Processing of big data in the detection of geochemical anomalies of rare-earth metal deposits

Laura Temirbekova

Processing of Big Data in the Detection of Geochemical Anomalies of Rare-Earth Metal Deposits

Laura Temirbekova

Abai Kazakh National Pedagogical University, Almaty, Kazakhstan

laura-nurlan@mail.ru

Abstract. The paper presents a numerical method for processing a large amount of data in detection of anomalies of the chemical elements distribution in polymetallic deposits. Mathematically, the problem reduces to solving the Fredholm integral equation of the first kind for large number of different right-hand parts, while the kernel remains unchanged. The algorithm consists of two stages: at the first stage a number of problems are solved, for which knowing the right part is not required, in the second stage, the previously obtained data are used to find solutions of the integral equation for different right-hand sides. This approach allows the data processing "in-situ"and to define prospects of the established abnormal areas, allocate the overriding areas for geological exploration. The algorithm is tested on model and real data.

INTRODUCTION

In the paper an algorithm for big data processing in the detection of anomalies of chemical elements' distribution on rare-earth metal deposits [1-5] is presented. Mathematically the problem is to solve the Fredholm integral equation of the first kind for a large the number of different right-hand sides as the kernel of the integral equation remains unchanged. Proposed in the paper algorithm for solving the problem is divided into two stages. At the first stage, a number of tasks, in which the right-hand side of the integral equation is not involved, are solved. Using the results of calculations of the first stage, at the second stage the desired solution of the integral equation for a particular right-hand side is calculated by means of two summations. Such an algorithm allows the first stage of calculations to be hold in advance, before getting "in-situ", the simplicity of the second stage allows data processing "in-situ" in real time mode.

Currently during the initial assessment of the distribution of chemical elements in order to determine the prospects of anomalous areas, distinguish the overriding areas for geological exploration approximate evaluation methods are used [5]. These formulas are simple, they allow you to process data and to conduct an "in-situ" assessment but the result can differ quite from the actual distribution of the useful element concentration. To refine the results, drilling in prospective areas is carried out. As it is known, drilling is sufficiently expensive procedure therefore the optimal determination of points for drilling is the vital task. To increase the accuracy of the forecast for surface data, one could apply mathematical methods based on the solution of the Fredholm integral equation of the first kind [6], however, as the solution of this equation is an ill-posed problem and may require a significant time to obtain an acceptable solution, so the more time is required to process data of the same type collected from a large area. Therefore, in order to preserve convenience of the application of the existing forecasting method (simplicity of the surface data processing formulas) and to increase its accuracy, the approach described in this paper was proposed.

An extensive literature on the solution of integral equations, including the Fredholm equation of the first kind, exists there. To get acquainted with the basic methods of solving integral equations of various types the recommended directories are [7,8] and an electronic resource [9].

International Conference on Analysis and Applied Mathematics (ICAAM 2018)
AIP Conf. Proc. 1997, 020072-1–020072-7; https://doi.org/10.1063/1.5049066
Published by AIP Publishing. 978-0-7354-1713-7/$30.00

STATEMENT OF THE PROBLEM AND ITS TRANSFORMATION

Let us consider the Fredholm integral equation of the first kind:

$$\int_a^b K(x,s)u(s)ds = f(x), \quad x \in [a,b]. \tag{1}$$

It is necessary to solve the same integral equation (1) for different right-hand sides.

Let us propose a numerical procedure for solving the integral equation (1), which will allow us to break up the numerical solution of the problem in two stages: the first does not depend on $f(x)$, the second depends on it.

The solution of the integral equation (1) is connected with the problem of revealing anomalies in the exploration of spatial distribution of chemical elements in rare-earth metal deposits [6]. When carrying out work ”field” it is necessary to process a large number of real data, which in equation (1) play the role of the right-hand sides $f(x)$, and the kernel $K(x,s)$ does not vary.

To transform the formulation of problem (1), let us use the method of the adjoint operator proposed in [10], and proved itself to be useful in solving the Cauchy problem for elliptic equations on real data [11-14].

Consider the conjugate equation:

$$\int_a^b K(s,x)v_k(s)ds = \alpha_k(x), \quad x \in [a,b]. \tag{2}$$

Here $\alpha_k(x)$ are known functions that will be defined later, k is the number, $v_k(x)$ is the solution of the conjugate equation (2), that depends on the number k.

