GEOPHYSICS, VOL. 66, NO. 2 (MARCH-APRIL 2001); P. 613–626, 10 FIGS.

A constrained global inversion method using an overparameterized scheme: Application to poststack seismic data

Xin-Quan Ma*

ABSTRACT

A global optimization algorithm using simulated annealing has advantages over local optimization approaches in that it can escape from being trapped in local minima and it does not require a good initial model and function derivatives to find a global minimum. It is therefore more attractive and suitable for seismic waveform inversion. I adopt an improved version of a simulated annealing algorithm to invert simultaneously for acoustic impedance and layer interfaces from poststack seismic data. The earth's subsurface is overparameterized by a series of microlayers with constant thickness in two-way traveltime. The algorithm is constrained using the low-frequency impedance trend and has been made computationally more efficient using this a priori information as an initial model. A search bound of each parameter, derived directly from the a priori information, reduces the nonuniqueness problem. Application of this technique to synthetic and field data examples helps one recover the true model parameters and reveals good continuity of estimated impedance across a seismic section. This approach has the capability of revealing the high-resolution detail needed for reservoir characterization when a reliable migrated image is available with good well ties.

INTRODUCTION

Seismic inversion is the calculation of the earth's structure and physical parameters from some sets of observed seismic data. The output of the seismic inversion can be *P*-wave and *S*-wave velocities, Poisson's ratio, or acoustic impedance volume. The fundamental interpretive benefit of any form of seismic inversion is that interface information is converted to interval information. Hence, the final presentation is representative of the geology, which is the study of rocks rather than rock boundaries. An impedance volume is more readily tied to well logs, for example, and more directly related to reservoir properties. It permits reservoir properties to be interpreted between wells on a layer-by-layer basis using calibrations at well locations. Acoustic impedance sections and volumes are a valuable asset for exploration, appraisal, and development since impedance is often a direct hydrocarbon indicator. The data are also useful for mapping fluids within the reservoir and improving volumetric estimation. Acoustic impedance modeling is also accepted as a valuable analytic tool for reservoir characterization and 4-D time-lapse seismic studies.

Historically, the most popular seismic inversion technique used to estimate acoustic impedance is recursive inversion. It is based on the well-known reflection coefficient formula in terms of the adjacent acoustic impedances. However, to make the algorithm work effectively, the seismic data must be free from the effects of the source wavelet, noise, multiples, spherical spreading, and transmission losses and only represent a band-limited, zero-phase sonic-log reflection coefficient series. Recursive inversion can only operate over the available seismic bandwidth. The upper frequency limit imposes a seismic resolution restriction on the estimated acoustic impedance. The lower limit indicates that inversion cannot generate absolute values of acoustic impedance but only relative ones. Absolute values of interval velocity are often determined from NMO and are combined with inverted velocity for a final output. However, NMO velocity is only used on frequencies of up to about 3 Hz.

Many inversion approaches are now based on forwarding modeling. Synthetic seismograms are generated from an initial subsurface model and compared to the real seismic data; the model is modified, and the synthetic data are updated and compared to the real data again. If after many iterations no further improvement is achieved, the updated model is the inversion result. A priori information is usually incorporated to improve the uniqueness of the output. Inversion is usually treated as a linear problem, that is, measurements are assumed to bear a linear relation to the parameters. However, many geophysical problems are nonlinear. Such problems are usually solved

Published on Geophysics Online August 28, 2000. Manuscript received by the Editor November 12, 1998; revised manuscript received December 13, 1999.
*Scott Pickford Group Ltd., Seismic Technology Division, 7th Floor, Leon House, 233 High Street, Croydon CR0 9XT, United Kingdom. E-mail: frankm@scopic.com.

by a sequence of linear approximations. A starting model is given, the misfit is used to perturb the model to make the misfit smaller, the new misfit is then used to further perturb the model, and so on. Iterations are stopped when the misfit is smaller than a prescribed convergence tolerance or when successive iterations fail to make any improvements. Typical examples of this inversion approach are the generalized linear inversion method (Cooke and Schneider, 1983) and the Marquardt-Levenberg method (Tarantola and Valette, 1982; Lines and Treitel, 1984). These conventional local optimization methods require the determination of function derivatives. They are prone to trapping in local minima, and their success depends heavily on the choice of starting model. This is because they rely on exploiting the limited information derived from a comparatively small number of models and avoid extensive exploration of the model space. Some local methods find the minimum by taking an increment along the steepest gradient to arrive at the next approximation, the step length often being proportional to the magnitude of the gradient. Convergence assumes that the objective function is continuous and that the initial estimate is already on the upper slopes of the correct global minimum. Local methods find the nearest minimum of the objective function, which may not be the global minimum or the true model. If the starting model has a misfit that is within the valley of the global minimum, then linearized inversion will descend the walls of that valley to that minimum by making small perturbations to the starting model. However, if you start in the wrong valley, the iterative search diverges from the global minimum. The linearized inversion normally ends up with a final model that differs only slightly from the starting model.

Because of the highly band-limited nature of seismic data, the objective function normally contains many valleys. We wish to find the lowest valley by searching widely through the model space. This problem belongs to the category of global optimization. Global methods, usually using a Monte Carlo random process, require no derivatives information and do not assume that the objective function has a particular shape. Simulated annealing is one of the most widely used global methods, operating analogously to thermodynamic systems. It is usually implemented by a drunkard's walk through the model space, where steps begin in random fashion but are progressively biased toward the global minimum (Corana et al., 1987; Sen and Stoffa, 1991; Goffe et al., 1994). Another class of global methods is called genetic algorithms. They try to evolve a trial population of models in a way similar to biological evolution (Goldberg, 1989; Stoffa and Sen, 1991). Global methods in general avoid the limitations of local methods and are particularly attractive in model-driven seismic waveform inversion.

Inversion of poststack seismic data using simulated annealing is described by Vestergaard and Mosegaard (1991). They use an improved version of the simulated annealing algorithm by Nulton and Salamon (1998) in which statistical information about the system to be optimized is used to improve the performance of the algorithm. They restrict the simulated annealing optimization to the two-way traveltime parameters to locate interfaces and perform a simple, linear optimization (least squares) to reflection coefficients. The objective function is a single-term L_2-norm misfit between the modelled and observed seismic traces. A weak reflection coefficient constraint is also applied to the algorithm using minimum and maximum reflection coefficients. The output from their optimization algorithm is reflection coefficients series; acoustic impedance traces must be obtained using a conventional recursive formula given a surface impedance value. While the use of a linear optimization to reflection coefficient parameters is computationally efficient, reflection coefficients, unlike acoustic impedance, are difficult to constrain because bounds for reflection coefficients as a function of two-way traveltimes are difficult to derive from a priori background information. Furthermore, the success of converting the finally optimized reflection coefficients into acoustic impedance traces critically depends on the starting impedance value.

In this paper I use an improved version of simulated annealing by Corana et al. (1987) and Goffe et al. (1994) to invert the acoustic impedance for 1-D earth models (applications to 2-D and 3-D models do not involve any fundamentally different theory). The global optimization applies not only to acoustic impedances but also to the two-way traveltimes (interfaces). The algorithm is constrained by including an additional term in the objective function prohibiting the final solution drifting away from the a priori background low-frequency trend. The computational efficiency is enhanced by starting the algorithm from an a priori impedance trend. Meanwhile, the nonuniqueness problem is reduced by restricting impedance perturbations within sensible bounds, which are derived directly from the background impedance trend. Since impedance values along with interfaces are solved simultaneously by the global optimization, the output are the absolute acoustic impedance values within the optimum interfaces; the surface impedance value, normally used by other algorithms for the recursive calculation of impedance from reflection coefficients, is not required.

To illustrate the robustness of the simulated annealing algorithm used in this paper, I examine the effects on the resolved impedance and interfaces of initial models and the selected seeds for the random number generator, and I show why the global optimization algorithm does not depend on the initial model and particular paths the random process has followed. I also show how to deduce a critical initial temperature for simulated annealing using statistical information during trial runs. The developed method is applied to a synthetic data example of ideally blocky layers and also to field data where well data are available for estimating the source wavelet and constructing initial models.

SIMULATED ANNEALING

Simulated annealing is a global optimization technique that mimics the physical process by which a crystal is grown by slow cooling of melt until the global minimum energy state is reached. It explores the objective function's surface and tries to optimize the function while moving both uphill and downhill. In standard annealing, a random point in the model space is selected and the energy f or misfit is calculated. The new model is accepted unconditionally if the energy associated with the new point f' is lower ($\Delta f = f' - f < 0$). If the new point has a higher misfit ($\Delta f > 0$), then it is accepted with the probability $p = \exp(-\Delta f/T)$, where T is a control parameter called acceptance temperature. The generation–acceptance process is repeated several times at a fixed temperature. Then the temperature is lowered following a cooling schedule, and the process

is repeated. The algorithm is stopped when the error does not change after a sufficient number of trials (see Appendix). This acceptance criterion is known as the Metropolis rule (Metropolis et al., 1953). Since the probability of accepting a step in an uphill direction is always greater than zero, the algorithm can climb out of a local minimum. This is in contrast to the local search methods in which a new model is accepted only if $\Delta f < 0$, i.e., it always searches in the downhill direction. Given a high starting temperature, simulated annealing first builds a rough view of the objective function surface. When the temperature is decreased, the algorithm progresses to reduce to zero the probability of accepting a bad step as the global minimum is reached (Figure 1).

