Application of the genetic algorithms to the identification of the hydrodynamic parameters

Wenddabo Olivier Sawadogo^{1*}, Noureddine Alaa², Kounhinir Somé¹, Blaise Somé¹

¹ University of Ouagadougou, Burkina Faso ² University of Cadi Ayyad, Morocco *Corresponding author E-mail: wenddabo81@gmail.com

Copyright ©2015 Sawadogo et.al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this work, we propose an adaptation of the algorithm Non-dominated Sorting Genetic Algorithm-II (NSGA-II) proposed by deb. et al. (2002) to solve multi-objective problems to the resolution of mono-objective problem. Contrary to the majority of the genetic algorithms, we did not define a probability of crossing.

After having applied our algorithm to functions test, we then used it to identify hydrogeologic parameters where the boundaries values and the source term are supposed to be unknown besides the permeability.

The direct problem was solved by using the finite elements of Galerkin on freefem++ and the genetic algorithm was programmed in Matlab. Then we carried out a coupling of the two codes to identify the parameters.

AMS Subject Classification: 35A15, 35R30, 65L09, 65L10, 65L12.

Keywords: Code coupling; Finite elements; Genetic algorithm; Hydrodynamic parameters; Inverse problem.

1. Introduction

For the numerical simulation of flow and transfer related problems of substances into the basement knowing the boundary values, the source terms and certain physical parameters of geological strata is capital. However, these factors which are generally experiment-based cause a number of difficulties.

In this work, we adapt the genetic algorithm Non-dominated Sorting Genetic Algorithm-II (NSGA-II) proposed by Deb and al(2002)[11] for the optimization of a multi-objective problem to the resolution of a mono-objective problem without constraints. First we tested it on examples before using it to solve an inverse problem of hydrogeology by basing us on actual data resulting from the studing site of project TRANSPOL II (INERIS 2003)[1].

2. Presentation of the algorithm used

This algorithm makes it possible to maximize a positive function f called fitness or evaluation function of the individual. The individuals represent the variables.

2.1. Coding and creation of the initial population

The real-type coding used consists in directly representing the actual values of the variable. We subdivided the eligible field in several under fields. And the initial population was created in a random way by using the uniform law in each under field. That makes it possible to have a diversified population from the beginning and convergency-accelerating. The size of the population is n. We create a table of n variables.

2.2. Operation of selection

We used the selection by caster of Goldberg [19]. The parents are selected according to their performance. In this method the probability p with which an individual i represented by a variable x_i of fitness f_i (evaluation of the function in x_i) reintroduced in a new population of size n is:

$$p = \frac{f_i}{\sum_{j=1}^n f_j}$$

2.3. Operation of crossing

The barycentric crossing is used but we did not use a probability of crossing. In this kind of crossing, two genes P1(i) and P2(i) are selected from each parent to the same position i. They define two new genes C1(i) and C2(i) by linear combination:

$$C1(i) = aP1(i) + (1-a)P2(i)$$
; $C2(i) = (1-a)P1(i) + aP2(i)$; $a \in]0,1[$.

In this document, we crossed the whole mother population to get a child population of size n.

2.4. Operation of mutation

Mutation of a Gaussian type is applied to the population. One selects an individual x under a probability p. If p is lower than the probability of mutation p_m , one adds a Gaussian noise to x i.e. one replaces x by $x + \epsilon$, where ϵ are a random value obtained according to the law from Gauss. The newly-created individual replaces the former one if it is better and if it is in the acceptable field.

2.5. New population

After the operations of selection, crossing and change, an intermediate population of size 2n is created by gathering the parent and child populations. The new parent population is obtained by keeping the N better individuals. Finally the algorithm used is:

Algorithm 1 Algorithm used

For each iteration t do

To calculate the score of each individual of P_t

To generate a new population of child Q_t by applying the operators of selection, crossing and mutation

 $R_t = P_t \cup Q_t \text{ (add } Q_t \text{ to } P_t)$

To classify the individuals of R_t from decreasing order according to the score of each individual

To keep n best individuals of R_t to form a new population of parent P_{t+1}

t = t + 1 (To increment the counter of the generation)

2.6. Application to the test functions

Before using our algorithm for the identification of the parameters in hydrogeology, we tested it on classical functions.

