

Simulated Annealing Algorithm for Model Reconstruction of the Four-Layer Medium with Elliptical Inclusion in the Third Layer

Dmitry Prokhorov[1(✉)], Galina Reshetova[2], and Denis Bratchikov[3]

[1] Sobolev Institute of Mathematics SB RAS, 4 Koptug Avenue, 630090 Novosibirsk, Russia
prokhorovdi@ipgg.sbras.ru

[2] Institute of Computational Mathematics and Mathematical Geophysics SB RAS, 6 Lavrentev Avenue, 630090 Novosibirsk, Russia
kgv@nmsf.sscc.ru

[3] Novosibirsk State University, 2 Pirogova Street, 630090 Novosibirsk, Russia
d.bratchikov@g.nsu.ru

Abstract. The paper presents an approach for reconstructing the properties of two-dimensional viscoelastic medium with defined geometry using the simulated annealing algorithm. The inverse problem solution requires a lot of computational resources because the direct seismic modeling is performed at each iteration. The staggered grid finite-difference scheme is implemented using CUDA technology to speed up the solution of the direct problem by parallelization. The choice of the simulated annealing method for solving inverse problem is due to the method's ability to avoid local minima of the target functional. However, the simulated annealing method needs a good coverage of the model space by realizations of random probing vectors. It leads to enormous computation time in the case of a four-layer medium with elliptical inclusion in the third layer, which has 37 parameters. Therefore, the sequential reconstruction of model parameters, where the simulated annealing algorithm searches for the parameters in 1 or 2-D subspace, is introduced. Nevertheless, the attenuation properties of medium were not reconstructed by simulated annealing. For their study, the deep convolutional neural network is used.

Keywords: simulated annealing · viscoelastic medium · seismic modeling

1 Introduction

Seismic exploration is one of the advanced tools used to study the internal geological structure of the Earth. The construction of a reliable depth model of the environment is a prerequisite for the search for oil and gas fields. The most important recoverable parameters in the study of hydrocarbon reservoirs are the speed of propagation of seismic waves since it allows us to determine the depth

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024 Workshops, LNCS 14817, pp. 334–351, 2024.
https://doi.org/10.1007/978-3-031-65238-7_23

of rocks and the absorption of seismic energy since it characterizes the fluid saturation of rocks.

In the 20th century, many seismic exploration methods were developed, ranging from reflected and refracted wave methods to various summation and migration procedures. However, in recent years, the requirements for the detail and reliability of predictions obtained during the interpretation of seismic data have increased significantly since the search for new deposits is conducted in areas with complex geological structures. The full waveform inversion method (FWI) [16] is one of the reliable and high-resolution methods used to estimate the parameters of complex medium by the results of measuring wave fields. It solves the inverse dynamic seismic problem by sequential solutions of the direct problem and minimizing the functional discrepancy between the observed and synthetic data. For the first time, the FWI in the time domain using the generalized least squares method was proposed in a series of papers by [1,13]. FWI was developed in the frequency domain [11] to accelerate the inversion. However, FWI has several disadvantages:

1. Ill-posedness of the inverse problem: non-uniqueness of the solution and instability to noisy data.
2. Strong nonlinearity of the target functional—convergence of the method to a local minimum.
3. The sensitivity of the solution to the presence of low-frequency information in seismic data.
4. The presence of the connection between parameters when solving a multiparametric inverse problem leads to incorrect parameter reconstruction.
5. The high computational cost of wave fields computation at each nonlinear least squares method iteration.

Based on the above, when developing an algorithm for restoring the depth model of velocity and attenuation parameters by full wave field, two main areas of optimization are distinguished: the speed of solving a direct problem and the ability of the algorithm to avoid local minima.

A large number of numerical methods are known to solve the problem of seismic wave propagation: the finite difference method [15], the finite volume method [18], the finite element method [2], the spectral element method [7]. The main advantages of the finite difference method include the simplicity of constructing the algorithm and its software implementation, lower computational resource requirements relative to other methods, and the ability to approximate environments of arbitrary geometric complexity with sufficient accuracy for seismic modeling [17]. The algorithms are easily parallelized using CUDA technology, and data exchange between RAM and video memory is minimal.

Global optimization methods can move more freely through the parameter space than gradient descent. Examples include simulated annealing [14], genetic algorithms, and particle swarm optimization [10]. The work of the simulated annealing method [6] is similar to the process occurring during the crystallization of a substance: the algorithm can assume a state with a higher cost than the current one. However, the longer the algorithm works, the lower the 'temperature", i.e., the probability of deciding with a higher cost. Thus, the main

advantage of the simulated annealing algorithm is the possibility of getting out of the local minimum state.

Two tasks have been set as part of the work:

1. To develop and programmatically implement using CUDA technology a finite difference scheme on a staggered grid [15] for a system of viscoelasticity equations.
2. Based on the method of simulated annealing to develop and implement an algorithm for solving the inverse dynamic problem of seismic.

We use the developed algorithm to reconstruct a four-layer medium's geometric and velocity parameters with elliptical inclusion in the third layer. Also, the order of their reconstruction under the condition of limited computing resources is established.

2 Problem Statement

The system of equations describing wave propagation in a two-dimensional isotropic medium with absorption, written in Cartesian coordinates, has the following "velocity-stress" formulation:

$$\begin{cases} \rho\dfrac{\partial u}{\partial t} = \dfrac{\partial \sigma_{xx}}{\partial x} + \dfrac{\partial \sigma_{xz}}{\partial z} + f(t,x,z), \\ \rho\dfrac{\partial w}{\partial t} = \dfrac{\partial \sigma_{zx}}{\partial x} + \dfrac{\partial \sigma_{zz}}{\partial z} + h(t,x,z), \\ \dfrac{\partial \sigma_{xx}}{\partial t} = (\lambda+2\mu)\dfrac{\partial u}{\partial x} + \lambda\dfrac{\partial w}{\partial z} + \xi_{xx}, \\ \dfrac{\partial \sigma_{zz}}{\partial t} = \lambda\dfrac{\partial u}{\partial x} + (\lambda+2\mu)\dfrac{\partial w}{\partial z} + \xi_{zz}, \\ \dfrac{\partial \sigma_{xz}}{\partial t} = \dfrac{\partial \sigma_{zx}}{\partial t} = \mu\left(\dfrac{\partial u}{\partial z} + \dfrac{\partial w}{\partial x}\right) + \xi_{xz}, \\ \dfrac{\partial \xi_{xx}}{\partial t} + \dfrac{1}{\tau_\sigma}\xi_{xx} = (\Lambda+2M)\dfrac{\partial u}{\partial x} + \Lambda\dfrac{\partial w}{\partial z}, \\ \dfrac{\partial \xi_{zz}}{\partial t} + \dfrac{1}{\tau_\sigma}\xi_{xx} = \Lambda\dfrac{\partial u}{\partial x} + (\Lambda+2M)\dfrac{\partial w}{\partial z}, \\ \dfrac{\partial \xi_{xz}}{\partial t} + \dfrac{1}{\tau_\sigma}\xi_{xx} = M\left(\dfrac{\partial u}{\partial z} + \dfrac{\partial w}{\partial x}\right), \end{cases} \tag{1}$$

