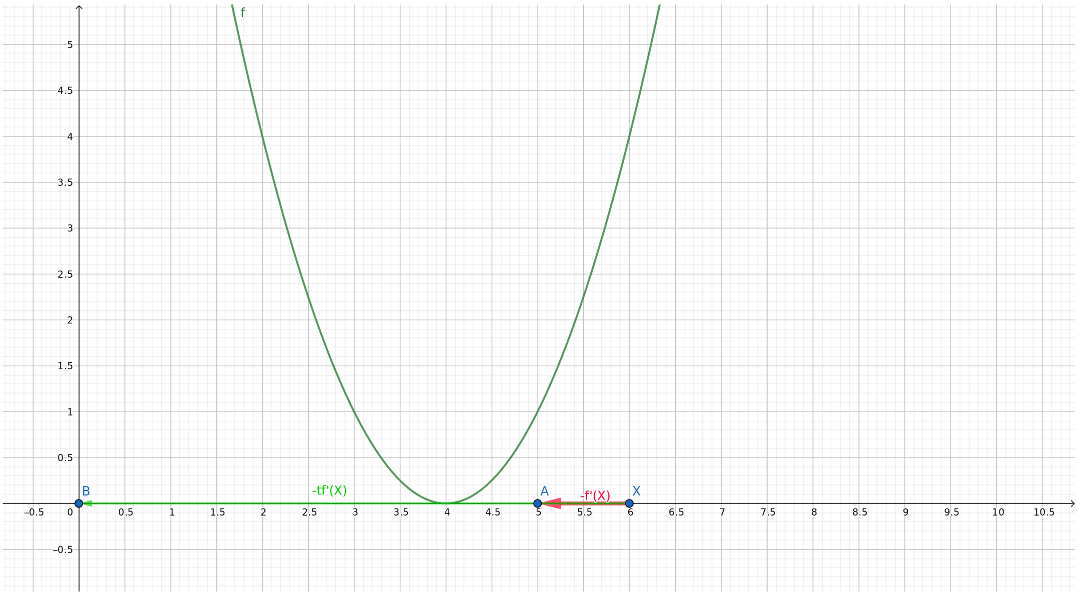
Optimisation convexe 2

Graphe de de $f : \mathbb{R} \to \mathbb{R}$



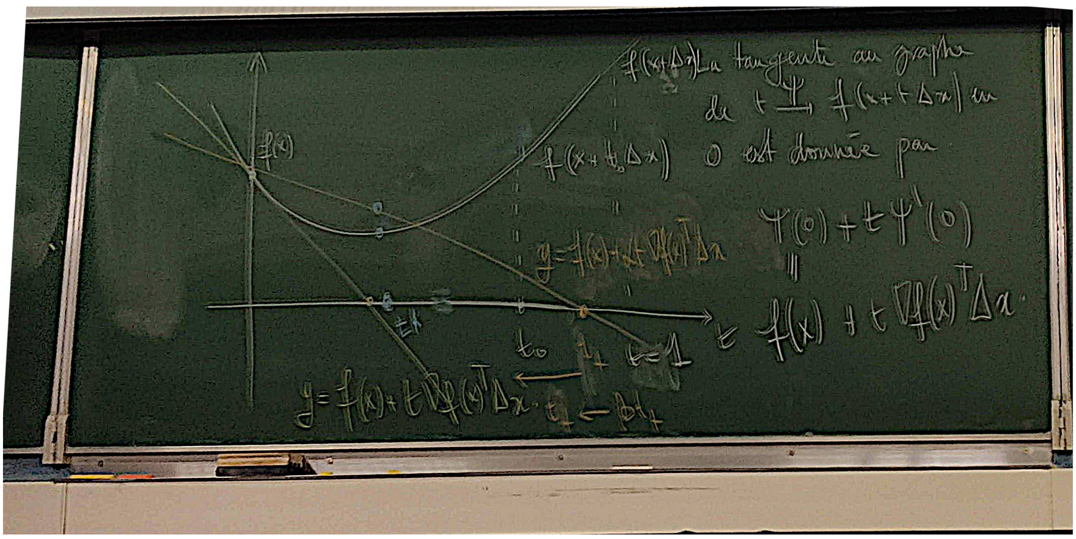
Calculer le pas

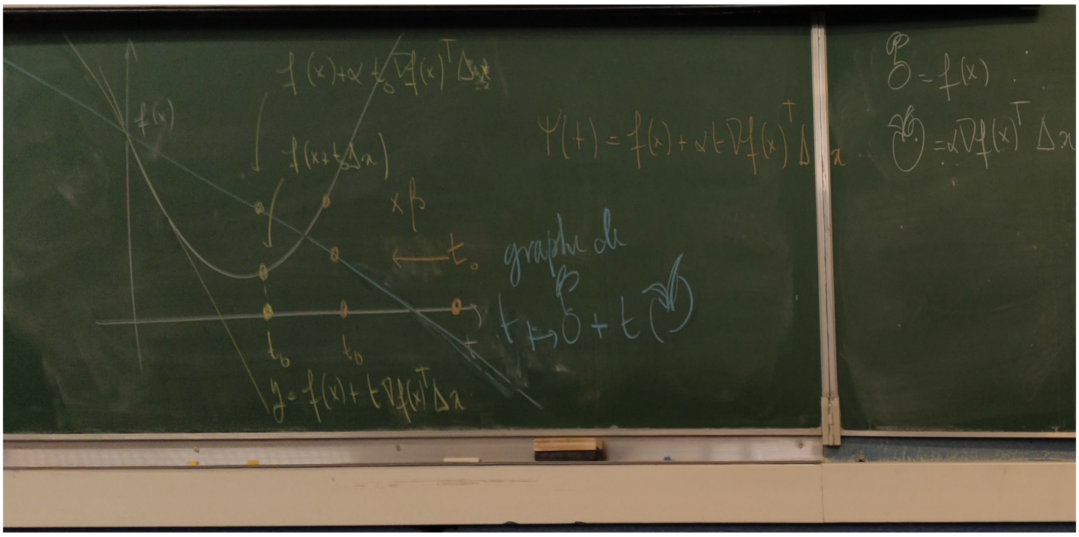
Calcul du pas optimal

trouver t^* par la minimisation de $f(x + t\Delta x)$

la tangente au graphe de $t \xrightarrow{\psi} f(x + t\Delta x)$ en 0 est donnée par