2.6.1. Function of Rosenbrock

The function of Rosenbrock is a unimodal function, nonconvex, of n dimensions, used like test for problems of mathematical optimization. It is defined by:

$$F_1(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2) + (x_i - 1)^2]$$

In the literature, this function is regarded as being a difficult problem because of the nonlinear interaction between the variables [20]. The global minimum is obtained as in point (1, ..., 1), for which the function is worth 0.

2.6.2. Function of Rastrigin

It is a function of n dimensions, strongly multimodal defined by:

$$F_2(x) = \sum_{i=1}^{n} (x_i^2 - 10 * \cos(2\pi x_i) + 10)$$

The local minima site is distributed regularly. The global minimum is in the beginning and the value of its function is equal to zero.

2.6.3. Results of tests

In all the examples, we used the same parameters given by the table below:

p_m	σ	size of the opulation	Number of iterations
0.00001	0.5	100	100

We applied our algorithm to these functions in dimension three. Then, we compared our results, with those obtained by the genetic algorithm provided with by Matlab (the function ga). For each function test, we carry out five simulations. The tables below, gives the results of optimization. It is noticed that the results got with our code are better than those obtained with the function ga of Matlab but our code is slower.

Table 1: Minimization of the function of Rosenbrock: our code

x_1	x_2	x_3	Values of the function
0.99994	0.99990	0.99980	1.784×10^{-8}
1	1	1	0
0.99256	0.98518	0.97107	2.974×10^{-4}
0.99999	0.99999	0.99999	0
1	1	1	0

Table 2: Minimization of the function of Rosenbrock: function ga

x_1	x_2	x_3	Values of the function
0.93595	0.87593	0.76760	1.950×10^{-2}
0.83982	0.70461	0.49370	1.137×10^{-1}
0.83796	0.69925	0.48648	1.181×10^{-1}
0.81582	0.66530	0.44308	1.459×10^{-1}
0.88482	0.78141	0.60889	6.156×10^{-2}

3. Application to the identification of hydrodynamic parameters

3.1. Description of the site

3.1.1. Geometry

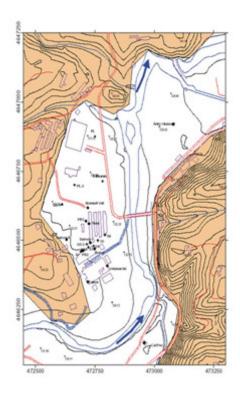
The site in question has a length of 500 meters in the direction of groundwater flow (SN) and a width of 300 meters. The unsaturated zone is 6 meters.

 ${\bf Table~3:}~{\bf Minimization~of~the~function~of~Rastrigin:~our~code$

x_1	x_2	x_3	Values of the function
-3.9×10^{-10}	2.1×10^{-10}	1.5×10^{-9}	0
6×10^{-11}	3.2×10^{-11}	-3.6×10^{-10}	0
-1.5×10^{-9}	2.4×10^{-10}	5.9×10^{-10}	0
-1.1×10^{-10}	-5.1×10^{-11}	10^{-9}	0
3.2×10^{-10}	-7.8×10^{-10}	3.3×10^{-10}	0

Table 4: Minimization of the function of Rastrigin : fonction ga

x_1	x_2	x_3	Values of function
8.418×10^{-4}	-4.565×10^{-4}	-2.163×10^{-3}	6.385×10^{-2}
1.002×10^{-4}	9.949×10^{-1}	7.115×10^{-6}	1.110×10^{-3}
3.598×10^{-5}	9.95×10^{-1}	-7.753×10^{-6}	9.949×10^{-1}
6.797×10^{-6}	-4.108×10^{-4}	-2.730×10^{-4}	9.949×10^{-1}
1.827×10^{-4}	-1.824×10^{-5}	9.949×10^{-1}	9.949×10^{-1}



 $\textbf{Figure 1: } \mathbf{Site}$

3.1.2. Boundary conditions

The map of figure 1 shows the domain to be modeled. The river flows southwest to the northeast. The upstream boundary condition is equated with the river and has a constant load over time. The downstream boundary condition corresponds to an imaginary line perpendicular to the stream. A constant charge will be imposed. The others side boundary and the lower limit corresponding to zero flow conditions.

