Optimisation sans Contraintes

Formulation du Problème

Problème d'optimisation sans contraintes : Soit $\theta \in \mathbb{R}^d$ le vecteur des paramètres réels inconnus, et $J : \mathbb{R}^d \to \mathbb{R}$ une fonction à minimiser, supposée différentiable.

$$(P) \quad \min_{\theta \in \mathbb{R}^d} J(\theta)$$

Exemple :
$\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ et $J(\theta) = \theta_1^2 + \theta_2^2$

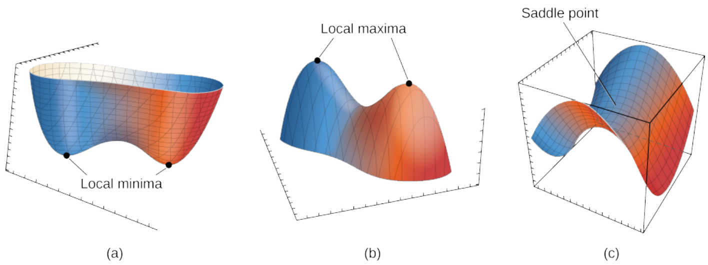
Conditions d'Optimalité

Solutions

- **Solution globale** : $J(\theta^*) \leq J(\theta)$ pour tout $\theta \in \mathrm{dom} J$.
- **Solution locale** : $J(\hat{\theta}) \leq J(\theta)$ pour tout θ tel que $\|\hat{\theta} - \theta\| \leq \epsilon$.

Conditions de Premier Ordre

Théorème : Si θ_0 est une solution, alors $\nabla J(\theta_0) = 0$. Tout point θ_0 qui satisfait cette condition est un point selle ou un point critique.



Conditions de Second Ordre

Théorème : Si θ_0 est un minimum local, alors $\nabla J(\theta_0) = 0$ et la matrice Hessienne $H(\theta_0)$ est définie positive.

- $H(\theta)$ est *définie positive* si ses valeurs propres sont positives.
- Si $H(\theta_0)$ est définie négative, alors θ_0 est un maximum local.
- Si $H(\theta)$ possède des valeurs propres positives et négatives, θ_0 est un point selle.

Algorithmes de Descente

Principe des Méthodes de Descente

À partir d'un point initial θ_0, une séquence de points $\{\theta_k\}$ est générée :

$$\theta_{k+1} = \theta_k + \alpha_k h_k$$

où h_k est une direction de descente et α_k le pas de descente.

Algorithme Général

1. Initialiser θ_0 aléatoirement ou avec des connaissances préalables du problème
2. Répéter :
   - Calculer une direction de descente h_k.
   - Rechercher un pas $\alpha_k > 0$ pour réduire J.
   - Mettre à jour : $\theta_{k+1} \leftarrow \theta_k + \alpha_k h_k$.
   - Arrêter lorsque $\|\nabla J(\theta_k)\| < \epsilon$.

Hugo Tondenier

Méthodes de Descente Principales

Méthode du Gradient

- **Direction de descente** : $h_k = -\nabla J(\theta_k)$.
- **Complexité** : $O(d)$.
- **Convergence** : linéaire.

Méthode de Newton

- **Approximation du second ordre** :

$$J(\theta + h) \approx J(\theta_k) + \nabla J(\theta_k)^\top h + \frac{1}{2} h^\top H(\theta_k) h$$

- Direction $h_k = -H(\theta_k)^{-1} \nabla J(\theta_k)$, avec $H(\theta_0)$ la matrice Hessienne(dérivées secondes).
- **Complexité** : $O(d^3)$.
- **Convergence** : quadratique.

Méthode Quasi-Newton

- **Direction de descente** : $h_k = -B(\theta_k)^{-1} \nabla J(\theta_k)$, avec $B(\theta_k)$ approximation de $H(\theta_k)$.
- **Complexité** : $O(d^2)$.
- **Convergence** : superlinéaire.

Recherche du Pas

Stratégies de Choix du Pas α_k

- **Pas fixe** : $\alpha_k = \alpha$ constant.
- **Pas optimal** : $\alpha_k = \arg\min_{\alpha > 0} J(\theta_k + \alpha h_k)$.
- **Pas variable** : adapté à chaque itération.

Règle d'Armijo (Pas Variable)

Rechercher un pas α_k satisfaisant :

$$J(\theta_k + \alpha_k h_k) \leq J(\theta_k) + c\alpha_k \nabla J(\theta_k)^\top h_k$$

où $c \in [10^{-5}, 10^{-1}]$.

Résumé

Méthode	Direction de descente	Complexité	Convergence
Gradient	$-\nabla J(\theta)$	$O(d)$	Linéaire
Quasi-Newton	$-B(\theta)^{-1} \nabla J(\theta)$	$O(d^2)$	Superlinéaire
Newton	$-H(\theta)^{-1} \nabla J(\theta)$	$O(d^3)$	Quadratique