NUMERICAL METHODS OF SOLVING A SYSTEM OF MULTI-DIMENSIONAL NONLINEAR EQUATIONS OF THE DIFFUSION TYPE

A. V. Agapov, B. I. Kolosov

NUMERICAL METHODS FOR SOLVING SYSTEMS OF MULTI-DIMENSIONAL NON-LINEAR EQUATIONS OF THE DIFFUSION TYPE

A. V. Agapov and B. I. Kolosov

SCITRAN
Box 5456
Santa Barbara, CA 93108

National Aeronautics and Space Administration
Washington, D.C. 20546


This article examines the numerical algorithms for solving non-linear equations of the diffusion type, used to formulate the principles of conservation and stability of difference schemes achieved using the iteration control method. For the schemes obtained of the predictor-corrector type, the conversion is proved for the control sequences of approximate solutions to the precise solutions in the Sobolev metrics. Within the framework of the mathematical system developed, algorithms are also proposed for reducing the differential problem to integral relationships, whose solution methods are known. The algorithms for the problem solution are classified depending on the non-linearity of the diffusion coefficients, and practical recommendations for their effective use are given.

Unclassified - Unlimited

Unclassified

23
INTRODUCTION

The non-steady state-equations of the diffusion type are the most widely used equations, to which the description of many important physical phenomena are reduced. The literature \[5, 6\] has made an extensive study of the linear class of these equations, and many numerical methods for solving them on computers in the one-dimensional and the multidimensional geometry have been developed and have become classical. The situation is different in the case of the nonlinear dependence of the diffusion coefficients on the initial solution, which frequently arise when studying nonlinear effects in the theory of quasi-linear plasma oscillations, in the problems of heat and mass transfer, and other present day physical applications. This article proposes a universal numerical algorithm for solving these problems, having the property of conservation and numerical stability. The formulation and study of the algorithms presented below are based on a priori information connected with the initial problem.

We should also note that since certain existing physical phenomena are examined, which are modeled by nonlinear equations of the diffusion type, the existence for solutions of these equations is a natural requirement, which indicates the useful information of the initial physical model. It is known that diffusion equations are based on a balance of the flow of physical quantities determined by a differential expression of the form \[ \frac{\partial f}{\partial t} \], where \( f \) is the physical quantity examined; \[ \frac{\partial f}{\partial y} \] -- diffusion coefficient. Since in modern physics the laws of conservation are only considered in geometric space, but in the phase space of velocity or the Fourier transformation space, in the following the variable

*Numbers in the margin indicate pagination of original foreign text.
\( \hat{y} = (y_1, y_2) \) will designate the parameters of the corresponding space [for example, \( (\bar{u}, \bar{v}) \) or \( (k_1, k_2) \)], in which the diffusion processes are studied. As a rule, numerical modeling of non-steady state phenomena is always reduced to discretization of the time variable for a sequence of evolutionary layers, for which the "conservation" of the solution algorithm is of great importance. This solution is the consequence of the necessary agreement between the a priori requirements of the differential problem and the properties of its difference analog. Here, "conservation" is reached when, when the stable algorithms are formulated, the values of \( f \) and \( \partial_y^p (f) \) are taken on the same evolutionary layer, which leads in its turn to implicit difference analogs of the initial equations with respect to the solution and, as a result, to systems of nonlinear equations which may be solved only using iteration approaches.

In conclusion, we should note that since numerical methods of solving multidimensional equations are examined here, whose realization falls within the framework of modern computers, one of the basic requirements imposed on the algorithms is a minimum number of computational operations for a given accuracy of the approximation to the solution.

1. FORMULATION AND BASIC PROPERTIES OF THE PROBLEM.

Let us now formulate the problem. For this purpose we shall consider Euclidean space

\[
\mathbb{R}^2 \times \mathbb{R}^3 \quad \mathbb{R} = (\mathbb{R}_{\mathbb{R}_{\mathbb{R}}}, \mathbb{R}^3 \mathbb{R}_{\mathbb{R}})
\]

Let us now assume that \( Q \subset \mathbb{R}^2 \) is a certain region in \( \mathbb{R}^2 \) and let us set \( Q = Q \times (0, T) \) and \( \partial_r = \partial x (0, T) \) is the lateral surface \( Q_r \). Now let us examine on \( Q_r \) a certain real function \( f \), having the first and second generalized derivative. For each \( t \in [0, T] \), let us introduce the Banach space \( X_e (Q) \) and the Sobolev space \( W^{1} (Q) \), which consists of the functions of the space \( L^2 (Q) \), having in \( Q \) the generalized derivatives summed with the square, \( W^{1} (Q) \) produces Hilbert space with respect to the norm

\[
(\hat{\Pi}^{1} \hat{\Pi}^{2}) = \left\{ \frac{1}{2} \left( \int_{Q} \sum_{i=1}^{2} \sum_{j=1}^{3} \partial_i \partial_j f \, \partial_i \partial_j f \, dQ + \sum_{i=1}^{2} \sum_{j=1}^{3} \partial_i \partial_j f \, \partial_i \partial_j f \, dQ \right) \right\}.
\]

Let us now assume \( f \) belongs to the class of these functions \( F \) in \( Q_r \), which for each \( t \in [0, T] \) produces Hilbert space \( H(Q) \) either with respect to the
norm $W^k_w(Q)$ or with respect to its equivalent.