where $\boldsymbol{u} = (u, w)$—velocity vector, $\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xz} \\ \sigma_{zx} & \sigma_{zz} \end{pmatrix}$—stress tensor, $\xi = \begin{pmatrix} \xi_{xx} & \xi_{xz} \\ \xi_{zx} & \xi_{zz} \end{pmatrix}$—memory variables, ρ—constant density, $\lambda, \mu, \Lambda, M, \tau_\sigma$—coefficients, characterizing the viscoelastic properties of the medium. In this case, the state of the viscoelastic medium $\boldsymbol{u}, \sigma, \xi$ depends on the spatial variables x, z and time t, and the parameters of the medium $\rho, \lambda, \mu, \Lambda, M, \tau$ depend only on spatial variable, the system 1 is defined in the half-plane $z \leq 0$.

For a system of Eqs. 1 the initial boundary value problem is formulated in the half-plane $\boldsymbol{R}^2 z \leq 0$ with zero initial data:

$$\begin{aligned} u|_{t=0} = 0, \quad w|_{t=0} = 0, \\ \sigma_{xx}|_{t=0} = 0, \quad \sigma_{xz}|_{t=0} = 0, \quad \sigma_{zz}|_{t=0} = 0, \end{aligned} \tag{2}$$

and the conditions at the boundary $z = 0$ corresponding to the stress-free surface

$$\sigma_{zz}|_{z=0} = 0, \quad \sigma_{xz}|_{z=0} = 0. \tag{3}$$

The search for a solution is carried out for the time interval $[0, T]$.

For the numerical solution of the initial boundary value problem for a system of Eqs. 1 with initial data 2 and boundary conditions 3, we use a staggered grid scheme [8,15] with second order approximation of finite-difference operators for both spatial and time variables. The finite difference scheme is implemented in C++ using CUDA technology.

3 Formulation of the Inverse Dynamic Seismic Problem

On the free surface $z = 0$ at the specified point $x = x_i (i = 1, ..., N)$ there is a system of receivers recording the coordinates of the velocity vector $\boldsymbol{u}^j(t_k, x_i, 0) = (u_i^j(t_k), w_i^j(t_k))$, $1 \leq i \leq N$, $1 \leq j \leq P$, $1 \leq k \leq K$ at time points t_k corresponding to the j impact on the system. At points $(x_j, 0)$, a Ricker wavelet with a dominant frequency f is applied:

$$R(t) = \left(1 - 2(\pi f(t - \frac{2}{f}))^2\right) e^{-\left(\pi f(t-\frac{2}{f})^2\right)}.$$

The symbol Y_j denotes the vector corresponding to the seismogram for the j impact:

$$Y_j = (u_1^j(t_1), w_1^j(t_1), u_2^j(t_1), w_1^j(t_2), ..., u_1^j(t_2), w_1^j(t_2), ..., u_N^j(t_K), w_1^N(t_K)),$$

and the symbol Y is:

$$Y = (Y_1 \cdot Y_2 \cdot ... \cdot Y_P),$$

where '$\cdot$' is a vector concatenation operation. Thus, Y is a vector of length $2 \times N \times K \times P$. On the vector space of seismograms $\boldsymbol{Y}$, a 2-norm is set ($||x||_2 = \sqrt{x_1^2 + ... + x_{2NKP}^2}$). Next, $\boldsymbol{Y}$ is used to denote the normed vector space of seismograms.

Let us introduce a model of a viscoelastic medium $m = (\rho, \lambda, \mu, L, M, \tau) \in M$ (M—data space) and define the operator of the direct problem $F(m) : M \to \boldsymbol{Y}$, matching the models m vector $F(m) \in \boldsymbol{Y}$, built according to the solution of the system 1 with the parameters m. Then the inverse problem can be formulated as follows [1]:

$$\begin{aligned} \text{Find } \tilde{m} \in M \mid E(\tilde{m}) = \min_{m \in M} E(m), \\ E(m) = ||F(m) - Y_d||, \end{aligned} \tag{4}$$

where Y_d is a vector that corresponds to the seismograms obtained by the observation system.

4 Data Space for a Four-Layer Medium with Elliptical Inclusion in the Third Layer

In general, M is a space of vector functions. However, in the case of a four-layer medium with an elliptical inclusion in the third layer (Fig. 1), the definition domain of the functional F is significantly simplified. Let $(\rho_i, \lambda_i, \mu_i, L_i, M_i, \tau_i)$ denote the parameters of the medium on the i-th layer for $1 \leq i \leq 4$, $i = 5$ corresponds to the parameters of the medium in an elliptical inclusion. The parameters h_1, h_2, h_3 are the thicknesses of the layers, (x_0, y_0)—the center of the ellipse, a, b—its semi-major and semi-minor axes. Thus, the model is completely described by a vector of 37 parameters, each a scalar.

It is more convenient to use another parameterization of the medium model using velocities and relaxation times instead of Lame coefficients [12]:

$$\begin{aligned} \lambda &= \pi_0 \frac{\tau_\epsilon^p}{\tau_\sigma} - 2\mu_0 \frac{\tau_\epsilon^s}{\tau_\sigma},\ \mu = \mu_0 \frac{\tau_\epsilon^s}{\tau_\sigma},\ \tau = \tau_\sigma, \\ \Lambda &= \frac{1}{\tau_\sigma}\left(\pi_0\left(1 - \frac{\tau_\epsilon^p}{\tau_\sigma}\right) - 2\mu_0\left(1 - \frac{\tau_\epsilon^s}{\tau_\sigma}\right)\right), \\ M &= \frac{1}{\tau_\sigma}\mu_0\left(1 - \frac{\tau_\epsilon^s}{\tau_\sigma}\right),\ \mu_0 = \rho V^{s2},\ \pi_0 = \rho V^{p2}. \end{aligned} \tag{5}$$

where τ_σ—stress relaxation time, $\tau_\epsilon^p, \tau_\epsilon^s$—relaxation times of deformations of P- and S-waves, respectively, μ_0—relaxation modulus for S-waves, analogous to the Lame coefficient μ in elasticity, π_0—relaxation modulus for P-waves, analogous to the sum of the Lame coefficients $\lambda + 2\mu$ in elasticity. In the viscoelastic case, V^p and V^s in the parametrization 5 do not coincide with the actual propagation velocities of P and S waves in the medium. Thus, the parameters of the medium in the layer (ellipse) are set by six parameters $(\rho_i, V^p, V^s, \tau_\epsilon^p, \tau_\epsilon^s, \tau_\sigma)$ and the length of the desired vector remains equal to 37. The simulated annealing algorithm is used to find it.

