Splitting and Alternating Direction Methods

G.I. Marchuk

Department of Numerical Mathematics
USSR Academy of Sciences
Ryleev Street 29
119034 Moscow, USSR

HANDBOOK OF NUMERICAL ANALYSIS, VOL. I
Finite Difference Methods (Part 1)—Solution of Equations in $\mathbb{R}^n$ (Part 1)
Edited by P.G. Ciarlet and J.L. Lions
© 1990, Elsevier Science Publishers B.V. (North-Holland)

Contents

PREFACE 203

CHAPTER I. Introduction 205

1. Approximation 206
2. Stability 216
3. Convergence 222
4. The Crank–Nicolson scheme 224

PART 1. ALGORITHMS FOR THE SPLITTING METHODS AND THE ALTERNATING DIRECTION METHODS 229

CHAPTER II. Componentwise Splitting (Fractional Steps) Methods 231

5. The splitting method based on implicit schemes of first-order accuracy 231
6. The componentwise splitting method based on Crank–Nicolson schemes: The case $A = A_1 + A_2$ 232
7. Multicomponent splitting of problems based on Crank–Nicolson schemes 234
8. A general approach to componentwise splitting based on elementary Crank–Nicolson schemes 235
9. A general formulation for the splitting method based on multilevel schemes 236
10. The two-level splitting scheme with weight coefficients 237
11. The splitting method for systems which do not belong to the Cauchy–Kovalevskaya class 238
12. Splitting schemes for the heat conduction equation: Local one dimensional schemes 239
    - 12.1. The splitting scheme for the heat conduction equation in an orthogonal coordinate system 239
    - 12.2. The splitting scheme for the heat conduction equation in an arbitrary coordinate system 240
    - 12.3. Local one-dimensional schemes 242

CHAPTER III. Two-Cycle Componentwise Splitting Methods 245

13. The two-cycle componentwise splitting methods: The case $A = A_1 + A_2$ 245
14. The two-cycle multicomponent splitting method 247
15. The two-cycle componentwise splitting method for quasi-linear problems 248
16. A general approach to the two-cycle componentwise splitting method 249
17. The two-cycle componentwise splitting scheme for the heat conduction equation 252

CHAPTER IV. Splitting Schemes with Factorization of the Operators 255

18. Schemes factorizing the operators 255
19. The implicit splitting scheme with approximate factorization of the operator 258
20. The stabilization method (explicit–implicit scheme with approximate factorization of the operator) 259

21. A general scheme for the method of approximate factorization of the operator 263
22. The scheme of approximate factorization for the parabolic equation 265

CHAPTER V. The Predictor-Corrector Method 269

23. The predictor-corrector method: The case $A=A_1+A_2$ 269
24. The predictor-corrector method: The case $A=\Sigma_{\alpha=1}^{n} A_\alpha$ 272
25. The predictor-corrector method for the parabolic equation 273

CHAPTER VI. The Alternating Direction and the Stabilizing Correction Methods 277

26. The alternating direction method 277
27. The stabilizing correction method 278
28. A general formulation for the stabilizing correction method 279
29. Application of the alternating direction scheme to the parabolic equation 280

CHAPTER VII. Methods of Splitting with Respect to Physical Processes 283

30. The method of splitting with respect to physical processes 283
31. The method of particles in a cell 285
32. The method of large particles 286

CHAPTER VIII. The Alternating Triangular Method and the Alternating Operator Method 289

33. The alternating triangular method 289
34. The alternating operator method 291
35. The generalized alternating operator method 292
36. The scheme of the alternating triangular method for the parabolic equation 292

CHAPTER IX. Splitting Methods and Alternating Direction Methods as Iterative Methods for Stationary Problems 295

37. The stationing method: General concepts of the theory of iterative methods 295
38. Iterative algorithms 297
39. Acceleration of the convergence of iterative methods 299

PART 2. METHODS FOR STUDYING THE CONVERGENCE OF SPLITTING AND ALTERNATING DIRECTION SCHEMES 301

CHAPTER X. Convergence Studies of the Splitting Schemes by Use of the Fourier Method (Spectral Method) 303

40. General statement of the Fourier method 303
41. The Fourier method and the convergence studies of splitting schemes for stationary problems 307
42. The Fourier method and the grounding of the splitting schemes for nonstationary problems 310

CHAPTER XI. The A Priori Estimates Method and the Convergence Studies of the Splitting Schemes 315

43. The simplest a priori estimates 315
44. A priori estimates for the splitting scheme of type $A_j\varphi^{j+1}=B_j\varphi^j+\tau_j f^j$ 319
45. The energy inequalities method for constructing a priori estimates 322

CHAPTER XII. The Splitting of the Evolutionary Problem for a System of Differential Equations 327

46. The splitting of problems defined on fractional intervals and the weak approximation method 327

47. The splitting of problems defined on the whole interval 331
48. The two-cycle splitting of the problem 332
49. Some results on convergence and stability 335

CHAPTER XIII. Convergence Studies and Optimization of Iterative Methods 339

50. Sufficient conditions for convergence 339
51. The choice of parameters in the commutative alternating direction method 344
52. The choice of parameters in the noncommutative alternating direction method 348
53. The convergence acceleration procedures for the alternating direction method 350
54. Generalizations 352

CHAPTER XIV. Splitting and Decomposition Methods for Variational Problems 355

55. Splitting and decomposition methods for classical variational problems 355
56. Decomposition of a general variational problem 356
57. A variational problem with restrictions 357
58. The convergence of decomposition algorithms 358

PART 3. APPLICATIONS OF SPLITTING METHODS TO PROBLEMS OF MATHEMATICAL PHYSICS 361

CHAPTER XV. The Heat Conduction Equation 363

59. The two-cycle componentwise splitting scheme for a parabolic equation with three spatial variables 363
60. Schemes of second-order accuracy for p-dimensional parabolic equations without mixed derivatives 366
61. Schemes for equations with mixed derivatives 368
62. Alternating direction schemes 370
63. Schemes of increased order of accuracy 371
64. Finite element method schemes and splitting schemes for two-dimensional parabolic equations 374

CHAPTER XVI. Equations of Hyperbolic Type 377

65. The stabilization scheme for the multidimensional equation of oscillations 377
66. Approximate factorization schemes for equations of oscillations 378
67. Local one-dimensional schemes for multidimensional hyperbolic equations 381
68. The splitting scheme for multidimensional hyperbolic systems of equations 383

CHAPTER XVII. Integro-Differential Transport Equations 389

69. Statement of the problem and the scheme of incomplete splitting 389
70. The scheme of complete splitting 392
71. The scheme of approximate factorization of the operator 392
72. The method of integral identities and the splitting method 394
73. The numerical scheme for the nonstationary transport equation in (x, y) geometry 399
    73.1. Approximation in spatial variables 400
    73.2. Approximation in the time variable 405
    73.3. Approximation in angular variables 406
    73.4. Numerical realization of the algorithm 407
74. Splitting methods as iterative algorithms for stationary transport equations 410

CHAPTER XVIII. The Splitting Method for Problems of Hydrodynamics 413

75. Splitting schemes for Navier–Stokes equations with ε-perturbation of incompressible fluid equations 413

76. Splitting schemes restoring divergence for incompressible fluid equations 416
77. The general principle for constructing splitting schemes for Navier–Stokes equations 419

CHAPTER XIX. Problems in Meteorology 427

78. Equations of atmosphere hydrothermodynamics 427
79. The general splitting method with respect to physical processes based on the separation of characteristic times 432
80. The approximation of equations in spatial variables and the discrete analogues of conservation laws 435
81. The method of splitting with respect to geometric variables and numerical realization 437

CHAPTER XX. Problems in Oceanology 441

82. Statement of the problem and the splitting of equations with respect to physical processes 441
83. The splitting of adaptation equations in planes and generalization for the nonhydrostatic case 444
84. The splitting of adaptation equations with respect to "topography" 445
85. The splitting of "shallow water" equations with respect to coordinates 447

REFERENCES 449

SUBJECT INDEX 461

Preface

We witness the advent of the age of parallel large-capacity computers and processors. Parallel computers enter our life as the realization of objective processes in aspiration for creating more and more powerful computers capable of solving most complicated computational and informational problems arising in a society. Since on each level of technology we approach our limits in computer capacity, this intention to accelerate the process of creating more and more powerful computers stimulates us permanently for transition to computational processes based on parallel and asynchronous data processing. Rapid progress in the electronic element base allows to construct such a multiprocessor computer system on VLSI crystals.

However, the new quality parallel computers require a completely new mathematical base for solving complicated problems on parallel computer systems. What is this base? No doubt such base is the adequate representation of problems in a form accessible to parallel processing inherent to the structure of the computer.

Numerical mathematics has already done much in this direction. First of all there are the methods of splitting complicated problems of mathematical physics into simpler systems realizable on parallel processors.

This fact inspired the author to give a survey of the splitting methods and to demonstrate the results achieved in this direction. The author tried to analyze all basic splitting algorithms, naturally focussing special attention on the Soviet school of numerical mathematics which has achieved important results both in formulating algorithms and in applying them to solutions of complicated problems. As for the rest of the world's experience, it was in our opinion reflected sufficiently while considering the algorithms as well as in the detailed bibliography presented in this work.

In preparing material for this article the author was assisted greatly by V.I. Agoshkov, G.R. Kontarev, V.I. Kuzin, Y.A. Kuznetsov, V.N. Lykossov, V.P. Shutyaev, V.B. Zalesny. The author is very much obliged to all these colleagues.

CHAPTER I

Introduction

The intensive development of the methods for solving linear algebraic equations with three-diagonal and block three-diagonal matrices in the fifties and sixties led to the creation of effective numerical algorithms for solving stationary problems based on the factorization of the difference operator. A special place among the *methods of factorization* (*sweep methods*) belongs to various versions of noniterative methods of factorization developed by KELDYSH [1942], GELFAND and LOKUTSIEVSKY [1962], RUSANOV [1960], GODUNOV [1962], ABRAMOV and ANDREEV [1963] and others.

Experience accumulated in solving one-dimensional problems by methods based on factorization had prepared the basis for the construction of the algorithms for solving more complicated problems in mathematical physics. And the beginning of the sixties was marked by a great contribution into numerical mathematics by the development of such algorithms. This contribution is connected with the names of Douglas, Peaceman and Rachford, who suggested the *alternating direction method* (see DOUGLAS and RACHFORD [1956], PEACEMAN and RACHFORD [1955]). The success of the method was ensured by a simple reduction of the multidimensional problem to a sequence of one-dimensional problems with three-diagonal matrices easily invertible by computers. This method affected significantly the construction of algorithms in various fields of applied mathematics. The theoretical studies of this and related methods are presented in the works by DOUGLAS [1962], DYAKONOV [1961–1967, 1970–1972], SAMARSKII [1961–1967, 1970–1971, 1977], BIRKHOF, VARGA and YOUNG [1962], WACHPRESS [1962], KELLOGG [1964], GUNN [1965], VOROBYOV [1968] and others.

The methods were developed based on homogeneous and inhomogeneous approximations. In the case of inhomogeneous approximation each of the auxiliary (intermediate) problems not necessarily approximates the original problem but in the whole in special norms this approximation takes place. These methods have been named the *splitting methods*; they have been developed in the works of the Soviet mathematicians YANENKO [1966, 1967], DYAKONOV [1963a, 1966], SAMARSKII [1962d, 1963a], SAUL'EV [1960], MARCHUK [1971] and others.

The splitting methods were widely used in problems of various kinds and they stimulated the more general approach to the solution of problems in mathematical physics based on the method of weak approximation developed by YANENKO [1964b, 1967] and SAMARSKII [1963a, 1965b]. It appeared that the splitting method may be understood as the method of weak approximation of the original equation

by a more simple equation. The convergence conditions for the method of weak approximation were formulated in the theorem by YANENKO and DEMIDOV [1966] and in the works by LEBEDEV [1977] and DYAKONOV [1962f, 1971d]. The method of weak approximation has found natural applications in the problems of hydrodynamics, meteorology, oceanology, in the theory of radiation transport, etc. (MARCHUK [1967, 1974a], YANENKO [1967]).

The original scheme of the predictor-corrector type by Lax and Wendroff found wide applications in problems of hydrodynamics, meteorology, oceanology, where the predictor was suggested in the form of an explicit difference scheme. This scheme is conditionally stable, it is easy in realization and has a second-order approximation in all variables. A detailed study of the scheme is presented in the book by RICHTMYER and MORTON [1972].

Various versions of the predictor-corrector method based on implicit difference approximations were proposed by BRAYN [1966], DOUGLAS [1961], SOFRONOV [1965] and MARCHUK and YANENKO [1966]. All these schemes proved to be equivalent in a certain sense and differed only in the realization technique. In the last of these works the implicit splitting scheme with factorized operator is used as a predictor, which has first-order accuracy. For the problems in hydrodynamics the implicit majorant schemes are used as predictors.

Of particular interest is the method of decomposition and decentralization formulated by LIONS and TEMAM [1966] ad also by BENSOUSSAN, LIONS and TEMAM [1975] which also is close to the splitting methods and methods of weak approximation.

In the sixties the method for solving multidimensional problems of mathematical physics was developed intensively which is connected with the name of HARLOW [1967]. This method was named the *method of large particles*. Today it is also considered as a splitting method. In the works by DYACHENKO [1965], BELOTSERKOVSKY and DAVYDOV [1978], and YANENKO, ANUCHINA, PETRENKO and SHOKIN [1971] various modifications of this method are given and schemes of their realization are considered.

In the present work *many of the splitting (fractional steps) and alternating direction methods are considered, some facts from the theory of convergence of these methods are presented and the algorithms for numerically solving a number of concrete applied problems are formulated* based on the applications of the methods discussed further. Main attention in this work is paid to nonstationary problems. In this connection the *splitting and alternating direction methods* are formulated as classes of finite difference algorithms for solving these problems. Nevertheless, in some sections they will be considered as iterative methods for solving stationary equations.

1. Approximation

This section introduces the general concepts of the theory of finite difference methods which will be used in the present work.

Consider a stationary problem of mathematical physics in the operator form:

$$\begin{aligned} &A\varphi = f \quad \text{in } \Omega, \\ &a\varphi = g \quad \text{on } \partial\Omega, \end{aligned} \tag{1.1}$$

where A is a linear operator, $\varphi \in \Phi$, $f \in F$. Here Φ and F are Hilbert spaces, whose elements are defined respectively on $\Omega + \partial\Omega = \bar{\Omega}$ ($\partial\Omega$ is the boundary of the domain Ω) and Ω, a is the linear boundary condition operator, $g \in G$, where G is the Hilbert space of the functions defined on $\partial\Omega$. For the sake of definiteness and simplicity of notation we will consider the functions from Φ, F, and G to be dependent only upon two variables x and y (which may be regarded as spatial).

Construct the finite-dimensional approximation of the problem (1.1) for example by the finite difference method. To this end consider a set of points (x_k, y_l), where k and l are arbitrary integers. The set of these points will be called a *grid* and the points will be called *grid points*. The distance between grid points will be described by the value $h > 0$ called the *step of the grid* (naturally, this distance may also be described by two parameters h_x and h_y—the grid steps in the x- and y-direction respectively). Denote by Ω_h the set of the grid points approximating in some sense the set of the points of the domain Ω, and by $\partial\Omega_h$ the set of the grid points approximating the boundary $\partial\Omega$. Further, the functions whose definition domain is a grid will be called *grid functions*. The set of the grid functions φ^h with definition domain Ω_h will be denoted by Φ_h. Each function $\varphi \in \Phi$ may be connected with the grid function $(\varphi)_h$ as follows: the value of $(\varphi)_h$ at the grid point (x_k, y_l) equals $\varphi(x_k, y_l)$ (of course, if the values $\{\varphi(x_k, y_l)\}$ have a meaning). This correspondence presents the linear operator acting from Φ to Φ_h; this operator is called a *projection* of the function φ on the grid. The function $\psi \equiv A\varphi$ can also be projected on the grid by taking $(\psi)_h = (A\varphi)_h$. The correspondence between $(\varphi)_h$ and $(A\varphi)_h$ is a linear operator defined on the grid functions.

By now applying approaches developed in the theory of the finite difference methods we construct the problem in the finite-dimensional grid function space:

$$\begin{aligned} &A^h\varphi^h = f^h \quad \text{in } \Omega^h, \\ &a^h\varphi^h = g^h \quad \text{on } \partial\Omega^h, \end{aligned} \tag{1.2}$$

which is the finite difference analogue of problem (1.1). Here A^h and a^h are linear operators depending on the grid step h, $\varphi^h \in \Phi_h$, $f^h \in F_h$, $g^h \in G_h$, and Φ_h, F_h and G_h are grid function spaces.

We introduce in Φ_h, F_h and G_h respectively the norms $\|\cdot\|_{\Phi_h}$, $\|\cdot\|_{F_h}$ and $\|\cdot\|_{G_h}$. Let $(\cdot)_h$ denote the linear operator which puts the element $\varphi \in \Phi$ in correspondence with the element $(\varphi)_h \in \Phi_h$ in such a way that

$$\lim_{h \to 0} \|(\varphi)_h\|_{\Phi_h} = \|\varphi\|_{\Phi}.$$

We shall say that the problem (1.2) *approximates* the problem (1.1) with *order* n on the solution φ, if there exist positive constants $\bar{h}$, M_1, M_2 such that for all $h < \bar{h}$ the

following inequalities hold

$$\|A^h(\varphi)_h - f^h\|_{F_h} \leqslant M_1 \cdot h^{n_1}, \qquad \|a^h(\varphi)_h - g^h\|_{G_h} \leqslant M_2 \cdot h^{n_2}, \tag{1.3}$$

and $n = \min(n_1, n_2)$.

In cases where the solution φ of problem (1.1) is sufficiently smooth it is convenient to find the order of approximation using the norm which is natural for the space of continuous and differentiable functions. To that end the expansion into Taylor series of the solution and other functions contained in the problem is usually applied.

Further we shall suppose that the reduction of problem (1.1) to problem (1.2) has been accomplished and, moreover, that the boundary condition of (1.2) has been used for excluding the values of the solution at the boundary points of the domain $\Omega_h + \partial\Omega_h$. As a result we have the equivalent problem

$$\tilde{A}^h \tilde{\varphi}^h = \tilde{f}^h. \tag{1.4}$$

The values of the solution at the boundary points can be found from (1.2) after solving (1.4). In some cases it is more convenient to write the approximating problem in the form (1.4) and in some other cases in the form (1.2). Thus, as a result of the reduction and with the required approximation taken into account, problem (1.1) with a continuous argument has been reduced to the problem of linear algebra (1.4), i.e., of solving the system of algebraic equations.

Further we shall mainly use Hilbert spaces of grid functions and, unless otherwise stated, the corresponding norm $\|\varphi^h\|_\Phi$ will be defined by the relationship

$$\|\varphi^h\|_\Phi = (\varphi^h, \varphi^h)_\Phi^{1/2}.$$

It must be noted, however, that many of the introduced concepts (approximation, etc.) also apply to the case of Banach spaces and in some statements and illustrating examples grid function norms will be introduced which are not connected with the inner product defined by the above relationship.

We shall illustrate the above concepts with an example of the following concrete problem

$$\begin{aligned} -\Delta\varphi &= f \quad \text{in } \Omega, \\ \varphi &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{1.5}$$

where

$$\Omega = \{(x, y): 0 < x < 1,\ 0 < y < 1\},$$

$f = f(x, y)$ is a smooth function,

$$\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2.$$

Let F be the Hilbert space of real functions $L_2(\Omega)$ with the inner product

$$(u, v) = \int_\Omega uv \, dx \, dy$$

and the norm

$$\|u\| = (u, u)^{1/2}.$$

Denote by Φ the set of functions that are continuous in $\bar{\Omega} = \Omega + \partial\Omega$ and that have continuous first and second derivatives in Ω. We take the same norm in Φ as in F, i.e. $\|\cdot\|_\Phi \equiv \|\cdot\|$. Choose as G the Hilbert space $L_2(\partial\Omega)$ of functions defined on $\partial\Omega$ with the norm

$$\|g\|_{L_2(\partial\Omega)} = \left(\int_{\partial\Omega} |g|^2 \, d\Gamma \right)^{1/2}.$$

If we introduce the operators $A\varphi = -\Delta\varphi$ and $a\varphi = \varphi|_{\partial\Omega}$, then problem (1.5) may be represented in the form (1.1) for $g \equiv 0$.

