Non-parametric identification of geological models

Marc Schoenauer[1], Andreas Ehinger[2] and Bertrand Braunschweig[3]
(1) : CMAP, CNRS-URA 756, Ecole Polytechnique, Palaiseau, France
(2) : Geophysics and Instrumentation, IFP, Pau, France
(3) : Computer Science and Applied Mathematics, IFP, Rueil-Malmaison, France
E-mails : Marc.Schoenauer@polytechnique.fr, {andreas.ehinger,bertrand.braunschweig}@ifp.fr

Abstract

Many problems to be solved in geophysical processing can be expressed in terms of identification of spatial geological models : given a function ϕ applied to a geological model γ, producing a result $\mathcal{R}$, the problem is to find γ^* such that $\phi(\gamma^*) = \mathcal{R}^*$, where $\mathcal{R}^*$ is the expected result : a seismogram, a pressure curve, a seismic cross-section etc.

The presented research deals with the joint use of evolutionary algorithms and Voronoï diagrams to address some non-parametric *instances of identification problems in geophysics, i.e. without a priori hypothesis about the geometrical layout of possible solutions. In this paper, a first application in velocity determination for seismic imaging demonstrates the ability of this approach to identify both the geometry and the velocities of the underground from experimental seismograms.*

1 Introduction

1.1 The geophysical problem

A seismic experiment starts with an artificial explosion at some point near the surface. Elastic waves propagate through the underground medium, eventually being reflected by interfaces between different media (see Figure 1). The reflected waves are measured at some points of the surface by some receptors recording the surface movement along time, called *seismograms*. The inverse problem then consists in determining the underground medium from the recorded seismograms.

For geophysical applications, the first subsurface parameter to be inverted for is the propagation velocity of compressional waves. Once this velocity field is known, classical depth imaging processes can be used to retrieve a correct seismic image of the underground that can then be interpreted by geologists.

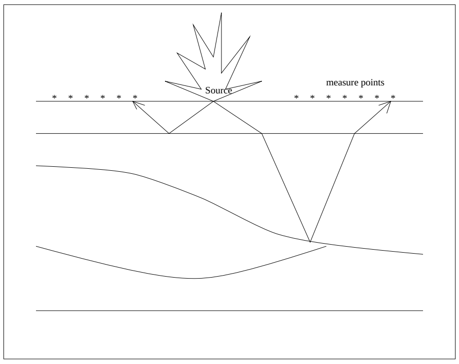


Figure 1: *A two-dimensional view of a typical seismic-reflection experiment.*

A straightforward approach to velocity model determination consists in defining an optimization problem: try to find an underground model such that the difference between modeled data and the data recorded in the field is minimal. For cost efficiency, the forward modeling can be approached by using an acoustic wave equation (as opposed to a fully elastodynamic wave equation). This optimization problem is known to be highly non-linear. A global optimization strategy is thus required to solve it.

1.2 Evolutionary approaches

A number of authors recently addressed seismic processing problems with EA techniques. In [17], a genetic algorithm with multiple crossover locations is developed for the inversion of plane-wave seismogram; in this work, the underground is represented by a user-defined series of layers. In [3], the 'pseudo-subspace' method for the inversion of geophysical data is proposed: the complexity of models is progressively increased, but following a user-defined evolution.

When applying evolutionary algorithms techniques to the problem of seismic velocity determination, one has to consider the following requirements :

- the velocity model must be defined by a sufficiently small number of parameters in order to have generations of a reasonable population size: for instance, it has been heuristically demonstrated [7] and theoretically proved [4, 5] that the population size of a binary genetic algorithm should increase linearly with the number of bits in the bitstrings;

- the performance index (the fitness function) must be fast, stable in the sense that optimal performance corresponds to the optimal model, and sensitive to non-optimality.

One of the possible choices for addressing the first requirement is to use a predefined model for the underground, as the layer-model in [17]. However, this approach relies on strong hypotheses on the geometrical layout of the underground. Another current alternative used to represent the underground with a smaller number of parameters is the spline model. however, control points for the splines must be equally distributed in the underground domain in the absence of problem-specific knowledge.

