Méthodes d'Optimisation Numérique Appliquées au Dataset California Housing

Nom et Prénom:Fatimetou/limam/Abeid
C16698

Université de NKTT

MiniProjet 2026

Plan

1. Introduction
2. Dataset California Housing
3. Phase 1 — Gradient déterministe
4. Phase 2 — Gradient stochastique
5. Phase 3 — Méthodes proximales
6. Résultats expérimentaux
7. Conclusion

Objectifs du projet

- Étudier les méthodes d'optimisation vues aux Chapitres 1 à 4
- Comparer :
  - Gradient déterministe
  - Gradient stochastique
  - Méthodes adaptatives
  - Méthodes proximales
- Application sur des données réelles
- Analyse des courbes Perte vs Temps

Présentation du dataset

- Recensement Californie – 1990
- Données agrégées par district
- Variables socio-économiques :
  - Revenu médian
  - Âge des logements
  - Population
  - Nombre de pièces
- Cible : valeur médiane des maisons

Prétraitement

- Transformation en classification binaire :

$$y_i = \begin{cases} +1 & \text{si MedHouseVal} > \text{médiane} \\ -1 & \text{sinon} \end{cases}$$

- Standardisation des variables
- Dimension : $n \times d$

Fonction objectif

$$F(w) = \frac{1}{n}\sum_{i=1}^{n} \log\left(1 + e^{-y_i x_i^T w}\right) + \frac{\lambda}{2}\|w\|^2$$

- Perte logistique
- Régularisation L2 (Ridge)

Propriétés théoriques

- $F \in C^2(\mathbb{R}^d)$
- Fonction convexe
- λ-fortement convexe :
$$\nabla^2 F(w) \succeq \lambda I_d$$
- Existence et unicité du minimiseur

Gradient et Lipschitzianité

$$\nabla F(w) = -\frac{1}{n}\sum_{i=1}^{n} \frac{y_i x_i}{1 + e^{y_i x_i^T w}} + \lambda w$$

$$L = \frac{1}{4n}\|X\|_2^2 + \lambda$$

- Pas de gradient : $\alpha \leq \frac{1}{L}$
- Convergence linéaire

Pourquoi SGD ?

- Gradient complet : $\mathcal{O}(nd)$
- SGD : $\mathcal{O}(d)$
- Adapté aux grands jeux de données
- Mise à jour aléatoire

Algorithme SGD

$$w_{k+1} = w_k - \alpha_k \left(-\frac{y_{i_k} x_{i_k}}{1 + e^{y_{i_k} x_{i_k}^T w_k}} + \lambda w_k \right)$$

$$\alpha_k = \frac{\alpha_0}{1 + k}$$

RMSProp et Adam

- RMSProp :
  - Normalisation adaptative
  - Réduction de variance
- Adam :
  - Momentum + RMSProp
  - Très rapide au début
  - Stabilité accrue

Problème non lisse

$$\Phi(w) = f(w) + \lambda \|w\|_1$$

- Norme L_1 non dérivable
- Utilisation du proximal
- Promotion de la sparsité

Opérateur proximal

$$\text{prox}_{\lambda\|\cdot\|_1}(v) = \text{soft-thresholding}(v)$$

$$(v_j - \lambda)_+ - (-v_j - \lambda)_+$$

ISTA et FISTA

- ISTA :

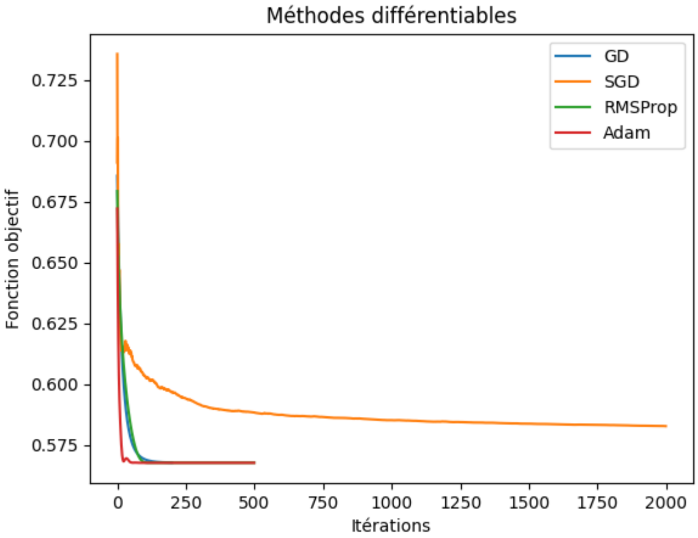
$$\mathcal{O}(1/k)$$

- FISTA (Nesterov) :

$$\mathcal{O}(1/k^2)$$

- Accélération significative observée

Courbes Perte vs Temps



Analyse

- Adam converge le plus vite initialement
- SGD plus bruité mais peu coûteux
- FISTA domine ISTA
- Impact clair du momentum

Conclusion

- Validation théorique des chapitres 1 à 4
- Comparaison complète des méthodes
- Données réelles et résultats observables
- Bon compromis biais–variance via L1