Under these conditions, let us examine the quasi-linear equation in the domain $Q_T$ with the divergent right side

$$\frac{\partial \rho}{\partial t} = \frac{2}{\alpha\beta} \frac{\partial}{\partial y_1} \frac{\partial}{\partial y_2} \rho,$$

for which $\rho$ is given on the surface $S_T$. Then for the known diffusion coefficients $\lambda_{\beta\beta}$ the equation (2) determines the unique solution. We shall also assume that the operator

$$\hat{\Lambda} = -\frac{2}{\alpha\beta} \frac{\partial}{\partial y_1} \frac{\partial}{\partial y_2},$$

exists in Hilbert space $L^2(Q)$, and the set $\varphi(\hat{\Lambda})$ is its domain of definition, consisting of the functions $\varphi \in L^2(Q)$ such that $\hat{\Lambda} \varphi \in L^2(Q)$, and it is not degenerate, i.e., for any non-zero vector $\xi = (\xi_1, \xi_2)$, the inequality is satisfied

$$\nu^2 \xi^2 \leq \sum_{\beta\beta} \frac{2}{\alpha\beta} \frac{\partial}{\partial y_1} \frac{\partial}{\partial y_2} \varphi(t, y_1, y_2) \xi_1 \xi_2 \mu^2 (t, y_1, y_2) \leq \mu^2 (t, y_1, y_2) \xi^2,$$

where $\nu$, $\mu$ are positive numbers. The condition (3) will be called the condition of strong ellipticity. If

$$D_{\beta\beta} = \lambda_{\beta\beta} (t, y_1, y_2, \dot{\varphi}, \ddot{\varphi}),$$

then the condition of strong ellipticity has the form

$$\nu(\varphi) \xi^2 \leq \sum_{\beta\beta} \frac{2}{\alpha\beta} \frac{\partial}{\partial y_1} \frac{\partial}{\partial y_2} \varphi(t, y_1, y_2, \dot{\varphi}, \ddot{\varphi}) \xi_1 \xi_2 \mu^2 (t, y_1, y_2) \xi^2,$$

where $\nu(\varphi)$, $\mu(\varphi)$ are certain continuous positive functions.

Now to complete the formulation of the problem, we must add additional conditions on equation (2) which determine the dependence of the coefficients $D_{\beta\beta}$ on the solution $f$. For this purpose, let us write the following equation in the domain $Q_T$

$$\hat{\Lambda} \varphi = \lambda_{\beta\beta} (t, \frac{2}{\alpha\beta}, \frac{\partial}{\partial y_1}, \frac{\partial}{\partial y_2}) \xi^2 \mu^2 \xi^2,$$

(5)
where \( \hat{L} \) in the general case is a certain linear operator which acts in Hilbert space \( H \), produced by \( \mathcal{W}^2_\tau(Q) \). We may give a sufficient criterion for ellipticity of the operator \( \hat{L} \) and as a result the parabolic nature of the system of equations (2) - (5). Let us assume \( \hat{L} \) exists and is positive. Then we have the following:

**Lemma 1.** If for \( \forall \tau \in \Omega \) the inequalities \( \beta_{\omega 0}(\theta_0, 0) > 0 \) and \( \beta_{\omega 0}(\theta_0, 0) > 0 \) hold, then the matrix \( D_{\omega \varphi} \) is positive definite.

Actually, it follows from the positive nature of \( \hat{L} \) that \( |L_{\hat{L}}| > |L_{\hat{L}}|^2 \), where \( K \) is the cone for \( Q \). Since according to the conditions of the Lemma

\[
\lambda = \frac{p_0}{2} + \frac{1}{2} p_2 \text{ and } \lambda = \frac{1}{2} \text{ in a similar way,} \]

from which we have \( t_{\alpha} = \frac{1}{2} \theta(0) = \varphi + \varphi \) or in view of \( \rho \leq 0 \), which designates the positive definiteness \( \{D_{\omega \varphi}\} \). Although the properties of the smoothness of the function \( \beta_{\omega 0}(\varphi, \frac{1}{2}, \varphi) \), with respect to which we shall assume that the condition of the Lipschitz continuity is satisfied at the norm \( \mathcal{W}^2_\tau(Q) \), mainly

\[
\|\beta_{\omega 0}(\varphi, \frac{1}{2}, \varphi) - \beta_{\omega 0}(\varphi, \frac{1}{2}, \varphi)\| \leq \frac{1}{2} \|\theta_0\| + \frac{1}{2} \|\theta_2\|, \quad (6)
\]

We should note that equation (5) means that from the functional equation determining the diffusion coefficients, we may distinguish a certain linear part for \( D_{\omega p} \). Thus, the system of equations (2) and (5) under the ellipticity conditions (3) and given on the surface \( S_T \) of the boundary conditions completely determines the solution \( \hat{L}(y, z) \) and \( D_{\omega p}(y, z) \).

Let us turn to a numerical solution of the problem. For this purpose, we write the semi-difference analog of the system of equations (2) - (5), which has the following form in view of the conservation requirements:

\[
\begin{align*}
\quad \hat{\beta}_{\omega 0}^{n+1} & = \hat{\beta}_{\omega 0}^n + \Delta t \frac{1}{2} \left( \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \hat{\beta}_{\omega 0}^n + \frac{1}{2} \hat{\beta}_{\omega 0}^n \hat{\beta}_{\omega 0}^n \\
\quad \hat{D}_{\omega p}^{n+1} & = \hat{D}_{\omega p}^n + \Delta t \hat{D}_{\omega p}^n
\end{align*}
\]

where \( \hat{\beta}_{\omega 0} \) designates the number of the evolutionary layer and the time layer in
the given case, and \( \Delta t \) — the discretization step of the variable \( \frac{v}{\Delta t} \in [0, 1] \).

Let us examine the elliptical operator in the domain \( Q \)

\[
\hat{\mathcal{L}} = \frac{1}{2} \frac{\partial^2}{\partial x^2} \frac{\partial^2}{\partial y^2} + \frac{1}{\Delta t^2},
\]

where \( \mathcal{D}_m^n = \mathcal{D}_m^n \) is determined on \( Q + \Gamma \) for \( \det(\mathcal{D}) > 0 \) for \( \forall y \in Q + \Gamma \).

Since the solution of the system of equations (7) for each value of the evolutionary layer \( n \) is reduced to inversion of the elliptical operators \( \hat{\gamma} \), which may be satisfied in view of the nonlinear nature only using iteration approaches, we shall now examine several properties of the operators which are not omitted (9).