Now, we introduce the finite-dimensional approximation of problem (1.5). To this end, we cover the square $\bar{\Omega} = \Omega + \partial\Omega$ by a uniform grid with step h in the x- and y-direction. The grid points of the domain will be identified by a pair of indices (k, l), where the first index k $(0 \leqslant k \leqslant N_x)$ corresponds to discretization of the x-coordinate and the second index l $(0 \leqslant l \leqslant N_y)$ corresponds to that of the y-coordinate. Consider the following approximations:

$$\frac{\partial^2\varphi}{\partial x^2} \equiv \varphi_{xx} \to \Delta_x \nabla_x(\varphi)_h, \qquad \frac{\partial^2\varphi}{\partial y^2} \equiv \varphi_{yy} \to \Delta_y \nabla_y(\varphi)_h,$$

where Δ_x, Δ_y, ∇_x and ∇_y are difference operators defined on grid functions φ^h with components $\varphi^h_{k,l}$ as follows:

$$(\Delta_x \varphi^h)_{k,l} = \frac{1}{h}(\varphi^h_{k+1,l} - \varphi^h_{k,l}),$$

$$(\nabla_x \varphi^h)_{k,l} = \frac{1}{h}(\varphi^h_{k,l} - \varphi^h_{k-1,l}),$$

$$(\Delta_y \varphi^h)_{k,l} = \frac{1}{h}(\varphi^h_{k,l+1} - \varphi^h_{k,l}),$$

$$(\nabla_y \varphi^h)_{k,l} = \frac{1}{h}(\varphi^h_{k,l} - \varphi^h_{k,l-1}).$$

Then problem (1.5) can be approximated by the following problem:

$$\begin{aligned} &A^h\varphi^h \equiv -[\Delta_x \nabla_x \varphi^h + \Delta_y \nabla_y \varphi^h] = f^h \quad \text{in } \Omega_h, \\ &\varphi^h = 0 \quad \text{on } \partial\Omega_h, \end{aligned} \tag{1.6}$$

where $\partial\Omega_h$ is the set of grid points that belong to the boundary $\partial\Omega$, and Ω_h is the set of internal grid points in Ω. Here the operator A^h is given by

$$A^h = A_x + A_y,$$

$$A_x = -\Delta_x \nabla_x, \qquad A_y = -\Delta_y \nabla_y,$$

i.e., A^h is a difference analogue of the operator $-\Delta$: $A^h \equiv -\Delta^h$. Now (1.6) can be rewritten in the form

$$\begin{aligned} -\Delta^h \varphi^h &= f^h \quad \text{in } \Omega_h, \\ \varphi^h &= 0 \qquad \text{on } \partial\Omega_h, \end{aligned} \tag{1.7}$$

where φ^h and f^h are vectors with the components $\varphi^h_{k,l}$ and $f^h_{k,l}$ and

$$(\Delta^h \varphi^h)_{k,l} = \frac{1}{h^2}(\varphi^h_{k+1,l} + \varphi^h_{k-1,l} + \varphi^h_{k,l+1} + \varphi^h_{k,l-1} - 4\varphi^h_{k,l}),$$

$$f^h_{k,l} = \frac{1}{h^2} \int_{x_{k-1/2}}^{x_{k+1/2}} \int_{y_{l-1/2}}^{y_{l+1/2}} f \, dx \, dy,$$

$$x_{k\pm 1/2} = x_k \pm \tfrac{1}{2}h, \qquad y_{l\pm 1/2} = y_l \pm \tfrac{1}{2}h.$$

Here and in the sequel we take as $f^h_{k,l}$ some average $f(x, y)$, calculated by the above formula. (This allows us to consider the difference schemes with a function $f(x, y)$ which, in general, is not assumed to be sufficiently smooth.)

We introduce the space Φ_h. Let the elements of Φ_h be defined in the domain

$$\bar{\Omega}_h = \Omega_h + \partial\Omega_h = \{(x_k, y_l): 0 \leqslant k \leqslant N_x = 1/h, \ 0 \leqslant l \leqslant N_y = 1/h\}.$$

Define the inner product and the norm as follows

$$(\varphi^h, \psi^h)_{\Phi_h} = \sum_{k=0}^{N_x} \sum_{l=0}^{N_y} h^2 \varphi^h_{k,l} \psi^h_{k,l},$$

$$\|\varphi^h\|_{\Phi_h} = \left(\sum_{k=0}^{N_x} \sum_{l=0}^{N_y} h^2 (\varphi^h_{k,l})^2 \right)^{1/2}$$

Choose as F_h the space of grid functions defined on

$$\Omega_h = \{(x_k, y_l): 1 \leqslant k \leqslant N_x - 1, \ 1 \leqslant l \leqslant N_y - 1\}$$

with the inner product and the norm given by:

$$(\varphi^h, \psi^h)_{F_h} = \sum_{k=1}^{N_x - 1} \sum_{l=1}^{N_y - 1} h^2 \varphi^h_{k,l} \psi^h_{k,l},$$

$$\|\varphi^h\|^2_{F_h} = \sum_{k=1}^{N_x - 1} \sum_{l=1}^{N_y - 1} h^2 (\varphi^h_{k,l})^2.$$

The space G_h of grid functions defined on $\partial\Omega_h$ is defined in a similar way.

Take as $(\varphi)_h$ a vector whose components are the values of the function φ at corresponding grid points. Then using the expansions of the functions $\varphi(x, y)$ and $f(x, y)$ into Taylor series we obtain

$$\| -\Delta^h (\varphi)_h - f^h \|_{F_h} \leqslant M_1 \cdot h^2, \tag{1.8}$$

where $M_1 = \text{const} < \infty$. The approximation of the boundary conditions in this example is accurate. From this fact and from the estimate (1.8) it follows that problem (1.7) approximates problem (1.5) with the second-order of accuracy on the solutions of problem (1.5) having bounded fourth-order derivatives.

Note, that if we require the grid functions of Φ_h to satisfy the condition $\varphi^h|_{\partial\Omega_h} = 0$, then for such functions the inner products in Φ_h and F_h (and therefore the generated norms) will coincide. Further, using the following identities analogous to the first and the second Green formulae

$$-\sum_{k=1}^{N_x-1} (\Delta_x \nabla_x \varphi^h)_{k,l} \cdot \psi^h_{k,l} = \sum_{k=1}^{N_x} (\nabla_x \varphi^h)_{k,l} \cdot (\nabla_x \psi^h)_{k,l},$$

$$-\sum_{k=1}^{N_x-1} (\Delta_x \nabla_x \varphi^h)_{k,l} \cdot \psi^h_{k,l} = -\sum_{k=1}^{N_x-1} (\Delta_x \nabla_x \psi^h)_{k,l} \cdot \varphi^h_{k,l}, \qquad \varphi^h, \psi^h \in \Phi_h$$

and similar identities for the sums over index l it is not difficult to show that

$$(A^h \varphi^h, \psi^h)_{F_h} = (\varphi^h, A^h \psi^h)_{F_h},$$

$$(A^h \varphi^h, \varphi^h)_{F_h} = h^2 \sum_{k=1}^{N_x} \sum_{l=1}^{N_y} [((\nabla_x \varphi^h)_{k,l})^2 + ((\nabla_y \varphi^h)_{k,l})^2] > 0$$

for $\varphi^h \not\equiv 0$

i.e., the operator A^h on the functions of Φ_h satisfying the condition $\varphi^h|_{\partial\Omega_h} = 0$ is a self-adjoint positive-definite operator. It is of interest to note also that here A^h is representable as the sum of symmetric positive-definite commutative operators A_x, A_y:

$$A^h = A_x + A_y, \qquad A_x A_y = A_y A_x.$$

So far the stationary problem and its approximation in spatial variables were considered. But we may consider in a similar way the problem of approximation of the evolutionary equation (i.e., the equation which has no time derivatives in A and may be resolved explicitly with respect to the first derivative in time):

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \qquad \Omega_t = (0, T), \\ &a\varphi = g \quad \text{on } \partial\Omega \times \Omega_t, \\ &\varphi = \varphi^0 \quad \text{in } \Omega \qquad \text{for } t = 0. \end{aligned} \tag{1.9}$$

Problem (1.9) will be approximated in two stages. First we shall approximate this problem in the domain $(\Omega_h + \partial\Omega_h) \times \Omega_t$ in spatial variables. As a result we obtain the differential equation in time which is a difference equation in spatial variables.

In some cases it is easy to exclude from the obtained differential-difference problem the values of the solution at the boundary points of the domain $(\Omega_h + \partial\Omega_h) \times \Omega_t$ using difference boundary conditions. Assuming this to be carried out we come to the evolutionary equation of the kind

$$d\varphi^h/dt + \Lambda\varphi^h = f^h, \tag{1.10}$$

where Λ, f^h and φ^h are functions of time t. Further we shall omit the index h in problem (1.10) assuming that we deal with the difference analogue of the original problem of mathematical physics in spatial variables.

Equation (1.10) presents a system of ordinary differential equations for the components of the vector φ^h.

Consider the following Cauchy problem:

$$\begin{aligned} &d\varphi/dt+\Lambda\varphi=f, \\ &\varphi=g \quad \text{for } t=0. \end{aligned} \tag{1.11}$$

Suppose that the operator Λ does not depend on time. Consider the simplest methods of approximation to problem (1.11) in time. Currently the difference schemes of the first and second order of approximation in t are used in most cases.

First, consider the simplest explicit scheme of the first-order of approximation on the grid Ω_τ:

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda\varphi^j=f^j, \qquad \varphi^0=g, \tag{1.12}$$

where $\tau=t_{j+1}-t_j$ and f^j is some projection of the function f. For the sake of simplicity we may now take $f^j=f(t_j)$.

If we consider the simplest implicit scheme, then we have

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda\varphi^{j+1}=f^j, \qquad \varphi^0=g, \tag{1.13}$$

and we choose f^j as $f(t_{j+1})$. The schemes (1.12) and (1.13) are first-order approximations in time. We may check it easily using expansion into Taylor series in time and assuming, for example, the existence of bounded second-order derivatives of the solution in time.

Resolving (1.12) and (1.13) with respect to φ^{j+1} we come to the recursive relation

$$\varphi^{j+1}=T\varphi^j+\tau Sf^j, \tag{1.14}$$

where the *step operator* T and the *source operator* S are defined as follows: for scheme (1.12)

$$T=E-\tau\Lambda, \qquad S=E;$$

for scheme (1.13)

$$T=(E+\tau\Lambda)^{-1}, \qquad S=T.$$

The difference schemes of type (1.14) for evolutionary equations will be called *two-level schemes.*

In applications the second-order approximation *Crank–Nicolson scheme* is widely used.

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda\frac{\varphi^{j+1}+\varphi^j}{2}=f^j, \qquad \varphi^0=g, \tag{1.15}$$

where $f^j = f(t_{j+1/2})$. Scheme (1.15) can also be represented in the form (1.14) with

$$T = (E + \tfrac{1}{2}\tau\Lambda)^{-1} \cdot (E - \tfrac{1}{2}\tau\Lambda), \qquad S = (E + \tfrac{1}{2}\tau\Lambda)^{-1}.$$

In some cases it is convenient to write the difference equations (1.12), (1.13) and (1.15) as a system of two equations, the first approximating in $\Omega_{h\tau}$ the equation itself, and the second approximating the boundary condition on $\partial\Omega_{h\tau}$. In this case the difference analogue of problem (1.9) has the form

$$\begin{aligned} &L^{h\tau}\varphi^{h\tau} = f^{h\tau} \quad \text{in } \Omega_{h\tau}, \\ &l^{h\tau}\varphi^{h\tau} = g^{h\tau} \quad \text{on } \partial\Omega_{h\tau}, \end{aligned} \tag{1.16}$$

where

$$\begin{aligned} &\Omega_{h\tau} = \Omega_h \times \Omega_\tau, \qquad \partial\Omega_{h\tau} = \Omega_h \times \{0\} \cup \partial\Omega_h \times \Omega_\tau, \\ &\|L^{h\tau}(\varphi)_{h\tau} - f^{h\tau}\|_{F_{h\tau}} \leqslant M_1 \cdot h^n + N_1 \cdot \tau^p, \\ &\|l^{h\tau}(\varphi)_{h\tau} - g^{h\tau}\|_{G_{h\tau}} \leqslant M_2 \cdot h^n + N_2 \cdot \tau^p. \end{aligned} \tag{1.17}$$

In these inequalities, just as in (1.3), $(\cdot)_{h\tau}$ is the operator of projection on the corresponding grid space.

By introducing vector functions and new operators acting in the domains $\Omega_h \times \Omega_\tau$, where Ω_τ is the set $\{t_j\}$, the difference equation in canonical form (1.14) can also be written in the form

$$\tilde{L}^{h\tau}\tilde{\varphi}^{h\tau} = \tilde{f}^{h\tau}. \tag{1.18}$$

Thus, the evolutionary equation with boundary conditions and initial data can be reduced to problem (1.18) of linear algebra.

In particular a boundary value problem of elliptic type, an integral equation, etc., may be reduced to equation (1.18). Here the approximation condition can again be written in the form (1.17) with the only approximation index h which is the maximum from the set $\{\Delta x_i\}$ of steps in spatial variables.

Consider the problem

$$\begin{aligned} &L\varphi \equiv \frac{\partial\varphi}{\partial t} + A\varphi \equiv \frac{\partial\varphi}{\partial t} - \Delta\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = 0 \quad \text{on } \partial\Omega \times \Omega_t, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0. \end{aligned} \tag{1.19}$$

Solutions are assumed to be defined on domain $(\Omega + \partial\Omega) \times \Omega_t$, where Ω is a square as before, and $\Omega_t = \{0 \leqslant t \leqslant T\}$. From $\Omega, \partial\Omega, \Omega_t$ we come to $\Omega_h, \partial\Omega_h, \Omega_\tau$ respectively. Let Ω_τ be the set of points t_j and let $t_{j+1} - t_j = \tau$. Then problem (1.19) can be approximated as follows:

$$\begin{aligned} &L^{h\tau}\varphi^{h\tau} = f^{h\tau} \quad \text{in } \Omega_h \times \Omega_\tau, \\ &\varphi^{h\tau} = 0 \qquad \text{on } \partial\Omega_h \times \Omega_\tau, \\ &\varphi^{h\tau} = g \qquad \text{in } \Omega_h \times \{0\}. \end{aligned} \tag{1.20}$$

Consider the simplest explicit approximation

$$(L^{h\tau}\varphi^{h\tau})^j_{k,l}\equiv\frac{\varphi^{j+1}_{k,l}-\varphi^j_{k,l}}{\tau}-(\Delta^h\varphi^j)_{k,l}, \tag{1.21}$$

$$f^j_{k,l}=\frac{1}{h^2}\int_{x_{k-1/2}}^{x_{k+1/2}}\int_{y_{l-1/2}}^{y_{l+1/2}} f(x,y,t_j)\,dx\,dy, \tag{1.22}$$

$$g_{k,l}=\frac{1}{h^2}\int_{x_{k-1/2}}^{x_{k+1/2}}\int_{y_{l-1/2}}^{y_{l+1/2}} g(x,y)\,dx\,dy, \tag{1.23}$$

where $\varphi^j_{k,l}=\varphi^{h\tau}(x_k, y_l, t_j)$. Then

$$\varphi^{j+1}_{k,l}=\varphi^j_{k,l}+\tau(\Delta^h\varphi^j)_{k,l}+\tau f^j_{k,l} \quad \text{in } \Omega_h\times\Omega_\tau. \tag{1.24}$$

Besides,

$$\begin{aligned}&\varphi^j_{k,l}=0 \qquad \text{on } \partial\Omega_h\times\Omega_\tau,\\ &\varphi^0_{k,l}=g_{k,l} \quad \text{in } \Omega_h\times\{0\}.\end{aligned} \tag{1.25}$$

The recursive relation (1.24) can be written as

$$\varphi^{j+1}=T\varphi^j+\tau f^j, \tag{1.26}$$

where φ^j is the grid function with the definition domain $\Omega_h+\partial\Omega_h$ and

$$\varphi^j(x_k, y_l)=\varphi^{h\tau}(x_k, y_l, t_j),$$

$$T=E+\tau\Delta^h=E-\tau(A_x+A_y)$$

is the step operator, and A_x and A_y are the same as in (1.6). For the sake of simplicity let here

$$F_h=\Phi_h \quad \text{and} \quad \|\varphi^h\|_{\Phi_h}=\left(\sum_{k,l}h^2|\varphi_{k,l}|^2\right)^{1/2}.$$

We estimate the norm of the operator T. To that end, we want to find its largest eigenvalue

$$\begin{aligned}&Tu=\lambda(T)u \quad \text{in } \Omega_h,\\ &u=0 \qquad\quad \text{on } \partial\Omega_h.\end{aligned} \tag{1.27}$$

The following relation is obviously valid

$$\lambda_n(T)=1+\tau\lambda_n(\Delta^h).$$

Noting that the orthonormal eigenvectors $\{u_{mp}\}$ of the operator A^h have the components

$$u^{kl}_{mp}=2\sin m\pi kh\cdot\sin p\pi lh,$$

$$m=1, 2, \ldots, N_x-1, \quad p=1, 2, \ldots, N_y-1,$$

and corresponding eigenvalues are

$$\lambda_{mp}(A^h)=\frac{4}{h^2}(\sin^2 \tfrac{1}{2}m\pi h+\sin^2 \tfrac{1}{2}p\pi h),$$

we easily find $\lambda_n(T)$ and the norm of the operator T:

$$\|T\|=\max\left\{\left|1-\frac{8\tau}{h^2}\cos^2 \tfrac{1}{2}\pi h\right|,\ \left|1-\frac{8\tau}{h^2}\sin^2 \tfrac{1}{2}\pi h\right|\right\} \tag{1.28}$$

and if $\tau/h^2<\frac{1}{4}$, then $\|T\|<1$.

Along with the explicit approximation of the first order in τ we may consider the implicit approximation of the first order in τ and of the second order in h^2. Then instead of (1.21) we will have:

$$(L^{h\tau}\varphi^{h\tau})^j_{k,l}\equiv\frac{\varphi^{j+1}_{k,l}-\varphi^j_{k,l}}{\tau}-(\Delta^h\varphi^{j+1})_{k,l}, \tag{1.29}$$

where $f^j_{k,l}$ and $g_{k,l}$ are defined by (1.22) and (1.23) respectively. In this case (1.20) can no longer be solved explicitly and we come to the operator equation

$$((E-\tau\Delta^h)\varphi^{j+1})_{k,l}=\varphi^j_{k,l}+\tau f^{j+1}_{k,l} \quad \text{in } \Omega_h\times\Omega_\tau, \tag{1.30}$$

whose solution must satisfy the following conditions:

$$\begin{aligned}&\varphi^j_{k,l}=0 \qquad \text{on } \partial\Omega_h\times\Omega_\tau,\\ &\varphi^0_{k,l}=g_{k,l} \quad \text{in } \Omega_h\times\{0\}.\end{aligned} \tag{1.31}$$

We write (1.30) in the form

$$\varphi^{j+1}_{k,l}=(T(\varphi^j_{k,l}+\tau f^j))_{k,l}, \tag{1.32}$$

where $T=(E-\tau\Delta^h)^{-1}$. In this case the norm of the operator T becomes

$$\|T\|=\max\left\{\frac{1}{1+(8\tau/h^2)\cos \tfrac{1}{2}\pi h},\ \frac{1}{1+(8\tau/h^2)\sin^2 \tfrac{1}{2}\pi h}\right\} \tag{1.33}$$

and therefore $\|T\|<1$ for any τ and h.

