Inverse Problems **14** (1998) 427–440. Printed in the UK PII: S0266-5611(98)92484-4

Exploring multidimensional landscapes without a map

Malcolm Sambridge

Research School of Earth Sciences, Institute of Advanced Studies, Australian National University, Canberra, ACT 0200, Australia

Received 13 March 1998

Abstract. A discussion of methodologies for nonlinear geophysical inverse problems is presented. Geophysical inverse problems are often posed as optimization problems in a finite-dimensional parameter space. An Earth model is usually described by a set of parameters representing one or more geophysical properties (e.g. the speed with which seismic waves travel through the Earth's interior). Earth models are sought by minimizing the discrepancies between observation and predictions from the model, possibly, together with some regularizing constraint. The resulting optimization problem is usually nonlinear and often highly so, which may lead to multiple minima in the misfit landscape. Global (stochastic) optimization methods have become popular in the past decade. A discussion of simulated annealing, genetic algorithms and evolutionary programming methods is presented in the geophysical context. Less attention has been paid to assessing how well constrained, or resolved, individual parameters are. Often this problem is poorly posed. A new class of method is presented which offers potential in both the optimization and the 'error analysis' stage of the inversion. This approach uses concepts from the field of computational geometry. The search algorithm described here does not appear to be practical in problems with dimension much greater than 10.

1. Introduction

All of our present knowledge about the Earth's interior is derived from observations collected at the surface. Therefore measurements are always indirect, and there exists an inverse problem in retrieving information about physical properties at depth.

Inverse problems are not restricted to the geo-sciences, and occur in nearly all scientific disciplines, e.g. medical imaging and nondestructive testing. In some cases the similarity between inverse problems has allowed innovative techniques to be applied to diverse situations. For example the success of x-ray (and later acoustic) tomography in medical imaging, in the 1970s, inspired the early attempts at two- (2D) and three-dimensional (3D) seismic tomography of Earth structure in the early 1980s. (See Nolet 1987, Thurber and Aki 1987, Iyer and Hirahara, 1993 for summaries. See also Trampert 1998.) Seismic tomography has since become one of the most commonly used techniques in both exploration and global seismology.

The amount of reliable information that can be retrieved on the Earth's structure depends heavily on the limitations of the underlying inversion techniques. Two important features of nearly all geophysical inverse problems is that they are nonunique and nonlinear. Non-unique means that more than one set of unknowns, and usually an infinite class, fit the data equally well and there is no direct way to decide which is closer to the truth. Nonlinear means that the observations (e.g. the arrival times of prominent phases on a seismogram) and the parameters of interest (e.g. seismic properties of rocks at depth) are such that small

0266-5611/98/030427+14$19.50

variations in the latter may induce large variations in the former. Each of these factors pose severe problems for characterizing the solution of an inverse problem (see Snieder 1998).

In the past two decades geophysical inversion has consisted, to a large extent, of the development of linearized techniques for nonlinear problems. Linearized techniques iteratively improve upon an initial guess at the unknown parameters using local derivative information (see Menke 1989). Often derivative calculations are very expensive and unfortunately there is no guarantee that the iterations will converge to an acceptable model (i.e. set of unknowns). Even if convergence is achieved the resulting model may be far from the true model, have a limited fit to the data, and a strong dependence on the initial guess at the unknown parameters. If the level of nonlinearity is severe, then two or more entirely different classes of model may exist which satisfy the data. In this case linearized methods usually fail, and more global (direct sampling) approaches must be used. (Reviews of geophysical inverse theory for linearizable problems can be found in Parker (1977), Tarantola (1987) and Menke (1989).)

Direct inversion methods based on sampling of a parameter space avoid the limitations of linearization and require only that a model can be assessed for its fit to the data. The set of unknowns is represented as a model vector in a parameter space. In nearly all cases a simple scalar misfit is used to measure the discrepancy between observations and theoretical predictions from the model. The misfit function forms a multidimensional 'landscape' through which an algorithm must navigate. Usually the inherent nonuniqueness of the problem and presence of errors in the data mean that if one model fits the data then an infinite number of models will fit the data. In a direct sampling approach one (hopes) to tackle the nonuniqueness of the problem explicitly by seeking out all classes of model that fit the observations. The second stage of the inversion would then be to extract information on model parameters from examining this complete class of potential solutions. For example this might be achieved by trying to determine properties or features shared by all satisfactory models which have been obtained in the search stage.

In practice most problems involve a large number of unknowns (i.e. tens–thousands of parameters), and even with a crude discretization of each parameter axis, this quickly leads to an astronomical number of feasible models, sometimes up to 10^{100}. In these cases it is not possible to sample all models and search techniques for satisfactory data fit based on randomized (stochastic) sampling are always more efficient. If the cost of evaluating the misfit function is large and the dimension of the space high, it becomes increasingly difficult to address questions of error analysis, nonuniqueness and trade-offs between unknowns. Often one is reduced to merely finding the best data fitting model and estimate errors using linearizing approximations. The question of how best to deal with the 'error analysis' stage of the inversion (or which hypotheses to test) in cases where direct sampling methods are appropriate is an active area of research.

This paper is concerned with stochastic sampling methods, usually called Monte Carlo methods. By this we mean any algorithm which uses probabilistic decisions to search a parameter space, usually for models with higher data fit. In the next section we summarize the development and use of these methods as applied to geophysical problems. The methods popular in the geo-sciences can be (loosely) divided into three types, namely, simulated annealing, genetic algorithms, and evolutionary programming. We briefly describe and compare each approach. In the subsequent section we discuss the underlying concepts upon which these methods are based and this motivates us to present a new class of stochastic sampling method which possesses some features similar to the above methods, but is conceptually simple and easier to use. To illustrate the concepts behind the new approach we design a simple algorithm and demonstrate its use with a numerical example.

We do not expect this particular algorithm to be practical for problems with dimension much greater than 10. However, the general concepts appear to hold much potential for problems of this kind. It is our aim to provide an overview of the previous work and suggest some future directions in direct sampling methods, as applied to geophysical inverse problems. In this way it is hoped that readers in other fields will contribute to the argument.

