# Inversion of acoustic data using a combination of genetic algorithms and the Gauss-Newton approach

Peter Gerstoft<sup>a)</sup>

SACLANT Undersea Research Centre, 19138 La Spezia, Italy

(Received 18 July 1994; accepted for publication 13 December 1994)

An inversion procedure for obtaining speeds, attenuation, densities, and thicknesses for a layered medium is described. The inversion is carried out using the least-squares technique and the forward modeling is based on SAFARI. The optimization is a hybrid method combining the global genetic algorithms and the local Gauss–Newton method. This is done by taking several gradient steps between each update of the object function for each "individual" in the population. The gradients for the Gauss–Newton method are computed analytically; this makes the computation faster and more stable than computing the gradients by numerical differentiation. The combination of a global and a local method makes the hybrid method faster and it gets closer to the global minimum than a pure global method. Examples based on both real and synthetic data in wave-number-frequency and range-frequency domains show that the method works well.

PACS numbers: 43.30.Ma, 43.30.Pc, 43.60.Pt, 43.40.Ph

### INTRODUCTION

Recently there has been a substantial increase in the use of global optimization techniques such as simulated annealing  $(SA)^{1-3}$  and genetic algorithms  $(GA)^{4,5}$  applied to inversion of underwater acoustic signals when the inversion is stated as an optimization problem. The reason for this is clear, as they are formulated independently of the forward model and thus very easy to apply. However, few successful inversions of global optimization have been reported for inversions with more than 30 parameters increases.

Local methods have also been used for seismoacoustic problems<sup>6,7</sup> and for seismic exploration.<sup>8-12</sup> These are based on the gradients of the object function and require computation of these quantities; provided that it is possible to compute the gradients, a local method will usually be able to descend quite efficiently toward a local minimum. However, local methods have the disadvantage of getting trapped in suboptimal minima, and the method can become unstable when computing the next step in the iteration.

The idea presented in this paper is to combine the two search methods, so that (1) the local method does not get trapped in local minima, (2) the computer time used for a global method is significantly reduced, and (3) larger problems can be handled than if only a global method was used.

Few examples of combinations of local and global methods exist, e.g., Refs. 13 and 14. In Ref. 14 it was suggested to use the covariance matrix of the gradient of the object function to reparametrize the parameter space. After this reparametrization the search proceeds using SA. This approach is efficient if some parameters in the parameter space are characterized by a few local minima with prominent features in one direction. In Ref. 13 a Monte Carlo method was used in combination with a gradient descent method in order to fit seismic waveforms of marine data. A Monte Carlo method was used first to explore the large parameter space using a coarse grid. When this search was estimated to be in the region of the global minimum, the search was sped up using a gradient descent method on the most plausible model. Given the good performance of guided Monte Carlo methods such as SA and GA in comparison with a simple Monte Carlo method, it seems natural to combine a guided Monte Carlo with a gradient search.

The goal of the inversion procedure is to find the model vector **m** that optimizes a cost function or object function  $\phi$ . Its formulation depends both on the problem at hand and on the measured data available. Here we choose a quadratic variation:

$$\boldsymbol{\phi}(\mathbf{m}) = [\mathbf{W}\mathbf{e}]^T \mathbf{W}\mathbf{e}, \quad \mathbf{e} = \mathbf{d}_{obs} - \mathbf{d}_{cal}(\mathbf{m}), \quad (1)$$

where  $\mathbf{d}_{obs}$  and  $\mathbf{d}_{cal}$  are normalized unit vectors containing the  $N_{obs}$  observed and calculated amplitude of the pressure field,  $\mathbf{m}$  is the model vector containing the physical parameters, and  $\mathbf{W}$  is a diagonal matrix containing the weighting for each observation point. Scaling of the data vectors allows us to work only with the shape of the pressure field.

SAFARI<sup>15,16</sup> is used as the forward model. Thus the environment is horizontally stratified and only the acoustic parameters are considered. The model vector  $\mathbf{m}$  with M elements is given by

$$\mathbf{m} = [\mathbf{c}_{p}^{T}, \boldsymbol{\gamma}_{p}^{T}, \boldsymbol{\rho}^{T}, \mathbf{z}^{T}]^{T},$$
(2)

with the four subvectors being the speeds  $c_p$ , attenuations  $\gamma_p$ , densities  $\rho$ , and thicknesses z for the layers. Naturally, not all the environmental parameters need to be unknown. Analytic computation of the gradients from a wave-theoretic model has been done by several authors in the seismic community, for example for acoustic<sup>11</sup> and for elastic media.<sup>12</sup>

This paper is organized as follows. In Sec. I the hybrid method is introduced, followed by an overview of both genetic algorithms and the Gauss–Newton method. It is important to regularize the solution, here shape functions are used,

<sup>&</sup>lt;sup>a)</sup>E-mail:gerstoft@saclantc.nato.int

and this is also described in Sec. I. In Sec. II the approach is applied to the inversion of geoacoustic parameters for both real and synthetic data.

## I. THE HYBRID APPROACH

A crucial point in deciding if a global or local method is most efficient for a given object function is the number of local minima in the search space. By a global method we mean a method, such as SA or GA, which does not use gradients. If there are only a few local minima a local method with a few random starting points is more efficient than a global method, because a global method is very inefficient in descending, as the gradients are not used. For noisier data the object function will show several suboptimal minima. For this scenario the global method will become more efficient as it will not get stuck in each of the minima. Finally, if the data becomes too noisy then only an exhaustive search will find the solution.

For a large class of inverse problems it is believed that the number of local minima in the search space is small. Here an efficient approach would be to find the local minima by a gradient method and then use a global method to choose between these local minima. This is what the hybrid method should do.

