

# Méthodes d'optimisation sans gradient pour des problèmes inverses en ingénierie pétrolière

**Laurent DUMAS**

avec F. Delbos, D. Ding (IFPEN), E. Echague, B. Marteau  
(doctorants IFPEN)

Laboratoire de Mathématiques de Versailles (LMV)  
Université de Versailles Saint Quentin en Yvelines (UVSQ)

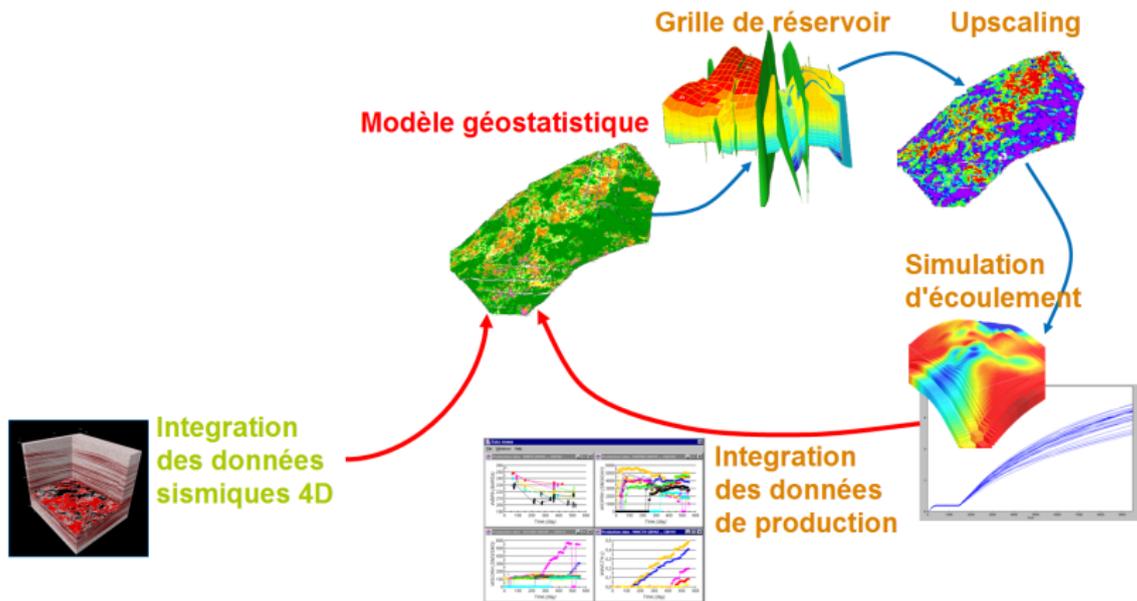
18 avril 2013

# Plan de l'exposé

- 1 Problèmes inverses en ingénierie pétrolière
- 2 Une nouvelle méthode de type surface de réponse
  - Principe des sparse grids
  - Méthode d'optimisation associée
  - Résultats numériques
- 3 Une nouvelle méthode de type région de confiance
  - Principe général
  - Adaptation au cas des fonctions partiellement séparables
  - Un résultat de convergence
  - Résultats numériques

- 1 Problèmes inverses en ingénierie pétrolière
- 2 Une nouvelle méthode de type surface de réponse
- 3 Une nouvelle méthode de type région de confiance

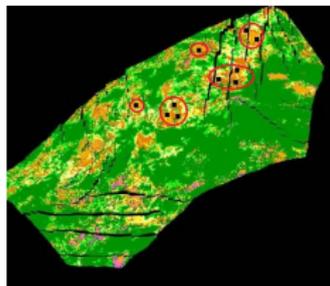
# Le calage d'historique



## Caractéristiques du problème inverse

- L'évaluation de la fonction objectif  $(x_1, \dots, x_n) \mapsto f(x_1, \dots, x_n)$  à optimiser nécessite la simulation d'écoulements complexes dans un modèle de réservoir, processus très coûteux.
- Les caractéristiques principales du problème à résoudre sont :
  - Dépendance en de nombreux paramètres
  - Evaluation coûteuse de la fonction objectif
  - Fonction objectif partiellement séparable