2. Monte Carlo methods in geophysics

Monte Carlo methods are not new to geophysics. Keilis-Borok and Yanovskaya (1967) first presented the approach, Press (1968) used a uniform Monte Carlo sampling procedure to derive one-dimensional (1D) density profiles within the Earth from free oscillation data, and Wiggins (1969) applied a Monte Carlo procedure to the inversion of seismic body-wave data. Anderssen (1970), and Anderssen and Senata (1971) considered its role in addressing questions of nonuniqueness and interpretation of results. These early works stimulated interest among geophysicists in Monte Carlo methods. However, they were largely based on uniform sampling of the parameter space which becomes inefficient when the dimensionality of the space (i.e. number of unknowns) is large.

The method known as simulated annealing (Kirkpatrick *et al* 1983) was introduced to geophysicists by Rothman (1985, 1986), and has become very popular. For a recent survey of geophysical papers see Sen and Stoffa (1995), and for detailed descriptions see Aarts and Korst (1989). This approach is motivated by a natural optimization process, namely thermodynamic annealing in crystalline solids. The basic method is to find the global minimum of the misfit function by generating a sequence of models in a way that mimics thermal equilibrium. A control parameter (usually called temperature) is used to control the sequence, and by 'slowly' lowering its value the system settles into its global minimum (in misfit). The powerful feature of this algorithm is its demonstrated ability to jump out of local minima in the data misfit function.

A 'move class' algorithm is needed to generate the next model in the sequence. Two types are common, known as the *Metropolis algorithm* (Metropolis *et al* 1953) and the *heat bath* algorithm (Geman and Geman 1984). In the former the next model in the sequence is generated by perturbing the previous model in a restricted region, which is usually along one parameter axis, say the ith. The perturbed model is accepted if its fit to the data is improved. Otherwise it may still be accepted depending on the result of a random decision. The probability of acceptance is determined by the temperature value. At high temperature poorer fits are readily accepted, while at lower temperature they are less likely to be accepted. If a perturbation is rejected then the previous model is retained and the sequence begins again, usually by considering the next, $(i + 1)$th parameter axis in the same way.

In contrast to the Metropolis algorithm, which involves two stages, the heat bath is a one-step procedure. In this case the next model vector in the sequence is chosen by assigning a probability density function (PDF) over a limited region about the current model. Again the most common type is along the ith parameter axis passing through the current model. A perturbed model is generated by drawing a random number according to this distribution and replacing the ith parameter with this new value. The process is then repeated for the next parameter and so on. For each unknown the probability density function depends on both the misfit value and the current temperature. Usually it is constructed by evaluating the misfit function at discrete intervals along the parameter axis. Note that at each step of the heat bath method the misfit function must be calculated a number of times, but a new model (with one parameter changed) is always generated, whereas in the Metropolis method the misfit function is evaluated only once, but the new model may be rejected. Rothman (1986)

concluded that for the residual statics problem in seismology, the heat bath approach was superior because of the extremely high number of rejections produced by the Metropolis algorithm. In both cases the misfit function must be generated a large number of times before a model vector is obtained with all its elements changed. After one cycle through all axes the heat bath approach produces a complete new model which is considered to be statistically uncorrelated with the previous model, whereas in the Metropolis approach the models are still correlated.

Many variants of the simulated annealing algorithm have been developed, (e.g. Ingber 1989), and work has been done on finding optimal strategies for lowering the temperature parameter (Nulton and Salamon 1988, Andersen *et al* 1988). A recognized drawback is the phenomenon known as 'critical slowing down' which occurs because the computation time of the algorithm can become dominated by the grid size used. Strategies have also been put forward to overcome this effect (Brower *et al* 1989).

The second type of direct sampling method to attract the attention of geophysicists was genetic algorithms. Originally developed by Holland (1975), they were popularized for geophysical problems by Stoffa and Sen (1991), Gallagher *et al* (1991), Wilson and Vasudevan, (1991) Sen and Stoffa (1992) and Sambridge and Drijkoningen (1992). Again many variants exist, but the basic approach is to drive an entire population of models towards better data fit simultaneously. Again stochastic decisions are involved, and the algorithm is motivated by another natural optimization process, namely evolution. One iteration of the basic method involves three stages usually called *selection*, *crossover* and *mutation*. Each of which involve some 'tuning' parameters to be chosen by the user. The overall effect is to produce a new population of offspring models which in some sense possess properties of the successful models of the previous generation (i.e. the ones with better data fit). One iteration of the procedure requires an evaluation of the misfit function for each model in the old population. The inclusion of a form of evolutionary pressure, or 'survival of the fittest', drives the population towards lower regions of the misfit landscape. The mechanisms used to implement selection pressure and combine characteristics of models can be crucially important to the overall effectiveness of the algorithm. For a discussion on many of the relevant issues see Goldberg (1989).

An important difference between this type of method and simulated annealing is that it works by driving a population of models simultaneously and not through a sequence of perturbations to a single model. This basic difference indicates that genetic algorithms try to make use of a more global sampling of the parameter space at each stage, whereas simulated annealing uses quasilocal information to meander through the misfit landscape in search of better fitting models. This is a general observation, it does not imply that one will necessarily be a better optimizer than the other, but merely that they work in different ways. Several comparisons have been made between the two approaches (see Ingber and Rosen 1992, Scales *et al* 1992, Sen and Stoffa 1995). Conclusions on relative performance differ between applications and authors. An unsatisfactory feature of many of the variants of simulated annealing and genetic algorithms is that they inevitably introduce extra tuning parameters, or design features, that must be chosen in an *ad hoc* manner. This is particularly true of genetic algorithms. Usually proposed variants work well in a particular problem but may not translate well to other inverse or global optimization problems.

Recently evolutionary programming methods (Fogel 1962, Fogel *et al* 1966) have begun to find applications in geophysics. Minster *et al* (1995) have used the approach for earthquake hypocentre location. This method is similar to a genetic algorithm in many respects except for an absence of the crossover stage. At each iteration a population of models are perturbed often using a Gaussian random variable in either one or more

dimensions of the parameter space. Models are then selected in a probabilistic manner (using data fit) from the set of perturbed models and the original (parent) models. As with a genetic algorithm competition forces the fitter models to survive on average while the poorer models in the population are removed. The absence of the crossover stage means that information is not shared between models, nevertheless it seems that the approach is capable of navigating towards good data fitting models in landscapes with multiple minima. However, at present experience with this approach in geophysical problems is very limited.