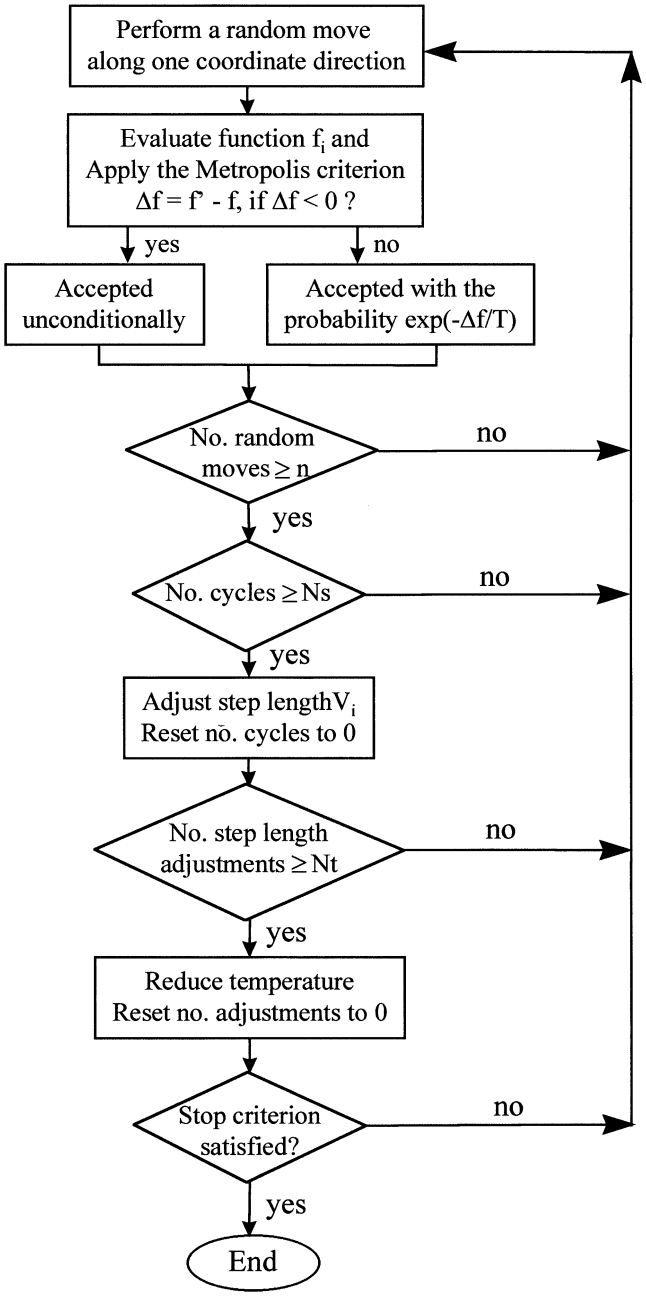

FIG. 1. A flow diagram of an improved simulated annealing optimization routine.

The modifications made by Corana et al. (1987) to the standard simulated annealing methods include a model perturbation controlled by a step length (refer to the Appendix). This step length is determined from the statistical information drawn during a number of iterations and is adjusted automatically to sample the objective function widely and to focus attention on the most promising area in the model space. The adjustment is made such that at least 50% of all moves in each coordinate direction are accepted. If a greater percentage of moves is accepted, the relevant element of a step-length vector is enlarged. This increases the number of rejections and decreases the percentage of acceptance. After many times through the above adjustment loops, T is reduced (Figure 1). A lower temperature makes a given uphill move less likely, so the number of acceptances decreases and the step lengths decline. The algorithm is stopped when the difference between several recorded optima is less than a given convergence tolerance (see the Appendix).

The step-length adjustment has many advantages over other versions of simulated annealing methods in that it provides valuable information about the function's flatness or roughness and increases the accuracy of parameter solutions. If an element of a step-length vector remains large, the function is flat in that parameter. Further extensions of the above algorithms are introduced by Goffe et al. (1994). One modification checks if the global optimum is indeed achieved by using different seeds for the random number generator. Another extension lets us determine a critical initial temperature for the algorithm by examining the step-length change during a trial run. This minimizes the execution time of the algorithm.

A proper selection of a starting temperature T_0 is crucial to implementing the simulated annealing method because the global method is computationally intensive. If the initial temperature is too low, the step length will be too small and the area containing the global optimum may be missed. If too high, then the step length will be too large and an excessively large area will be searched, wasting computing resources.

FORWARD MODELING

To apply optimization algorithms to the inversion problem, one must find a model to generate synthetic seismograms so that a misfit between the observed and synthetic can be determined. In laterally homogeneous acoustic media, seismic propagation can be approximately modelled by the convolution theory:

$$S(t) = \int_{-\infty}^{\infty} R(\tau) W(t - \tau)\, d\tau, \qquad (1)$$

where $S(t)$ is a seismic trace as a function of two-way traveltime, $W(t)$ is the seismic wavelet, and $R(t)$ is the reflection coefficients series. This mathematical model establishes a relationship between the poststack seismic data and the unknown model parameters such as velocity, density, acoustic impedance, and locations of interfaces. However, a proper analysis of the seismic data must include the effects of geometrical divergence, anelastic absorption, dispersion of wavelet, transmission losses across the boundaries of the layered media, and multiple reflections. For this paper in which the convolutional model is applied in a relatively small time window,

these complications are ignored and I assume the convolutional model is adequate.

The seismic source wavelet is usually unknown but can be estimated using various techniques—for example, the partial coherence method of White (1980). In this paper, I use a zero-phase Ricker wavelet of 35 Hz for testing the inversion of synthetic earth models but a wavelet estimated by the partial coherence method using both seismic and well data for testing the inversion of the real data. The synthetic seismic data are calculated in the frequency domain using forward and inverse fast Fourier transforms (FFTs).

OBJECTIVE FUNCTION

Optimization procedures require a misfit criterion or objective function to evaluate the quality of each trial model. The optimum model is determined when the objective function is in the global minimum or maximum, dependent on which type of objective function is used. The choice of the measure of misfit is crucial to the success of the process. Two types of objective functions are widely used in seismic waveform inversion. One is the correlation coefficient type, representing the resemblance of the observed and synthetic seismic data, which leads to a maximization problem (Sen and Stoffa, 1991). The other one is the least-squares type, representing the difference between the observed and synthetic data (the L_2-norm), which defines a minimization problem (Lines and Treitel, 1984). While the advantage of the correlation-based objective function is that the inversion results are less sensitive to noise, the least-squares type has proven to show much better results (Huang, 1996). An objective function may contain multiple terms, allowing different types of constraints to be built into the model. This is crucially important in improving the completeness and uniqueness of inverted results. In seismic waveform inversion, the low-frequency impedance and lateral continuity constraints are commonly used to reduce the nonuniqueness problem and improve the lateral coherence of the solution. In this paper, I use the L_1-norm error function, the least absolute deviation between the observed and modelled seismic trace. The L_1-norm error function can avoid overweighing large residuals as the L_2-norm does. I also build in an a priori information constraint as a second term in the objective function, which forces the solution to be close to the low-frequency impedance trend. The objective function is expressed as

$$\Delta f = W_1 \sum_{i=1}^{n} \left| S^i_{\text{obs}} - S^i_{\text{mod}} \right| + W_2 \sum_{i=1}^{m} \left| P^i_{\text{pri}} - P^i_{\text{mod}} \right|, \quad (2)$$

where S^i_{obs} is the observed seismic data, S^i_{mod} is the synthetic seismic data, P^i_{pri} is an a priori low-frequency impedance trend, P^i_{mod} is the modelled impedance, n is the number of samples in the seismic trace, m is the number of microlayers in the initial model, and W_1 and W_2 are weights applied to the two terms, respectively.

LAYER PARAMETERIZATION

The earth's impedance is a continuous function in depth. It is often advantageous to make a discrete approximation to this continuous function. If lithology boundaries are known, an obvious discrete approximation is to sample the earth medium as separate layers; the layer thickness is assumed to be known, and we solve for the impedance in each layer (we often refer to this impedance representation as exact parameterization). When lithology boundaries are unknown, the earth model can be sampled as a number of microlayers with constant thickness, and both layer interfaces and impedance need to be solved. If the total number of microlayers is larger than that of the true earth medium, this representation is defined as overparameterization. Another category is the full-scale parameterization by which the earth model is sampled at the same interval of equal two-way traveltime as in seismic data. This is the extreme case of overparameterization, requiring a very large number of impedance unknowns—but no layer boundaries—to solve.

While the exact parameterized scheme appears to be superior in terms of its simplicity, in practice it is difficult to implement because the layer interfaces are not generally known exactly; we are often required to invert uninterpreted seismic data and to extract more detailed information, including the layer boundaries. Even if the data have been interpreted, some human error is certain to be introduced because of the bandwidth limitation of seismic data and structures complicated by thinning beds, faulting, and intrusive features. Since the vertical resolution of the inverted impedance profile is expected to be greater than the seismic data, the use of exact parameterization based on preinterpreted horizons may limit the vertical resolution of the inverted earth impedance profile. The full-scale parameterization also appears to be appealing because one needs to concentrate on impedance only. However, the main drawback of this scheme is its high computational cost from the increasing number parameters. Furthermore, this scheme introduces instability into final solutions resulting from the nonuniqueness problem (Sen and Stoffa, 1991).