Finally, we consider the approximation in the Crank–Nicolson scheme. Define the operators and functions in problem (1.20) as follows:

$$(L^{h\tau}\varphi^{h\tau})^j_{k,l}\equiv\frac{\varphi^{j+1}_{k,l}-\varphi^j_{k,l}}{\tau}-\left(\Delta^h\frac{\varphi^{j+1}+\varphi^j}{2}\right)_{k,l}, \tag{1.34}$$

$$\begin{aligned}f^j_{k,l}&=\frac{1}{h^2}\int_{x_k-1/2}^{x_k+1/2}\int_{y_l-1/2}^{y_l+1/2} f(x,y,t_{j+1/2})\,dx\,dy,\\ g_{k,l}&=\frac{1}{h^2}\int_{x_k-1/2}^{x_k+1/2}\int_{y_l-1/2}^{y_l+1/2} g(x,y)\,dx\,dy.\end{aligned} \tag{1.35}$$

Then we obtain the problem

$$(E-\tfrac{1}{2}\tau\Delta^h)\varphi_{k,l}^{j+1}=(E+\tfrac{1}{2}\tau\Delta^h)\varphi_{k,l}^{j}+\tau f_{k,l}^{j} \quad \text{in } \Omega_h\times\Omega_\tau, \tag{1.36}$$

$$\begin{aligned}&\varphi_{k,l}^{j}=0 \quad \text{on } \partial\Omega_h\times\Omega_\tau,\\ &\varphi_{k,l}^{0}=g_{k,l} \quad \text{in } \Omega_h\times\{0\}.\end{aligned} \tag{1.37}$$

In this case (1.36) can be formally solved with respect to the unknown $\varphi_{k,l}^{j+1}$ in the form

$$\varphi_{k,l}^{j+1}=(T\varphi^j)_{k,l}+\tau(Sf^j)_{k,l}, \tag{1.38}$$

where

$$T=(E-\tfrac{1}{2}\tau\Delta^h)^{-1}(E+\tfrac{1}{2}\tau\Delta^h), \quad S=(E-\tfrac{1}{2}\tau\Delta^h)^{-1}.$$

The norm of the step operator is

$$\|T\|=\max\left\{\left|\frac{1-(4\tau/h^2)\cos^2\frac{1}{2}\pi h}{1+(4\tau/h^2)\cos^2\frac{1}{2}\pi h}\right|, \left|\frac{1-(4\tau/h^2)\sin^2\frac{1}{2}\pi h}{1+(4\tau/h^2)\sin^2\frac{1}{2}\pi h}\right|\right\}. \tag{1.39}$$

Hence $\|T\|<1$.

2. Stability

Consider the next important concept of the theory of finite difference methods: *stability*.

To clarify basic definitions and concepts of the theory of stability consider first the explicit difference scheme (1.12)

$$\varphi^{j+1}=(E-\tau\Lambda)\varphi^j+\tau f^j, \qquad \varphi^0=g \tag{2.1}$$

The solution φ^j is found for $0\leqslant\tau j\leqslant T$.

Assume that the operator Λ is positive and has a complete system of eigenfunctions $\{u_n\}$ and a set of eigenvalues $\{\lambda_n>0\}$ corresponding to the spectral problem

$$\Lambda u=\lambda u.$$

We introduce the following Fourier series:

$$\varphi^j=\sum_n \varphi_n^j u_n, \qquad f^j=\sum_n f_n^j u_n, \qquad g=\sum_n g_n u_n, \tag{2.2}$$

where

$$\varphi_n^j=(\varphi^j, u_n^*), \qquad f_n^j=(f^j, u_n^*), \qquad g_n=(g, u_n^*);$$

u_n^* are the eigenfunctions of the adjoint spectral problem. Substituting (2.2) into (2.1) and taking the inner product of the result and the vectors u_n^* we obtain the following expression for the Fourier coefficients:

$$\varphi_n^{j+1}=(1-\tau\lambda_n)\varphi_n^j+\tau f_n^j. \tag{2.3}$$

Since

$$\varphi^0=\sum_n g_n u_n,$$

the initial condition has the form

$$\varphi_n^0=g_n. \tag{2.4}$$

By successively excluding the unknowns we obtain

$$\varphi_n^j=r_n^j g_n+\tau\sum_{i=1}^{j} r_n^{j-i} f_n^{i-1}, \tag{2.5}$$

where

$$r_n=1-\tau\lambda_n. \tag{2.6}$$

From (2.5) it follows that for $\tau>0$

$$|\varphi_n^j|\leqslant|r_n|^j\cdot|g_n|+\tau\sum_{i=1}^{j}|r_n|^{j-i}\cdot|f_n^{i-1}|.$$

We strengthen the latter inequality by substituting $|f_n^{i-1}|$ for $|f_n|=\max|f_n^j|$ under the summation symbol and we obtain

$$|\varphi_n^j|\leqslant|r_n|^j|g_n|+\frac{1-|r_n|^j}{1-|r_n|}\cdot\tau|f_n|. \tag{2.7}$$

Von Neumann has introduced the so-called *spectral criterion of stability*, the essence of which is as follows. If for each Fourier coefficient φ_n^j of (2.2) the following relation holds

$$|\varphi_n^j|\leqslant C_{1n}|g_n|+C_{2n}|f_n|,\quad n=1,2,\ldots, \tag{2.8}$$

where C_{1n}, C_{2n} are constants bounded uniformly for $0\leqslant j\tau\leqslant T$, then the difference scheme (2.1) is *computationally stable*. Consider conditions on the parameters of the difference scheme (1.12) sufficient for relation (2.8) to be valid. An analysis of (2.7) shows that the stability criterion (2.8) is satisfied if the following restriction is imposed on the parameter r_n

$$|r_n|\leqslant 1,\quad n=1,2,\ldots. \tag{2.9}$$

Assume that the spectrum of the operator Λ is contained in the interval

$$0<\alpha(\Lambda)\leqslant\lambda_n(\Lambda)\leqslant\beta(\Lambda).$$

Then, according to (2.6), if

$$\tau\leqslant 2/\beta(\Lambda), \tag{2.10}$$

then relation (2.9) will hold. Thus inequality (2.10) becomes a constructive stability condition for the difference scheme (2.1). Note that condition (2.10) is a sufficient stability condition; the scheme remains stable, for instance, when

$$\tau=2/\beta(\Lambda).$$

Relation (2.7) in this case becomes obviously

$$|\varphi_n^j| \leqslant |g_n| + j\tau |f_n|. \tag{2.11}$$

But $j\tau \leqslant T$, where T is fixed. This means that for small τ a large number of steps j is required, and $j \to \infty$ as $\tau \to 0$, but such that the upper bound of the time interval T remains fixed. Again, we come to schemes which are stable in the sense of von Neumann.

We consider now some other difference schemes based on the implicit difference approximations. In the case of the implicit scheme of the first-order approximation (1.13) we obtain an expression similar to (2.7):

$$|\varphi_n^j| \leqslant |r_n|^j \cdot |g_n| + \frac{1-|r_n|^j}{1-|r_n|} \cdot \tau |r_n| |f_n|, \tag{2.12}$$

where

$$r_n = 1/(1+\tau\lambda_n(\Lambda)).$$

This difference scheme for $\lambda_n(\Lambda) > 0$ is stable for any $\tau > 0$, since

$$|r_n| < 1, \quad n = 1, 2, \ldots.$$

Stability of this kind will be called the *absolute* stability.

In the case of the Crank–Nicolson scheme (1.15) the estimate for the Fourier coefficients of the solution has the form

$$|\varphi_n^j| \leqslant |r_n|^j \cdot |g_n| + \frac{1-|r_n|^j}{1-|r_n|} \tau |\mu_n| |f_n|, \tag{2.13}$$

where

$$r_n = \frac{1-\frac{1}{2}\tau\lambda_n(\Lambda)}{1+\frac{1}{2}\tau\lambda_n(\Lambda)}, \qquad \mu_n = \frac{1}{1+\frac{1}{2}\tau\lambda_n(\Lambda)}.$$

Therefore $|r_n| < 1$ for any $\tau > 0$ provided that $\lambda_n(\Lambda) > 0$.

It should be noted that, firstly, the stability in von Neumann's sense is based on the spectral analysis of the operator of the problem. This means that for this approach the computation of the largest eigenvalue or the estimation of its upper bound are necessary parts of algorithm. Secondly, the spectral stability criterion ascertains stability of the solution with respect to each of the harmonics from the Fourier series but it says nothing about the stability of the solution in terms of energy norms. At the same time the norm of the solution φ^j often happens to be the only characteristic of the problem's solution. All of this made the researchers offer some other definitions of stability which would be related to the norms of the operators. It should be emphasized, however, that stability analysis of von Neumann type continues to play a prominent role in applications.

We now come to a more general definition of the concept of computational stability. To that end we consider the problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{for } t = 0, \end{aligned} \tag{2.14}$$

which is approximated by the difference problem

$$\begin{aligned} &\varphi^{j+1} = T\varphi^j + \tau S f^j \quad \text{in } \Omega_h \times \Omega_\tau, \\ &\varphi^0 = g. \end{aligned} \tag{2.15}$$

We shall say that the difference scheme (2.15) is *stable*, if for any parameter h characterizing the difference approximation and for $j \leqslant T/\tau$ the following relation holds

$$\|\varphi^j\|_{\Phi_h} \leqslant C_1 \cdot \|g\|_{G_h} + C_2 \|f^{h\tau}\|_{F_{h\tau}}, \tag{2.16}$$

where the constants C_1 and C_2 are uniformly bounded on $0 \leqslant t \leqslant T$ and do not depend on τ, h, g and f; G_h denotes the space to which g of (2.15) belongs.

The definition of computational stability is closely related to the concept of well-posed problems with a continuous argument. One may say that computational stability implies continuous dependence of the solution on the input data in the case of problems with a discrete argument. It is easy to see that the definition of stability in the form (2.16) already relates the solution itself with the a priori knowledge concerning the input data for the problem. This definition is often more suitable for analyzing the stability of many problems and more informative than Neumann's definition. Consider the stability of the scheme (1.12). To this end rewrite the recursive relation (2.1) as

$$\varphi^{j+1} = T\varphi^j + \tau f^j, \qquad \varphi^0 = g, \tag{2.17}$$

where

$$T = E - \tau \Lambda, \tag{2.18}$$

and Λ is the operator approximating A. The formal solution of the problem (2.17) has the form

$$\varphi^{j+1} = T^j g + \sum_{i=1}^{j} T^{j-i} f^{i-1}. \tag{2.19}$$

We assume[1] $G_h = F_h$, and by estimating the norm of the solution of (2.19) we obtain

$$\|\varphi^{j+1}\|_{\Phi_h} \leqslant \|T\|^j \|g\|_{G_h} + \tau \sum_{i=1}^{j} \|T\|^{j-i} \|f^{i-1}\|_{F_h}. \tag{2.20}$$

We substitute for $\|f^{i-1}\|$ under the summation symbol its maximal value for all j from the fixed time interval. Let

$$\|f^{h\tau}\| = \max_j \|f^j\|_{F_h},$$

[1]This assumption is made for the sake of simplicity. Otherwise instead of one norm $\|T\| \equiv \|T\|_{F_h \to \Phi_h} = \sup_{\varphi \in F_h} (\|T\varphi\|_{\Phi_h} / \|\varphi\|_h)$ we should introduce two norms: $\|T\|_{F_h \to \Phi_h}$ and $\|T\|_{G_h \to \Phi_h} = \sup_{\varphi \in G_h} (\|T\varphi\|_{\Phi_h} / \|\varphi\|_{G_h})$.

then

$$\|\varphi^j\|_{\Phi_h} \leqslant \|T\|^j \|g\|_{G_h} + \frac{1-\|T\|^j}{1-\|T\|}\tau\|f^{h\tau}\|. \tag{2.21}$$

If we assume that

$$\|T\| < 1, \tag{2.22}$$

then scheme (1.12) will be stable in the sense of definition (2.16). Naturally, (2.22) is a sufficient condition for stability. More sophisticated conditions could be obtained using the norms of powers of the step operators $\|T^i\|$ $(i=1,2,\ldots,j)$. Weakening the condition however brings additional difficulties in the constructive procedure of ascertaining the stability criterion. As a rule, the sufficient condition (2.22) is used in practice.

The stability of the implicit difference equations (1.13) and (1.15) may be considered in a similar way. In these cases we have

$$\|\varphi^j\|_{\Phi_h} \leqslant \|T\|^j \cdot \|g\|_{G_h} + \frac{1-\|T\|^j}{1-\|T\|}\tau \cdot \|S\| \cdot \|f\|_{F_h},$$

where for scheme (1.13)

$$T=(E+\tau\Lambda)^{-1}, \qquad S=(E+\tau\Lambda)^{-1}$$

and for scheme (1.15)

$$T=(E+\tfrac{1}{2}\tau\Lambda)^{-1}(E-\tfrac{1}{2}\tau\Lambda), \qquad S=(E+\tfrac{1}{2}\tau\Lambda)^{-1}.$$

It is easy to show that these difference schemes will be absolutely stable in the sense of definition (2.16) provided that $\Lambda > 0$ and

$$\|\varphi^j\|_{\Phi_h} = \left[\sum_{k,l} |\varphi^j_{k,l}|^2 h^2\right]^{1/2}.$$

It may be done using the estimate

$$\|(E+\tau\Lambda)^{-1}\| \leqslant 1 \quad \text{for } \Lambda \geqslant 0, \quad \tau > 0 \tag{2.23}$$

resulting from the relations:

$$\frac{\|(E+\tau\Lambda)^{-1}\varphi\|^2}{\|\varphi\|^2} = \frac{\|\psi\|^2}{\|(E+\tau\Lambda)\psi\|^2}$$

$$= \frac{\|\psi\|^2}{\|\psi\|^2 + \tau((\Lambda+\Lambda^*)\psi,\psi) + \tau^2\|\Lambda\psi\|^2} \leqslant 1$$

and from the following theorem (which is very useful in the stability analysis of many schemes).

THEOREM 2.1 (KELLOGG [1963]). *If the operator A acting in real Hilbert space Φ is positive-semidefinite and the numerical parameter σ is nonnegative, then*

$$\|(E-\sigma A)\cdot(E+\sigma A)^{-1}\| \leqslant 1. \tag{2.24}$$

Indeed, introduce the notation $T=(E-\sigma A)\cdot(E+\sigma A)^{-1}$. Then we have

$$\|T\|^2=\sup_{\varphi}\frac{((E-\sigma A)(E+\sigma A)^{-1}\varphi,(E-\sigma A)(E+\sigma A)^{-1}\varphi)_\Phi}{(\varphi,\varphi)_\Phi}$$

$$=\sup_{\psi}\frac{((E-\sigma A)\psi,(E-\sigma A)\psi)_\Phi}{((E+\sigma A)\psi,(E+\sigma A)\psi)_\Phi}$$

$$=\sup_{\psi}\frac{(\psi,\psi)_\Phi-2\sigma(A\psi,\psi)_\Phi+\sigma^2(A\psi,A\psi)_\Phi}{(\psi,\psi)_\Phi+2\sigma(A\psi,\psi)_\Phi+\sigma^2(A\psi,A\psi)_\Phi}\leqslant 1.$$

REMARK 2.1. In the case when $A>0$ and $\sigma>0$ we would have instead of (2.24) the inequality of the kind

$$\|(E-\sigma A)(E+\sigma A)^{-1}\|<1. \tag{2.25}$$

Discuss briefly the limit transitions. In solving the difference analogues of the evolutionary problems of mathematical physics we have to consider approximations in time with step τ as well as in space with characteristic step h. This means that the transition operator $T=T(\tau,h)$ depends both on τ and on h.

The construction of a stable algorithm for a given approximation method is usually reduced to establishing the relation between τ and h which ensures the computational stability. If the difference scheme happens to be stable for any $\tau>0$ and $h>0$, then it is declared *absolutely stable*. And if the scheme happens to be stable only in the case where there is a certain relation between τ and h then such a scheme is called *conditionally stable*.

Assume that τ and h are related according to the inequality

$$\tau\leqslant C\cdot h^p, \tag{2.26}$$

where C and p are given constants independent of τ and h. It is worthy to note that such relations are usually established while considering the amplitudes of the "shortest" perturbation. As a rule they reflect the connection between the minimal spatial and time scales of the phenomena simulated by the difference scheme.

Of course, large perturbations (say, of the order of several h) will then be described more precisely.

Suppose that we need to increase the accuracy of the solution formally by decreasing the grid step h. Then we must decrease simultaneously the time step τ so that the above inequality is again satisfied with these new parameters. This means that we may even allow the limit transition as $\tau\to 0$ and $h\to 0$ if we satisfy the condition (2.26), for example, in the form

$$\tau/h^p=\text{const}\leqslant C.$$

Along with the above definitions of the computational stability other definitions are used in the literature allowing the expansion of the class of difference schemes which are of interest in applications. For example, the scheme is called stable if

$$\|T\|\leqslant 1+O(\tau). \tag{2.27}$$

For small τ such a definition allows for the exponential growth of round-off errors in time.

Note that if the approximation of the evolutionary equation is studied in the spaces of grid functions defined in $\Omega_h \times \Omega_\tau$ then it is useful to give the definition of the stability in terms of the same spaces. Indeed, let the difference problem have the form

$$\begin{aligned} &L^{h\tau}\varphi^{h\tau} = f^{h\tau} \quad \text{in } \Omega_h \times \Omega_\tau, \\ &l^{h\tau}\varphi^{h\tau} = g^{h\tau} \quad \text{on } \partial\Omega_h \times \Omega_\tau. \end{aligned} \tag{2.28}$$

We introduce the stability criterion in the form:

$$\|\varphi^{h\tau}\|_{\Phi_{h\tau}} \leqslant C_1 \|f^{h\tau}\|_{F_{h\tau}} + C_2 \|g^{h\tau}\|_{G_{h\tau}}, \tag{2.29}$$

where C_1 and C_2 are constants on the interval $0 \leqslant t \leqslant T$ independent of h, τ, $f^{h\tau}$ and $g^{h\tau}$.

Assume that the original problem is approximated by the difference equation with the boundary conditions already taken into account. Then it is convenient to introduce the stability criterion in the form:

$$\|\varphi^{h\tau}\|_{\Phi_{h\tau}} \leqslant C \|f^{h\tau}\|_{F_{h\tau}}, \tag{2.30}$$

where C is bounded on the interval $0 \leqslant t \leqslant T$.

3. Convergence

We turn now to the formulation of the major result in the theory of the finite difference algorithms: *the convergence theorem.* The study of the convergence of the difference solution to the solution of the original problem for the stationary and evolutionary problems of mathematical physics is based on the same principles. This allows us to follow the main idea of the proof on the example of the stationary problem (1.1) which is approximated by the *difference scheme* (1.2) (i.e., by the system of equations approximating both the equation and the boundary condition of (1.1)).

The following convergence theorem is valid (GODUNOV and RYABENKII [1970], FILIPPOV [1955]).

THEOREM 3.1. *Let the following conditions hold:*

(1) *The difference scheme* (1.2) *approximates the original problem* (1.1) *on the solution* φ *with order* n.

(2) A^h *and* a^h *are linear operators.*

(3) *The difference scheme* (1.2) *is stable in the sense of* (2.29), *i.e. there exist positive constants* $\bar{h}, C_1, C_2$ *such that for all* $h < \bar{h}$, $f^h \in F_h$, $g^h \in G_h$ *there exists a unique solution* φ^h *of problem* (1.2) *which satisfies the inequality*

$$\|\varphi^h\|_{\Phi_h} \leqslant C_1 \|f^h\|_{F_h} + C_2 \|g^h\|_{G_h}. \tag{3.1}$$

Then the solution of the difference problem φ^h *converges to the solution* φ *of the original problem, i.e.,*

$$\lim_{h \to 0} \|(\varphi)_h - \varphi^h\|_{\Phi_h} = 0,$$

when the following estimate for the convergence rate is valid

$$\|(\varphi)_h - \varphi^h\|_{\Phi_h} \leqslant (C_1 M_1 + C_2 M_2) h^n, \tag{3.2}$$

where M_1 *and* M_2 *are the same constants as in* (1.3).