3. Ensemble inversion rather than optimization

In the discussion of direct sampling methods above, the objective of the inverse problem was stated as to somehow 'extract information from all models found that fit the data adequately'. Global optimization techniques were presented as ways of generating an ensemble of 'good' models. If one or more models can be found which fit the data adequately then it is tempting to simply examine the best data fitting model. However, it has long been known that this is at best only one part of an inverse problem (and in some cases it may be a meaningless part). The nonuniqueness, trade-offs and constraints placed on the estimated unknowns are often more important to understand than the features of the best data fitting model. It is often more useful to answer questions such as 'What properties do all acceptable models share?' rather than 'What does the best model look like?' (Parker 1977). Kennett and Nolet (1978), Kennett (1978), Dosso and Oldenburg (1991) and Vasco *et al* (1993) have all taken this view and proposed methods for analysing the ensemble of acceptable models, primarily using cluster analysis techniques.

An alternative framework used to address 'error analysis' issues is Bayesian inference. This statistical view of inverse problems is fully described by Tarantola (1987) and Duijndam (1988a, 1988b). In essence the problem becomes one of replacing the misfit landscape with an *a posteriori* probability density (PPD) function on the unknown parameters. This function incorporates all *a priori* information and our best estimates of all types of error present in the problem. Quantities of interest are then expressed as integrals over this multidimensional space. For example each element of the *a posteriori* model co-variance matrix becomes,

$$C_{i,j} = \frac{1}{\nu} \int_M m_i m_j f(\boldsymbol{m}) \, \mathrm{d}\boldsymbol{m} - \langle m^i \rangle \langle m^j \rangle \tag{1}$$

where $\langle m^i \rangle$ denotes the mean value of the ith unknown, which is itself an integral,

$$\langle m^i \rangle = \int_M m_i f(\boldsymbol{m}) \, \mathrm{d}\boldsymbol{m} \tag{2}$$

$f(\boldsymbol{m})$ is the PPD, and ν is the normalizing constant,

$$\nu = \int_M f(\boldsymbol{m}) \, \mathrm{d}\boldsymbol{m}. \tag{3}$$

Each of the sampling methods above results in an ensemble of models, hopefully biased towards the parts of parameter space with low misfit (or high PPD). In this case each model is assigned a value of probability density and contributes to the numerical estimates of the integrals, which contrasts with the previous approach where one analyses only the 'satisfactory models'. Since all of the above integrals are of the form,

$$I = \int_M g(\boldsymbol{m}) f(\boldsymbol{m}) \, \mathrm{d}\boldsymbol{m} \tag{4}$$

where $g(\boldsymbol{m})$ represents the remaining parts of the integrands in equations (1)–(3), then for N samples the Monte Carlo estimate of (4) is,

$$I_{\mathrm{MC}} = \frac{1}{N}\sum_{k=1}^{N}\frac{g(\boldsymbol{m}_k)f(\boldsymbol{m}_k)}{\rho(\boldsymbol{m}_k)} \tag{5}$$

where $\rho(\boldsymbol{m})$ is the density function of the ensemble of models. The closer the density of the samples is to the PPD the faster the Monte Carlo integral will converge to the true integral, i.e. as a function of N. Ideally we would like the sampling density to be equal to the PPD. In this case all integrals of the form (4) can be easily evaluated, otherwise an estimate of the density must be available. Therefore to be useful for the inference problem rather than just the optimization problem the objective of direct sampling methods should be to importance sample the PPD.

A number of authors have estimated integrals of this kind (Cary and Chapman 1995, Sen and Stoffa 1991, Nolte and Frazer 1994, Mosegaard and Tarantola 1995, Lomax and Snieder 1995, Mosegaard 1998). In those cases where uniform Monte Carlo sampling is used the density is just a constant and disappears from the integrals, however, the sampling becomes extremely inefficient in high dimensions. For genetic algorithms the density distribution of the samples is not known, and will certainly vary, depending on how the algorithm is implemented. For simulated annealing it is known that under ideal conditions an ensemble can be collected which is distributed according to the PPD (see e.g. Rothman 1985). When viewed in this light the simulated annealing is often called a Gibb's sampler or Markov chain Monte Carlo technique (see Sen and Stoffa 1995, Mosegaard and Tarantola 1995). In practice the number of models in the (PPD) ensemble produced by simulated annealing is often very much smaller than the models for which the misfit function has been evaluated. The reason for this 'loss factor' is that for the models to be distributed according to the PPD they must be *statistically independent*. The heat bath algorithm in SA produces independent samples directly, whereas a further selection stage (i.e. pruning) is required to produce independent samples with the Metropolis algorithm. All intervening models are statistically dependent and cannot be used. Both methods, therefore, have an inherent loss factor, which can easily be 1000 to 1 or more depending on the dimensionality of the parameter space.

Methods to generate samples according to a desired multidimensional density function is an active area of research in statistics (for a review see Gelfand and Smith 1990, Smith 1991). A feature of all methods is the high number of PPD evaluations per final model in the ensemble. This inherent difficulty in generating PPD samples has forced geophysicists to look for other, more heuristic means of assessing the constraint on unknowns. These include sequentially displaying a large number of acceptable models (Koren *et al* 1991), plotting the spread of the acceptable data fitting models (Lomax and Snieder 1995, Shibutani *et al* 1996) and other projection methods (Basu and Frazer 1990).

4. The neighbourhood approximation

Here we outline some new concepts in global optimization and inversion which are motivated by the following question: 'How can we make use of all models in parameter space, regardless of their distribution, for which the misfit function has been evaluated?'. All of the methods described above either use inefficient sampling (uniform Monte Carlo, or grid search), produce an unknown distribution (genetic algorithms and evolutionary programming), or throw away a large proportion of samples in order to approximate the distribution as the PPD (simulated annealing or Gibbs sampler). Furthermore, in the last

case there is often little way of telling how accurate the approximation is. The answer to the question is that in order to take all information into account in evaluating integrals of the form (4), then there must be some way of evaluating the density of the distribution at each sample. Density estimation is an under-determined problem. The 2D case is much studied in statistics. However, little work has been carried out in higher dimensions. Here we suggest that the geometrical construct known as a 'Voronoi cell' may provide an answer to this problem.