I concentrate on the overparameterized scheme, which is more practical for real applications. The earth model is described by a series of interconnected microlayers; each microlayer is described by its thickness or two-way traveltime and impedance. In this scheme, both impedance and interfaces are considered as unknown parameters and are solved simultaneously by the global optimization procedure.

SYNTHETIC DATA EXAMPLE

An earth model is assumed to consist of ten layers overlying a half-space. Each layer is assigned a constant impedance value and layer thickness. The impedance ranges from 2 to 8 ($g/cm^3 \times km/s$), while the thickness in two-way traveltime ranges from 20 to 60 ms (Figure 2a). The observed seismic data in this experiment are calculated using the reflection coefficients of the true earth model convolved with a zero-phase Ricker wavelet of 35 Hz (Figure 2b). We have also obtained the low-frequency impedance curve by band-pass filtering (0–5 Hz) the true impedance data. The objective function consists of the least absolute deviation between the observed and synthetic seismic data and the least absolute deviation between the low-frequency and modelled impedance. The former is the misfit function; the latter is an a priori impedance constraint. Layer parameterization is performed assuming the earth is made up of 21 horizontal microlayers including the underlying half-space. Since the number of microlayers is greater than that of the model, this is an overparameterized scheme. The unknown parameters to be optimized consist of 21 impedances

and 20 interfaces. This is to say that we have an objective function with 41 variables (parameters) and need to resolve them, given the observed seismic data and a wavelet, so that the resultant 41 parameters produce a global minimum of the objective function.

Global optimization is achieved using the modified simulated annealing routine described above. To apply the algorithm, one needs to supply the objective function, lower and upper parameter boundaries, initial model parameters (or a starting model), initial annealing temperature, and a convergence tolerance. The initial model parameters consist of 20 interfaces at microlayer boundaries and 21 impedance values taken from the low-frequency impedance trend. The simulated annealing algorithm begins by calculating the objective function using the initial model parameters and then accepting or rejecting this move using the Metropolis criteria. Within a given temperature many iterations are allowed to be performed, and the step length is adjusted according to the statistics drawn from those iterations. The temperature is lowered until the convergence tolerance is satisfied. The outputs are the optimized impedance values and locations of layer interfaces.

While the global method does not depend heavily on the starting model, it does requires physically sensible bounds to restrict the parameter searching space as a result of the multimodal nature of the objective function. A good selection of parameter boundaries reduces the nonuniqueness problem and minimizes the computing time. In this experiment, we allow the impedance parameters to lie within ± 2.5 (g/cm^3 $\times$ km/s) around the low-frequency trend and the two-way traveltimes within ± 12 ms around microlayer boundaries. The convergence tolerance ε should be chosen with reference to the objective function values, for example, 0.1% of the objective function value at the first iteration in simulated annealing.

The inverted impedance and interfaces are shown in Figure 3a. The inversion results match well with the true earth model, with relative errors less than 3%. The synthetic seismic trace calculated using the inverted impedance profile has also been plotted along with the observed and error traces (Figure 3b). They indicate that nearly all energy in the observed trace has been inverted, resulting in an almost zero-error trace. To illustrate the evolution of the simulated annealing procedure, I show four intermediate results from annealing. The initial annealing temperature is taken as 0.1 and is subsequently reduced according to the rule described in the Appendix. The inverted impedance profiles are plotted against the true impedance depth function after 1000, 5000, 10 000, and 15 000 iterations (Figure 4a). Although the match at the 1000th iteration is very poor, the impedance function follows the general trend of the low-frequency impedance profile because of the initial model and parameter search bounds imposed. After 5000 iterations, the impedance values within layers 2, 3, and 9 get close to the true impedance. After 10 000 iterations, not only impedance but also interfaces of all 11 layers are near to the optimum solution. The last 5000 iterations tune the results further, and by the end of 15 000 iterations an optimum solution is found. To reveal how the misfit is reduced through the evolution of annealing, I plot the synthetic seismic trace against the observed seismic data after 1000, 5000, 10 000, and 15 000 iterations (Figure 4b). The quality of match increases as the number of iterations increases. By the end of 15 000 iterations, almost all energy in the observed seismic trace has been recovered, resulting in an optimum impedance solution.

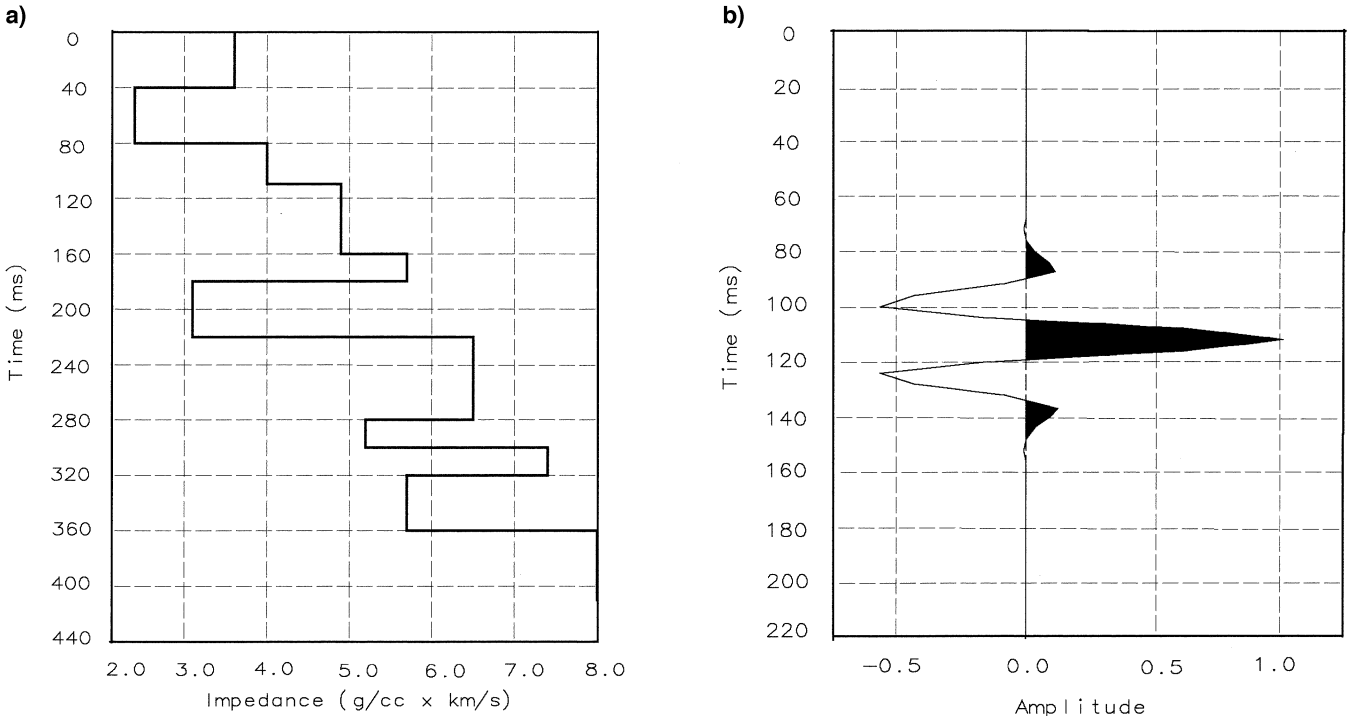

FIG. 2. (a) A 1-D earth model consisting of 10 layers overlying a half-space. Each layer is described by constant acoustic impedance and thickness in two-way traveltime. (b) A Ricker wavelet with a principal frequency of 35 Hz.

Many factors determine the total number of iterations required for the algorithm to reach the final equilibrium. Among them are the initial temperature T_0 and convergence tolerance ε. There are other parameters: N_s, N_t, N_ε, and r_T, critically control the efficiency of the algorithm. The meanings of these parameters are fully explained in the Appendix. While Corana et al. (1987) use $N_s = 20$, $N_t = \max(100, 5n)$, $N_\varepsilon = 4$, and $r_T = 0.85$ in their function tests, I found that in this seismic inversion exercise the parameters can be smaller. My tests show that I can choose $N_s = 10$, $N_t = 3$, $N_\varepsilon = 3$, and $r_T = 0.5$ and get equally satisfactory results. This modification substantially reduces the total number of iterations without sacrificing accuracy. This means that to invert the impedance profile in Figure 3a using the newly modified parameters requires 18 700 iterations, in contrast to 32 800 iterations if Corana's parameter values are used. I also found that setting the initial step lengths equal to 25% of the initial model parameters will speed up the convergence.

