Analyse, Classification et Indexation des données (ACID)

Descente de gradient

et autres méthodes d'optimisation

Akka Zemmari

LaBRI, Université de Bordeaux

2024 - 2025

Introduction

Problème d'optimisation :

- **Définition :** Un problème d'optimisation consiste à trouver les valeurs des paramètres qui minimisent (ou maximisent) une fonction objectif.

Introduction

Problème d'optimisation :

- **Définition :** Un problème d'optimisation consiste à trouver les valeurs des paramètres qui minimisent (ou maximisent) une fonction objectif.
- **Exemples :** Minimiser le coût de production, maximiser le profit, ajuster les paramètres d'un modèle pour minimiser l'erreur de prédiction.

Introduction

Problème d'optimisation :

- **Définition :** Un problème d'optimisation consiste à trouver les valeurs des paramètres qui minimisent (ou maximisent) une fonction objectif.
- **Exemples :** Minimiser le coût de production, maximiser le profit, ajuster les paramètres d'un modèle pour minimiser l'erreur de prédiction.
- **Importance :** L'optimisation est au cœur de nombreuses applications, de la logistique à la finance, en passant par le **machine learning.**

Introduction

La descente de gradient :

En une phrase :

Méthode <u>itérative</u> utilisée pour trouver le <u>minimum d'une fonction</u> en se déplaçant progressivement dans la direction opposée au gradient (pente).

Introduction

Lien avec le Machine Learning :

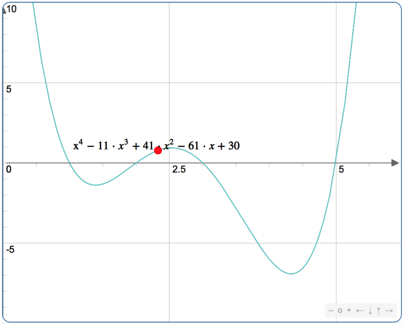
- Entraînement de Modèles : en ML, la descente de gradient est utilisée pour ajuster les poids des modèles (régression, réseaux de neurones, etc.) afin de minimiser une fonction de perte (erreur entre les prédictions et les valeurs réelles).
- Optimisation des Paramètres : C'est le principal algorithme d'optimisation pour les modèles d'apprentissage supervisé, non supervisé et par renforcement.

Intuition

Soit la fonction :

$$f(x) = (x-1)(x-2)(x-3)(x-5) = x^4 - 11x^3 + 41x^2 - 61x + 30.$$

et son graphique sur l'intervalle $[0,6]$:



Intuition

- Question : Trouver x^* tel que

$$f(x^*) = \min_{x \in [0,6]} \{f(x)\}$$

Intuition

- Question : Trouver x^* tel que

$$f(x^*) = \min_{x \in [0,6]} \{f(x)\}$$

Ce que l'on peut formuler par

$$x^* = argmin_{x \in [0,6]} (f(x))$$

Intuition

- Question : Trouver x^* tel que

$$f(x^*) = \min_{x \in [0,6]} \{f(x)\}$$

Ce que l'on peut formuler par

$$x^* = argmin_{x \in [0,6]} (f(x))$$

- (une) réponse : calculer $f'(x)$:

$$f'(x) = 4x^3 - 33x^2 + 82x - 61$$

et résoudre $f'(x) = 0$ ... dur dur ...

Intuition

- Question : Trouver x^* tel que

$$f(x^*) = \min_{x \in [0,6]} \{f(x)\}$$

Ce que l'on peut formuler par

$$x^* = argmin_{x \in [0,6]} (f(x))$$

- (une) réponse : calculer $f'(x)$:

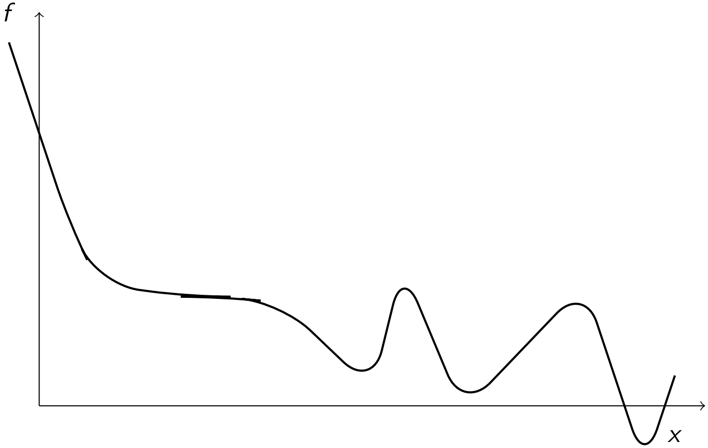
$$f'(x) = 4x^3 - 33x^2 + 82x - 61$$

et résoudre $f'(x) = 0$ ... dur dur ...

- Une solution : Descente du gradient (voir l'explication sur le `notebook`).

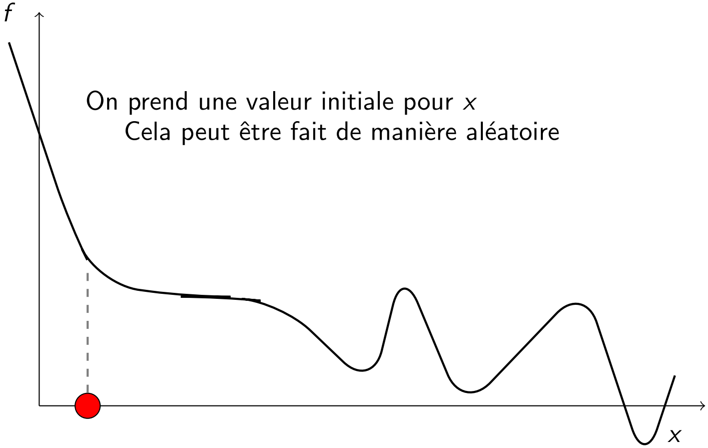
Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.