Voronoi cells have been known for many years (Voronoi 1908). They have been rediscovered in a number of different fields and are primarily studied in the field of computational geometry. A survey of their many properties and variants can be found in Okabe *et al* (1992). Recently they have been used in 2D and 3D geophysical problems (Braun and Sambridge 1995, 1997, Sambridge *et al* 1995). For any distribution of points (nodes) Voronoi cells are defined as the region about each node (in any number of dimensions) which is closer to that node than any other node. For Euclidean distance in 2D they are polygons whose edges are perpendicular bi-sectors between pairs of nodes, in 3D they are convex polyhedra, in higher dimensions they are convex polytopes.

Even though Voronoi cells are conceptually simple, they possess some powerful properties. Regardless of the dimension, or how uneven or anisotropic the nodal distribution may be, the Voronoi cells always form local 'neighbourhoods' about each node, whose size (area, volume, etc) automatically grows, shrinks or changes shape depending on the local nodal density. (Examples in 2D are shown in figures 2(*a*)–(*c*).) This means that they can be used to produce an approximation to the PPD in the entire parameter space, i.e. by setting the value of the PPD in each cell to be equal to the known PPD at the node defining that cell. We call this the 'neighbourhood approximation'. Since the approximation is based on Voronoi cells it takes on their properties, i.e. where the cells are large it represents only long wavelength variations, and where the cells are small (high nodal density) it can rapidly respond to changes in the PPD. Using this approximation we have,

$$\rho(\boldsymbol{m}_k) = V_k^{-1} \tag{6}$$

i.e. the nodal density at any point is given by the reciprocal of the Voronoi cell volume. In which case equation (5) becomes,

$$I_{\mathrm{MC}} = \frac{1}{N} \sum_{k=1}^{N} g(\boldsymbol{m}_k) f(\boldsymbol{m}_k) V_k^{-1}. \tag{7}$$

Evaluation of (7) requires the ratio of the volumes of the Voronoi cells to be determined, which can be done with numerical integration methods (e.g. Flournay and Tsutakawa 1989). However, this is likely to become computationally expensive in high dimensions. This is a nontrivial task and the subject of ongoing work. For the remainder of this paper we restrict our attention to the use of Voronoi cells for the global optimization part of the inverse problem.

Since the neighbourhood approximation is completely space filling, and is uniquely defined for any number of samples, it may be used for global optimization in the following way. We first generate a set of n samples from any method (e.g. uniform sampling) and then generate a new set of n samples so that they are independently distributed according to the current neighbourhood approximation of the PPD. The new samples will be concentrated in the regions of parameter space which have lower misfit (i.e. higher PPD), as given by the current 'best estimate' of the PPD based on all available information. We then update the approximation and generate n more samples. As iterations proceed the PPD approximation

will be refined and each new generation will be concentrated in the 'better' parts of the PPD landscape.

This is a conceptually simple approach containing only one tunable parameter, i.e. the sample size, n. As with previous methods the algorithm is stochastic, but in contrast each new population is directly guided by all previous samples (through the approximate PPD). A number of methods are available for generating a sample according to the approximate PPD (see Gelfand and Smith 1990, Mosegaard and Tarantola 1995). Note that in this procedure the evaluation of the misfit function is effectively replaced with a solution to the nearest neighbour problem, i.e. find the node (previous model) which is closest to a given point in parameter space. In geophysical problems it is almost always much faster to solve the latter than the former. To illustrate how the algorithm works we present a numerical example.

4.1. A numerical example

To illustrate how a neighbourhood algorithm can be used for global optimization we use a simple 2D example (following Lomax and Snieder 1995). Figure 1 shows a synthetic PPD landscape generated using the topography of the Snowy Mountains region in south east Australia. Figure 2(*a*) shows an initial set of 100 samples generated using a Sobol–Antonov–Saleev quasirandom sequence. (This is a better uniform 'random' generator of points than the common pseudorandom number generators commonly in use. See Press *et al* 1992, for details.) At the next iteration 100 new samples are generated (according to the 100 node neighbourhood PPD approximation) using the heat bath algorithm. To do this we start a random walk from the centre of the space and construct the intersection of the PPD with the x-axis passing through this point. (In practice we approximate this 1D function by discretizing the axis and find the Voronoi cell at each point along the axis. In this case we use 100 points along the axis.) We can then generate a new x-value for the current point according to our 1D conditional PPD approximation using standard techniques (Press *et al* 1992). The process is then repeated along the y-axis through the updated point and so on until 100 new points have been generated. The next 100 points are generated in a similar fashion using the first 200 Voronoi cells, and so on.

The situation after 500 new points have been generated is shown in figure 2(*b*), and after 1000 new points in figure 2(*c*). Note that as the new cells are generated they

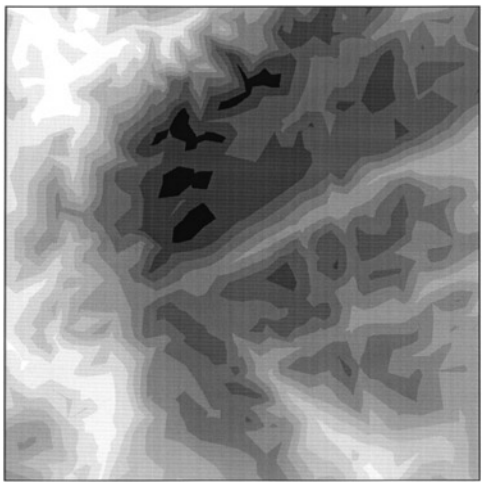

Figure 1. Synthetic PPD surface used in the numerical example. The surface was generated using the topography field in south eastern Australia. Darker contours represent higher values of the PPD. The global maximum lies in the middle of the three dark contours in the central part of the figure.

