Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function. We assume that

1. $f$ is convex ;
2. $f$ has a global minimizer $x_*$ ;
3. $f$ is differentiable and, for any $x \in \mathbb{R}^d$,

$$||\nabla f(x)||_2 \leq 1.$$

We fix a starting point $x_0$ and run gradient descent from this point, with a sequence of positive stepsizes $(h_k)_{k\in\mathbb{N}}$ :

$$x_{k+1} = x_k - h_k \nabla f(x_k).$$

1. a) Show that, for any $k \in \mathbb{N}$,

$$f(x_k) - f(x_*) \leq \langle \nabla f(x_k), x_k - x_* \rangle .$$

b) Show that, for any $k \in \mathbb{N}$,

$$||x_{k+1} - x_*||_2^2 \leq ||x_k - x_*||_2^2 - 2h_k(f(x_k) - f(x_*)) + h_k^2 ||\nabla f(x_k)||_2^2.$$

c) Show that, for any $n \in \mathbb{N}$,

$$2\sum_{k=0}^{n} h_k(f(x_k) - f(x_*)) \leq ||x_0 - x_*||_2^2 - ||x_{n+1} - x_*||_2^2 + \sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2.$$

d) For any $n$, let $k_n \in \{0, \dots, n\}$ be such that

$$f(x_{k_n}) = \min_{k=0,\dots,n} f(x_k).$$

Show that, for any $n$,

$$2(f(x_{k_n}) - f(x_*)) \left( \sum_{k=0}^{n} h_k \right) \leq ||x_0 - x_*||_2^2 - ||x_{n+1} - x_*||_2^2 + \sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2.$$

e) Show that, for any $n$,

$$2(f(x_{k_n}) - f(x_*)) \left( \sum_{k=0}^{n} h_k \right) \leq ||x_0 - x_*||_2^2 + \sum_{k=0}^{n} h_k^2.$$

2. In this question, we assume that, for any $k$, $h_k = \frac{1}{\sqrt{k+1}}$. Show that, for any $n$,

$$f(x_{k_n}) - f(x_*) \leq \frac{||x_0 - x_*||_2^2 + 2 + \log(n)}{\sqrt{n+2}}.$$

Hint : You can use the fact that, for any $n$,

$$\sum_{k=1}^{n+1} \frac{1}{k} \leq 2 + \log(n) \quad \text{and} \quad \sum_{k=1}^{n+1} \frac{1}{\sqrt{k}} \geq \frac{\sqrt{n+2}}{2}.$$

3. In this question, we assume that the sequence of stepsizes is constant : there exists $\eta > 0$ such that, for any $k \in \mathbb{N}$, $h_k = \eta$.

Give an example of a function $f$ satisfying properties 1, 2, 3, and a starting point $x_0$ such that

$$f(x_{k_n}) - f(x_*) \overset{n \to +\infty}{\not\to} 0.$$

Hint : Define

$$\begin{aligned} f : x \in \mathbb{R} \quad \to \quad & |x| - \frac{\epsilon}{2} && \text{if } |x| \geq \epsilon; \\ & \frac{x^2}{2\epsilon} && \text{if } |x| \leq \epsilon, \end{aligned}$$

for some $\epsilon > 0$ small enough.

1. a) Let $k$ be fixed. We apply the characterization of convexity for differentiable functions : at $x_*$, $f$ is above its tangent at $x_k$, that is

$$f(x_*) \geq f(x_k) + \langle \nabla f(x_k), x_* - x_k \rangle \,,$$

which is equivalent to the desired inequality.

b) For any $k$,

$$\begin{aligned}
||x_{k+1} - x_*||_2^2 &= ||x_k - x_* - h_k \nabla f(x_k)||_2^2 \\
&= ||x_k - x_*||_2^2 - 2h_k \langle \nabla f(x_k), x_k - x_* \rangle + h_k^2 ||\nabla f(x_k)||_2^2 \\
&\overset{1.a)}{\leq} ||x_k - x_*||_2^2 - 2h_k(f(x_k) - f(x_*)) + h_k^2 ||\nabla f(x_k)||_2^2.
\end{aligned}$$

c) We deduce from the previous question that, for any $k \in \mathbb{N}$,

$$2h_k(f(x_k) - f(x_*)) \leq ||x_k - x_*||_2^2 - ||x_{k+1} - x_*||_2^2 + h_k^2 ||\nabla f(x_k)||_2^2.$$

Therefore, for any $n \in \mathbb{N}$,

$$\begin{aligned}
2\sum_{k=0}^{n} h_k(f(x_k) - f(x_*)) &\leq \sum_{k=0}^{n} (||x_k - x_*||_2^2 - ||x_{k+1} - x_*||_2^2) + \sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2 \\
&= ||x_0 - x_*||_2^2 - ||x_{n+1} - x_*||^2 + \sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2.
\end{aligned}$$

d) Let $n$ be fixed. For any $k \leq n$, we have, from the definition of $k_n$, $f(x_{k_n}) \leq f(x_k)$. As a consequence, for any $k \leq n$,

$$2h_k(f(x_{k_n}) - f(x_*)) \leq 2h_k(f(x_k) - f(x_k)).$$

and

$$\begin{aligned}
2(f(x_{k_n}) - f(x_*)) \left( \sum_{k=0}^{n} h_k \right) &\leq 2\sum_{k=0}^{n} h_k(f(x_k) - f(x_k)) \\
&\overset{1.c)}{\leq} ||x_0 - x_*||_2^2 - ||x_{n+1} - x_*||^2 + \sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2.
\end{aligned}$$