The hybrid approach states quite simply that for each new child generated by the GA, in addition to the crossover and mutation operator,  $N_g$  Gauss-Newton steps are applied to the child in order to increase the fitness. The choice of  $N_g$ is problem dependent. For a local problem it is optimal to use a large  $N_g$  so that the minimum is reached quickly. However, if the minimum is not reached during the first set of iterations, it can be reached in the next with little extra computational cost. For very noisy data it is optimal not to perform any Gauss-Newton steps. Experimentally, we found the  $N_g=5$  seems to be a good choice.

## A. Genetic algorithms

Genetic algorithms (GA) are based on an analogy with biological evolution. While these have already provided promising results in the seismic community,<sup>17-19</sup> an application to ocean acoustic problems has only recently been published.<sup>4</sup> The basic principle of GA is simple: From all possible model vectors, an initial population size of q member is selected. The fitness of each member is computed based on the fit between the observed and computed data. Then through a set of evolutionary steps the initial population evolves in order to become fitter. An evolutionary step consists of selecting a parental distribution from the population based on the individual's fitness. The parents are then combined in pairs and operators are applied to them to form a set of children. Traditionally the crossover and mutation operators have been used, but for the hybrid method the "Gauss-Newton" operator is also applied. Finally, the children replace part of the population to get a fitter population.

The environment is discretized into M parameters in a model vector **m**. Each parameter  $m_j$ , j = 1, ..., M, has a binary parameter string of length  $n_j$  and can take  $2^{n_j}$  discrete values according to an *a priori* probability distribution (Gaussian, rectangular or based on *a priori* information).

۱ <sub>j</sub> =	α <sub>nj-1</sub>	αnj-2	·	•	•	α <sub>3</sub>	α2	α <sub>1</sub>	
0 =	0	0			•	0	0	0	m <sup>min</sup> j
1 =	0	0		•		0	0	1	$m_j^{min} + 1 \Delta m$
2 =	0	0	•	•	•	0	1	0	m <sup>imin</sup> +2∆m
3 =	0	0	·	•	•	0	1	1	$m_{j}^{min}$ + 3 $\Delta m$
2 <sup>n</sup> j <sup>-1</sup> =	1	1				1	1	1	m <sup>max</sup>

FIG. 1. Binary coding of model parameters.  $n_i$  is the number of bits.

Here a rectangular distribution between a lower and upper bound  $[m_i^{\min}, m_i^{\max}]$  is used. We have (see Fig. 1)

$$m_j^i = m_j^{\min} + i_j \Delta m_j, \quad i_j = 0, 1, 2, \dots, 2^{n_j} - 1,$$
 (3)

where

$$\Delta m_{j} = \frac{m_{j}^{\max} - m_{j}^{\min}}{2^{n_{j}} - 1}.$$
(4)

A major difference between SA and GA is that GA uses q model vectors at the same time, where q is the population size, while SA uses only one. GA consists essentially of three operators: Selection, crossover, and mutation.

Selection: In order to establish the next generation, a subset of the current population must be selected as parents. Selection is based on the fitness of the individual models. The probability that the *k*th member is selected as parent is made according to a normalized Boltzmann distribution:

$$p_k = \frac{\exp[-\phi(\mathbf{m}^k)/T]}{\sum_{l=1}^{q} \exp[-\phi(\mathbf{m}^l)/T]}, \quad k = 1, \dots, q.$$
(5)

The introduction of temperature T, as in SA, gives us the opportunity to stretch the probability and to improve the performance of the algorithm.<sup>17</sup> Indeed, at the first stage of the procedure, by stretching the fitness we avoid choosing as parents only those members with the better fit, which would otherwise tend to dominate the population; later in the optimization this stretching leads to better discrimination between models with close fitness. As with SA, the choice of the temperature T is difficult. It must neither be too high nor too low. A good compromise is a temperature of the same magnitude as the object function, here  $T = \min[\phi(\mathbf{m}^k)]$ . During the optimization, the fitness increases and the temperature decreases.

Crossover: For each set of parents, each consisting of a model vector, two children are constructed. For each parameter in the model vector each child may either be a direct copy of one parent, with probability  $1-p_x$ , or it can be a bit crossover of the two parents with crossover probability  $p_x$ , see Fig. 2. The crossover point is chosen randomly in the interval [1,N-1], where N is the number of bits used in the coding. Different techniques are available to perform this crossover where the entire chromosome is used once, and multiple-point crossover (as used here) where the chromosome is divided into genes related to each parameter on which the crossover is applied.



FIG. 2. Crossover is a binary exchange of l bits between the binary codes for two model parameters. l is chosen randomly.

*Mutation*: This is a random change of one bit in the model vector, performed with probability  $p_m$  in order to better explore the search space, see Fig. 3.

It is possible that a run of a GA will approach a local minimum. In order to increase the probability of finding the global minimum, several independent populations are started. This is also advantageous for collecting statistical information to estimate the *a posteriori* probabilities. For advice on how to set the optimization parameters, see Ref. 4.