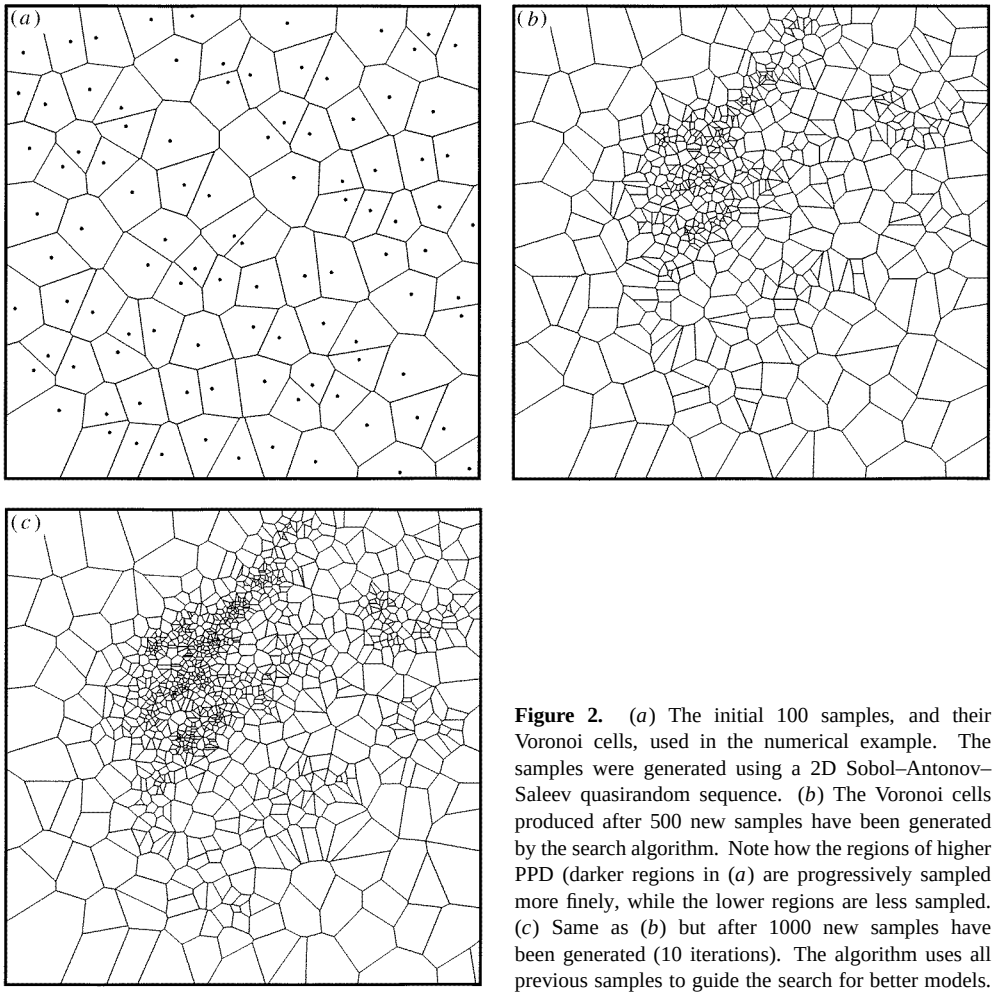

Figure 2. (*a*) The initial 100 samples, and their Voronoi cells, used in the numerical example. The samples were generated using a 2D Sobol–Antonov–Saleev quasirandom sequence. (*b*) The Voronoi cells produced after 500 new samples have been generated by the search algorithm. Note how the regions of higher PPD (darker regions in (*a*) are progressively sampled more finely, while the lower regions are less sampled. (*c*) Same as (*b*) but after 1000 new samples have been generated (10 iterations). The algorithm uses all previous samples to guide the search for better models.

concentrate in regions of high topography (this is a maximization problem). Figure 3 shows the neighbourhood approximation to the landscape and should be compared with figure 1. All main peaks have been identified and the global maximum (middle of three dark contours) closely sampled. A gradual improvement in the approximation of the PPD guides the algorithm to sample about all maxima in the landscape. The first question to ask is how dependent is the performance on the size of the sample? Figures 4(*a*) and (*b*) show the results of the same procedure applied to a sample size of just two ($n = 2$). As one might expect the regions of the space with lower PPD are sampled less, but interestingly enough the algorithm does not appear to have become trapped in any one local maximum. All are sampled equally well. As n increases the algorithm will favour exploration of the space over exploitation of the current information in the PPD. However, in this 2D example the lower PPD parts of the landscape are reproduced crudely while the higher parts are sampled with increasing resolution (figure 4(*a*)). In this case then the sample size does not appear to have affected the overall performance.

This is merely a simple 2D example that illustrates how the algorithm works. It is designed to importance sample the PPD. Optimization is, in a sense, a byproduct of this.

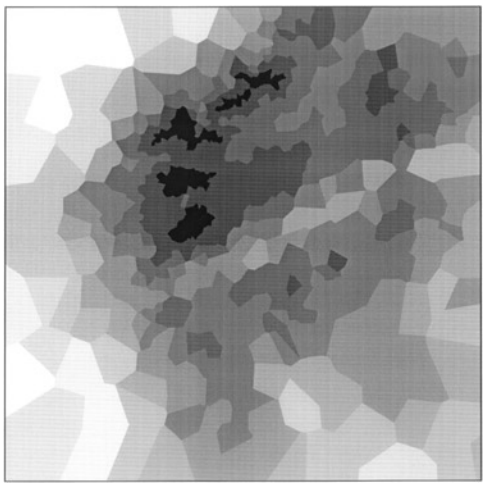

Figure 3. The approximate PPD produced by the 1100 samples in figure 2(*c*). Compare with the true PPD in figure 1. The resolution is increased in the regions with higher PPD.

