Polytechnique Montréal
Gerad

Pca-Mads - Réduction de dimension dans Mads

Montréal, 2 avril 2020

Romain Vanden Bulcke

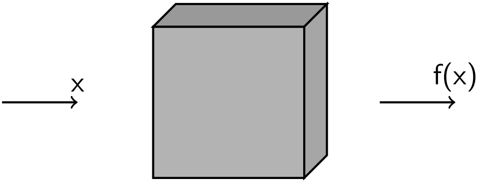
Directeurs : Charles Audet
Sébastien Le Digabel

Superviseur : Miguel Diago Martinez

Introduction

Contexte

- Optimisation sans dérivées / boîtes noires
- Problèmes en grande dimension
- Algorithme séquentiel
- MADS



Introduction

Contexte

- Optimisation sans dérivées / boîtes noires
- Problèmes en grande dimension
- Algorithme séquentiel
- Mads

Idée

- Exploiter les informations connues de la boîte noire $\rightarrow$ cache
- Identifier les combinaisons de variables les plus influentes
- Suite des projets d'Imen et Nadir

Plan

Réduction de dimension en Optimisation Bayesienne I

Optimisation Baysesienne

- Méthode d'optimisation basée sur des modèles aléatoires
- Distribution *a priori* sur F, $p(F)$
- Données D et modèle de vraisemblance $p(D|F)$
- Distribution *a posteriori* $p(F|D) \propto p(D|F)p(F)$
- Construction d'une fonction d'acquisition u, facile à optimiser (probabilité d'amélioration, espérance d'amélioration, ...)
- Mise à jour des distributions et des modèles

Réduction de dimension en Optimisation Bayesienne II

Dimension effective [5]

Une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}$ a une dimension effective $d \leq n$ s'il existe un sous-espace linéaire T de dimension d tel que pour tout $x_\top \in T$ et $x_\perp \in T^\perp$,

$$f(x_\top + x_\perp) = f(x_\top),$$

où d est le plus petit entier satisfaisant cette relation.

Exemple [5]

La fonction $f(x) = \sin^2(x_1 - x_2 - 0.5)$
Sous-espace actif $(x_1, x_2) = (1, -1)^\top y, y \in \mathbb{R}$
Ensemble de ses minima $\{(1,1)^\top t - (0, 0.5 + \pi k)^\top : t \in \mathbb{R}, k \in \mathbb{Z}\}$.

Réduction de dimension en Optimisation Bayesienne III

Idée

- Problème original :

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{1}$$

- Construire un nouveau problème :

$$\min_{y \in \mathcal{Y} \subset \mathbb{R}^p} \quad f(Ay) \tag{2}$$

  où $A \in \mathbb{R}^{n \times p}$, $p \ll n$
- A est une matrice aléatoire qui définit un sous-espace actif
- Répéter avec plusieurs matrices aléatoires
- Choix de A et du domaine $\mathcal{Y}$ critiques [3, 4, 5]

Réduction de dimension dans MADS

Algorithme STATS-MADS

- Mémoire d'Imen [1, 6]
- Identifier les variables les plus importantes via une analyse de variance, pour fixer les autres variables
- Alterner entre MADS avec toutes les variables et MADS avec les variables les moins influentes fixées

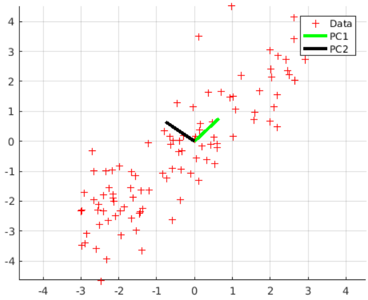
Analyse de sensibilité dans PSD-MADS

- Mémoire de Nadir [2]
- Version parallèle de MADS
- Utiliser la même analyse de sensibilité pour attribuer des variables aux processeurs

Plan

Analyse en composante principale

- Transformation de n variables aléatoires corrélées entre elles X en n variables aléatoires $\alpha_i^\top X$, $i = 1, 2, \dots, n$, variables non corrélées
- Les nouvelles variables $\alpha_i^\top X$ sont les composantes principales
- Les $p \ll n$ premières composantes principales reprennent la majeure partie de l'information



Calcul des composantes principales

- Vecteurs propres de la matrice de covariance Σ (ou corrélation) des variables aléatoires X
- Possibilité de calcul à partir d'un nuage de N points $D \in \mathbb{R}^{N\times(n+1)}$
- Possibilité d'approximer la matrice de covariance Σ par

$$S_{ij} = \frac{1}{N-1}\sum_{l=1}^{N}(D_{li} - \bar{D}_i)(D_{lj} - \bar{D}_j),$$

où $\bar{D}_j = \frac{1}{N}\sum_{i=1}^{N} D_{ij}$ est la moyenne des réalisations de la j^e variable aléatoire.