5 The Use of the Simulated Annealing Algorithm for Solving the Inverse Dynamic Seismic Problem

The minimum is searched in the rectangular subdomain M_{sa} of the space M. The subdomain M_{sa} is defined by vectors of upper and lower bounds:

$$\begin{aligned} m^u &= (\rho_1^u, \lambda_1^u, ..., a^u, b^u), \\ m^l &= (\rho_1^l, \lambda_1^l, ..., a^l, b^l), \\ M_{sa} &= [\rho_1^l, \rho_1^u] \times [\lambda_1^l, \lambda_1^u] \times ... \times [a^l, a^u] \times [b^l, b^u]. \end{aligned}$$

The vectors of the upper and lower bounds are selected based on some a priori estimates of the medium parameters.

Description of the simulated annealing algorithm in terms of the inverse problem 4:

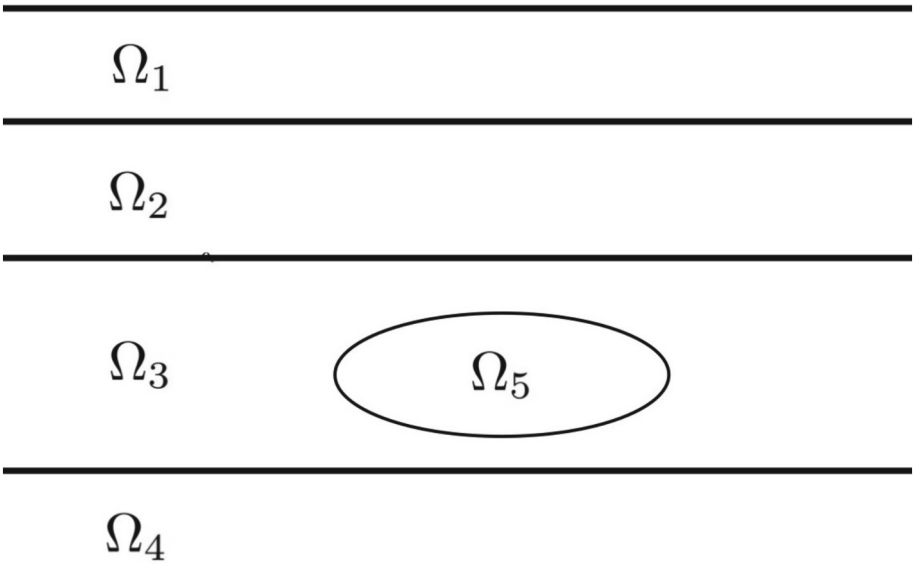


Fig. 1. Four-layer medium with elliptical inclusion in the third layer

1. *Initialization:* calculate the cost of $E(m_1)$ for a random vector of medium parameters $m_1 \in M_{sa}$. Initialize the best state $m_b = m_1$. Assign $k = 0, n = 0$. Set the initial values of the acceptance temperature $T_k^{ac} = T_0^{ac}$ (affects the probability of accepting a state with a higher cost) and the generation temperature $T_k = T_0$ (affects the correlation between the current vector and the probing one).
2. *Generation:* generate a probing vector $m_2 = m_1 + \epsilon$, where $\epsilon = g(T_k)$ is a random variable. In the case of $m_2 \notin M_{sa}$, consider the components of the vector m_2 that go beyond the range to be equal to the corresponding values at the boundary. Find the cost of the probing vector $E(m_2)$.
3. *Acceptance:* Assign $m_1 = m_2$ if $E(m_2) \leq E(m_1)$, otherwise assign $m_1 = m_2$ with probability $A(m_1 \rightarrow m_2, T_k^{ac})$. In the case of $E(m_2) \leq E(m_b)$ assign $m_b = m_2$. Increase $n = n + 1$. If $n < N$, go to step 2, otherwise to step 4.
4. *Temperature decreasing:* reduce temperatures according to schedules $T_{k+1} = U(T_k, k)$, $T_{k+1}^{ac} V(T_k^{ac}, k)$. Assign $m_1 = m_b$, $n = 0$. Increase $k = k + 1$. If $k < K$, go to step 2. Otherwise, the algorithm terminates with the best state of m_b.

Schedules U, V, distribution g, and probability of acceptance A are the algorithm's parameters. The following expressions are used to solve the inverse problem:

$$A(m_1 \rightarrow m_2, T_k^{ac}) = \exp\left(-\frac{E(m_2) - E(m_1)}{T_k^{ac}}\right),$$

$$U(T_k) = \alpha T_k, V(T_k^{ac}) \quad = \alpha^{ac} T_k^{ac}.$$

$$g(T_k) = (\epsilon_1 \epsilon_2 T_k, ..., \epsilon_1 \epsilon_2 T_k),$$

where ϵ_1 is a random variable taking the value 1 with probability p and the value 0 with probability $1 - p$, ϵ_1 is a random variable with a standard normal distribution. The ability to use the same expressions for all components of the vector ϵ is achieved by normalizing the search ranges of the components of the vector m.

The algorithm has seven parameters:

1. N—the number of iterations of the inner loop.
2. K—the number of iterations of the outer loop.
3. T_0—generation temperature.
4. T_0^{ac}—acceptance temperature.
5. α— the decrease "rate" of the generation temperature.
6. α^{ac}—the decrease "rate" of the acceptance temperature.
7. p—the probability of changing the component of the current state when generating a probing solution.

The algorithm takes a vector of upper and lower bounds as input. The output is the vector $m_b \in \Omega_{sa}$ with the lowest cost $E(m_b) \leq (m)$, among all vectors $m \in \Omega_{sa}$ tested by the algorithm.

6 Medium Parameters Reconstruction

The simulated algorithm belongs to the group of Monte Carlo methods. Therefore, to accurately restore the model parameters, the probing vectors must cover the domain Ω_{sa} good enough. For example, if the probing vectors are not random but located in nodes of a uniform grid with ten nodes for each dimension, then 10^{37} solutions to the direct problem would be required. Carrying out such many calculations in practice is impossible in the current state of computing technologies.

Based on the simulated annealing algorithm analysis, it was found to primarily restore the model's most significant (in terms of error E) parameters. Example: the Rayleigh wave makes the most significant error, the speed of which depends on the speed of the S-wave in the first layer. Therefore, we assume that the model parameters can be found sequentially.