### B. Gauss-Newton

A minimum of Eq. (1) is obtained by linearizing the function  $\mathbf{d}_{cal}(\mathbf{m})$  in the neighborhood of some working point, and a solution to the linearized problem is then found. This procedure is iterated, using for the new working point the solution found at the previous step, until a stable solution is determined. The Gauss-Newton approach described in this section is standard; for further detail see the published literature.<sup>6-9,11,20</sup>

For the linearization, we need to compute the derivative of the calculated pressures  $\mathbf{J} = \nabla \mathbf{d}_{cal}(\mathbf{m})$ , where the *ij*th element of the  $N_{obs} \times M$  Jacobian matrix is given by

$$J_{ij} = \frac{\partial d_{\text{cal},i}(\mathbf{m})}{\partial m_{i}},\tag{6}$$

see the Appendix. For a model perturbation  $\partial \mathbf{m}$  from the current model  $\mathbf{m}$ , a Taylor expansion with terms up to the second order of the object function, Eq. (1), is performed

$$\phi(\mathbf{m} + \delta \mathbf{m}) = \phi(\mathbf{m}) + \mathbf{g}^T \delta \mathbf{m} + \frac{1}{2} \delta \mathbf{m}^T \mathbf{H} \delta \mathbf{m}, \tag{7}$$

where the vector of first derivatives, the gradient  $\mathbf{g}$ , and the matrix of second derivatives, the Hessian  $\mathbf{H}$ , of the object function are given by

$$\mathbf{g} = 2[\mathbf{W}\mathbf{J}]^T \mathbf{W}\mathbf{e},\tag{8}$$

$$\mathbf{H} = 2[\mathbf{W}\mathbf{J}]^T \mathbf{W}\mathbf{J} + 2[\mathbf{W}(\nabla \mathbf{J})]^T \mathbf{W}\mathbf{e}, \tag{9}$$

$$\approx 2[\mathbf{W}\mathbf{J}]^T \mathbf{W}\mathbf{J}.$$
 (10)



FIG. 3. Mutation is a random change of one bit.

Neglecting the second term in Eq. (9) is the Gauss– Newton approximation and corresponds to assuming that the residuals **e** are locally linear with respect to the parameter changes. We minimize Eq. (7) subject to a regularization term, and obtain the Gauss–Newton regularized iteration:

$$\delta \mathbf{m} = ([\mathbf{W}\mathbf{J}]^T \mathbf{W}\mathbf{J} + \lambda \mathbf{B}^T \mathbf{B})^{-1} [\mathbf{W}\mathbf{J}]^T \mathbf{W}\mathbf{e}.$$
(11)

The matrix **B** is the regularization matrix, and  $\lambda$  is a Lagrange multiplier. In general, **B** will impose some smoothness constraint on the solution. In the present study **B** is the identity matrix, thus the iteration in Eq. (11) corresponds to the Levenberg-Marquardt method. An error analysis is independent of the particular iteration method chosen, but does depend on the regularization operator.

The best tool for solving and analyzing the solution of a local method is the singular-value decomposition (SVD) of WJ.<sup>7-9</sup> Given orthogonal matrices U, V, and a diagonal matrix  $\Sigma$  such that  $\mathbf{J}=\mathbf{U}\Sigma\mathbf{V}^{T}$ , the solution of Eq. (11) for  $\lambda=0$  belongs to the space spanned by the columns v of V (*parameter eigenvectors*); the corresponding diagonal terms  $\sigma_i$  in  $\Sigma$  (*singular values*) are an indication of the relative importance of each eigenvector for the given problem. Using the singular-value decomposition, Eq. (11) is written

$$\delta \mathbf{m} = \sum_{i=1}^{M} \frac{\sigma_i}{\sigma_i^2 + \lambda} \mathbf{v}_i(\mathbf{u}_i^T \mathbf{e}) = \sum_{i=1}^{M} \frac{\sigma_i}{\sigma_i^2 + \lambda} \mathbf{v}_i(\alpha_i).$$
(12)

From Eq. (12) it is seen that small singular values will indicate low sensitivity of the field to the corresponding parameter eigenvector, and moreover may cause numerical instability in the computation. The effect of the regularization term in Eq. (11) is to damp the smaller singular values by shifting all the singular values away from the origin, so that the computation becomes stable. However the information in the parameter subspace spanned by the corresponding eigenvectors cannot be retrieved accurately.

The use of SVD is straightforward when the parameter vector **m** has elements with the same physical meaning, e.g., sound speed as a function of depth within the seafloor. If **m** is a combination of parameters with different physical dimensions, a coordinate transform would be introduced so that the SVD is independent from the physical units of the different parameters, and thus is invariant to scaling. But even if this is done correctly, several problems will arise when using a local method in several parameters having different physical meaning.

In this application a simple scaling has been used: A diagonal matrix containing the current values of the model vector  $\mathbf{D} = \text{diag}[m_1^{-1}, m_2^{-1}, \dots, m_M^{-1}]$  has been used to scale the model vector. The Jacobian matrix **J** should then be replaced by **JD** in Eq. (11). The resulting solution  $\mathbf{m}^D$  must then be rescaled to obtain the model vector  $\mathbf{m} = \mathbf{D}^{-1}\mathbf{m}^D$ . Other methods for determining the scaling matrix exist **D**, see Ref. 8.

We first solve Eq. (12) with an insignificant value of  $\lambda(\lambda=10^{-5}\sigma_1^2)$ ; if this new model vector does not give an improved value of the object function we multiply  $\lambda$  by a factor 10 and resolve Eq. (12). This continues until an improved value of the object function is found or the stop cri-

terion is satisfied. The stop criterion is that the maximum number of iterations is reached or that for all parameters i=1,...,M:

$$\delta m_i < \Delta m_i \epsilon, \tag{13}$$

where  $\epsilon = 0.2$  and  $\Delta m_i$  is the discretization step size given by Eq. (4).

An important step in obtaining a solution by a local method is the determination of the derivatives of the pressure field with respect to the model parameters. These can be computed in three ways:<sup>12</sup> finite difference, adjoint state technique, and analytical differentiation.

The adjoint state is based directly on the differential equation, see Ref. 20.

