



Méthode du Gradient

Ce chapitre explore une classe importante d'algorithmes pour résoudre des problèmes d'optimisation sans contrainte. Le concept central est la direction de descente, qui consiste à trouver un chemin vers le minimum d'une fonction.

 par **Lamia Triqui**

Principe de la Méthode

L'objectif est de minimiser une fonction f . On commence par un point de départ arbitraire x_0 . Pour trouver le point suivant x , il faut s'approcher du minimum de f , c'est-à-dire $f(x) < f(x_0)$. On recherche alors x sous la forme: $x = x_0 + \rho d$, où d est un vecteur non nul de $\mathbb{R}^n$ et ρ est un réel strictement positif.

Direction de Descente

On cherche ρ et d tels que $f(x_0 + \rho d) < f(x_0)$. Si d existe, on l'appelle direction de descente.

Pas de Descente

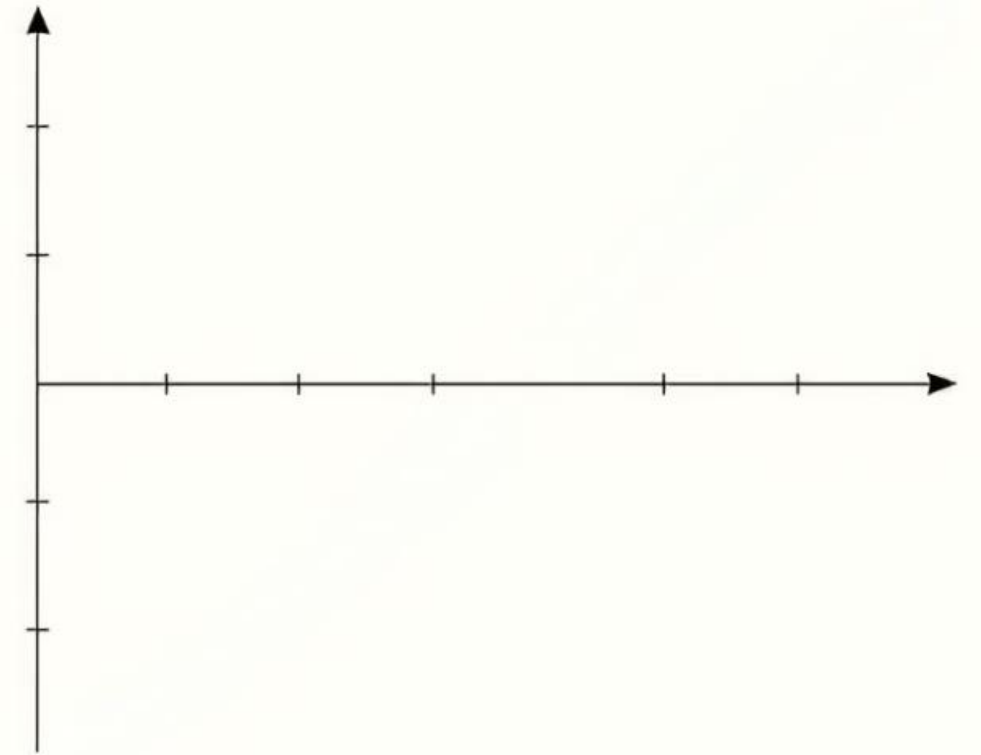
ρ représente le pas de descente, qui détermine la taille du déplacement dans la direction de descente.

Schéma Général

Le schéma général d'une méthode de descente s'écrit comme suit:

$$\begin{cases} x_0 \in \mathbb{R}^n \\ x_{k+1} = x_k + \rho_k d_k, \quad d_k \in \mathbb{R}^n \setminus \{0\}, \quad \rho_k \in \mathbb{R} \end{cases}$$

où ρ_k et d_k sont choisis de manière à ce que $f(x_k + \rho_k d_k) < f(x_k)$.