Multiplying (1) by $v_k(x)$ and integrating the obtained result with respect to x, we have:

$$\int_a^b v_k(x) \int_a^b K(x,s)u(s)\,ds\,dx = \int_a^b v_k(x)f(x)\,dx.$$

Changing the order of integration and taking into account (2), we obtain

$$\int_a^b u(s)\alpha_k(s)\,ds = \int_a^b v_k(x)f(x)\,dx. \tag{3}$$

Supposing that $\alpha_k(x)$ are the basis functions on the interval $[a,b]$, then the first integral in (3) is the k-th coefficient in the decomposition of the function $u(x)$ in a Fourier series, i.e. from (3) it follows that

$$u_k = \int_a^b v_k(x)f(x)\,dx, \quad u(x) = \sum_{k=1}^{N} u_k\alpha_k(x). \tag{4}$$

Thus, the solution of problem (1) is divided into two stages:

1. solving equations (2);
2. calculation of the Fourier coefficients and summation (4).

It is easy to see that step 1 is complex, but it can be performed in advance, before obtaining the data $f(x)$; for the solution of problems (2), any chosen method that would solve these problems with satisfactory accuracy can be applied. Stage 2 is a technological one, as it represents only two summations; such actions can be performed ”in-situ” in real time mode.

Thus, stage 1 is preparatory, at this stage functions $v_k(x)$ are calculated and stored. Stage 2 can be performed in real time mode at the $f(x)$ data acquisition site. At this stage, when processing the newly obtained functions $f(x)$ we use the known functions $v_k(x)$.

Problem Parameters

When processing large areas, the area is divided into standard sections, on which the data $f(x)$ is measured [1]. In this case, since the experimental data are presented in the Gauss-Krueger coordinates, we have $a = 1.45 \cdot 10^7$ and $b = 1.472612 \cdot 10^7$ m. It is convenient to carry out the nondimensionalization, then $a = 1.45$, $b = 1.472612$.

The parameter N (the number of basis functions) can be selected from the following conditions. Since the accuracy with which the data $f(x)$ is measured, is known in practice, so the parameter N can be chosen from the condition of the following inequality being fulfilled:

$$\left\| f(x) - \sum_{k=1}^{N} f_k \alpha_k(x) \right\| \leq \delta, \tag{5}$$

where f_k the decomposition coefficients of the function $f(x)$ into a Fourier series, and $\alpha_k(x)$ are the basis functions on the interval $[a, b]$. Since all the functions $f(x)$ are measured with approximately the same accuracy, it is sufficient to take several functions $f(x)$, decompose it into a Fourier series, and choose the optimal N, that satisfies inequality (5). When expanding the function $f(x)$ into a Fourier series, it is necessary to remember with what interval the values of the function $f(x)$ on the profile are recorded, this determines the maximum frequency (Nyquist frequency) when the given function is ecomposed in a series (see, for example [15]). Further, to process the data the already known value of the parameter N can be used.

The kernel has the form [6]:

$$K(x, s) = \frac{h}{\pi(x-s)^2 + h^2},$$

where h is the level located below the surface of the earth.

From the particularity of the formulation and solution of the geochemical problem of the anomalies detection in the space dimensional distribution of chemical elements on rare-earth metal deposits it is known [1,6]

$$u(a) = u(b) = 0. \tag{6}$$

Equations (6) allow us to choose basis functions on the interval $[a, b]$ in the following form:

$$\alpha_k(x) = \sin \frac{\pi k(2x - (b+a))}{b-a}. \tag{7}$$

Testing the Proposed Approach

Let us consider the integral equation (1), in which

$$\begin{aligned} K(x, s) &= \frac{1}{x^2 + s^2}, \\ u(x) &= \frac{2x - (b+a)}{b-a}, \\ f(x) &= \frac{1}{b-a} \ln \frac{x^2 + b^2}{x^2 + a^2} - \frac{b+a}{b-a} \frac{1}{x} \left(\arctan \frac{b}{x} - \arctan \frac{a}{x} \right), \\ \alpha_k(x) &= \sin \frac{\pi k(2x - (b+a))}{b-a}, \end{aligned} \tag{8}$$

and $a = -1$, $b = 1$. The example is taken from [16]. It is clear that the function $u(x)$ does not satisfy condition (6); nevertheless, it can be decomposed in a series by the functions $\alpha_k(x)$ (see [16, p. 52, No. 41]).

