A New Derivative-Free Optimization Method for Partially Separable Functions

Laurent DUMAS

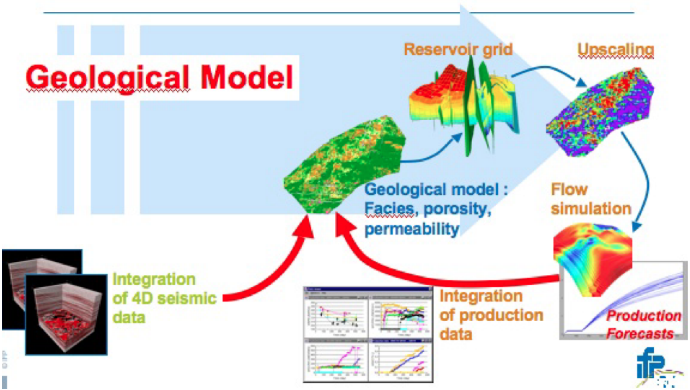
with D. Ding (IFPEN), B. Marteau (PhD , IFPEN)

Laboratoire de Mathématiques de Versailles (LMV)
Université de Versailles
France

CORS/ INFORMS 2015, June, 15th

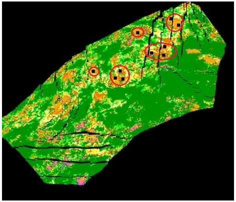
Outline of the talk

History matching in reservoir engineering



Main characteristics of the inverse problem

- High computational cost of the objective function : each evaluation of $(x_1, , ..., x_n) \mapsto f(x_1, ..., x_n)$ needs the simulation of a costly geological model.
- Large number of parameters.
- Partial separability of the cost function.



Partial separability of the cost function

The objective function can be written in the following form :

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

with $\forall i,\ n_i \leq n$

Objectives

- The inverse problem can be summarized as follows : *obtain the best history matching solution, given a small and fixed number of cost function evaluations*.
- Some questions arise at each stage of the inverse problem :
  - Parametrization : find the most adapted parameters of the problem.
  - Initialization : find the best initialization process.
  - Optimisation : exploit the specific properties of the cost function.

Objectives

- The lack of information about the cost function gradient lead to consider Derivation Free Optimization (DFO) methods.
- Among DFO methods (direct methods, evolutionary algorithms, response surface, trust region, etc...), trust region methods have been chosen.

1. An inverse problem in oil industry

2. Trust region methods
   - Description
   - The NEWUOA method
   - A self improving geometry principle

3. The new algorithm for partially separable functions

4. Numerical results

Description of a trust region method

A simple algorithm

1. **Initialisation** Construction of an initial quadratic model m_0 of the cost function on $B(x_0, \Delta_0)$.
2. **Iteration** k
   - Compute x_k^+, the minimum of m_k on $B(x_k, \Delta_k)$.
   - Remplace the current most far interpolation point by x_k^+
   - If x_k^+ is 'good', then $x_{k+1} := x_k^+$ and increase the trust region radius, else $x_{k+1} := x_k$ and reduce the trust region radius.
   - Build the new interpolation model m_k.
3. **Stopping condition**. Stop if the gradient model at x_k is small.

The acception criterion

- The acception/rejection criterion of x_k^+ is based on the following rate :

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

- The model is said to be 'good' if $\rho_k \geq \nu$ with $\nu \in]0, 1[$.

The quadratic model

- In a DFO approach, the quadratic model on the trust region is obtained by quadratic Lagrange interpolation with $p = \frac{(n+1)(n+2)}{2}$ points.
- It is possible to reduce the number of interpolation points by solving a least square minimization problem for some of the polynomial coefficients (Powell, *NEWUOA algorithm 2004*).