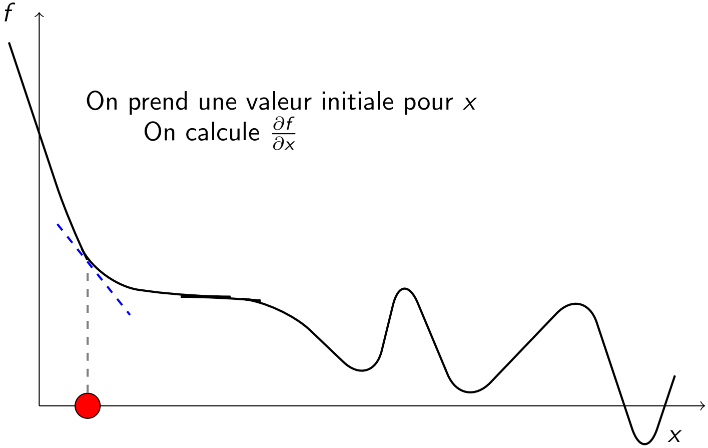


On prend une valeur initiale pour x

Cela peut être fait de manière aléatoire

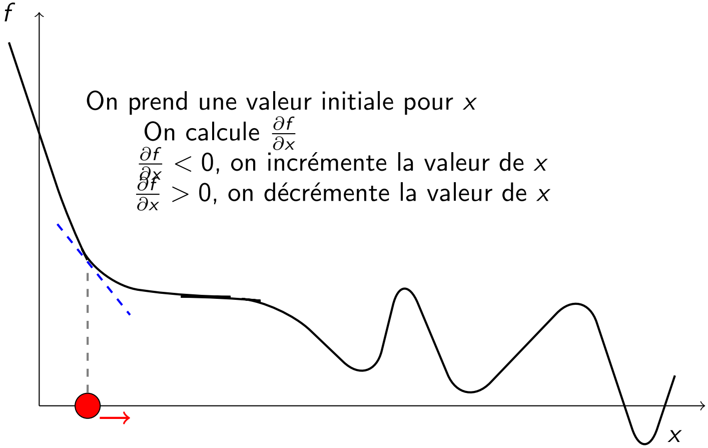
Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



On prend une valeur initiale pour x

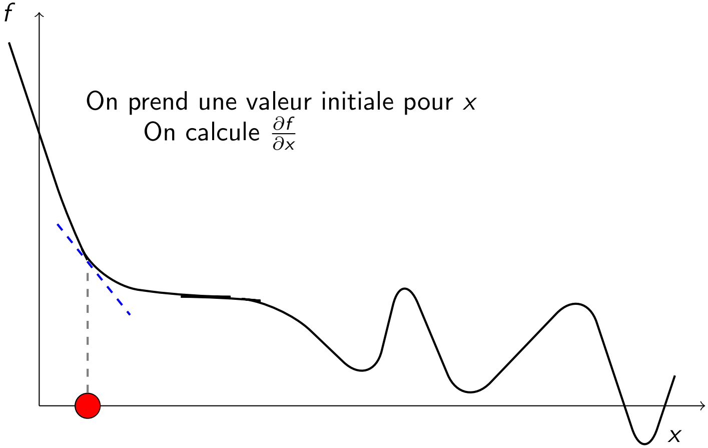
On calcule $\frac{\partial f}{\partial x}$

$\frac{\partial f}{\partial x} < 0$, on incrémente la valeur de x

$\frac{\partial f}{\partial x} > 0$, on décrémente la valeur de x

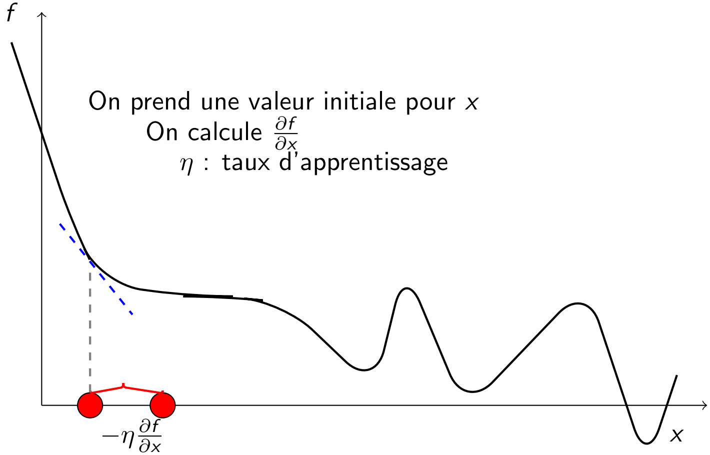
Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



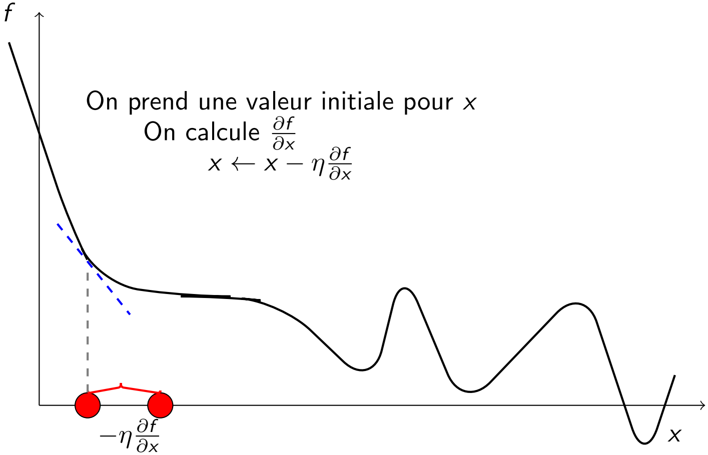
Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.