For the finite difference method, the main problem is to determine the step size. This should be determined by some adaptive method in order to get a stable result. It is also computationally demanding; for a problem of M parameters a finite difference solution requires at least M+1 forward models. This is in contrast to both adjoint and analytical differentiation where the computation of derivatives does not increase the computer time significantly. Thus finite difference should only be used for small problems. In addition, finite difference will have problems in obtaining stable derivatives; it cannot easily be automated.

Analytic differentiation has been used in seismic reflectivity modeling in the frequency-wave-number domain by Refs. 11 and 12 for seismic exploration problems, but for each formulation of the forward model the analytic differentiation is different.

In the present application we derive analytical expressions for the gradient using the direct global matrix method (DGM) as implemented in SAFARI. These are essential in getting a Gauss-Newton method to work and are developed in the Appendix.

#### C. Regularization

Experience with synthetic and real data<sup>5,21</sup> had shown that it is advantageous to regularize the solution in order to have *a priori* well-behaved solutions. Regularization is introduced via shape functions:

$$\mathbf{m} = \sum_{i}^{M_s} \mu_i \mathbf{h}_i, \qquad (14)$$

where  $\mathbf{h}_i$  is the *i*the shape function or basis function,  $\mu_i$  is the coefficient associated with the shape function, and  $M_s$  is the number of shape functions used. The advantages of shape functions are (1) To constrain the solution to belong to a certain class of expected profiles. (2) To describe the variation of the parameters with fewer coefficients and in so doing to reduce the number of unknowns. This can also constrain the solution to be more physically correct. By linking the velocities in the sediment together we can describe them by a smoother function and, further, the number of unknowns is reduced. (3) To link correlated parameters. For example the source of receiver depth is often specified in depth from the surface. Alternatively, using shape functions they can be specified from the bottom.

TABLE I. Comparison of inversion results for the three examples. CPU time is for a DEC-3000/800 workstation.

Case	Obj fn	CPU (min)	Model calls (×10 <sup>3</sup> )	New individuals (×10 <sup>3</sup> )
A1-GA	0.003	14	50	50
A1-hy	0.0001	0.6	1	0.2
A1-GN	0.006	0.03	0.022	
A2-GA	0.0003	14	50	50
A2-hy	0.000007	9	10	1
B-GA	0.56	40	10	10
B-hy	0.62	14	1	0.2

Regularization using shape functions may decrease the correlation between parameters and this would improve the inversion results. This was done in Ref. 5 where we inverted for the slope and the offset of the water sound speed instead of inverting for the absolute sound speed at discrete points.

#### **II. EXAMPLES**

The examples will compare the hybrid approach in both wave number or range domain on both synthetic and real data to the classical genetic algorithms and Gauss–Newton methods.

For the GA the "standard" parameters<sup>4,5</sup> were used: Population size 32, crossover rate 0.8, mutation probability 0.05, and reproduction size 0.5. A full GA run will consist of several parallel populations each with 2000 forward models. For the hybrid method, a run will consist of one population with 1000–5000 forward models and five Gauss–Newton iterations or five forward models for each new individual. This corresponds to evaluating 200–1000 new individuals.

The examples will show comparisons between GA and the hybrid method. The main findings are summarized in Table I. The basis for comparison of the two methods is how fast the final estimate is obtained. How well the final estimate fits the data is expressed through the value of the object function, Eq. (1). How fast the solution is obtained is best expressed through the CPU-time, which depends on the actual implementation of the code. It could also be expressed by the number of forward modeling calls, but because the GN also needs a gradient the forward modeling calls are more CPU intensive for this method (about a factor 2-5 depending on the number of gradients to be computed). Finally, the speed can also be expressed as number of new individuals. For the hybrid method, computation of a new individual requires  $N_g$  forward models; using  $N_g=5$  makes a new individual a factor  $5 \times 2 = 10$  more intensive than the simple GA. This extra effort is often well spent.

It is instructive to bear in mind some limiting cases: (1) For a purely local problem GN is faster than the hybrid method which is faster than GA. The GN does not waste any time exploring the model space. (2) If the GA and hybrid method have the same number of new individuals the hybrid will always outperform the GA because it does additional GN steps which will give better or at worst the same fitness. (3) On very noisy data the hybrid method will just waste time computing gradients.



FIG. 4. The inversion result (dashed line) for the sound-speed profile using (a) GA, (b) Gauss-Newton, and (c) the hybrid method. The true solution is the solid line.

#### A. Inversion in wave-number domain

In order to describe the approach we first invert for the sound-speed profile given the simple environmental model given in Table II, which is similar to the environment given in Ref. 4 except that we have no shear. The bottom is parametrized into 10 layers, each 10 m thick, plus a basement layer; in total 11 layers. In order to obtain maximum response from the bottom, both the source and the receivers are placed on the seabed. The source frequency is 100 Hz and the magnitude of the horizontal wave-number spectrum is computed at 64 points equally spaced in the phase velocity range from 1200 to 3000 m/s. The amount of information available to retrieve the parameters here is rather limited in order to make the inversion more difficult. These examples are purely synthetic. Thus there is no consideration of how we should obtain the input data from actual experiments. For a discussion of how to transform real data to wave-number domain, see Refs. 11 and 12.

In the first example (case A1) we will only invert for the sound-speed profile in the bottom; all other parameters are kept at their correct values. First we solved the problem using GA alone with 25 populations, each consisting of 2000 forward models, which gives a total of 50 000 forward model calculations. The search interval for each parameter is 1500–3100 m/s. The best solution obtained is displayed in Fig. 4(a). The first layers are well determined, whereas the lower layers are less well determined. One advantage of this method is that we can obtain as estimate of the accuracy of the solution by plotting the marginal *a posteriori* probabilities,<sup>4</sup> Fig. 5. These show that as we get down to the lower layers the solution becomes more uncertain, and there is no well-defined peak.