Numerically integral equations (2) were solved by the method proposed and implemented in [17]. Let us briefly describe it here.

Equation (2) is subject to the regularization of Lavrentiev M.M. [18]. Integral equations of the following form are considered

$$\mu v_k(x) + \int_a^b K(s, x) v_k(s) ds = \alpha_k(x), \quad k = \overline{1, N}, \tag{9}$$

where μ is Lavrentiev's small regularization parameter.

The discretization is carried out with the uniform pitch $h = (b-a)/M$ $x_i = a+(i-0.5)h, s_j = a+(j-0.5)h$ $(i, j = \overline{1, M})$, M is the number of points of the interval $[a, b]$ partition. In this case, it is necessary to solve a system of linear algebraic equations

$$Aw_k = \beta_k, \quad k = 1, 2, ..., N, \tag{10}$$

where

$$A = \mu E + h \cdot \begin{bmatrix} K(s_1, x_1) & K(s_1, x_2) & \cdots & K(s_1, x_M) \\ K(s_2, x_1) & K(s_2, x_2) & \cdots & K(s_2, x_M) \\ \cdots & \cdots & \cdots & \cdots \\ K(s_M, x_1) & K(s_M, x_2) & \cdots & K(s_M, x_M) \end{bmatrix}, \quad w_k = \begin{bmatrix} v_k(x_1) \\ v_k(x_2) \\ \vdots \\ v_k(x_M) \end{bmatrix}, \quad \beta_k = \begin{bmatrix} \alpha_k(x_1) \\ \alpha_k(x_2) \\ \vdots \\ \alpha_k(x_M) \end{bmatrix}.$$

In papers [17, 19], the square root method and the iterative method were used for numerical solution equations (10)

$$B\frac{w_k^{n+1} - w_k^n}{\tau} + Aw_k^n = \beta_k,$$

where B is an easily inversible matrix.

The results of numerical calculations for data (8) using the approach proposed above are shown in Fig. 1.

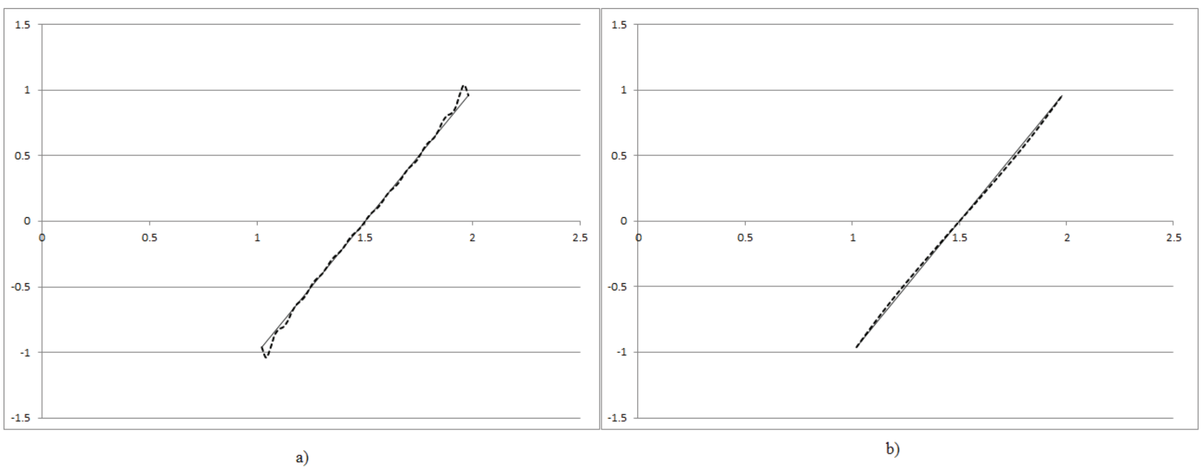

FIGURE 1. The exact (continuous line) and the approximate solution (dotted line) u for a) $M = 60$, b) $M = 1000$

When testing the proposed approach $\mu = 0.04$, the maximum M was 1000, the parameter $N = 50$, the value of w_k was calculated with the accuracy of 10^{-4}, which required 90 to 100 thousand iterations. The absolute error between the exact $u(x)$ and the calculated (see formula (4)) solutions did not exceed 0.03, the relative error was – 3%. Calculation time on the PC ASUS X541U, Windows 10, Intel Core I5-7200U CPU 2.50 GHz reaches 44 minutes.

