

DOI: 10.18721/JCSTCS.111102

UDC 316.452

RECONSTRUCTION OF MEDIUM REFLECTIVITY COEFFICIENTS BASED ON SEISMIC DATA THROUGH MACHINE LEARNING

F.V. Krasnov¹, A.V. Butorin¹, A.V. Mikheyenkov²

¹Gazpromneft STC, St. Petersburg, Russian Federation

²Institute for High Pressure Physics RAS, Moscow, Troitsk, Russian Federation

Geological models of Digital oil fields (DoF) require information about structural properties of subsurface media. 3D models of structural properties of subsurface media are based on data from field seismic survey. Seismic survey is one of the few universal geophysical methods of obtaining information of the Earth subsurface. A reflected signal as a part of seismic data provides information of the properties of a medium through which it has passed. Reflectivity coefficients are determined by fluctuation of the medium's elastic properties and serve as a basis for interpretation of seismic data as well as for prediction of geological structures. We have developed a new method of processing seismic data which allows to locate reflecting planes and compute values of reflectivity coefficients with high degree of precision. To resolve this problem, we have used the Semi-supervised learning method. The machine learning method made it possible to develop a mathematical model, optimize its parameters for synthetic data in order to further use the model for unmarked-up seismic data. The main novelty is in developing a learning algorithm using signal convolution and reflectivity coefficients' regularization. The model we have developed demonstrated high precision for synthetic seismic data with high density of reflecting planes (103 planes per a second of trace). The resulting low level of errors allows significant improving of quantitative understanding of the subsurface structure based on seismic data and is a firm basis for building geological models.

Keywords: seismic data; machine learning; optimization problem; reflection plane position; signal processing; dictionary learning.

Citation: Krasnov F.V., Butorin A.V., Mikheyenkov A.V. Reconstruction of medium reflectivity coefficients based on seismic data through machine learning. St. Petersburg State Polytechnical University Journal. Computer Science. Telecommunications and Control Systems, 2018, Vol. 11, No. 1, Pp. 18–27. DOI: 10.18721/JCSTCS.111102

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕКОНСТРУКЦИИ СВОЙСТВ ОТРАЖАЮЩИХ ПЛОСКОСТЕЙ СРЕДЫ ПО СЕЙСМИЧЕСКИМ ДАННЫМ

Ф.В. Краснов¹, А.В. Буторин¹, А.В. Михеенков²

¹Научно-технический центр «Газпром нефть»,
Санкт-Петербург, Российская Федерация;

²Институт физики высоких давлений имени Л.Ф. Верещагина РАН,
Москва, г. Троицк, Российская Федерация

Геологические модели месторождения создаются с использованием информации о структурных свойствах среды. Получение свойств среды в трехмерном

виде основано на данных об исследовании площади месторождения с помощью сейсморазведки. Коэффициенты отражения среды определяются изменением упругих свойств среды и являются основой для интерпретации сейсмических данных, а также прогноза геологического строения. В статье описан новый метод обработки сейсмических данных, который позволяет определить положение отражающих плоскостей и значения коэффициентов отражения с высокой точностью. Для решения задачи использована методика машинного обучения. Применение методов машинного обучения позволило создать математическую модель, оптимизировать ее параметры на синтетических данных для дальнейшего применения на неразмеченных сейсмических данных. Основным новшеством стала разработка алгоритма обучения, использующего свертку сигнала и регуляризацию коэффициентов отражения. Полученная модель показала высокую точность на синтетических сейсмических данных с высокой плотностью отражающих плоскостей.

Ключевые слова: сейсмические данные; методы машинного обучения; оптимизационная задача; положение отражающих плоскостей; обработка сигналов.

Ссылка при цитировании: Краснов Ф.В., Буторин А.В., Михеенков А.В. Применение методов машинного обучения для реконструкции свойств отражающих плоскостей среды по сейсмическим данным // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. 2018. Т. 11. № 1. С. 18–27. DOI: 10.18721/JCSTCS.11102

Introduction

Modern methods of retrieving decomposition coefficients for signal with known components are based on the Matching Pursuit approach which is proposed in the work [1], such as Batch Orthogonal Matching Pursuit [2], Stabilized Orthogonal Matching Pursuit [3] and Hierarchical Matching Pursuit [4].