PROOF. Let $\bar{h}$ be the constant introduced in the definitions of approximation and stability. The stability implies that for any right-hand sides f^h and g^h for $h < \bar{h}$ there exists a unique solution φ^h, i.e. for $h < \bar{h}$ we may consider the difference $(\varphi)_h - \varphi^h$. Since A^h is linear we have

$$A^h[(\varphi)_h - \varphi^h] = A^h(\varphi)_h - A^h\varphi^h = A^h(\varphi)_h - f^h.$$

Similarly,

$$a^h[(\varphi)_h - \varphi^h] = a^h(\varphi)_h - g^h.$$

Since $h < \bar{h}$, owing to the stability and approximation of the relations

$$A^h[(\varphi)_h - \varphi^h] = A^h(\varphi)_h - f^h, \qquad a^h[(\varphi)_h - \varphi^h] = a^h(\varphi)_h - g^h, \tag{3.3}$$

it follows that

$$\begin{aligned}\|(\varphi)_h - \varphi^h\|_{\Phi_h} &\leqslant C_1 \|A^h(\varphi)_h - f^h\|_{F_h} + C_2 \|a^h(\varphi)_h - g^h\|_{G_h} \\ &\leqslant C_1 M_1 h^{n_1} + C_2 M_2 h^{n_2} \leqslant (C_1 M_1 + C_2 M_2) h^n.\end{aligned}$$

While obtaining the last inequality we may suppose without loss of generality that $h < 1$.

In the case of the evolutionary problem consider

$$\begin{aligned}\delta f^{h\tau} &\equiv L^{h\tau}[(\varphi)_{h\tau} - \varphi^{h\tau}] \equiv L^{h\tau}(\varphi)_{h\tau} - f^{h\tau}, \\ \delta g^{h\tau} &\equiv l^{h\tau}[(\varphi)_h - \varphi^{h\tau}] \equiv l^{h\tau}(\varphi)_{h\tau} - g^{h\tau}.\end{aligned} \tag{3.4}$$

From (3.4) and from the stability condition (2.29) we have

$$\|(\varphi)_{ht} - \varphi^{ht}\|_{\Phi_{ht}} \leqslant C_1 \|\delta f^{ht}\|_{F_{ht}} + C_2 \|\delta g^{ht}\|_{G_{ht}},$$

or, taking (1.17) into account,

$$\|(\varphi)_{h\tau} - \varphi^{h\tau}\|_{\Phi_{h\tau}} \leqslant K_1 h^n + K_2 \tau^p, \tag{3.5}$$

where

$$K_1 = C_1 M_1 + C_2 M_2, \qquad K_2 = C_1 N_1 + C_2 N_2.$$

Estimate (3.5) proves the convergence of the difference solution to the exact one and gives a clear comprehension of the convergence with respect to the spatial grid step h as well as to the time step τ. □

The assumptions of the theorem include the rather rigid requirement that C_1 and C_2 are independent of h and τ. The condition that C_1 and C_2 must be independent of h is especially unpleasant since in some cases C_1 and C_2 may tend to infinity as $h \to 0$. Let

$$C_1^h = C_1/h^m, \qquad C_2^h = C_2/h^m,$$

where $m \geqslant 0$. The rate of convergence of the approximate solution to the exact one will then be estimated as follows:

$$\|(\varphi)_{h\tau} - \varphi^{h\tau}\|_{\Phi_{n\tau}} \leqslant Mh^{k-m} + N\tau^p h^{-m}.$$

If $k > m$ and $\tau^p \theta^{-m} \to 0$ as $\tau \to 0, h \to 0$, then the convergence takes place. Of course, the convergence theorem can be formulated even in the case when C_1 and C_2 depend both on h and τ.

4. The Crank–Nicolson scheme

The Crank–Nicolson scheme will play a significant role in the next chapters. So we shall consider this scheme in more detail.

Consider the evolutionary equation

$$\begin{aligned} &\frac{\partial \varphi}{\partial t} + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0, \end{aligned} \tag{4.1}$$

where $A \geqslant 0$ (i.e. $(A\varphi, \varphi) \geqslant 0$) and the solution φ is sufficiently smooth. We shall assume that the solution satisfies certain boundary conditions on $\partial\Omega$. We shall assume as well that (4.1) already presents the finite difference approximation to the corresponding original evolutionary problem in all variables excluding t (i.e. A is a matrix, φ is the grid function depending on t, etc., as we specified above, the index "h" of A, φ, Ω and other symbols will be often omitted for the sake of simplicity). Suppose first that A does not depend on t. Then it is easy to check that the difference equation of the form

$$\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\frac{(\varphi^{j+1} + \varphi^j)}{2} = 0, \qquad \varphi^0 = g \tag{4.2}$$

approximates problem (4.1) with second-order accuracy in τ. This difference scheme, as was already noted, is usually referred to as the *Crank–Nicolson scheme*. It is of interest to note that scheme (4.2) is the result of alternately applying explicit and implicit first-order schemes written on the intervals $t_j \leqslant t \leqslant t_{j+1/2}$ and $t_{j+1/2} \leqslant t \leqslant t_{j+1}$ respectively (provided that A is a linear operator independent of t):

$$\begin{aligned} &\frac{\varphi^{j+1/2} - \varphi^j}{\frac{1}{2}\tau} + A\varphi^j = 0, \\ &\frac{\varphi^{j+1} - \varphi^{j+1/2}}{\frac{1}{2}\tau} + A\varphi^{j+1} = 0. \end{aligned} \tag{4.3}$$

Excluding the unknowns $\varphi^{j+1/2}$ from this system of difference equations we come to the Crank–Nicolson scheme.

Assume that the operator A is independent of time and is approximated in the

problem (4.1) by the difference operator, which will be denoted by Λ. Then we will deal with the problem in linear algebra of the form

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda^j\frac{\varphi^{j+1}+\varphi^j}{2}=0, \qquad \varphi^0=g, \tag{4.4}$$

where

$$(\Lambda^j\varphi, \varphi)\geqslant 0 \tag{4.5}$$

for all functions from the subspace Φ.

Resolving (4.4) with respect to φ^{j+1} we obtain

$$\varphi^{j+1}=(E+\tfrac{1}{2}\tau\Lambda^j)^{-1}\cdot(E-\tfrac{1}{2}\tau\Lambda^j)\varphi^j, \tag{4.6}$$

$$\varphi^{j+1}=T^j\varphi^j, \tag{4.7}$$

where T^j is a step operator:

$$T^j=(E+\tfrac{1}{2}\tau\Lambda^j)^{-1}\cdot(E-\tfrac{1}{2}\tau\Lambda^j). \tag{4.8}$$

In order to prove the computational stability we need not to estimate the norm of the step operator T^j. Taking the inner product of (4.4) and $\frac{1}{2}(\varphi^{j+1}+\varphi^j)$ we obtain

$$\frac{(\varphi^{j+1}, \varphi^{j+1})_\Phi-(\varphi^j, \varphi^j)_\Phi}{2\tau}+\left(\Lambda^j\frac{\varphi^{j+1}+\varphi^j}{2}, \frac{\varphi^{j+1}+\varphi^j}{2}\right)_\Phi=0. \tag{4.9}$$

If Λ^j is positive-semidefinite, then

$$\|T^j\|\leqslant 1 \tag{4.10}$$

and, therefore,

$$\|\varphi^{j+1}\|_\Phi\leqslant\|\varphi^j\|_\Phi, \tag{4.11}$$

i.e., the scheme is stable.

If the operator Λ^j is skew-symmetric, i.e., the following equality holds:

$$(\Lambda^j\varphi, \varphi)_\Phi=0,$$

then we have instead of (4.10) the equality

$$\|\varphi^{j+1}\|_\Phi=\|\varphi^j\|_\Phi. \tag{4.12}$$

And so here

$$\|T^j\| = 1. \tag{4.13}$$

We now consider the approximation with the Crank–Nicolson scheme in the case where the operator A is time-dependent. We define the operator L by the equality

$$L\varphi\equiv\partial\varphi/\partial t+A\varphi \tag{4.14}$$

and the operator L_τ by the equality

$$(L_\tau\varphi)^j=\frac{(\varphi)^{j+1}-(\varphi)^j}{\tau}+\Lambda^j\frac{(\varphi)^{j+1}+(\varphi)^j}{2}, \tag{4.15}$$

where $(\varphi)^j$ is the projection of a function φ on the grid Ω_τ. Further, we introduce the norm

$$\|(L_\tau\varphi)\|_{C_\tau}=\max_{t_j\in\Omega_\tau}\|(L_\tau\varphi)^j\|, \tag{4.16}$$

where $\|\cdot\|$ is some norm in the space to which $(L_\tau\varphi)^j$ belongs. In order to estimate the norm (4.16) we expand the solution of the original equation (4.1) into a Taylor series, and we have:

$$(\varphi)^{j+1}=(\varphi)^j+\tau(\varphi_t)^j+\tfrac{1}{2}\tau^2(\varphi_{tt})^j+\cdots. \tag{4.17}$$

Taking into account the obvious relations

$$\begin{aligned}\varphi_t&=-A\varphi,\\ \varphi_{tt}&=A^2\varphi-A_t\varphi,\end{aligned} \tag{4.18}$$

where $A_t=\partial A/\partial t$, we transform Taylor series (4.17) into the form

$$(\varphi)^{j+1}=(\varphi)^j-\tau A^j(\varphi)^j+\tfrac{1}{2}\tau^2[(A^j)^2(\varphi)^j-A_t^j(\varphi)^j-\cdots]. \tag{4.19}$$

Substituting (4.19) into (4.16) and taking into account (4.15) we obtain

$$\begin{aligned}&\|(f-L_\tau\varphi)\|_{C_\tau}\\ &\quad=\max_{t_j}\|\Lambda^j(\varphi)^j-A^j(\varphi)^j+\tfrac{1}{2}\tau\{(A^j)^2-A_t^j-\Lambda^jA^j\}(\varphi)^j+O(\tau^2)\|,\end{aligned} \tag{4.20}$$

where f^j is the right-hand side of (4.4) which, in this case, equals zero.

If we choose as the approximation of the operator Λ the operator

$$\Lambda^j=A^j=A(t_j), \tag{4.21}$$

then it follows from (4.20) that

$$\|(f-L_\tau\varphi)\|_{C_\tau}=\tfrac{1}{2}\tau\max_{t_j\in\Omega_\tau}\|A_t^j(\varphi)^j\|+O(\tau^2)$$

and we have the first-order approximation. Note that in the special case where A is independent of t the approximation in the form (4.21) ensures second-order accuracy in τ.

Suppose now that the approximating operator has been chosen in the form

$$\Lambda^j=A^j+\tfrac{1}{2}\tau A_t^j. \tag{4.22}$$

In this case we have

$$\|(f-L_\tau\varphi)\|_{C_\tau}=O(\tau^2).$$

Note that the approximation with the Crank–Nicolson scheme will also have second-order accuracy in τ if the operator Λ^j is chosen in the form

$$\Lambda^j=A^{j+1/2} \tag{4.23}$$

or

$$\Lambda^j=\tfrac{1}{2}(A^{j+1}+A^j). \tag{4.24}$$

The three forms (4.22), (4.23) and (4.24) of the approximation of the operator A ensuring second-order accuracy are used in various applications, in particular in numerically solving the quasi-linear equations.

We now consider the inhomogeneous equations

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0. \end{aligned} \tag{4.25}$$

The difference approximation to problem (4.25) based on the Crank–Nicolson scheme has, under the assumptions formulated above, the form

$$\frac{\varphi^{j+1} - \varphi^j}{\tau} + \Lambda^j \frac{\varphi^{j+1} + \varphi^j}{2} = f^j, \tag{4.26}$$

where

$$f^j = f(t_{j+1/2}).$$

It is not difficult to check that the difference problem (4.26) approximates (4.25) with second-order accuracy in τ. Write the solution of problem (4.26) in the form

$$\varphi^{j+1} = T^j \varphi^j + \tau(E + \tfrac{1}{2}\tau\Lambda^j)^{-1} f^j. \tag{4.27}$$

From (4.27) it follows that

$$\|\varphi^{j+1}\|_\Phi \leqslant \|T^j\| \cdot \|\varphi^j\|_\Phi + \tau \|(E + \tfrac{1}{2}\tau\Lambda^j)^{-1}\| \cdot \|f^j\|_F. \tag{4.28}$$

Since $\tau > 0$ and

$$(\Lambda^j \varphi^j, \varphi^j)_\Phi \geqslant 0, \tag{4.29}$$

we obtain

$$\begin{aligned} &\|T^j\| \leqslant 1, \\ &\|(E + \tfrac{1}{2}\tau\Lambda^j)^{-1}\| \leqslant 1. \end{aligned} \tag{4.30}$$

Therefore, (4.28) leads to

$$\|\varphi^{j+1}\|_\Phi \leqslant \|\varphi^j\|_\Phi + \tau \|f^j\|_F. \tag{4.31}$$

Taking

$$\|\varphi^0\|_\Phi = \|g\|, \qquad \|f\| = \max_j \|f^j\|_F$$

and using the recursive relation (4.31) we obtain

$$\|\varphi^j\|_\Phi \leqslant \|g\| + j\tau \|f\|, \quad j\tau \leqslant \text{const}. \tag{4.32}$$

Thus relation (4.32) proves the stability of the difference scheme. Besides, this relation presents the a priori estimate for the norm of the solution.

Thus we have introduced some concepts which will be used in considering the splitting and alternating direction methods. Besides, we have shown how to reduce this problem using the finite difference method and approximating the original

evolution problem in all variables (excluding t) to the solution of the system of ordinary differential equations of the form

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t=0, \end{aligned} \tag{4.33}$$

which may be solved using splitting or alternating direction methods. (*Note that it is this approach of successive approximation to the problem that is often used in practice!*)

Further, the finite difference method for the approximation of problems has been chosen here since the present work is devoted mainly to this method. However, currently systems of difference equations may be obtained in various ways. Thus, the finite element method also leads to a certain system which may be considered a finite difference system (STRANG and FIX [1977], BABUŠKA [1970], MARCHUK and AGOSHKOV [1981]). But in this case as a rule we obtain instead of (4.33) the system

$$\begin{aligned} &B\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &B\varphi = g \quad \text{for } t=0, \end{aligned} \tag{4.34}$$

where the matrix B is symmetric positive-definite and presents a Gram matrix (if the original problem has in the main equation a distinguished derivative "$\partial\varphi/\partial t$"). But here it is often possible to approximately replace this matrix in (4.34) by a certain diagonal matrix $\tilde{B}$ (with positive elements on the diagonal). As a result from (4.34) one comes to the problem

$$\begin{aligned} &\tilde{B}\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\tilde{B}\varphi = g \quad \text{for } t=0, \end{aligned} \tag{4.35}$$

and the order of approximation of the original problem by problem (4.35) remains the same as while considering the system (4.34) (see STRANG and FIX [1977]). Now it is easy to notice that by introducing the vector $\Phi = \tilde{B}^{1/2}\varphi$ and the matrix $\tilde{A} = \tilde{B}^{-1/2} A \tilde{B}^{-1/2}$ we may obtain from (4.35) a system of the form of (4.33). And therefore it is possible to expand the splitting and alternating direction algorithms described below for problems like (4.34) arising in the finite element method. Note also that we may arrive at systems of type (4.33) (or (4.34)) if we approximate the problem in spatial variables using Bubnov–Galerkin or Galerkin–Petrov methods or some other methods (see MARCHUK and AGOSHKOV [1981]). In this context it is of interest to formulate the splitting and the alternating direction algorithms for direct application to (4.33) as it will be done in the sequel.

PART 1

Algorithms for the Splitting Methods and the Alternating Direction Methods

In Part 1 of the paper some algorithms will be considered which are widely used currently for the solution of complex multidimensional problems of mathematical physics. In order to concentrate on the formulations of the algorithms we will consider as a basic object for study the evolutionary problem of mathematical physics given by

$$\frac{\partial \varphi}{\partial t} + A\varphi = f \quad \text{in } \Omega \times \Omega_t,$$

$$\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0$$

with a positive-semidefinite operator A (i.e., $A \geqslant 0$ or $(A\varphi, \varphi) \geqslant 0$). We suppose here that the approximation to the original problem in all variables except t has already been carried out, Ω is a grid domain, A is matrix, φ, t and g are grid functions (see Section 1). For the sake of simplicity the index h of the grid parameter is omitted. We suppose that the solution of the problem satisfies already given boundary conditions on $\partial\Omega$, and A and f are constructed with this fact taken into account. Let also the spaces Φ and F of grid functions coincide and, if not otherwise specified, have the same norm $\|\varphi\|_\Phi = (\varphi, \varphi)_\Phi^{1/2}$.

This problem may also be written in the form

$$\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t,$$

$$\varphi = g \quad \text{for } t = 0$$

implying that the first equation is considered in $\Omega \times \Omega_t$ and the second in $\Omega \times \{0\}$. Let us agree to use mostly the latter form.

Note that if it is desirable we may consider the above equations as the equations of the original problem with no preliminary approximation, and that the algorithms formulated below are applied directly to this problem. However, the description of such algorithms should often be considered as formal and their theoretical grounding in such cases is far more complicated.

CHAPTER II

Componentwise Splitting (Fractional Steps) Methods

The splitting (fractional steps) methods are based on the idea to split a complex operator in a sequence of the simplest ones and as a result to reduce the integration of the initial equation to successively integrating equations of a more simple structure. Since the splitting schemes must satisfy the conditions of approximation and stability only on the whole, we have the possibility for a flexible construction of the schemes essentially for all basic equations of mathematical physics.

5. The splitting method based on implicit schemes of first-order accuracy

Consider the problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0, \end{aligned} \tag{5.1}$$

where

$$A = \sum_{\alpha=1}^{n} A_\alpha, \qquad A_\alpha \geqslant 0, \quad n \geqslant 2. \tag{5.2}$$

Suppose that operators A_α are time-independent. The splitting algorithm based on using implicit schemes of first-order accuracy in τ (MARCHUK [1980], p. 269) has the form:

$$\begin{aligned} &\frac{\varphi^{j+1/n} - \varphi^j}{\tau} + A_1 \varphi^{j+1/n} = 0, \\ &\quad \vdots \\ &\frac{\varphi^{j+1} - \varphi^{j+(n-1)/n}}{\tau} + A_n \varphi^{j+1} = 0, \\ &\quad j = 0, 1, \ldots, \\ &\varphi^0 = g. \end{aligned} \tag{5.3}$$

This algorithm is absolutely stable, and the system (5.3) approximates problem (5.1) with first-order accuracy in τ (MARCHUK [1980], p. 270).

REMARK 5.1. Let us agree that if there are no special notes the approximation of the schemes is studied on solutions which are sufficiently smooth.

For the inhomogeneous problem,

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t=0, \end{aligned} \tag{5.4}$$

the splitting scheme of type (5.3) is

$$\begin{aligned} &\frac{\varphi^{j+1/n}-\varphi^{j}}{\tau} + A_1\varphi^{j+1/n} = 0, \\ &\quad\vdots \\ &\frac{\varphi^{j+1}-\varphi^{j+(n-1)/n}}{\tau} + A_n\varphi^{j+1} = f^j, \\ &\quad j=0,1,\ldots, \\ &\varphi^0 = g. \end{aligned} \tag{5.5}$$

This scheme is absolutely stable and has first-order accuracy in τ, when the following estimate holds:

$$\|\varphi^{j+1}\|_\Phi \leqslant \|g\|_G + \tau j \max_j \|f^j\|_F. \tag{5.6}$$

The realization of algorithms (5.3) and (5.5) consists of successively solving the equations of (5.3) and (5.5). And if A is split into A_α ($\alpha=1,\ldots,n$) so that it is easy to invert the operators $(E+\tau A_\alpha)$ (for example, if A_α and $(E+\tau A_\alpha)$ are three-diagonal or triangular matrices), then it is easy to find an approximate solution φ^{j+1} of the problem corresponding to $t=t_{j+1}$.

Algorithms (5.3) and (5.5) permit obvious generalizations in the case, where A is time-dependent. Then, in the cycle of computations of the splitting scheme, instead of A, a suitable approximation A^j of A should be taken on each interval $t_j \leqslant t \leqslant t_{j+1}$.