Real Data Processing

The IRGETAS laboratory (East Kazakhstan State Technical University named after D. Serikbaev, Ust-Kamenogorsk, Kazakhstan) provided geochemical mapping data on non-coherent sediments within the Kalba-Narymskaya structural-metallogenic zones at a scale of 1 : 500000 with the selection of lithic samples on secondary blue caps. In the course of field exploration, samples of soil have been collected from the caves at a depth of up to 20 centimeters along the referential network of 5×5 kilometers. The sampling was carried out in accordance with State Standard 28168. The fixing of points was carried out using the GPS navigator GARMIN. As a result, 777 soil samples were taken from a territory of more than 40 thousands square kilometers. Selected geochemical samples were analyzed by ICP-MS spectroscopy for 70 elements.

According to the algorithm described in the paper, the following functions were calculated for each depth $v_k(x_i)$ ($k = \overline{1,N}$, $k = \overline{1,M}$), $N = 51$, $M = 1000$. The amount of data required to carry out the work "in-situ" makes up $0,4$ MB. At present, the Kalba-Narymskaya zone has been well studied [2, 4]. The data confirmed by drilling was used to verify the results of numerical calculations. The results of comparison on the distribution of the anomaly of the Li element on the surface and at a depth of $h = 300$ m is given below (see Fig. 2).

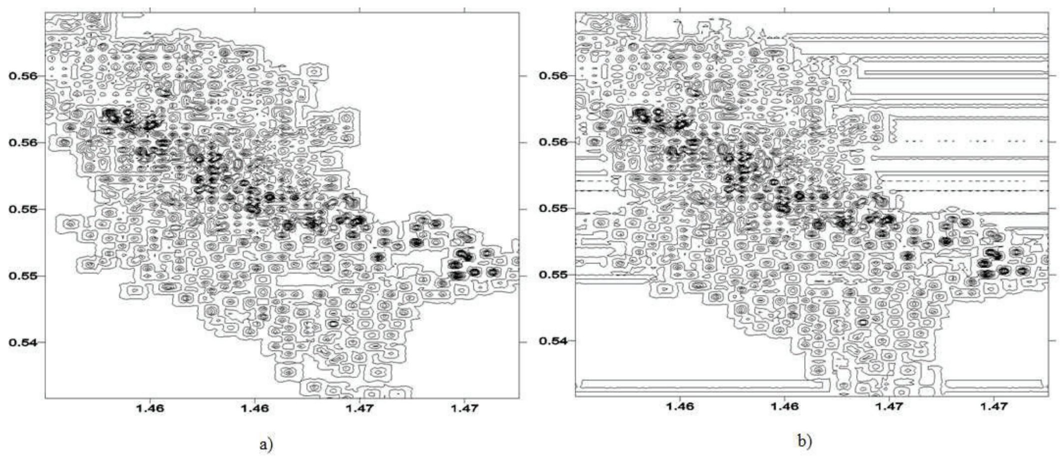

FIGURE 2. An example of the distribution of *Li* anomalies on the surface and at a depth $h = 300$ m, obtained by the proposed method

The concentration of *Li* at the surface and at depth is given in Tab. 1.

TABLE 1. Concentration of *Li* (gg per kkk)

	Sample name	Surface concentration	Concentration at a depth of 300 m	The result of numerical calculations for a depth of 300 m
1	G-71	105.8	107.7	110.2
2	G-115	132.7	128.4	130.3
3	G-169	126.6	127.8	130.4
4	G-181	133	131.5	130.8
5	G-185	141	142.3	140.7
6	G-498	104.1	107.6	110.3
7	Q-498	156.975	158.5	160.1
8	G-173	122.2	119.3	120.4

The results of comparisons of the concentration values for other elements (*Nb*, *Be*, *Sn* etc.) of other samples from different depths (100–500 m) showed a similar result: the discrepancy between the calculated data and the data from the wells is not exceeding 2–4 %.

CONCLUSIONS

Modern approved geochemical exploration methods only allow to sufficiently reliable estimate geochemical anomalies. Given results represents a qualitative and a quantitative characterization of the geochemical parameters of the detected anomaly, on the basis of which its nature (ore, barren, technogenic), productivity, type of predicted mineralization level, anomaly opening degree (erosion cut degree), industrial significance. The final result of the geochemical anomaly estimation is the calculation of the predicted (prospective) resources of minerals up to the technologically expedient depth. The task of interpretation of the geochemical anomaly distribution to depth can be solved only in case of intersections presence with anomalous contents of chemical elements at various hypsometric levels, which requires expensive drilling operations. The more of such intersections, the higher will be the reliability in the interpretation of geochemical anomalies to depth. The proposed method allows interpretation of the spread of the geochemical anomaly to the depth only according to the sampling data from the surface and ”in-situ”. This significantly reduces both the time and financial costs associated with conducting geological exploration work.