## Séparabilité partielle de la fonction objectif



La fonction objectif peut s'écrire sous la forme :

$$\begin{aligned} f(x_1, \dots, x_n) &= \frac{1}{2} \sum_{i=1}^{n_1} \frac{\omega_i^P}{N_P(i)} \sum_{j=1}^{N_P(i)} \left( \frac{P_{i,j}^{obs}(x) - P_{i,j}^{sim}(x)}{\sigma_{i,j}^P} \right)^2 \\ &= \sum_{i=1}^p f_i(x_1, \dots, x_n) \\ &\approx \sum_{i=1}^p f_i(x_{1_i}, \dots, x_{n_i}) \end{aligned}$$

avec  $\forall i, n_i \leq n$

# Objectifs

- Le problème général à résoudre peut se formuler de la manière suivante : *obtenir le meilleur calage d'historique à partir d'un nombre maximal fixé de simulations d'écoulement.*
- Plusieurs questions se posent alors :
  - Paramétrisation : *déterminer les paramètres les plus pertinents.*
  - Initialisation du problème : *proposer la meilleure initialisation possible.*
  - Optimisation : *exploiter au mieux les connaissances sur la fonction.*

## Comparaisons des méthodes d'optimisation

- De nombreux tests ont été réalisés dans la littérature et à l'IFPEN pour le domaine pétrolier.
- Il en ressort que les méthodes sans dérivées sont en général à privilégier en l'absence d'estimation simple du gradient.
- Quatre grandes familles de méthodes d'optimisation sans gradient sont disponibles :
  - Méthodes directes de type simplexe (Nelder Mead)
  - Méthodes de type évolutionnaires (recuit simulé, algorithmes génétiques, CMA-ES)
  - Méthodes de type surfaces de réponse (RBF, krigeage)
  - Méthodes de type région de confiance (SQP, NEWUOA)

- 1 Problèmes inverses en ingénierie pétrolière
- 2 Une nouvelle méthode de type surface de réponse
  - Principe des sparse grids
  - Méthode d'optimisation associée
  - Résultats numériques
- 3 Une nouvelle méthode de type région de confiance
  - Principe général
  - Adaptation au cas des fonctions partiellement séparables
  - Un résultat de convergence
  - Résultats numériques

## Principe des méthodes de type surface de réponse

- Une méthode de type surface de réponse consiste à construire de manière itérative un modèle approché de la fonction objectif.
- Les surfaces de réponse classiques consistent en des combinaisons linéaires de fonctions de base (méthode RBF, krigeage).
- On choisit d'utiliser ici une méthode d'interpolation, appelée sparse grid, appliquée initialement pour le calcul d'intégrale en grande dimension ou l'étude de sensibilité.

# Construction hiérarchique en dimension 1

On cherche ici à interpoler la fonction

$$f : [0, 1] \rightarrow \mathbb{R}.$$

Pour  $i \in \mathbb{N}$ , on note  $X^i$  l'ensemble des points de Chebychev sur  $[0, 1]$  de niveau  $i$ . Plus précisément,  $x_j^i \in X^i$  est tel que :

$$x_j^i = \begin{cases} \frac{1}{2} & \text{pour } j = 1 \text{ et } i = 1 \\ \frac{1}{2}(-\cos(\pi \frac{j-1}{n_i-1})) + 1 & \text{pour } j = 1, \dots, 2^{i-1} + 1 \text{ et } i > 1. \end{cases}$$

# Construction hiérarchique en dimension 1

Soit  $a_j^i$  la famille des polynômes de Lagrange associés à chaque  $x_j^i \in X^i$  :

$$a_j^i(x) = \begin{cases} 1 & \text{pour } j = 1 \text{ et } i = 1 \\ \prod_{k=1, k \neq j}^{n_i} \frac{x - x_k^i}{x_j^i - x_k^i} & \text{pour } j = 1, \dots, n_i \text{ et } n_i > 1. \end{cases}$$

Le modèle d'interpolation de niveau  $i$  de  $f$ , noté  $m_i(f)$ , s'écrit :

$$m_i(f) = \sum_{x_j^i \in X^i} f(x_j^i) a_j^i. \quad (1)$$

## Construction hiérarchique en dimension 1

Soit  $\Delta^i$  l'écart entre les modèles de niveau  $i$  et  $i - 1$  :

$$\Delta^i = m_i(f) - m_{i-1}(f) = \sum_{x_j^i \in X^i} (f(x_j^i) - m_{i-1}(f)(x_j^i)) \cdot a_j^i.$$

En notant  $X_\Delta^i = X^i \setminus X^{i-1}$ , on a

$$\Delta^i = \sum_{x_j^i \in X_\Delta^i} \underbrace{(f(x_j^i) - m_{i-1}(f)(x_j^i))}_{w_j^i} \cdot a_j^i, \quad (2)$$

