

# Aplikace matematiky

---

Václav Červ; Karel Segeth

A comparison of the accuracy of the finite-difference solution to boundary value problems for the Helmholtz equation obtained by direct and iterative methods

*Aplikace matematiky*, Vol. 27 (1982), No. 5, 375–390

Persistent URL: <http://dml.cz/dmlcz/103983>

## Terms of use:

© Institute of Mathematics AS CR, 1982

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

## A COMPARISON OF THE ACCURACY OF THE FINITE-DIFFERENCE SOLUTION TO BOUNDARY-VALUE PROBLEMS FOR THE HELMHOLTZ EQUATION OBTAINED BY DIRECT AND ITERATIVE METHODS

VÁCLAV ČERV, KAREL SEGETH

(Received November 17, 1980)

### 1. INTRODUCTION

The development of iterative methods for solving linear algebraic systems [14], [16] has brought the question of when the employment of these methods is more advantageous than the use of the direct ones. This question can be considered from various points of view from which the time requirements, the storage requirements, and the achieved accuracy of the solution undoubtedly seem to be the most important.

The final decision in the contention between the advocates of direct and iterative methods has not yet been taken. In fact, such a decision cannot be taken if our question is formulated quite generally. Nevertheless, solving problems from exactly determined classes (in particular, boundary-value problems for partial differential equations) and employing clearly chosen criteria, we can — at least experimentally — compare direct and iterative methods. We have attempted such a comparison in this paper. Our test problems are taken from geophysics, namely, from geoelectric research, and they are concerned with the numerical modeling of the electromagnetic field.

The numerical modeling of the electromagnetic field in two-dimensionally horizontally inhomogeneous media, which forms the theoretical foundation for the interpretation of the geophysical magneto-telluric and magneto-variational measurements, has been recently paid close attention. The numerical methods most frequently used are the finite difference and finite element methods [2], [4], [5], [6], [9], which make it possible to solve the problem for considerably general models of two-dimensionally horizontally inhomogeneous structures. The numerical treatment consists in a transformation of the Helmholtz partial differential equation to a linear algebraic system. The system obtained, consisting usually of several thousands equations, is then to be solved by a suitable numerical method. In this paper we are concerned with some problems arising in the computation of the solution of the above

problem, especially in solving the obtained linear algebraic system by direct and iterative methods.

In Section 2 we formulate the boundary-value problem the solution of which is studied in the paper. The medium is supposed to be two-dimensionally horizontally inhomogeneous, i.e., the electric conductivity is supposed to be independent of one horizontal coordinate, say  $x$ . The Maxwell equations then possess two independent solutions,  $E$ -polarization and  $H$ -polarization. Their components  $E_x$  and  $H_x$  satisfy the Helmholtz equation and fulfil the corresponding boundary and interface conditions.

In the conclusion of this section we construct the finite-difference approximation for both the problems considered with the help of the Green theorem.

The results of numerical experiments and their discussion are given in Section 3. The computations have been carried out for two models, which differ from each other by the type of dependence of the conductivity on  $y$  and  $z$ , i.e. by the absolute term in the Helmholtz equation.

For model 1 (Section 3.1) the values of  $E_x$  and  $H_x$  are computed. From these quantities we further numerically determine  $\partial E_x/\partial y$ ,  $\partial E_x/\partial z$  and  $\partial H_x/\partial z$ , i.e. the derived secondary quantities important for the practical application of the results obtained. In addition, the values of  $\partial H_x/\partial z$  are computed by another independent method for comparison.

For model 2 (Section 3.2) only the values of  $E_x$  (and the secondary quantities, i.e.  $\partial E_x/\partial y$  and  $\partial E_x/\partial z$ ) are computed.

Our further detailed discussion is concerned with the rate of convergence of the quantity  $\partial E_x/\partial y$  (Section 3.3) and the dependence of the convergence of  $E_x$  on the overrelaxation factor  $\omega$  (Section 3.4).

In conclusion we survey the storage and time requirements of the methods used for solving linear algebraic systems in Section 3.5.

## 2. FORMULATION OF THE PROBLEM

Let the domain  $\Omega$  be such a section through a two-dimensionally horizontally inhomogeneous structure that the boundary  $\Gamma_1$  corresponds to the upper boundary of the air layer above the Earth's surface ( $z = 0$ ). We assume that the parameters of the medium (electric conductivity) do not depend on one horizontal coordinate, say  $x$ . The  $x$ -axis then represents the axis of the geoelectric homogeneity of the model. The electric conductivity changes in the  $yz$ -plane orthogonal to the axis of homogeneity (see Figures 1 and 2).

Assuming the time factor  $\exp(i\omega_f t)$  and neglecting the displacement currents, we can determine two independent solutions,  $H$ - and  $E$ -polarization, from the general system of Maxwell equations. We have either the component  $H_x$  or  $E_x$  of the source field parallel to the axis of homogeneity. For both the polarizations we solve the

equation

$$(2.1) \quad \Delta u(y, z) - i\eta u(y, z) = 0, \quad \eta = \omega_f \mu \sigma,$$

in the domain  $\Omega$ , where the solution  $u(y, z) = u_1(y, z) + iu_2(y, z)$  is either  $H_x$  or  $E_x$ . The quantities  $\omega_f$ ,  $\mu$  and  $\sigma$  represent the angular frequency, magnetic permeability in  $\text{Hm}^{-1}$  and electric conductivity in  $\Omega^{-1}\text{m}^{-1}$ , respectively.