Our approach differs as it uses a non-parametric approach based on the Voronoï representation for modeling the subsurface: the complexity of the representation (the number of "parameters") is not only reduced, but also locally tuned by the algorithm itself, leading to a greater flexibility.

All these types of models can easily be translated into a regular mesh which is the input data to the finite difference wave propagation algorithm. As an example, a grid of 400x400=160,000 points can be represented as a full 160,000-long vector in the full parametric approach, packed into 40x20=800 spline coefficients, or into a few dozens of Voronoï cells, each involving 3 real coefficients (Section 2 below).

2 Voronoï representation

The underlying hypothesis in this paper is that the underground domain is made of homogeneous regions, i.e. regions where the velocity is constant. Mathematically, the velocity is supposed piecewise constant in the underground. The underground domain is thus partitioned into regions of constant velocity, and the

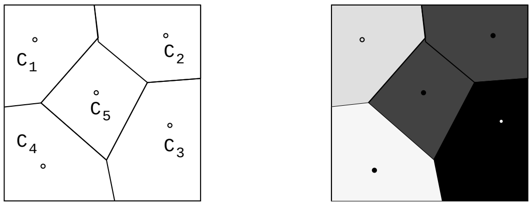


Figure 2: *The Voronoï representation: Voronoï sites define a partition of the domain into polygonal subsets in (a). Associate a velocity v_i to each site C_i to obtain the piecewise constant function shown in (b).*

idea is to evolve both the partition and the velocities.

A possible way of representing partitions of a given domain comes from computational geometry, more precisely from the Voronoï diagram theory (see [11, 2] for a detailed introduction to algorithmic geometry and Voronoï diagrams).

The representation of partitions by Voronoï diagrams to describe their evolutionary has been introduced to tackle Optimum Design problems [13] and identification problems [15] in Structural Mechanics. However, both these applications used boolean partitions of the design domain, whereas real-valued coefficients are needed here.

2.1 Voronoï diagrams

Consider a finite number of points $V_0, \dots, V_N$ (the *Voronoï sites*) of a given subset of $\mathbb{R}^n$ (the design domain). To each site V_i is associated the set of all points of the design domain for which the closest Voronoï site is V_i, termed *Voronoï cell.* The *Voronoï diagram* is the partition of the design domain defined by the Voronoï cells (see Figure 2-a). Each cell is a polyhedral subset of the design domain, and any partition of a domain of $\mathbb{R}^n$ into polyhedral subsets is the Voronoï diagram of at least one set of Voronoï sites.

2.2 The genotype

Consider now a (variable length) list of Voronoï sites, in which each site C_i is given a real valued coefficient v_i. The corresponding Voronoï diagram represents real-valued piecewise constant function if all points in each Voronoï cell are given the velocity value of the associated site (here the Voronoï diagram is supposed regular, i.e. to each cell corresponds exactly one site). For instance, the genotype of the example of Figure 2 is made of 5 sites (described by their 2 coordinates) and the associated 5 velocities.

Note that this representation does not depend in any way on the discretization that will be used to compute the response of the model to wave propagation. Furthermore, Voronoï diagrams being defined in any dimension, the extension of this representation to $\mathbb{R}^3$ and $\mathbb{R}^n$ is straightforward.

An important remark is that this representation presents a high degree of *epistasis* (the influence of one site on the physical shape is modulated by all neighbor sites). This will be discussed in more details in Section 3.4.

2.3 Evolution operators

In order to apply evolutionary algorithms to the Voronoï representation, one needs to define some evolution operators, namely crossover and mutation operators.