On prend une valeur initiale pour x

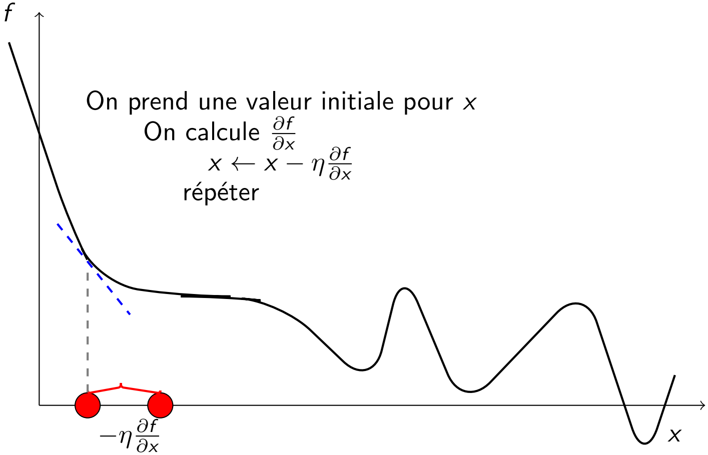
On calcule $\frac{\partial f}{\partial x}$

$$x \leftarrow x - \eta \frac{\partial f}{\partial x}$$



Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



On prend une valeur initiale pour x

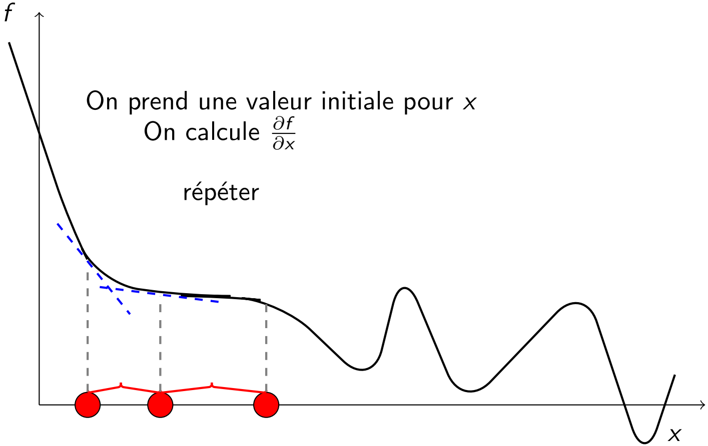
On calcule $\frac{\partial f}{\partial x}$

$x \leftarrow x - \eta\frac{\partial f}{\partial x}$

répéter

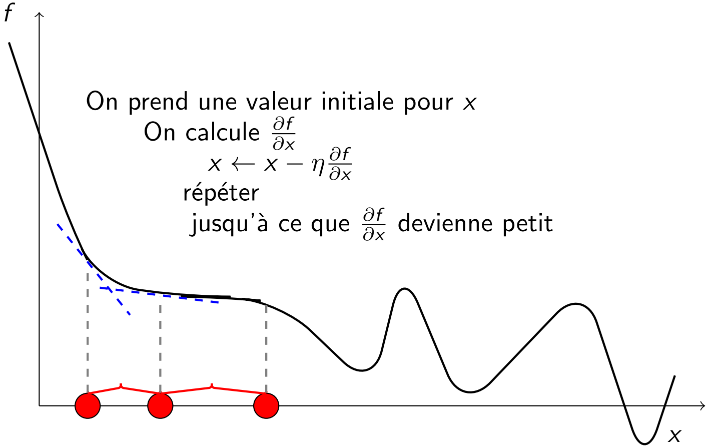
Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



Descente du gradient

Trouver le(s) paramètre(s) x qui minimise(nt) la fonction f.



On prend une valeur initiale pour x

On calcule $\frac{\partial f}{\partial x}$

$x \leftarrow x - \eta \frac{\partial f}{\partial x}$

répéter

jusqu'à ce que $\frac{\partial f}{\partial x}$ devienne petit

Descente du gradient

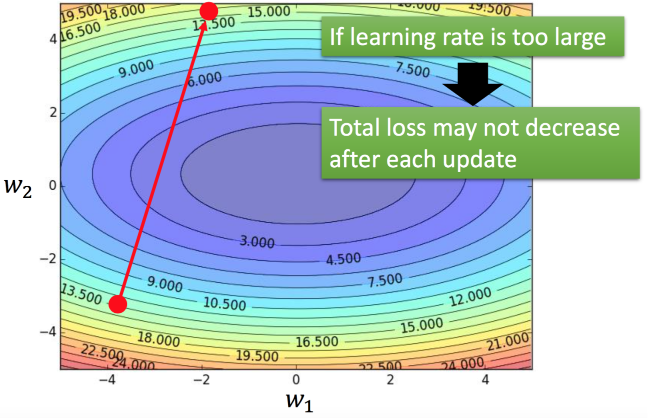
L'algorithme

1. Initialiser avec x_0 (au hasard);
2. Répéter
$$x_{t+1} \leftarrow x_t - \eta \nabla f(x_t);$$
3. Jusqu'à convergence

- au hasard car quasiment impossible de trouver une valeur optimale
- $\nabla f(x_t)$ il s'agit du gradient (voir les rappels, chapitre précédent)
- η est un paramètre qui permet de moduler la correction
- Question : pourquoi le -?
- convergence : Nombre d'itérations fixé, ou différence entre valeurs successives de x_t ou $\nabla f(x_t)$ très petit

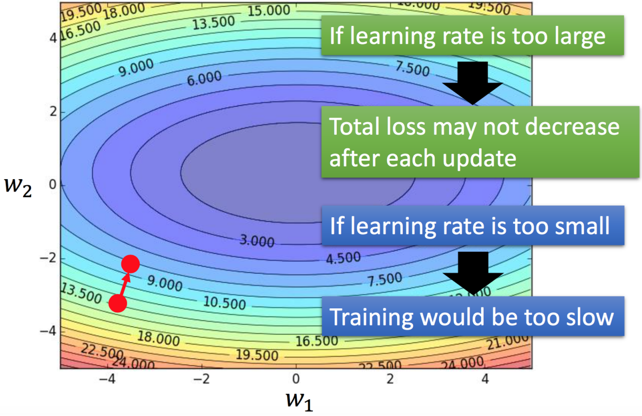
Descente du gradient

de l'importance du choix du taux d'apprentissage η :



Descente du gradient

de l'importance du choix du taux d'apprentissage η :



Descente du gradient

de l'importance du choix du taux d'apprentissage η :

- Un taux trop petit garantit une convergence vers un minimum mais le temps de convergence peut être trop grand ....
- Un taux trop grand peut faire "sauter" le minimum ..