6. The componentwise splitting method based on Crank–Nicolson schemes: The case $A=A_1+A_2$

We suppose that the operator $A=A(t)$ in (5.1) is representable as the sum of two positive-semidefinite matrices:

$$A(t)=A_1(t)+A_2(t), \qquad A_1(t)\geqslant 0, \quad A_2(t)\geqslant 0, \tag{6.1}$$

which are sufficiently smooth along t-elements. We consider the approximation of these matrices on the interval $t_j \leqslant t \leqslant t_{j+1}$ in the form

$$A_\alpha^j = A_\alpha(t_{j+1/2}), \quad \alpha=1,2. \tag{6.2}$$

The componentwise splitting scheme for problem (5.1) (MARCHUK [1980], p. 256) has the form

$$\begin{aligned}
&\frac{\varphi^{j+1/2}-\varphi^j}{\tau}+\Lambda_1^j\frac{\varphi^{j+1/2}+\varphi^j}{2}=0,\\
&\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau}+\Lambda_2^j\frac{\varphi^{j+1}+\varphi^{j+1/2}}{2}=0,\\
&j=0,1,\ldots,\\
&\varphi^0=g.
\end{aligned}\tag{6.3}$$

By excluding auxiliary functions $\varphi^{j+1/2}$ the system of difference equations (6.3) may be reduced to one equation,

$$\varphi^{j+1}=T^j\varphi^j,\tag{6.4}$$

where

$$\begin{aligned}
T^j=&(E+\tfrac{1}{2}\tau\Lambda_2^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_2^j)\\
&\cdot(E+\tfrac{1}{2}\tau\Lambda_1^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_1^j).
\end{aligned}\tag{6.5}$$

Assume that

$$\tfrac{1}{2}\tau\|\Lambda_\alpha^j\|<1,\quad \alpha=1,2.\tag{6.6}$$

Then scheme (6.3) is absolutely stable. It has a second-order approximation in τ if Λ_1^j and Λ_2^j commute, and a first-order approximation in τ if they do not.

Indeed, expanding the operator T^j into a power series in τ we obtain

$$T^j=E-\tau\Lambda^j+\tfrac{1}{2}\tau^2((\Lambda_1^j)^2+2\Lambda_2^j\Lambda_1^j+(\Lambda_2^j)^2)-\cdots.$$

If the operators Λ_α^j commute, that is if $\Lambda_1^j\Lambda_2^j=\Lambda_2^j\Lambda_1^j$, then this expansion may be written as

$$T^j=E-\tau\Lambda^j+\tfrac{1}{2}\tau^2(\Lambda^j)^2-\cdots.$$

Then (6.4) may be represented as follows:

$$\varphi^{j+1}=\varphi^j-\tau\Lambda^j\varphi^j+\tfrac{1}{2}\tau(\Lambda^j)^2\varphi^j-\cdots\tag{6.7}$$

or

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda^j(\varphi^j-\tfrac{1}{2}\tau\Lambda^j\varphi^j+\cdots)=0.$$

Substituting the expression $\tau\Lambda^j\varphi^j$ from (6.7) into the last equality we have

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda^j\left(\frac{\varphi^{j+1}+\varphi^j}{2}-\tau^2(\Lambda^j)^2\varphi^j+\cdots\right)=0.$$

Comparing this scheme with the Crank–Nicolson scheme (1.15) we conclude that the order of approximation of (6.3) differs from the order of approximation of (1.15) by a quantity $O(\tau^2)$. Hence, the scheme (6.3) has a second-order approximation in

τ in the case $\Lambda_1^j \Lambda_2^j = \Lambda_2^j \Lambda_1^j$ and only a first-order approximation in τ in the case $\Lambda_1^j \Lambda_2^j \neq \Lambda_2^j \Lambda_1^j$.

The absolute stability of scheme (6.3) follows from the inequality,

$$\|T^j\| \leqslant \|(E+\tfrac{1}{2}\tau\Lambda_2^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_2^j)\| \cdot \|(E+\tfrac{1}{2}\tau\Lambda_1^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_1^j)\|,$$

and from the estimates

$$\|(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_\alpha^j)\| = \|(E-\tfrac{1}{2}\tau\Lambda_\alpha^j)(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)^{-1}\| \leqslant 1,$$

resulting from Theorem 2.1.

Realization of scheme (6.3) can be carried out in the following way:

$$\begin{aligned}
&\xi^{j+1/4} = (E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^j,\\
&(E+\tfrac{1}{2}\tau\Lambda_1^j)\xi^{j+2/4} = \xi^{j+1/4},\\
&\xi^{j+3/4} = (E-\tfrac{1}{2}\tau\Lambda_2^j)\xi^{j+2/4},\\
&(E+\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j+1} = \xi^{j+3/4},\\
&\quad j=0,1,\ldots,\\
&\varphi^0 = g.
\end{aligned} \tag{6.8}$$

If A is split into A_1 and A_2 in such a way that the effective solution of the equations with matrices $(E-\frac{1}{2}\tau\Lambda_\alpha^j)$ is possible, then the realization of the whole algorithm will also be effective.

REMARK 6.1. To formulate scheme (6.3) with a varying time step it is enough to substitute τ in (6.3) and (6.8) for $\tau_j = t_{j+1} - t_j$.

7. Multicomponent splitting of problems based on Crank–Nicolson schemes

We suppose that in (5.1) we have

$$A = \sum_{\alpha=1}^{n} A_\alpha, \qquad A_\alpha \geqslant 0, \quad n \geqslant 2. \tag{7.1}$$

We introduce on the interval $t_j \leqslant t \leqslant t_{j+1}$ the approximations Λ^j of the operators Λ_α^j so that

$$\Lambda^j = \sum_{\alpha=1}^{n} \Lambda_\alpha^j, \quad \Lambda_\alpha^j \geqslant 0. \tag{7.2}$$

The multicomponent splitting scheme based on elementary Crank–Nicolson scheme is the system of equations of the form

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+\alpha/n} = (E-\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+(\alpha-1)/n},\\
&\quad \alpha = 1,2,\ldots,n, \quad j=0,1,\ldots,\\
&\varphi^0 = g.
\end{aligned} \tag{7.3}$$

In the case where $\Lambda_\alpha^j \geqslant 0$ are commutative and

$$\Lambda_\alpha^j = A_\alpha(t_{j+1/2}) \quad \text{or} \quad \Lambda_\alpha^j = \tfrac{1}{2}(A_\alpha(t_{j+1}) + A_\alpha(t_j)), \tag{7.4}$$

the given scheme is absolutely stable and has a second-order approximation in τ. For noncommutative operators Λ_α^j scheme (7.3) will be, generally speaking, the scheme of first-order accuracy in τ (MARCHUK [1980], p. 266–267).

8. A general approach to componentwise splitting based on elementary Crank–Nicolson schemes

Let for problem (5.1) the assumptions (7.1) hold where the A_α are time-independent and the solution φ is sufficiently smooth in time. Suppose that (5.1) is approximated in a weak sense (weak approximation will be studied in Part 2 of this paper) on each interval $\theta_j = \{t_j \leqslant t \leqslant t_{j+1}\}$ by the problem

$$\begin{aligned} &\partial\varphi_\alpha/\partial t + A_\alpha\varphi_\alpha = 0 \quad \text{in } \theta_j, \\ &\varphi_\alpha^j = \varphi_{\alpha-1}^{j+1} \quad \text{for } t = t_j, \end{aligned} \qquad \alpha = 1, \ldots, n. \tag{8.1}$$

We introduce the notations:

$$\varphi_0^{j+1} = \varphi^j, \qquad \varphi_n^{j+1} = \varphi^{j+1}. \tag{8.2}$$

Applying a Crank–Nicolson scheme to each of the equations in (8.1) we obtain the system of difference equations:

$$\frac{\varphi^{j+\alpha/n} - \varphi^{j+(\alpha-1)/n}}{\tau} + A_\alpha \frac{\varphi^{j+\alpha/n} + \varphi^{j+(\alpha-1)/n}}{2} = 0, \quad \alpha = 1, \ldots, n, \tag{8.3}$$

where

$$\varphi^{j+\alpha/n} = \varphi_\alpha^{j+1}, \qquad \varphi^{j+1} = \varphi_n^{j+1}. \tag{8.4}$$

We suppose that each operator A_α is in its turn representable in the form

$$A_\alpha = \sum_{\beta=1}^{m_\alpha} A_{\alpha\beta} \tag{8.5}$$

where $A_{\alpha\beta} \geqslant 0$.

REMARK 8.1. The question arises: Is it worthwhile to split first the operator A into A_α and then in turn the operators A_α into $A_{\alpha\beta}$? Is it not easier to represent the operator A as a set of operators $A_{\alpha\beta}$ right away? In this context it should be remarked that though these two approaches seem equivalent, in many cases it is more convenient first to decompose a complex problem of mathematical physics into simpler ones which further can be independently reduced to elementary problems.

Taking into account (8.5) we state for system (8.1) the following splitting scheme

$$\frac{\varphi_\alpha^{j+\beta/m\alpha} - \varphi_\alpha^{j+(\beta-1)/m\alpha}}{\tau} + A_{\alpha\beta} \frac{\varphi_\alpha^{j+\beta/m\alpha} + \varphi_\alpha^{j+(\beta-1)/m\alpha}}{2} = 0, \tag{8.6}$$

$$\alpha = 1, 2, \ldots, n, \quad \beta = 1, 2, \ldots, m_\alpha,$$

where

$$\varphi_1^j=\varphi^j, \qquad \varphi_\alpha^j=\varphi_{\alpha-1}^{j+1} \quad (\alpha>1),$$
$$\varphi_n^{j+1}=\varphi^{j+1}.$$

This scheme approximates system (8.1) with second-order accuracy in τ, if the operators $A_{\alpha\beta}$ commute (MARCHUK [1980], p. 274). This result remains true in the case where operators $A_{\alpha\beta}$ are time-dependent. In this case it is necessary to construct a second-order approximation in τ of the operators $A_{\alpha\beta}=\Lambda_{\alpha\beta}^j$ on each interval $t_j\leqslant t\leqslant t_{j+1}$.

9. A general formulation for the splitting method based on multi-level schemes

Consider the problem

$$\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega_t, \tag{9.1}$$
$$\varphi=g \quad \text{for } t=0,$$

where

$$A(t)=\sum_{\alpha=1}^{n} A_\alpha(t)$$

is a representation of the operator $A(t)$ as the sum of n operators $A_1,\ldots,A_n$. Approximate the operators $A_1,\ldots,A_n$ by operators $\Lambda_{\alpha\beta}^j$ on the interval $t_j\leqslant t\leqslant t_{j+1}$ in such a way that the operator $\Sigma_{\beta=0}^{\alpha}\Lambda_{\alpha\beta}^j$ approximates the operator A_α. We write the following splitting scheme:

$$\frac{\varphi^{j+1/n}-\varphi^j}{\tau}+\Lambda_{10}^j\varphi^j+\Lambda_{11}^j\varphi^{j+1/n}=F_1, \tag{9.2}$$

$$\frac{\varphi^{j+2/n}-\varphi^{j+1/n}}{\tau}+\Lambda_{20}^j\varphi^j+\Lambda_{21}^j\varphi^{j+1/n}+\Lambda_{22}^j\varphi^{j+2/n}=F_2,$$

$$\vdots$$

$$\frac{\psi^{j+1}-\varphi^{j+(n-1)/n}}{\tau}+\Lambda_{n0}^j\varphi^j+\Lambda_{n1}^j\varphi^{j+1/n}+\cdots+\Lambda_{nn}^j\varphi^{j+1}=F_n,$$

where

$$F_\alpha=B_\alpha f$$

and the sum $\Sigma_{\alpha=1}^n B_\alpha$ approximates the identity operator E. Notice that if

$$\Lambda_{\alpha\beta}^j=0 \quad \text{for } \alpha<\beta-1, \tag{9.3}$$

then we obtain a two-level splitting scheme. In particular, the homogeneous

two-level scheme ($f\equiv 0$, $F_\alpha=0$, $\Lambda_{\alpha\beta}=0$ for $\alpha<\beta-1$) has the form

$$\frac{\varphi^{j+k/n}-\varphi^{j+(k-1)/n}}{\tau}+\Lambda_{kk-1}\varphi^{n+(k-1)/n}+\Lambda_{kk}\varphi^{j+k/n}=0, \tag{9.4}$$

$$k=1,2,\ldots,n.$$

If the operators Λ^j_{kk} and $\Lambda^j_{k,k-1}$ are commutative, then the scheme (9.4) approximates problem (9.1) and is stable provided that

$$\|C_nC_{n-1}\cdots C_1\|\leqslant 1+\text{const}\,\tau, \tag{9.5}$$

where

$$C_k=(E+\tau\Lambda^j_{kk})^{-1}(E-\tau\Lambda^j_{kk-1}),$$

$$k=1,2,\ldots,n$$

(see YANENKO [1967], p. 151).

REMARK 9.1. It follows from (9.5) that for the stability of the whole scheme the stability of each elementary scheme from (9.4) is not necessary at all.

REMARK 9.2. In the case where all operators $\Lambda^j_{\alpha\beta}$ and B_α are commutative the statements about the approximation of scheme (9.2) and its stability (and therefore the convergence conditions; see Theorem 3.1 in the Introduction) may be formulated as well.

REMARK 9.3. In scheme (9.2) the operators $\Lambda^j_{\alpha\beta}$ may have an arbitrary structure including difference, differential and integral-differential operators. The schemes of type (9.2) were developed in the works by YANENKO [1960, 1967] and by MARCHUK and YANENKO [1966] in which the theory and applications of such schemes are studied in more detail.

REMARK 9.4. If we take in (9.2) that $\Lambda^j_{kk}=\Lambda^j_k$ is the approximation of operator $A_k(t)$ (and $\Lambda^j_{\alpha\beta}\equiv 0$ for $\alpha\neq\beta$) and $B_\alpha=0$ for $\alpha<n$, $B_n=1$, then we obtain scheme (5.5).

10. The two-level splitting scheme with weight coefficients

Let the operators Λ^j_{kk-1}, Λ^j_{kk} in scheme (9.2) have the form

$$\begin{aligned}&\Lambda^j_{kk}=\sigma\Lambda^j_k,\\ &\Lambda^j_{kk-1}=(1-\sigma)\Lambda^j_k,\\ &0<\sigma=\text{const}<1,\end{aligned} \tag{10.1}$$

where Λ^j_k is the approximation of the operator $A_k(t)$ on $t_j\leqslant t\leqslant t_{j+1}$, and other operators $\Lambda^j_{\alpha\beta}$ are zero operators. Then we obtain the splitting scheme with weight

coefficients

$$\begin{aligned}
&\frac{\varphi^{j+1/n}-\varphi^{j}}{\tau}+\Lambda_1^j((1-\sigma)\varphi^j+\sigma\varphi^{j+1/n})=F_1,\\
&\frac{\varphi^{j+2/n}-\varphi^{j+1/n}}{\tau}+\Lambda_2^j((1-\sigma)\varphi^{j+1/n}+\sigma\varphi^{j+2/n})=F_2,\\
&\quad\vdots\\
&\frac{\varphi^{j+1}-\varphi^{j+(n-1)/n}}{\tau}+\Lambda_n^j((1-\sigma)\varphi^{j+(n-1)/n}+\sigma\varphi^{j+1})=F_n,\\
&\quad j=0,1,\ldots,\\
&\varphi^0=g.
\end{aligned}\tag{10.2}$$

If we take in (10.2) $\sigma=\frac{1}{2}$, then in the case where the operators Λ_k^j are commutative and Λ_k^j approximates A_k with the order $O(\tau^2)$, the whole scheme (10.2) will have a second-order approximation in τ as well. Besides, if $\Lambda_k^j\geqslant 0$ then the scheme will be absolutely stable.

11. The splitting method for systems which do not belong to the Cauchy–Kovalevskaya class

The splitting method allows a special formulation for systems which do not belong to the Cauchy–Kovalevskaya class (irregular systems) (see MARCHUK [1964, 1967], YANENKO [1967], MARCHUK and YANENKO [1966]). Let

$$\partial\varphi/\partial t+A\varphi+L\psi=f,\quad A=\sum_{\alpha=1}^{n}A_\alpha \tag{11.1}$$

and

$$K_1\varphi+K_2\psi+g=0 \tag{11.2}$$

be irregular system for which

$$\partial\varphi/\partial t+A\varphi=F,\quad F=f-L\psi \tag{11.3}$$

is regular for a given subsystem F. For integrating (11.1) and (11.2) the following scheme may be used:

$$\begin{aligned}
&\frac{\varphi^{j+1/n}-\varphi^{j}}{\tau}+\Lambda_{10}^j\varphi^j+\Lambda_{11}^j\varphi^{j+1/n}=B_1F^j,\\
&\frac{\varphi^{j+2/n}-\varphi^{j+1/n}}{\tau}+\Lambda_{20}^j\varphi^j+\Lambda_{21}^j\varphi^{j+1/n}+\Lambda_{22}^j\varphi^{j+2/n}=B_2F^j,\\
&\quad\vdots\\
&\frac{\varphi^{j+1}-\varphi^{j+(n-1)/n}}{\tau}+\Lambda_{n0}^j\varphi^j+\Lambda_{n1}^j\varphi^{j+1/n}+\cdots+\Lambda_{nn}^j\varphi^{j+1}=B_nF^{j+1},
\end{aligned}\tag{11.4}$$

where the operators $A_{\alpha\beta}^{j}$ and B_{α} were defined in the formulation of (9.2). While solving system (11.4) on the last fractional step (i.e., the last equation from (11.4)) it is necessary to use for definiteness of the algorithm the difference analogue of the relationship (11.2)

$$K_1^j \varphi^{j+1} + K_2^j \psi^{j+1} + g^{j+1} = 0, \tag{11.5}$$

where K_α^j is the approximation of operator K_α, g^{j+1} approximates g, and F^{j+1} (from (11.4)) is the approximation of $(f - L\psi)$.

12. Splitting schemes for the heat conduction equation: Local one-dimensional schemes

We consider some of the splitting schemes that are applied to the heat conduction equation.

12.1. The splitting scheme for the heat conduction equation in an orthogonal coordinate system

We consider the three-dimensional problem for the heat conduction equation:

$$\begin{aligned} &\partial\varphi/\partial t - \Delta\varphi = 0 \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = 0 \quad \text{on } \partial\Omega, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0, \end{aligned} \tag{12.1}$$

where

$$\Omega = \{(x, y, z): 0 < x, y, z < 1\}, \qquad \Omega_t = (0, T),$$

g is a given function, and

$$\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2.$$

Approximating (12.1) by the finite difference method in variables x, y and z in the same way as it was done in problems (1.5) and (1.19) and taking into acount the given boundary conditions from (12.1) we arrive at the matrix evolutionary equation

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0, \end{aligned} \tag{12.2}$$

where

$$A = A_x + A_y + A_z,$$

$$A_x = -(\Delta_x \nabla_x) \equiv \Lambda_1, \qquad A_y = -(\Delta_y \nabla_y) \equiv \Lambda_2, \qquad A_z = -(\Delta_z \nabla_z) \equiv \Lambda_3,$$

and g and φ are vectors ($\varphi = \varphi(t)$ here takes into the account given boundary conditions). The operator A is considered in the Hilbert space $\Phi = F$ with norm $\|\varphi\|_\Phi = (\varphi, \varphi)_\Phi^{1/2}$ defined as

$$\|\varphi\|_\Phi = \left(\sum_{k=1}^{N_x - 1} \sum_{l=1}^{N_y - 1} \sum_{p=1}^{N_z - 1} h_x h_y h_z \cdot \varphi_{k,l,p}^2 \right)^{1/2}.$$

Scheme (5.5) applied to problem (12.1) has the form (YANENKO [1969a]

$$\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau}+\Lambda_1\varphi^{j+1/3}=0,$$

$$\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau}+\Lambda_2\varphi^{j+2/3}=0,$$

$$\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau}+\Lambda_3\varphi^{j+1}=0, \tag{12.3}$$

$$j=0,1,\ldots,$$

$$\varphi^0=g.$$

Each equation from (12.3) may be easily solved by the factorization (sweep) method. Scheme (12.3) is absolutely stable and has a first-order approximation in τ, and therefore the convergence theorem holds (of the type of Theorem 3.1).