The evolution operators on the real-valued Voronoï representation are very similar to those defined on the boolean instances of Voronoï representations: the crossover takes into account the geometrical proximity of Voronoï sites (see [10, 14] for a detailed discussion of those operators); and the mutation operators are standard Gaussian mutations for real-valued evolutionary algorithms coming from Evolution Strategies [16]. More precisely,

- The crossover operators exchange Voronoï sites on a geometrical basis, as illustrated by Figure 3. Moreover, this mechanism easily extends to any dimension [8].
- A first mutation operator performs a Gaussian mutation on the coordinates and/or on the velocities of the sites. As in Evolution Strategies, the standard deviation of that Gaussian mutation is adaptive, using Rechenberg's $1/5^t h$ rule [12], or self-adapted, as proposed by Schwefel [16].
- Variable-length representations require mutation operators performing random addition and destruction of some sites.

3 Experiments and results

3.1 The fitness function

The fitness of a tentative solution – a velocity model represented by a list of Voronoï sites – is obtained by a comparison between some reference seismograms and seismograms obtained under the same experimental conditions but using the underground velocity model at hand.

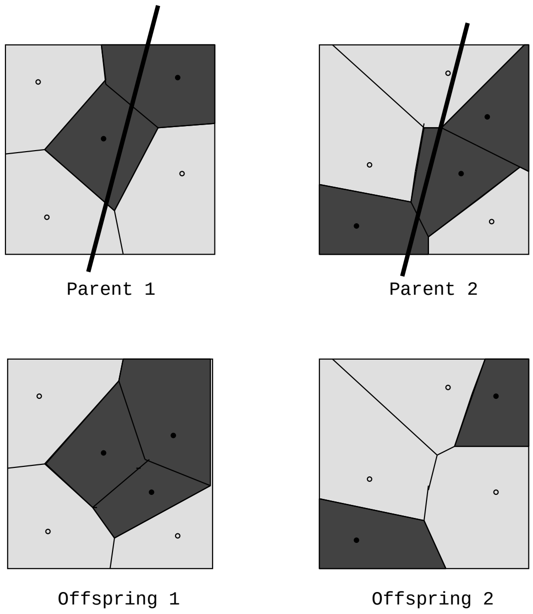


Figure 3: *The Voronoï representation crossover operator. A random line is drawn across both diagrams, and the sites on one side are exchanged.*

Hence the computation of the fitness goes through a numerical simulation of the wave equation in the – finite – underground domain. A finite difference method has been used, together with artificial absorbing boundary conditions on the underground boundaries to simulate an infinite domain. All technical details about the numerical method can be found in [6].

For these preliminary experiments, the "experimental" seismograms were actually computed using the same numerical simulation on a known velocity model: though using thereafter real-world data can show significant differences in the behavior of the algorithm, such "cheating" allows better understanding of the behavior of the algorithm.

A – coarse – 10×10 discretization was chosen for the $1000m \times 1000m$ underground domain. The Ricker-shaped source was supposed to lie at $20m$ depth in the middle of the domain. The 10 seismograms at each surface node were "recorded" during $1s$, with time steps of $1/1000s$. All velocities are given in meter per second and rounded to integer values in the following figures. All runs were allowed 10000 fitness evaluations, which took about 5-6 hours on a Pentium200.

3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
3000	3000	3000	3000	3000	3000	3000	3000	3000	3000
3000	3000	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	3000	3000
3000	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	3000
5000	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**	**5000**
5000	**5000**	1500	1500	1500	1500	1500	1500	1500	**5000**
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500

Figure 4: *The "dome" reference velocity model, on the* 10×10 *discretized domain. The bold has been added only to help distinguishing the three regions.*

Three test cases were chosen: a two layer model, with interface at half depth, velocity 2000 and 3000 respectively above and below that line; a "corner" model, in which the lower right quarter of the domain has velocity 5000 and the remaining of the domain velocity 3000; And a simplified "dome" model, made of three different regions, resembling interesting regions of the underground as far as petroleum prospection is concerned (see Figure 4).

3.2 The evolutionary algorithm

A standard genetic algorithm scheme was used in these preliminary experiments: parents are selected for reproduction by a 2-tournament, generate offspring with crossover rate of 0.6 and mutation rate of 0.3. Mutation of the velocities and the positions of the sites are 3 times more frequent than addition or destruction of a site. The Gaussian mutations of the velocities and the positions of the sites are adjusted following the $1/5^th$ rule [12]. All offspring replace all parents, with elitism.

