1.1 ELI5

Imagina que estás jugando un videojuego de shooters
como Fortnite, pero el objetivo es encontrar y
capturar una bandera que está en el punto más
alto del mapa. En este mapa hay mucha niebla y
es muy difícil ver a lo lejos, pero tenemos algunas
herramientas para guiarnos. Decides que la estrategia
más inteligente es ir subiendo todos los montes
que nos encontremos y esperamos que sean los más
altos. Para llegar más rápido usamos un jetpack para
brincar entre puntos en vez de caminar. Entonces, ¿Cómo
subimos montes? No podemos ver muy lejos pero podemos
ver el punto en el que estamos parados bastante bien.
Con poner sobre el suelo algo como una pelota o un
nivelador podemos ver la inclinación: hacia dónde
es y qué tan inclinado está. Entonces podemos hacer
eso y decidimos lo lógico: tiene sentido brincar
en "sentido contrario" a la dirección en la que
rodó la pelota. Entonces usamos una regla simple:

1. Si no se movió la pelota no hay inclinación y ent.
   no hay hacia dónde subir. No tenemos porqué movernos.
   esperamos haber llegado al punto más alto.

2. Si hay inclinación, brincamos con el jetpack
   en dirección "hacia arriba" del monte.

Pero ¿Qué tan lejos brincamos?

Si brincamos demasiado lejos podemos brincarnos la cima! Entonces queremos brincar tan lejos como se pueda sin pasarnos para no desperdiciar gasolina del jetpack. Para ayudarnos, lanzamos piedritas en la dirección en la que vamos a saltar (idea cruda de backtracking. Nota para examen no para la niña). Aventamos piedritas más y más lejos cada vez hasta que ya no vuelvan a bajar rodando. La última distancia a la que tiramos es a la que vamos a brincar.

Una vez habiendo brincado volvemos a checar la inclinación. Paramos hasta que no haya inclinación y estemos en piso plano. Cuando hayamos llegado a piso plano habremos llegado a la cima que buscábamos, o al menos llegamos a ~~una~~ cima. No sabemos si es la más alta y es difícil saber si lo es.

[1.2] DEMO (CH 3 - pg. 42 Convergence rate of steepest descent)

f cuadratica convexa $f(x) = \frac{1}{2} x^T Q x - b^T x$

P.D. el minimizador de una dim. sobre la línea

$x_k + \alpha p_k$ es $\quad \alpha_k = - \dfrac{\nabla f_k^T p_k}{p_k^T Q p_k}$

Queremos minimizar sobre $x_k + \alpha p_k$. Ent. basta con encontrar los puntos estacionarios

$$\frac{\partial}{\partial \alpha} \left(f(x_k + \alpha p_k) \right) = 0 .$$

Tomamos $f(x) = \frac{1}{2} x^T Q x - b^T x$

$$= \frac{1}{2} (x_k + \alpha p_k)^T Q (x_k + \alpha p_k) - b^T (x_k + \alpha p_k)$$

$$= \frac{1}{2} x_k^T Q x_k + \frac{1}{2} x_k^T Q \alpha p_k + \frac{1}{2} \alpha p_k^T Q x_k + \frac{1}{2} \alpha^2 p_k^T Q p_k - b^T x_k$$

$$- \alpha b^T p_k$$

$$= \frac{1}{2} x_k^T Q x_k + \alpha x_k^T Q p_k + \frac{1}{2} \alpha^2 p_k^T Q p_k - b^T x_k - \alpha b^T p_k$$

$$\therefore \frac{\partial}{\partial \alpha} \left(f(x_k + \alpha p_k) \right) = x_k^T Q p_k + \alpha p_k^T Q p_k - b^T p_k$$

$$= \cancel{p_k} \left(x_k^T Q + \alpha p_k^T Q - b^T \right) p_k$$

$$= \left(\alpha p_k^T Q + \underbrace{x_k^T Q - b^T}_{(*)} \right) p_k$$

Q simétrica por f convexa

El libro señala que $\nabla f(x) = Qx - b \Rightarrow \nabla f(x^T) = Qx^T - b^T = x_k^T Q - b^T$

$$\cancel{\left(\alpha p_k^T Q + \nabla f(x_k^T) \right) p_k = 0}$$

$$\therefore \quad \Leftrightarrow \quad \cancel{\alpha p_k^T Q = - \nabla f(x_k^T) \Rightarrow \alpha = - \frac{\nabla f(x_k^T)}{p_k^T Q}}$$

No puedo dividir entre cosas que no se si son números reales (i.e $p_k^T Q$)

$$\frac{\partial}{\partial \alpha} f(x+\alpha p_k) = x_k^T Q p_k + \alpha p_k^T Q p_k - b^T p_k$$

$$= \alpha p_k^T Q p_k + (x_k^T Q - b^T) p_k$$

Como Q simétrica porque f convexa

$x_k^T Q - b^T = Q x_k^T - b^T = \nabla f(x_k^T)$ (lo dice el libro).

$$\frac{\partial}{\partial \alpha} f(x+\alpha p_k) = \alpha p_k^T Q p_k + \nabla f(x_k^T) p_k = 0$$

$$\Leftrightarrow \quad \alpha p_k^T Q p_k = - \nabla f(x_k^T) p_k$$

$p_k^T Q p_k$ está bien definido y es escalar dado

$$\alpha = - \frac{\nabla f(x_k^T) p_k}{p_k^T Q p_k} \quad \text{q.e.d} \quad \blacksquare$$