The Gauss-Newton method is always sensitive to the starting values. We ran it with the initial model vector being

TABLE II. Environmental model for inversion.

Medium	Depth span (m)	Speed (m/s)	Atten. (dB/λ)	Density (kg/m <sup>3</sup> )
water	0-50	1500	0	1000
sediment 1	50-100	1600	0.1	1600
sediment 2	100 - 150	1800	0.1	2000
basement	150–∞	2800	0.1	2200

uniformly 1500 m/s. After 22 iterations the optimization did not improve and the result is given in Fig. 4(b). A food match is obtained down to 100 m, but below the match is unstable. This could indicate some stability problems with numerically insignificant gradients. A singular value decomposition was done at the correct values for all the eleven P-velocities (as a local uncertainty measure depends on the analysis point), Fig. 6. The inverse solution is a weighted sum of all the modes, see Eq. (12). Thus by inspection of the figure the deeper layers are seen to be mainly determined by the higher eigenvectors and the upper layers are determined by the lower eigenvectors. The lower eigenvectors are best determined and thus the upper layers would be more precisely determined by a local method. The deeper layers, corresponding to the higher eigenvectors (about 9-11), are harder to retrieve and, considering the associated singular values, may cause numerical instability if they are not damped.

For the same problem, the hybrid approach was run with five iterations for the Gauss-Newton method on each new individual. This was done for one population with 1000 forward models or 1000/5=200 new individuals in total. In this



FIG. 5. The marginal *a posteriori* distribution for each of the eleven sound speeds.



FIG. 6. The singular values and the corresponding eigenvectors.

case even the sound speed in the lowest layer is matched quite well, Fig. 4(c).

We also tried to increase the population size to 256 and let the Gauss-Newton run until convergence. This corresponds to selecting random starting points and using a local method from all these points. The fit by this method was about the same as for one Gauss-Newton iteration. Thus it seems better to use a guided random search than to use random starting points for the hybrid method.

The CPU times and values of the object function for this example (A1) are given in Table I. For this example the hybrid method is both much faster and more accurate. The actual convergence in the value of the object function versus the number of sampled models is displayed in Fig. 7. The number of sampled models is stacked since several populations are used for the GA.

Next (case A2) we show that we can successfully invert a many-parameter problem. This is done by parametrizing the sub-bottom into 11 layers and then inverting for all four of the physical parameters in each layer, i.e., eleven velocities (range 1500–3100 m/s), eleven attenuations (range 0–1 dB/ $\lambda$ ), eleven densities (range 1000–3000 kg/m<sup>3</sup>), and ten thicknesses (range 0–50 m). The initial values are selected randomly from the search interval. Each of the 43 parameters could take 256 values; this corresponds to a search space of



FIG. 7. Performance of the GA (solid line) and the hybrid (dashed line) for the inversions for sound speed. The only difference between (a) and (b) is the horizontal scale.

 $256^{43} = 10^{129}$ . Admittedly, this over-parametrization would normally be avoided, but the example shows the flexibility of the method.

For the hybrid approach we used one population with 1000 new individuals, or  $1000 \times 5 = 5000$  forward models. The GA used 25 parallel populations each with a size of 32 and running 2000 forward runs. In total 50 000 forward models were run. We did try to increase the number of populations drastically, but it did not improve the performance. The CPU times and values of the object function for this example (A2) are given in Table I. The retrieved parameters are as shown in Fig. 8. For the hybrid method only the attenuation and density below 150 m are not well determined, but this is due to their lack of importance for wave propagation. The solution by GA has minor deviations for most of the parameters.



FIG. 8. Inversion for thickness, speed, attenuation, and density: True solution (solid line), GA (dashed line), and hybrid (dotted line).



FIG. 9. Measurement configuration.

## B. Inversion of real transmission loss data

As an additional test of the algorithm we inverted real shallow water transmission loss data,<sup>22</sup> case B. The data are quite noisy and have only limited information. This data set has been analyzed by Lambert<sup>21</sup> using the author's GA code.

The experiment Tellaro took place in the Gulf of La Spezia in June 1992. The measurement configuration is sketched in Fig. 9: Four receivers are moored to the bottom at several depths (2, 5, 10, and 15 m) and a 330 Hz source is towed at a fixed depth of 7 m. The recording made with the 5-m-deep receiver was chosen because of its low noise compared to the others. The data, however, are still very noisy and we are obliged to smooth them by averaging the signal from six neighboring range values (Fig. 10).

From CTD measurements we have precise information on the water sound-speed profile and the water depth. It is reasonable to assume a constant water sound speed of 1523 m/s and a water depth of 16 m. Since the data are very noisy, especially at longer range, we use the signal between ranges 270 and 1600 m. In order to decrease the influence of the long-range data, they are weighted by  $1/\sqrt{r}$ , where r is the distance between source and receiver.

To describe the sediment, 10 layers of different thicknesses were used. Velocities and attenuation in each layer had to be retrieved, in total 20 parameters. The velocities and attenuations in the layers are coupled together using the shape functions in Fig. 11. Thus the actual profile is a weighted sum of the four shape functions, as described in



FIG. 10. Range-averaged signal for a receiver at 5-m depth.



FIG. 11. Shape functions for the speed and attenuation in the sediment.

Sec. I B. The first shape function gives a constant offset to the sediment sound speed, and shape functions 3 and 4 are used to describe the slope of the sediment sound speed. Shape function 2 was introduced to give more freedom to the first meter of the sediment, as it is very important. By using these 4 shape functions the number of unknowns is reduced from 20 to  $2 \times 4 = 8$ .