On obtient alors la construction hiérarchique de  $m_i(f)$  :

$$\begin{aligned} m_i(f) &= m_{i-1}(f) + \Delta^i \\ &= \sum_{k=1}^i \Delta^k. \end{aligned}$$

## Construction hiérarchique en dimension $n$

Dans le cas général où on cherche à interpoler la fonction  $f : [0, 1]^n \rightarrow \mathbb{R}$ , on obtient par tensorisation :

$$m_{(i_1, i_2, \dots, i_n)}(f) = \sum_{x_{j_1}^{i_1} \in X^{i_1}} \dots \sum_{x_{j_n}^{i_n} \in X^{i_n}} f(x_{j_1}^{i_1}, \dots, x_{j_n}^{i_n}) (a_{j_1}^{i_1} \otimes \dots \otimes a_{j_n}^{i_n}).$$

De même, l'expression hiérarchique en dimension  $n$  sur une grille pleine s'écrit :

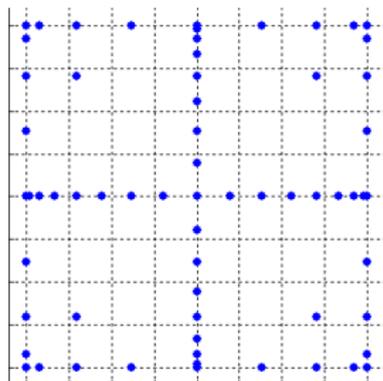
$$m_{(i_1, i_2, \dots, i_n)}(f) = \sum_{k_1=1}^{i_1} \dots \sum_{k_l=1}^{i_l} \dots \sum_{k_n=1}^{i_n} \underbrace{(\Delta^{k_1} \otimes \dots \otimes \Delta^{k_l} \otimes \dots \otimes \Delta^{k_n})}_{\Delta^k},$$

## Approximation par sparse grids

L'interpolation par sparse grids, proposée par Smolyak, consiste à négliger les fonctions de base aux supports les plus "petits" :

$$SG_N(f) = \sum_{|\mathbf{k}|_1 \leq d+N-1} (\Delta^{k_1} \otimes \dots \otimes \Delta^{k_d}).$$

soit par exemple en dimension 2 et pour le niveau  $N = 4$  :



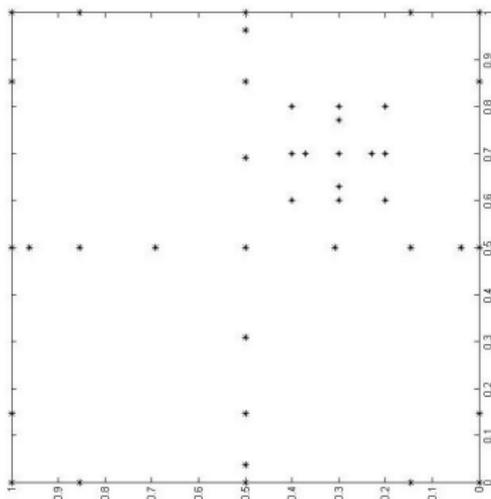
## Approximation par sparse grids

La méthode d'interpolation par sparse grids permet de réduire considérablement le nombre de points d'interpolation utilisés tout en gardant une très bonne approximation de la fonction  $f$  :

Niveau $N$	Points d'interpolation	erreur en norme $L^2$
Grille pleine	$\mathcal{O}(N^n)$	$\mathcal{O}(N^{-2})$
Sparse Grid	$\mathcal{O}(N(\log N)^{n-1})$	$\mathcal{O}(N^{-2}(\log N)^{n-1})$

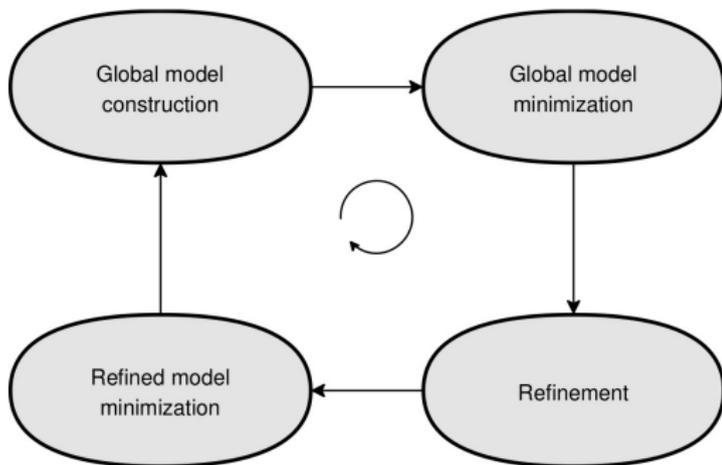
## Raffinement local

Un procédé de raffinement local, toujours par interpolation de type sparse grid, permet d'améliorer le modèle localement en des points d'intérêt :