Let us assume \( \mathcal{D}_m \) are continuous at \( Q + \Gamma \) for \( \mathcal{L} = \mathcal{L}_m \), and we shall assume \( \mathcal{D}_m \) is a factor of all the finite functions in Hilbert space \( H \), which produce \( \mathcal{L}_m(a) \), \( \mathcal{D}_m^* = \mathcal{L}^* \) at \( \mathcal{D}_m \). We shall use \( \hat{\mathcal{L}}_m \) to designate the closure of \( \hat{\mathcal{L}}_m \). Then \( \hat{\mathcal{L}}_m \) is called the minimum operator producing \( \hat{\mathcal{L}}_m \). The definition domain \( \hat{\mathcal{L}}_m \) is the space \( \mathcal{W}_m(a) \) at \( \mid \mathcal{L}_m \mid = 0 \) and \( \mid \mathcal{D}_m \mid = 0 \). By the expansion of the operator \( \hat{\mathcal{L}}_m \), we designate the operator \( \hat{\mathcal{L}}_m^* = \hat{\mathcal{L}}_m^* \) at the set \( \mathcal{W}_m(a) \), where \( \mathcal{W}_m(a) \) is the set \( \mathcal{W}_m(a) \) such that \( \mathcal{L}_m(a) \). Then, the solution of the equation

\[
\hat{\mathcal{L}}_m f = \mathcal{P}(y)
\]

is called, as is known, the solution of the first boundary value problem. For it the following estimate holds

\[
C_1 \| f \| \mathcal{W}_m(a) \leq \| f \| \mathcal{L}_m(a) \leq C_2 \| f \| \mathcal{W}_m(a).
\]

2. ITERATION SOLUTION OF A SYSTEM OF NONLINEAR ELLIPTICAL EQUATIONS.

Iteration methods are widely used for inversion of linear elliptical systems of the form (7) - (8). However, here the requirement for them is dictated also by the nonlinearity of the system, and in contrast to the linear case, a study of the convergence of the iteration process is of great importance. For convenience, we shall examine the initial approximation individually
and the iteration continuation of the system (11) - (12) for the solution (7) - (8):

\[ \Delta u^m_{\alpha \beta} = \rho_{\alpha \beta} \left( \frac{\partial^2 u^m_{\alpha \beta}}{\partial x^2} \right) + \alpha \beta \Delta t \]

\[ \rho^m_{\alpha \beta} - \rho^{m-1}_{\alpha \beta} = \frac{\Delta t}{\epsilon} \left( \frac{\partial^2 u^m_{\alpha \beta}}{\partial x^2} \right) + \Delta^t \]

The advantage of distinguishing the initial approximation is due to the fact that for certain functions \( \rho_{\alpha \beta} \), the unbounded operators \( LQV \), and the presence of a large amount of a priori information regarding the solution is sufficient to confine ourselves to a one-step process (11) - (12), if the solution of the equation (11) is found with a high order of accuracy. This approach, which is related to the introduction of non-control grids, will be examined below. In the case of bounded operators \( L \), a system of the predictor-corrector type (13) - (14) is preferable. We shall study its convergence.

Let us assume \( L \) is positive definite in \( H \) and a bounded operator, i.e., \( L^* L \geq \sum_{ij} c_{ij} A_{ij} \geq \sum_{ij} c_{ij} A_{ij} \). Then the following is valid:

**Theorem 1.** The iteration process (13) - (14) converges for each \( n \) to the solution of the system (7) - (8), if the inverse operator \( L^{-1} \) is positive and the inequalities
\[ \sum_{ij} c_{ij} A_{ij} \left( \sum_{ij} c_{ij} A_{ij} \right)^{-1} \leq m_L \]
hold.

In view of the positive nature of \( L \), there is an inverse operator, and \( \| L^{-1} \| \leq \frac{1}{m_L} \), from which the solution of the equation (8) follows.

Let us introduce the notation
\[ \Delta u^m_{\alpha \beta} = \rho^m_{\alpha \beta} - \rho^{m-1}_{\alpha \beta}, \quad \Delta_{\alpha \beta} = \frac{\partial^2 u^m_{\alpha \beta}}{\partial x^2} \]

Then, subtracting equation (13) from equation (7), with allowance for the notation in (15), we obtain
Now let us determine the rigid expansion determined by the operator
\[
\hat{\mathcal{L}}_m u_m = \frac{2}{\sqrt{\rho}} \frac{2}{\sqrt{\rho}} \mathcal{D}_{\mathcal{L}} \frac{2}{\sqrt{\rho}} f(m).
\]  
(16)

Since the matrices \( \mathcal{D}_{\mathcal{L}} \) and \( \mathcal{D}_{\mathcal{L}}^\theta \), in view of the conditions imposed on the function \( \mathcal{L}_{\mathcal{L}} \), have positive minimum eigenvalues, the inequality (10) follows.

We have the following from Lemma 1
\[
C(\Omega) \| p_m \|_{w^1} \leq C(\Omega) \| p_m \|_{L^1} \leq C(\Omega) \| p_m \|_{w^2}
\]
and (13) \( C(\Omega) \| p_m \|_{w^2} = \| p_m \|_{w^2} \). Consequently \( \| p_m \|_{w^2} \) is bounded.