Trouver une Direction de Descente

Une idée est d'utiliser le développement de Taylor à l'ordre 2 de la fonction f entre deux itérés x_k et $x_{k+1} = x_k + \rho_k d_k$:

$$f(x_k + \rho_k d_k) = f(x_k) + \rho_k (\nabla f(x_k), d_k) + o(\rho_k d_k)$$

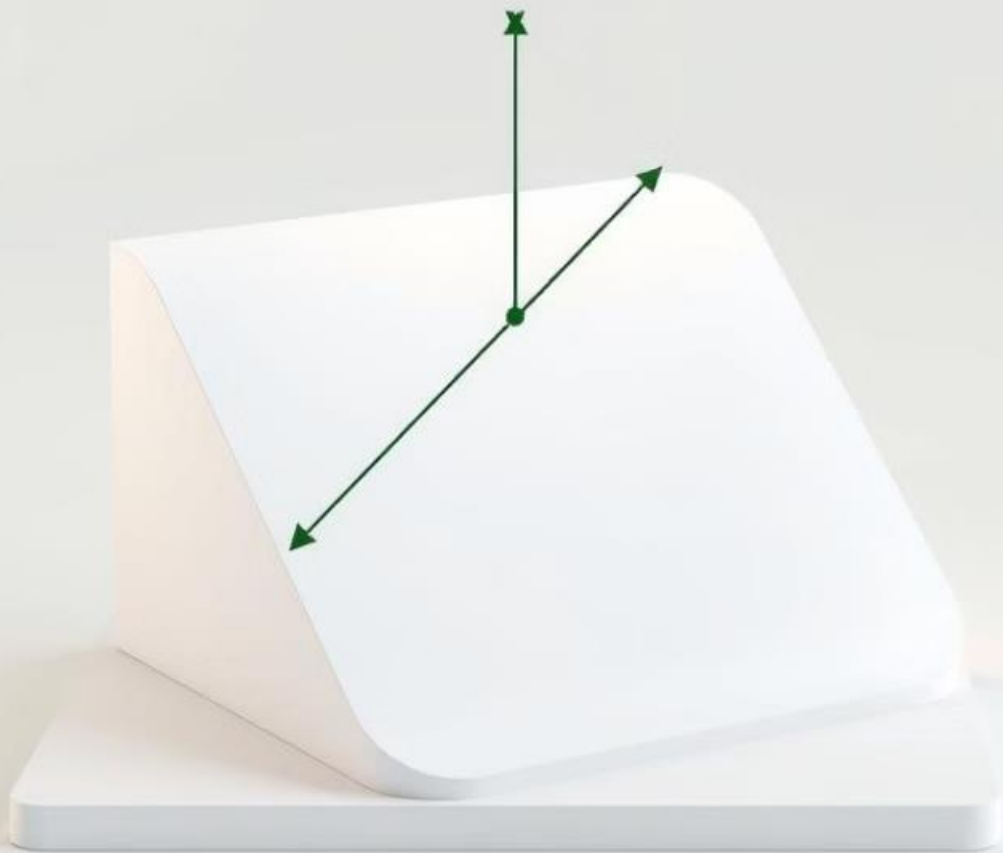
Pour que $f(x_k + \rho_k d_k) < f(x_k)$, on peut choisir:

1 *1. Algorithme du Gradient*

$$d_k = -\nabla f(x_k)$$

2 *2. Algorithme de Newton*

$$d_k = [H(x_k)]^{-1} \nabla f(x_k)$$



Méthode de Gradient

L'algorithme du gradient est un algorithme d'optimisation différentiable itératif qui procède par améliorations successives.

À chaque point courant, un déplacement est effectué dans la direction opposée au gradient, ce qui fait décroître la fonction. L'algorithme du gradient est également connu sous le nom d'algorithme de la plus forte pente (steepest descent) car le gradient est la pente de la fonction linéarisée au point courant et est donc, localement, sa plus forte pente.

Gradient Descent





Méthode de Gradient à Pas Fixe

On peut utiliser un pas fixe $\rho > 0$ pour toutes les itérations, ce qui donne la méthode de gradient simple:

$$\begin{cases} d_k = -\nabla f(x_k) \\ x_{k+1} = x_k + \rho d_k \end{cases}$$

Pour $f \in C'$, cette méthode converge si ρ est choisi suffisamment petit.

Le Choix de Pas

Le choix du pas ρ a un impact significatif sur la performance de la méthode de gradient.