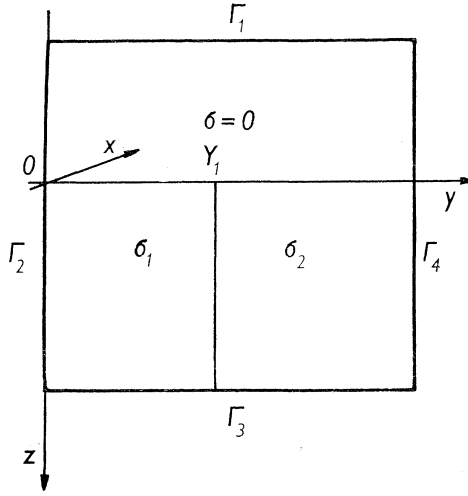


Fig. 1. The domain  $\Omega$  for model 1.

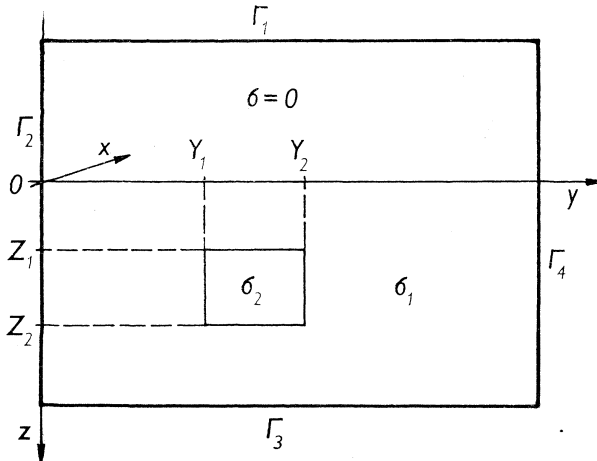


Fig. 2. The domain  $\Omega$  for model 2.

On the boundary of the domain  $\Omega$  we put

$$(2.2) \quad u = g = g_1 + ig_2,$$

where  $g$  is a given function defined and continuous on the boundary.

We suppose that the conductivity  $\sigma_i$  is constant in the individual subdomains  $\Omega_i$  of the domain  $\Omega$ . We require that the following conditions be fulfilled on the interfaces between the subdomains:

- (i) the continuity of the solution  $u$  (for both the polarizations);
- (ii) the continuity of the normal derivative  $\partial u/\partial n$  (for  $E$ -polarization);
- (iii) the continuity of the function  $\sigma^{-1} \partial u/\partial n$  (for  $H$ -polarization).

We assume that a geoelectric inhomogeneity is situated inside the domain  $\Omega$  and that the boundary of the domain  $\Omega$  is sufficiently far from this inhomogeneity so that the influence of the inhomogeneity can be neglected on the boundaries  $\Gamma_2$  and  $\Gamma_4$ . Examining the Maxwell equations, we can show that the magnetic field is constant on and above the Earth's surface (the air layer with  $\sigma = 0$ ) in the case of  $H$ -polarization [5], [9]. We can put  $H_x = \text{const} = 1$  for  $z = 0$  and solve (2.1) only in the conductive domain representing the Earth conductor. In the case of  $E$ -polarization it is necessary to consider a sufficiently thick air layer with zero conductivity above the Earth's surface, in which the secondary field of the conductive half-space is damped [10].

Introducing now the coordinate system according to Figure 1 or 2, we cover the domain  $\Omega$  by two mutually orthogonal systems of lines  $y = y_i$  ( $i = 0, 1, \dots, M + 1$ ) and  $z = z_j$  ( $j = 0, 1, \dots, N + 1$ ). Their intersections for  $i = 0, M + 1$  and  $j = 0, N + 1$  are boundary nodes, the other intersections are interior nodes. Put  $h_i = y_{i+1} - y_i$  and  $k_j = z_{j+1} - z_j$ . The grid is chosen in such a way that the interfaces coincide with some parts of some lines of the grid.

Let  $A_{ij}$  be an interior node of the domain  $\Omega$  at which several subdomains of various conductivity touch each other (see Fig. 3). Integrating the equation (2.1) over the rectangle with center at  $A_{ij}$  and sides  $(h_i + h_{i+1})/2$  and  $(k_j + k_{j+1})/2$ , using the Green theorem, and replacing the normal derivatives by differences [5], we obtain the difference analog of the equation (2.1) at the node  $A_{ij}$ ,

$$(2.3) \quad -P_{ij}U_{i+1,j} - Q_{ij}U_{i-1,j} - R_{ij}U_{i,j+1} - S_{ij}U_{i,j-1} + (T_{ij} + iV_{ij})U_{ij} = 0.$$

For  $H$ -polarization we put

$$\begin{aligned} P_{ij} &= (2h_i)^{-1} (\sigma_2^{-1}k_{j-1} + \sigma_4^{-1}k_j), \\ Q_{ij} &= (2h_{i-1})^{-1} (\sigma_1^{-1}k_{j-1} + \sigma_3^{-1}k_j), \\ R_{ij} &= (2k_j)^{-1} (\sigma_3^{-1}h_{i-1} + \sigma_4^{-1}h_i), \\ S_{ij} &= (2k_{j-1})^{-1} (\sigma_1^{-1}h_{i-1} + \sigma_2^{-1}h_i), \end{aligned}$$

$$T_{ij} = P_{ij} + Q_{ij} + R_{ij} + S_{ij},$$

$$V_{ij} = \frac{1}{4}\omega_f\mu(h_{i-1} + h_i)(k_{j-1} + k_j).$$

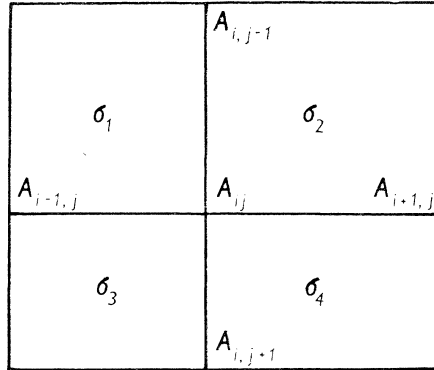


Fig. 3. An interior node  $A_{ij}$  on the boundary of four subdomains with different conductivity.