Let us also assume \( \| p_m \|_{w^2} < \infty \). Determining (16) \( \| \hat{\mathcal{L}}_m u_m \|_{L^1} = \| \hat{\mathcal{L}}_m u_m \|_{L^2} \), we obtain
\[
\| \hat{\mathcal{L}}_m u_m \|_{L^1} \leq 2 \| \hat{\mathcal{L}}_m u_m \|_{w^2} \left\{ \| \Delta_{\mathcal{L}} \|_{w^1} + \| \Delta_{\mathcal{L}}^\theta \|_{w^1} + \| \Delta_{\mathcal{L}}^m \|_{w^1} \right\}.
\]

In its turn, from (17) we have
\[
C(\Omega) \| u_m \|_{w^1} \leq C(\Omega) \| u_m \|_{w^2} \left\{ \| \Delta_{\mathcal{L}} \|_{w^1} \right\},
\]
Now, subtracting equation (8) from (14) and applying the condition of the Lipschitz continuity (6) for the function \( \mathcal{L}_{\mathcal{L}} \), we find
\[
\| \Delta_{\mathcal{L}} \| = \mathcal{L}_{\mathcal{L}} (f(m), \frac{\partial \mathcal{L}}{\partial \theta}, \mathcal{L}^m) - \mathcal{L}_{\mathcal{L}} (f(m), \frac{\partial \mathcal{L}}{\partial \theta}, \mathcal{L}^m)
\]
\[
\| \Delta_{\mathcal{L}} \| \leq \| \Delta_{\mathcal{L}} \| \left\{ \sum_{\alpha, \beta} \mathcal{A}_{\mathcal{L}}^\alpha \| u_m \|_{L^1} + \sum_{\alpha, \beta} \mathcal{A}_{\mathcal{L}}^\alpha \| \Delta_{\mathcal{L}}^m \|_{L^1} \right\},
\]
from which, by summing the inequalities (18) with respect to \( \alpha \) and \( \beta \), we obtain the conditions of the theorem, which completes the proof.

A necessary condition for the implementation of the iteration process (13) - (14) is, as is known, its stability with respect to small perturbations. Since the stability is due to the correct nature of the difference problem, we shall examine a certain general approach to the control of the system (13) - (14) for an arbitrary positive-definite and bounded operator \( \mathcal{L} \), and then its specific modifications as applied to the integral operator \( \mathcal{L} \) and the differential (unbounded) operator \( \mathcal{L} \).
3. GENERAL (CONSERVATIVE) ALGORITHM OF THE ITERATION CONTROL.

Let us consider the numerical implementation of the process (13) - (14), when \( L \) is a positive definite operator. For this purpose, we use the following control iteration process. We introduce a certain positive self-conjugate operator \( K \) in Hilbert space \( H \), in which \( L \) exists. Let us assume \( K : H \to H \), \( M_k \to (k, x, x) = m_k \); \( M_k < \infty \).

Let us set \( M_A = \inf \{ m_k \} \). Let us examine the following iteration process [1]

\[
L^m D_{x+}^{m} + K D_{x+}^{m} = \sum_{n=0}^{m} e_{x+}^{n} + L^m D_{x+}^{m},
\]

(19)

\[
D_{x+}^{m} = D_{x+}^{m+1}.
\]

(20)

If \( L^m D_{x+}^{m} \) belongs to the region of values \( L^h \), then the following theorem is valid [1].

Theorem 2. For \( D_{x+}^{m} \in H \), the sequence \( \{ \tilde{D}_{x+}^{m} \} \) from (19) - (20) converges to the solution of the equation (14).

To prove this statement, instead of the space examined \( H \), i.e., \( L_t(Q) \) with \( (X, Y) = \int x(t) y(t) d\sigma \), a certain new space \( H_t \) is introduced for \( L_t(Q) \); namely, \( (X, Y) = (L_t^{-1} x, y) \). It may be readily shown that the operator \( (L_t^{-1} + Q)^{-1} \) is self-conjugate in \( H_t \). Let us set

\[
C = (L_t^{-1} + Q)^{-1} K
\]

and then

\[
0 \leq (C X, X) \leq (X, X).
\]

(21)

We should note that the norms produced by the scalar products \( (X, Y)_t^{-1} \) and \( (X, Y)_t \) are equivalent. It follows from (12) that \( \rho (C) \leq \rho \). The proof of Theorem 2 follows from this fact and the Krasnosel'skii theorem.

However, Theorem 2 does not make it possible to apply the process (19) - (20) to the solution of the system (13) - (14). In its turn, the possibility of a numerical approximation of the process (19) - (20) establishes Theorem 2 [1]. Let us examine the process

\[
(\lambda^m + \Delta L^m) D_{x+}^{m} + K D_{x+}^{m} = \sum_{n=0}^{m} e_{x+}^{n} + L^m D_{x+}^{m} + \Delta D_{x+}^{m},
\]

(22)
i.e., we assume the operator \( \mathcal{L}^\beta \mathcal{A} \) is known with a certain non-self-conjugate error \( \Delta \mathcal{L} \), and we assume the same hold for \( \mathcal{L}^\beta \mathcal{A} \). We shall assume that 
\[
\| \mathcal{L}^\beta \mathcal{A} \| < N, \quad \| \mathcal{L}^\beta \mathcal{A} \| < N, \quad \| (\mathcal{L}^\beta \mathcal{A} + K) \Delta \mathcal{L} \| < \frac{1}{N}
\]
and the equation is satisfied.

Theorem 3. Let \( \mathcal{D}_{\beta}^{m} \) be the solution of equation (8). Then the following estimate holds:

1. If \( \| \mathcal{D}_{\beta}^{m} \| < N \), then
\[
\| \mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0} \| \leq \frac{1}{N} \left[ \frac{M N}{m_k \lambda} + \frac{N}{m_k} + \frac{1}{N} + O(\delta) \right],
\]
where \( \lambda \leq \lambda_m \sim \lambda_n \), \( \lambda_m \sim \lambda_n \rightarrow \infty \).

2. If \( \lambda = 1 \), then
\[
\| \mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0} \| \leq \lambda_m (\delta) + \frac{1}{m_k} \left[ \frac{M N}{m_k \lambda} + \frac{N}{m_k} + \frac{1}{N} + O(\delta) \right].
\]

The proof of Theorem 3 follows from the proof of the preceding estimate. Let us consider the convergence rate of the iteration processes (11) - (14).