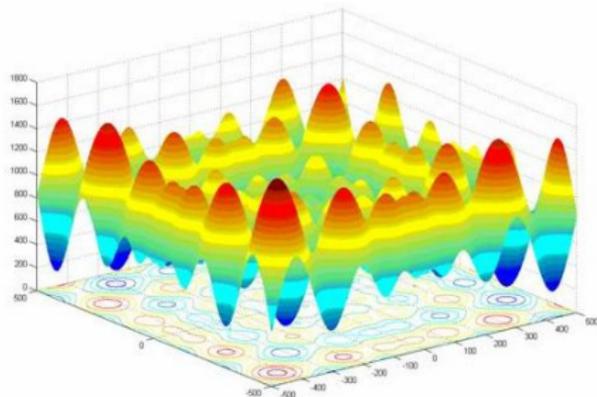
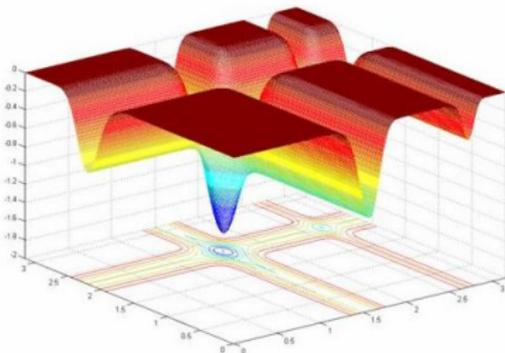
## Principe de la méthode d'optimisation

- La méthode d'optimisation associée à la surface de réponse de type sparse grid consiste à construire de manière hiérarchique le modèle d'interpolation.
- Un raffinement local autour des points d'intérêt au niveau hiérarchique  $N$  (minimum du modèle) est auparavant réalisé.



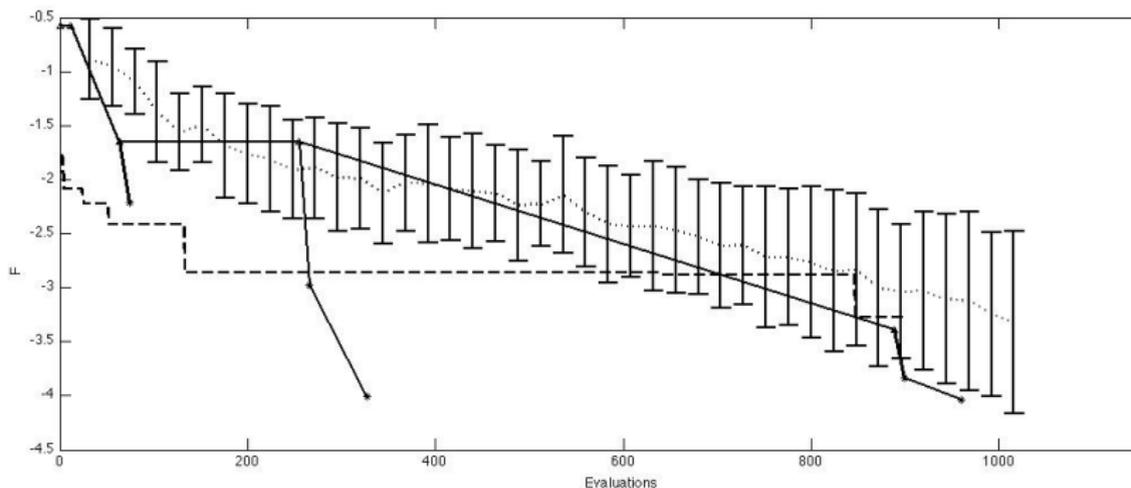
## Résultats numériques sur des fonctions classiques

La méthode précédente, nommée GOSGrid, a été testée sur des benchmarks de fonctions classiques comme les fonctions de Michalewicz et de Schwefel :