As a result of solving a series of inverse problems, it was found that the developed algorithm cannot restore the parameters ρ_i. For known density values, the algorithm restores the values of the relations $\frac{\tau_\epsilon^p}{\tau_\sigma}$ and $\frac{\tau_\epsilon^s}{\tau_\sigma}$, but not the absolute value of τ_σ. At an unknown density, restoring the ratio of relaxation times is impossible. Therefore, when solving a model problem, the medium is assumed to be elastic, and $\rho = const$ in the entire computational domain.

In order to find the upper and lower boundaries of the search and determine the hierarchy of the model parameters, seismograms are analyzed by the researcher. The coordinates of the center of the ellipse along the x axis are found by analyzing the zero-offset seismogram (Fig. 2). The flat boundaries of

the layers give straight lines of reflections at zero-offset; the arcs correspond to reflections from the ellipse. Their center of symmetry is $x = 2600$—the center of the ellipse. A seismogram recorded for the source $x = 4000$ (Fig. 3) is used to restore the parameters of the layers, and to restore the geometric parameters of the inclusion, a seismogram recorded for the source $x = 2600$ (Fig. 4), that is, when calculating the cost of E, only one seismogram is used in each experiment.

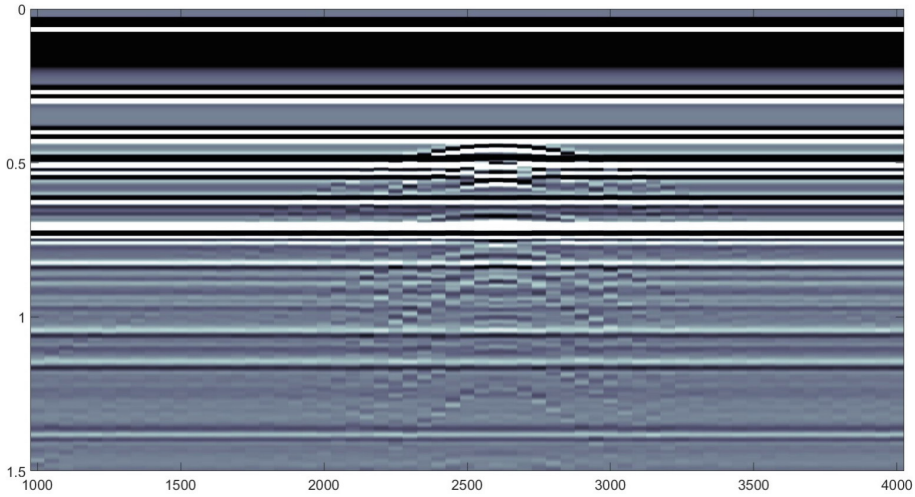


Fig. 2. Zero-offset seismogram

In order to restore the parameters of the medium, eight numerical experiments were performed. The parameters of the simulated annealing algorithm for each of them are shown in the Table 1.

Table 1. Parameters of the simulated annealing algorithm for eight numerical experiments

Number	N	K	T_0	T_0^{ac}	α	α^{ac}	p
1	10	3	0.15	0.5	0.5	0.25	1
2	10	3	0.15	0.5	0.5	0.25	1
3	15	3	0.15	0.5	0.5	0.25	1
4	20	3	0.15	0.05	0.5	0.25	1
5	30	3	0.15	0.5	0.5	0.25	1
6	20	2	0.15	0.05	0.5	0.25	1
7	30	3	0.15	0.05	0.5	0.25	1
8	100	2	0.15	0.05	0.5	0.25	1

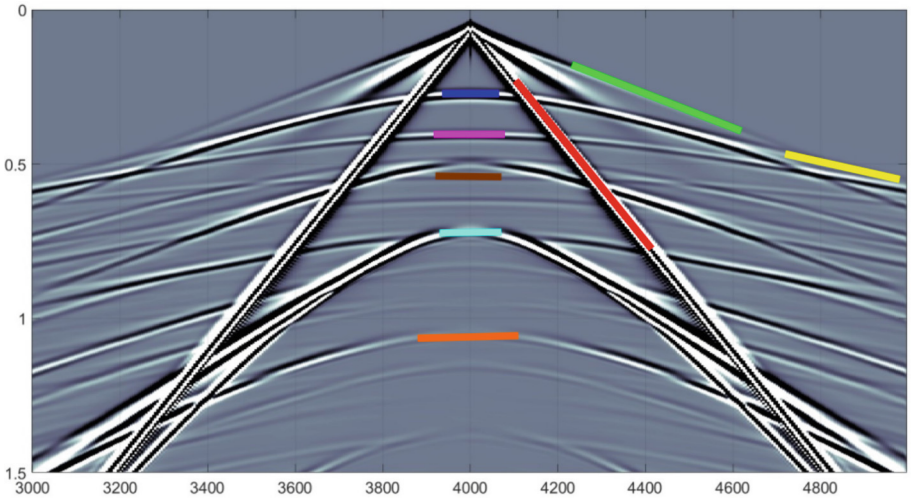


Fig. 3. Seismogram for a source with offset 4000 m. Red—Rayleigh wave. Green—direct P-wave. Blue—P-wave reflected from the second layer. Yellow—P-wave refracted by the second layer. Magenta—P-wave reflected from the third layer. Cyan—S-wave reflected from the second layer. Orange—S-wave reflected from the third layer. Brown—P-wave reflected from the fourth layer (Color figure online).

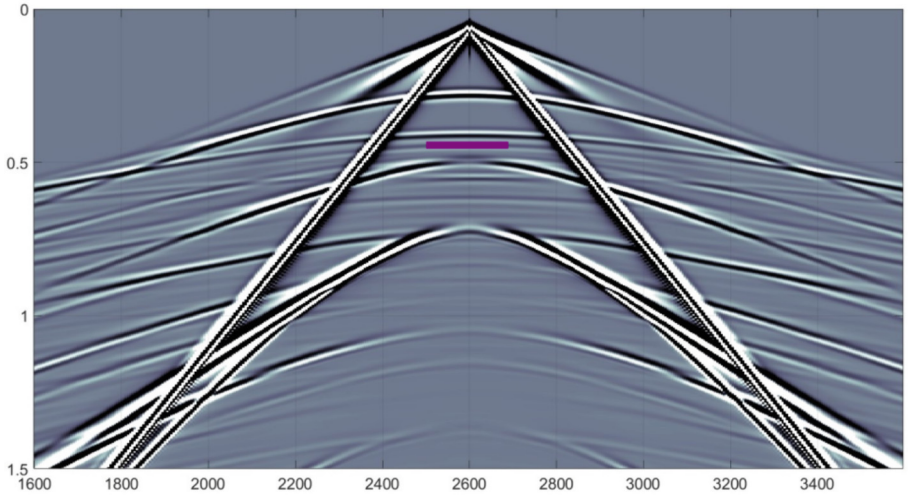


Fig. 4. Seismogram for a source with offset 2600 m. Purple—P-wave reflected from the ellipse top boundary (Color figure online).

