Résumé du cours : Optimisation et Analyse convexe

M2 Informatique – Parcours IAAA
Aix-Marseille Université
Enseignant : Thomas Schatz
Étudiants : Atchory Rameaux MEL & Joe HUBI

Table des matières

Introduction

Dans ce résumé de cours, nous allons couvrir :

- l'optimisation lisse (avec et sans contraintes),
- la convexité et ses propriétés,
- la dualité lagrangienne,
- les conditions d'optimalité et d'existence du minimum,
- les algorithmes d'optimisation.

1 Convexité

Un ensemble X est dit **convexe** si, pour tout $x, y \in X$, le segment reliant x et y est entièrement contenu dans X :

$$[x, y] = \{tx + (1-t)y \mid t \in [0, 1]\} \subseteq X.$$

Exemple intuitif : pour représenter la convexité, on peut tirer un trait entre deux points dans l'ensemble. Si tous les points de ce trait appartiennent aussi à l'ensemble, celui-ci est convexe.

Une fonction $f : X \to \mathbb{R}$ est **convexe** si :

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y), \quad \forall x, y \in X, \ \forall t \in [0, 1].$$

Si l'inégalité est stricte, f est **strictement convexe**.

Propriétés classiques

- Les fonctions linéaires sont convexes.
- Les combinaisons linéaires positives de fonctions convexes sont convexes.
- Le maximum (point par point) d'un ensemble de fonctions convexes est convexe.

Caractérisation par la Hessienne

Si f est deux fois différentiable (C^2), alors :

- f est convexe si et seulement si sa Hessienne $\nabla^2 f$ est semi-définie positive.
- f est strictement convexe si $\nabla^2 f$ est définie positive.

Cela permet de vérifier la convexité par des calculs analytiques, ce qui est essentiel pour des fonctions de grande dimension.

2 Conditions d'optimalité

Premier ordre

Si x^* est un minimum local et f est C^1, alors :

$$\nabla f(x^*) = 0.$$

Cela exprime que la pente est nulle au point optimal. Intuitivement, au minimum, il n'y a pas de direction de descente immédiate.

Second ordre

Si x^* est un minimum local et f est C^2, alors :

$$\nabla f(x^*) = 0 \quad \text{et} \quad \nabla^2 f(x^*) \succeq 0.$$

Si $\nabla^2 f(x^*)$ est définie positive, x^* est un minimum strict. Ici, la Hessienne joue le rôle d'une matrice de courbure : une matrice définie positive signifie que la fonction est « courbée vers le haut » dans toutes les directions.

3 Existence et unicité du minimum

- Si f est strictement convexe sur un ensemble convexe X, elle admet **au plus un minimum global**.
- Si f est convexe et coercive sur un ensemble fermé, elle admet **au moins un minimum global**.
- Tout minimum local d'une fonction convexe est aussi un minimum global. C'est une propriété clé de l'optimisation convexe qui simplifie grandement la recherche de solutions.

4 Dualité lagrangienne

Considérons un problème d'optimisation :

$$\min_{x \in X} f(x) \quad \text{sous les contraintes } e_i(x) = 0,\ c_j(x) \leq 0.$$

On définit le **Lagrangien** :

$$L(x, \mu, \lambda) = f(x) + \mu^T e(x) + \lambda^T c(x),$$

avec $\mu \in \mathbb{R}^r$ et $\lambda \in \mathbb{R}^s,\ \lambda \geq 0$.

La fonction duale est :

$$q(\mu, \lambda) = \inf_{x \in X} L(x, \mu, \lambda).$$

Le **problème dual** consiste à maximiser $q(\mu, \lambda)$ sous les contraintes $\lambda \geq 0$.

Lien entre primal et dual

- La solution du problème dual est **toujours inférieure ou égale** à celle du problème primal (**dualité faible**).
- Quand les deux coïncident ($q^* = f^*$), on parle de **dualité forte**. Cela survient dans des conditions de régularité (par ex. conditions de Slater).
- La **relaxation convexe** est une technique qui consiste à transformer un problème non convexe en un problème convexe (souvent en élargissant l'ensemble de contraintes), afin de bénéficier de la puissance des méthodes de dualité.

5 Algorithmes d'optimisation

Un point essentiel du cours concerne les méthodes pratiques de recherche de minimum.

Descente de gradient (pas fixe)

À partir d'un point initial x_0, on met à jour :

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

où $\alpha > 0$ est un pas fixe. Ce schéma converge si le pas est choisi correctement.

Descente de gradient avec backtracking (Armijo)

On choisit le pas $\alpha_k = \beta^{m_k} s$ avec $0 < \beta < 1$, de sorte que :

$$f(x_k) - f(x_{k+1}) \geq \gamma \alpha_k \|\nabla f(x_k)\|^2,$$

où $\gamma \in (0, 1)$. Cela garantit une diminution suffisante de la fonction à chaque itération.