Simulated annealing, unlike local optimization methods such as gradient descent and downhill simplex, does not rely on the initial parameters. As a matter of fact, most conventional simulated annealing codes use randomly distributed parameters to start with, and the optimization continues until the convergence tolerance is satisfied. To illustrate the dependency of solutions upon the initial model, I rerun the optimization routine for the earth model shown in Figure 2a but use different initial parameters. The results using an initial model that deviates from the general trend of true impedance is shown in Figure 5a. Although the initial model is close to the lower parameter boundary, the impedance and interface are still resolved with high accuracy. If the initial model is set to be close to the upper parameter boundary, the results are hardly affected (Figure 5b). Therefore, it is advantageous to use the global optimization method for seismic waveform inversion in cases where well control is sparse or where the correlation between seismic events and nearby wells is made difficult by fault zones, thinning beds, or the presence of strong noise. However, the independence of initial models is achieved partially by using a relatively high initial temperature, which allows the algorithm to search a wide area (more uphill moves accepted) but as a consequence increases the computational burden. In practice, if near-optimal solutions or a priori information are available, it is always beneficial to use this information in the starting models. This is because starting from a near-optimum model does not require a high initial temperature and saves a lot of computing time.

As mentioned earlier, in running the simulated annealing program, one needs an initial temperature T_0, a key parameter controlling the efficiency of the simulated annealing operator and the resolution of modelled parameters. While a large T_0 permits a wide space to be searched, it produces more acceptances and fewer rejections. This may impose a large computational burden and make the method less efficient. On the other hand, a small T_0 may make it difficult for the current solution to escape from the local minimum, which risks that the solution you have found is not in the global optimum. While trial and error is always suggested to obtain the optimal T_0, the improved simulated annealing routine presented in this paper allows us to

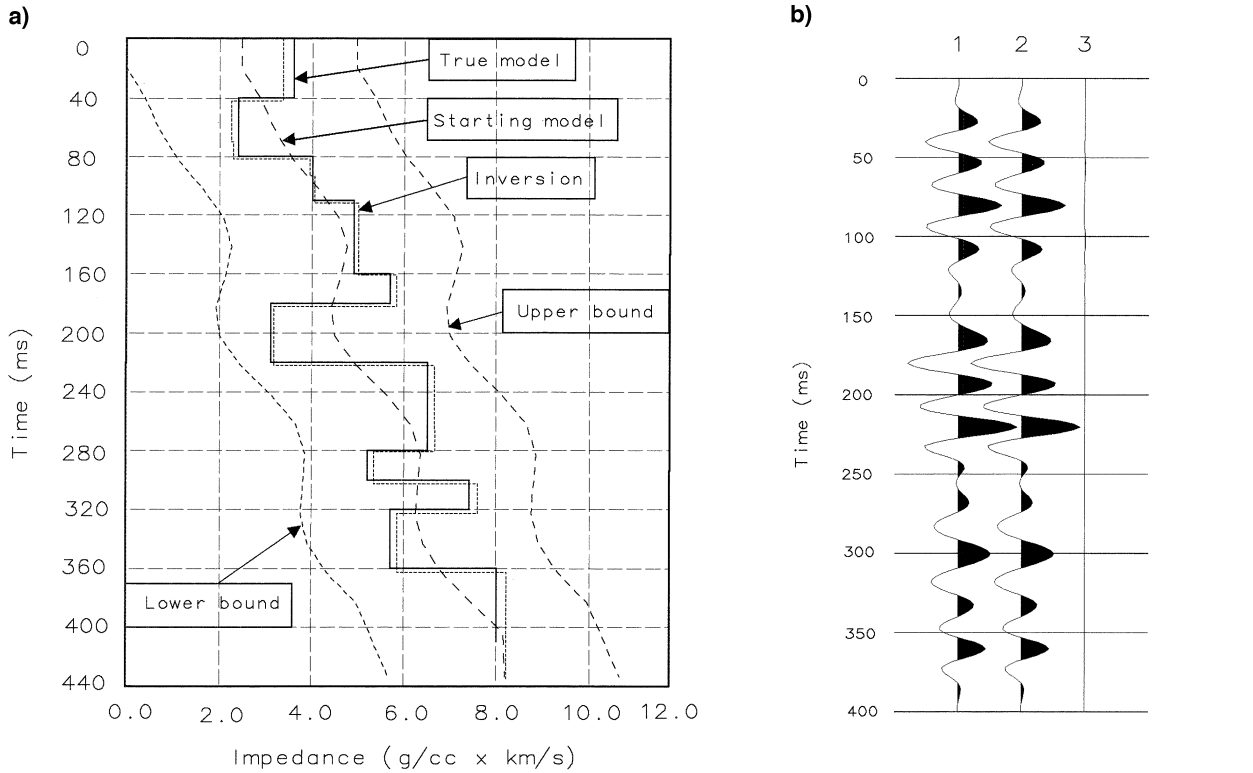

FIG. 3. (a) Comparison between the true earth impedance profile (solid line) and that inverted by an overparameterized scheme (dotted line). An initial model contains 20 microlayers with impedance values defined by the low-frequency trend. Initial temperature for simulated annealing is $T_0 = 0.1$. (b) The original, inverted, and error seismic traces.

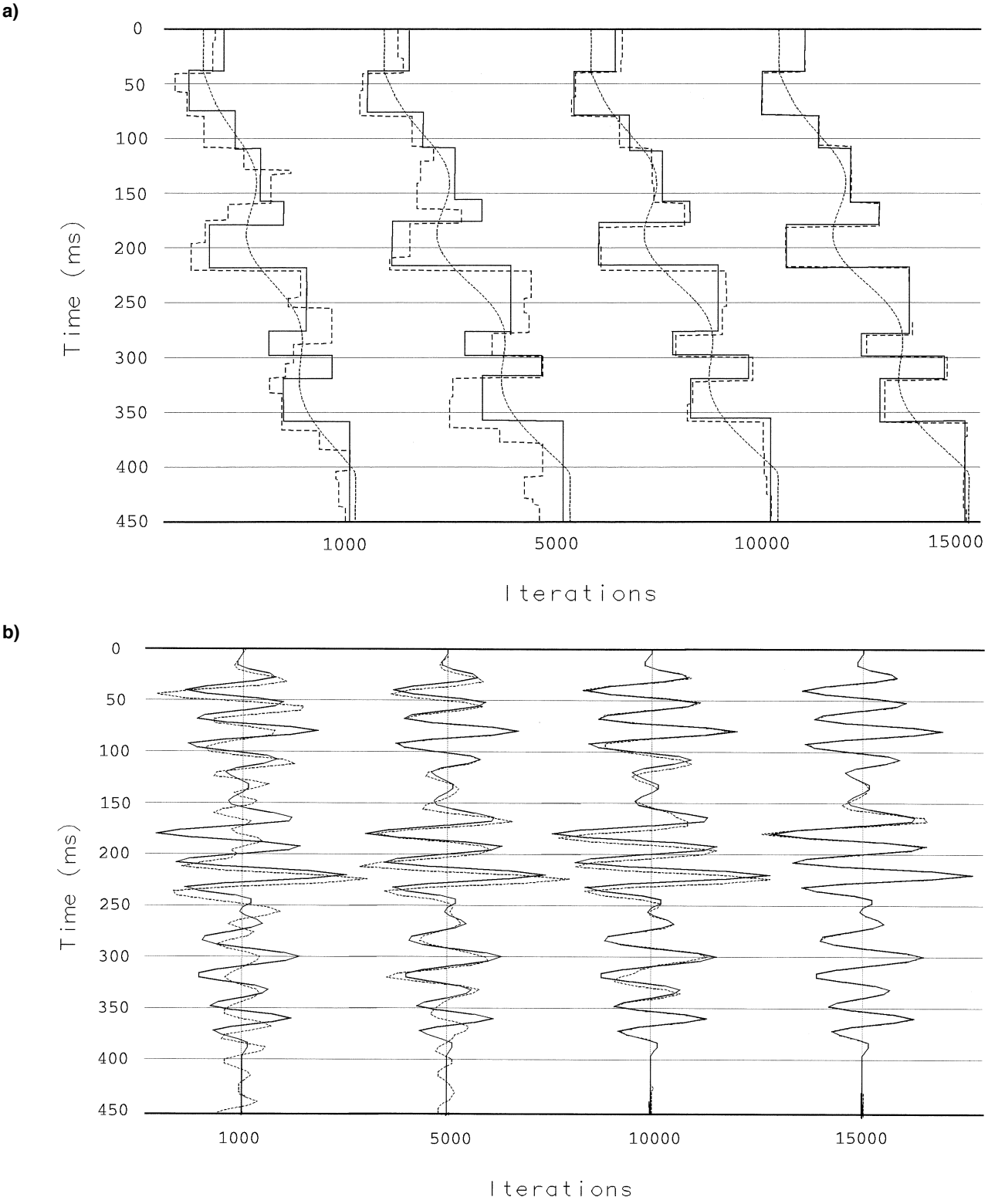

FIG. 4. (a) Four intermediate stages of simulated annealing. Solid lines are the true impedance profiles, while dashed lines are the inverted impedance profiles. Iterations 1000, 5000, 10 000, and 15 000 correspond to the annealing temperatures at 1.00×10^{-1}, 1.25×10^{-2}, 1.56×10^{-3}, and 9.76×10^{-5}, respectively. (b) The match between the synthetic seismic trace and the observed seismic trace at the four stages.

