Up from now we assume the gradient of f to fulfill as so called Lipschitz condition:
For some $L > 0$ we assume for all x, y to have

$$|\nabla f(x) - \nabla f(y)| \leq L|x - y| \tag{1}$$

This condition is often fulfilled in practical applications.

We aim to show $f(x_{n+1}) < f(x_n)$ for certain choice of τ:

We denote $h_n := -\nabla f(x_n)$. By the mean-value theorem there is $\theta \in [0, 1]$ such that

$$f(x_{n+1}) = f(x_n) + \nabla f(x_n + \theta\tau h_n) \cdot \tau h_n$$

We can add and subtract $\nabla f(x_n) \cdot \tau h_n$:

$$f(x_{n+1}) = f(x_n) + \nabla f(x_n) \cdot \tau h_n + (\nabla f(x_n + \theta\tau h_n) - \nabla f(x_n)) \cdot \tau h_n$$

We estimate this by using Cauchy-Schwarz inequality:

$$f(x_{n+1}) \leq f(x_n) + \nabla f(x_n) \cdot \tau h_n + \tau|\nabla f(x_n + \theta\tau h_n) - \nabla f(x_n)||h_n|$$

We back-insert $h_n := -\nabla f(x_n)$ and we make use of (1):

$$f(x_{n+1}) \leq f(x_n) - \tau|\nabla f(x_n)|^2 + \tau^2 L|\nabla f(x_n)|^2$$

By choosing $\tau < 1/L$ we have $-\tau + \tau^2 L < 0$, and thus

$$f(x_{n+1}) \leq f(x_n) + (-\tau + \tau^2 L)|\nabla f(x_n)|^2 < f(x_n)$$

This was our aim to show.