These parameters were adjusted by running the same algorithm using the same representation, but with the – computationally much cheaper – fitness of least square comparison with the target model. Only the population size was lowered when it came to the actual seismographic fitness, and a small population size of 5 was preferred, as the number of fitness evaluations had to be severely limited: Larger populations proved better long-term results on the simplified fitness, but for as many as 100000 evaluations. A few runs with population size of 50 seemed to confirm those results. Further details can be found in [1]. For the above parameter settings, 10 independent runs were performed on each of the three test cases.

3010	3000	3000	3000	3000	3000	3000	3000	3000	3000
3075	3010	3000	3000	3000	3000	3000	3000	3000	3000
3075	3010	3010	3000	3000	3000	3000	3000	3000	3000
3075	3035	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	3000	3000
3075	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**
5028	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**	**5028**
1500	1500	1500	1500	1500	1500	1500	1500	1500	**5028**
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500
1500	1500	1500	1500	1500	1500	1500	1500	1500	1500

Figure 5: *The best dome obtained in 10 independent runs of 10000 fitness evaluations*

3743	2999	2999	2999	2999	2999	2999	3140	3140	3140
3743	3743	2999	2999	2999	2999	2999	3140	3140	3140
3743	3743	3743	2999	2999	2999	2999	3140	3761	3761
3743	3743	3743	3743	**5174**	**5174**	**5174**	3761	3761	3761
3743	3743	3743	2333	**5174**	**5174**	**5174**	3761	3761	3761
3743	3743	3743	2333	**5174**	**5174**	**5174**	**5174**	3761	3761
3743	4298	4298	2333	2333	**5174**	**5174**	**5174**	**5174**	3761
4298	4298	4298	2333	2333	**5174**	**5174**	**5174**	1563	1563
4298	4298	4298	2333	2333	**5174**	**5174**	1563	1563	1563
4298	4298	4298	2333	2333	2333	1563	1563	1563	1563

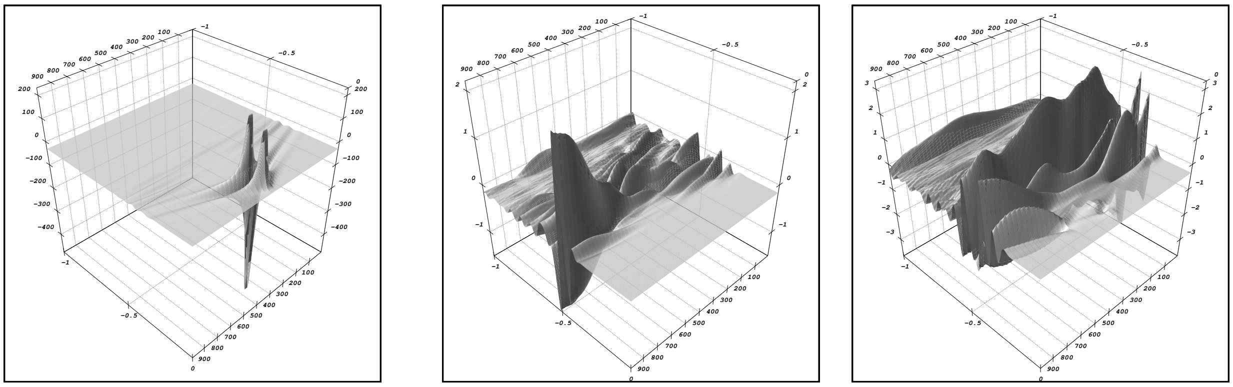
Figure 6: *The worst dome obtained in 10 independent runs of 10000 fitness evaluations*

3.3 First results

For the two-layer model, all runs but one gave an almost perfect result: perfect localization of the interface between both layers and velocities within 5 m/s from the actual values. Moreover, they all had only 2 Voronoï sites, even though no *parsimony* factor was added to prevent too many sites to appear. On the other hand, the unsuccessful run ended up very far from the target, with 24 Voronoï sites.