Pas Optimal

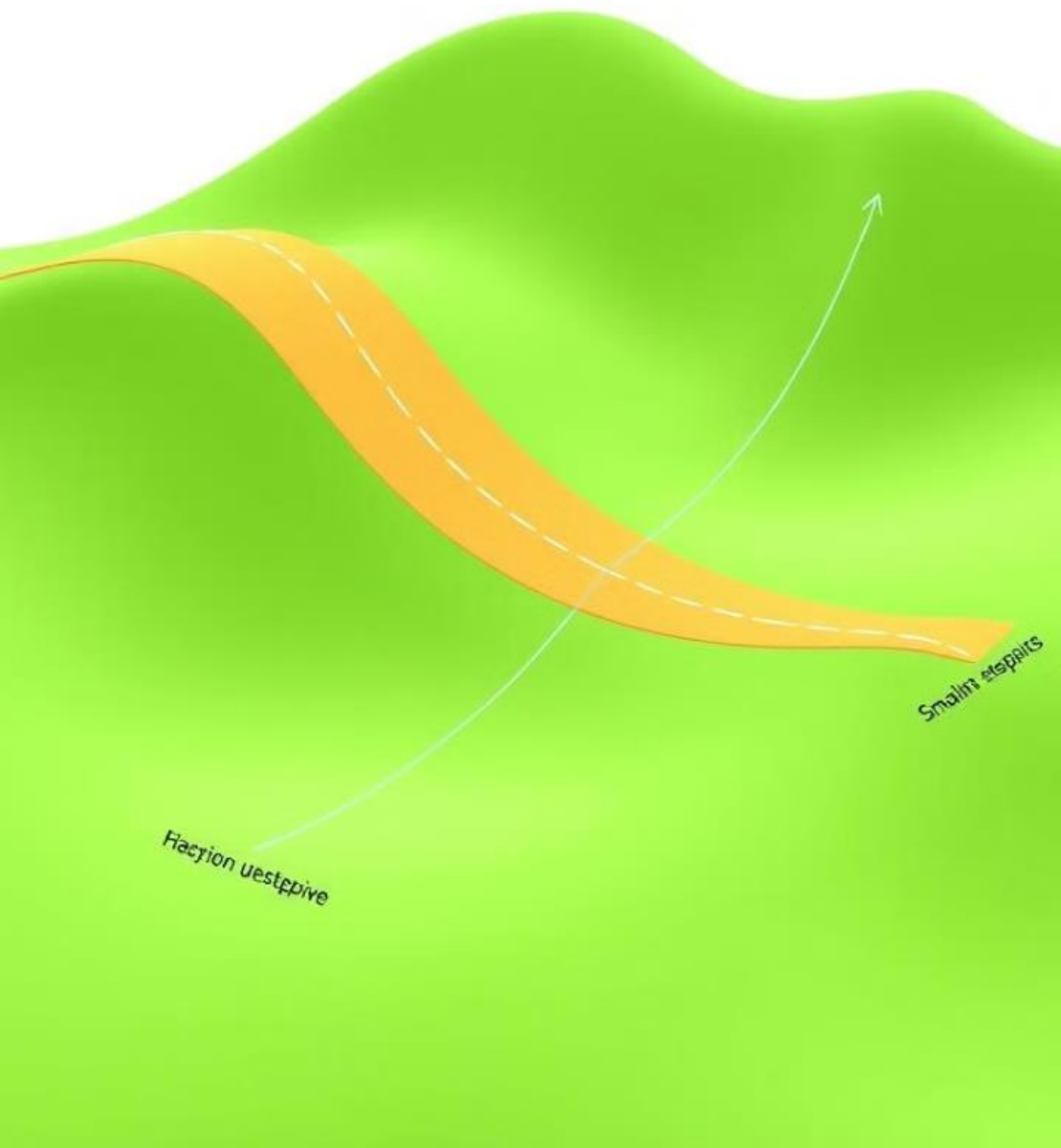
Un pas bien choisi donne des résultats similaires à la méthode de la plus profonde descente.

Pas Petit

Un pas plus petit atténue les zigzags des itérés mais augmente le nombre d'itérations.

Pas Grand

Un pas trop grand peut faire diverger la méthode.



Méthode de Gradient à Pas Optimal

Cette méthode consiste à calculer le pas ρ_k à chaque itération de manière à ce que la fonction de coût soit minimale le long de la direction de descente choisie. On calcule ainsi:

$$x_{k+1} = x_k - \rho_k \nabla f(x_k)$$

Pour illustrer la méthode, on considère le problème mathématique sans contrainte suivant: $\min f(x)$, avec $f(x) = 4x_1^2 + 6x_2^2 + 6x_1x_2 + 3x_1 + 4x_2 + 6$. On souhaite calculer le minimum directement (méthode analytique: $\nabla f(x) = 0$) et développer deux itérations des algorithmes de gradient à pas prédéterminé ($\rho = 0.1$) et à pas optimal pour $x_0 = (\dots)$.

Flogive take t dally, 1 pat this stesp the
is porte cfant patter Adjusted lotes the
fo' eagh, iitedtom or each bottnom.