Let us take a closer look at Matching Pursuit based algorithms for purposes of seismic surveys data. The OMP algorithm [2] solves the problem of finding decomposition coefficients for a reference signal based on a dictionary (for example with wavelets [5, 6]). Whereby the

OMP algorithm is based on a concept of residual decomposition error.

The result of OMP algorithm's functioning in Fig. 1 leads to achieving a very high precision of signal decomposition into sparse coefficients for a preset dictionary. To demonstrate the advantages and disadvantages of the OMP algorithm for purposes of seismic data, let us make a dictionary D of one 30 Hz Ricker wavelet. The resulting dictionary is shown in Fig. 2.

It should be noted that this dictionary has been made for 250 samples long traces with 2 ms sample rate. Using the OMP algorithm and such a dictionary, we can find same shape

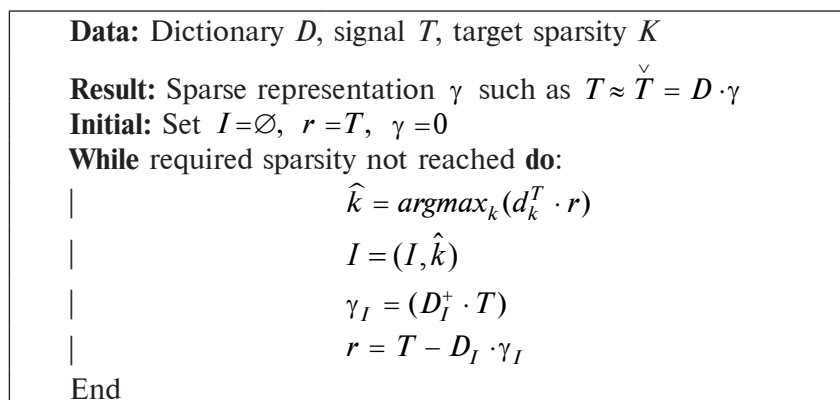


Fig. 1. OMP algorithm based on [2]

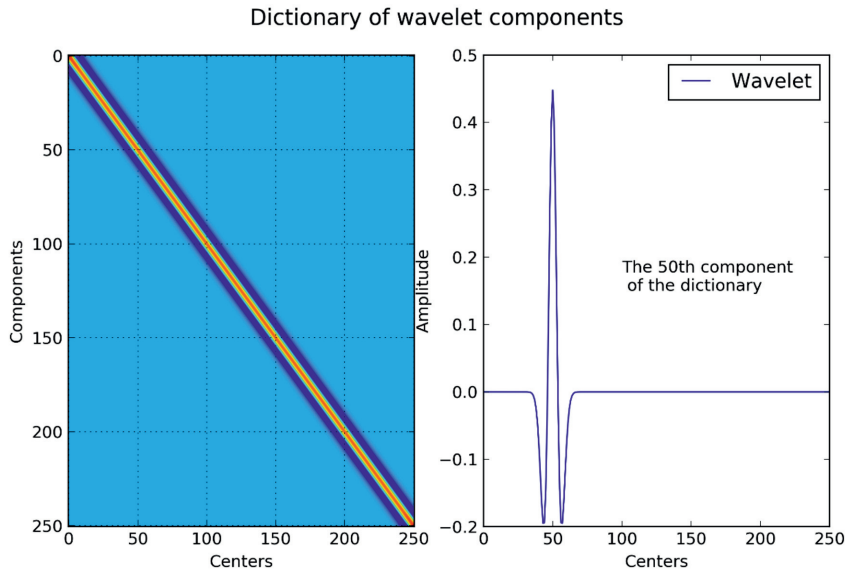


Fig. 2. Dictionary of 30 Hz wavelets.
 One wavelet of the dictionary component No. 50 is shown on the right side

wavelets on each of 250 samples of a trace. The OMP algorithm determines the sparse coefficient for each wavelet of the dictionary for a particular seismic trace (signal).

Let us have a closer look at the particulars of the resulting OMP decomposition based on the dictionary D (Fig. 2) on one seismic trace's example.

Fig. 3 shows one seismic trace resulted from convolution of the medium reflection coefficients and the 30 Hz Ricker wavelet. Further, a condition of nonzero coefficients [7] of decomposition was used for decomposition of the synthetic trace by OMP method and the root-mean-square error (RMSE) was measured. As we can see the decomposition error is quite significant (145.07) for five coefficients (Fig. 3 b), although initially the synthetic trace has been built based on these five reflection coefficients.