Descente du gradient

de l'importance du choix du taux d'apprentissage η :
Une solution simple : réduire le taux par un facteur toutes les K itérations

- Initialement, on est loin de la destination, on utilise donc un pas assez grand,
- Après quelques itérations, on est proche de la destination, on réduit donc le pas.
- En général, on utilise la mise à jour suivante :

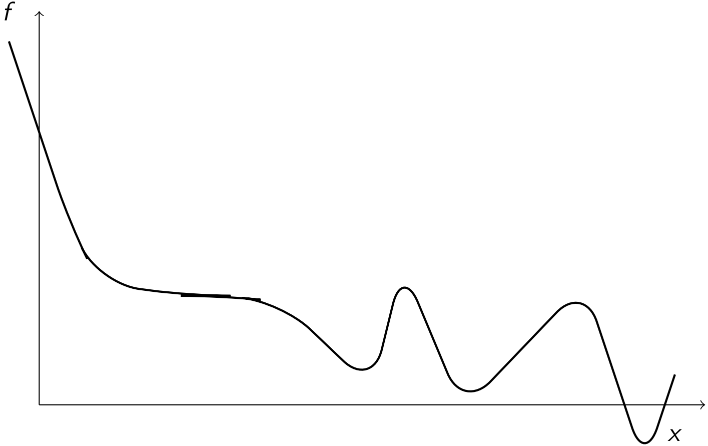
$$\eta^{(t)} = \frac{\eta}{\sqrt{t+1}}$$

Descente du gradient

- La descente du gradient ne garantie pas de trouver un minimum global,

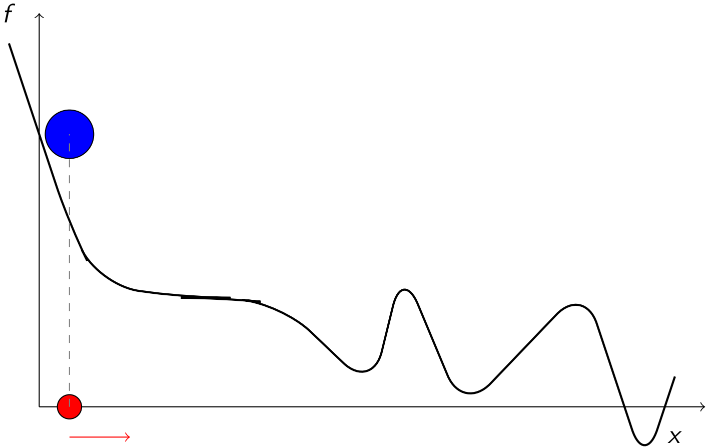
Descente du gradient

Problème : minimum mais local ....



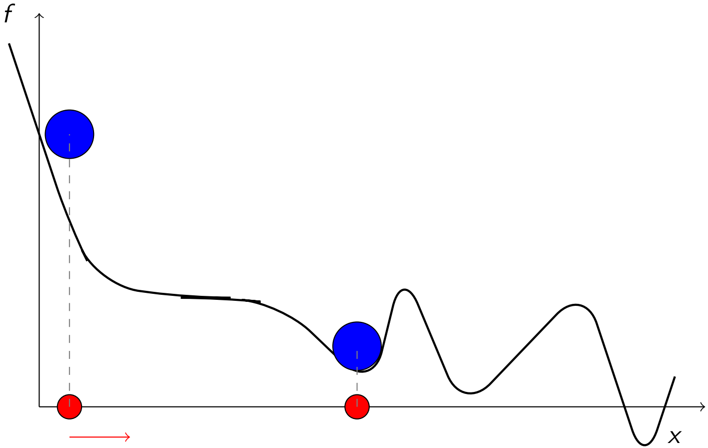
Descente du gradient

Problème : minimum mais local ...



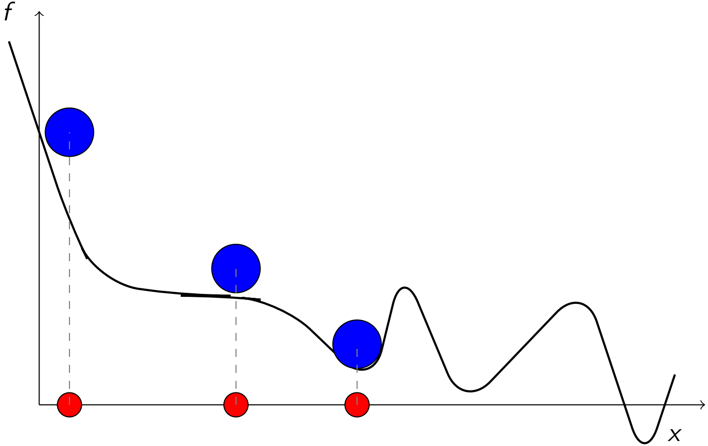
Descente du gradient

Problème : minimum mais local ....



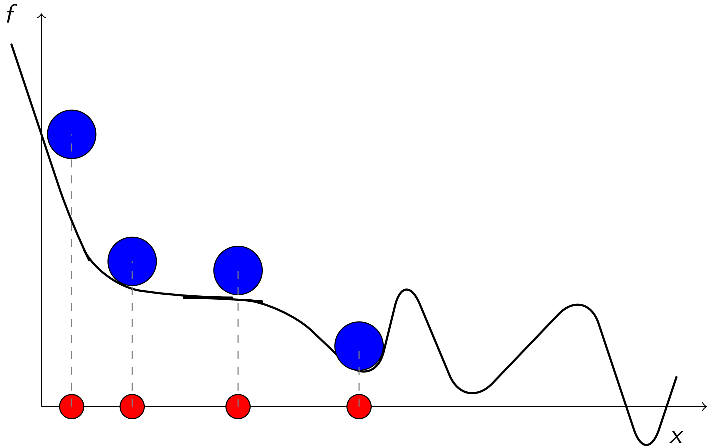
Descente du gradient

Problème : minimum mais local ....