REFERENCES

[1] O. D. Gavrilenko, *Methodology for assessing the prospects of geochemical anomalies* (Vestnik of EKSTU named after D. Serikbaev and Computing technologies, joint release, Ust-Kamenogorsk, Kazakhstan, Part 3, 2013), 79-85 pp. http://rmebrk.kz/journals/1879/59164.pdf (in Russian).

[2] O. N. Kuzmina, B. A. Dyachkov, A. G. Vladimirov, M. V. Kirillov, and Yu. O. Redin, Geology and Geophysics **54(12)**, 1889-1904 (2013), in Russian.

[3] B. A. Dyachkov, Z. I. Chernenko, I. E. Mataibaeva, and O. V. Frolova, Vestnik of KazNTU, Earth Sciences **4**, 101-109 (2013), in Russian.

[4] Report on the results of the geological appraisal of 1:200000 scale sheets M-44-XXII, XXIII (interfluve of Shar and Irtysh rivers) by the works of 2006-2008, report on research (final): MEMR of The Republic of Kazakhstan Committee of Geology and Subsoil Use East Kazakhstan Interregional Department of Geology and Subsoil Use Public institution ”Vostkaznedra” ”TOPAZ” LLP EC; Head is Bagadayev B.A.; executive is Klepikov O.N. [and others]. - Ust-Kamenogorsk, 2008, 325 p. (No. GR3VK-06-013/RGSR-25/10) (in Russian).

[5] B. A. Dyachkov, O. D. Gavrilenko, and A. N. Bubnyak, Geology and Susoil protection **64(3)**, 31-37 (2017), in Russian.

[6] A. N. Tikhonov, *Mathematical Geophysics*, Moscow: IPE RAS, 1999, in Russian.

[7] A. V. Manzhirov and A. D. Polyanin, *Integral Equations. Methods of Solution*, Reference Book, Moscow: Faktorial Press, 2000, in Russian.

[8] A. F. Verlaine and V. S. Sizikov, *Integral Equations: Methods, Algorithms, Programs*, Reference Book, Kiev, Naukova Dumka, 1986, in Russian.

[9] EqWorld, http://eqworld.ipmnet.ru

[10] A. L. Karchevsky, Eurasian Journal of Mathematical and Computer Applications **1(2)**, 4-20 (2013).

[11] A. L. Karchevsky, I. V. Marchuk, and O. A. Kabov, Applied Mathematical Modelling **40(2)**, 1029-1037 (2016).

[12] I. Marchuk, A. Karchevsky, A. Surtaev, and O. Kabov, International Journal of Aerospace Engineering **2015**, Article ID 391036, 1-5 (2015).

[13] V. V. Cheverda, I. V. Marchuk, A. L. Karchevsky, E. V. Orlik, and O. A. Kabov, Thermophysics and Aeromechanics **23(3)**, 415-420 (2016).

[14] V. V. Cheverda, A. L. Karchevsky, I. V. Marchuk, and O. A. Kabov, Thermophysics and Aeromechanics **24(5)**, 803-806 (2017).

[15] M. A. Basarab, E. G. Zelkin, V. F. Kravchenko, and V. P. Yakovlev, *Cifrovaya obrabotka signalov na osnove teoremy Uitte-kera-Kotel’nikova-Shennona [Digital signal processing based on the Whittaker-Kotelnikov-Shannon theorem]*, Moscow, Radio Engineering Publ., 2004, in Russian.

[16] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Sums, Series and Products*, Moscow: Gos. Izd-vo in fiz.-mat. lit-ry, 1962, in Russian.

[17] L. N. Temirbekova and G. Dairbaeva, Applied and Computational Mathematics **12(2)**, 234-246 (2013).

[18] M. M. Lavrentiev, *Ill-posed Problems for Differential Equations*, Novosibirsk: NSU, 1981, in Russian.
[19] G. Dairbaeva, L. N. Temirbekova, Izvestiya NAS RK, Physical-Mathematical Series **1**, 3-9, (2012), in Russian.