For increasing the accuracy of scheme (12.3) the weighted scheme (10.2) may be applied (YANENKO [1967], p. 33)

$$\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau}+\Lambda_1((1-\sigma)\varphi^{j}+\sigma\varphi^{j+1/3})=0,$$

$$\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau}+\Lambda_2((1-\sigma)\varphi^{j+1/3}+\sigma\varphi^{j+2/3})=0,$$

$$\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau}+\Lambda_3((1-\sigma)\varphi^{j+2/3}+\sigma\varphi^{j+1})=0, \tag{12.4}$$

$$j=0,1,\ldots,$$

$$\varphi^0=g.$$

With $\sigma=\frac{1}{2}$ the scheme has a second-order approximation in τ, and the whole order of accuracy of this scheme is $O(\tau^2+h^2)$, where $h=\max(h_x,h_y,h_z)$. It is absolutely stable. Equations (12.4) again may be solved by a one-dimensional sweep method.

12.2. The splitting scheme for the heat conduction equation in an arbitrary coordinate system

We consider a problem

$$\partial\varphi/\partial t+A\varphi=0 \quad \text{in } \Omega\times\Omega_t,$$

$$\varphi=0 \quad \text{on } \partial\Omega, \tag{12.5}$$

$$\varphi=g \quad \text{in } \Omega \qquad \text{for } t=0,$$

where Ω is a bounded two-dimensional domain and the differential operator A has the form

$$A=\sum_{i,j=1}^{2} a_{ij}\frac{\partial^2}{\partial x_i\partial x_j}, \quad a_{ij}=\text{const}, \tag{12.6}$$

$$a_{11}a_{22}-a_{12}^2>0.$$

We introduce in Ω a rectangular grid with step h_i in the x_i-direction ($i=1,2$) and the following approximations

$$\Lambda_{11}=-a_{11}\frac{(\Delta_{x_1}\nabla_{x_1})}{h_1^2} \quad \text{for } A_{11}=-a_{11}\frac{\partial^2}{\partial x_1^2},$$

$$\Lambda_{12}=\Lambda_{21}=-a_{12}\frac{(\Delta_{x_1}+\nabla_{x_1})(\Delta_{x_2}+\nabla_{x_2})}{4h_1h_2} \quad \text{for } A_{12}=-a_{12}\frac{\partial^2}{\partial x_1\partial x_2},$$

$$\Lambda_{22}=-a_{22}\frac{(\Delta_{x_2}\nabla_{x_2})}{h_2^2} \quad \text{for } A_{22}=-a_{22}\frac{\partial^2}{\partial x_2^2}$$

(taking into account the boundary conditions). We write the following scheme (YANENKO, SUCHOV and POGODIN [1959])

$$\begin{aligned}&\frac{\varphi^{j+1/2}-\varphi^j}{\tau}+\Lambda_{11}\varphi^{j+1/2}+\Lambda_{12}\varphi^j=0,\\&\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau}+\Lambda_{21}\varphi^{j+1/2}+\Lambda_{22}\varphi^{j+1}=0,\end{aligned} \tag{12.7}$$

which is a special case of the splitting scheme (9.2). It is not difficult to notice that here in the first fractional step "half" of the operator A, i.e. $A_{11}+A_{12}$, is approximated, where A_{11} is approximated on the time layer $j+\frac{1}{2}$ and A_{12} is approximated on the layer j; in the second fractional step the second "half" of A, i.e., $A_{21}+A_{22}$, is approximated, where $A_{21}=A_{12}$ again is approximated on the layer $j+\frac{1}{2}$ and A_{22} is approximated on the layer $j+1$. Scheme (12.7) approximates problem (12.5) and is stable, therefore the convergence theorem holds.

Under stronger conditions than the ellipticity condition for the three-dimensional heat conduction equation given by

$$\frac{\partial\varphi}{\partial t}-\sum_{i,j=1}^{3}a_{ij}\frac{\partial^2\varphi}{\partial x_i\partial x_j}=0. \tag{12.8}$$

YANENKO [1959b, 1964] suggested the scheme of the form

$$\begin{aligned}&\frac{\varphi^{j+1/6}-\varphi^j}{\tau}+\tfrac{1}{2}\Lambda_{11}\varphi^{j+1/2}+\Lambda_{12}\varphi^j=0,\\&\frac{\varphi^{j+2/6}-\varphi^{j+1/6}}{\tau}+\Lambda_{21}\varphi^{j+1/6}+\tfrac{1}{2}\Lambda_{22}\varphi^{j+2/6}=0,\\&\frac{\varphi^{j+3/6}-\varphi^{j+2/6}}{\tau}+\tfrac{1}{2}\Lambda_{11}\varphi^{j+3/6}+\Lambda_{13}\varphi^{j+2/6}=0,\\&\frac{\varphi^{j+4/6}-\varphi^{j+3/6}}{\tau}+\Lambda_{31}\varphi^{j+3/6}+\tfrac{1}{2}\Lambda_{33}\varphi^{j+3/6}=0,\\&\frac{\varphi^{j+5/6}-\varphi^{j+4/6}}{\tau}+\tfrac{1}{2}\Lambda_{22}\varphi^{j+5/6}+\Lambda_{23}\varphi^{j+4/6}=0,\\&\frac{\varphi^{j+1}-\varphi^{j+5/6}}{\tau}+\Lambda_{32}\varphi^{j+5/6}+\tfrac{1}{2}\Lambda_{33}\varphi^{j+1}=0.\end{aligned} \tag{12.9}$$

This scheme approximates (12.8) and is stable if the matrix with elements $b_{ij}=a_{ij}$ for $i\neq j$ and $b_{ii}=\frac{1}{2}a_{ii}$ is positive-definite.

12.3. *Local one-dimensional schemes*

We consider problem (5.1) and system (8.1) which approximates (5.1). If in (8.1) the operators A_α are one-dimensional differential operators (or their approximations), then the corresponding difference scheme is also called *local one-dimensional* (SAMARSKII [1971], p. 407). Thus, if in (8.3) the operators A_α are one-dimensional then this scheme may be called local one-dimensional.

The theory of local one-dimensional schemes for some differential equations is presented in the works by SAMARSKII [1971] and SAMARSKII and NIKOLAEV [1978]. Here we will consider only one such scheme in application to the following problem of the heat conduction equation:

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega\times\Omega_t,\\ &\varphi|_{\partial\Omega} = \varphi_{(\Gamma)}(x,t),\\ &\varphi = g(x) \quad \text{in } \Omega \qquad \text{for } t=0, \end{aligned} \tag{12.10}$$

where

$$A=\sum_{\alpha=1}^{n} A_\alpha, \quad A_\alpha = -\partial^2/\partial x_\alpha^2,$$

$$x=(x_1,\ldots,x_n)\in\Omega=\{0<x_\alpha<1:\alpha=1,\ldots,n\}.$$

We suppose that problem (12.10) has a unique sufficiently smooth solution.

In the layer $t_j\leqslant t\leqslant t_{j+1}$ we will solve instead of (12.10) the sequence of equations

$$\begin{aligned} &\partial\varphi_\alpha/\partial t + A_\alpha\varphi_\alpha = f_\alpha, \qquad x\in\Omega, \quad t\in\theta_j\\ &\alpha=1,\ldots,n, \quad \theta_j=\{t_j\leqslant t\leqslant t_{j+1}\} \end{aligned} \tag{12.11}$$

with the initial conditions

$$\begin{aligned} &\varphi_1(x,0)=g,\\ &\varphi_1(x,t_j)=\varphi_n(x,t_j), \quad j=1,2,\ldots,\\ &\varphi_\alpha(x,t_j)=\varphi_{\alpha-1}(x,t_{j+1}), \qquad j=0,1,\ldots, \quad \alpha=2,3,\ldots,n, \end{aligned}$$

where we set

$$\varphi(x,t_{j+1})=\varphi_n(x,t_{j+1}).$$

Here the f_α are arbitrary functions such that $\Sigma_{\alpha=1}^{n} f_\alpha = f$. Boundary conditions for φ_α are set only on the part $\partial\Omega_\alpha$ of the boundary $\partial\Omega$ consisting of the sides $x_\alpha=0$ and $x_\alpha=1$. We introduce in Ω a uniform grid with step h along each variable. Define the difference approximation of the operator A_α as follows:

$$\Lambda_\alpha = -(\Delta_{x_\alpha}\nabla_{x_\alpha})/h^2, \quad \alpha=1,\ldots,n.$$

Coming from (12.11) to difference approximations of the problems in spatial

variables (when φ_α and f_α are vectors and Λ_α are matrices, Ω is a grid domain and $\partial\Omega_\alpha$ is the grid on the sides $x_\alpha=0$ and $x_\alpha=1$) and approximating in t with the help of a two-layer implicit scheme of first-order accuracy in τ as well, we obtain the following local one-dimensional scheme:

$$\frac{\varphi_\alpha^{j+1}-\varphi_\alpha^j}{\tau}+\Lambda_\alpha\varphi_\alpha^{j+1}=f_\alpha(t_{j+1/2}) \tag{12.12}$$

$$\alpha=1,\dots,n, \quad j=0,1,\dots$$

with the initial conditions

$$\begin{aligned}&\varphi_1^0=g, \qquad \varphi_1^j=\varphi_\alpha^j, \quad j=1,2,\dots,\\ &\varphi_\alpha^j=\varphi_{\alpha-1}^{j+1}, \qquad \alpha=2,3,\dots,n, \quad j=0,1,2,\dots\end{aligned} \tag{12.13}$$

and the boundary conditions

$$\varphi_\alpha^{j+1}|_{\partial\Omega_\alpha}=\varphi_{(\Gamma)}(t_{j+1}). \tag{12.14}$$

Each problem in (12.12)–(12.14) (for each fixed α) is a one-dimensional first boundary value problem and may be solved by the sweep method.

In order to find the approximate values of the solution of the initial problem on the time layer t_{j+1} using data from the layer t_j, it is necessary to solve successively a sequence of one-dimensional problems. The local one-dimensional scheme (12.12)–(12.14) is stable in the metrics

$$\|\varphi\|_C=\max_{x_i\in\Omega}|\varphi_i|$$

(i.e. uniformly stable) with respect to the initial and boundary data and to the right-hand side.

If problem (12.10) has a unique solution $\varphi=\varphi(x,t)$ continuous in $\Omega\times\Omega_t$ and there exist in $\Omega\times\Omega_t$ derivatives

$$\frac{\partial^2\varphi}{\partial t^2}, \quad \frac{\partial^4\varphi}{\partial x_\alpha^2\partial x_\beta^2}, \quad \frac{\partial^3\varphi}{\partial+\partial x_\alpha^2}, \quad \frac{\partial^2 f}{\partial x_\alpha^2}, \quad 0\leqslant\alpha,\beta\leqslant n,$$

then scheme (12.12)–(12.14) converges uniformly with rate $O(h^2+\tau)$ (it has first-order accuracy in τ and second-order accuracy in h) so that

$$\|\varphi^j-\varphi(t_j)\|_C\leqslant M(h^2+\tau), \quad j=1,2,\dots,$$

where $M=\text{const}$ does not depend on τ and h (SAMARSKII [1971], p. 423).

REMARK 12.1. SAMARSKII ([1971], p. 425) presented and studied the local one-dimensional schemes for the parabolic equation with varying coefficients in the domain of complex shape. The local one-dimensional schemes for the quasi-linear parabolic equation were considered by FRYAZINOV [1964]. SAMARSKII [1970, 1971] gives a bibliography of the works devoted to this class of schemes.

CHAPTER III

Two-Cycle Componentwise Splitting Methods

Let us construct a class of splitting methods, in which the requirement of commutativeness of the operators $\{A_\alpha\}$ in representation $A=\Sigma_{\alpha=1}^{n} A_\alpha$ is taken off.

13. Two-cycle componentwise splitting methods: The case $A=A_1+A_2$

Consider the problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t=0 \end{aligned} \tag{13.1}$$

under the assumptions (6.1). We will approximate operators $A_1(t)$ and $A_2(t)$ not on the interval $t_j \leqslant t \leqslant t_{j+1}$ (as in (6.3)) but on the interval $t_{j-1} \leqslant t \leqslant t_{j+1}$. Let $A_\alpha^j = A_\alpha(t_j)$.

Write the following two systems of the difference equations:

$$\begin{aligned} &\frac{\varphi^{j-1/2}-\varphi^{j-1}}{\tau} + A_1^j \frac{\varphi^{j-1/2}+\varphi^{j-1}}{2} = 0, \\ &\frac{\varphi^{j}-\varphi^{j-1/2}}{\tau} + A_2^j \frac{\varphi^{j}+\varphi^{j-1/2}}{2} = 0, \end{aligned} \tag{13.2}$$

$$\begin{aligned} &\frac{\varphi^{j+1/2}-\varphi^{j}}{\tau} + A_2^j \frac{\varphi^{j+1/2}+\varphi^{j}}{2} = 0, \\ &\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau} + A_1^j \frac{\varphi^{j+1}+\varphi^{j+1/2}}{2} = 0. \end{aligned} \tag{13.3}$$

The cycle of computations consists of the alternative application of the schemes (13.2) and (13.3). Excluding $\varphi^{j+1/2}$ from these schemes we get on the whole step of computations

$$\varphi^{j+1} = T^j \varphi^{j-1}, \tag{13.4}$$

where

$$\begin{aligned} T^j &= (E+\tfrac{1}{2}\tau\Lambda_1^j)^{-1}(E+\tfrac{1}{2}\tau\Lambda_1^j)(E+\tfrac{1}{2}\tau\Lambda_2^j)^{-1} \\ &\quad \times (E-\tfrac{1}{2}\tau\Lambda_2^j)(E+\tfrac{1}{2}\tau\Lambda_2^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_2^j)(E+\tfrac{1}{2}\tau\Lambda_1^j)^{-1} \\ &\quad \times (E-\tfrac{1}{2}\tau\Lambda_1^j) \\ &= E-2\tau\Lambda^j+\tfrac{1}{2}(2\tau)^2(\Lambda^j)^2-\cdots. \end{aligned} \tag{13.5}$$

Here and further we suppose that restrictions (6.6) are satisfied.

If $A_\alpha(t)\geqslant 0$, then the system of the difference equations (13.2)–(13.3) is absolutely stable and scheme (13.4) approximates the initial problem (13.1) with second-order accuracy in τ, provided that the solution φ of problem (5.1) and the elements of the matrices $\{A_\alpha(t)\}$ are sufficiently smooth.

Indeed, from the expansion (13.5) of the step operator T^j we conclude that it coincides up to a quantity of order $O(\tau^2)$ with the step operator of the Crank–Nicolson scheme applied to the doubled time interval. Hence, independently of commutativeness of the operators $\{A_\alpha\}$ scheme (13.4) has a second-order approximation in τ.

The stability of the scheme follows from the estimates

$$\|\varphi^{j+1}\| \leqslant \|T^j\|\,\|\varphi^{j-1}\| \leqslant \|\varphi^{j-1}\| \leqslant \cdots \leqslant \|g\|,$$

since $\|T^j\| \leqslant 1$ as follows from Theorem 2.1.

For the inhomogeneous problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t=0 \end{aligned} \tag{13.6}$$

the two-cycle componentwise splitting scheme has the form:

$$\begin{aligned} &(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-1/2} = (E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-1}, \\ &(E+\tfrac{1}{2}\tau\Lambda_2^j)(\varphi^j-\tau f^j) = (E-\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j-1/2}, \\ &(E+\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j+1/2} = (E-\tfrac{1}{2}\tau\Lambda_2^j)(\varphi^j+\tau f^j), \\ &(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+1} = (E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+1/2}, \end{aligned} \tag{13.7}$$

where $f^j = f(t_j)$. Solving these equations with respect to φ^{j+1} we obtain

$$\varphi^{j+1} = T^j\varphi^{j-1} + 2\tau T_1^j T_2^j f^j, \tag{13.8}$$

where

$$\begin{aligned} &T^j = T_1^j T_2^j T_2^j T_1^j, \\ &T_\alpha^j = (E+\tfrac{1}{2}\tau\Lambda_\alpha^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_\alpha^j). \end{aligned}$$

Scheme (13.7) approximates problem (13.6) with second-order accuracy in τ and is absolutely stable with the assumptions of Theorem 2.1 (MARCHUK [1980], p. 259).

The system of equations (13.7) may also be written in the following equivalent

form:

$$
\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-2/3}=(E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-1},\\
&(E+\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j-1/3}=(E-\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j-2/3},\\
&\varphi^{j+1/3}=\varphi^{j-1/3}+2\tau f^j,\\
&(E+\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j+2/3}=(E+\tfrac{1}{2}\tau\Lambda_2^j)\varphi^{j+1/3},\\
&(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+1}=(E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+2/3},
\end{aligned}
\tag{13.9}
$$

Excluding unknown quantities with fractional indices we come to the resolved equation in the form

$$
\varphi^{j+1}=T_1^jT_2^jT_2^jT_1^j\varphi^{j-1}+2\tau T_1^jT_2^jf^j, \tag{13.10}
$$

which coincides with (13.8). In some cases representation of equations in the form (13.9) is more preferable than in the form (13.7).

14. The two-cycle multicomponent splitting method

Suppose that in (13.1) the operator A has the form $A(t)=\Sigma_{\alpha=1}^{n}A_\alpha(t)$, $A_\alpha(t)\geqslant 0$. Let

$$
\Lambda=\sum_{\alpha=1}^{n}\Lambda_\alpha^j,
$$

where all operators Λ_α^j are positive-semidefinite: $\Lambda_\alpha^j\geqslant 0$.

Consider the special construction of the splitting method which solves the Cauchy problem for the positive-semidefinite and noncommutative operators Λ_α^j and has a second-order approximation. It appears to be in some sense the complete solution of the splitting problem. To construct such a scheme it is necessary to replace (7.3) by the scheme

$$
\varphi^j=\prod_{\alpha=1}^{n}T_\alpha^j\varphi^{j-1},\qquad \varphi^{j+1}=\prod_{\alpha=n}^{1}T_\alpha^j\varphi^j, \tag{14.1}
$$

where

$$
T_\alpha^j=(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)^{-1}(E-\tfrac{1}{2}\tau\Lambda_\alpha^j).
$$

This means algorithmically that first we solve the system of equations (7.3) on the interval $t_{j-1}\leqslant t\leqslant t_j$ for $\alpha=1, 2,\ldots,n$ and then the same system on the interval $t_j\leqslant t\leqslant t_{j+1}$, but in the backward sequence ($\alpha=n, n-1,\ldots,1$):

$$
\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+\alpha/n-1}=(E-\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+(\alpha-1)/n-1},\quad \alpha=1,2,\ldots,n,\\
&(E+\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+1-(\alpha-1)/n}=(E-\tfrac{1}{2}\tau\Lambda_\alpha^j)\varphi^{j+1-\alpha/n},\quad \alpha=n, n-1,\ldots 1.
\end{aligned}
\tag{14.2}
$$

Obviously, for the whole cycle (14.2) we have

$$
\Phi^{j+1}=T^j\Phi^{j-1}
$$

where

$$T^j=\prod_{\alpha=1}^{n} T_\alpha^j\cdot\prod_{\alpha=n}^{1} T_\alpha^j=E-2\tau\Lambda^j+\tfrac{1}{2}(2\tau)^2(\Lambda^j)^2+O(\tau^3).$$

Thus, on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$ scheme (14.2) has second-order accuracy in τ, provided that we take as Λ_α^j one of the next two operators:

$$\Lambda_\alpha^j=A_\alpha(t_j)+\frac{\tau}{2}\frac{\partial}{\partial t}A_\alpha(t_j), \tag{14.3}$$

$$\Lambda_\alpha^j=\tfrac{1}{2}(A_\alpha(t_{j+1})+A_\alpha(t_j)). \tag{14.4}$$

Besides, scheme (14.2) is absolutely stable for $\Lambda_\alpha^j\geqslant 0$ (MARCHUK [1980], p. 268).