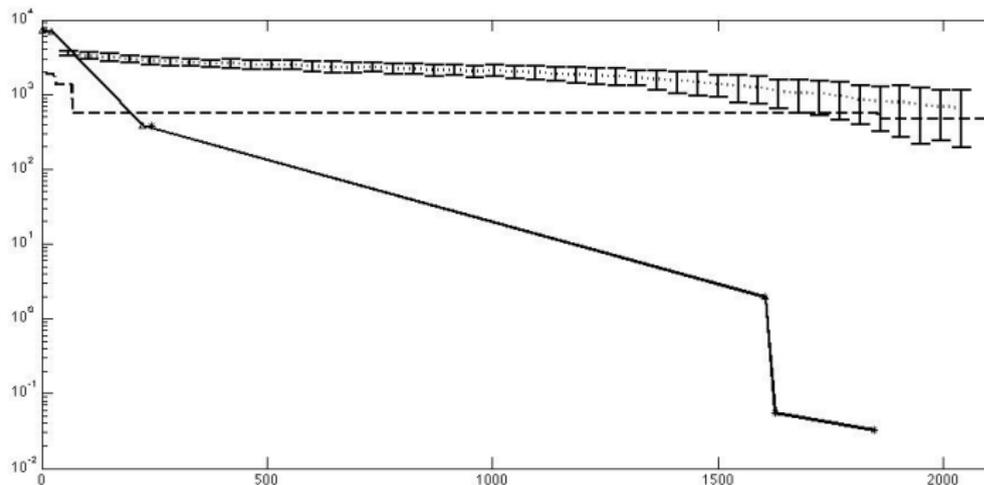
## Résultats sur la fonction de Michalewicz, $n=5$

La figure ci-dessous présente une comparaison de la méthode GOSGrid avec une méthode de krigeage (EGO) et un algorithme évolutionnaire (CMA-ES) :



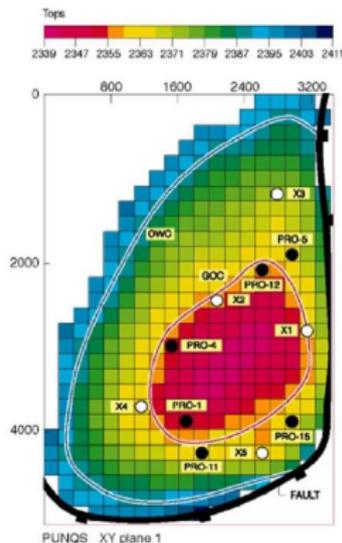
## Résultats sur la fonction de Schwefel, $n=10$

La figure ci-dessous présente une comparaison de la méthode GOSGrid avec une méthode de krigeage (EGO) et un algorithme évolutionnaire (CMA-ES) :



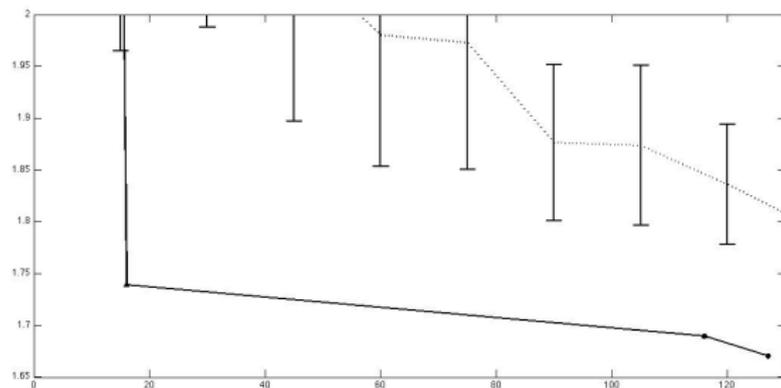
## Résultat numérique sur un cas test réservoir

La méthode GOSGrid a également été appliquée sur un cas test de réservoir synthétique, avec des données de production bruitées (7 paramètres géologiques à déterminer et 6 puits de mesure) :



## Résultat numérique sur un cas test réservoir

Les résultats obtenus sont meilleurs que ceux issus du calcul avec une méthode évolutionnaire lorsque la structure particulière de la fonction (somme de carrés) est utilisée :



- 1 Problèmes inverses en ingénierie pétrolière
- 2 Une nouvelle méthode de type surface de réponse
  - Principe des sparse grids
  - Méthode d'optimisation associée
  - Résultats numériques
- 3 **Une nouvelle méthode de type région de confiance**
  - Principe général
  - Adaptation au cas des fonctions partiellement séparables
  - Un résultat de convergence
  - Résultats numériques

