By the Cauchy-Schwarz inequality we have

$$|g'(x)| = |\nabla f(x) \cdot h| \leq |\nabla f(x)||h| = |\nabla f(x)| \qquad (1)$$

This shows, the maximal absolute value of our optimization is restricted by $|\nabla f(x)|$.

But on the contrary if we set $h := \nabla f(x)/|\nabla f(x)|$ then we see

$$|\nabla f(x) \cdot h| = \left|\nabla f(x) \cdot \frac{\nabla f(x)}{|\nabla f(x)|}\right| = \frac{|\nabla f(x)|^2}{|\nabla f(x)|} = |\nabla f(x)|$$

This shows, we can attain the maximum in (1) with this specific value for $h$.

But by knowing how to maximize the absolute value, provides us with the $h$ which minimizes the value itself. In detail, by setting $h = -\nabla f(x)/|\nabla f(x)|$ we compute

$$g'(0) = -\nabla f(x) \cdot \frac{\nabla f(x)}{|\nabla f(x)|} = -\frac{|\nabla f(x)|^2}{|\nabla f(x)|} = -|\nabla f(x)|$$

The r.h.s is the smallest value which can be achieved, otherwise it would contradict above found maximal absolute value.

We keep this all as a result:

**Theorem 1** *The negative gradient is the direction of steepest fall.*