The parameters of the finite-difference scheme for solving the direct problem for each numerical experiment are given in the Table 2. The width of the absorbing layers is 20 points.

Table 2. Parameters of a finite-difference scheme for solving a direct problem for eight numerical experiments

Number	N_x	N_z	h_x (m)	h_z (m)	τ (s)	T (s)
1	2201	200	1	1	0.0001	0.35
2	2201	300	1	1	0.0001	0.35
3	2201	300	1	1	0.0001	0.35
4	2201	500	1	1	0.0001	0.45
5	2201	500	1	1	0.0001	1.5
6	2201	850	1	1	0.0001	0.7
7	2201	850	1	1	0.0001	0.5
8	2201	850	1	1	0.0001	0.7

Step 1. Constant values are set for calculations: $\rho = 1500, \tau_\sigma = \tau_\epsilon^p = \tau_\epsilon^s = 0.005$. The dominant frequency of the Ricker wavelet is calculated $f = 30$ Hz.

Step 2. The approximate value of the Rayleigh wave velocity is found by the slope coefficient k_1 of the line 1 (Fig. 3).

$$V^r = \frac{1}{k_1} = 565.$$

It is assumed that the velocity of the S-wave on the first layer is close to the velocity of the Rayleigh wave. Therefore, visual analysis of seismograms fails to establish a line corresponding to the S-wave.

Similarly, the approximate value of the velocity of the P-wave in the first layer is found by the slope coefficient k_2 of the line 2 (Fig. 3).

$$V_1^{p,init} = \frac{1}{k_1} = 1840.$$

The found velocity values are used as an approximate model of the medium and to set the search boundaries in the simulated annealing algorithm.

Experiment 1. Reconstruction of the first layer S-wave velocity. Medium model: homogeneous medium, P-wave velocity—$V_1^{p,init}$, S-wave velocity V_1^s—unknown parameter. The subdomain of the seismogram is involved in calculating E: $t \leq 0.35$. The search range of the simulated annealing algorithm is $V_s^1 \in [V^r, 1.1V^r]$. Found value: $V_1^s = 603$.

Experiment 2. Reconstruction of the first layer P-wave velocity. Medium model: homogeneous medium, P-wave velocity—unknown parameter, S-wave velocity—V_1^s. The subdomain of the seismogram is involved in calculating E: $t \leq 0.5$ with a Rayleigh wave cut out. Search range: $V_1^p \in [0.9V_1^{p,init}, 1.1V_1^{p,init}]$. Found value: $V_1^p = 1874$.

Step 3. The measured arrival time of the reflected P-wave from the second layer (curve 3, Fig. 3) is 0.277. The start time t_0 of the source's impact on the medium is assumed to be 0.077. Therefore, the doubled time of the P-wave run through the first layer $t_1 = 0.277 - 0.77 = 0.2$. The time t_1 is used to estimate the thickness of the first layer $h_1 = \frac{t1}{2} V_1^p = 187$. Also, based on the slope of the refracted wave, the velocity of the P-wave in the second layer is estimated using the formula 6

$$h_1 = \frac{t_{ref} V_2^p V_1^p}{2((V_2^p)^2 - (V_1^p)^2)^{1/2}},$$
$$V_2^p = \sqrt{\frac{(V_1^p)^2}{1 - \frac{t_{ref}^2}{4h_1^2}(V_1^p)^2}}, \tag{6}$$

where $t_{ref} = 0.156$—the time-coordinate of intersection of line 4 (Fig. 3) with the line $x = 4000$. Therefore, the value of the P-wave velocity on the second layer is estimated as $V_2^p = 2992$. However, attempts to reconstruct V_2^p using simulated annealing were unsuccessful. The algorithm finds $V_2^p \approx .2800$, synthetic seismograms with this value differ from the observed ones. This algorithm's behavior is because attenuation is not considered in this experiment: the amplitude of the refracted wave on the synthetic seismogram significantly exceeds the amplitude of the refracted wave on the observed one. Because of this, the state with a lower cost is where the arrival of the refracted wave coincides with the arrival of the reflection from the third layer (Fig. 5). Therefore, in this experiment, the optimizer is not the simulated annealing algorithm but a researcher who, based on the results of a series of numerical solutions to a direct problem, chooses such a value V_2^p so that the peaks of the amplitudes of reflected waves on synthetic and observed data coincide. Found value: $V_2^p = 3070$.

The measured arrival time of the P-wave reflection from the third layer (curve 5, Fig. 3): 0.411. The doubled travel time of the P-wave through the second layer $t_2 = 0.411 - t1 - t0 = 0.134$.

The values t_1, t_2 and V_2^p are used to restore the thickness of the first and second layers and the velocity of the S-wave in the second layer.

Experiment 3. Reconstruction of the first layer thickness. Medium model: two-layer medium, velocities of the first layer—(V_1^p, V_1^s), thickness of the first layer h_1—unknown parameter, velocities of the second layer—$(V_p^2, \frac{V_2^p}{\sqrt{3}})$. The subdomain of the seismogram is involved in calculating E: $t \leq 0.35$ with a cut Rayleigh wave. Search range: $h_1 \in [0.9\frac{t_1}{2}V_1^p, 1.1\frac{t_1}{2}V_1^p]$.
Found value $h_1 = 199$.

Experiment 4. Reconstruction of the second layer thickness. Medium model: three-layer medium, velocities of the first layer—(V_1^p, V_1^s), thickness of the first layer—h_1, velocities of the second layer —$(V_p^2, \frac{V_2^p}{\sqrt{3}})$, thickness of the second layer h_2—unknown parameter, speed of the third layer—$(1.33V_2^p, \frac{1.33V_2^p}{\sqrt{3}})$. The subdomain of the seismogram is involved in calculating E: $t \leq 0.7$ with a Rayleigh wave cut out. Search range: $h_2 \in [0.9\frac{t_2}{2}V_2^p, 1.1\frac{t_2}{2}V_2^p]$. The found value is $h_2 = 202$.