In this work [2] it is recommended to use the number of decomposition coefficients equal to 10 % of a signal length. If the trace is 250 samples long it means 25 coefficients. As the number of decomposition coefficients is increased to 25 (Fig. 3 c), the trace reconstruction error becomes more admissible (10.67); however, decomposition coefficients and synthetic reflection coefficients still do not match.

In the result of decomposition into 50 nonzero coefficients (Fig. 3 d) within the OMP algorithm, there are many small coefficients, which decrease the RMSE but do not change the already selected coefficients.

Insufficient precision and emergence of fake reflecting planes make decomposition into coefficients based on the dictionary with the OMP unacceptable for the purposes of retrieving reflectivity coefficients. Thus, we can make a preliminary conclusion of the non-physical nature of decomposition coefficients resulting from using the OMP algorithm.

This study is focused on searching a wavelet composition which would minimize trace reconstruction error with preset position of coefficients.

We have come up with the following research hypothesis:

Hypothesis

There is an algorithm of seismic trace decomposition into decomposition coefficients, which quite precisely match medium reflectivity coefficients in terms of quantity, amplitude and position on a trace.

Further, we have explored the feasibility of building algorithms using machine learning methods, have developed the learning method

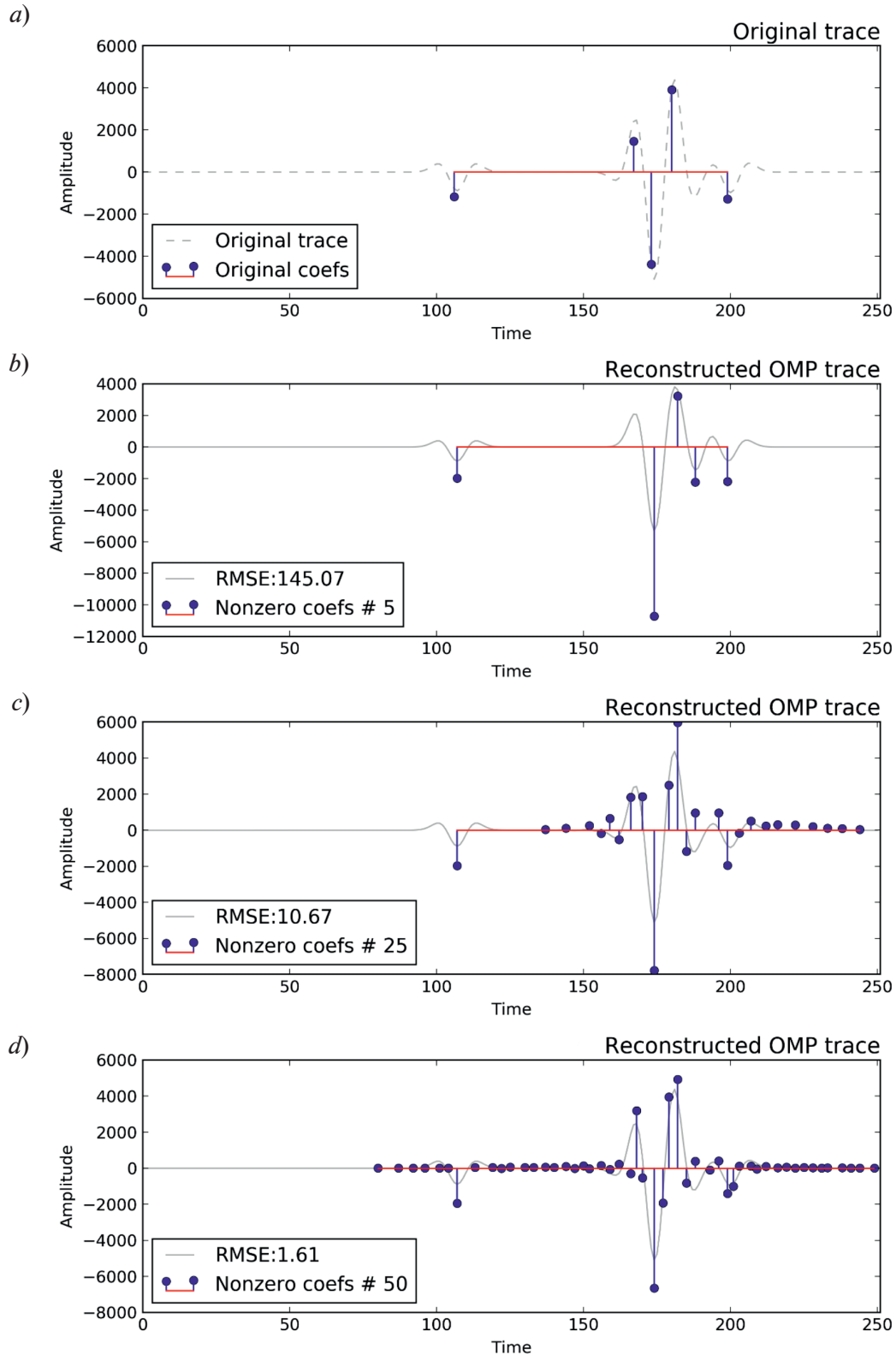


Fig. 3. The figure *a* shows a synthetic trace. Figs *b*, *c*, *d* show the coefficients resulting from the OMP decomposition and the traces reconstructed based on these coefficients

with consideration to a medium's physical properties and have performed several digital experiments on synthetic seismic traces.

This article is composed of introduction, methodological part, experiment outcome and conclusion.

Method

According to [8], the problem we are going to solve in this study falls into the category of inverse coefficient problems. Suppose a process studied through experiment can be modeled by the problem's solution:

$$L_0[u] = g(x, \theta), \quad x \in X \subset \mathbb{R}^K \quad (1)$$

with additional settings

$$l_0[u] = h(x, \theta), \quad x \in \delta X. \quad (2)$$

Here is the set of so-called controllable variables $x = \{x_1, x_2, \dots, x_k\}$, $\theta \in \Omega$ a set of certain parameters, $L_0[\cdot]$ is the determined differential operator depending on θ , \mathbb{R}^K is the Euclidean space with K dimensions, δX is the boundary of a set X .