Here, the following Theorem 3 is valid [2]. If \( \mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0} \) belongs to the region of values for the operator \( \mathcal{L}^\beta \mathcal{A} \) and \( \mathcal{L}^\beta \mathcal{A} \), then
\[
\| \mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0} \| < \frac{1}{\lambda}
\]

The proof of the theorem is based on the following
\[
\mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0} = [\mathcal{L}^\beta \mathcal{A} + K]^{-1} (\mathcal{D}_{\beta}^{m} - \mathcal{D}_{\beta}^{m,0}).
\]

Theorem 3 may be refined for finite-dimensional space, namely, for finite-dimensional space as applied to the process (12). It is known that \( \mathcal{L} \) are two matrices \( \mathcal{Z} \) and \( \mathcal{Z}^{\mathcal{A}} \), so that
\[
\mathcal{Z} \mathcal{L} \mathcal{Z} = \text{diag} (\lambda_1, ..., \lambda_\beta) \]
\[
\mathcal{Z} (\mathcal{L} \mathcal{L}^\beta) \mathcal{Z} = \text{diag} (\lambda_1^2, ..., \lambda_\beta^2).\]

Here we may assume that the following estimate is satisfied [1].
Then
\[ \| \mathbf{Z}_i \| = \| \mathbf{Z}_j \| = \mathbf{M}_k \delta_i \quad i = 1, \ldots, p. \]

Then
\[ \| \mathbf{Z}_i \| = \| \mathbf{Z}_j \| = \mathbf{M}_k \delta_i \quad i = 1, \ldots, p. \]

where
\[ \psi = \frac{1}{1 + \lambda^{\text{min}}} < 1. \]

Let us refine the form of the matrices \( \mathbf{Z} \) and \( \mathbf{Z}_i \). \( \mathbf{Z} \) is the matrix whose columns are \( K \)-normed vectors \( \mathbf{Z}_i = K^T \mathbf{U}_i \), where \( \mathbf{U}_i \) are the eigenvectors of the symmetric matrix \( K \mathbf{L}_{\lambda} K^T \); \( p \) — dimensionality of the matrix; \( K \) — the quantity \( \lambda = 0. \)

Thus, the solution of the system (7) — (8) may be reduced to two iteration processes: the process (13) — (14) and the process (19) — (20). Both of these processes cannot be implemented in practice, since the two-dimensional nature of the parabolic equation makes the time for the solution of (7) — (8) beyond the limits of present day computers. This article proves the convergence of both processes to one, namely, the following processes examined

\[
\begin{align*}
\delta^n - z^n &= \delta^n - z^n \quad (23) \\
\Delta \mathbf{D}_{\alpha}^n + K \mathbf{D}_{\alpha}^n &= \Delta \mathbf{D}_{\alpha}^n + K \mathbf{D}_{\alpha}^n \\
\delta = K \mathbf{D}_{\alpha}^n \quad (24) \\
\end{align*}
\]


Let us consider the process (23) — (25), where the operator \( K \) is selected as follows:

1. \( (Kx, x) \geq m(x, x), M_k(x, x) \geq (Kx, x), \)

2. \( K^{-1} \) acts in the cone, i.e., it is positive.
Let us examine the new operator $iK$, where $\varepsilon \supset \forall$ is a given number. We shall select $\varepsilon$ so that the conditions for the convergence of the process (23) - (25) are satisfied with respect to $i(L^2+k)^{-2}$, i.e., this operator changes a cone into a cone. In addition, $K \supset \hat{\varepsilon} \hat{K}$. Under these conditions in space $H$, which produces $W_t(x)$, the following is valid:

**Theorem 4.** In the space $H$ examined, the process (23) - (25) converges if

$$\alpha > \beta > 0 / \alpha, \beta \in \mathbb{R}, \beta \in (0, 1),$$

$$\sum_{n=0}^{\infty} \frac{\alpha^n}{(\beta^n)^2} < 1.$$ 

Let us subtract equation (3) from equation (12)

$$\delta_{\alpha, \beta} = (L^2+k)^{-2} \delta_{\alpha, \beta} = \frac{\alpha}{\beta} \frac{\alpha^n}{(\beta^n)^2}.$$ 

It may be readily shown that $\delta_{\alpha, \beta}$ satisfies the ellipticity condition

$$\sum_{n=0}^{\infty} \frac{\alpha^n}{(\beta^n)^2} < 1.$$ 

Let us set

$$\delta = (L^2+k)^{-2} \delta_{\alpha, \beta},$$

and

$$(L^2+k)^{-2} \delta_{\alpha, \beta} = (L^2+k)^{-2} \delta_{\alpha, \beta} > 0.$$ 

These equations are strictly larger than zero on $\delta_{\alpha, \beta}$. A similar statement is valid for $\forall m$, which is readily proven by induction. Actually $\delta_{\alpha, \beta} (L^2+k)^{-2} \delta_{\alpha, \beta} > 0$,

$$\delta_{\alpha, \beta} > 0,$$

but

$$\delta_{\alpha, \beta} = \delta_{\alpha, \beta}, \delta_{\alpha, \beta} = \delta_{\alpha, \beta}, \delta_{\alpha, \beta} = \delta_{\alpha, \beta},$$

since for the first iteration indicated above, the induction is valid. At the same time we should note that the minimum eigenvalue of the matrix

$$\lambda_{\min} > \delta > 0 \forall n \lambda = \frac{\beta_2 (\beta_2 - 1)}{2 \beta_2 \beta_2} > 0.$$ 

Let us introduce the expansion of $\delta_{\alpha, \beta}$ for the operator
for which, as was shown above, the following estimate is valid (10)
\[ c_f(p, q, p^c) \| \phi \|_{W^p_{L}} \leq \| \phi \|_{L^p} \leq C \| \phi \|_{N^p_{L}}. \]