examine the step length of each parameter and hence guides us to choose an appropriate starting temperature through a few trial runs. I have rerun the above model using a very high starting temperature, $T_0 = 1000$, and monitored the step-length variation for parameters 2, 10, and 20 with falling temperature. The step length here corresponds to the impedance parameter for a specific layer. The relationship between step length and temperature is shown in Figure 6a. One can see that the step length changes very little at the earlier stage of annealing ($T < 1.0$) and starts to fall only when the temperature gets

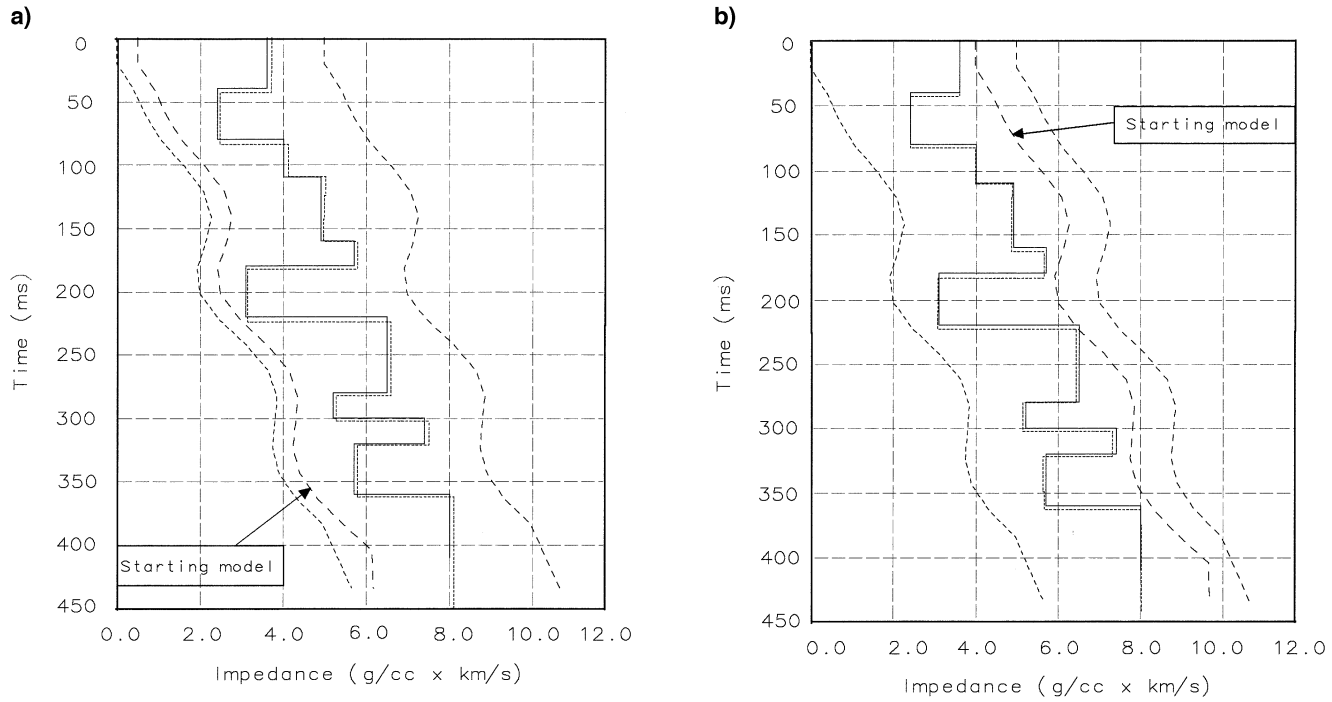

FIG. 5. (a) Same as Figure 3a except that the starting impedance parameters are chosen to be close to the lower parameter bound. (b) Starting parameters are close to the upper parameter bound.

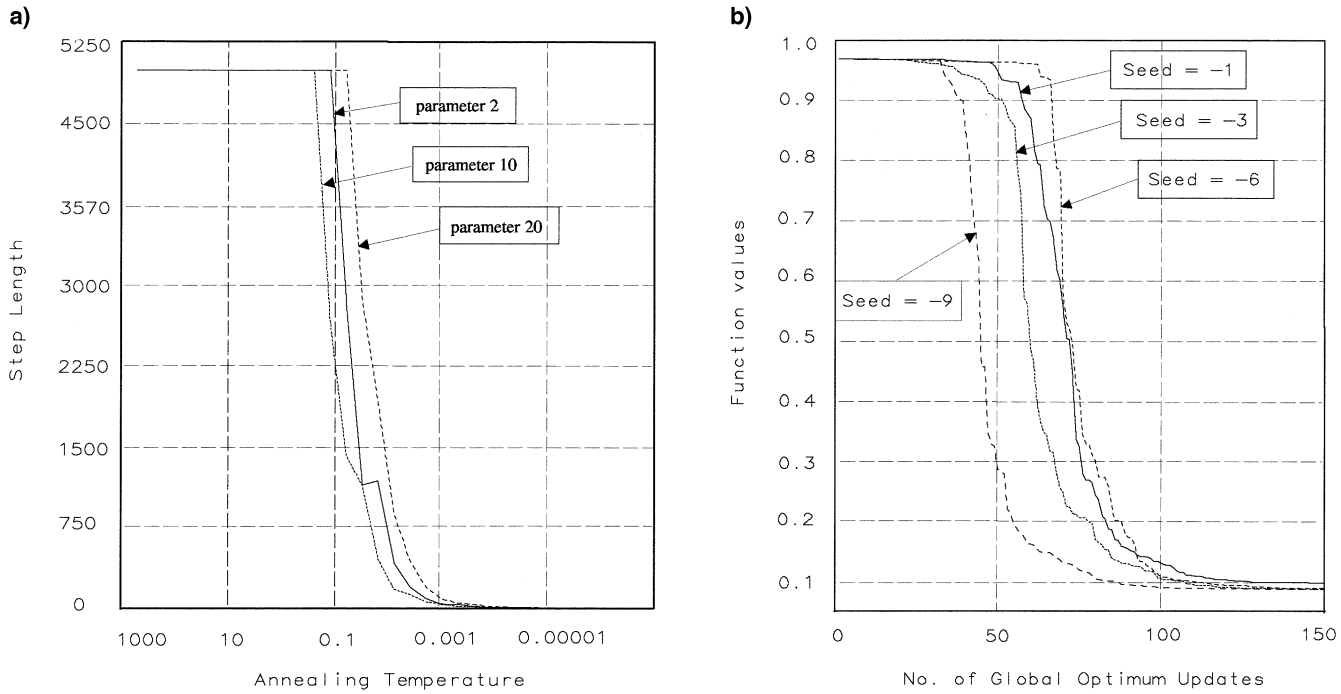

FIG. 6. (a) Relationship between step length and annealing temperature for three parameters. This is used to determine the critical temperature for the simulated annealing procedure. (b) Relationship between the objective function and number of global minimum updates as a function of seeds for the random number generator. This shows the convergence of the solution through different paths.

around 0.1. This temperature ($T = 0.1$ in this case) is often termed as a critical temperature from which more effective searches really begin. Prior to the critical temperature, many function evaluations are carried out and too many acceptances are made, resulting in a considerable waste of computing resources. In this example, an appropriate starting temperature should be chosen as $T_0 = 0.1$. However, the critical temperature determined may also be dependent on which parameter is being monitored. One should always choose the maximum critical temperature among all the parameters monitored as a starting temperature for the annealing procedure.

Apart from the influences of an initial temperature T_0 and starting models, the simulated annealing procedure is also affected by the use of different seeds for the random number generator. This is because a different seed enables the algorithm to follow a completely different path. In theory, no matter what seed is used, the convergence to the same global minimum should take place. In practice a slight difference in solutions should be expected because of the different paths the algorithm has followed. To illustrate this point, I have run the optimization routine four times using four different seeds (Figure 3a). At each run, the number of updates to the global minimum and their associated objective function values are plotted in Figure 6b. In the earlier updates, the objective function values are almost the same; later, they converge to the global minimum through different paths. This is because in the earlier stage, for a fixed temperature, the algorithm searches parameters near the initial model, which gives similar function values. As the temperature decreases, the objective function is reduced toward the global minimum. The exact path the algorithm follows depends on the nature of the objective function and the type of the random generator. However, one can rerun the algorithm with the same initial temperature and different seeds. If the same optimum is found within an accepted tolerance, there is a high degree of confidence in the global optimum found.

FIELD DATA EXAMPLE

The seismic data were from a North Sea 3-D survey and had been processed using conventional data processing methods including spherical divergence correction, deconvolution, multiple suppression and prestack migration. Four wells had P-wave velocity logs available. Following are inversion results for a single seismic trace near the well and for an in-line seismic cross-section.

A wavelet estimate is required before optimization procedures can be implemented. We assume the seismic trace is equivalent to the earth's reflectivity convolved with a stationary wavelet plus an amount of noise. The well data give us the earth's reflectivity at the well locations. Comparing the seismic data at these locations, to a synthetic trace will yield an estimate of the wavelet. The wavelet is the matching filter between the log and seismic segments and is computed using the auto-, cross-, and power spectra of both segments (White, 1980). This method also quantifies the quality of match between the synthetic seismogram and the data and hence indicates the validity of the convolutional model. The wavelet estimated at one of the wells is shown in Figure 7.

