Gradiente Descendiente y Newton Rapson.

$$\frac{\partial L}{\partial \vec{x}} = \frac{2}{N}\left(A^T A x - A^T b\right)$$

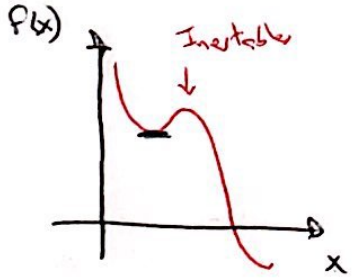
$$\frac{\partial^2 L}{\partial \vec{x}^2} = \frac{2}{N}\left(A^T A\right)$$ "Definida positiva" ← Tiene mínimo Global.

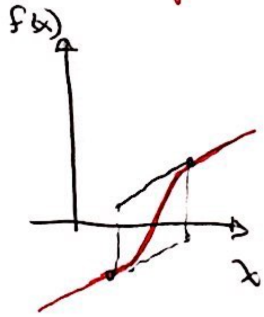
→ Dado un $\vec{x}_0$, por ejemplo $\vec{0}$, yo puedo usar Newton Rapson para encontrar el siguiente $\vec{x}_1$ que minimiza L ⇒ ¡Encontrar una raiz!

$$L(\hat{b}, b_{obs}) \geq l_{min} \Rightarrow L(\hat{b}, b_{obs}; \vec{x}) - l_{min}$$ | tiene al menos 1 raiz

$$\vec{x}_j = \vec{x}_{j-1} - \underbrace{\frac{\partial L}{\partial \vec{x}}}\epsilon$$ con ϵ un parámetro pequeño "Tasa de aprendizaje"

Recordemos que Newton Rapson tiene sus limitaciones inherentes al problema.





Actualizaciones dependen de la pendiente y si es (a) ¿cero...? o (b) ¿si el peso ϵ es muy grande?

¿Alternativas? Expandir la función de costo a segundo orden

$$L(\hat{b}, b_{obs}; \vec{x} + \delta\vec{x}) \simeq L(\hat{b}, b_{obs}; \vec{x}) + \frac{\partial L}{\partial \vec{x}}\bigg|_{\vec{x}}^{T} \cdot \delta\vec{x} + \frac{1}{2}\delta\vec{x}^T \hat{H}\, \delta\vec{x} + O(\delta x^3)$$

donde $\hat{H}$ se conoce como el Hessiano.

$$\mathbb{H} = \begin{array}{c|c} & \begin{array}{cccc} x_1 & x_2 & \cdots & x_N \end{array} \\ \hline \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_N \end{array} & \begin{pmatrix} \frac{\partial^2 L}{\partial x_1 \partial x_1} & \frac{\partial^2 L}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 L}{\partial x_1 \partial x_N} \\ \frac{\partial^2 L}{\partial x_2 \partial x_1} & \frac{\partial^2 L}{\partial x_2^2} & \cdots & \frac{\partial^2 L}{\partial x_2 \partial x_N} \\ \vdots & \vdots & & \vdots \end{pmatrix} \end{array}$$

Apliquémoslo a nuestro problema!

$$L(\mathbb{A}, \mathbb{b}_{obs}; \vec{x}) = \frac{1}{N} \left(\mathbb{A} \cdot \vec{x} - \mathbb{b}_{obs}\right)^2$$

Partimos de $\vec{x} = \vec{0}$

$$\frac{\partial L}{\partial \vec{x}} = \frac{2}{N} \mathbb{A}^T \left(\mathbb{A} \cdot \mathbb{x} - b_{obs}\right)$$

Iterativamente hasta que $\mathbb{x} = \mathbb{x}_{opt}$; $L <$ cutoff (10^{-10})