Problem 1 (Stochastic Gradient Descent). In this problem we will try to analyze stochastic gradient descent algorithm for strongly convex functions.

Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) is a \( L \)-smooth, \( \mu \)-strongly convex function with optimal point at \( x^* \). In particular
\[
\langle \nabla f(x), x - x^* \rangle \geq \frac{\mu}{2} \|x - x^*\|^2 + \frac{1}{2L} \|\nabla f(x)\|^2.
\]

We will try to optimize this function by running a stochastic gradient descent algorithm:

**Algorithm 1** Stochastic Gradient Descent

for \( t = 0 \) to \( k - 1 \) do
    \( x^{(t+1)} = x^{(t)} - \eta_t (\nabla f(x^{(t)}) + \epsilon_t) \).
end for

In the algorithm, \( \eta_t \) is a step size that we will choose later. The vector \( \nabla f(x^{(t)}) + \epsilon_t \) is a stochastic gradient for \( f \) at \( x^{(t)} \), in particular, \( \epsilon_t \) is a random variable that only depends on \( x^{(t)} \), and for every \( x \)

\[
\mathbb{E}[\epsilon|x] = 0, \mathbb{E}[\|\epsilon\|^2_2|x] \leq \sigma^2.
\]  

(a) (5 points) Suppose \( \nabla f(x) + \epsilon \) is a stochastic gradient for \( f \) at \( x \) that satisfies Equation (1). Show that
\[
\mathbb{E}[\|\nabla f(x) + \epsilon\|^2_2] = \|\nabla f(x)\|^2 + \sigma^2.
\]

(b) (5 points) Let \( r_t = \mathbb{E}[\|x^{(t)} - x^*\|^2_2] \), show that when \( \eta \leq \frac{1}{L} \),
\[
r_{t+1} \leq (1 - \eta \mu) r_t + \eta^2 \sigma^2.
\]

(Hint: Consider \( r_{t+1} = \mathbb{E}[\|(x^{(t)} - x^*) - \eta (\nabla f(x^{(t)}) + \epsilon_t)\|^2_2] \), and expand out the square.)

(c) (5 points) Show that when \( r_t \geq \frac{2\sigma^2}{\mu} \), we can choose \( \eta_t = \frac{1}{t} \), and get \( r_{t+1} \leq (1 - \frac{\mu}{2L}) r_t \).

(d) (10 points) Suppose \( r_{t_0} = \frac{4\sigma^2}{\mu^2 k} \) for some integer \( k \), and \( k \geq \frac{2L}{\mu} \). Show that we can choose \( \eta_t \) appropriately to ensure \( r_{t_0 + t} \leq \frac{4\sigma^2}{\mu^2 (k+1)} \) for all integer \( t > 0 \).

(Hint: The bound in (b) is quadratic in \( \eta \), optimize that to get a good choice of step size.)