For  $E$ -polarization we put

$$P_{ij} = (2h_i)^{-1}(k_{j-1} + k_j), \quad Q_{ij} = (2h_{i-1})^{-1}(k_{j-1} + k_j),$$

$$R_{ij} = (2k_j)^{-1}(h_{i-1} + h_i), \quad S_{ij} = (2k_{j-1})^{-1}(h_{i-1} + h_i),$$

$$T_{ij} = P_{ij} + Q_{ij} + R_{ij} + S_{ij},$$

$$V_{ij} = \frac{1}{4}\omega_f\mu(\sigma_1 h_{i-1} k_{j-1} + \sigma_2 h_i k_{j-1} + \sigma_3 h_{i-1} k_j + \sigma_4 h_i k_j).$$

The presented approximation is well-founded for  $E$ -polarization and, moreover, for  $H$ -polarization if  $\sigma_1 = \sigma_2$  and  $\sigma_3 = \sigma_4$  (vertical contact) or if  $\sigma_1 = \sigma_3$  and  $\sigma_2 = \sigma_4$  (horizontal contact). For a completely general choice of conductivities  $\sigma_1, \sigma_2, \sigma_3$  and  $\sigma_4$ , a derivative of the solution may have a singularity at the node  $A_{ij}$ . Employing our approximation (2.3) we thus introduce a certain error. As shown in [1], however, this error is limited only to the closest neighborhood of the critical node and cannot influence the approximate solution in the domain in general.

### 3. RESULTS OF NUMERICAL EXPERIMENTS

The problem formulated in Section 2 is thus transformed into solving a linear algebraic system, where the unknowns  $U_{ij}$  are the values of the approximate solution at nodes. The system is of a relatively high order (in our experimental models about 1000–1500 equations) and its matrix is complex, sparse, and symmetric. If the nodes are properly numbered the matrix of the system is block tridiagonal. Its diagonal blocks are tridiagonal, its off-diagonal blocks are diagonal.

Both the direct and iterative methods can be efficiently used for solving such a system. In our numerical experiments we have concentrated on the Gaussian elimination (GE) and the successive (point) overrelaxation (SOR) with factor  $\omega$  (in the particular case of  $\omega = 1$  it is usually called the Gauss-Seidel method) [12], [14], [16]. We are not able to guarantee the convergence of the SOR method generally for our problem by any simple criterion. In our experiments we have employed SOR either with various fixed values of  $\omega$  or with the value of  $\omega$  successively optimized in the course of computation [11].

In practical considerations, the first derivatives of the sought solution  $u$  of the equation (2.1), i.e. the functions  $\partial u/\partial y$  and  $\partial u/\partial z$ , have a physical meaning, rather than the function  $u$  itself. We compute these derivatives from the finite-difference solution by numerical differentiation, i.e., we construct an interpolation polynomial and differentiate it. The results of the numerical differentiation employing interpolation polynomials of various degrees can differ from each other. The results presented in the paper have been obtained with the help of a polynomial of the second degree (before differentiation) unless otherwise stated.

The most important values for the practical application of the results are  $E_x$ ,  $\partial E_x/\partial y$ ,  $\partial E_x/\partial z$  and  $\partial H_x/\partial z$ , all with  $z = 0$ . In the following discussion we always speak about these values (with  $y$  variable) unless otherwise stated. All the coordinates are given in kilometers.

We wish to point out that all the comparisons made in the following are considered from the practical point of view. The computation of geophysical fields serves mainly as a theoretical foundation for the interpretation of the measured data. The quantities usually measured in the field are proportional to the quotient of the both first partial derivatives of the function  $u$  or to the quotient of a derivative and the function  $u$  itself. The result of measurements is an average obtained from a great number of measured values.

The interpretation consists in estimating the agreement of the experimental and the theoretical curve and comparing their character. In this connection we should consider the fact that the inverse problem, i.e. the problem to find the coefficients of the equation (2.1) for a given function  $u$ , is not well-posed. In general we can say that the values of  $u$  should be determined with such an accuracy that the values of the derived secondary practical quantities have the accuracy of several per cent.

The fact that the result of our numerical modeling is — in addition to the function  $u$  and its derivatives — also the quotient of these derivatives, influences the employment of iterative methods. Examining the course of convergence of the iterative process with respect to the function  $u$  sought, we can find out that its derivatives as well as their quotient differ substantially, as far as two consecutive iteration steps are concerned, even if the values of the approximated function  $u$  itself change very little. Similarly, if the values of the approximated function  $u$  computed by a direct and an iterative method differ from each other very little, the difference may be considerable for the derivatives of  $u$  and their quotient.

A numerical process starting from a proper variational formulation and employing the finite element method with elements of degree at least 3 would apparently be efficient for computation of the solution of our problem and its first derivatives [13]. We have not tried such a process in practice since it is obviously both time and storage consuming.

All the computations have been carried out on an IBM System/370 Model 135 computer. Single precision is concerned unless otherwise stated.