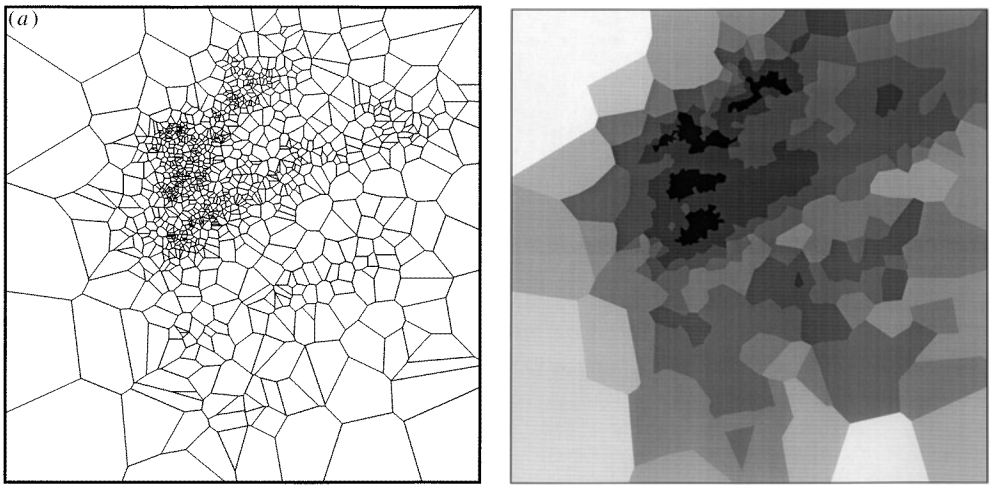

Figure 4. (*a*) Voronoi cells of the 1002 samples produced with an initial sample size of 2 (i.e. 500 iterations). The figure is similar to figure 2(*c*) although the edge cells are larger. (*b*) The approximate PPD produced by the cells.

The only lengthscales the algorithm 'sees' in the parameter space are given by the sizes of the Voronoi cells. Therefore it does not use truly local information in the way that a gradient optimization method does. Nevertheless its optimization performance in two dimensions is encouraging.

If this algorithm is to be useful for geophysical problems, then it must be practical in higher dimensions (between 5 and 100 dimensions would be useful). It turns out that there is a major problem with extending this particular algorithm to higher dimensions. Before expanding on this point it is worthwhile to first consider the cost of the nearest neighbour search in high dimensions.

Contrary to what one might expect, the search for nearest neighbours in high dimensions proves not to be the limiting factor. This is because we avoid the obvious brute force search, which would be expensive. A much better approach is to do the following. We first store the sum of the square distances of each node to the initial 1D axis. The distance of any node

to any point along the axis can then be found by simply summing two scalars, i.e. there is no loop over dimension. (Obviously the need to calculate square roots can be avoided by simply comparing the squares of distances to find the nearest neighbour.) The stored set of perpendicular squared distances can be updated for the next 1D axis using a single loop over dimension. In this way the 'nearest neighbour' search problem is drastically reduced in cost, and becomes practical even in high dimensions. This means that performing a random walk through a Voronoi cell in high dimensions may remain quite practical.

The main difficulty which affects the sampling algorithm is that the minimum sample size required for the algorithm to work is very sensitive to dimension. Even though the smallest sample size of 2 seemed to work well in the 2D example, this is not true in higher dimensions. Our experiments show that much larger sample sizes are required because the algorithm relies on a global sampling of the space to obtain a good enough approximation to the PPD that will drive further models. It turns out that in higher dimensions it is very difficult to obtain a good enough global sampling of the parameter space without drastically increasing the sample size.

To illustrate the point we perform an experiment. We generate uniform random points in a multidimensional space and count the average number of natural neighbours about each node. Two nodes are natural neighbours if their Voronoi cells share a common boundary (see figure 2). Figures 5(*a*) and (*b*) show the results of tests in up to 10 dimensions. As the number of nodes increases the space begins to saturate and the average number of natural neighbours tends to a constant. It is clear from figure 5(*b*) that the number of points required to reach saturation increases exponentially with dimension. For example, in five dimensions the space has filled up with only about 200 points, but for 10 dimensions the space is still far from saturation (hence undersampled) with 1000 points. For the algorithm to perform well, then (ideally) the sample size should be big enough to achieve saturation, and this becomes increasingly difficult as the dimension increases. In effect the density of the nodes becomes very small in high dimensions and the approximation of the PPD used by the algorithm becomes relatively featureless. Convergence is therefore extremely slow.

This particular neighbourhood algorithm is therefore unlikely to be useful in very high-dimensional parameter spaces. Lomax and Snieder (personal communication) have reported convergence of a very similar approach in nine dimensions, and figure 5(*b*) suggests that the sample size will need to grow rapidly in higher dimensions. Nevertheless, the neighbourhood approximation to the PPD contains some very appealing properties and provides a basis for future work in this area. Sambridge (1998) has extended the work presented here and developed alternative algorithms for the optimization and error analysis stage of the inversion based on these concepts, which appear to be suitable for much higher dimensional problems.

5. Summary

Our objective in this paper has been to present some of the relevant issues in nonlinear (Monte Carlo) inversion algorithms as known by the geophysical community. We have briefly discussed the popular Monte Carlo methods and suggested that they all have deficiencies in that they do not make use of all the information available for the assessment of trade-offs and resolution. In this paper we have provided no answers but some suggestions of where future progress might be made. The Voronoi cell concept is simple and powerful. We have shown how it might lead to a method for solving the commonly used multidimensional error analysis integrals, based on any sampling of the parameter space. We have also shown how it can be used as the basis of a simple global optimization method. This particular