The NEWUOA initialization

- The NEWUOA algorithm uses $m = 2n + 1$ points for the initialization step.
- The initial interpolation points are chosen around the starting point x_0 :

$$\forall i \in \{1, 2, ..., n\} \left\{ \begin{array}{l} y_{i+1} = x_0 + \rho e_i \\ y_{i+n+1} = x_0 - \rho e_i \end{array} \right.$$

Poisedness of the interpolation set

- It is necessary to check at each iteration the well poisedness of the interpolation set.
- The set $Y = \{y_1, \dots y_m\}$ is said to be Λ poised on B if its associated Lagrange polynomial family is such that :

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

- It is possible to ensure at each iteration the Λ-poisedness, but with an additional computational cost.

A self improving geometry principle

- A self improving geometry principle has been introduced by Scheinberg and Toint (2009).
- It is based on the following lemma :

Lemma

Let $\beta \in (0,1)$. For all $\Lambda > 1$, there exists k_Λ such that if iteration k is unsuccessfull, that is $\rho_k < \nu$, and :

$$F_k := \{y_{k,j} \in Y_k \text{ s.t. } ||y_{k,j} - x_k|| > \beta\Delta_k \text{ and } l_{k,j}(x_k^+) \neq 0\} = \emptyset$$

with $\Delta_k < k_\Lambda ||\nabla m_k(x_k)||$, then the set

$$C_k := \{y_{k,j} \in Y_k \backslash \{x_k\} \text{ s.t. } ||y_{k,j} - x_k|| \leq \beta\Delta_k \text{ and } l_{k,j}(x_k^+) \geq \Lambda\}$$

is non-empty.

A self improving geometry principle

Trust Region algorithm with geometry self correction

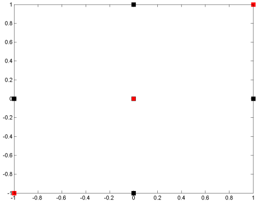
1. **Initialization**
2. **Criticality test**
3. **Iteration** k
   - Compute x_k^+ the minimum of m_k on $B(x_k, \Delta_k)$.
   - If x_k^+ is good, replace x_k by x_k^+, increase Δ_k and replace the most far interpolation point by x_k^+.
   - If x_k^+ is not good, keep x_k and Δ_k and replace one point in F_k (far from x_k^+) or in C_k (near x_k^+) by x_k^+, if one among them is non empty, else reduce Δ_k.
   - Build the new interpolation model m_k.

The initialization principle

- In the case of a cost function f of the type :

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2)$$

the initialization can be done with 3 points instead of 5 :



- More generally, initialization is done with a reduced number of points by exploiting the independance of variables (using a coloured graph).

The quadratic model

- A different quadratic model m_i for each partial objective function is created :

$$m_k(x_1, \ldots, x_n) = \sum_{i=1}^{p} m_k^i(x_{1_i}, \ldots, x_{n_i})$$

- It is thus expected to give a more accurate model, with less interpolation points.
- However, a strategy for improving at least one submodel at each iteration is needed.

The dominant submodel

- Define the improvement rate of the i-th submodel :

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

- Among them, the dominant submodel is the one maximizing its improvement :

$$j = arg\max_i (m_k^i(x_k^+) - m_k^i(x_k))$$

A self improving geometry principle

A self improving geometry principle has been proved for the dominant submodel :

Lemma

Let F_k^i and C_k^i the extensions of F_k et C_k to submodel i. At iteration k, for all $\Lambda > 1$, if j is the index of the dominant submodel, and :

$$\begin{cases} \Delta_k \leq k_\Lambda \|g_k\| \\ \rho_k^j < \nu \\ F_k^j = \emptyset \end{cases}$$

then $C_k^j \neq \emptyset$.