To compare GE and SOR, we used two models that show a typical behavior. Discussing the results, we do not consider the discretization error, which is the same for GE as for SOR. We analyse only the error of the numerical solution of the linear algebraic system, which influences the accuracy of both  $u$  and its derivatives. We put  $T = 10$  s, i.e.  $\omega_f \mu = 8\pi^2 10^{-8} \text{ Hm}^{-1} \text{ s}^{-1}$ , in the equation (2.1).

### 3.1. Model 1

We put  $\sigma_1 = 10^{-1} \Omega^{-1} \text{ m}^{-1}$ ,  $\sigma_2 = 10^{-3} \Omega^{-1} \text{ m}^{-1}$  and  $Y_1 = 150$  km for model 1 (Fig. 1). We computed both  $E_x$  and  $H_x$  from the equation (2.1) for this model.

Computing  $E_x$ , we choose the rectangle  $[0, 450] \times [-300, 200]$  or  $[0, 380] \times [-300, 200]$  for the domain  $\Omega$ . These two cases differ from each other also by the choice of the grid (grid 1:  $30 \times 35$  meshes, grid 2:  $36 \times 35$  meshes). The comparison of the direct and the iterative method, however, presents the same picture in both

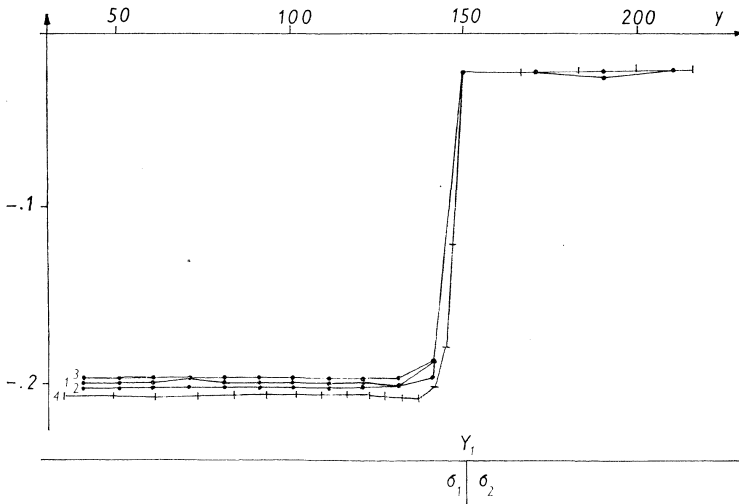


Fig. 4. Model 1, function  $\text{Re } \partial H_x(y, 0)/\partial z$ . 1 — computation according to (3.1); 2 — GE, double precision, grid 1; 3 — SOR ( $\omega = 1$ , 200 steps), interpolation of degree 3, grid 1; 4 — GE, double precision, grid 2. The function  $\text{Im } \partial H_x(y, 0)/\partial z$  has a similar behavior (cf. also Table 2). In this as well as the following figures, a single line is plotted where two or more lines merge in each other.



the grids. The behavior of  $E_x(y, 0)$  (in dependence on  $y$ ) is smooth. At both the left and right-hand ends of  $\Omega$ ,  $E_x$  approaches the value that corresponds to the field of the homogeneous Earth with the conductivity  $\sigma_1$  and  $\sigma_2$ , respectively.

Both the methods (GE and SOR with  $\omega = 1$  and 200 iteration steps) give results the components of which differ at most by  $4^0/00$  (if the difference is related to  $|E_x(y, 0)|$  for the corresponding  $y$ , cf. Table 1). Similar results are obtained also for  $\partial E_x/\partial y$  and  $\partial E_x/\partial z$ . GE has been tested in single as well as double arithmetic and has given the same results.

Computing  $H_x$ , we choose the rectangle  $[0, 450] \times [-260, 0]$  or  $[0, 380] \times [-255, 0]$  for the domain  $\Omega$ . We consider these two cases with different grids on different rectangles (grid 1:  $30 \times 25$  meshes, grid 2:  $36 \times 23$  meshes). The behavior of  $\partial H_x(y, 0)/\partial z$  on the left as well as on the right of the interface  $y = Y_1$  is almost constant. Here, this function is equal to the value corresponding to the field of the homogeneous Earth with the respective conductivity  $\sigma$ . The function  $\partial H_x/\partial z$  has a jump for  $y = Y_1$  (Fig. 4). The difference between the results obtained by GE and SOR ( $\omega = 1$ , 200 steps) is negligible, in components at most about  $4^0/00$ , if the difference is related to  $|\partial H_x(0, 0)/\partial z|$  (Table 2). Also the results obtained by the same method in two grids are different since the solution  $H_x$  grows rapidly at the left-hand end of the rectangle and a sufficiently fine grid is needed to approximate it.

For comparison we also determine the approximate values of  $\partial H_x/\partial z$  by numerical quadrature from the formulae presented in [15]:

$$(3.1) \quad \frac{\partial H_x}{\partial z}(y, z) = -v_1 \exp(-v_1 z \sqrt{i}) -$$

$$- v_1^2(v_2^2 - v_1^2) \frac{2i}{\pi} \int_0^\infty \frac{\xi^2 \exp(\mu_1(y - Y_1)) \cos \xi z}{\mu_1^2 \mu_2 (\mu_1 v_2^2 + \mu_2 v_1^2)} d\xi, \quad y \leq Y_1,$$

$$\frac{\partial H_x}{\partial z}(y, z) = -v_2 \exp(-v_2 z \sqrt{i}) +$$

$$+ v_2^2(v_2^2 - v_1^2) \frac{2i}{\pi} \int_0^\infty \frac{\xi^2 \exp(-\mu_2(y - Y_1)) \cos \xi z}{\mu_1 \mu_2^2 (\mu_1 v_2^2 + \mu_2 v_1^2)} d\xi, \quad y > Y_1,$$

where  $v_j^2 = 4\pi\sigma_j\omega_j$ ,  $\mu_j^2 = \xi^2 + iv_j^2$ ;  $j = 1, 2$  (Fig. 4, Table 2). The integral in (3.1) is replaced by an integral to  $B$  for a sufficiently large  $B$  and then computed with the help of repeated application of the Gauss quadrature formula (with 6 abscissae) to small subintervals. Establishing the formula (3.1), we neglect some small quantities. The total error caused by this fact is not examined in [15]. Nonetheless, we can say that the formula (3.1) has proven useful for our model.