3.2. Mathematical model of the flow

The fluid is considered incompressible and monophasic and the medium Ω porous and is saturated, and we place ourselves in permanent mode. The law of conservation of the mass and the law of Darcy [13] applied to our site gives:

$$div(u) = f \quad \text{in} \quad \Omega$$

$$u = -K\nabla p \quad \text{in} \quad \Omega$$

$$p = d \quad \text{on} \quad \Gamma^{0}$$

$$u.\nu = -K\frac{\partial p}{\partial r} = 0 \quad \text{on} \quad \Gamma^{1}$$
(1)

where:

- u(x): Darcy velocity,
- p(x): hydraulic potential,
- f:source term,
- K: the hydraulic conductivity (constant)
- Ω represent the domain.
- \bullet Γ^0 : limits upstream and downstream where the hydraulic potential are constant.
- Γ^1 : limits of worthless flows.
- d(x): Dirichlet boundary conditions.

3.3. Direct problem solving

Theorem 1 Under the assumptions of flow, the problem (1) admits one and only solution.

Proof By eliminating u, the problem (1) is equivalent to:

$$-div(K\nabla p) = f \quad \text{in} \quad \Omega$$

$$p = d \quad \text{on} \quad \Gamma^{0}$$

$$-K\frac{\partial p}{\partial n} = 0 \quad \text{on} \quad \Gamma^{1}$$
(2)

Let V be the space be defined by

$$V = \{ v \in H^1(\Omega), \ v = 0 \text{ on } \Gamma^0 \}.$$

Since $mes(\Gamma^0) > 0$, according to [21], we can choose

$$||v||_V = \left(\int_{\Omega} |\nabla v|^2\right)^{1/2}$$
 as norm on V .

Let $v \in V$ be a test function. On multiplying (2) by v and integrating by parts, the variational formulation associated to the problem (2) is:

$$\begin{cases}
\operatorname{Find} p \in H^1(\Omega) \text{ such that } p = d \text{ on } \Gamma^0 \text{ and such that} \\
\int_{\Omega} K \nabla p \cdot \nabla v = \int_{\Omega} f v, \, \forall v \in V
\end{cases}$$
(3)

We denote by γ_0 the trace operator. Let $r_d \in H^1(\Omega)$ such as $\gamma_0(r_d) = d$ and we denote $p_0 = p - r_d$. The variational formulation becomes:

$$\begin{cases}
\operatorname{Find} p_0 \in V \text{ such as} \\
\int_{\Omega} K \nabla p_0 \cdot \nabla v = \int_{\Omega} K \nabla r_d \cdot \nabla v + \int_{\Omega} f v, \, \forall v \in V
\end{cases}$$
(4)

 $f \in L^2(\Omega)$

Let the bilinear form $\alpha: V \times V \longrightarrow \mathbb{R}$ be defined by:

$$\alpha(p_0, v) = \int_{\Omega} K \nabla p_0 . \nabla v$$

Let the linear form $L: V \longrightarrow \mathbb{R}$ be defined by:

$$L(v) = \int_{\Omega} K \nabla r_d . \nabla v + \int_{\Omega} f v$$

The space V is a Hilbert space for the Hilbertian norm $\|.\|_V$.

The bilinear form α is continuous, coercive and the linear form L is also continuous. Thus the theorem of Lax-Milgram [15] ensures the existence and uniqueness of a solution to the variational problem (4) and consequently the existence and uniqueness of a solution of (2).

Let T_h be a triangulation of Ω . Let P_1 denote the space of continuous, piewise affine function in Ω i.e the space of continuous functions which are affine in x, y on each triangle of T_h . We pose $V_h = P_1 \cap V$. V_h is a linear vector space of finite dimension. We denote N its dimension and ϕ_1, \ldots, ϕ_N a basis. The approximated problem is:

find
$$p_h \in V_h$$
, such that $\alpha(p_h, v_h) = L(v_h)$ for all $v_h \in V_h$. (5)

Let

$$p_h(x,y) = \sum_{i=1}^{N} p_i \phi_i(x,y)$$

and take $v_h = \phi_i$ for i = 1, ..., N; equation (6) is equivalent to

$$\alpha(\sum_{i=1}^{N} p_j \phi_j, \phi_i) = L(\phi_i) , i = 1, ..., N$$
(6)

This gives the system Ax = b, where:

$$A_{ij} = \int_{\Omega} K \nabla \phi_i . \nabla \phi_j = \sum_{T \in T_i} \int_{T} K \nabla \phi_i . \nabla \phi_j$$

and

$$\begin{split} b_i &= \int_{\Omega} f \phi_i + \int_{\Omega} K \nabla r_d. \nabla \phi_i \\ &= \sum_{T \in T_c} \int_{T} f \phi_i + \sum_{T \in T_c} \int_{T} K \nabla r_d. \nabla \phi_i \end{split}$$

3.4. Inverse problem

3.4.1. Data

The location of monitoring points are given in figure 1 and observed values are given in the table below.

