Convex Functions
Recall that a function \( f \) over \( \mathbb{R}^n \) is convex if for any two inputs \( x, y \in \mathbb{R}^n \) and any \( \lambda \in [0, 1] \) we have
\[
    f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).
\]
In other words, the line segment from \((x, f(x))\) to \((y, f(y))\) stays “above” the function. Alternatively, if the function is differentiable then it is convex iff for all \( x, y \), we have
\[
    f(y) \geq f(x) + \langle \nabla f(x), y-x \rangle.
\]
Moreover, recall that a set \( K \subseteq \mathbb{R}^n \) is convex if for any two points \( x, y \in K \) and any \( \lambda \in [0, 1] \) we have that the point \( \lambda x + (1-\lambda)y \) \( \in K \). In other words, the line segment from \( x \) to \( y \) stays inside the set.

Gradient Descent
In the lecture notes (Theorem 7) we show that starting with the point \( x_0 \) and using the gradient descent rule \( x_{t+1} \leftarrow x_t - \eta \nabla f_t(x_t) \) at each step, we get that for any fixed point \( x^* \), we have the following bound.
\[
    \sum_{t=0}^{T-1} f_t(x_t) \leq \sum_{t=0}^{T-1} f_t(x^*) + \frac{\eta}{2} G^2 T + \frac{1}{2\eta} \left\| x_0 - x^* \right\|^2, \tag{1}
\]
where \( G \) is an upper bound on the norm of the gradient \( \left\| \nabla f(x) \right\| \). Let’s see how to use this to find a point \( \hat{x} \) at which the function value is very close to the minimum value.

1. Suppose we get a fixed function \( f_t = f \) at each step. From the above expression, show that if we set \( \hat{x} = \frac{1}{T} \sum_{t=0}^{T-1} x_t \), then
\[
    f(\hat{x}) \leq f(x^*) + \frac{\text{regret}(T)}{T}.
\]

**Solution:** Use the first definition of convexity to show that
\[
    f(\hat{x}) = f \left( \frac{1}{T} \sum_{t=0}^{T-1} x_t \right) \leq \frac{1}{T} \sum_{t=0}^{T-1} f(x_t).
\]
This is \( 1/T \) times the LHS of (1). Which is at most \( 1/T \) times the RHS of (1), which is \( f(x^*) + \text{regret}(T)/T \).
2. Suppose we set the “learning rate” \( \eta = \frac{\|x_0 - x^*\|}{G \sqrt{T}} \). Show that \( \text{regret}(T) \leq G \|x_0 - x^*\| \sqrt{T} \).

**Solution:** Substituting and doing some simple algebra.

3. Combining the above two parts, show that after \( T = \left( \frac{G \|x^* - x_0\|}{\varepsilon} \right)^2 \) steps, the function value \( f(\hat{x}) \leq f(x^*) + \varepsilon \).

**Solution:** Substituting and doing some simple algebra.

Now let’s see what we can get in a setting like that for the experts algorithm. Suppose you know the function \( f(x) = \sum_i c_i x_i \) for some \( c = (c_1, \ldots, c_n) \in [0, M]^n \) (i.e., \( f \) is linear) and suppose we have some convex body \( K \) contained within the unit cube: i.e., \( K \subseteq \{x \mid 0 \leq x_i \leq 1 \ \forall i \in \{1, 2, \ldots, n\} \} \).

4. What is the diameter of \( K \)? (The diameter is the maximum Euclidean distance between two points in \( K \).)

**Solution:** The maximum distance is bounded by the max-distance between \((0, 0, \ldots, 0)\) and \((1, 1, \ldots, 1)\), which is \( \sqrt{1^2 + 1^2 + \ldots + 1^2} = \sqrt{n} \).

5. If you start with some \( x_0 \in K \), give an upper bound on \( \|x_0 - x^*\| \).

**Solution:** \( \|x_0 - x^*\| \) is at most the diameter of the cube, so setting \( D := \sqrt{n} \) suffices.

6. What is the maximum value of \( \|\nabla f(x)\| \) at any point \( x \in K \)?

**Solution:** \( \nabla f(x) = \nabla (c_1 x_1 + \ldots + c_n x_n) = c \), so \( \|\nabla f(x)\| = \|c\| \leq M \sqrt{n} \). Hence you can set \( G = M \sqrt{n} \).

7. Plugging these values in, what expressions do you get for \( T, \eta \)?

**Solution:** Recall \( T = \left( \frac{\|x_0 - x^*\|G}{\varepsilon} \right)^2 \). Substituting, \( \eta = \frac{\|x_0 - x^*\|}{G \sqrt{T}} = \frac{\varepsilon}{M^2 n} \).