Рассмотрим произвольную бесконечно дифференцируемую функцию $J(x)$, $x = (x_1, ..., x_n)$ — вектор параметров, и произвольное приращение аргумента Δx. Пользуясь формулой Тейлора для функции многих переменных, запишем:

$$J\left(x + \Delta x\right) - J\left(x\right) = J\left(x\right) + \left(\Delta x, \nabla\right) J\left(x\right) + o\left(\|\Delta x\|\right) - J\left(x\right) = \\ = \left(\Delta x, \nabla\right) J\left(x\right) + o\left(\|\Delta x\|\right) = \left(\Delta x, \nabla J\right) + o\left(\|\Delta x\|\right),$$

где $\|\Delta x\| = \sqrt{\sum\limits_{i=1}^{n} \left(\Delta x_i\right)^2}$. Мы получили, что при достаточно малом приращении Δx изменение функции приближённо равно скалярному произведению приращения аргумента и градиента функции: $(\Delta x, \nabla J)$. Далее будем считать, что приращение аргумента имеет фиксированную достаточно малую длину ϵ: $\|\Delta x\| = \epsilon$.

Для произвольного приращения аргумента верно следующее (использовано неравенство Коши-Буняковского):

$$J(x + \Delta x) - J(x) = (\nabla J, \Delta x) \leq \|\nabla J\| \cdot \|\Delta x\| = \|\nabla J\| \cdot \epsilon.$$

Рассмотрим приращение вдоль градиента:

$$\Delta x' = \frac{\nabla J}{\|\nabla J\|} \epsilon.$$

Тогда в силу определения скалярного произведения получим:

$$J(x + \Delta x') - J(x) = (\nabla J, \Delta x') = \left(\nabla J, \frac{\nabla J}{\|\nabla J\|} \epsilon\right) = \|\nabla J\| \cdot \epsilon.$$

Очевидно, что $\Delta x'$ доставляет максимум приращению функции, т.е. градиент задаёт направление наибольшего роста функции:

$$J(x + \Delta x) - J(x) \leq \|\nabla J\| \cdot \epsilon = J(x + \Delta x') - J(x).$$