Earlier, a seismic broadband analysis at the Tellaro site was carried out to estimate the bottom sound-speed profile. The S-wave sound speed was estimated to be about 100 m/s<sup>7</sup> and therefore its influence on the transmission loss can be neglected. The P-wave sound-speed profile was estimated with the Herglotz–Wiechert analysis of the refracted arrival<sup>23,24</sup> and provided the broadband estimated soundspeed profile in Fig. 12. The bandwidth was 10–200 Hz. This profile is used for reference but it is not necessarily the correct one. As it is a broadband experiment it probably has a deeper penetration depth than single-frequency measurements because the lower frequencies penetrate further into the sediment.

For the inversion using GA we used 1000 forward modeling runs in ten parallel populations, in total 10 000 forward modeling runs. For the hybrid inversion we used only 1000 forward modeling in one population. Increasing the size of the population did not have any effect in this case. The value of the object function as found by the two methods are nearly



FIG. 12. Sound-speed profile for the Tellaro site determined by the broadband estimate (solid line), GA (dashed line), and the hybrid (dotted line).



FIG. 13. Pressure field for the Tellaro site: The data (solid line) and result (dashed line) by (a) GA and (b) the hybrid method. The y axis is arbitrarily scaled.

the same, 0.56 (GA) and 0.62 (hybrid). The inverted sound-speed profiles are shown in Fig. 12, and the corresponding fields for the two methods are shown in Fig. 13(a) and (b).

In this example the GA and the hybrid performed about equal, which is probably due to the rather noisy data, and thus the many local minima in the object function. In this example the CPU time used for the two examples is about equal, see Table I. For the hybrid method we have to compute the gradient for each parameter and then combine the gradients to obtain the derivatives of the shape functions. Thus we are actually computing 20 gradients and not just 8; the shape functions also call for a lot of computer bookkeeping. However, the shape functions are very useful in regularizing the solution.

The results of this inversion seem to support the conclusion of Ref. 21 that for a frequency of 330 Hz only the first few meters of the sediment are important for wave propagation. Therefore, information about the deeper layers can only be retrieved with considerable uncertainty. The fact that we obtain the same sound-speed profile for the first few meters using two different measurement methods and two different inversion approaches gives us a high degree of confidence in the solution for the first few meters of the sediment.

## **III. CONCLUSIONS**

A hybrid optimization method combining the global genetic algorithm and the local Gauss-Newton method has been developed. This method takes several gradient steps between each update of the object function for each "individual" in the population. For a large class of optimization problems the optimization is both faster and more accurate than by using the genetic algorithm or Gauss–Newton alone. This approach makes it feasible to solve problems with more parameters than if a global optimization was used alone, and we have the ability to use both local and global uncertainty estimates.

Due to the analytic derivation of the gradients a local method is limited to the particular forward model used, here SAFARI. It is, however, feasible to derive exact gradients for most forward models. In this respect the global optimization method is more flexible since the forward modeling method can be easily replaced.

## ACKNOWLEDGMENTS

This paper is based on inspiration from and collaboration with Andrea Caiti, Don Gringras, Finn Jensen, and Marc Lambert. Special thanks goes to Andrea Caiti for providing the experimental data.

### **APPENDIX: DERIVATIVES OF THE PRESSURE**

In this Appendix we derive analytical expressions for the pressure gradient using the direct global matrix method as implemented in SAFARI. Here only an outline of the DGM approach is given, focusing on those details necessary to obtain the derivatives. For further particulars see the published literature.<sup>15,16</sup>

#### 1. Green's functions for stratified media

We express the displacements and pressure for a field point x=(x,z) and of time dependence  $\exp(i\omega t)$  in vector form,

$$\mathbf{f}(\mathbf{x}) = \{u_z, -p\}^T(\mathbf{x}). \tag{A1}$$

The solution for a stratified medium is based on the integral transform representation of the Green's function, for cylindrical geometry the Hankel transformation,

$$\mathbf{f}(x) = \int_0^\infty \tilde{\mathbf{f}}(k) J_0(kx) k \ dk, \tag{A2}$$

$$\tilde{\mathbf{f}}(k) = \int_0^\infty \mathbf{f}(x) J_0(kx) x \ dx. \tag{A3}$$

The depth-dependent Green's function  $\mathbf{f}(k)$  satisfies the depth-separated wave equation.

The DGM approach solves the depth separated wave equation by expressing the total Green's function in layer l as a superposition of the free-field Green's function for sources in the layer, and the solutions to the homogeneous wave equation,

$$\tilde{\mathbf{f}}_l(k) = \tilde{\mathbf{f}}_l(k) + \tilde{\mathbf{f}}_l^*(k), \tag{A4}$$

where both the source contribution  $f_l(k)$ , and the homogeneous solution  $\tilde{f}_l^*(k)$  in a layer are of the form

$$\mathbf{f}_{l}^{*}(k) = \mathbf{K}_{l}(k)\mathbf{E}_{l}(z,k)\mathbf{D}_{l}(k), \qquad (A5)$$

where  $K_l(k)$  is a coefficient matrix which transforms the potentials to displacements and stresses. The diagonal matrix  $E_l(z,k)$  is the phase matrix which propagates the potentials from the interfaces. It contains all the depth dependence,

$$\mathbf{E}_{l}(z,k) = \operatorname{diag}[e^{-\alpha z^{-}}, e^{-\alpha z^{+}}], \qquad (A6)$$

where  $z^-=z-z_{l-1}$ ,  $z^+=z_l-z$  are the distances from the field point to the interfaces bounding the layer above and below, respectively, and the vertical wave number  $\alpha$  given by

$$\alpha = \sqrt{k^2 - h_I^2}.$$
 (A7)

Here, h is the compressional wave number given by

$$h = \frac{\omega}{c_p} (1 - i\,\delta_p) = \frac{\omega}{c_p} \left( 1 - i\frac{\gamma}{40\,\pi\,\log\,e} \right),\tag{A8}$$

where  $\gamma_p$  is the attenuation per wavelength.  $\mathbf{D}_l$  in Eq. (A5) is a vector containing wave field amplitudes, known for the source contribution and determined from the boundary conditions for the homogeneous solution.

