boldsymbol{x}_{p^n-1}) = f(\overline{\boldsymbol{x}}) = 0$ for arbitrary $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p^n-1}, \overline{\boldsymbol{x}}$ yet $\min_{\boldsymbol{x} \in [0,1]^n} f(\boldsymbol{x}) \leq -\epsilon$. Note that (argue this) by the Pidgeonhole Principle there exists an ℓ^∞ ball of radius $\frac{1}{2p}$ around a point $\boldsymbol{x}^\star$ which doesn't intersect with any of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{p^n-1}, \overline{\boldsymbol{x}}$. Try $f(\boldsymbol{x}) = \min\{0, L\|\boldsymbol{x} - \boldsymbol{x}^\star\|_\infty - \epsilon\}$. ∎

---

3 Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, and that the eigenvalues $\lambda_i(x)$ of the Hessian $\nabla^2 f(x)$ are uniformly bounded between $0 < \ell < \lambda_i(x) < L$ across all of $\mathbb{R}^n$. Such a function is known as ℓ-*strongly convex* Consider the following algorithm, known as *gradient descent*, which finds an approximation x_T to a global minimizer x^* of f.

> **Data:** An arbitrary starting point $x_0 \in \mathbb{R}^n$, a step size $0 < \alpha < 2/L$, and a maximum number of iterations $T \in \mathbb{N}$.
> **Result:** An approximate global minimizer $x_T \in \mathbb{R}^n$ to f.
> **for** $t = 1, 2, \ldots, T$ **do**
> $\quad$ | $\quad x_t = x_{t-1} - \alpha \nabla f(x_{t-1})$
> **end**

Prove that if $\frac{1-q}{\ell} \leq \alpha \leq \frac{1+q}{L}$ for some $0 < q < 1$ then the global minimizer x^* is unique and

$$\|x_T - x^*\|_2 \leq \frac{\alpha q^T}{1-q}\|\nabla f(x_0)\|_2 \leq \frac{2}{L}\frac{q^T}{1-q}\|\nabla f(x_0)\|_2.$$

That is to say that if we want $\|x_T - x^*\|_2 \leq \epsilon$ then we can just set

$$T \geq \log(\tfrac{1}{q})^{-1} \log\left(\frac{2\|\nabla f(x_0)\|_2}{\epsilon L(1-q)} \right)$$

to guarantee this[a]. *Hint* Define $F(x) = x - \alpha \nabla f(x)$. Prove that the eigenvalues $\gamma_i(x)$ of the Jacobian $\nabla F(x)$ are always bounded above by q: $|\gamma_i(x)| \leq q$. Apply the Banach Fixed Point Theorem from analysis.

---

[a]It turns out that this is the best convergence rate you could hope for with this class of functions f. [Arjevani2016]

Note that $x_t = F(F(\cdots F(x_0))) = F^n(x_0)$ and $\nabla^2 F = I - \alpha \nabla^2 f$. Follow the hint. ∎