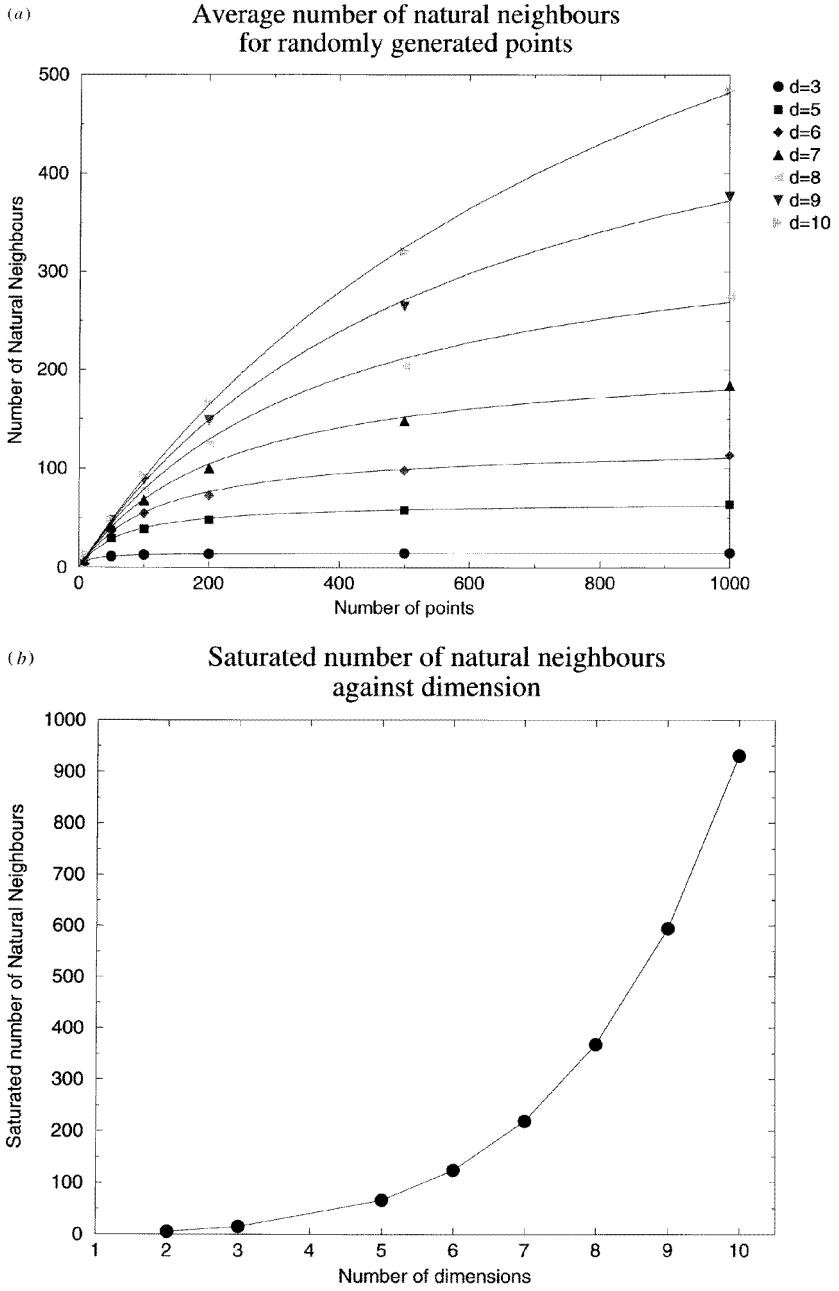

Figure 5. (*a*) A plot of the average number of natural neighbours against the number of randomly generated points in a space (see text). Calculations were repeated for seven different dimensions. The solid symbols are the calculated values and the curves are best fit curves of the form $\frac{ax}{(1+x)}$. Note how each curve tends to an asymptote when the space is saturated. (*b*) A plot of the estimated saturation level (a) of the previous curves as a function of dimension. The dependence of the saturation level on dimension is exponential, making it extremely difficult to achieve global sampling of high-dimensional spaces.

method suffers from a slowing down of the convergence rate as the dimension increases and is unlikely to be practical in dimensions above 10. However, we have suggested that the properties of Voronoi cells to adapt to any distribution of nodes in any number of dimensions makes them worthy of further study.

Acknowledgments

This work has benefited from numerous conversations with Jean Braun, Kerry Gallagher, Oli Gudmundsson, Anthony Lomax and Roel Snieder. Oli Gudmundsson and John Scales provided useful comments on the manuscript.

References

Aarts E and Korst J 1989 *Simulated Annealing and Boltzmann Machines* (Chichester: Wiley)

Andersen B, Hoffman K H, Mosegaard K, Nulton J, Pedersen J M and Salamon P 1988 *J. Phys.* **49** 1485

Anderssen R S 1970 The character of nonuniqueness in the conductivity modelling problem for the Earth *Pure Appl. Geophys.* **80** 238–59

Anderssen R S and Senata E 1971 A simple statistical estimation procedure for Monte Carlo inversion in geophysics *Pageoph* **91** 5–13

Basu A and Frazer L N 1990 Rapid determination of critical temperature in simulated annealing inversion *Science* **249** 1409–12

Braun J and Sambridge M 1995 Dynamic Lagrangian remeshing (DLR): a new algorithm for solving large strain deformation problems and its applications to fault-propagating folding *Earth Planet. Sci. Lett.* **124** 211–20

Braun J and Sambridge M 1997 Modelling landscape evolution on geological time scales: a new method based on irregular spatial discretization *Basin Res.* **9** 27–52

Brower R C Giles R, Moriarty K J M and Tamayo P 1989 *J. Comput. Phys.* **80** 472

Cary P W and Chapman C H 1988 Automatic 1D waveform inversion of marine seismic refraction data *Geophys. J.* **93** 527–46

Dosso S E and Oldenburg D W 1991 Magnetotelluric appraisal using simulated annealing *Geophys. J. Int.* **106** 370–85

Duijndam A J W 1988a Bayesian estimation in seismic inversion part I: principles *Geophys. Prospect.* **36** 878–98

——1988b Bayesian estimation in seismic inversion part II: uncertainty analysis *Geophys. Prospect.* **36** 899–918

Flournay N and Tsutakawa R K (ed) 1989 *Statistical Multiple Integration (Proc. AMS-IMS-SIAM Summer Research Conf. on Statistical Multiple Integration)* (Providence, RI: American Mathematical Society)

Fogel L J 1962 Autonomous automata *Ind. Res.* **4** 14–19

Fogel L J, Owens A J and Walsh M J 1966 *Artificial Intelligence through Simulated Evolution* (New York: Wiley)

Gallagher K, Sambridge M S and Drijkoningen G G 1991 Genetic algorithms: an evolution on Monte Carlo methods in strongly non-linear geophysical optimization problems *Geophys. Res. Lett.* **18** 2177–80

Gelfand A E and Smith A F M 1990 Sampling based approaches to calculating marginal densities *J. Am. Stat. Assoc.* **85** 398–409

Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images *IEEE Trans. Pattern. Anal. Mach. Intell.* **6** 721–41