In practice variables θ are unknown which leads to the following inverse problem: to evaluate initial parameters θ and response $u = (x, \theta)$ function for the equations (1), (2) based on experimental data if the experiment produces some functional $b[u]$ of response u .

In this study only seismic traces and a wavelet are available as experimental data based on which it is necessary to find a medium's reflectivity coefficients with acceptable degree of precision.

To solve this problem we have used the machine learning method. Machine learning approach for geophysics tasks was used in our previous study [8].

A synthetic trace constructed based on the preset synthetic reflection coefficients is used as marked up data. The algorithm learns to select decomposition coefficients matching synthetic coefficients based on which synthetic trace is constructed. An algorithm, named $\mathcal{A}0$, learns to select reflection coefficients based on seismic trace. Labeled data in this case are the synthetic reflection coefficients based on which trace was created.

Formally the inverse problem is defined as follows: there is a discrete synthetic signal (trace) shown as a vector $T \in \mathbb{R}^N$. Trace T

is constructed using the Ricker wavelet W and K reflecting planes through convolution function. Each reflecting plane k is defined by a discrete number $t \in \mathbb{N}^K$ and a certain reflection coefficient $r \in \mathbb{R}^K$. Then the process of constructing a trace can be described by the following equation:

$$T = \sum_i^K W(t_i) * r_i. \quad (3)$$

Based on the trace T data only it is required to develop an algorithm $\mathcal{A}0$ which would determine that reflection coefficients $\tau \in \mathbb{N}^M$ and $\rho \in \mathbb{R}^M$ meet the following criteria:

1. $\check{T} \approx T$, где $\check{T} = \sum_i^M W(\tau_i) * \rho_i$
2. $K \approx M$
3. $\tau \approx t$
4. $r \approx \rho$

Let us have a closer look at criteria 1-4 from the point of view of quantitative evaluation of errors E_{1-4} . Error E_1 is the degree of mismatch of two traces and is computed as Root-mean-square deviation, $\text{RMSE} \left(T, \check{T} \right)$. Deviation in

the number of initial and resulting reflection planes (E_2) is measured as a module of their difference, $\text{abs}(K-M)$. Wiggle of the reflecting planes' positions (E_3) is determined through F1-score metrics. Differences in amplitudes of reflection coefficients (E_4) are computed only for the reflecting planes fitting the positions $\text{RMSE} \left(T[t = \tau], \check{T}[t = \tau] \right)$.