As described in the synthetic data example, we also need the low-frequency impedance trend to constrain our inversion. This low-frequency trend was generated using all the well data and the seismic horizon data. A velocity function of the form $V = V_0 + kZ$ was determined for each of the interpreted units between the horizons, where V_0 is the velocity at the horizontal surface datum, V is the velocity at a depth Z below the datum plane, and k is a constant whose value is generally between 0.3/s and 1.3/s . This function was then interpolated for all CDP bins; for each of these, a velocity trace in time was generated. The velocity volume was then converted to impedance using Gardner's equation.

The results from the inversion of a single seismic trace near a well are shown in Figure 8. Only a 400-ms data window is used. The low-frequency impedance is used as a constraint in the objective function and to determine the lower and upper bounds of impedance within which the impedance parameters are allowed to perturb. Using an overparameterized scheme as described in the synthetic data example, the earth is assumed to consist of 23 microlayers including the underlying half-space, which yields 22 interfaces. These 45 parameters are optimized using the simulated annealing procedure with an initial temperature $T_0 = 0.1$; results are shown in Figure 8a. The major horizons at 120, 225, 250, and 285 ms have been well predicted. The impedance match between the inverted seismic section and log is reasonably good except at 360–390 ms, where larger impedance is predicted. The mismatch may be attributable to either inaccuracy in log data or an unremoved multiple event in seismic data. The original, synthetic, and error seismic traces are shown in Figure 8b. Some residuals still exist within the error trace.

I also applied this single trace inversion technique to some of the in-line sections in this field—for example, section 595 (Figure 9a). Optimization parameters such as the initial annealing temperature, interface perturbation window, and

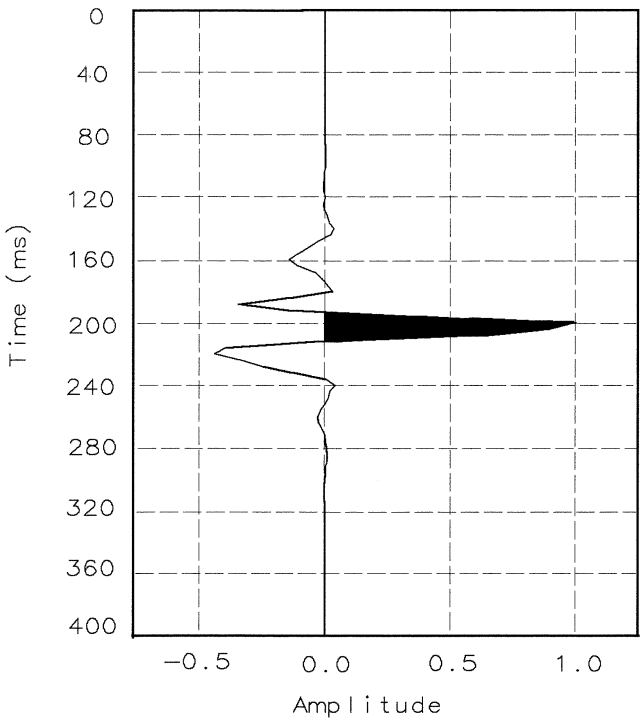

FIG. 7. A source wavelet estimated using the partial coherence matching of synthetic seismograms with real seismic data.

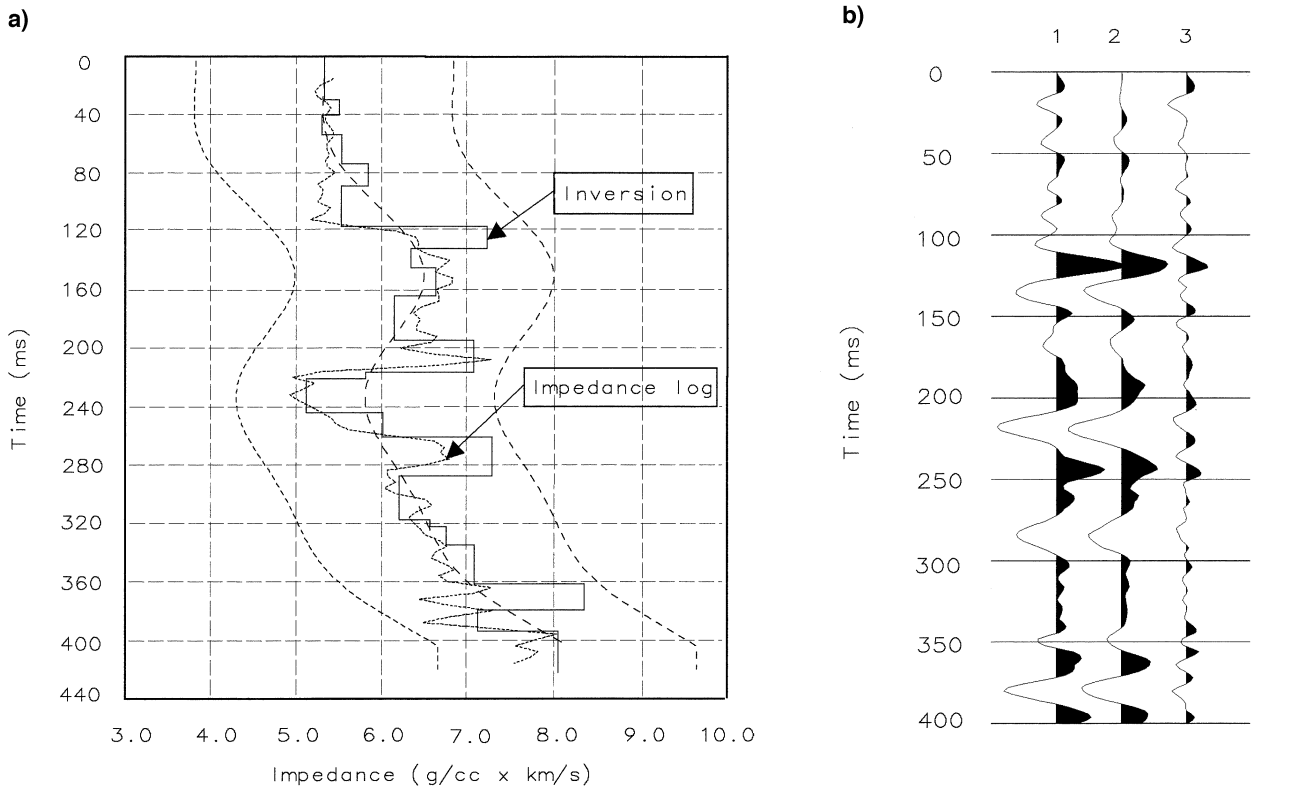

FIG. 8. (a) Comparison between impedance log (dotted line) and inversion by an overparameterized scheme (solid line). Impedance is in g/cm^3 × km/s. An initial model contains 22 microlayers with impedance values defined by the low-frequency trend. Initial temperature for simulated annealing is 0.1, and the total number of iterations is around 19 000. (b) The original, synthetic, and error seismic traces. The starting time of the data window is from 0, which corresponds to 2000 ms in the real data set.

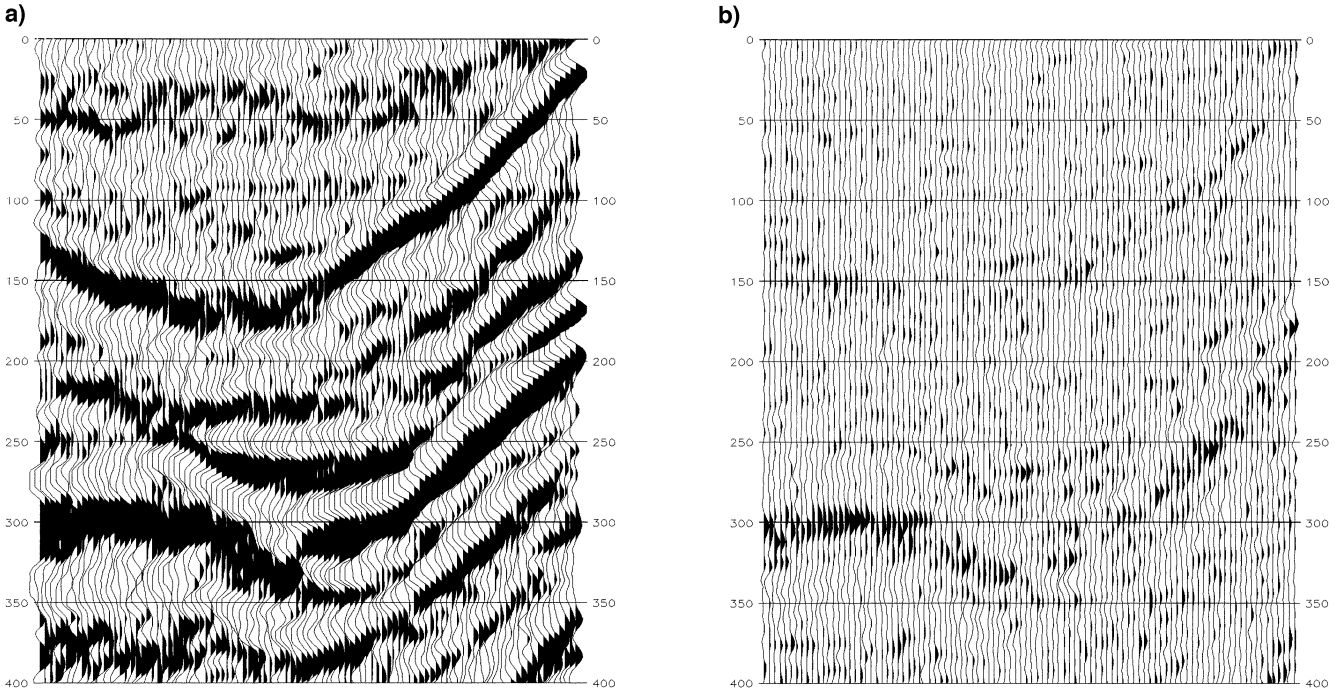

FIG. 9. (a) An in-line seismic section from a North Sea 3-D survey. (b) Error seismic traces generated by subtracting the synthetic seismograms from the original seismic data shown in (a). The starting time of the data window is from 0, which corresponds to 2000 ms in the real data set.