Points	X(m)	Y(m)	Observed values
P_{33}	472782	4646714	125.01
P_{32}	473080	4646920	123.23
P_{31}	472749	4646470	125.51
P_{30}	472802	4646408	125.58
P_{28}	472712	4646570	125.4
P_{27}	472696	4646462	125.52
P_{26}	472700	4646464	125.51
P_{24}	472792	4646520	125.4
P_{23}	472762	4646500	125.47
P_{22}	472745	4646480	125.51
P_{20}	472722	4646475	125.53
P_{19}	472691	4646460	125.54
P_{18}	472750	4646875	124.88
P_{17}	472720	4646618	125.29
P_{16}	472664	4646700	125.08
P_{14}	472716	4646487	125.51

Table 5: Observed data

3.4.2. Parametric identification problem

We suppose that the permeability K is constant and unknown. We also assume that the Dirichlet boundary conditions (three values) values and the value of the source term are unknown. The problem is to find the values of these constants that minimize J defined by

$$J = \frac{1}{2} \sum_{i=1}^{nobs} (p_s(x_i, y_i) - p_{ob}(x_i, y_i))^2$$
(7)

with:

 $p_s(x_i, y_i)$ is the simulated pressure head at the point (x_i, y_i) , $p_{ob}(x_i, y_i)$ is the pressure head observed at the point (x_i, y_i) . nobs number of observations.

We use the genetic algorithm to minimize J.

3.5. Algorithm of calculus of the function cost

Since the points (x_i, y_i) of measurement do not correspond necessarily to the points of discretization where the solution is calculated, it is thus necessary to seek the approximation of $h_i(x_i, y_i)$. The points P_{17} and P_{30} were not used in the procedure of identification. They will be used as point tests.

Algorithm 2 Algorithm to calculate the function cost

 $J \leftarrow 0$

for i=1 to nobs-2 do

Determine the triangle T_i such as $(x_i, y_i) \in T_i$

Determine the approximate solution $P_i(x,y)$ so that $(x,y) \in T_i$

 $J \longleftarrow J + \frac{1}{2} (P_i(x_i, y_i) - P_{ob}(x_i, y_i))^2$

endfor

3.6. Inverse problem solving: coupling Freefem++/Matlab

We programmed this algorithm by using Freefem++ and matlab. The function cost which uses the solution of the direct problem is calculated in FreeFem++. The communication between FreeFem++ and Matlab is made through two files. With each iteration, a file which contains the new population (population.dat) is created by the

program Matlab and a file which contains the scores (scores.dat) of the various individuals is created by FreeFem++. Therefore, with each iteration, the FreeFem++ program calculates the scores of the various individuals, by solving the direct problem. The data are the values of the various parameters being in the file *population.dat*. The Matlab program uses the file *scores.dat* for the process of optimization by genetic algorithm.

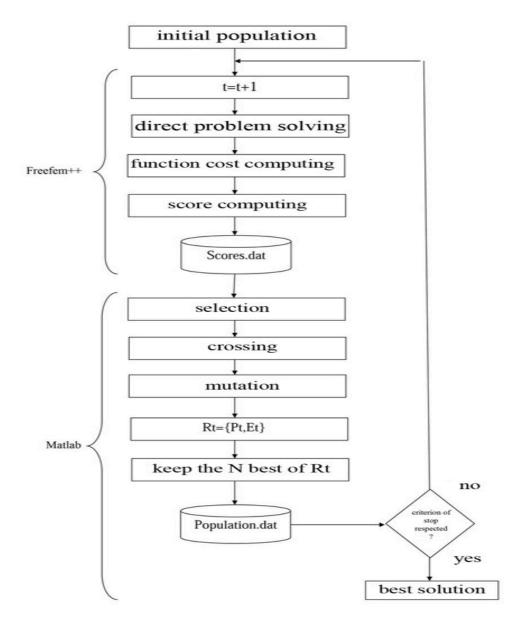


Figure 2: Coupling Freefem++/Matlab

3.7. Results

3.7.1. Configuration of the genitic algorithm

Pm	sigma	size	generations
0.001	0.5	100	30

3.7.2. Identification

We denote:

d1: boundary condition upstream (west),

d2: boundary condition upstream (east),

d3: boundary condition downstream

The identified values are given in table 2.