Errors E_{2-4} are quantitative evaluations of the algorithm's functioning but cannot be applied to optimization ones. In order to find an optimal solution it is necessary to vary the reflection coefficients to minimize the E_1 error.

In addition to the condition of the E_1 error minimization, we also include the following physical criteria in the optimization process:

- Reflecting planes should not be too close to each other (U_1);
- Number of reflecting planes should be minimal (U_2).

Physical criteria U_1 and U_2 are quantified as penalty functions $F_1(U_1)$, $F_2(U_2)$ in the appraisal of optimization progress. Thus, a meta-algorithm of optimization can be described as

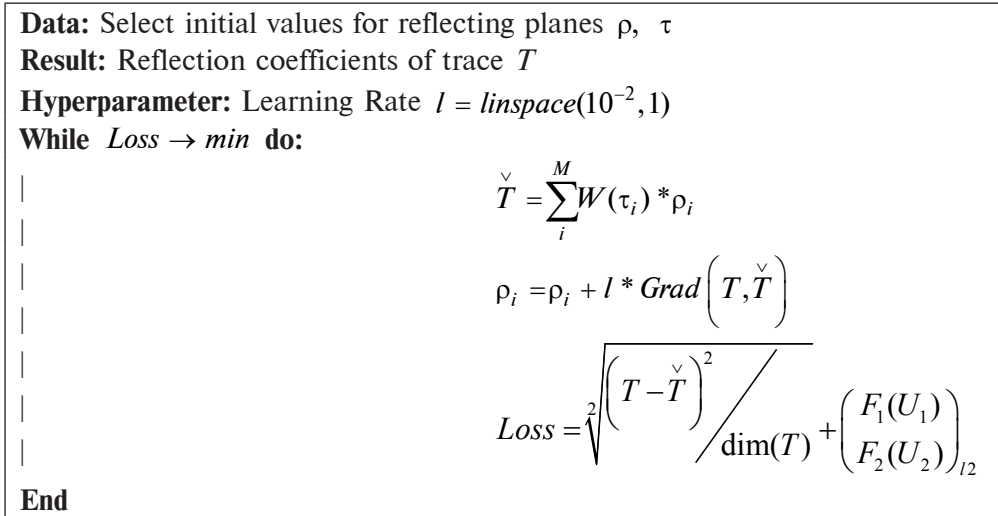


Fig. 4. Meta-algorithm of optimization \mathcal{A}_0

follows (Fig. 4).

The algorithm \mathcal{A}_0 uses the information of traces mistie to determine direction and scope of reflecting planes' modulation. For this reason \mathcal{A}_0 can fall into the category of variable optimization algorithms aimed at searching for a global minimum. The heuristic approaches to the optimization problem based on the work [6] and tested by us have proven less efficient.

Experiment outcome

To check the above-mentioned method we have performed experiments with synthetic

traces of various density of reflecting planes. Based on the outcome of the experiments optimal values of learning rate, gravity of penalties from U_1 and U_2 have been determined.

Initialization of a vector of reflecting planes has been studied separately. Three types of initialization have been tested:

- Initialization by random numbers in normal distribution;
- Initialization by trace amplitude values with a scale factor;
- Initialization by trace's extremes amplitudes.