# Description d'une méthode de type région de confiance

## ALGORITHME SIMPLE

- 1 **Initialisation** Construire un modèle quadratique initial  $m_0$  de la fonction
- 2 **Itération  $k$** 
  - 1 Calculer  $x_k^+$  le minimum de  $m_k$  sur une région de confiance
    - remplacer le point d'interpolation le plus éloigné du point courant par  $x_k^+$
    - si  $x_k^+$  est bon alors  $x_{k+1} := x_k^+$  et augmenter la région de confiance, sinon  $x_{k+1} := x_k$  et réduire la région de confiance
  - 2 Construire le nouveau modèle  $m_k$  de la fonction.
- 3 **Condition d'arrêt** S'arrêter lorsque le gradient du modèle devient trop petit

## Critère d'acceptation du nouveau point

- Le critère d'acceptation/rejet du point  $x_k^+$  est basé sur le calcul du rapport :

$$\rho_k = \frac{m_k(x_k^+) - f(x_k)}{f(x_k^+) - f(x_k)}$$

- Le modèle est supposé bon lorsque  $\rho_k \geq \rho$  avec  $\rho \in ]0, 1[$ .

## Choix de la méthode

- Dans le cas d'une méthode sans gradient, le modèle quadratique de la région de confiance est obtenu par interpolation et résolution d'un problème aux moindres carrés.
- On choisit de travailler ici sur une méthode proche de la méthode NEWUOA de Powell (2004).

## Calcul du modèle quadratique approché dans NEWUOA

- Dans la méthode NEWUOA, l'obtention du modèle quadratique  $m_k$  nécessite un ensemble de points d'interpolation comprenant au minimum  $n + 1$  points.
- Le modèle quadratique approché  $m_k$  est obtenu en résolvant le problème de minimisation suivant :

$$\min(\|\alpha_Q\|_F^2)$$

sous la contrainte :  $m_k(x_j) = f(x_j)$  pour tout  $j \in \{1, \dots, m\}$  où

$$m_k(x) = f(x_k) + \langle \alpha_L, x \rangle + \frac{1}{2} \langle x, \alpha_Q x \rangle$$

- De bons résultats sont en général obtenus avec  $m = 2n + 1$  points.

## Etape de vérification de la géométrie dans NEWUOA

- Cette méthode possède une étape supplémentaire indispensable de vérification de la géométrie des points utilisés dans la construction du modèle quadratique.
- Ces points sont choisis de telle sorte que l'ensemble soit unisolvant : on dit que l'ensemble  $Y = \{y_1, \dots, y_m\}$  est  $\Lambda$  unisolvant sur  $B$  ssi la famille des polynômes de Lagrange associée à  $Y$  est telle que

$$\Lambda \geq \max_{j=1, \dots, m} \max_{x \in B} |l_j(x)|$$

- Dans l'étape d'initialisation, les points d'interpolation sont choisis à partir du point courant  $x_0$  et d'un réel  $\rho$  :

$$\forall i \in \{1, 2, \dots, n\} \begin{cases} x_{i+1} = x_0 + \rho e_i \\ x_{i+n+1} = x_0 - \rho e_i \end{cases}$$

# Description de la méthode NEWUOA

## ALGORITHME NEWUOA

- 1 **Initialisation** Construire un modèle quadratique initial  $m_0$  de la fonction (à partir de  $m$  points où  $m \geq n + 1$ )
- 2 **Itération  $k$** 
  - 1 Si le modèle  $m_k$  n'est pas assez précis, remplacer un point d'interpolation par un point améliorant la géométrie du modèle
  - 2 Sinon, calculer  $x_k^+$  le minimum de  $m_k$  sur une région de confiance
    - remplacer le point d'interpolation le plus éloigné du point courant par  $x_k^+$
    - si  $x_k^+$  est bon alors  $x_{k+1} := x_k^+$  et augmenter la région de confiance, sinon  $x_{k+1} := x_k$  et réduire la région de confiance
- 3 **Condition d'arrêt** S'arrêter lorsque le gradient du modèle devient trop petit

## Auto-correction de la géométrie

- Il a été montré qu'il n'est pas nécessaire d'avoir une étape d'amélioration de la géométrie du modèle à chaque itération : le point d'essai  $x_k^+$  permet de s'en affranchir la plupart du temps.
- L'étape de correction de la géométrie s'effectue automatiquement en considérant un critère d'élimination d'un point basé non pas la distance usuelle  $d(y_j, x_k^+)$  mais sur la valeur  $d(y_j, x_k^+) \cdot |l_j(x_k^+)|$  pour  $y_j \in Y$ .
- Ce principe permet de démontrer la convergence de l'algorithme associé (Scheinberg, Toint, 2009).