In addition, from this equation, taking into account (26), we obtain
\[ \| L_{p}^{n} \left( \frac{\hat{p} - p}{\hat{q} - q} \right) \|_{L^{m} (a)} \leq \frac{2}{q - q} \left( D_{p}^{m} - D_{p}^{n} \right) \| p \|_{L^{m} (a)} \cdot \| \phi \|_{L^{m} (a)}. \]

Now, taking the fact into account that the Sobolev spaces \( \mathcal{W}^d \) or \( d \leq \mathcal{N} \) form a Hilbert scale of space, we have the inequality
\[ \| L_{p}^{m} \left( \frac{\hat{p} - p}{\hat{q} - q} \right) \|_{L^{m} (a)} \leq \frac{2}{q - q} \left( D_{p}^{m} - D_{p}^{n} \right) \| p \|_{L^{m} (a)} \cdot \| \phi \|_{L^{m} (a)} \cdot \left\{ \sum_{\alpha \beta} \| L_{p}^{m} - D_{p}^{m} \|_{\mathcal{W}^p_{L}} \right\}. \]

Let us subtract (4) from (13). We thus have
\[ (L_{p}^{n} + \varepsilon K)(D_{p}^{m} - D_{p}^{n}) = k \left( \frac{\hat{p} - p}{\hat{q} - q} \right) + L_{p}^{n} \left( \frac{\hat{p} - p}{\hat{q} - q} \right) \left( D_{p}^{m} - D_{p}^{n} \right) \| p \|_{L^{m} (a)} \cdot \| \phi \|_{L^{m} (a)} \cdot \left\{ \sum_{\alpha \beta} \| L_{p}^{m} - D_{p}^{m} \|_{\mathcal{W}^p_{L}} \right\}. \]

Let us determine the norm (27), from which we use the Lipschitz continuity condition
\[ \| \Delta_{p}^{m} \| \leq \varepsilon \| \Delta_{p}^{m} \| + \frac{A_{p}^{m} \| \phi \|_{L^{m} (a)} \| \phi \|_{L^{m} (a)} \| \Delta_{p}^{m} \| \}, \]
where we set \( \Delta_{p}^{m} = \hat{p} - p \).

Finally, combining the inequalities obtained, we have
\[ \sum_{\alpha \beta} \| \Delta_{p}^{m} \| \leq \sum_{\alpha \beta} \| \Delta_{p}^{m} \| \left\{ \| \frac{\hat{p} - p}{\hat{q} - q} \|_{L^{m} (a)} \| \phi \|_{L^{m} (a)} \| \Delta_{p}^{m} \| \right\}, \]
which must be proven.

We should note that everywhere for \( \mathcal{W}^p_{L} \), the following norm is selected:
\[ \| x \|_{\mathcal{W}^p_{L}} = \| x \|_{L^{p}} + \| D^x \|_{L^{p}} + \ldots + \| D^x \|_{L^{p}}. \]

For this norm, all of the computations examined above may be satisfied most simply.

The proven theorem establishes the convergence of the iteration process in Sobolev space. However, certain estimates, for the purpose of distinguishing the control
process from the non-control process, may be obtained in the metric \( C \), which will be given below.

5. **Stable Iteration Algorithm for Solving the System** \((13)-(14)\), **where** \( L \) **is the integral operator of the first kind.**

Let us examine the system \((13)-(14)\) for a specific form of the operator \( L \), namely:

\[
\omega \beta = \sum_{\alpha} P_\alpha (x_\alpha, y_\alpha, I) D_\beta (y_\beta),
\]

where \( x_\alpha \in Q \) \( \cap \) \( P_\alpha (x_\alpha, y_\alpha, I) \) is a certain positive kernel. Then the system \((13)-(14)\) assumes the form

\[
\begin{align*}
\frac{P_\alpha}{I_{m+1}} &= \frac{2}{I_{m+1}} \frac{2}{I_{m+1}} \frac{2}{I_{m+1}}, \\
\sum_{\alpha} P_\alpha (x_\alpha, y_\alpha, I) D_\beta (y_\beta) &= \beta_{\alpha \beta} (x_\alpha, y_\alpha, I).
\end{align*}
\]

where

\[
\begin{align*}
P_{\alpha \beta} (x_\alpha, y_\alpha, I) &= P_{\alpha \beta} (x_\alpha, y_\alpha, I); \\
\beta_{\alpha \beta} (x_\alpha, y_\alpha, I) &= \beta_{\alpha \beta} (x_\alpha, y_\alpha, I).
\end{align*}
\]

The presence of the Fredholm operator of the first kind does not make it possible to immediately solve the equations \((13)-(14)\). Therefore, we shall use the algorithm \((23)-(25)\). Let us reduce \((29)\) to a system of integral relationships, namely, let us introduce in the domain \( Q \) the difference nodes and we shall designate them by letters in the text. Then for \( \bar{x}_i : (i \leq j \leq l) \), \( \bar{x}_1, \bar{x}_2, \ldots \), \( \bar{x}_e \in \bar{Q} \), we obtain the following integral relationships:

\[
\begin{align*}
\sum_{\alpha} P_{\alpha \beta} (x_\alpha, y_\alpha, I) D_\beta (y_\beta) &= \beta_{\alpha \beta} (x_\alpha, y_\alpha, I),
\end{align*}
\]

where

\[
\begin{align*}
P_{\alpha \beta} (x_\alpha, y_\alpha, I) &= P_{\alpha \beta} (x_\alpha, y_\alpha, I); \\
\beta_{\alpha \beta} (x_\alpha, y_\alpha, I) &= \beta_{\alpha \beta} (x_\alpha, y_\alpha, I).
\end{align*}
\]