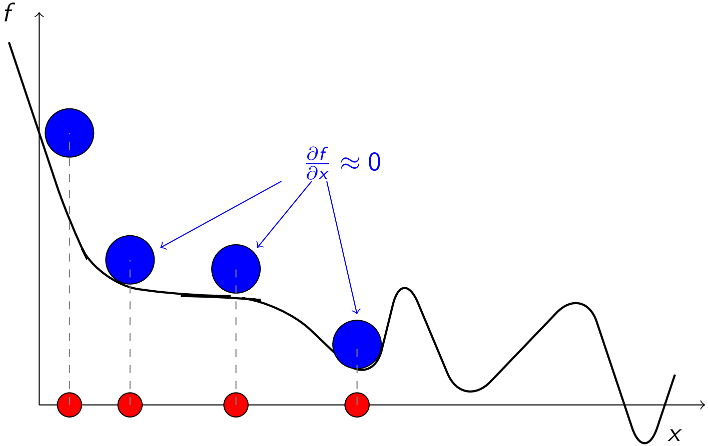
Descente du gradient

Problème : minimum mais local ....



Descente du gradient

Problème : minimum mais local ....

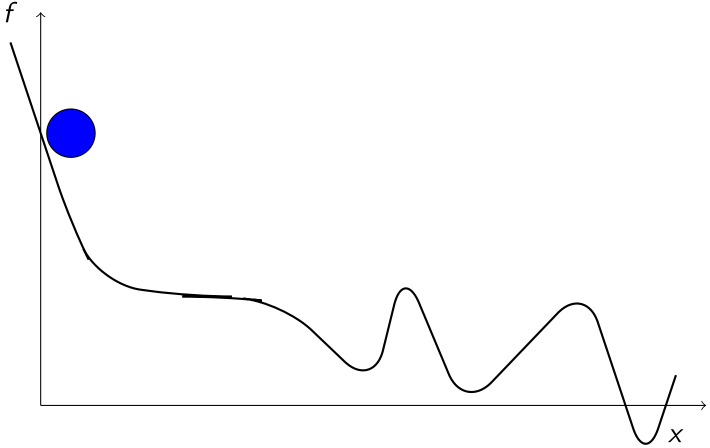


Descente du gradient

Problème : minimum mais local ....
Idée : s'inspirer de la physique et d'une balle qui dévale une pente

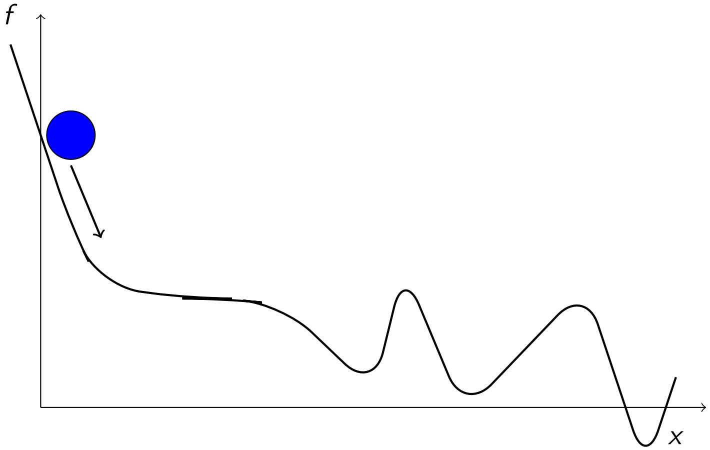
Descente du gradient

Méthode du Momentum :



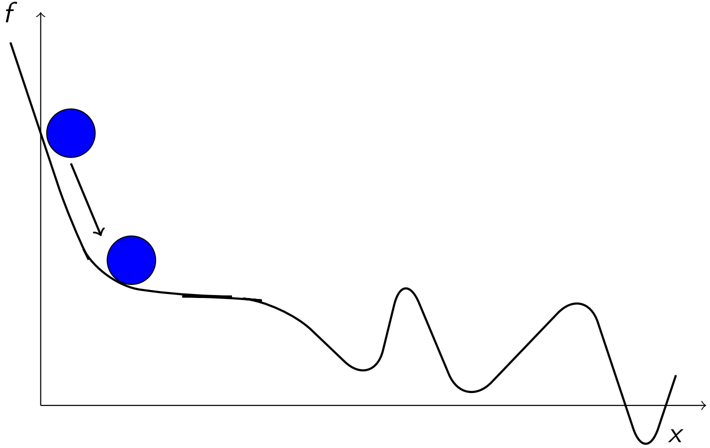
Descente du gradient

Méthode du Momentum :



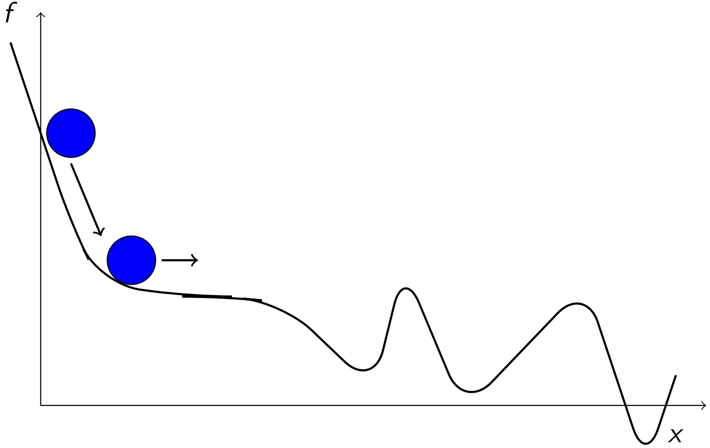
Descente du gradient

Méthode du Momentum :