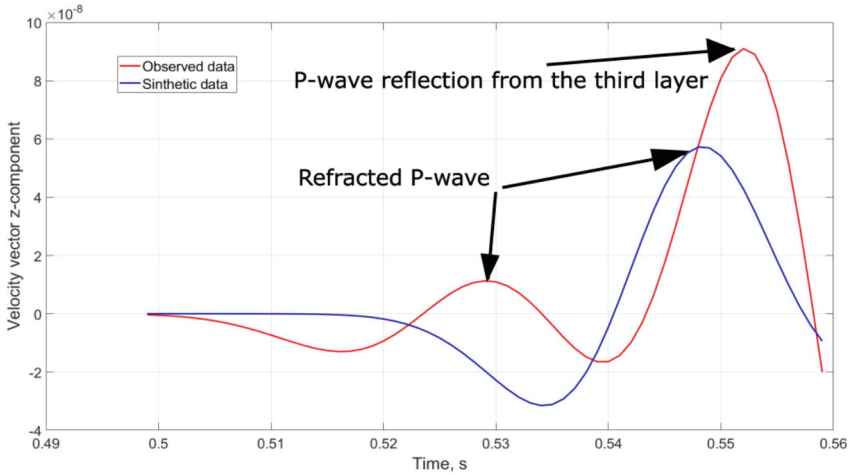


Fig. 5. Trace $x = 3100$ for the source $x = 4000$

Experiment 5. Reconstruction of the the second layer S-wave velocity. Medium model: three-layer medium, velocities of the first layer—(V_1^p, V_1^s), thickness of the first layer—h_1, P-wave velocity on the second layer—V_2^p, S-wave velocity on the second layer V_2^s—the desired parameter, the thickness of the second layer—h_2, the speed of the third layer—$1.33V_2^p$, $\frac{1.33V_2^p}{\sqrt{3}}$). The subdomain of the seismogram is involved in calculating E: the entire seismogram with a cut Rayleigh wave. Search range: $U_1 \in [V_1^s, \frac{V_2^p}{\sqrt{2}}]$. The found value is $V_2^s = 1370$.

Step 4. The measured arrival time of the P-wave reflection from the fourth layer (curve 8, Fig. 3): 0.529. The doubled travel time of the P-wave through the third layer $t_3 = 0.529 - t2 - t1 - t0 = 0.118$. The upper bound of the search for the velocity of the P-wave V^{max} is considered equal to 6000. The P-wave speed in the third layer and the layer thickness are reconstructed in two stages. First, the travel time is specified t_3:

Experiment 6. Refinement of the P-wave travel time in the third layer. Medium model: four-layer medium, speeds of the first layer—(V_1^p, V_1^s), thickness of the first layer—h_1, speeds of the second layer—(V_p^2, V_2^s), thickness of the second layer—h_2, the P-wave velocity in the third layer V_3^p— the desired parameter, the velocity of the S-wave in the third layer—$\frac{1.33V_2^p}{\sqrt{3}}$, thickness of the third layer h_3—unknown parameter, speed of the fourth layer—$(1.1V_{max}, \frac{1.1V_{max}}{\sqrt{3}})$. The subdomain of the seismogram involved in calculating E: $0.5 \le t \le 0.7$ with a Rayleigh wave cut out. Search range: $\frac{h_3}{V_3^p} \in [0.9\frac{t3}{2}, 1.1\frac{t3}{2}]$. The found value is $\frac{h_3}{V_3^p} = 0.0649$.

Then the values of h_3 and V_3^p themselves are found, provided $h_3 = 0.0649V_3^p$

Experiment 7. Reconstruction of the P-wave velocity in the third layer and the thickness of the third layer. Medium model: four-layer medium, speeds of

the first layer—(V_1^p, V_1^s), thickness of the first layer—h_1, speeds of the second layer—(V_p^2, V_2^s), thickness of the second layer—h_2, the velocity of the P-wave in the third layer V_3^p—the desired parameter, the velocity of the S-wave in the third layer—$\frac{1.33V_2^p}{\sqrt{3}}$, thickness of the third layer h_3—unknown parameter, speed of the fourth layer—$(1.1V_{max}, \frac{1.1V_{max}}{\sqrt{3}})$. The subdomain of the seismogram involved in calculating E: $0.5 \leq t \leq 0.7$ with a Rayleigh wave cut out. Search range: $V_3^p \in [V_2^p, V^{max}]$. The values found are $V_3^p = 4707,\ h_3 = 306$.

Step 5. The measured arrival time of the P-wave reflection from the upper boundary of the ellipse (curve 1, Fig. 4): 0.456. The doubled travel time of the P-wave from the boundary of the second layer to the boundary of the ellipse $t_e = 0.456 - t2 - t1 - t0 = 0.045$. Since only the geometric inclusion parameters are restored, the velocities inside the ellipse are set as $V_p^e = 2200, V_s^e = 1200$, corresponding to highly porous sandstone.

It is assumed that the distance from the lower boundary of the ellipse to the fourth layer is at least 5 m, the semi-major axis of the ellipse is parallel to the x axis and does not exceed 1000 m, and the length of the semi-minor axis is at least 10 m.

Experiment 8. Reconstruction of the geometric properties of an elliptical inclusion. Medium model: four-layer medium, speeds of the first layer—(V_1^p, V_1^s), thickness of the first layer—h_1, speeds of the second layer—(V_p^2, V_2^s), thickness of the second layer—h_2, speeds of the third layer—$(V_p^3, \frac{1.33V_2^p}{\sqrt{3}})$, thickness of the third layer—h_3, speeds of the fourth layer—$(1.1V_{max}, \frac{1.1V_{max}}{\sqrt{3}})$, elliptical inclusion velocities—(V_p^e, V_s^e), center of elliptical inclusion —$x = 2600$, coordinate z—unknown parameter, semi-major and semi-minor axes a, b—unknown parameters. The subdomain of the seismogram is involved in calculating E: $t \leq 0.7$ with a Rayleigh wave cut out. Search range: $a \in [\frac{h_3 - \frac{t_4}{2}V_p^3 - 5}{2}, 1000],\ b \in [10, \frac{h_3 - \frac{t_4}{2}V_p^3 - 5}{2}]$. The values found are $a = 98,\ b = 64$, the center of the ellipse—$(2600, 571)$.

In total, the following medium parameters were found (Fig. 6):

- First layer: $V^p = 1874$m/s, $V^s = 603$m/s, $h = 199$m.
- Second layer: $V^p = 3070$m/s, $V^s = 1370$m/s, $h = 202$m.
- Third layer: $V^p = 4707$m/s, $h = 306$m.
- Elliptical inclusion: offset of the ellipse center along the x axis—2600 m, offset of the ellipse center along the z axis—571 m, length of the semi-major axis (parallel to the x axis)—98 m, length of the semi-minor axis—64 m.