For the corner model, only 3 runs found a best model with exact localization of the corner, and velocities within 10 m/s of the actual values. However, all runs found almost the right value for the velocity of the first 200m in depth. Moreover, 5 runs show something like an "extended" corner with high velocity, and all runs give fairly good results compared to the worst run of the layered model. The number of Voronoï sites of the best solution of each run seems somehow related to its fitness: 2 of the best 3 runs have the optimal 3 sites while the other uses 5 sites.

The dome case is even more difficult: only one run identified almost exactly the dome-structure of the model (see Figure 5), using 6 Voronoï sites. The worst of the 10 runs is also shown in Figure 6, as a matter of comparison: even some points close to the surface are not given an accurate velocity. Note that this worst result uses only 8 Voronoï sites, while, for instance, the second best result has as many as 18 Voronoï sites. Moreover, allowing more fitness evaluations in that case considerably improved the result.



(a): *The "experimental" seismograms* (b): *Error for the best solution* (c): *Error for the worst solution*

Figure 6: *Plot (a) shows the pressure at the surface [0,1000], along time [0,1] after the explosion, for the dome model. Actual seismograms would be cross-sections of that plot at given surface positions. Plots (b) and (c) are the errors on the seismograms when using the best solution of Figure 5 and the worst solution of Figure 6 respectively. Notice the range of the pressures in plots (b) and (c), compared to that of plot (a).*

But the most puzzling issue arose when comparing the seismograms of both the exact model with those obtained by both the best and the worst results of the 10 runs. Figure 6 shows the "experimental" seismograms, together with the differences plots. Please note the scale of these plots: whereas the values of the seismograms range from -400 to 200, the differences do not go above 3 in the worst case! This seems to indicate that the inverse problem, at least with such coarse discretization and artificial data, has many local quasi-optima.

3.4 Discussion and future work

These first results first justify the use of the non-parametric Voronoï representation: the resulting algorithm did indeed discover both the geometry of the underground and the associated velocities.

The main limitation remains the computing cost of this method: Ongoing work is concerned with finer discretizations of the underground domain during the computation of the fitness, in order to achieve better accuracy (and to hopefully reduce the number of quasi-optimal different solutions for a given set of seismograms). But a more powerful computer (e.g. a parallel super-computer) will have to be used.

Of course another mandatory further work is to use real experimental data rather than the above "cheated" artificially generated data.

One interesting remark about the above results is that the best results have been obtained using a rather high mutation rate (0.3). It might be that the velocities in the real-valued Voronoï representation can only be adjusted by the corresponding mutation operator. This draws a very different picture than, for instance, in the parametric real-valued evolutionary algorithms, where the crossover (e.g. the arithmetic crossover in real-valued GA, or the intermediate crossover in ES) also contribute to the precise adjustment of the real coefficients. It might be useful to design another crossover operator that would actually merge also the velocity values of two Voronoï diagrams.

The presented representation is the first actual extension to the real valued case of the boolean Voronoï representation already successful in Structural Mechanics [13, 15]. In that line, another representation for boolean partition of a given domain is also under investigation, based on variable length list of rectangles, each adding a component to the velocity of the points it covers. But having more than one possible representation for the same problem raises the question of the a priori choice of a representation for a given instance of problem: Another line of research deals with the study and comparison of problem difficulties [9].

4 Conclusion

This paper has introduced an original approach to the problem of identification of the velocity of underground regions from seismograms, based on the representation of the underground model by Voronoï diagrams.

From the geophysical point of view, this approach does not make any strong a priori hypothesis on the geometry of underground model (e.g. "it is composed of 3 homogeneous layers'): the Voronoï representation seems suited to any piecewise constant velocity model. Moreover, the presented results have demonstrated that artificial models like the corner model, or the simplified dome model, can indeed be identified both geometrically and in terms of velocity values.