## Adaptation à la séparabilité partielle de la fonction objectif

- L'idée est de créer un modèle quadratique  $m^i$  pour chaque sous fonction objectif. On espère ainsi :
  - Obtenir des modèles plus précis
  - Construire des modèles à partir de moins de points d'interpolation
- Un problème apparaît pour l'étape d'amélioration des modèles. Sans modifications, il serait nécessaire d'améliorer tous les sous modèles, ce qui aurait un coût prohibitif.

## Adaptation à la séparabilité partielle de la fonction objectif

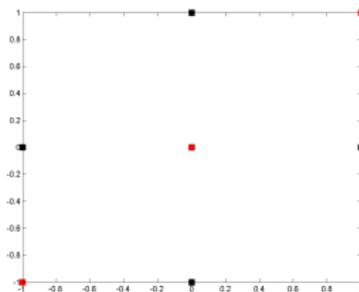
- On définit la notion de sous modèle dominant : il s'agit du sous-modèle maximisant la valeur

$$\rho_k^i = \frac{m_k^i(x_k^+) - f(x_k)}{f(x_k^+) - f(x_k)}$$

- Un résultat d'auto-correction de la géométrie similaire au cas général peut être obtenu en imposant que  $x_k^+$  améliore au moins le sous modèle dominant à chaque itération.

## Adaptation de l'initialisation

- Dans le cas où  $f$  s'écrit :  $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$ , l'initialisation se fait grâce à trois points et non cinq points :



- De manière générale, l'initialisation se fait avec un nombre réduit de points exploitant l'indépendance partielle des variables.

## Convergence de la méthode

- Hypothèses : on suppose que
  - (i) la fonction  $f$  est différentiable et  $\nabla f$  continue et Lipschitzienne.
  - (ii)  $f$  est minorée sur un ensemble contenant toutes les itérations de l'algorithme.
- Conclusion : la suite d'itérations  $(x_k)_{k \in \mathbb{N}}$  de la méthode issue de NEWUOA et exploitant la séparabilité partielle de la fonction  $f$  est telle que

$$\lim_{k \rightarrow +\infty} \nabla f(x_k) = 0$$

## Idée de la preuve

- La démonstration repose sur le résultat d'approximation suivant :

**Proposition** : Etant donné une boule fermée  $B(x, \Delta)$  et un ensemble d'interpolation  $Y$   $\Lambda$ -unisolvant, il existe des constantes  $K_1$  et  $K_2$  telle que

$$\|f(y) - m(y)\| \leq K_1 \sum_{j=1}^m \|y_j - y\|^2 |l_j(y)|$$

et

$$\|\nabla f(y) - \nabla m(y)\| \leq K_1 \Lambda \Delta$$

- On utilise aussi un résultat de décroissance du modèle :

$$m_k(x_k) - m_k(x_k^+) \leq K_3 \|\nabla m_k(x_k)\| \min\left(\frac{\|\nabla m_k(x_k)\|}{1 + \|\nabla^2 m_k(x_k)\|}, \Delta_k\right)$$

## Fonctions tests

Des tests de la nouvelle méthode, nommée PSOF, ont été réalisés sur des fonctions analytiques partiellement séparables : pour  $x = (x_1, \dots, x_n)$

$$DQDRTIC(x) = \sum_{i=1}^{n-2} (x_i^2 + x_{i+1}^2 + x_{i+2}^2)$$

$$LIARWHD(x) = \sum_{i=1}^n (4(x_i^2 - x_1)^2 + (x_i - 1)^2)$$

$$BDQRTIC(x) = \sum_{i=1}^{n-4} ((-4x_i + 3)^2 + (x_i^2 + x_{i+1}^2 + x_{i+2}^2 + x_{i+3}^2) + 5x_n^2)$$

$$ARWHEAD(x) = \sum_{i=1}^{n-1} ((x_i^2 + x_n^2)^2 - 4x_i + 3)$$

$$ROSENBROCK(x) = \sum_{i=1}^{n-1} (100(x_i - x_{i+1})^2 + (x_i - 1)^2)$$

## Résultats numériques, comparaison avec NEWUOA

- Le tableau suivant compare la minimisation des fonctions tests précédentes réalisées une méthode de type NEWUOA ou avec la nouvelle méthode PSOF.
- Les mêmes paramètres de départ sont pris pour les 2 méthodes, on compare dans ce tableau le nombre de simulations pour arriver à convergence