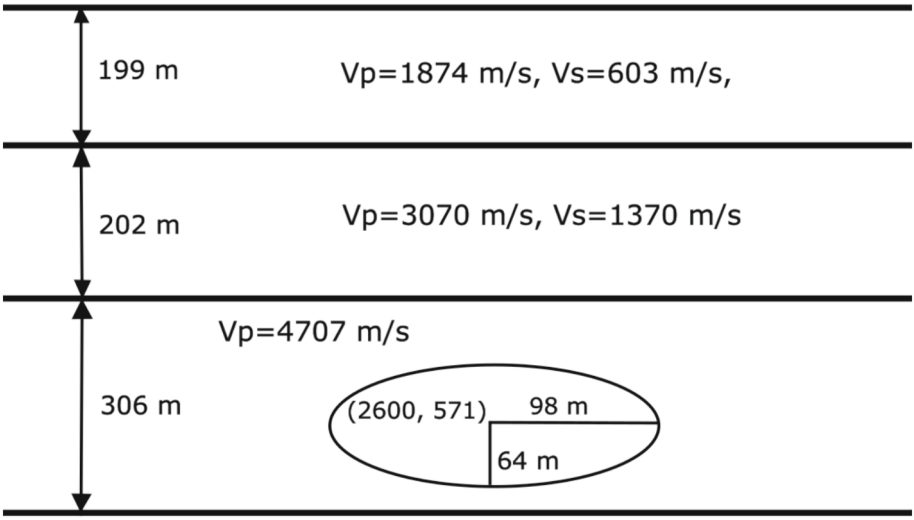


Fig. 6. Reconstructed medium parameters

7 Reconstruction of the Attenuation Parameters of a Viscoelastic Medium Using Deep Learning Methods

Since the simulated annealing algorithm fails to restore the attenuation parameters of the medium, an alternative method for solving the inverse problem based on data is considered. Assuming that there is a data sample that describes the relationship between seismograms and the viscoelastic model of the medium well enough, we can try to reconstruct the operator of the inverse problem F^{-1} using supervised learning on the given data set by minimizing the loss function

$$F_*^{-1} = \arg\min_{F^{-1} \in U} L(m, F^{-1}(d)), \tag{7}$$

where L is a loss function expressing the discrepancy between the model and the action of the operator of the inverse problem on the data, and U is the space of functions used to approximate the operator of the inverse problem.

The approximation of the inverse problem operator F^{-1} is carried out using deep learning methods. A deep convolutional neural network (CNN) is used as an approximator. According to the universal approximation theorem of CN [19], any continuous function of several variables can be approximated by a deep convolutional neural network with any predetermined accuracy. Therefore, if the sample is sufficiently representative, then seismic data mapping into the medium model can be learned with sufficiently high accuracy. After training, the operator $\hat{F}^{-1}$ is used on new seismic data d_{new} to reconstruct the medium model m_{target}.

To translate data from the seismogram space into the model space, a convolutional neural network of the ResNet [3] architecture with a residual block of full preactivation [4] is used. It consists of 18 convolutional blocks that isolate features from input seismograms, residual units that prevent the attenuation of gradients in the learning process, and a fully connected layer that acts as a regression model based on features highlighted by convolutional layers. Each convolutional block consists of a convolutional layer, batch normalization to regularize the learning process, and the activation function ReLU [9] to introduce non-linearity into the approximation of the operator.

Since the thickness of the layers and the velocities of the longitudinal and transverse waves are restored by simulated annealing, deep learning is used only to reconstruct the Q-factors. In order to simplify the learning process, the Q-factors of only the longitudinal waves in the first two layers are reconstructed.

A synthetic training sample of seismograms is prepared by the developed direct problem solver. The velocity and geometric characteristics of the medium are assumed to be equal to the values found by the simulated annealing method. It is assumed that in the first layer, the density of ρ is fixed; its value is— $1500\,\mathrm{kg/m^3}$. In the second and third layers, densities range from $1000\,\mathrm{kg/m^3}$ to $8000\,\mathrm{kg/m^3}$ in $250\,\mathrm{kg/m^3}$ increments. The Q-factor of the longitudinal waves Q_p varies from 10 to 100 in increments of 10 in the first and second layers. The relaxation times for the 5 model are calculated from the Q-factor by the following formulas:

$$\begin{aligned} Q &= \frac{1+\omega^2\tau_\sigma\tau_\epsilon}{\omega(\tau_\epsilon-\tau_\sigma)}, \\ \omega &= 2\pi f. \end{aligned} \tag{8}$$

Q reaches the minima for $\omega = \frac{1}{\sqrt{\tau_\sigma\tau_\epsilon}}$. Then, assuming that $\tau_\sigma \approx \tau_\epsilon$, $\tau_\sigma = \frac{1}{\omega} = 0.0053052$ for $f = 30$ Hz. Therefore,

$$\tau_\epsilon = \tau_\sigma + \frac{2}{Q\omega}. \tag{9}$$

It is assumed that there is no attenuation in the third layer. For each model, seismograms of the vertical u_z and horizontal u_x components of the velocity vector are calculated. Computational domain — rectangle with sides of 2000 m and 450 m, sampling step in space $h_x = h_z = 1$ m, finite difference grid (with PML) contains 2200×500 nodes, sampling step in time $\tau = 0.0001$ s. The point source is at a point with coordinates $x = 1000$ m, $z = 0.5$ m. Source type — expansion center (30 Hz Ricker wavelet). The receivers are on a free surface in 5 m increments (400 receivers). Recording time is 0.45 s with a sampling step of 0.001 s. Thus, 84100 training pairs (d, m) were obtained to restore the Q-factor of longitudinal waves Q_p in the first and second layers according to seismograms u_x, u_z. The calculation time of one seismograms pair is 2 s (NVIDIA RTX 3090 Ti 24 GB), and the calculation time of the entire sample is 47 h.

The data is normalized before training. The seismograms d are trimmed by the values $[-2e^{-7}, 2e^{-7}]$ for u_z, $[-1e^{-7}, 1e^{-7}]$—for u_x. Then, the seismograms

are compressed to size 256×256 and scaled to the segment $[0, 1]$. The m models are also scaled to the $[0, 1]$ segment.

The test sample for evaluating the quality of training consists of 130 seismograms calculated for known models: the density values in the second and third layers in the test sample models have a uniform distribution over the segment $[1000, 8000]$, the Q-values in the first and second layers have a uniform distribution over the segment $[10, 100]$.

The average quadratic error is selected as a loss function for training a neural network.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(m_i^{true} - m_i^{pred} \right)^2, \tag{10}$$

where m_i^{true} is the true model, $m_i^{pred} = F^{-1}(d_i^{true})$ is the predicted model, N is the size of the training dataset.

The search for optimal weights of the neural network is carried out by minimizing the loss function. To minimize the loss function, an optimizer with an adaptive learning rate Adam [5] is selected - an optimization algorithm combining momentum accumulation principles and conservation of the gradient change frequency.