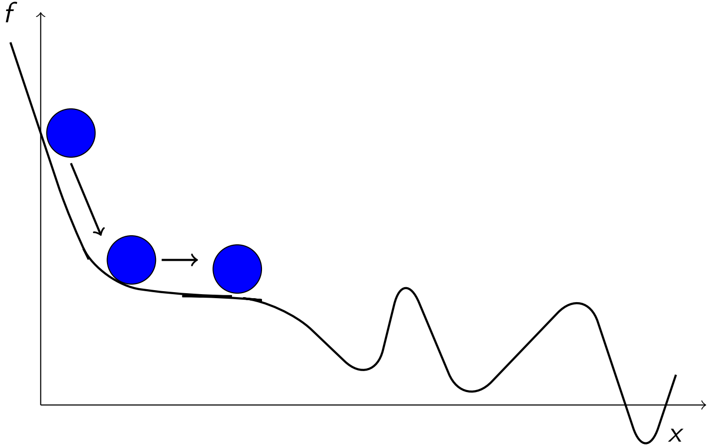
Descente du gradient

Méthode du Momentum :



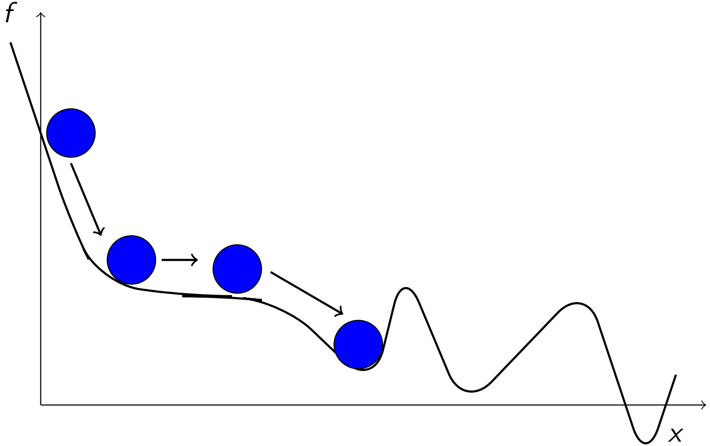
Descente du gradient

Méthode du Momentum :



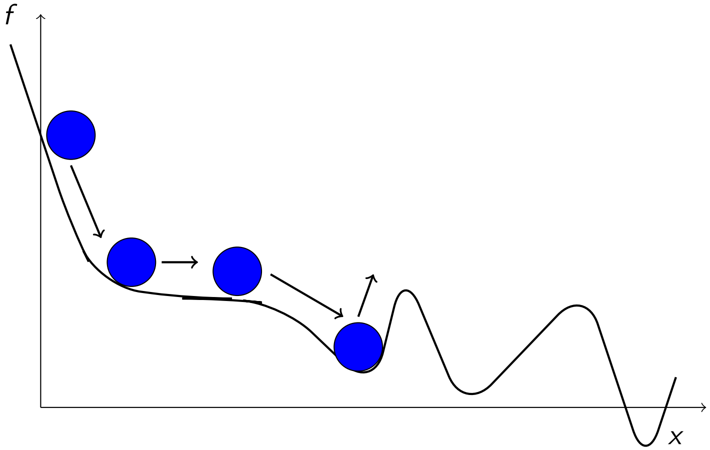
Descente du gradient

Méthode du Momentum :



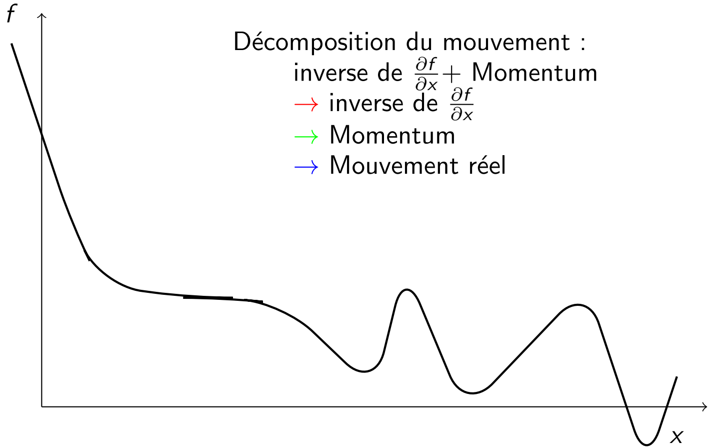
Descente du gradient

Méthode du Momentum :



Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

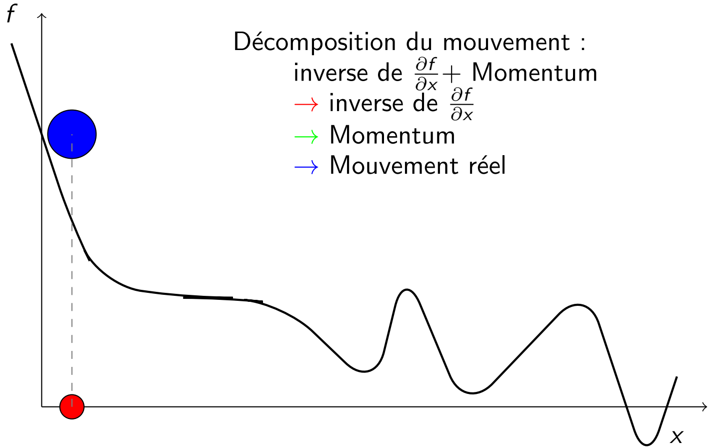
$\rightarrow$ inverse de $\frac{\partial f}{\partial x}$