The study \([2]\) was devoted to solving the system of the form \((30)\). If the system \((30)\) is examined in operator form, then the operator will act from \( L_2 \) (where \( D_m (y) \in L_2 (Q) \) in \( R^e - \bar{L} \) — dimensional Euclidean space, i.e.,

\[
\]
Thus, we may readily see that the operator \( \left( P_{\alpha \beta}^{m-1} \right)^* \) which is conjugate to \( P_{\alpha \beta}^{m-1} \) will act from \( R_{L} \) in \( L_{L}(\Omega) \) and we may show that the operator \( \left( P_{\alpha \beta}^{m-1} \right)^* P_{\alpha \beta}^{m-1} \) is the integral operator with the kernel

\[
\tilde{T}_{\alpha \beta}^{m-1}(x, y') = \sum_{j=1}^{l} P_{\alpha \beta}^{m-1}(x) P_{\alpha \beta}^{m-1}(y').
\]

If we now take \( K = \delta I \) as the operator \( K = \delta K \), then equation (24) may be written in the form

\[
\Phi^{m-1}(x, y^j) = \sum_{j=1}^{l} P_{\alpha \beta}^{m-1}(x) P_{\alpha \beta}^{m-1}(y^j).
\]

In its turn, equation (31) has the following analytical solution:

\[
\tilde{T}_{\alpha \beta}^{m-1}(x) = \tilde{T}_{\alpha \beta}^{m-1}(x) + \sum_{j=1}^{l} P_{\alpha \beta}^{m-1}(x) \frac{\partial}{\partial y_j}.
\]

where \( d = (d_1, \ldots, d_l) \) is found from the inversion of the matrix \( \frac{1}{\delta + \epsilon} \) or

\[
\left( A + \delta \right) \bar{d} = \bar{q}_0 - \bar{d}^{\omega-1}, \quad \bar{A} = (a_{i,j}),
\]

where

\[
\bar{q}_0 = \sum_{j=1}^{l} \tilde{T}_{\alpha \beta}^{m-1}(y) P_{\alpha \beta}^{m-1}(y), \quad \bar{d}_j = \sum_{j=1}^{l} P_{\alpha \beta}^{m-1}(y) \frac{\partial}{\partial y_j}.
\]

Formula (32) has a very simple form. However, the analytical solution may also be obtained for the operator \( K = \delta I \). We should also note that there are methods making it possible to have a system of integral relationships and equation (28). Thus, the integral relationships apparently are the finite form to which the systems of nonlinear parabolic equations are reduced. If we take a certain more complex operator, as compared with that examined here, as the operator \( K = K \), then in this case it is first advantageous to reduce the Fredholm operator to matrix form and then to use the algorithm (24) in the projection on the grid space.
6. CONSERVATIVE SINGLE STEP ALGORITHM (11) - (12), USING A KNOWLEDGE OF THE SYSTEM PREHISTORY.

The iteration algorithm examined below must be effective when \( \mathcal{L} \) is a bounded, poorly defined operator. When \( \mathcal{L} \) is a non-bounded operator in \( \mathcal{L}(\mathbb{R}) \), for example, the differentiation operator with respect to the evolutionary variable, the iteration process is directly used. We shall take its conservation to mean conservation within an accuracy of the system approximation. However, with a knowledge of the system prehistory, which supplies certain a priori information, a single step process is avoided. The quantities \( f \) and \( D \) are retained in the computer memory not only in the n-1 layer, but also at certain previous layers up to n-S inclusively. It is natural to assume that the older is the derivative of the function, the smoother it is. This is the basic a priori assumption here. Then, instead of the process (13) - (14), we may regard the system (11) - (12), where instead of equation (12)

\[
\mathcal{L} \tilde{z}_p^l = \ell_p \left( \frac{\partial^{n+1} z_p}{\partial t^{n+1}}, \frac{\partial^n z_p}{\partial t^n}, \frac{\partial z_p}{\partial t} \right)
\]

we shall solve the equation

\[
\mathcal{L} \tilde{z}_p^l = \ell_p \left( \frac{\partial^{n+1} z_p}{\partial t^{n+1}}, \frac{\partial^n z_p}{\partial t^n}, \frac{\partial z_p}{\partial t} \right)
\]

The derivatives of all the functions used here may be calculated using the following formulas

\[
\frac{\partial \theta}{\partial z^l} = \frac{\partial^{n+1} \theta}{\partial t^{n+1}} \\
\frac{\partial \chi}{\partial z^l} = \frac{\partial^{n+2} \chi}{\partial t^{n+2}} - 2 \frac{\partial^{n+1} \chi}{\partial t^{n+1}} \frac{\partial^{n+1} \chi}{\partial t^{n+1}}
\]

where we set \( \theta = \{ \ell, \ell_p, \ell_p \} \). It is apparent that the better known are the derivatives, the greater the amount of the prehistory for solving the system, which is retained in the computer memory. We should note that if the dependence \( \ell_p \left( \frac{\partial z_p}{\partial t}, \frac{\partial^l z_p}{\partial t^l} \right) \) is strongly nonlinear, namely

\[
\ell_p = \ell_p \left( \frac{\partial z_p}{\partial t}, \frac{\partial^l z_p}{\partial t^l} \right) \theta_p \quad ; \quad \theta_p > 0,
\]

where \( \ell_p \) is a smooth function of \( \frac{\partial z_p}{\partial t} \), in this case all the methods of directly calculating \( \mathcal{L}z_p^l \) for the differential operator \( \mathcal{L} \) give an increasing error, which disturbs the conservation. Let us examine the approach making it possible to retain a solution which is constant with respect to the error even in the case of strong nonlinearity. For this purpose, we turn to conjugate equations for which
Theorem 6. The relative error of the diffusion coefficient is retained when changing to conjugate equations. This control of the basic system differs primarily from the control of that examined above related to the addition of $\xi_k$. The change to conjugate equations for the diffusion coefficient contains two basic approaches, namely, the change to an implicit characteristic and then the change to the conjugate equation. If the operator $\hat{L}_a^k = \frac{d^k}{dt^k}$, then $\hat{L}_a^k \hat{P} : \hat{L}_a^q \approx \frac{d^k}{dt^k}$.