The neural network is trained for 100 epochs with an initial learning rate of 0.01 and a data packet size 32. Every 15 epochs, the learning rate decreases by a factor of 10. The weights obtained at the 34th epoch are chosen as the final weights since they give the smallest error in the test sample—0.7%. Network training over 34 epochs takes 24 h.

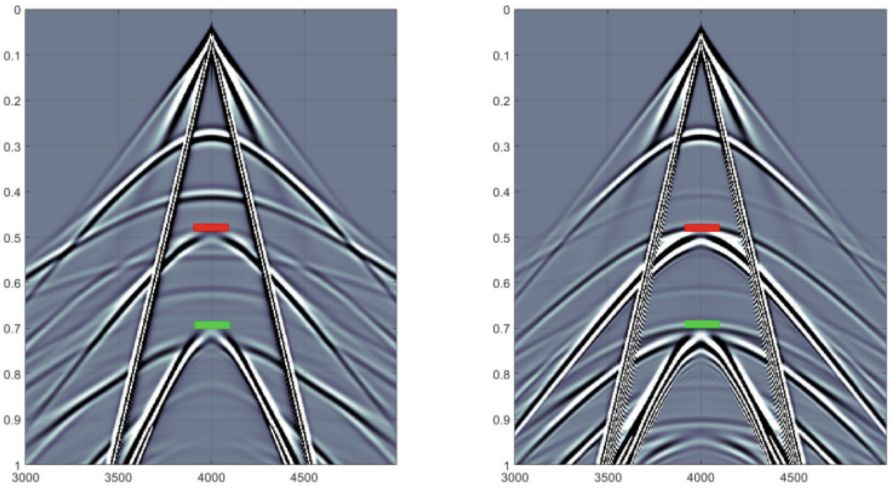


Fig. 7. Seismogram u_z for the source $x = 4000$ (left), seismogram u_z calculated from predicted Q-values (right). Red—the second P-wave reflection from the second layer. Green—the third P-wave reflection from the second layer. (Color figure online)

The trained neural network is applied to the original seismograms for the source $x = 4000$ m, trimmed to $t = 0.45$c. Prediction results: $Q_1^p = Q_2^p = 100$, which can be interpreted as the absence of attenuation in the first and second layers. However, when visually comparing the initial seismograms and seismograms constructed according to the predicted values (Fig. 7), there are differences in attenuation of repeated reflections of the P wave from the boundary of the second layer, which leads to the conclusion that the Q-factor of longitudinal waves in the first layer is significantly lower than predicted.

8 Conclusion

To solve the inverse dynamic seismic problem, an algorithm based on the simulated annealing method has been developed. The algorithm successfully reconstructs the velocity and geometric characteristics of a layered medium with elliptical inclusion in the presence of their approximations. Since numerical experiments have shown that it is impossible to restore the absorbing properties of the medium using the developed algorithm, deep learning methods are used to restore them, namely a convolutional neural network with the ResNet architecture. Although the values predicted by the network seem to differ from the true ones, this approach can give positive results with additional data processing or alternative parameter selection. For example, a Rayleigh wave can be cut from the training sample's seismograms or the seismograms' recording time can be increased so that repeated reflections are observed on the seismograms. However, this will lead to a significant increase in the calculation time of the training sample.

Both proposed approaches are related to the calculation of many direct problems. Therefore, the optimization of the direct problem solver is seen as the most relevant direction for further research. Such optimization requires an excellent scientific base and much experience in parallel computing. A significant advantage here is the use of CUDA technology, which (although this opinion is subjective) is easier to implement and, at the same time, more effective for parallelizing finite difference schemes, compared with parallelization technologies for the CPU (MPI, OMP). Moreover, the memory capacity of modern graphics accelerators makes it possible to use CUDA technology to solve three-dimensional problems without dividing the computing area into several physical nodes, which is a significant advantage in the further solver development.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgments. The research was supported by the Russian Science Foundation grant no. 19–77-20004.

References

1. Bamberger, A., Chavent, G., Hemon, C., Lailly, P.: Inversion of normal incidence seismograms. Geophysics **47**(5), 757–770 (1982). https://doi.org/10.1190/1.1441345
2. Becache, E., Joly, P., Tsogka, C.: A new family of mixed finite elements for the linear elastodynamic. SIAM J. Numer. Anal. **39**, 2109–2132 (2002)
3. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
4. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
5. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. Int. Conf. Learn. Representations (2014)
6. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by simulated annealing. Science **220**(4598), 671–680 (1983). https://doi.org/10.1126/science.220.4598.671
7. Komatitsch, D., Vilotte, J.P.: The spectral element method; an efficient tool to simulate the seismic response of 2D and 3D geological structures. Bull. Seismol. Soc. Am. **88**(2), 368–392 (1998)
8. Levander, A.R.: Fourth-order finite-difference P-SV seismograms. Geophysics **53**(11), 1425–1436 (1988)
9. Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the 30th International Conference on Machine Learning (2013)
10. Mojica, O., Santos, P., Santos, R., Revelo, D.: Building starting models for full-waveform inversion using global optimization methods: a comparison between particle swarm optimization and genetic algorithm (2019). 10.22564/16cisbgf2019.078
11. Pratt, R.G., Shin, C., Hick, G.J.: Gauss-newton and full newton methods in frequency-space seismic waveform inversion. Geophys. J. Int.**133**(2), 341–362 (05 1998). 10.1046/j.1365-246X.1998.00498.x
12. Robertsson, J.O.A., Blanch, J.O., Symes, W.W.: Viscoelastic finite-difference modeling. Geophysics **59**(9), 1444–1456 (1994). https://doi.org/10.1190/1.1443701
13. Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. Geophysics **49**(8), 1259–1266 (1984). https://doi.org/10.1190/1.1441754
14. Tran, K., Hiltunen, D.: Two-dimensional inversion of full waveforms using simulated annealing. J. Geotech. Geoenvironmental Eng. **138**, 1075–1090 (2012). 10.1061/(ASCE)GT.1943-5606.0000685
15. Virieux, J.: P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics **51**(4), 889–901 (1986)
16. Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics **74**(6), WCC1–WCC26 (2009)
17. Vishnevsky, D., Lisitsa, V., Tcheverda, V., Reshetova, G.: Numerical study of the interface errors of finite-difference simulations of seismic waves. Geophysics **79**(4), T219–T232 (2014)
18. Zhang, J.: Quadrangle-grid velocity-stress finite-difference method for elastic-wave-propagation simulation. Geophys. J. Int. **131**(1), 127–134 (1997)
19. Zhou, D.: Universality of deep convolutional neural networks. CoRR **abs/1805.10769** (2018)