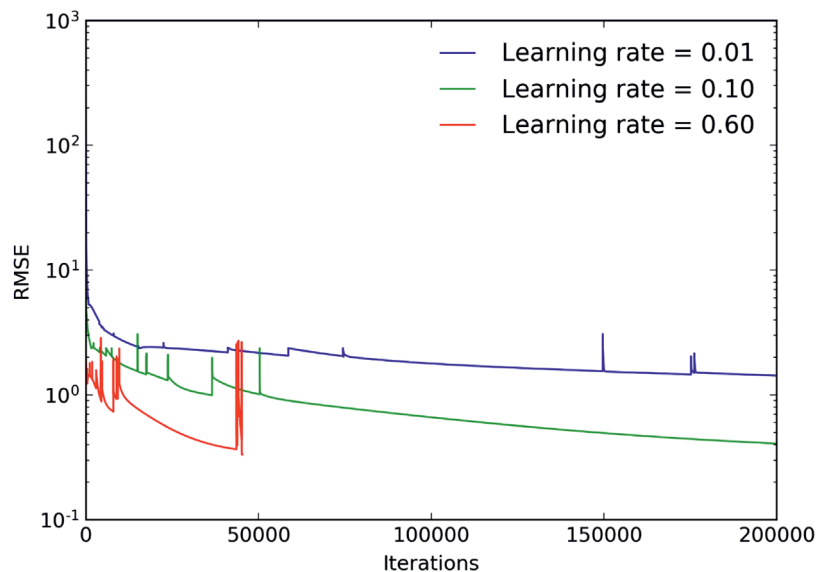


Fig. 5. Loss functions depending on various learning rates

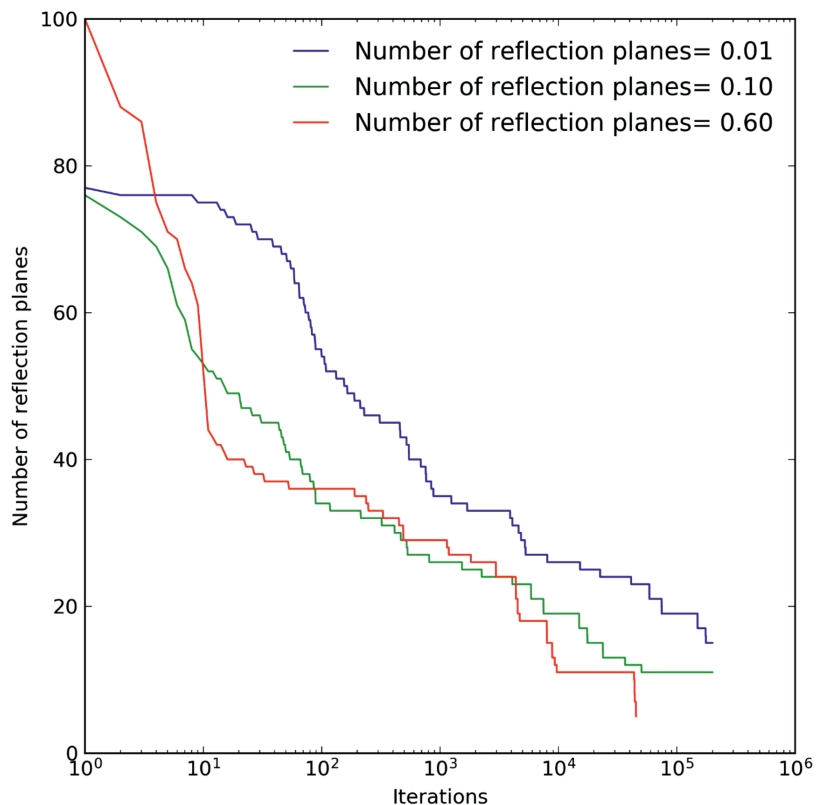


Fig. 6. Number of reflecting planes depending on the number of iterations

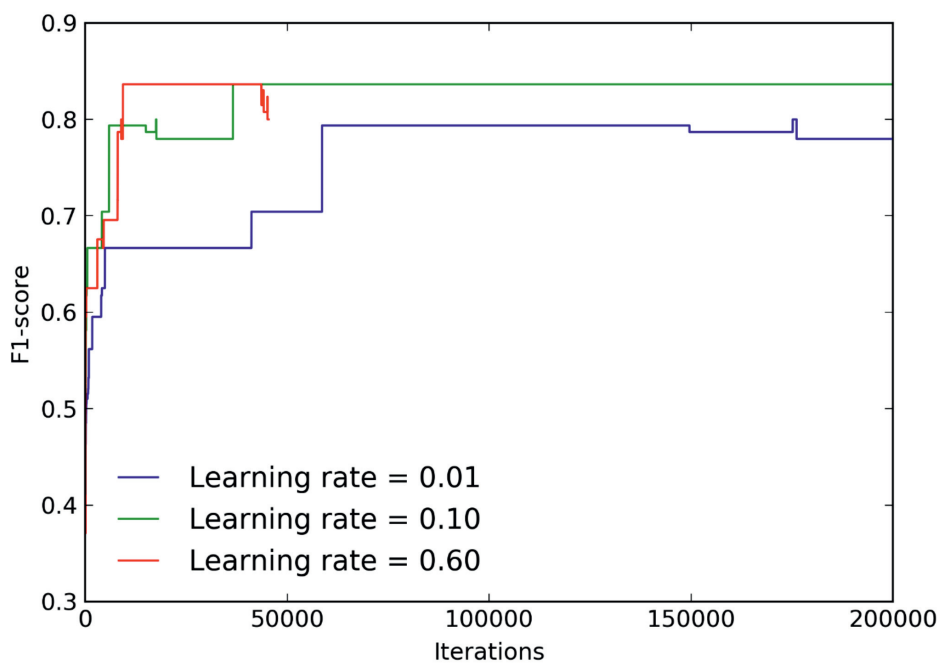


Fig. 7. Dependence of the F1-score metrics on various rates of the A_0 algorithm's learning



The fastest convergence was achieved in the initialization by trace's extremes amplitudes. Fig. 5 shows the loss function's dependence on various values of the learning rate parameter.

The dependences in Fig. 5 show a standard pattern: increasing learning rate leads to faster convergence. Applied regularization by the number of reflecting planes (U_2) is shown in Fig. 5 as excursions in transition to fewer reflecting planes. This effect is shown in more detail in Fig. 6.

It can be seen from Fig. 6 that the number of reflecting planes stops changing at a certain number of iterations. Achievement of a minimum constant is one of the signs that the algorithm should be stopped, along with the decreasing RMSE.

Wiggle of reflecting planes' positions (E_2) is determined through the F1-score metrics. Fig. 7 shows dependence of the F1-score metrics on various learning rates of the algorithm.

Dependence of the F1-score metrics on various rates of the $\mathcal{A}0$ algorithm's learning rates. We have separately examined the dependences for errors E_1 (Fig. 5), E_2 (Fig. 6), E_3 (Fig. 7). Table 1 shows comparison of the E_{1-4} errors for the OMP algorithm and $\mathcal{A}0$.

It is significant that using the algorithm $\mathcal{A}0$ allows making significantly fewer errors for traces with five reflecting planes. Table 2 shows comparison of errors for the OMP and $\mathcal{A}0$ for the trace with 103 reflection planes (500 samples with 2 ms sampling rate).