On the other hand, these results also show that the inverse problem of velocity identification from seismograms can have many different quasi-optimal solutions, i.e. different layout of the underground leading to similar responses. More experimental results (e.g. involving recordings from two explosions), finer discretizations and/or longer runs of the evolutionary algorithm are probably necessary to reach the global optimum with more reliability in the case of complex geometries.

From the evolutionary point of view, it brings yet another argument to the importance of the choice of a representation for a given problem: being independent of any discretization of the underground, the variable length Voronoï representation demonstrates an self-adaptive complexity, that will be adjusted by the algorithm itself for each new instance of problem.

Acknowledgments

The authors wish to thank F. Clément and P. Joly from INRIA who kindly provided the program solving the acoustic wave equation with absorbing boundary conditions.

References

[1] S. Berthebaud. Algorithmes génétiques et diagrammes de voronoï : Optimisation et identification de modèles géologiques. Rapport de DEA - CMAP - Ecole Polytechnique, Sept. 1997.

[2] J.-D. Boissonnat and M. Yvinec. *Géométrie algorithmique.* Ediscience International, 1995.

[3] F. Boschetti. *Application of genetic algorithms to the inversion of geophysical data.* PhD thesis, University of Western Australia, 1995.

[4] R. Cerf. *Une théorie assymptotique des algorithmes génétiques.* PhD thesis, Université de Montpellier II, March 1994.

[5] R. Cerf. An asymptotic theory of genetic algorithms. In J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution*, volume 1063 of *LNCS*, pages 37–53. Springer Verlag, 1996.

[6] F. Clément. *Une formulation en temps de parcours par migration pour la détermination des vitesses de propagation acoustique à partir de données sismiques bidimensionnelles.* PhD thesis, INRIA - TU-0398, Dec. 1994.

[7] D. E. Goldberg, K. Deb, and J. H. Clark. Genetic algorithms, noise and the sizing of populations. *Complex Systems*, 6:333–362, 1992.

[8] A. B. Kahng and B. R. Moon. Toward more powerful recombinations. In L. J. Eshelman, editor, *Proceedings of the* 6^{th} *International Conference on Genetic Algorithms*, pages 96–103. Morgan Kaufmann, 1995.

[9] L. Kallel and M. Schoenauer. Fitness distance correlation for variable length representations. Technical Report 363, CMAP, Ecole Polytechnique, November 1996. Submitted.

[10] C. Kane and M. Schoenauer. Genetic operators for two-dimensional shape optimization. In J.-M. Alliot, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution*, number 1063 in LNCS. Springer Verlag, Septembre 1995.

[11] F. P. Preparata and M. I. Shamos. *Computational Geometry: an introduction.* Springer Verlag, 1985.

[12] I. Rechenberg. *Evolutionstrategie: Optimierung Technisher Systeme nach Prinzipien des Biologischen Evolution.* Fromman-Holzboog Verlag, Stuttgart, 1973.

[13] M. Schoenauer. Representations for evolutionary optimization and identification in structural mechanics. In J. Périaux and G. Winter, editors, *Genetic Algorithms in Engineering and Computer Sciences*, pages 443–464. John Wiley, 1995.

[14] M. Schoenauer. Shape representations and evolution schemes. In L. J. Fogel, P. J. Angeline, and T. Bäck, editors, *Proceedings of the* 5^{th} *Annual Conference on Evolutionary Programming*, pages 121,129. MIT Press, 1996.

[15] M. Schoenauer, L. Kallel, and F. Jouve. Mechanical inclusions identification by evolutionary computation. *European Journal of Finite Elements*, 5(5-6):619–648, 1996.

[16] H.-P. Schwefel. *Numerical Optimization of Computer Models.* John Wiley & Sons, New-York, 1981. 1995 – 2^{nd} edition.

[17] P.L. Stoffa and M.K. Sen. Nonlinear multiparameter optimization using genetic algorithms : inversion of plane-wave seismograms. *Geophysics*, 56, 1991.