The new DFO-PSOF method

DFO-PSOF algorithm

1. **Initialization**
2. **Criticality test**
3. **Iteration** k
   - Compute x_k^+ the minimum of x_k on a trust region set, then compute ρ_k and each ρ_k^i.
   - Treatment of each submodel for the 3 cases :
     - $\rho_k > \nu$,
     - $\rho_k < \nu$ and $\rho_k^i < \nu$ with the sets F_k^i and C_k^i (if non empty)
     - $\rho_k < \nu$ and $\rho_k^i > \nu$
   - If no improvement is done at the previous stage (only the last case) : improvement of at least one non dominant model.
   - Update Δ_k.
   - Build the new model m_k and submodels m_k^i.

A convergence result

Theorem

Assume that :

(i) the cost function f is differentiable and ∇f is Lipschitz continuous on a set $\mathcal{V}$ including all the iterations,

(ii) f is bounded from below on $\mathcal{V}$,

(iii) for all $k \in \mathbb{N}$, $||H_k|| \leq C$

then, the sequence $(x_k)_{k\in\mathbb{N}}$ of the DFO-PSOF algorithm is such that :

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

Numerical results on analytic test functions

Some tests have first been made on classical partially separated test functions : for $x = (x_1, \ldots, x_n)$, define :

$$DQDRTIC(x) = \sum_{i=1}^{n-2} (x_i^2 + x_{i+1}^2 + x_{i+2}^2)$$
$$LIARWHD(x) = \sum_{i=1}^{n} (4(x_i^2 - x_1)^2 + (x_i - 1)^2)$$
$$BDQRTIC(x) = \sum_{i=1}^{n-4}((-4x_i + 3)^2 + (x_i^2 + x_{i+1}^2 + x_{i+2}^2 + x_{i+3}^2) + 5x_n^2)$$
$$ARWHEAD(x) = \sum_{i=1}^{n-1}((x_i^2 + x_n^2)^2 - 4x_i + 3)$$
$$ROSENBROCK(x) = \sum_{i=1}^{n-1}(100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2)$$

Numerical results on analytic test functions

- The following table compares the DFO-PSOF results with a NEWUOA type method (number of evaluations for reaching convergence) :

Function	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

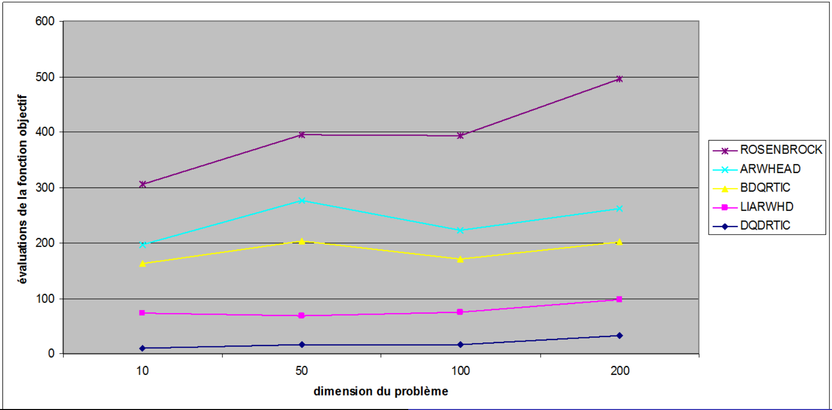
Numerical results on analytic test functions

The cost function value at the end of the algorithm is the following :

Function	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

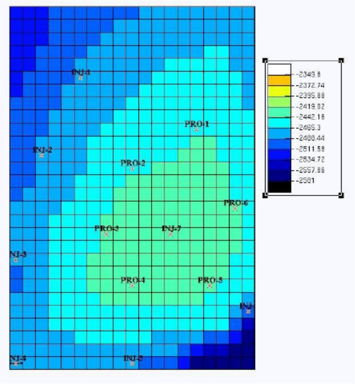
Numerical results on analytic test functions

The DFO-PSOF results are almost independent of the parameters number :



Application in reservoir engineering

- The PUNQ test case is a synthetic test case respresentative of a real oil field :