As we can see from Table 2, the algorithm $\mathcal{A}0$ allows minimizing errors of decomposition below the OMP level. However, the algorithm $\mathcal{A}0$ maintains physical significance of decomposition coefficients.

Conclusion

We have developed an algorithm that allows incorporating physical laws into machine learning methods. To compare efficiency of the proposed algorithm the authors have developed a complex precision metrics containing four components:

- E_1 is the degree of two traces' mistie computed as a normalized sum of squared residuals, $\text{RMSE}\left(T, \check{T}\right)$.
- Difference in numbers of initial and resulting reflecting planes as a result of the algorithm's application (E_2).
- Wiggle of reflecting planes' positions (E_3)

Table 1

Comparison of the E_{1-4} errors for the OMP algorithm and $\mathcal{A}0$ for the trace with five reflecting planes

Algorithm / Error	OMP, 5 coefficients	OMP, 25 coefficients	OMP, 50 coefficients	$\mathcal{A}0$
E_1	145.07	10.67	1.61	0.59
E_2	0	20	45	0
E_3	0.2	0.06	0.04	0.38
E_4	829.43	788.61	752.37	10.03

Table 2

Comparison of the E_{1-4} errors for the OMP algorithm and $\mathcal{A}0$ for the trace with 103 reflecting planes

Algorithm / Error	OMP, 50 coefficients	OMP, 103 coefficients	OMP, 150 coefficients	$\mathcal{A}0$
E_1	39.84	9.40	0.92	0.87
E_2	53	0	47	2
E_3	0.23	0.25	0.24	0.33
E_4	1644.11	1562.16	1517.16	100.23

is determined through the F1-score metrics.

- Differences in reflection coefficients amplitudes (E_4).

The performed experiments have shown that the trained algorithm allows making lower value errors than the OMP [2] and maintaining

physical significance of resulting decomposition coefficients.

It is worthwhile to continue further studies in this direction on real data instead of synthetic ones, for a particular deposit with sufficient number of investigated wells.