$\psi(0) + t\psi'(0) = f(x) + t$
$= f(x) + t\nabla f(x)^T \Delta x$

On s'arrete des qu'on trouve:
t^* tq
$f(x + t^*\Delta x) \leq f(x) + \alpha t^* \nabla f(x)^T \Delta x$





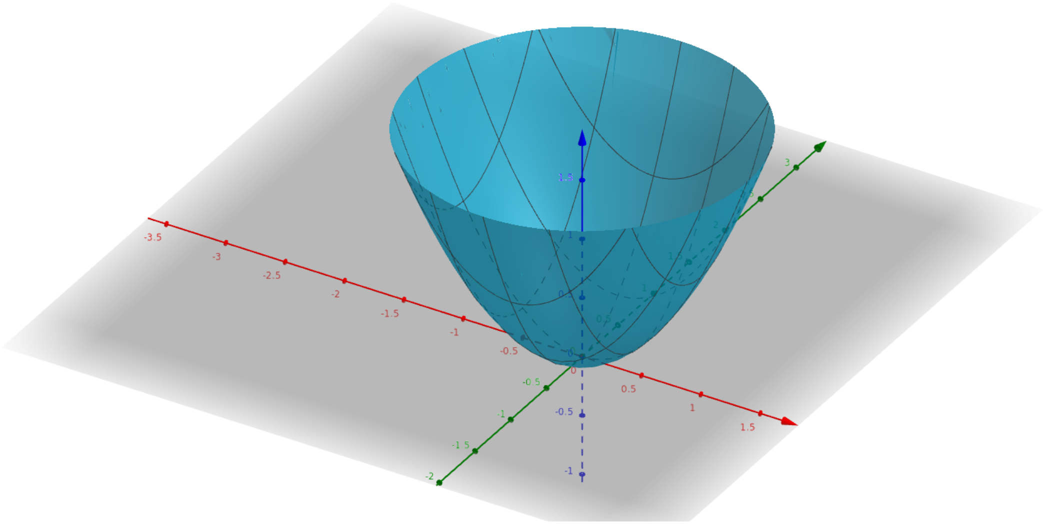
La hessienne d'une fonction f en un point definit une fonction quadratique.

$$X \longmapsto X^T \underbrace{\nabla^2 f(a)}_{\text{matrice carre (hessienne de f en a)}} X$$