Goldberg D E 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning* (Reading, MA: Addison-Wesley)

Holland J H 1975 *Adaptation in Natural and Artificial Systems* (Ann Arbor, MI: University of Michigan Press)

Ingber L 1989 Very fast simulated reannealing *Math. Comput. Modelling* **12** 967–93

Ingber L and Rosen B 1992 Genetic algorithms and simulated reannealing: a comparison *Math. Comput. Modelling* **16** 87–100

Iyer H M and K Hirahara (ed) 1993 *Seismic Tomography Theory and Practice* (London: Chapman and Hall)

Keilis-Borok V I and Yanovskaya T B 1967 Inverse problems of seismology *Geophys. J.* **13** 223–34

Kennett B L N 1978 Some aspects of nonlinearity in inversion *Geophys. J. R. Astron. Soc.* **55** 373–91

Kennett B L N and Nolet G 1978 Resolution analysis for discrete systems *Geophys. J. R. Astron. Soc.* **53** 413–25

Kirkpatrick S C, Gelatt D and Vecchi M P 1983 Optimization by simulated annealing *Science* **220** 671–80

Koren Z, Mosegaard K, Landa E, Thore P and Tarantola A 1991 Monte Carlo estimation and resolution analysis of seismic background velocities *J. Geophys. Res.* **96** 20 289–99

Lomax A and Snieder R 1995 Identifying sets of acceptable solutions to nonlinear geophysical inverse problems which have complicated misfit functions *Nonlinear Process. Geophys.* **2** 222–7

Minster J-B H, Williams N P, Masters T G, Gilbert J F and Haase J S 1995 Application of evolutionary programming to earthquake hypocenter determination *Proc. 4th Ann. Conf. Evol. Prog.* ed J R McDonnell *et al* (Cambridge, MA: MIT) pp 3–17

Menke W 1989 *Geophysical Data Analysis: Discrete Inverse Theory* (San Diego, CA: Academic)

Metropolis N, Rosenbluth M N, Rosenbluth A W, Teller A H and Teller E 1953 Equation of state calculations by fast computing machines *J. Chem. Phys.* **21** 1087–92

Mosegaard K 1998 Resolution analysis of general inverse problems through inverse Monte Carlo sampling *Inverse Problems* **14** 405–26

Mosegaard K and Tarantola A 1995 Monte Carlo sampling of solutions to inverse problems *J. Geophys. Res.* **100** 12 431–47

Nolet G 1987 Seismic wave propagation and seismic tomography *Seismic Tomography* ed G Nolet (Dordrecht: Reidel)

Nolte B and Fraser L N 1994 Vertical seismic profile inversion with genetic algorithms *Geophys. J. Int.* **117** 162–79

Nulton J D and Salamon P 1988 Statistical mechanics of combinatorial optimization *Phys. Rev.* A **37** 1351–6

Okabe A, Boots B and Sugihara K 1992 *Spatial Tessellations Concepts and Applications of Voronoi Diagrams* (New York: Wiley)

Parker R L 1977 Understanding inverse theory *Ann. Rev. Earth Planet. Sci.* **5** 35–64

Press F 1968 Earth models obtained by Monte Carlo inversion *J. Geophys. Res.* **73** 5223–34

Press W H, Flannery B P, Saul A T and Vetterling W T 1992 *Numerical Recipes* 2nd edn (Cambridge: Cambridge University Press)

Rothman D H 1985 Nonlinear inversion statistical mechanics, and residual statics corrections *Geophys.* **50** 2784–96

——1986 Automatic estimation of large residual statics corrections *Geophys.* **51** 332–46

Sambridge M S and Drijkoningen G G 1992 Genetic algorithms in seismic waveform inversion *Geophys. J. Int.* **109** 323–42

Sambridge M S 1988 unpublished manuscript

Sambridge M, Braun J and McQueen H 1995 Geophysical parametrization and interpolation of irregular data using natural neighbours *Geophys. J. Int.* **122** 837–57

Scales J A, Smith M L and Fischer T L 1992 Global optimization methods for multimodel inverse problems *J. Comput. Phys.* **103** 258–68

Sen M and Stoffa P L 1991 Nonlinear one-dimensional seismic waveform inversion using simulated annealing *Geophysics* **56** 1624–38

——1992 Rapid sampling of model space using genetic algorithms *Geophys. J. Int.* **108** 281–92

——1995 Global optimization methods in geophysical inversion *Advances in Exploration Geophysics* vol 4 (Amsterdam: Elsevier)

Shibutani T, Sambridge M and Kennett B 1996 Genetic algorithm inversion for receiver functions with application to crust and uppermost mantle structure beneath Eastern Australia *Geophys. Res. Lett.* **23** **14** 1829–32

Smith A F M 1991 Bayesian computational methods *Phil. Trans. R. Soc.* A **337** 369–86

Snieder R 1998 The role of nonlinearity in inverse problems *Inverse Problems* **14** 387–404

Stoffa P L and Sen M K 1991 Nonlinear multiparameter optimization using genetic algorithms: inversion of plane wave seismograms *Geophys.* **56** 1794–810

Tarantola A 1987 *Inverse Problem Theory* (Amsterdam: Elsevier)

Thurber C H and Aki K 1987 Three-dimensional seismic imaging *Ann. Rev. Earth Planet. Sci.* **15** 115–39

Trampert J 1998 Global seismic tomography: the inverse problem and beyond *Inverse Problems* **14** 371–85

Vasco D W Johnson L R and Majer E L 1993 Ensemble inference in geophysical inverse problems *Geophys. J. Int.* **117** 711–28

Voronoi M G 1908 Nouvelles applications des parametres continus a la theorie des formes quadratiques *J. Reine Angew. Math.* **134** 198–287

Watson D F 1992 *Contouring: A Guide to the Analysis and Display of Spatial Data* (Oxford: Pergamon)

Wiggins R A 1969. Monte Carlo inversion of body wave observations *J. Geophys. Res.* **74** 3171–81

Wilson W G and Vasudevan K 1991 Application of the Genetic Algorithm to Residual Statics Estimation *Geophys. Res. Lett.* **18** 2181–4