The results computed by SOR ( $\omega = 1$ , 200 steps) and by the employment of interpolation polynomials of degree 2 or 3 for numerical differentiation (Table 2) differ from each other in components by as much as  $18^0/00$  (if related to the value

Table 1. Model 1, grid 1, the function  $E_x(y, 0)$  at some nodes.

$y$	0	100	110	120	130	140	150	170	190	210	230	250	450
GE	Re	1.000	.994	1.010	1.033	1.088	1.335	2.397	5.177	6.847	7.874	8.518	8.927
	Im	.000	-.295	-.316	-.351	-.428	-.577	-.370	1.277	1.571	1.469	1.254	1.027
SOR, $\omega = 1$ , 200 steps	Re	1.000	.993	1.009	1.032	1.088	1.334	2.396	5.169	6.834	7.858	8.499	8.907
	Im	.000	-.295	-.316	-.351	-.427	-.575	-.367	1.289	1.591	1.497	1.288	1.065

Table 2. Model 1, grid 1, the function  $\partial H_x(y, 0)/\partial z$  at some nodes.

$y$	10	110	120	130	140	150	170	190	210	430
GE, double precision	Re	-.201	-.201	-.201	-.201	-.190	-.022	-.021	-.020	-.020
	Im	-.190	-.190	-.190	-.191	-.198	-.024	-.020	-.020	-.020
SOR, $\omega = 1$ , 200 steps	Re	-.201	-.201	-.201	-.201	-.190	-.023	-.021	-.020	-.020
	Im	-.190	-.190	-.190	-.191	-.198	-.023	-.020	-.020	-.020
computation according to (4.1)	Re	-.199	-.199	-.199	-.200	-.199	-.023	-.021	-.020	-.020
	Im	-.199	-.199	-.199	-.198	-.211	-.024	-.020	-.020	-.020
SOR, $\omega = 1$ , 200 steps, interpolation of degree 3	Re	-.199	-.198	-.198	-.199	-.188	-.023	-.021	-.020	-.020
	Im	-.195	-.195	-.195	-.196	-.203	-.023	-.020	-.020	-.020

$|\partial H_x(0, 0)/\partial z|$ ), which shows that the procedures used for numerical differentiation are questionable.

### 3.2. Model 2

Model 2 (Fig. 2) is more interesting from the point of view of our comparison. We put  $\sigma_1 = 10^{-3} \Omega^{-1} \text{m}^{-1}$ ,  $\sigma_2 = 10^{-1} \Omega^{-1} \text{m}^{-1}$ ,  $Y_1 = 240 \text{ km}$ ,  $Y_2 = 280 \text{ km}$ ,  $Z_1 = 6 \text{ km}$  and  $Z_2 = 16 \text{ km}$ . We compute only  $E_x$ , namely in two grids on the rectangle  $[0, 520] \times [-226, 200]$  (both grids with  $38 \times 38$  meshes). The functions  $E_x$  and  $\partial E_x/\partial z$  are symmetric with respect to the point  $y = 260$  while the function  $\partial E_x/\partial y$  is skew-symmetric. The former two functions reach their extremum at  $y = 260$  whereas the function  $\partial E_x/\partial y$  has a zero at this point and its local extrema are skew-symmetrically located near the point  $y = 260$ . The comparison of the methods is very similar for both the grids.

For model 2 we also employ the decomposition of the sought solution  $E_x$  into the sum of the known primary field  $E_x^p$  of the homogeneous Earth and a secondary field  $E_x^s$  (for which a homogeneous boundary condition is prescribed on  $\Gamma$ ), i.e.

$$(3.2) \quad E_x(y, z) = E_x^p(y, z) + E_x^s(y, z).$$

Substituting (3.2) into (2.1), we obtain a differential equation of the type (2.1) for  $E_x^s$ . Computing the derivatives of the function  $E_x$ , we can employ the known analytic expression for  $E_x^p$  and numerically differentiate only  $E_x^s$ .

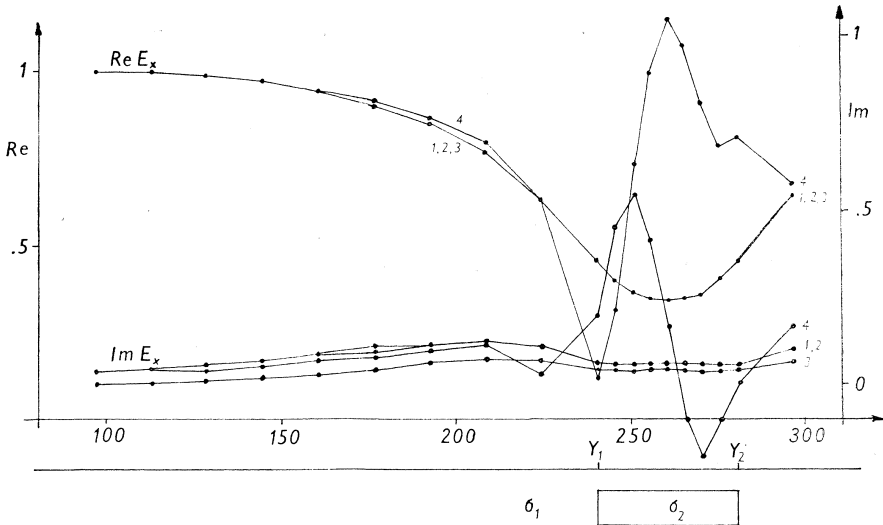


Fig. 5. Model 2, grid 1, function  $E_x(y, 0)$ . 1 — GE, the decomposition (3.2) is used; 2 — GE; 3 — SOR ( $\omega = 1$ , 200 steps); 4 — SOR ( $\omega = 1.6$ , 200 steps).