The boundary conditions to be satisfied simultaneously at all horizontal interfaces (L-1) then take the form

$$\tilde{\mathbf{f}}_{l-1}^{*} - \tilde{\mathbf{f}}_{l}^{*} = \hat{\tilde{\mathbf{f}}}_{l} - \hat{\tilde{\mathbf{f}}}_{l-1}, \quad l = 1, \dots, (L-1).$$
 (A9)

This is a global system of linear equations in the wave-field amplitudes  $\mathbf{D}_l$ , the solution of which is unconditionally stable with the particular choice of local coordinate systems used in the depth solutions, Eq. (A6). Once solved, the solution vector  $\mathbf{D}_l$  is inserted into Eq. (A5) and superposed with the source contributions.

Expressed in matrix form the system of equations in Eq. (A9) can be written as

$$\mathbf{AD} = \mathbf{F},\tag{A10}$$

where  $\mathbf{F}$  is the source vector and  $\mathbf{D}$  is the wave-length amplitudes. The coefficient matrix  $\mathbf{A}$  is a block-banded matrix. For each layer it has the structure

$$\mathbf{A}_{l} = \begin{bmatrix} \alpha & -\alpha E \\ \rho \omega^{2} & \rho \omega^{2} E \\ -\alpha E & \alpha \\ -\rho \omega^{2} E & -\rho \omega^{2} \end{bmatrix}, \qquad (A11)$$

where  $E = e^{-\alpha z_l}$ .

#### 2. The Jacobian

In the derivation of the derivatives we will assume that these are not taken in a source or receiver layer. Thus we will only consider the inhomogeneous contribution, Eq. (A5). The source and receiver layers are trivial extensions. The derivatives in an elastic medium have also been derived; these are considerably more complicated due to the doubling of the matrices involved.

To find the derivatives with respect to the medium properties we differentiate Eq. (A2),

$$\frac{\partial \mathbf{f}(x)}{\partial \theta} = \int_0^\infty \frac{\partial \bar{\mathbf{f}}(k)}{\partial \theta} J_0(kx) k \ dk. \tag{A12}$$

Thus the derivatives in the range domain is found by integrating the contributions from each wave number. In the following we concentrate on obtaining the derivatives for one wave number. Neglecting the inhomogeneous contribution, the derivative of the field can be expressed by differentiation of Eq. (A5)

$$\frac{\partial \mathbf{\tilde{f}}(k)}{\partial \theta} = \frac{\partial \mathbf{K}}{\partial \theta} \mathbf{E} \mathbf{D} + \mathbf{K} \frac{\partial \mathbf{E}}{\partial \theta} \mathbf{D} + \mathbf{K} \mathbf{E} \frac{\partial \mathbf{D}}{\partial \theta}.$$
 (A13)

By differentiation of Eq. (A10) the derivative of the potential function  $\partial D/\partial \theta$  is found by solving the system of equations

$$\mathbf{A}\frac{\partial \mathbf{D}}{\partial \theta} = \frac{\partial \mathbf{F}}{\partial \theta} - \frac{\partial \mathbf{A}}{\partial \theta} \mathbf{D}.$$
 (A14)

Note that the coefficient matrix A is unchanged. Thus to find the derivatives we just have to solve the system of equations for an extra right-hand side for each parameter in the parameter vector  $\theta$ . This can be done at little extra cost.

For the derivatives of the coefficient matrix with respect to the layer parameters in layer l,  $\partial A/\partial \theta_l$ , only that part of the matrix representing layer l,  $\partial A_l/\partial \theta_l$ , is nonzero. The following derivatives are obtained:

$$\frac{\partial \mathbf{A}_{l}}{\partial z} = -\alpha E \begin{bmatrix} 0 & -\alpha \\ 0 & \rho \omega^{2} \\ -\alpha & 0 \\ -\rho \omega^{2} & 0 \end{bmatrix}, \quad (A15)$$
$$\frac{\partial \mathbf{A}_{l}}{\partial \rho} = \omega^{2} \begin{bmatrix} 0 & 0 \\ 1 & E \\ 0 & 0 \\ -E & -1 \end{bmatrix}, \quad (A16)$$
$$\frac{\partial \mathbf{A}_{l}}{\partial \alpha} = \begin{bmatrix} 1 & -(1-\alpha z)E \\ 0 & -\rho \omega^{2} zE \\ -(1-\alpha z)E & 1 \end{bmatrix}. \quad (A17)$$

From Eq. (A17) we can find the derivatives of the velocities and attenuation by computing the derivatives of the vertical wave number

 $\rho \omega^2 z E$ 

$$\frac{\partial \alpha}{\partial h} = -\frac{h}{\alpha},\tag{A18}$$

0

$$\frac{\partial \alpha}{\partial c_p} = \frac{\partial \alpha}{\partial h} \frac{\partial h}{\partial c_p} = \frac{-h}{\alpha} \frac{-h}{c_p} = \frac{h^2}{\alpha c_p},$$
(A19)

$$\frac{\partial \alpha}{\partial \gamma} = \frac{\partial \alpha}{\partial h} \frac{\partial h}{\partial \gamma} = \frac{-h}{\alpha} \frac{\omega}{c_p} \frac{-i}{40\pi \log e} = \frac{h}{\alpha} \frac{\omega}{c_p} \frac{i}{40\pi \log e}.$$
(A20)