For the inhomogeneous problem (13.6) where $A(t)\geqslant 0$ and

$$A(t)=\sum_{\alpha=1}^{n} A_\alpha(t),\quad A_\alpha(t)\geqslant 0$$

on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$ the two-cycle multicomponent splitting scheme has the form:

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-(n-1)/n}=(E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j-1},\\
&(E+\tfrac{1}{2}\tau\Lambda_n^j)(\varphi^j-\tau f^j)=(E-\tfrac{1}{2}\tau\Lambda_n^j)\varphi^{j-1/n},\\
&(E+\tfrac{1}{2}\tau\Lambda_n^j)\varphi^{j+1/n}=(E-\tfrac{1}{2}\tau\Lambda_n^j)(\varphi^j+\tau f^j),\\
&(E+\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+1}=(E-\tfrac{1}{2}\tau\Lambda_1^j)\varphi^{j+(n-1)/n},
\end{aligned} \tag{14.5}$$

where

$$\Lambda_\alpha^j=A_\alpha(t_j).$$

It is not difficult to see that this scheme with the necessary smoothness assumption has second-order approximation in τ and is absolutely stable (MARCHUK [1980], p. 269).

In the same way as in the case $n=2$, the multicomponent system of equations (14.5) may be written in an equivalent form

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_\alpha)\varphi^{j-(n+1-\alpha)/(n+1)}=(E-\tfrac{1}{2}\tau\Lambda_\alpha)\varphi^{j-(n+1-\alpha+1)/(n+1)},\\
&\quad\alpha=1,2,\ldots,n,\\
&\varphi^{j+1/(n+1)}=\varphi^{j-1/(n+1)}+2\tau f^j,\\
&(E+\tfrac{1}{2}\tau\Lambda_{n-\alpha+2})\varphi^{j+\alpha/(n+1)}=(E-\tfrac{1}{2}\tau\Lambda_{n-\alpha+2})\varphi^{j+(\alpha-1)/(n+1)},\\
&\quad\alpha=2,3,\ldots,n+1.
\end{aligned} \tag{14.6}$$

15. The two-cycle componentwise splitting method for quasi-linear problems

Consider an evolutionary problem with operator A which depends upon the time and upon the solution of the problem

$$\begin{aligned}
&\partial\varphi/\partial t+A(t,\varphi)\varphi=0\quad\text{in }\Omega\times\Omega_t\\
&\varphi=g\quad\text{in }\Omega\qquad\text{for } t=0.
\end{aligned} \tag{15.1}$$

Suppose that $A(t, \varphi)$ is nonnegative, sufficiently smooth and has the form

$$A(t, \varphi) = \sum_{\alpha=1}^{n} A_\alpha(t, \varphi), \quad A_\alpha(t, \varphi) \geqslant 0. \tag{15.2}$$

Suppose further that the solution φ is also a sufficiently smooth function of time. Consider on the interval $t_{j-1} \leqslant t \leqslant t_{j+1}$ the splitting scheme

$$\begin{aligned}
&\frac{\varphi^{j+1/n-1} - \varphi^{j-1}}{\tau} + A_1^j \frac{\varphi^{j+1/n-1} + \varphi^{j-1}}{2} = 0, \\
&\quad \vdots \\
&\frac{\varphi^{j} - \varphi^{j-1/n}}{\tau} + A_n^j \frac{\varphi^{j} + \varphi^{j-1/n}}{2} = 0, \\
&\frac{\varphi^{j+1/n} - \varphi^{j}}{\tau} + A_n^j \frac{\varphi^{j+1/n} + \varphi^{j}}{2} = 0, \\
&\quad \vdots \\
&\frac{\varphi^{j+1} - \varphi^{j+(n-1)/n}}{\tau} + A_1^j \frac{\varphi^{j+1} + \varphi^{j+(n-1)/n}}{2} = 0,
\end{aligned} \tag{15.3}$$

where

$$\begin{aligned}
&A^j = A(t_j, \tilde{\varphi}^j), \\
&\tilde{\varphi}^j = \varphi^{j-1} - \tau A(t_{j-1}, \varphi^{j-1})\varphi^{j-1}, \\
&\tau = t_j - t_{j-1}.
\end{aligned} \tag{15.4}$$

It may be proved by methods used earlier for linear operators depending only upon time that splitting scheme (15.3) under conditions (15.4) has a second-order approximation in τ and is absolutely stable. The splitting method for inhomogeneous quasi-linear equations is defined in a similar way. This opens up broad vistas for applications of the componentwise splitting schemes to the nonstationary quasi-linear problems in hydrodynamics, meteorology, oceanology and other important fields of natural sciences.

16. A general approach to the two-cycle componentwise splitting method

We saw in Section 8 that, if the evolutionary problem (5.1) where $A = \Sigma_{\alpha=1}^{n} A_\alpha$ is reduced to partial problems (8.1) and these are then regarded as the set of new evolutionary problems, then the approximation to the initial problem (5.1) will be of first-order accuracy in τ if at least one of the elementary evolutionary problems is reduced to a difference scheme of first-order accuracy. But, if each such problem has a second-order approximation, then in the framework of the two-cycle procedure on α and β we obtain an approximation of second-order accuracy in τ. Note, that if the operators $A_{\alpha\beta}$ are noncommutative then, without the two-cycle procedure, we obtain a first-order approximation to problem (5.1). In this case the initial problem

will be the following:

$$\begin{aligned} &\partial\varphi/\partial t+\sum_{\alpha=1}^{n} A_\alpha\varphi=0 \quad \text{in } \Omega\times\theta_j, \\ &\varphi=\varphi^j \quad \text{in } \Omega \qquad \text{for } t=t_j. \end{aligned} \tag{16.1}$$

Reduce problem (16.1) to the system

$$\partial\varphi_\alpha/\partial t+A_\alpha\varphi_\alpha=0, \quad \varphi_\alpha^j=\varphi_{\alpha-1}^{j+1}, \qquad \alpha=1, 2, \ldots, n. \tag{16.2}$$

Let $A_\alpha=\Sigma_{\beta=1}^{m_\alpha} A_{\alpha\beta}$. Then we will solve each problem in (16.2) with the two-cycle method:

$$\begin{aligned} &\frac{\varphi_\alpha^{j+\beta/2m_\alpha}-\varphi_\alpha^{j+(\beta-1)/2m_\alpha}}{\frac{1}{2}\tau}+A_{\alpha\beta}\frac{\varphi^{j+\beta/2m_\alpha}+\varphi^{j+(\beta-1)/2m_\alpha}}{2}=0, \\ &\beta=1, 2, \ldots, m_\alpha, \\ &\frac{\varphi_\alpha^{j+\beta/2m_\alpha}-\varphi_\alpha^{j+(\beta-1)/2m_\alpha}}{\frac{1}{2}\tau}+A_{\alpha,2m_\alpha+1-\beta}\frac{\varphi^{j+\beta/2m_\alpha}+\varphi^{j+(\beta-1)/2m_\alpha}}{2}=0, \\ &\beta=m_\alpha+1, m_\alpha+2, \ldots, 2m_\alpha. \end{aligned} \tag{16.3}$$

Initial conditions for each system in (16.3) will be respectively taken in the form

$$\varphi_1^j=\varphi^j, \qquad \varphi_\alpha^j=\varphi_{\alpha-1}^{j+1}, \quad \alpha=2, \ldots, n. \tag{16.4}$$

It is not difficult to check that on the interval $t_j\leqslant t\leqslant t_{j+1}$ problem (16.3) approximates problem (16.2) with accuracy up to τ^2.

In order to obtain the solution of problem (5.1) with accuracy up to τ^2 it is necessary to alternate the basic cycles as well. So instead of (16.2) on the interval $t_{j-1}\leqslant t\leqslant t_j$ there should be

$$\begin{aligned} &\partial\varphi_\alpha/\partial t+A_\alpha\varphi_\alpha=0, \quad \alpha=1, 2, \ldots, n, \\ &\varphi_1=\varphi^{j-1}, \qquad \varphi_\alpha^{j-1}=\varphi_{\alpha-1}^j, \quad \alpha>1, \\ &\varphi^j=\varphi_n^j; \end{aligned} \tag{16.5}$$

and on the interval $t_j\leqslant t\leqslant t_{j+1}$

$$\begin{aligned} &\partial\varphi_\alpha/\partial t+A_{n-\alpha+1}\varphi_\alpha=0, \quad \alpha=1, 2, \ldots, n, \\ &\varphi_1^j=\varphi^j, \qquad \varphi_\alpha^j=\varphi_{\alpha-1}^{j+1}, \quad \alpha>1, \\ &\varphi^{j+1}=\varphi_n^{j+1}. \end{aligned} \tag{16.6}$$

It is supposed that each of the problems (16.5) and (16.6) is solved with the two-cycle method of the kind (16.3). Note that with $A_{\alpha\beta}\geqslant 0$ the componentwise splitting method is absolutely stable.

Let us give the general splitting scheme for the inhomogeneous equation

$$\begin{aligned} &\partial\varphi/\partial t+\sum_{\alpha=1}^{n} A_\alpha\varphi=f, \\ &\varphi=g \quad \text{for } t=0 \end{aligned} \tag{16.7}$$

on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$ based on the two-cycle method.

On the interval $t_{j-1}\leqslant t\leqslant t_j$ we set

$$\begin{aligned} &\partial\varphi_\alpha/\partial t+A_\alpha\varphi_\alpha=0, \quad \alpha=1, 2, \ldots, n-1, \\ &\partial\varphi_n/\partial t+A_n\varphi_n=f+\tfrac{1}{2}\tau A_n f; \end{aligned} \tag{16.8}$$

and on the interval $t_j\leqslant t\leqslant t_{j+1}$

$$\begin{aligned} &\partial\varphi_{n+1}/\partial t+A_n\varphi_{n+1}=f-\tfrac{1}{2}\tau A_n f, \\ &\partial\varphi_{n+\alpha}/\partial t+A_{n-\alpha+1}\varphi_{n+\alpha}=0, \quad \alpha=2, 3, \ldots, n \end{aligned} \tag{16.9}$$

provided that

$$\begin{aligned} &\varphi_1(t_{j-1})=\varphi(t_{j-1}), \\ &\varphi_{\alpha+1}(t_{j-1})=\varphi_\alpha(t_j), \quad \alpha=1, 2, \ldots, n \end{aligned} \tag{16.10}$$

and, respectively,

$$\varphi_{\alpha+1}(t_j)=\varphi_\alpha(t_{j+1}), \quad \alpha=n+1, n+2, \ldots, 2n. \tag{16.11}$$

Using now Crank–Nicolson schemes for solving equations (16.8)–(16.11) on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$ with $f=f^j$ we obtain the systems (14.5) for (16.9).

Along with the system of equations (16.8) and (16.9) let us consider the following. On the interval $t_{j-1}\leqslant t\leqslant t_j$

$$\begin{aligned} &\partial\varphi_1/\partial t+A_1\varphi_1=0, \\ &\quad\vdots \\ &\partial\varphi_n/\partial t+A_n\varphi_n=0; \end{aligned} \tag{16.12}$$

on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$

$$\partial\varphi_{n+1}/\partial t=f; \tag{16.13}$$

and on the interval $t_j\leqslant t\leqslant t_{j+1}$

$$\begin{aligned} &\partial\varphi_{n+2}/\partial t+A_n\varphi_{n+2}=0, \\ &\quad\vdots \\ &\partial\varphi_{2n+1}/\partial t+A_1\varphi_{2n+1}=0. \end{aligned} \tag{16.14}$$

The initial conditions for system (16.12) will be

$$\begin{aligned} &\varphi_1(t_{j-1})=\varphi(t_{j-1}), \\ &\varphi_\alpha(t_{j-1})=\varphi_{\alpha-1}(t_j) \quad \alpha=2, 3, \ldots, n; \end{aligned} \tag{16.15}$$

for equation (16.13)

$$\varphi_{n+1}(t_{j-1})=\varphi_n(t_j); \tag{16.6}$$

and for the system of equations (16.14)

$$\varphi_\alpha(t_j)=\varphi_{\alpha-1}(t_{j+1}), \quad \alpha=n+2, n+3, \ldots, 2n+1. \tag{16.17}$$

The approximation and the stability of these schemes ensure the convergence of the approximate solution to the exact one (see Introduction).

17. The two-cycle componentwise splitting scheme for the heat conduction equation

Consider the two-dimensional heat conduction problem

$$\begin{aligned} &\partial\varphi/\partial t-\Delta\varphi=f \quad \text{in } \Omega\times\Omega_t,\\ &\varphi=0 \quad \text{on } \partial\Omega,\\ &\varphi=g \quad \text{in } \Omega \qquad \text{for } t=0, \end{aligned} \tag{17.1}$$

where

$$\Omega=\{0<x, y<1\}, \qquad \Omega_t=\{0<t<T<\infty\}.$$

Introducing the grid with steps $h_x=1/N_x$, $h_y=1/N_y$ along the x- and y-direction respectively and approximating problem (17.1) as it was described in the Introduction, we obtain the matrix evolutionary problem

$$\begin{aligned} &\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega\times\Omega_t,\\ &\varphi=g \quad \text{in } \Omega \qquad \text{for } t=0, \end{aligned} \tag{17.2}$$

where φ, f, g are vectors and A is a matrix:

$$A=A_1+A_2,$$

$$A_1\equiv A_x=-(\Delta_x\nabla_x)/h_x^2, \qquad A_2\equiv A_y=-(\Delta_y\nabla_y)/h_y^2.$$

We suppose that the boundary conditions are taken into account in (17.2) and that Ω is the grid domain consisting of the internal grid points of the initial domain.

The operators A_α, $\alpha=1, 2$, in (17.2) are positive-definite, $A_1>0$, $A_2>0$, so with $A_1\equiv\Lambda_1$, $A_2\equiv\Lambda_2$ the two-cycle componentwise splitting scheme (on the interval $t_{j-1}\leqslant t\leqslant t_{j+1}$)

$$\begin{aligned} &(E+\tfrac{1}{2}\tau\Lambda_1)\varphi^{j-1/2}=(E-\tfrac{1}{2}\tau\Lambda_1)\varphi^{j-1},\\ &(E+\tfrac{1}{2}\tau\Lambda_2)(\varphi^j-\tau f^j)=(E-\tfrac{1}{2}\tau\Lambda_2)\varphi^{j-1/2},\\ &(E+\tfrac{1}{2}\tau\Lambda_2)\varphi^{j+1/2}=(E-\tfrac{1}{2}\tau\Lambda_2)(\varphi^j+\tau f^j),\\ &(E+\tfrac{1}{2}\tau\Lambda_1)\varphi^{j+1}=(E-\tfrac{1}{2}\tau\Lambda_1)\varphi^{j+1/2},\\ &\quad j=0, 1, \ldots,\\ &\varphi^0=g \end{aligned} \tag{17.3}$$

is absolutely stable and has a second-order approximation in τ for smooth solutions. Equations (17.3) may be written in componentwise form as follows:

$$\left(1+\frac{\tau}{h_x^2}\right)\xi_{k,l}^{j-1/2}-\frac{\tau}{2h_x^2}(\xi_{k+1,l}^{j-1/2}+\xi_{k-1,l}^{j-1/2})=F_{k,l}^{j-1},$$

$$\left(1+\frac{\tau}{h_y^2}\right)\xi_{kl}^{j}-\frac{\tau}{2h_y^2}(\xi_{k,l+1}^{j}+\xi_{k,l-1}^{j})=F_{k,l}^{j-1/2},$$

$$\left(1+\frac{\tau}{h_y^2}\right)\xi_{k,l}^{j+1/2}-\frac{\tau}{2h_y^2}(\xi_{k,l+1}^{j+1/2}+\xi_{k,l-1}^{j+1/2})=F_{k,l}^{j+1/2}, \tag{17.4}$$

$$\left(1+\frac{\tau}{h_x^2}\right)\varphi_{k,l}^{j+1}-\frac{\tau}{2h_x^2}(\varphi_{k+1,l}^{j+1}+\varphi_{k-1,l}^{j+1})=F_{k,l}^{j+1},$$

$$\xi_{k,l}^{j-1/2}\equiv\varphi_{k,l}^{j-1/2},\qquad \xi_{k,l}^{j}\equiv\varphi_{k,l}^{j}-\tau f_{k,l}^{j},\qquad \xi_{k,l}^{j+1/2}\equiv\varphi_{k,l}^{j+1/2}$$

where

$$F_{k,l}^{j-1}=\left(1-\frac{\tau}{h_x^2}\right)\varphi_{k,l}^{j-1}+\frac{\tau}{2h_x^2}(\varphi_{k+1,l}^{j-1}+\varphi_{k-1,l}^{j-1}),$$

$$F_{k,l}^{j-1/2}=\left(1-\frac{\tau}{h_y^2}\right)\xi_{k,l}^{j-1/2}+\frac{\tau}{2h_y^2}(\xi_{k,l+1}^{j-1/2}+\xi_{k,l-1}^{j-1/2}),$$

$$F_{k,l}^{j+1/2}=\left(1-\frac{\tau}{2h_y^2}\right)(\xi_{k,l}^{j}+2\tau f_{k,l}^{j})+\frac{\tau}{2h_y^2}(\xi_{k,l+1}^{j}+2\tau f_{k,l+1}^{j}+\xi_{k,l-1}^{j}+2\tau f_{k,l-1}^{j}),$$

$$F_{k,l}^{j+1}=\left(1-\frac{\tau}{h_x^2}\right)\xi_{k,l}^{j+1/2}+\frac{\tau}{2h_x^2}(\xi_{k+1,l}^{j+1/2}+\xi_{k-1,l}^{j+1/2}).$$

Note, that the right-hand sides in (17.4) are calculated explicitly. Thus we deal here with a system of equations represented in the form of three-point schemes, and the solution of these equations may be easily and effectively obtained by the sweep method.

The two-cycle componentwise splitting method is used for solving parabolic equations in the papers by MARCHUK [1967], MARCHUK and SARKISYAN [1980] and others.

Note that in the paper by MARCHUK and KUZIN [1983] (see also Chapter XIV) this method is used for the realization of the finite element approximations of a two-dimensional parabolic equation with mixed derivatives. It is proved that the scheme is stable and has a second-order approximation.

CHAPTER IV

Splitting Schemes with Factorization of the Operators

18. Schemes factorizing the operators

Consider the evolutionary problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t=0 \\ &\varphi \in \Phi, \qquad f \in F, \end{aligned} \tag{18.1}$$

where Φ and F are Hilbert spaces and the difference operator A does not depend on time. Suppose that problem (18.1) is solved using a time difference scheme which may be written in the form:

$$B\varphi^j = F^j, \quad j=1, 2, \ldots, \tag{18.2}$$

where $F^j = F^j(\varphi^{j-1}, \varphi^{j-2}, \ldots; \tau)$ is a given function of τ, φ^{j-1}, $\varphi^{j-2}, \ldots$ (in case of a two-level scheme $F^j = F^j(\tau, \varphi^{j-1})$ and B is some operator constructed in such a way that it allows a factorized representation

$$B = B_1 B_2 \cdots B_p. \tag{18.3}$$

The operators B_α have a more simple structure than B and may be easily inverted. Difference schemes allowing representation (18.3) are called *factorized schemes* (BAKER and OLIPHANT [1960], SAMARSKII [1971], p. 368) or *factorization schemes* of the difference operator (YANENKO [1967], p. 37).

Numerical solution of equation (18.2) on each time level may be carried out by way of successive solution of more simple equations

$$\begin{aligned} &B_1 \varphi^{j+1/p} = F^j, \\ &B_2 \varphi^{j+2/p} = \varphi^{j+1/p}, \\ &\quad \vdots \\ &B_p \varphi^{j+1} = \varphi^{j+(p-1)/p}. \end{aligned} \tag{18.4}$$

If $p=2$ and B_1, B_2 are triangular matrices, then the relationships (18.4) represent an explicit ("running") calculation scheme (see SAMARSKII [1971], p. 368).

We now consider some ways of constructing factorized operators and corresponding factorized schemes applied to the problems of parabolic type.