The results for  $E_x$  in grid 1 (Fig. 5) show a relatively good agreement of all methods. SOR with  $\omega = 1.6$  and 200 steps apparently does not converge. The results with this  $\omega$  are not symmetric and after 400 steps (not shown in the figure) the results differ from the exact solution even more. GE shows the same behavior in single as well as double arithmetic.

After differentiation, the results obtained in grid 1 differ from each other more than in the case of  $E_x$  itself. A typical behavior is represented by  $\partial E_x(y, 0)/\partial z$  in Fig. 6.

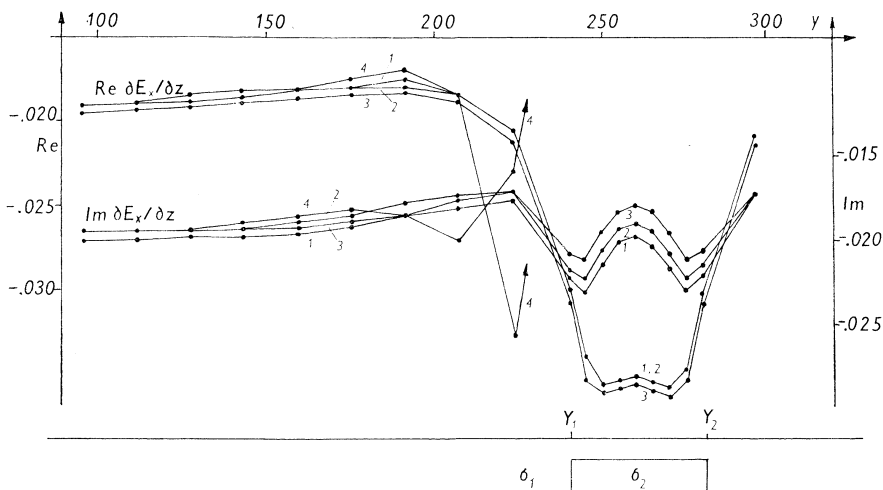


Fig. 6. Model 2, grid 1, function  $\partial E_x(y, 0)/\partial z$ . 1 — GE, the decomposition (3.2) is used; 2 — GE; 3 — SOR ( $\omega = 1$ , 200 steps); 4 — SOR ( $\omega = 1.6$ , 200 steps). Curve 4 runs out of the scope of the figure.

### 3.3. Convergence of Derivatives

Model 2 (grid 2) has been studied also from the point of view of the rate of convergence of the derivatives of the solution  $E_x$ . The most typical behavior is represented by the quantity  $|\partial E_x(y, 0)/\partial y|$  (see Fig. 7).

The results demonstrate a considerable dependence of  $\partial E_x/\partial y$  on the method used for solving the system. They are worst for SOR with  $\omega = 1$  and 200 steps. The results are less dependent on the method used for solving the system at those parts of  $\Omega$  where the mesh size is less. Quite good results are reached already for  $\omega = 1$  and 400 steps in the neighborhood of the inhomogeneity where small mesh size is chosen since this is an important part of the profile  $z = 0$  from the point of view of interpretation. The results with  $\omega = 1.65 \approx \omega_{opt}$  seem to be the best along the whole profile. The results obtained by GE are very close to the best ones. With  $\omega = 1.2$  we obtain, after 400 steps of SOR, results practically equivalent to the computation

with  $\omega = 1$  and 600 steps. The results with  $\omega = 1.4$  and 400 steps are better than those computed with  $\omega = 1$  and 800 steps.

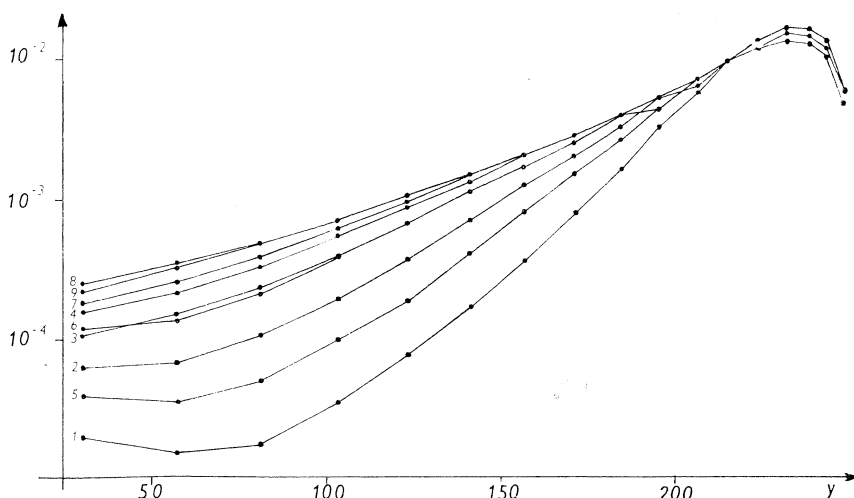


Fig. 7. Model 2, grid 2, function  $|\partial E_x(y, 0)/\partial y|$ . 1 — SOR ( $\omega = 1$ , 200 steps); 2 — SOR ( $\omega = 1$ , 400 steps); 3 — SOR ( $\omega = 1$ , 600 steps); 4 — SOR ( $\omega = 1$ , 800 steps); 5 — SOR ( $\omega = 1.2$ , 200 steps); 6 — SOR ( $\omega = 1.2$ , 400 steps); 7 — SOR ( $\omega = 1.4$ , 400 steps); 8 — SOR ( $\omega = 1.65 \approx \omega_{opt}$ , 400 steps); 9 — GE.