$\rightarrow$ Momentum

$\rightarrow$ Mouvement réel

Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

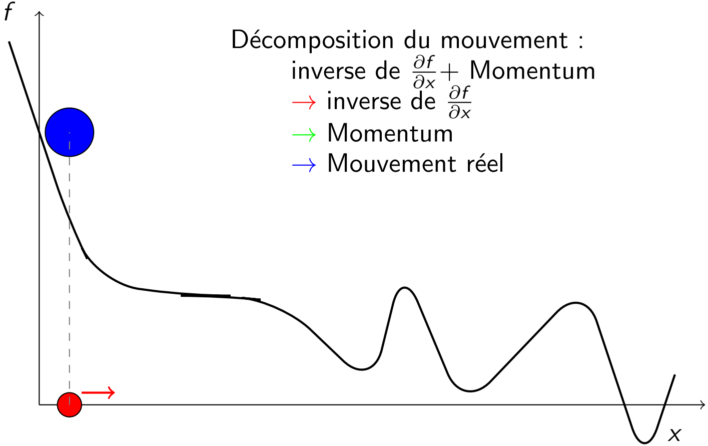
$\rightarrow$ inverse de $\frac{\partial f}{\partial x}$

$\rightarrow$ Momentum

$\rightarrow$ Mouvement réel

Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

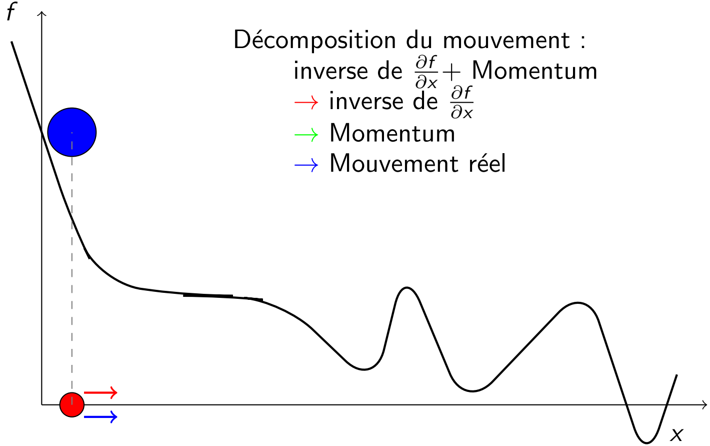
$\rightarrow$ inverse de $\frac{\partial f}{\partial x}$

$\rightarrow$ Momentum

$\rightarrow$ Mouvement réel

Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

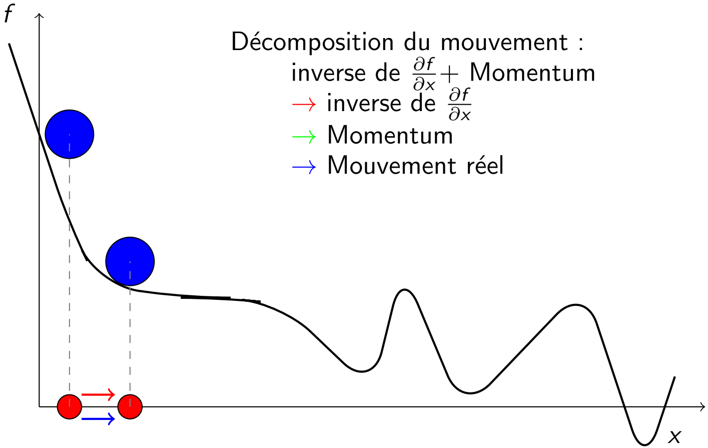
$\rightarrow$ inverse de $\frac{\partial f}{\partial x}$

$\rightarrow$ Momentum

$\rightarrow$ Mouvement réel

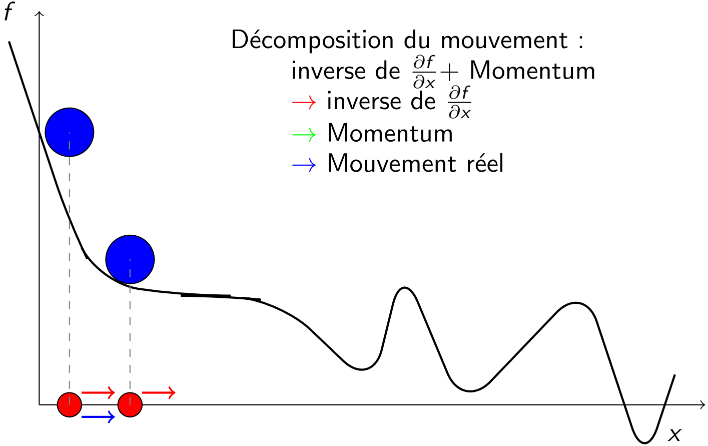
Descente du gradient

Problème : minimum mais local ....



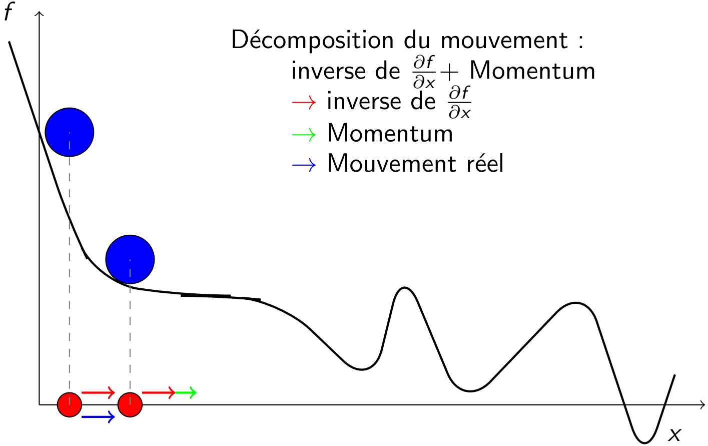
Descente du gradient

Problème : minimum mais local ....