e) From our third assumption on $f$, $||\nabla f(x_k)||_2 \leq 1$ for any $k \in \mathbb{N}$. Therefore, for any $n \in \mathbb{N}$,

$$\sum_{k=0}^{n} h_k^2 ||\nabla f(x_k)||_2^2 \leq \sum_{k=0}^{n} h_k^2.$$

Since, in addition, $-||x_{n+1} - x_*||_2^2 \leq 0$, we deduce from question 1.d) that

$$2(f(x_{k_n}) - f(x_*)) \left( \sum_{k=0}^{n} h_k \right) \leq ||x_0 - x_*||_2^2 + \sum_{k=0}^{n} h_k^2.$$

2. For any $n \in \mathbb{N}$,

$$\sum_{k=0}^{n} h_k^2 = \sum_{k=1}^{n+1} \frac{1}{k} \leq 2 + \log(n);$$
$$\sum_{k=0}^{n} h_k = \sum_{k=1}^{n+1} \frac{1}{\sqrt{k}} \geq \frac{\sqrt{n+2}}{2}.$$

Plugging these inequalities into the one established at question 1.e) yields

$$(f(x_{k_n}) - f(x_*))\sqrt{n+2} \leq 2(f(x_{k_n}) - f(x_*)) \left( \sum_{k=0}^{n} h_k \right) \leq ||x_0 - x_*||_2^2 + \sum_{k=0}^{n} h_k^2$$
$$\leq ||x_0 - x_*||_2^2 + 2 + \log(n).$$

Therefore,

$$f(x_{k_n}) - f(x_*) \leq \frac{||x_0 - x_*||_2^2 + 2 + \log(n)}{\sqrt{n+2}}.$$

3. We set $\epsilon = \frac{\eta}{2}$ and define $f$ as suggested :

$$f : x \in \mathbb{R} \quad \to \quad \begin{cases} |x| - \dfrac{\epsilon}{2} & \text{if } |x| \geq \epsilon; \\ \dfrac{x^2}{2\epsilon} & \text{if } |x| \leq \epsilon, \end{cases}$$

Let us show that $f$ satisfies properties 1, 2, 3.

We start with property 2. For any $x \in \mathbb{R}$ such that $|x| \geq \epsilon$,

$$f(x) \geq \epsilon - \frac{\epsilon}{2} > 0.$$

For any $x \in \mathbb{R}$ such that $|x| < \epsilon$,

$$f(x) = \frac{x^2}{2\epsilon} \geq 0.$$

Therefore, $f$ is nonnegative over $\mathbb{R}$. Since $f(0) = 0$, it implies that $x_* = 0$ is a global minimizer of $f$.

Let us now show that $f$ is differentiable and compute its derivative. The function $|.|$ is differentiable over $\mathbb{R} - \{0\}$ so $f$ is differentiable over $]-\infty; -\epsilon] \cup [\epsilon; +\infty[$, with derivative

$$f'(x) = -1 \quad \forall x \in ]-\infty; -\epsilon];$$
$$f'(x) = 1 \quad \forall x \in [\epsilon; +\infty[.$$

(The derivative is only a left derivative when $x = -\epsilon$ and a right derivative when $x = \epsilon$.)

The square function is differentiable over $\mathbb{R}$ so $f$ is differentiable over $[-\epsilon; \epsilon]$, with derivative

$$f'(x) = \frac{x}{\epsilon} \quad \forall x \in [-\epsilon; \epsilon].$$

(The derivative is only a right derivative when $x = -\epsilon$ and a left derivative when $x = \epsilon$.)

Since the left and right derivatives coincide in $x = -\epsilon$ and $x = \epsilon$, the function $f$ is differentiable at $-\epsilon$ and $\epsilon$ and therefore differentiable over $\mathbb{R}$.

For any $x$ such that $|x| \geq \epsilon$, we have $|f'(x)| = 1$ and, for any $x$ such that $|x| \leq \epsilon$, we have $|f'(x)| = \frac{|x|}{\epsilon} \leq 1$. As a consequence, the norm of the gradient (that is, in this case, the derivative), is always at most 1 and Property 3 holds.

Now that we have computed the derivative, we can easily show that $f$ is convex : its derivative is continuous, nondecreasing (actually constant) over $]-\infty; -\epsilon]$, increasing over $[-\epsilon; \epsilon]$, nondecreasing again over $[\epsilon; +\infty[$. Therefore, the derivative is nondecreasing over $\mathbb{R}$ and $f$ is convex.

We consider the starting point $x_0 = \frac{\eta}{2} = \epsilon$. With this definition,

$$\begin{aligned} x_1 &= x_0 - h_0 f'(x_0) \\ &= \epsilon - \eta \times 1 \\ &= -\epsilon \end{aligned}$$

and

$$
\begin{aligned}
x_2 &= x_1 - h_0 f'(x_1) \\
&= -\epsilon - \eta \times (-1) \\
&= \epsilon.
\end{aligned}
$$

We can iteratively reapply this result and we obtain that $x_k = -\epsilon$ for all odd $k$ and $x_k = \epsilon$ for all even $k$. In particular, $x_k \not\to x_* = 0$ when $k \to +\infty$.