### 3.4. Dependence of the Convergence on Factor $\omega$

The behavior of the maximum of the magnitude of the change of the solution in one iteration step as a function of the number of steps of SOR is presented in Fig. 8 for the same model and grid as in Section 3.3. Although the difference of the corresponding quantities (as far as two consecutive iteration steps are concerned) is very small for all the quantities after 400 steps (see Fig. 7), the values of the derivatives of the solution need not be satisfactory in the whole  $\Omega$ .

The value  $\omega = 1.65$  has been experimentally and approximately determined as the optimal one. The iterative process diverges already for  $\omega = 1.73$ .

To determine  $\omega_{opt}$  we have also tried the algorithm proposed in [11] and originating in [3]. Several parameters have to be a priori chosen in this algorithm.

With regard to our little experience in this matter and an apparently unsuitable choice of these parameters, our results obtained with the help of this procedure are not satisfactory. In general we could say that if the values of the solution do not vary too much in the domain considered, the value of  $\omega_{opt}$  reaches 1.8 to 1.9. If the solution varies very much in the domain considered the value of  $\omega_{opt}$  is close to 1.0 to 1.1. With  $\omega = 1$ , our iterative process has always converged. The convergence gets

better when  $\omega$  grows up to the value  $\omega_{opt}$ . For the factor  $\omega$  greater than  $\omega_{opt}$ , SOR usually diverges.

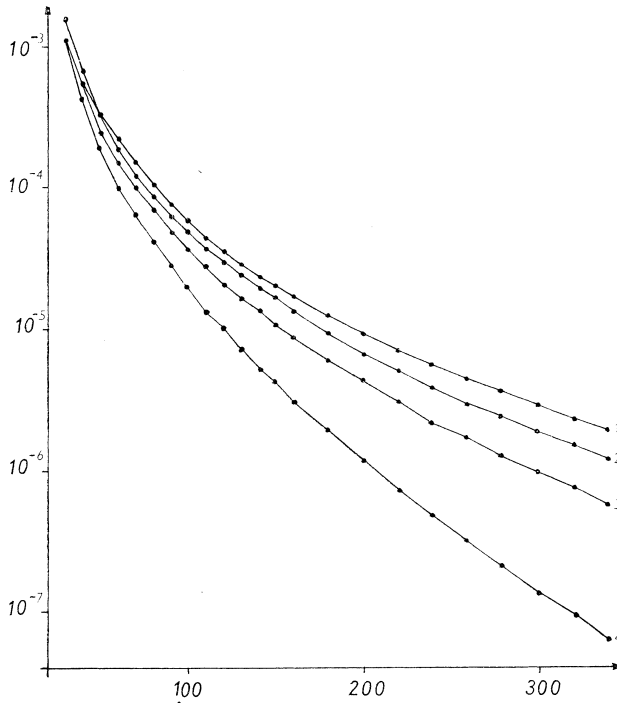


Fig. 8. Model 2, grid 2, the maximum of the magnitude of the difference between  $E_x(y, z)$  in two consecutive iteration steps (over all  $y$  and  $z$ ). SOR, 400 steps. 1 -  $\omega = 1$ ; 2 -  $\omega = 1.2$ ; 3 -  $\omega = 1.4$ ; 4 -  $\omega = 1.65$ .

We can see in Fig. 7 that the quality of results considerably depends on the choice of the grid. For a suitable, sufficiently fine grid, the results do not depend too much on the choice of the method for solving the system. If the grid is inadequate there is a great difference between the results obtained by various methods and the quality of the results (influenced, in addition, by the discretization error) is always questionable.

### 3.5. Storage and Time Requirements

The comparison of direct and iterative methods would not be complete without a survey of the storage and time requirements of these methods.

The storage requirements can be easily determined. If the matrix of the system is considered in GE to be a band matrix, the bandwidth for our models is  $2N + 1$

and the total number of equations (and unknowns) is  $MN$ . The number of words necessary for storing the non-zero elements of the matrix is thus proportional to  $MN^2$ . Since only  $N + 1$  rows of the matrix of the system are necessary for performing one step of GE, the amount of the main storage words required for storing the non-zero elements of the matrix is proportional to  $N^2$ . Each row is stored on disk after the corresponding elimination step has been finished and a new row is generated in the main storage. The storage requirements on disk are thus of order  $MN^2$  words. A reduction of storage requirements can be achieved at the cost of detriment of the need of time.

For the SOR method we need not store the matrix at all. The values of its elements are computed only at the moment when they are needed. The main storage requirements are thus of order  $MN$  words, namely for the  $MN$  unknowns.

We wish to remark that the so-called fast direct methods for solving systems [7], [8] would deserve a special attention. They cannot, however, be applied to the equation (2.1) directly. It is necessary to solve a sequence of simpler problems the solution of which converges to the solution of (2.1). Our experience with the employment of the fast direct methods is not yet sufficient for us to be able to judge its efficiency. The principal drawback of the methods of this kind, i.e. the requirement to choose a constant mesh size, plays, however, an important role in solving problems of our type.