Table 6: Results of the identification

	Interval	Identified values
K(m/j)	[100, 180]	155.187
$f(m^3/j)$	[0.01, 1]	0.108
d1(m)	[100, 150]	127.9
d2(m)	[100, 150]	124.12
d3(m)	[100, 150]	124.3

Value of the function cost J: 0.734

3.7.3. Simulated groundwater level

Groundwater levels simulated with the identified values are given in the table below

Table 7: Simulated values and observed value

Pt	X(m)	Y(m)	Observed	Simulated	Residue
pt_{33}	472782	4646714	125.01	125.0805	0.0705
pt_{32}	473080	4646920	123.23	124.5968	1.3668
p_{31}	472749	4646470	125.51	125.353	-0.157
p_{30}	472802	4646408	125.58	125.5042	-0.0758
p_{28}	472712	4646570	125.4	125.2194	-0.187
p_{27}	472696	4646462	125.52	125.333	-0.029
p_{26}	472700	4646464	125.51	125.3326	-0.1774
p_{24}	472792	4646520	125.4	125.2979	-0.1021
p_{23}	472762	4646500	125.47	125.3155	-0.1545
p_{22}	472745	4646480	125.51	125.3341	-0.1759
p_{20}	472722	4646475	125.53	125.3265	-0.2035
p_{19}	472691	4646460	125.54	125.3328	-0.2072
p_{18}	472750	4646875	124.88	124.8548	-0.0252
p_{17}	472720	4646618	125.29	125.1826	-0.1074
p_{16}	472664	4646700	125.08	125.1215	0.0415
p_{14}	472716	4646487	125.51	125.3058	-0.2042

3.7.4. Results of INERIS project

In the project INERIS [1], this is to determine the coefficient of permeability K and the source term f. The upstream boundary conditions are imposed at 130 m (west) and 127 m (east). The downstream boundary condition is imposed at 125 m. The method of finite differences of size variable was used for solving the direct problem. The results are summarized in the table below.

Table 8:	Simulated values(Project INERIS)
K	$= 150m/j$ and $f = 0.1m^3/j$

Pt	X(m)	Y(m)	Observed	Simulated	Residue
pt_{33}	472782	4646714	125.01	125.0805	0.0705
pt_{32}	473080	4646920	123.23	124.5968	1.3668
p_{31}	472749	4646470	125.51	125.353	-0.157
p_{30}	472802	4646408	125.58	125.5042	-0.0758
p_{28}	472712	4646570	125.4	125.2194	-0.187
p_{27}	472696	4646462	125.52	125.333	-0.029
p_{26}	472700	4646464	125.51	125.3326	-0.1774
p_{24}	472792	4646520	125.4	125.2979	-0.1021
p_{23}	472762	4646500	125.47	125.3155	-0.1545
p_{22}	472745	4646480	125.51	125.3341	-0.1759
p_{20}	472722	4646475	125.53	125.3265	-0.2035
p_{19}	472691	4646460	125.54	125.3328	-0.2072
p_{18}	472750	4646875	124.88	124.8548	-0.0252
p_{17}	472720	4646618	125.29	125.1826	-0.1074
p_{16}	472664	4646700	125.08	125.1215	0.0415
p_{14}	472716	4646487	125.51	125.3058	-0.2042

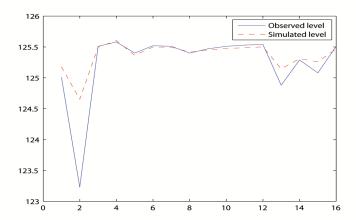


Figure 3: Curve of observed and simulated level to measurement points

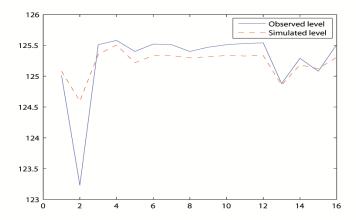


Figure 4: Curve of observed and simulated level to measurement points(Project INERIS)