Algorithme PCA-MADS I

Contexte

- Problèmes en grandes dimensions
- Pas de parallélisme

Structure et idée

- Structure de MADS normale, avec une étape de recherche et une étape de sonde
- Idée : optimisation d'un problème en petite dimension à l'étape de recherche

Algorithme PCA-MADS II

Etape de recherche

- Reprendre un ensemble de points de la cache
- Appliquer une analyse en composante principale sur ce nuage de points
- Identifier les combinaisons de variables qui semblent influencer fortement l'objectif dans une matrice $P \in \mathbb{R}^{n \times p}$, $p \leq n$
- Lancer une instance de MADS sur le problème

$$\min_{y \in \mathbb{R}^p} f(x_k + Py)$$

- Projeter les points à évaluer sur le treillis

Plan

Suite de problèmes

Fonctions Rosenbrock

- $$f(x) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2$$
- Petites dimensions : $2, 3, \ldots, 20$
- Grandes dimensions : $100, 200, \ldots, 500$

Suite bbob de coco

- Suite de 24 fonctions
- Dimensions : $2, 3, 5, 10, 20, 40$
- 16 instances par fonctions, dans chaque dimension

Méthodes

Méthodes

- `Pca-Mads` : implémentation personnelle de l'algorithme, avec la dimension du sous-problème $p = \lceil \frac{n}{5} \rceil$, en Matlab
- `Mads` : implémentation personnelle en Matlab, la même que `Pca-Mads` sans étape de recherche
- `Nomad-default` : Nomad 3.9.1 avec tous les paramètres mis à leur valeur par défaut
- `Nomad-basic` : Nomad 3.9.1 avec toutes les options désactivées

Résultats : Rosenbrock en petites dimensions

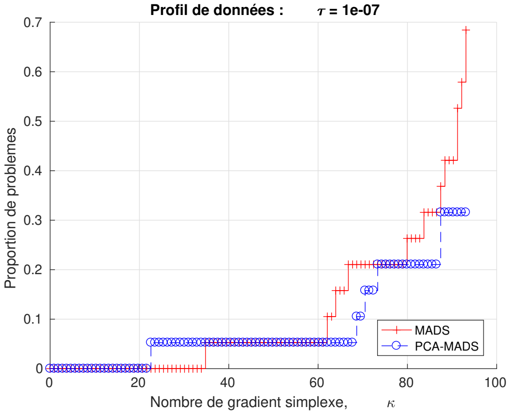


Figure: Profil de données pour l'ensemble de fonctions Rosenbrock en petites dimensions, avec un budget de $100n$

Résultats : Rosenbrock en grandes dimensions

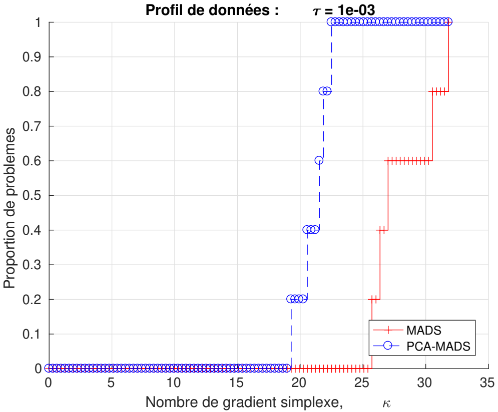


Figure: Profil de données pour l'ensemble de fonctions Rosenbrock en grandes dimensions, avec un budget de $50n$