The discussion of the time requirements of the individual methods is far from being easy. An exact time measurement is not feasible on computers with multiprogramming and virtual storage. The results of the measurements of the CPU time can differ from each other even more than by 10%, even when always the same problem is solved. Rough time measurements show that, for about 1000 to 1500 equations, the computation employing GE takes about the same time as 200 steps of SOR. The number of operations in GE is proportional to the quantity  $MN^3$  and the computation time thus grows as  $MN^3$  as well.

The number of operations needed for a step of SOR is proportional to  $MN$ . The total time for iterative methods is thus proportional to  $KMN$ , where  $K$  is the number of steps necessary to reach the prescribed accuracy. This number certainly depends on the number of equations and grows with it. For a rather moderate number of equations (1000 to 1500), direct methods seem to be more advantageous than iterative ones. For a medium number of equations (several thousand), both the methods are comparable. It is advantageous to use  $\omega > 1$  for iterative methods. If the value of  $\omega$  is close to  $\omega_{opt}$ , usually only 1/3 iterations are needed to reach the required accuracy as compared with the Gauss-Seidel method ( $\omega = 1$ ). For a greater number of equations, iterative methods get superior to direct ones. It is primarily due to the fact that, when employed, they can still operate in the main storage of computer while direct (elimination) methods are bound to use the auxiliary storage substantially, which makes the computation more time-consuming.

## References

- [1] *I. Babuška, R. B. Kellog*: Numerical solution of the neutron diffusion equation in the presence of corners and interfaces. Numerical Reactor Calculations. International Atomic Energy Agency, Vienna 1972, 473—486.
- [2] *V. Bezdová, K. Segeth*: Mathematical modeling of electromagnetic fields. The Use of Finite Element Method and Finite Difference Method in Geophysics. (Proceedings of Summer School, Liblice 1977.) Geofyzikální ústav ČSAV, Praha 1978, 329—332.
- [3] *B. A. Carre*: The determination of the optimum accelerating factor for successive over-relaxation. Comput. J. 4 (1961), 73—78.
- [4] *V. Červ*: Numerical modelling of geoelectric structures using the Galerkin and finite element methods. Studia geoph. et geod. 22 (1978), 283—294.
- [5] *V. Červ, O. Praus*: Numerical modelling in laterally inhomogeneous geoelectrical structures. Studia geoph. et geod. 22 (1978), 74—81.
- [6] *J. H. Coggon*: Electromagnetic and electrical modelling of induction effects in laterally non-uniform conductors. Phys. Earth Planet. Inter. 10 (1975), 265.
- [7] *P. Concus, G. H. Golub*: Use of fast direct methods for the efficient numerical solution of nonseparable elliptic equations. SIAM J. Numer. Anal. 10 (1973), 1103—1120.
- [8] *F. W. Dorr*: The direct solution of the discrete Poisson equation on a rectangle. SIAM Rev. 12 (1970), 248—263.
- [9] *F. W. Jones, L. J. Pascoe*: A general computer program to determine the perturbation of alternating electric currents in a two-dimensional model of a region of uniform conductivity with an embedded inhomogeneity. Geophys. J. Roy. Astronom. Soc. 23 (1971), 3.
- [10] *F. W. Jones, A. T. Price*: The perturbation of alternating geomagnetic fields by conductivity anomalies. Geophys. J. Roy. Astronom. Soc. 20 (1970), 317.
- [11] *I. Marek*: On the SOR method for solving linear equations in Banach spaces. Wiss. Z. Tech. Hochsch. Karl-Marx-Stadt 11 (1969), 335—341.
- [12] *A. Ralston*: A first course in numerical analysis. McGraw-Hill, New York 1965.
- [13] *G. Strang, G. J. Fix*: An analysis of the finite element method. Prentice-Hall, Englewood Cliffs, N. J., 1973.
- [14] *R. S. Varga*: Matrix iterative analysis. Prentice-Hall, Englewood Cliffs, N. J., 1962.
- [15] *J. T. Weaver*: The electromagnetic field within a discontinuous conductor with reference to geomagnetic micropulsations near a coastline. Canad. J. Phys. 41 (1963), 484—495.
- [16] *D. M. Young*: Iterative solution of large linear systems. Academic Press, New York 1971.

## Souhrn

### SROVNÁNÍ PŘESNOSTI SÍŤOVÉHO ŘEŠENÍ OKRAJOVÝCH ÚLOH PRO HELMHOLTZOVU ROVNICI, ZÍSKANÉHO PŘÍMÝMI A ITERAČNÍMI METODAMI

VÁCLAV ČERV, KAREL SEGETH

Rozvoj iteračních metod pro řešení soustav lineárních algebraických rovnic přinesl otázku, kdy jsou tyto metody výhodnější než metody přímé. V článku jsme se pokusili pomocí numerického experimentu porovnat přímé a iterační metody při řešení



jisté třídy okrajových úloh pro parciální diferenciální rovnice, které slouží k numerickému modelování elektromagnetického pole v geofyzice.

V odstavci 2 je formulována úloha a její aproximace pomocí konečných diferencí (sítí). Odstavec 3 je věnován výsledkům numerických experimentů a porovnání časových a paměťových nároků metod a dosažené přesnosti řešení. Srovnání jsou ilustrována tabulkami a obrázky.

*Authors' addresses:* Václav Červ, CSc., Geofyzikální ústav ČSAV, Boční II, 141 31 Praha 4.  
RNDr. Karel Segeth, CSc., Matematický ústav ČSAV, Žitná 25, 115 67 Praha 1.