Below, the indices $\alpha, \beta$ are omitted for purposes of simplicity. Then

$$\frac{dD}{dt} = \beta(t) D^2.$$ (33)

Now let us multiply (33) by a certain function $E$ (regulator)

$$\frac{dDE}{dt} - D \frac{dE}{dt} = \beta(t) D^2 E D.$$ (34)

We require that the equation $DE = \text{const} \cdot C$ is satisfied. Then the equation (33) is equivalent to the following equation:

$$\frac{dE}{dt} = \beta(t) \frac{C^2}{E^2} E Ah \frac{E_{in} - E_{in}}{At} \frac{E_{in} - E_{in}}{At} \cdot \frac{C^2}{E_{in}^2} E Ah.$$ (35)

We should note that the diffusion coefficient $D$ in (35) may be taken from the preceding evolutionary layer, since the function $E$ decreases and, consequently, its error does not increase. In turn, to calculate $E$, we must use an implicit scheme. The accuracy of determining it will be $0(\Delta t)$, where $\Delta t < \Delta t$. For this purpose, let us introduce the implicit characteristics (Fig. 1).
In conclusion, we should note that in connection with Fig. 1, the characteristics may be considered according to an implicit scheme, if their directions only depend on \( t_i \) and simultaneously we use the step \( \Delta t_i \) as the step \( \Delta t \).

7. ALGORITHMS FOR SOLVING THE SYSTEM OF EQUATIONS (7) – (8), USING THE APPROXIMATION OF HIGH ACCURACY.

As was shown above, the conservation requirement imposed on numerical algorithms for solving systems of nonlinear parabolic equations, is reduced to complex computational schemes with iteration control. However, in certain cases this requirement may be reduced, if we solve the equation (11) very precisely, for example, using the introduction of non-control difference grids or grids with "frequency increase," if this does not lead to a great increase in the amount of operations. This can be done if when solving equation (11) we use the method of variable directions. Briefly, we shall show how to retain the advantages of the method of variable directions for an increased frequency difference grid (Fig. 2). These variable directions consist of reducing the numerical solution to satisfying the subsequent "trial run."

For simplicity, we shall assume that \( D_{\rho} \) has a diagonal form. Let us examine (11)

\[
\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial x} \left( \frac{\partial}{\partial y} D_{\rho} \frac{\partial \phi}{\partial y} \right).
\]

Its difference analog in finite dimensional space, produced by the grid in the domain \( Q \), will be a linear system which within the framework of the variable direction method may be written by the recursion relationship

\[
\left( \hat{L}_k - \omega_k \right) \hat{\phi}^{n+1} = \hat{Q} - \left( \hat{L}_k - \omega_k \right) \hat{\phi}^n,
\]

\[
\left( \hat{L}_k - \omega_k \right) \hat{\phi}^{n+1} = \hat{Q} - \left( \hat{L}_k - \omega_k \right) \hat{\phi}^n.
\]
Generalizing the method of variable direction to a class of non-regular difference grids was performed in [3]; where it was shown how to separate the main operator $L_k = L_k^1 + L_k^2$ into the operators $L_k^1$ and $L_k^2$ for grids with partial frequency increase in a two-dimensional domain. The method introduced in [3] makes it possible to reduce all the operator inversions to satisfying the subsequent trial runs, with a small increase in the number of equations.

8. SCHEMATIC CLASSIFICATION OF NUMERICAL ALGORITHMS FOR SOLVING SYSTEMS OF NONLINEAR PARABOLIC EQUATIONS,

Now let us briefly give certain practical results of applying the approach given above to solving systems of equations with nonlinear diffusion. Since this problem is very cumbersome, each of the algorithms proposed for the solution may not be used completely, but only partially. For example, we shall assume that the dependence of the diffusion coefficients $D_{p}(y)$ on the solution is known. In this case, we should be able to expand $D_{p}$ in Fourier series. However, another approach is advantageous here, namely -- in accordance with the laws of conservation -- from the parabolic system (2), we separate the retained quantities such as energy, energy flux, etc. Then $D_{p}$ may be subjected to iteration, so that they satisfy the laws of conservation. In conclusion, we would like to give a certain schematic classification of the numerical algorithms, which may be useful when selecting a method for solving the system (2) -- (5).
Fig. 3. Practical diagram for selecting the algorithm for the numerical solution of systems of nonlinear parabolic equations.

In conclusion, we would like to express our appreciation to V. G. Zolotukhin for assistance in the work and valuable comments, and we would also like to thank V. G. Popov for useful discussions.

*Illegible in foreign text.
REFERENCES


CONTENTS

INTRODUCTION  1
1. FORMULATION AND BASIC PROPERTIES OF THE PROBLEM  2
2. ITERATION SOLUTION OF A SYSTEM OF NONLINEAR ELIPTICAL EQUATIONS  5
3. GENERAL (CONSERVATIVE) ALGORITHM OF THE ITERATION CONTROL  8
4. CERTAIN SUFFICIENT CONDITIONS FOR THE CONVERGENCE OF THE CONTROL PROCESS (23) - (25)  10
5. STABLE ITERATION ALGORITHMS FOR SOLVING THE SYSTEM (13) - (14)
   WHERE L IS THE INTEGRAL OPERATOR OF THE FIRST KIND  13
6. CONSERVATIVE SINGLE STEP ALGORITHM (11) - (12), USING A KNOWLEDGE
   OF THE SYSTEM PREHISTORY  15
7. ALGORITHMS FOR SOLVING THE SYSTEM OF EQUATIONS (7) - (8), USING
   THE APPROXIMATION OF HIGH ACCURACY  17
8. SCHEMATIC CLASSIFICATION OF NUMERICAL ALGORITHMS FOR SOLVING
   SYSTEMS OF NONLINEAR PARABOLIC EQUATIONS  18