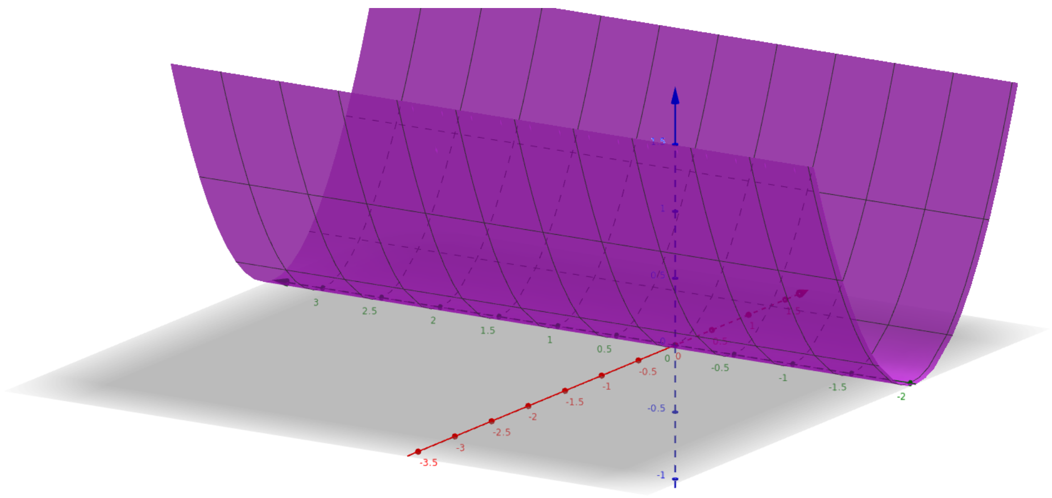
Dire que $a \mapsto \nabla^2 f(a)$ est majoree sur un lieu S de son domaine de definition est equivalent au fait de dire que les valeurs propres (fonctions en a) de la hessienne sont des fonctions majorees.

2. $\forall a \in S$ on a que les valeurs propres de la hessienne de f sont minorees par une meme constante $m > 0$.

Strictement convexe:



Non Strictement convexe:



Generaliser la descente classique

$f(x+v) = f(x) + \nabla f(x)^T v + o(v)$

- On remplace $f(x+v)$ par son approximation au 1er ordre; cad $f(x) + \nabla f(x)^T v$
- On cherche la direction (cad les vecteurs de norme 1 ($\|.\|_2$)) tq $f(x) + \nabla f(x)^T v$ est minimal. On cherche donc a calculer $v^* = \text{argmin}\{\nabla f(x)^T v | \|v\|_2 = 1\}$

Rq: Si v^* minimise $\nabla f(x)^T v$ ssi $-v^*$ maximise $\nabla f(x)^T v$ pour $\|v\|_2 = 1$

Rappel: Cauchy Schwarz : $|\nabla f(x)^T v| \leq \|\nabla f(x)\|_2 \|v\|_2$

$|\nabla f(x)^T v| \leq \|\nabla f(x)\|_2$

En prenant $v = \frac{\nabla f(x)}{\|\nabla f(x)\|_2}$ (hyp denominateur != 0)

$$\nabla f(x)^T v^* = \left(\nabla f(x)^T \nabla f(x)\right) x \frac{1}{\|\nabla f(x)\|_2} = \frac{\|\nabla f(x)\|_2^2}{\|\nabla f(x)\|_2} = \|\nabla f(x)\|_2$$

Donc pour que v^* maximise $\nabla f(x)^T v^*$ pour $\|v\|_2 = 1$
Ainsi $\frac{-\nabla f(x)}{\|\nabla f(x)\|_2}$ minimise $\nabla f(x)^T v$ pour $\|v\|_2 = 1$

Au lieu d'utiliser une direction normalisee, pour la mise a jour, on regarde plutot
$\Delta_{x_{sd}} = \|\nabla f(x)\|_2$

> nsd => normalized steepest descent

Pour la norme 1 on s'interesse au calcul de:
$\text{argmin}\{\nabla f(x)^T v \mid \|v\|_1 = 1\}$

$\text{argmin}\{\nabla f(x)^T v \mid \|v\|_1 \leq 1\}$

Ceci est un programme lineaire, ces points optimaux sont des points extremaux du lieu admissible

Ces points extremaux sont les points:
$\mathcal{F}_{e_i}$ pour $i \in \{1, \ldots, n\}$

$\text{argmin}\left\{\nabla f(x)^T v \mid \|v\|_1 \leq 1\right\} = Ie_i$ pour un certain $i \in \{1, \ldots, n\}$

$\nabla f(x)^T e_i$ est la projection de $\nabla f(x)$ sur e_i cad $\frac{\partial f(x)}{\partial x_i}$ Le e_i qui realise l'argmin.

$\Delta x_{sd} =$ la projection de $\nabla f(x)$ le long de Δx_{nsd}
$\Rightarrow \Delta x_{sd} = -\frac{\partial f(x)}{\partial x_i} e_i$

$\Delta x_N = -(\nabla^2 f(x))^{-1}(\nabla f(x))$

$f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v = \Psi(v)$

Ψ est minimum ssi $\nabla \Psi(v) = 0$

$$\nabla \Psi(v) = \nabla f(x) + \nabla^2 f(x) v => \nabla \Psi(v) = 0 <=> \nabla^2 f(x) v = -\nabla f(x)$$