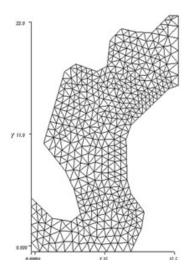


Figure 5: Mesh and Simulated levels on the all domain

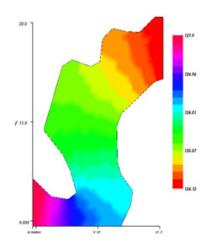


Figure 6: Simulated levels on the all domain

4. Conclusion

In this work, we adapted a genetic algorithm allowing to optimize positive functions effectively. Tests carried out on certain functions made it possible to prove the effectiveness of the algorithm to find the optimum total.

We applied this algorithm to identify hydrodynamic parameters in which the boundary conditions were part of the unknown factors to be identified. This was made by using a coupling of codes carried out on FreeFem++ and Matlab.

By comparing the measured and calculated levels, we see that the method used in this work is much more precise. The levels calculated and measured at points 17 and 30 which were not used in the process of identification as well as the curves of figure 2 confirm that.

But the algorithm that we used remains slow. A parallelization of the code is thus necessary to make it faster.

References

- [1] Adrian Dan, Patrick Goblet Programme TRANSPOL II (INERIS 2003), Rapport final Mai 2004.
- [2] J.H Holland, Adaptation in Natural and Artificial Systems, University of Michigan press, 1975.
- [3] F. Hetch, O. Pironneau FreeFem++, http://www.freefem.org.
- [4] P.J.M. van Laarhoven and E.H.L. Aarts, Simulated annealing: Theory and Applications, Kluwer Academic Publishers, 1987. ISBN: 90-277-2513-6.
- [5] Anatoly A. Zhigljavsky, Theory of Global Random Search, Kluwer Academic Plubishers, 1991.
- [6] R. Cerf, Une Théorie Asymptotique des Algorithmes Génétiques, PhD thesis, Université Montpellier II (France), 1994.
- [7] D.E Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Reading MA Addison Wesley, 1989.
- [8] D.E Goldberg, Real-coded genetic algorithms, virtual alphabets and blocking, Complex Systems, 5:139–167, 1991.
- [9] Latifa Oulladji -Ales Janka -Jean Antoine Désidéri -Alain Dervieux, Optimisation aérodynamique par algorithmes génétiques hybrides: application à la réduction d'un critère de bang sonique, Rapport de recherche INRIA No 4884, Juillet 2003.
- [10] Nicolas Durand, Nicolas Alech, Jean-Marc Alliot, and Marc Schoenauer, Genetic algorithms for optimal air traffic conflict resolution, In Proceedings of the Second Singapore Conference on Intelligent Systems. SPICIS, 1994.
- [11] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan, A Fast Elitist Multiobjective Genetic Algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation 6 (2002), no. 2, 182? 197.
- [12] Owe Axelsson, Vincent A. Barker, Finite element solution of boundary value problems: theory and computation, Society for Industrial and Applied Mathematics, 2001.
- [13] E. Ledoux, Modèles mathématiques en hydrogéologie, Centre d'Informatique Géologique École Nationale Supérieure des Mines de Paris, 2003.
- $[14]\,$ G. de Marsily, Cours~d'hydrog'eologie, Universit\'e Paris VI, Septembre 2004 .
- [15] Guillaume CARLIER, ANALYSE FONCTIONNELLE, Notes de cours ENS, 2009-2010.
- [16] Philippe G. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, 1980.
- [17] Owe Axelsson, Vincent A. Barker, Finite element solution of boundary value problems: theory and computation, Society for Industrial and Applied Mathematics, 2001.
- [18] Alexandre Ern, Jean-Luc Guermond, Éléments finis: théorie, applications, mise en œuvre, Numéro 36 dans Mathématiques et Applications. Springer, 2002.
- [19] D.E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Reading MA Addison Wesley, 1989.
- [20] D. Ortiz-Boyer and C. Hervs-Martinez and N. Garc?a, A crossover operator for evolutionary algorithms based on population features, Journal of Artifical Intelligence Research, 2005.
- [21] P. G. Ciarlet, Introduction ? l'Analyse Numérique Matricielle et á l'Optimisation, Masson, Paris, 1982.