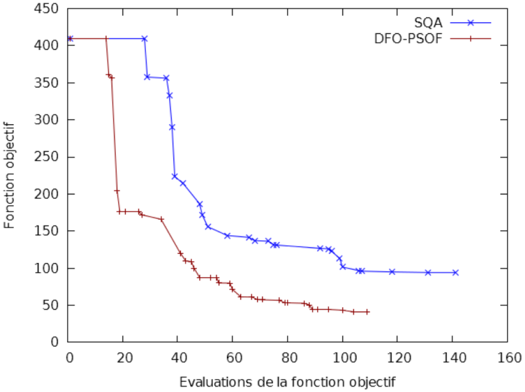
Application in reservoir engineering

- The PUNQ test case includes 6 producing wells and 7 injectors.
- The following partial separability is assumed :

f_{PRO-1}	X_{PRO-1}	X_{INJ-1}	X_{PRO-2}	X_{INJ-6}	X_{INJ-7}	
f_{INJ-1}	X_{PRO-1}	X_{INJ-1}	X_{PRO-2}	X_{INJ-2}		
f_{PRO-2}	X_{PRO-1}	X_{INJ-1}	X_{PRO-2}	X_{INJ-2}	X_{PRO-2}	X_{INJ-7}
f_{INJ-2}	X_{INJ-1}	X_{INJ-2}	X_{PRO-2}	X_{INJ-3}	X_{PRO-3}	
f_{PRO-3}	X_{INJ-2}	X_{PRO-3}	X_{INJ-3}	X_{PRO-4}	X_{INJ-4}	X_{INJ-7}
f_{INJ-3}	X_{PRO-2}	X_{INJ-3}	X_{PRO-3}			
f_{PRO-4}	X_{INJ-3}	X_{PRO-4}	X_{INJ-4}	X_{INJ-5}	X_{INJ-7}	
f_{INJ-4}	X_{PRO-2}	X_{PRO-3}	X_{INJ-4}			
f_{PRO-5}	X_{PRO-4}	X_{PRO-5}	X_{INJ-5}	X_{INJ-6}	X_{INJ-7}	
f_{INJ-5}	X_{PRO-4}	X_{PRO-5}	X_{INJ-5}	X_{INJ-6}		
f_{PRO-6}	X_{PRO-1}	X_{PRO-5}	X_{PRO-6}	X_{INJ-6}	X_{INJ-7}	
f_{INJ-6}	X_{PRO-5}	X_{PRO-6}	X_{INJ-6}			
f_{INJ-7}	X_{PRO-2}	X_{PRO-3}	X_{PRO-4}	X_{PRO-5}	X_{PRO-6}	X_{INJ-7}

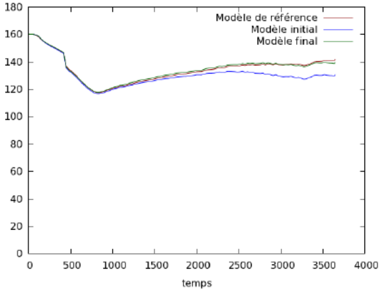
Application in reservoir engineering

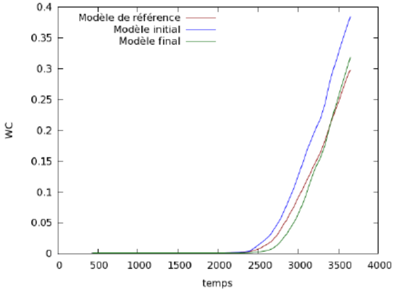
- The obtained results show a large improvement by using PSOF, compared to NEWUOA :



Application in reservoir engineering

- The following figure displays the improvement after optimization of history matching on an arbitrary well :





Conclusions

- A new DFO method has been developped to solve an inverse problem in oil industry.
- The new method consists in adapting a trust region method to the case of partially separable cost functions.
- By exploiting the partial separability of the cost function at the initialization stage and during the optimization process, the convergence rate has been improved compared to other classical trust region methods.