Fonction	10 param.		50 param.	
	NEWUOA	PSOF	NEWUOA	PSOF
DQDRTIC	204	21	+1300	20
LIARWHD	174	51	1215	66
BDQRTIC	231	149	+2000	169
ARWHEAD	368	43	+1300	45
ROSENBROCK	244	128	+1600	233

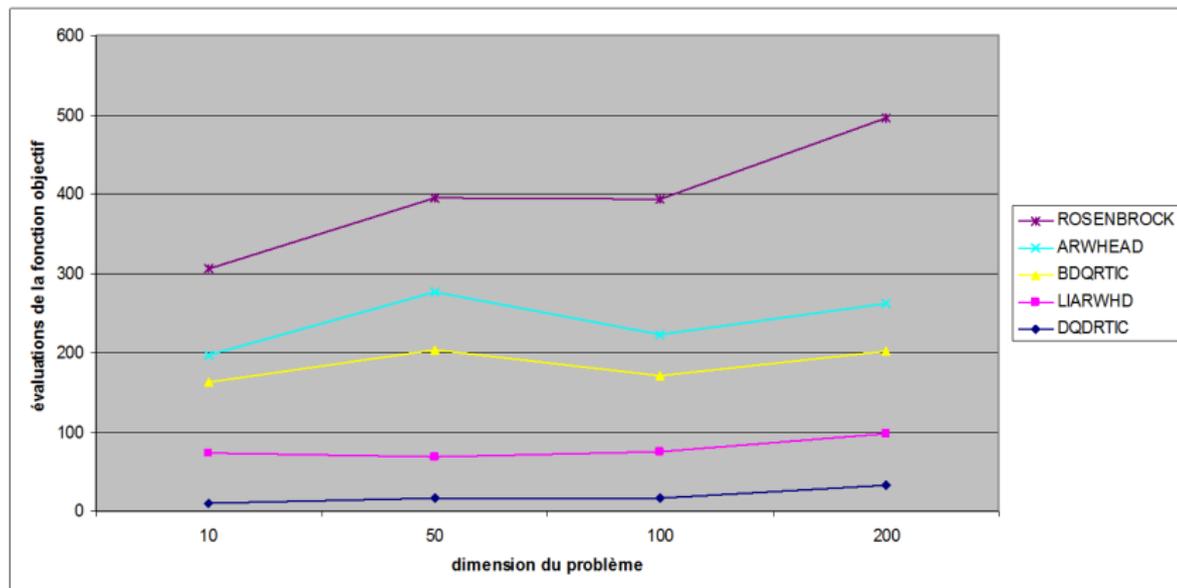
## Résultats numériques, comparaison avec NEWUOA

Ce tableau compare la valeur de la fonction objectif à l'arrêt des algorithmes :

Fonction	10 param.		50 param.	
	NEWUOA	PSOF	NEWUOA	PSOF
DQDRTIC	$1.9 * 10^{-4}$	$8.0 * 10^{-17}$	13000	$9.3 * 10^{-18}$
LIARWHD	0.01	$6.0 * 10^{-9}$	0.01	$3.6 * 10^{-7}$
BDQRTIC	37.9	18.5	312	178.9
ARWHEAD	2.77	$7.7 * 10^{-9}$	2.9	$1.8 * 10^{-8}$
ROSENBROCK	1.8	$6.0 * 10^{-5}$	129	0.04

## Dépendance au nombre de variables d'optimisation

Les résultats de la méthode PSOF sont quasiment indépendants du nombre de variables :



## Conclusion et perspectives

- Deux types de méthodes d'optimisation sans gradient ont été développés pour résoudre des problèmes inverses dans le domaine pétrolier (ingénierie de réservoir).
- La première méthode consiste à construire une surface de réponse de type sparse grids et de raffiner celle-ci localement autour des points d'intérêt.
- La seconde méthode consiste à adapter un procédé de type région de confiance au cas de fonctions partiellement séparables.
- Ces deux approches, de nature complémentaire, permettent d'améliorer les méthodes existantes, en particulier dans le cas d'un grand nombre de variables à optimiser.