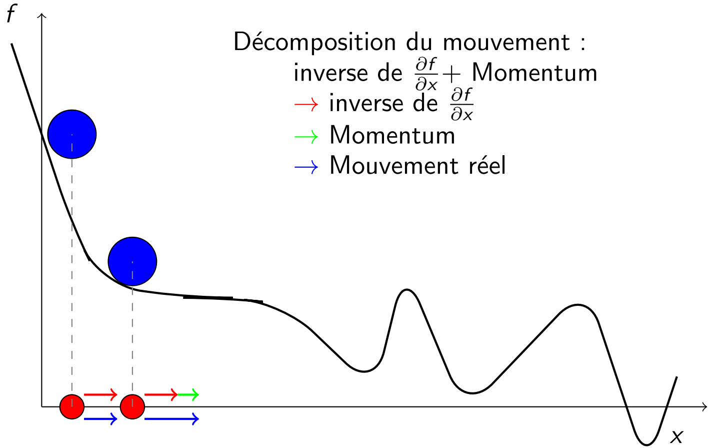
Descente du gradient

Problème : minimum mais local ....



Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

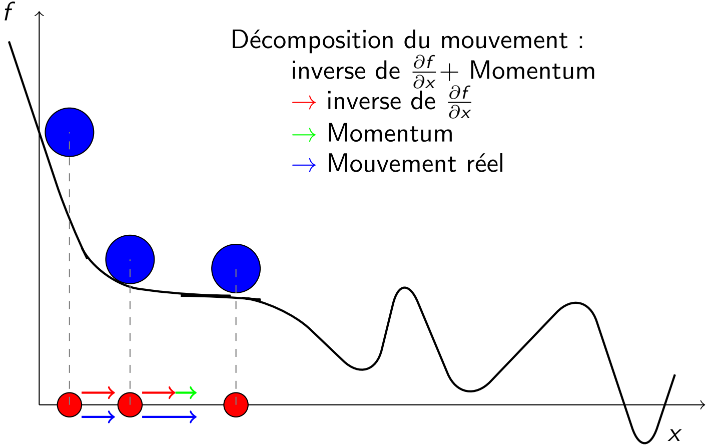
$\color{red}{\rightarrow}$ inverse de $\frac{\partial f}{\partial x}$

$\color{green}{\rightarrow}$ Momentum

$\color{blue}{\rightarrow}$ Mouvement réel

Descente du gradient

Problème : minimum mais local ....



Décomposition du mouvement :

inverse de $\frac{\partial f}{\partial x}$ + Momentum

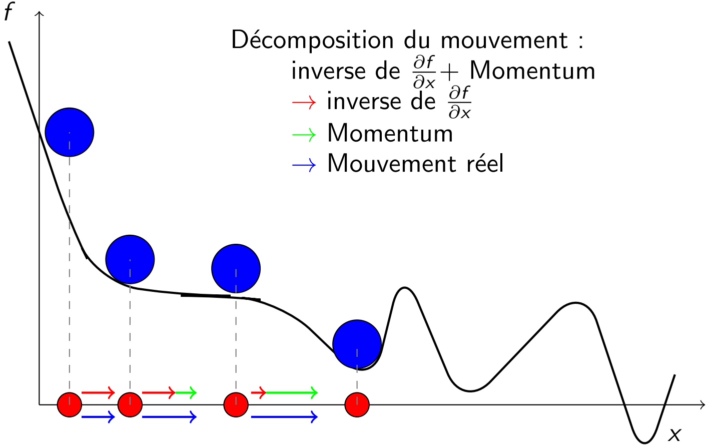
$\rightarrow$ inverse de $\frac{\partial f}{\partial x}$

$\rightarrow$ Momentum

$\rightarrow$ Mouvement réel

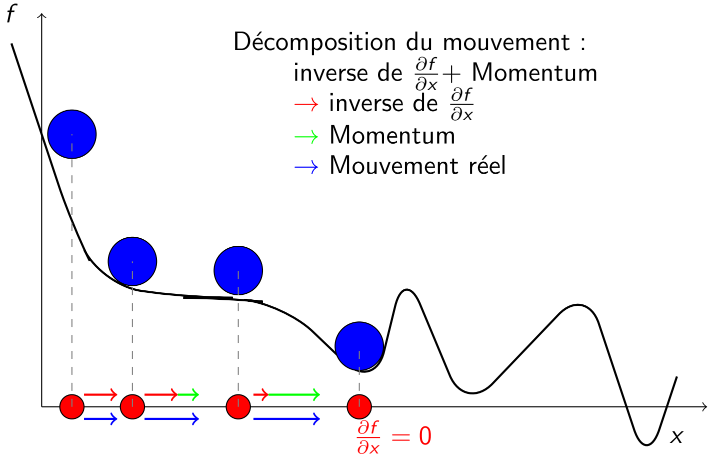
Descente du gradient

Problème : minimum mais local ....



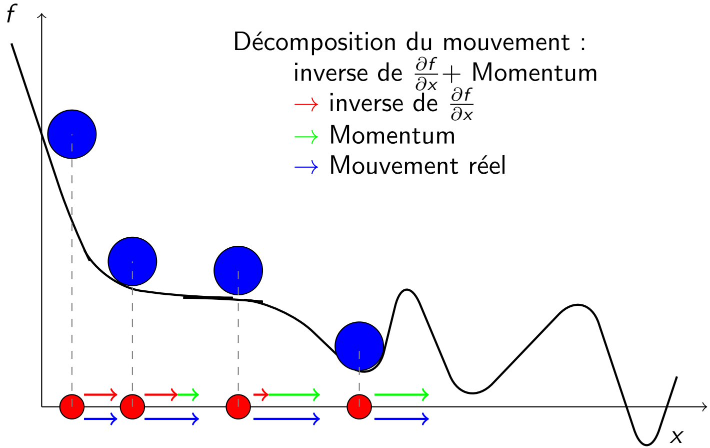
Descente du gradient

Problème : minimum mais local ....



Descente du gradient

Problème : minimum mais local ....



Descente du gradient

Problème : minimum mais local ....

Décomposition du mouvement :

La formule devient :

$\nu_t = \gamma \nu_{t-1} + \eta \frac{\partial f}{\partial x}$

$x \leftarrow x - \nu_t$