convergence tolerance are determined from the above exercise for the single trace near the well location (Figure 8) in which the true impedance log is available to control the quality of inversion. The low-frequency impedance trend (Figure 10a) is used as the initial model and defines constraints for the objective function. The absolute impedance inverted and error seismic traces are shown in Figures 10b and 9b, respectively. The inverted impedance corresponds well to seismic events, and lateral variations of seismic impedance values within layers are evident. Trace-to-trace continuity of inverted impedance is very high, demonstrating the stability of the method. Consistent layer boundaries have also been derived.

In the above real-data example, I chose 23 microlayers over a 400-ms data window as an initial model. The result shows that the match between the inverted and log impedance is reasonably good. But is this overparameterized? How do we choose the number of microlayers if there is little direct knowledge about this parameter from real data?

The first criterion for choosing the total number of microlayers is that the layers should be thin enough for the synthetic data to mimic the real data. An underparameterized scheme, in which the microlayers modelled are thicker than the layering of the real medium, is more likely to produce a large misfit between the synthetic and observed seismic data (Sen and Stoffa, 1991). On the other hand, while very thin microlayers are more likely to achieve a low misfit or a better match, the impedance solution may not necessarily be unique (the nonuniqueness problem is discussed in detail at the end of this paper). Furthermore, a consequence of using very thin microlayers is that a large number of parameters are generated, which makes the algorithm less efficient. This leads to the second criterion, i.e., the microlayers should not be so thin that the computational time becomes impractical and the solutions unstable. The instability problem caused by using very thin layers or too many layers within a short time distance is discussed in detail by Sen and Stoffa (1991). In their test examples, 21 microlayers represent a four-layer model. Their results show that although the correlation values obtained are very high (low misfit), the constructed models differ considerably from the true model. The inversion algorithm picked many contrasting velocity and density values in the homogeneous part of the model and generated a poor match of impedances at the real layer boundaries. They conclude that when the model is overparameterized, many different models explain the observed seismic data.

To determine the exact number of microlayers as an initial model satisfying the above two criteria is difficult in practice. However, a guideline can be adopted from analyzing the frequency spectrum of seismic data. The higher frequency of component the seismic data carries, the higher resolution and hence more details about the earth structure it reveals. Therefore, we should use thinner layers for high-frequency data than for low-frequency data. When impedance logs are available, seismic traces near well locations should be inverted first using different microlayer thicknesses. This helps us determine an optimum microlayer thickness through the quality control of an inverted impedance trace against an impedance log.

NONUNIQUENESS

Seismic inversion is nonunique, that is, a number of different models can lead to the same set of observations. This is true partly because measurements are incomplete and also because they involve uncertainties. In model-driven seismic inversion, the quality of a solution is determined by comparing the observed seismic trace to a synthetic trace generated from the solution. If these two are the same, then the solution is exact but not necessarily unique. For the overparameterized scheme presented in this paper, the earth is parameterized by a series of impedances and interfaces. Perturbation of either parameters will change the waveform of synthetic data. Without any constraints, this could lead to the situation that there are many combinations of impedance and interfaces, all satisfying the perfect match between the observed and synthetic data. The acoustic impedance solution from seismic inversion is extremely nonunique in the frequency ranges outside the bandwidth of the source wavelet. This phenomenon can be viewed from the low and high ends of frequency bandwidths. Consider that the main frequencies recorded range from approximately 10 to 60 Hz. Very low-frequency information about the acoustic impedance, say, <8 Hz, is not directly derivable from band-limited seismic data. Since reflection coefficients are negligible derived from a smooth, low-frequency impedance trace, its contribution to the synthetic seismic amplitudes after the convolution with a band-limited wavelet is negligible, too. In a model-driven inversion scheme without constraints, many solutions with different low-frequency trends all equally satisfy the observation. The inversion process also rejects any frequencies higher than the wavelet bandwidth, say, >60 Hz. Consider a sufficiently thin bed embedded in a homogeneous half-space. Such an impedance structure produces two reflection coefficients of opposite polarity and equal amplitude. The convolution of these reflection coefficients with a band-limited wavelet contributes little to the synthetic seismic amplitudes. Any number of thin beds can be added to the acoustic impedance profile without significantly affecting the fit to the data. The very high-frequency content is therefore impossible to recover through inversion.

Despite the nonuniqueness problem encountered in seismic inversion, constraints usually limit the physical properties such that the set of possible solutions exists only within narrow bounds. For my inversion scheme, these constraints are achieved by using a priori impedance information, that defines the parameter search boundaries and also guides the solution, moving toward the physically meaningful trend. Interfaces are also constrained by allowing them to perturb within a narrow time window. The low-frequency impedance data can be obtained using both well logs and seismic processing velocity data. From the optimization perspective, while global methods such as simulated annealing and genetic algorithms are less dependent on the initial model, they do require sensible parameter bounds. Incorporating a priori information into the inversion scheme can therefore speed up the convergence and stabilize the solution considerably.

CONCLUSIONS

Global optimization methods, based on a stochastic search mechanism, do not require derivatives information and a good initial model to find the global optimum. They can escape from being trapped in local minima by moving both downhill and uphill. This is in contrast to conventional local methods, which use the gradient information and need a good initial guess to

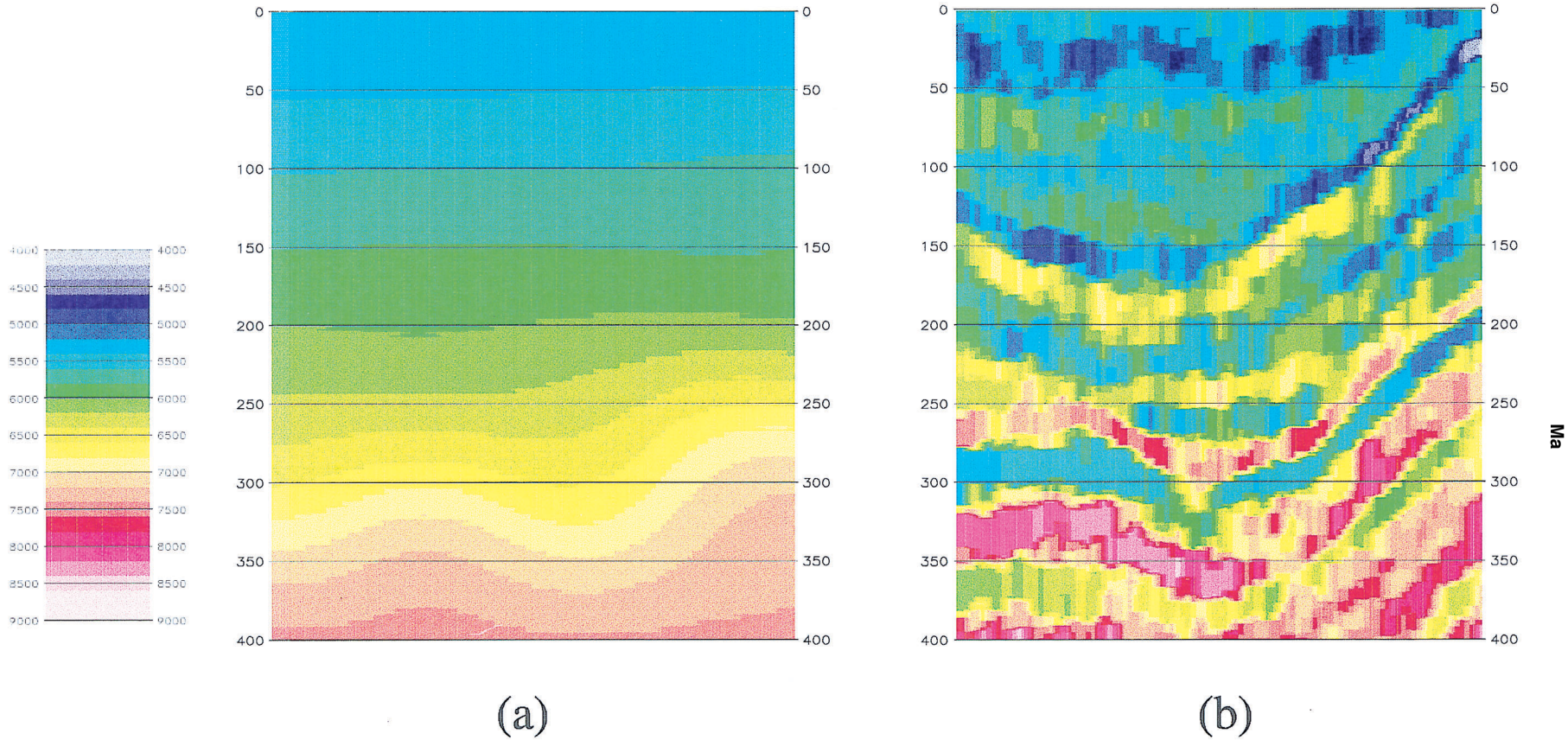

FIG. 10. (a) Low-frequency impedance trend obtained through interpolation of four wells in the field. This is used to define impedance constraints and to determine parameter search bounds for the inversion process. (b) Inverted acoustic impedance section for the seismic section shown in Figure 9a. Note that impedances in (a) and (b) are in $g/cm^3 \times m/s$.

guide their search. Global methods such as simulated annealing and genetic algorithms are therefore more attractive and suitable for model-driven seismic waveform inversion.