REFERENCES

1. **Mallat S.G., Zhang Z.** Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, Vol. 41, No. 12, Pp. 3397–3415.
2. **Rubinstein R., Zibulevsky M., Elad M.** *Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit Technical Report*. CS Technion, April 2008.
3. **Saadat S.A., Safari A., Needell D.** Sparse reconstruction of regional gravity signal based on Stabilized Orthogonal Matching Pursuit (SOMP). *Pure and Applied Geophysics*, 2016, Vol. 173, No. 6, Pp. 2087–2099.
4. **Bo L., Ren X., Fox D.** Unsupervised feature learning for RGB-D based object recognition. *Experimental Robotics*. Springer, Heidelberg, 2013, Pp. 387–402.
5. **Butorin A.V., et al.** Spectral inversion methods and its application for wave field analysis. *SPE Russian Petroleum Technology Conference*. Society of Petroleum Engineers, 2017. (rus)
6. **Butorin A.V., Krasnov F.V., et al.** Spectral inversion methods and its application for wave field analysis. *SPE Russian Petroleum Technology Conference*. Society of Petroleum Engineers, 2017. (rus)
7. **Sanyi Yuan, Shangxu Wang, Ming Ma, Yongzhen Ji, Li Deng.** Sparse bayesian learning-based time-variant deconvolution. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, Vol. 55(11), Pp. 6182–6194.
8. **Mirzadjanadze A.H., Khasanov M.M., Bakhtizin R.N.** *Modelling of oil and gas extraction*. Institute Computer Science, 2005.
9. **Krasnov F., Glavnov N., Sitnikov A.** A machine learning approach to enhanced oil recovery prediction. *In International Conference on Analysis of Images, Social Networks and Texts*, Springer, 2017, Pp. 164–171.
10. **Magnus Erik Hvass Pedersen.** *Tuning & simplifying heuristical optimization*. PhD thesis, University of Southampton, 2010.

Received 11.11.2017.

СПИСОК ЛИТЕРАТУРЫ

1. **Mallat S.G., Zhang Z.** Matching pursuits with time-frequency dictionaries // *IEEE Transactions on signal processing*. 1993. Vol. 41. No. 12. Pp. 3397–3415.
2. **Rubinstein R., Zibulevsky M., Elad M.** Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit Technical Report. CS Technion, Apr. 2008.
3. **Saadat S.A., Safari A., Needell D.** Sparse reconstruction of regional gravity signal based on Stabilized Orthogonal Matching Pursuit // *Pure and Applied Geophysics*. 2016. Vol. 173. No. 6. Pp. 2087–2099.
4. **Bo L., Ren X., Fox D.** Unsupervised feature learning for RGB-D based object recognition // *Experimental Robotics*. Springer, Heidelberg, 2013. Pp. 387–402.
5. **Butorin A.V., et al.** Spectral Inversion Methods and its Application for Wave Field Analysis // *SPE Russian Petroleum Technology Conf*. Society of Petroleum Engineers, 2017.
6. **Butorin A.V., Krasnov F.V., et al.** Spectral inversion methods and its application for wave field analysis // *SPE Russian Petroleum Technology Conf*. Society of Petroleum Engineers, 2017.
7. **Sanyi Yuan, Shangxu Wang, Ming Ma, Yongzhen Ji, Li Deng.** Sparse bayesian learning-based time-variant deconvolution // *IEEE Transactions on Geoscience and Remote Sensing*, 2017. Vol. 55(11). Pp. 6182–6194.
8. **Mirzadjanadze A.H., Khasanov M.M., Bakhtizin R.N.** *Modelling of oil and gas extraction*. Institute Computer Science, 2005.
9. **Krasnov F., Glavnov N., Sitnikov A.** A machine learning approach to enhanced oil recovery prediction // *Internat. Conf. on Analysis of Images, Social Networks and Texts*. Springer, 2017. Pp. 164–171.
10. **Magnus Erik Hvass Pedersen.** *Tuning & simplifying heuristical optimization*. PhD thesis. University of Southampton, 2010.

Статья поступила в редакцию 11.11.2017.

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

KRASNOV Fedor V.

КРАСНОВ Федор Владимирович

E-mail: krasnov.fv@gazprom-neft.ru

BUTORIN Alexander V.

БУТОРИН Александр Васильевич

E-mail: Butorin.av@gazpromneft-ntc.ru

MIKHEYENKOV Andrey V.

МИХЕЕНКОВ Андрей Витальевич

E-mail: mikheen@bk.ru