Résultats : coco en petites dimensions

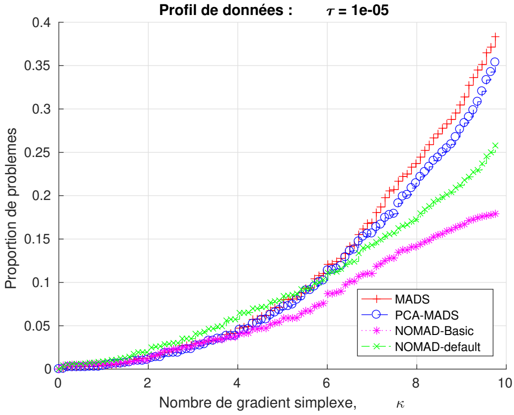


Figure: Profil de données pour l'ensemble de fonctions bbob de coco en petites dimensions, avec un budget de $10n$

Tests sur le paramètre p

Tests sur la dimension réduite

Différentes possibilités pour choisir la dimension du problème réduit p

- $p = n/5$
- $p = n/10$
- $p = n/20$
- Stratégie avec p variable qui divise les composantes principales en deux groupes, à l'aide de l'algorithme *k-means*

Profil pour le paramètre p II

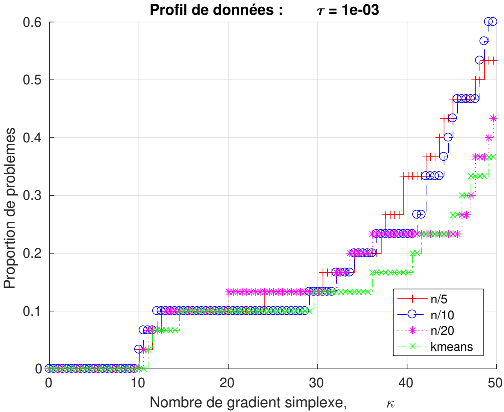


Figure: Comparaison de différentes stratégies de réduction de dimension dans `Pca-Mads` sur la suite *bbob-largescale* avec un budget de $50n$

Profil pour le paramètre p III

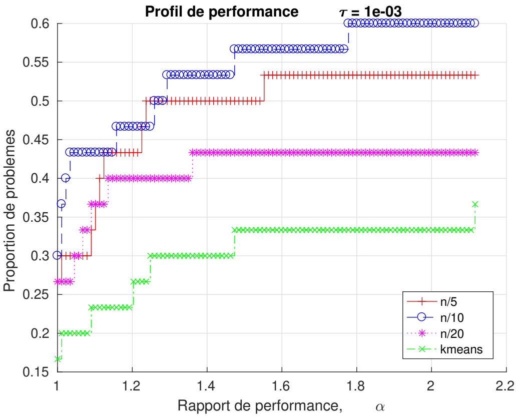


Figure: Comparaison de différentes stratégies de réduction de dimension dans `Pca-Mads` sur la suite *bbob-largescale* avec un budget de $50n$

Plan

Conclusion et suite

Conclusion

- PCA-MADS pas efficace pour des problèmes de petites tailles mais semblent intéressants en plus grandes dimensions
- Plusieurs paramètres peuvent influencer les performances de l'algorithme
- Pour l'instant, assez peu de résultats en grandes dimensions

Suite

- Etudier l'influence de chaque paramètre
- Faire plus de tests en grandes dimensions
- Comparer avec d'autres algorithmes en grande dimension
- Tester l'algorithme sur des boîtes noires réelles

Références I

L. Adjengue, C. Audet, and I. Ben Yahia B.
A variance-based method to rank input variables of the mesh adaptive direct search algorithm.
Optimization Letters, 8(5):1599–1610, 2014.

N. Amaioua.
Modèles quadratiques et décomposition parallèle pour l'optimisation sans dérivées.
PhD thesis, École Polytechnique de Montréal, 2018.

M. Binois, D. Ginsbourger, and O. Roustant.
On the choice of the low-dimensional domain for global optimization via random embeddings.
Journal of Global Optimization, 2019.

Références II

M. G. B. Blum, M. A. Nunes, D. Prangle, S. A. Sisson, et al.
A comparative review of dimension reduction methods in approximate Bayesian computation.
Statistical Science, 28(2):189–208, 2013.

C. Cartis and A. Otemissov.
A dimensionality reduction technique for unconstrained global optimization of functions with low effective dimensionality.
2020.

I. Ben Yahia.
Identification statistique de variables importantes pour l'optimisation de boîtes noires.
PhD thesis, École Polytechnique de Montréal, 2012.