The above describes the basic details for obtaining the derivatives. Often we are not interested in the derivative of the complex pressure itself, but only the derivative of the amplitude and/or the unit pressure vector:

Amplitude: The above is based on a complex pressure field  $P=P_r+iP_i$ . Often the optimization is only based on the amplitude  $|P| = \sqrt{P_r^2 + P_i^2}$ . Then the derivatives are

$$\frac{\partial |P|}{\partial \theta} = \left(\frac{\partial P_r}{\partial \theta} P_r + \frac{\partial P_i}{\partial \theta} P_i\right) / |P|.$$
 (A21)

Unit-pressure vector: The unit pressure is defined as

$$\hat{\mathbf{p}} = \frac{\mathbf{p}}{(\mathbf{p}^T \mathbf{p})^{1/2}}.$$
 (A22)

The derivative is

$$\hat{\mathbf{J}} = \frac{\partial \hat{\mathbf{p}}}{\partial \theta} \tag{A23}$$

$$= \frac{\mathbf{J}}{(\mathbf{p}^T \mathbf{p})^{1/2}} - \hat{\mathbf{p}} \operatorname{Re}\left\{ \hat{\mathbf{p}}^T \frac{\mathbf{J}}{(\mathbf{p}^T \mathbf{p})^{1/2}} \right\},$$
(A24)

where the superscript T indicates conjugate transpose.

- <sup>1</sup>M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean-bottom properties," J. Acoust. Soc. Am. 92, 2770-2783 (1992).
- <sup>2</sup>C. E. Lindsay and N. R. Chapman, "Matched field inversion for geophysical parameters using adaptive simulated annealing," IEEE J. Oceanic Eng. 18, 224-231 (1993).
- <sup>3</sup>S. E. Dosso, M. L. Yeremy, J. M. Ozard, and N. R. Chapman, "Estimation of ocean bottom properties by matched-field inversion of acoustic field data," IEEE J. Oceanic Eng. 18, 232-239 (1993).
- <sup>4</sup>P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions," J. Acoust. Soc. Am. 95, 770-782 (1994)
- <sup>5</sup>P. Gerstoft, "Global inversion by genetic algorithms for both source position and environmental parameters," J. Comp. Acoust. 2, 251-266 (1994).
- <sup>6</sup>S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow water," J. Acoust.Soc. Am. 82, 998-1017 (1987).
- <sup>7</sup>A. Caiti, T. Akal, and R. D. Stoll, "Estimation of shear wave velocity in shallow marine sediments," IEEE J. Oceanic Eng. 19, 58-72 (1994).
- <sup>8</sup>L. R. Lines and S. Treitel, "A review of least squares inversion and its application to geophysical problems," Geophys. Prosp. 32, 159-186 (1984).

- <sup>9</sup>J. A. Scales, P. Docherty, and A. Gersztenkorn, "Regularisation of nonlinear inverse problems: imaging the near-surface weathering layer," Inverse Problems 6, 115-131 (1990).
- <sup>10</sup>F. Kornendi and M. Dietrich, "Nonlinear waveform inversion of planewave seismogram in stratified elastic media," Geophysics 56, 664-674 (1991).
- <sup>11</sup>L. Amundsen and B. Ursin, "Frequency-wavenumber inversion of acoustic data," Geophysics 56, 1027-1039 (1991).
- <sup>12</sup>H. Zhao, B. Ursin, and L. Amundsen, "Frequency-wavenumber elastic inversion of marine seismic data," Geophysics 59, 1868-1881 (1994). <sup>13</sup>P. W. Cary and C. H. Chapman, "Automatic 1D waveform inversion of
- marine seismic refraction data," Geophys. J. 93, 527-546 (1988).
- <sup>14</sup>L. Fishman and M. D. Collins, "Characterizing parameter spaces," in Second European Conference on Underwater Acoustics, edited by L. Bjørnø (European Commission, Brussels, Belgium, 1994).
- <sup>15</sup>H. Schmidt and F. B. Jensen, "A full wave solution for propagation in multilayered viscoelastic media with application to Gaussian beam reflection at fluid-solid interfaces," J. Acoust. Soc. Am. 77, 813-825 (1985).
- <sup>16</sup>H. Schmidt and G. Tango, "Efficient global matrix approach to the computation of synthetic seismograms," Geophys. J. R. Astron. Soc. 84, 331-359 (1986).
- <sup>17</sup>P. L. Stoffa and M. K. Sen, "Multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms," Geophysics 56, 1794-1810 (1991).
- <sup>18</sup>J. A. Scales, M. L. Smith, and T. L. Fisher, "Global optimization methods for highly nonlinear inverse problems," J. Comp. Physics 103, 258-268 (1992).
- <sup>19</sup>M. Sambridge and G. Drijkoningen, "Genetic algorithms in seismic waveform inversion," Geophys. J. Int. 109, 323-343 (1992).
- <sup>20</sup>A. Tarantola, Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation (Elsevier, Amsterdam, 1987).
- <sup>21</sup>M. Lambert, "Inversion of seismo-acoustic real data using genetic algorithms," SM-276, SACLANT Undersea Research Centre, La Spezia, Italy (1994).
- <sup>22</sup> The transmission loss data were taken by the Seafloor Acoustics Group at SACLANTCEN; the broadband estimate was done by Andrea Caiti.
- <sup>23</sup>K. Aki and P. G. Richards, Quantitative Seismology: Theory and Methods (Freeman, San Francisco, 1980).
- <sup>24</sup>T. Akal, A. Caiti, and R. D. Stoll, "Remote sensing in shallow-water marine sediments," in Acoustic Classification and Mapping of the Seabed (Institute of Acoustics, Bath, UK, 1993).