The improved simulated annealing algorithm described in this paper has many advantages over the standard simulated annealing algorithm in terms of its robustness and computational efficiency. The step-length adjustment adopted during the cooling schedule allows us to monitor and study the function behavior and helps us find the critical initial temperature. This algorithm can find optimum solutions with high precision and is particularly robust in solving seismic inversion problems. The algorithm has been made computationally efficient by tuning annealing parameters to fit the seismic inversion scheme.

I have applied the simulated annealing algorithm to the inversion of a 1-D earth model using an overparameterized scheme in which the earth's subsurface is parameterized by a series of microlayers with constant impedance within each layer. The algorithm has been accelerated by using the low-frequency impedance trend as the starting point, and the nonuniqueness problem has been reduced by assigning a boundary to each parameter. Since the impedance bounds are derived directly from the low-frequency trend, the output is the absolute impedance values within optimized microlayer boundaries. Application of this scheme to synthetic and field data examples reveal a good match between inverted and true impedance for a single trace and impedance profiles with good continuity and stability. This technique has the capability of extracting maximum resolution of the acoustic impedance over typical reservoir intervals, making it a suitable tool for reservoir characterization and 4-D time-lapse seismic studies.

The inversion approach described in this paper can be used whenever forward modeling can be implemented, for example, prestack seismic inversion via amplitude variation with offset (AVO).

ACKNOWLEDGMENTS

The author thanks Paul Haskey and Adrian Pelham of Scott Pickford Group Ltd. for stimulating discussions and reading the manuscript.

REFERENCES

Cooke, D. A., and Schneider, W. A., 1983, Generalized linear inversion of reflection seismic data: Geophysics, **48**, 665–676.

Corana, A., Marchesi, M., Martini, C., and Ridella, S., 1987, Minimizing multimodal functions of continuous variables with the simulated annealing algorithm: ACM Trans. of Math. Soft., **13**, 262–280.

Goffe, W. L., Ferrier, G. D., and Rogers, J., 1994, Global optimisation of statistical functions with simulated annealing: J. Economet., **60**, 65–100.

Goldberg, D. E., 1989, Genetic algorithms in search, optimisation and machine learning: Addison-Wesley Publ. Co.

Huang, X., 1996, Seismic inversion using heuristic combinatorial algorithm: A hybrid scheme: 66th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1963–1966.

Lines, L. R., and Treitel, S., 1984, A review of least-squares inversion and its application to geophysical problems: Geophys. Prosp., **32**, 159–186.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., 1953, Equation of state calculations by fast computing machines: J. Chem. Phys., **21**, 1087–1092.

Nulton, J. D., and Salamon, P., 1988, Statistical mechanics of combinatorial optimisation: Physics Rev., **A37**, 1351–1356.

Sen, M. K., and Stoffa, P. L., 1991, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: Geophysics, **56**, 1624–1638.

Stoffa, P. L., and Sen, M. K., 1991, Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane-wave seismograms: Geophysics, **56**, 1974–1810.

Tarantola, A., and Valette, B., 1982, Generalised nonlinear inverse problems solved using the least squares criterion: Rev. Geophys. Space Phys., **20**, 219–232.

Vestergaard, P. D., and Mosegaard, K., 1991, Inversion of post-stack seismic data using simulated annealing: Geophys. Prosp., **39**, 613–624.

White, R. E., 1980, Partial coherence matching of synthetic seismograms with seismic traces: Geophys. Prosp., **28**, 333–358.

APPENDIX

AN IMPROVED SIMULATED ANNEALING ALGORITHM

Let $\mathbf{X}$ be a vector and $(x_1, x_2, \ldots, x_n)$ its components. Let $f(x)$ be the objective function to minimize, and let $a_1 < x_1 < b_1, \ldots, a_n < x_n < b_n$ be its n variables, each ranging in a finite, continuous interval. Simulated annealing proceeds interactively. Starting from a given point $\mathbf{X}_0$, it generates a succession of points: $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_i, \ldots$ tending to the global minimum of the objective function. Let us assume that the currently accepted solution is $\mathbf{X}$ and the function value is f. New candidate points are generated around the current point x_i, in turn applying random moves along each coordinate direction (Corana et al., 1987):

$$x' = x_i + r\,v_i,$$

where r is a uniformly distributed random number from $[-1, 1]$ and v_i is element i of a step-length vector $\mathbf{V} = (v_1, v_2, \ldots, v_n)$. If the point falls outside the definition boundary, a new point is randomly generated until a point belonging to the definition domain is found:

$$x' = a_i + r(b_i - a_i).$$

The function value f' is then computed. If f' is less than f, $\mathbf{X}'$ is accepted unconditionally and the algorithm moves downhill. If this is the smallest function value, $\mathbf{X}'$ is recorded as the best current value of the optimum. If f' is greater or equal to f, the Metropolis criterion decides on acceptance with the probability

$$p = e^{-(f'-f)/T}$$

where T is temperature and p is compared to p', a uniformly distributed random number from $[0, 1]$. If $p < p'$, the new point $(\mathbf{X}')$ is accepted and the algorithm moves uphill; otherwise, $\mathbf{X}'$ is rejected. Both a low temperature and a large difference in the function values decrease the probability of an uphill move.

After N_s steps through all elements of $\mathbf{X}$, the step-length vector $\mathbf{V}$ is adjusted to better follow the function behavior. The aim of these variations in step length is to maintain the average percentage of accepted moves at about one-half of the total number of moves (Corana et al., 1987). For each direction i the new step-vector component v_i' is

$$v_i' = v_i\left(1 + c_i\frac{n_i/N_s - 0.6}{0.4}\right) \quad \text{if } n_i > 0.6\,N_s,$$

$$v_i' = \frac{v_i}{1 + c_i\dfrac{0.4 - n_i/N_s}{0.4}} \quad \text{if } n_i < 0.4\,N_s,$$

$$v_i' = v_i \quad \text{otherwise.}$$

The value v_i is the current step length, c_i is a parameter controlling the step variation along each ith direction, and n_i is the number of accepted moves along the ith direction as a result of total N_s trials. The ratio n_i/N_s is therefore restricted to the interval [0, 1].

If a greater percentage of moves is accepted for x_i, then the relevant element of $\mathbf{V}$ is enlarged. For a given temperature this increases the number of rejections and decreases the percentage of acceptances. From an optimization view, a high number of accepted moves with respect to rejected ones means the function is explored with too small steps. On the other hand, a high number of rejected moves means that new trial points are generated too far from the current point. A 1 : 1 ratio between accepted and rejected moves ensures that the algorithm is following the function behavior well.

After $N_t \cdot N_s$ cycles of moves along every direction with N_t step-length adjustments (refer to Figure 1), T is reduced. The new temperature is given by

$$T' = r_T \cdot T,$$

where r_T is the temperature reduction factor set between 0 and 1. A lower temperature makes a given uphill move less likely, so the number of rejections increases and the step lengths decline. In addition, the first point tried at the new temperature is the current optimum. The smaller step length and starting at the current optimum focuses attention on the most promising area.

The process ends by comparing the last N_ε values of the smallest function values from the end of each temperature reduction with the most recent one and the current optimum function value. If all these differences are less than a prescribed tolerance, the process terminates. If these conditions are not satisfied, the algorithm starts again through all steps described above.

Apart from the initial temperature and convergence tolerance, other factors such as N_s, N_t, c_i, N_ε, and r_T also control the efficiency of the simulated annealing. Some test optimizations of simple functions made by Corana et al. (1987) suggest the following values should be used:

$$\begin{aligned} N_s &= 20, \\ N_t &= \max(100, 5n), \\ c_i &= 2, \quad i = 1, 2, \ldots, n, \\ N_\varepsilon &= 4, \\ r_T &= 0.85. \end{aligned}$$

However, much smaller values for r_T and N_t were chosen by Goffe et al. (1994) in the optimization tests, and their results were equally satisfactory.