BAKER and OLIPHANT [1960] consider for the problem

$$\begin{aligned} &\frac{\partial\varphi}{\partial t}-a^2\left(\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}\right)=0,\\ &\varphi=\varphi_{(\Gamma)} \quad \text{on } \partial\Omega,\\ &\varphi=g \qquad \text{for } t=0, \end{aligned} \tag{18.5}$$

where $\Omega=\{0<x,y<1\}$, and $a=\text{const}$, the three-layer scheme

$$\frac{\frac{3}{2}\varphi^{j+1}-2\varphi^j+\frac{1}{2}\varphi^{j-1}}{\tau}-\Lambda\varphi^{j+1}=0, \tag{18.6}$$

in which the operator Λ was constructed to approximate the operator $A=-a^2(\partial^2/\partial x^2+\partial^2/\partial y^2)$ with second-order accuracy; at the same time the operator

$$B=\tfrac{3}{2}E+\tau\Lambda \tag{18.7}$$

was represented as the product of two three-point operators B_1 and B_2. It was shown by BAKER and OLIPHANT [1960] that such an operator may be constructed; as a result the realization of scheme (18.6) is reduced to solving (18.2) with $F^j=F^j(\varphi^{j-1},\varphi^j;\tau)=2\varphi^j-\frac{1}{2}\varphi^{j-1}$. For $p=2$, (18.2) also is decomposed in two equations (18.4), which may be easily solved using the three-point sweep method.

BAKER [1960] describes the factorization schemes in the case of a multidimensional heat conduction equation with constant coefficients.

But if a two-dimensional parabolic equation with varying coefficients (diffusion equation) is considered, then it is impossible to represent B exactly as the product of two three-point operators. In this case additional iterations are needed (BULEYEV [1970], OLIPHANT [1961]). For such problems BULEYEV [1970] (see also MARCHUK [1980], pp. 235–237) formulated *a scheme of incomplete factorization.* Thus, let (18.2) be represented as a two-dimensional difference equation

$$\begin{aligned} &a_{i,k}\varphi_{i-1,k}+c_{i,k}\varphi_{i+1,k}+b_{i,k}\varphi_{i,k-1}+d_{i,k}\varphi_{i,k+1}-p_{i,k}\varphi_{i,k}=-f_{i,k},\\ &\quad i=1,2,\ldots,m, \quad k=1,2,\ldots,n,\\ &a_{1,k}=c_{m,k}=b_{i,1}=d_{i,n}=0,\\ &p_{i,k}\geqslant a_{i,k}+c_{i,k}+b_{i,k}+d_{i,k} \end{aligned} \tag{18.8}$$

(the index j is omitted for simplicity). Having added the vector $C\varphi$ to the left-hand and right-hand sides of the equation we obtain

$$(B+C)\varphi=C\varphi+F. \tag{18.9}$$

We now choose the matrix C in such a way that the matrix $B+C$ can be represented as the product of two simple triangular matrices S_1, S_2 and a diagonal matrix K:

$$B+C=K\cdot S_1\cdot S_2.$$

Then equation (18.9) is replaced by the system

$$S_1 Z = K^{-1}(C\varphi + F), \qquad S_2 \varphi = Z, \tag{18.10}$$

which will be solved by the method of successive approximations. The realization formulae of this iteration algorithm applied to (18.10) can be found in the form

$$\begin{aligned} Z_{i,k} &= \alpha_{i,k} Z_{i-1,k} + \beta_{i,k} Z_{i,k-1} + \gamma_{i,k}[f_{i,k} + D_{i,k}(\varphi) - q_{i,k}\varphi_{i,k}], \\ \varphi_{i,k} &= \varepsilon_{i,k}\varphi_{i+1,k} + \eta_{i,k}\varphi_{i,k+1} + Z_{i,k} \end{aligned} \tag{18.11}$$

(the index of the iteration step is omitted; $D_{i,k}(\varphi) = (D\varphi)_{i,k}$, where the matrix $D = KC + \operatorname{diag} q_{i,k}$ and $\operatorname{diag} q_{i,k}$ is a diagonal matrix with elements $q_{i,k}$).

Comparing (18.11) and (18.8) we obtain for coefficients $\alpha_{i,k}$, $\beta_{i,k}$, $\varepsilon_{i,k}$, $\eta_{i,k}$, $\gamma_{i,k}$ formulae which are analogous to the formulae of the sweep method:

$$\begin{aligned} &\alpha_{i,k} = \gamma_{i,k} a_{i,k}, \qquad \beta_{i,k} = \gamma_{i,k} b_{i,k}, \\ &\varepsilon_{i,k} = \gamma_{i,k} c_{i,k}, \qquad \eta_{i,k} = \gamma_{i,k} d_{i,k} \\ &\gamma_{i,k} = (p_{i,k} - q_{i,k} - a_{i,k} c_{i-1,k}\gamma_{i-1,k} - b_{i,k} d_{i,k-1}\gamma_{i,k-1})^{-1}. \end{aligned}$$

For the iteration operator $D_{i,k}(\varphi)$ we obtain the expression

$$D_{i,k}(\varphi) = a_{i,k}\eta_{i-1,k}\varphi_{i-1,k+1} + b_{i,k}\varepsilon_{i,k-1}\varphi_{i+1,k-1}.$$

The coefficient $q_{i,k}$ is taken as

$$q_{i,k} = \theta_{i,k}(a_{i,k}\eta_{i-1,k} + b_{i,k}\varepsilon_{i,k-1}), \quad 0 \leqslant \theta_{i,k} < 1.$$

To ensure the convergence of the iterative process (18.11) for $\theta > 0$ these equations are supplemented with a simple iteration

$$\begin{aligned} \tilde{\varphi}_{i,k}^{(l)} = \frac{1}{\gamma_{i,k} p_{i,k}} &(\alpha_{i,k}\varphi_{i-1,k}^{(l)} + \beta_{i,k}\varphi_{i,k-1}^{(l)} + \varepsilon_{i,k}\varphi_{i+1,k}^{(l)} \\ &+ \eta_{i,k}\varphi_{i,k+1}^{(l)}\gamma_{i,k} f_{i,k}), \end{aligned} \tag{18.12}$$

so that for obtaining the $(l+1)$th approximation of the function $\varphi_{i,k}$ the lth approximation of the function $\tilde{\varphi}_{i,k}$ is to be substituted into the equation for $Z_{i,k}^{(l+1)}$. If the problem is ill-conditioned, then the expression in square brackets in (18.11) for the grid points belonging to the right and upper boundaries of the domain should be replaced by

$$f_{i,k} + D_{i,k}(\varphi) - q_{i,k}\varphi_{i,k} + x_{i,k}(a_{i,k} + b_{i,k})\varphi_{i,k} \tag{18.13}$$

where $0 \leqslant x_{i,k} \leqslant 1$.

It is not difficult to show using induction method that the coefficients $\alpha_{i,k}$, $\beta_{i,k}$, $\varepsilon_{i,k}$, $\eta_{i,k}$, $\gamma_{i,k}$ satisfy the conditions

$$\begin{aligned} &\varepsilon_{i,k} + \eta_{i,k} \leqslant 1; \\ &\alpha_{i,k} + \beta_{i,k} \leqslant (a_{i,k} + b_{i,k}) \\ &\qquad \cdot \{c_{i,k} + d_{i,k} + (1 - \theta_{i,k})(\alpha_{i,k}\eta_{i-1,k} + \beta_{i,k}\varepsilon_{i,k-1}) \\ &\qquad + x_{i,k}(a_{i,k} + b_{i,k})\}^{-1}. \end{aligned}$$

Choosing parameters $\theta_{i,k}$ and $x_{i,k}$ it is always possible to satisfy the condition $\alpha_{i,n}+\beta_{i,n}\leqslant 1$, i.e. to achieve computational stability of the scheme (18.11). Note that MARCHUK ([1980], pp. 236–242) presented another scheme of incomplete factorization.

19. The implicit splitting scheme with approximate factorization of the operator

We consider for the problem (18.1) with operator $A>0$ the implicit scheme of the form

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda\varphi^{j+1}=0, \quad j=0,1,2,\ldots, \qquad \varphi^0=g, \tag{19.1}$$

where

$$\Lambda=\sum_{\alpha=1}^{n}\Lambda_\alpha, \quad \Lambda_\alpha\geqslant 0. \tag{19.2}$$

We rewrite (19.1) in the form

$$(E+\tau\Lambda)\varphi^{j+1}=\varphi^j \tag{19.3}$$

and factorize the operator $(E+\tau\Lambda)$ approximately up to quantities of order $O(\tau^2)$. To that end we replace the operator $(E+\tau\Lambda)$ in (19.3) with factorized operator

$$(E+\tau\Lambda_1)(E+\tau\Lambda_2)\cdots(E+\tau\Lambda_n)=E+\tau\Lambda+\tau^2R, \tag{19.4}$$

where

$$R=\sum_{i<j}\Lambda_i\Lambda_j+\tau\sum_{i<j<k}\Lambda_i\Lambda_j\Lambda_k+\cdots+\tau^{n-2}\Lambda_i\cdots\Lambda_n.$$

As a result we obtain an implicit scheme with an approximate factorization of the operator B (YANENKO [1967], p. 38):

$$B\varphi^{j+1}=\varphi^j, \quad j=0,1,2,\ldots, \qquad \varphi^0=g, \tag{19.5}$$

where

$$B=\prod_{\alpha=1}^{n}B_\alpha, \quad B_\alpha=E+\tau\Lambda_\alpha.$$

The solution of equations (19.5) may be carried out by successively solving equations

$$\begin{aligned}
&(E+\tau\Lambda_1)\varphi^{j+1/n}=\varphi^j,\\
&(E+\tau\Lambda_2)\varphi^{j+2/n}=\varphi^{j+1/n},\\
&\quad\vdots\\
&(E+\tau\Lambda_n)\varphi^{j+1}=\varphi^{j+(n-1)/n},\\
&\quad j=0,1,2,\ldots,\\
&\varphi^0=g.
\end{aligned} \tag{19.6}$$

If the operators Λ_α in representation (19.2) have a simple structure and allow cheap inversion, then system (19.6) may also be solved rather simply.

It is easy to see that scheme (19.6) has a first-order approximation in τ. It is absolutely stable for $\Lambda_\alpha \geqslant 0$ in the metrics of the grid function space Φ_h ($\|\varphi\|_{\Phi_h} = (\varphi, \varphi)_{\Phi_h}^{1/2}$), since, in that case, $\|E+\tau\Lambda_\alpha)^{-1}\|_{\Phi_h} \leqslant 1$ (see (2.23) and Theorem 2.1).

We will illustrate an application of the scheme with approximate factorization of the operator to scheme (18.6), which was suggested by BAKER and OLIPHANT [1960] for the heat conduction equation (18.5).

We take $\Lambda = \Lambda_1 + \Lambda_2$, where Λ_α is a difference approximation of the one-dimensional differential operator $A_\alpha = -a^2\partial^2/\partial^2 x_\alpha^2$ ($x_1 = x, x_2 = y$). We rewrite scheme (18.6) in the form

$$(\tfrac{3}{2}E + \tau\Lambda)\varphi^{j+1} = F^j, \quad F^j = 2\varphi^j - \tfrac{1}{2}\varphi^{j-1} \tag{19.7}$$

and replace the operator $(\frac{3}{2}E + \tau\Lambda)$ with the factorized one obtained by the implicit scheme with approximate factorization of the operator (YANENKO [1967], p. 39). We obtain the scheme

$$B\varphi^{j+1} \equiv \tfrac{3}{2}(E + \tfrac{2}{3}\tau\Lambda_1)(E + \tfrac{2}{3}\tau\Lambda_2)\varphi^{j+1} = F^j, \tag{19.8}$$

where B is a nine-point factorized operator. (Note, that scheme (19.8) coincides with the scheme suggested by BAKER and OLIPHANT [1960].)

20. The stabilization method (explicit–implicit scheme with approximate factorization of the operator)

We consider the homogeneous problem (18.1) with $A = A_1 + A_2$, $A_\alpha \geqslant 0$, $f \equiv 0$ and the following two-layer scheme

$$\begin{aligned} &(E + \tfrac{1}{2}\tau A_1)(E + \tfrac{1}{2}\tau A_2)\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^j = 0, \\ &\quad j = 0, 1, 2, \ldots, \\ &\varphi^0 = g. \end{aligned} \tag{20.1}$$

If $A_1 \geqslant 0$, $A_2 \geqslant 0$, and the solution of problem (18.1) is smooth enough, then the difference scheme (20.1) has a second-order approximation in τ and is absolutely stable, the following estimate being valid

$$\|\varphi^{j+1}\|_{C_2} \leqslant \|g\|_{C_2}, \quad j = 0, 1, 2, \ldots, \tag{20.2}$$

where

$$\|\psi\|_{C_2} = (C_2\psi, \psi)_\Phi^{1/2}, \quad C_2 = (E + \tfrac{1}{2}\tau A_2^*)(E + \tfrac{1}{2}\tau A_2).$$

Indeed, using simple transformations, equation (20.1) is reduced to the form

$$\begin{aligned} &(E + \tfrac{1}{4}\tau^2 A_1 A_2)\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\frac{\varphi^{j+1} + \varphi^j}{2} = 0, \\ &\varphi^0 = g. \end{aligned} \tag{20.3}$$

It is easy to notice that if the solution is sufficiently smooth then the order of approximation in τ of the difference equation (20.3) coincides with that of the Crank–Nicolson scheme

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\frac{\varphi^{j+1}+\varphi^j}{2}=0, \qquad \varphi^0=g, \tag{20.4}$$

i.e. second-order in τ.

We now move to the stability analysis of the difference equation (20.1). To this end we transform (20.1) to the form

$$\begin{aligned}&(E+\tfrac{1}{2}\tau A_1)(E+\tfrac{1}{2}\tau A_2)\varphi^{j+1}=(E-\tfrac{1}{2}\tau A_1)(E-\tfrac{1}{2}\tau A_2)\varphi^j,\\ &\varphi^0=g.\end{aligned} \tag{20.5}$$

Resolving the difference equation (20.5) with respect to φ^{j+1} we get

$$\varphi^{j+1}=(E+\tfrac{1}{2}\tau A_2)^{-1}(E+\tfrac{1}{2}\tau A_1)^{-1}(E-\tfrac{1}{2}\tau A_1)(E-\tfrac{1}{2}\tau A_2)\varphi^j.$$

From the unknowns φ^j we compute ψ^j using the following formula

$$\psi^j=(E+\tfrac{1}{2}\tau A_2)\varphi^j.$$

Then for a new unknown ψ^j we obtain the relationship

$$\psi^{j+1}=T\psi^j,$$

where the transition operator T is given by

$$T=(E+\tfrac{1}{2}\tau A_1)^{-1}(E-\tfrac{1}{2}\tau A_1)(E-\tfrac{1}{2}\tau A_2)(E+\tfrac{1}{2}\tau A_2)^{-1}\equiv\prod_{\alpha=1}^{2}T_\alpha,$$

$$T_\alpha=(E-\tfrac{1}{2}\tau A_\alpha)(E+\tfrac{1}{2}\tau A_\alpha)^{-1}.$$

It follows from Theorem 2.1 that

$$\|T_\alpha\|=\sup_{\varphi}\|T_\alpha\varphi\|_\Phi/\|\varphi\|_\Phi\leqslant 1, \quad \|\varphi\|_\Phi=(\varphi,\varphi)_\Phi^{1/2}.$$

Therefore, $\|T\|\leqslant 1$, which means that

$$\|\psi^j\|_\Phi=\|\varphi^j\|_{C_2}\leqslant\|\psi^{j-1}\|_\Phi\leqslant\cdots\leqslant\|g\|_{C_2}.$$

The difference scheme (20.1) of the stabilization method allows convenient realization on the computer. Indeed, the difference equation (20.1) may be written in the form

$$\begin{aligned}&F^j=A\varphi^j,\\ &(E+\tfrac{1}{2}\tau A_1)\varepsilon^{j+1/2}=-F^j,\\ &(E+\tfrac{1}{2}\tau A_2)\varepsilon^{j+1}=\varepsilon^{j+1/2},\\ &\varphi^{j+1}=\varphi^j+\tau\varepsilon^{j+1}.\end{aligned} \tag{20.6}$$

Here $\varepsilon^{j+1/2}$ and ε^{j+1} are some auxiliary quantities ensuring the reduction of problem (20.1) to the sequence of elementary problems (20.6). Note that the first and the last

equations in (20.6) are explicit relationships. This means that we have to invert operators only in the second and third equations, where only the elementary operators A_1 and A_2 are present.

We now consider the inhomogeneous problem (18.1), where $A=A_1+A_2$, $A_1\geqslant 0$, $A_2\geqslant 0$, $f\not\equiv 0$. In this case the stabilization scheme is written as follows:

$$\begin{aligned}&(E+\tfrac{1}{2}\tau A_1)(E+\tfrac{1}{2}\tau A_2)\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\varphi^j=f^j,\\&\varphi^0=g,\end{aligned}\tag{20.7}$$

where

$$f^j=f(t_{j+1/2}).\tag{20.8}$$

If the elements of the matrices A_α do not depend on time and the solution and the function f of problem (18.1) are sufficiently smooth, then scheme (20.8) is absolutely stable and approximates the original problem with second-order accuracy in τ (MARCHUK [1980], pp. 251–252).

REMARK 20.1. As we have seen the scheme of the stabilization method may be obtained by introducing into the Crank–Nicolson scheme (i.e. explicit–implicit scheme) the operator $B=E+\frac{1}{4}\tau^2 A_1A_2$. After that the operator of the problem becomes factorizable. Therefore, the stabilization scheme may also be called the *explicit–implicit scheme with approximate factorization of the operator.*

Let us formulate the stabilization scheme for the inhomogeneous problem (18.1) where

$$A=\sum_{\alpha=1}^{n}A_\alpha,\qquad n>2,\quad A_\alpha\geqslant 0\tag{20.9}$$

and the operators A_α are time-independent. With assumptions (20.9) the stabilization method may be represented in the form

$$\prod_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau A_\alpha)\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\varphi^j=f^j,\qquad \varphi^0=g,\tag{20.10}$$

where

$$f^j=f(t_{j+1/2}).$$

The scheme of realization of the algorithm has the following form:

$$\begin{aligned}&F^j=-A\varphi^j+f^j,\\&(E+\tfrac{1}{2}\tau A_1)\varepsilon^{j+1/n}=F^j,\\&(E+\tfrac{1}{2}\tau A_2)\varepsilon^{j+2/n}=\varepsilon^{j+1/n},\\&\quad\vdots\\&(E+\tfrac{1}{2}\tau A_n)\varepsilon^{j+1}=\varepsilon^{j+(n-1)/n},\\&\varphi^{j+1}=\varphi^j+\tau\varepsilon^{j+1}.\end{aligned}\tag{20.11}$$

With the assumption of sufficient smoothness the stabilization method (20.10) has a second-order approximation in τ. Computational stability will be ensured provided that

$$\| T \| \leqslant 1, \tag{20.12}$$

where the transition operator T is defined by the formula

$$T = E - \tau \prod_{\alpha=n}^{1} (E + \tfrac{1}{2}\tau A_\alpha)^{-1} A. \tag{20.13}$$

Unfortunately, the condition $A_\alpha \geqslant 0$ does not imply here the stability in some norm, as was the case for $n = 2$.

To ensure stability the following simple algorithmic tool is sometimes used. Set f^j equal to zero and reduce equation (20.10) resolved with respect to φ^{j+1} to the following form:

$$\varphi^{j+1} = T\varphi^j. \tag{20.14}$$

Since the operator T is assumed to be time-independent (i.e. does not depend on j), we will keep an eye only on the norm $\| \varphi^j \|$ when solving problem (20.14) with initial condition

$$\varphi^0 = g \tag{20.15}$$

and fixed parameter τ ensuring necessary approximation. If this norm is not increasing then we may be sure that computational stability takes place. After this we may move to the solution of the inhomogeneous problem. We rewrite equation (20.10) as follows:

$$\varphi^{j+1} = T\varphi^j + \tau \prod_{\alpha=n}^{1} (E + \tfrac{1}{2}\tau A_\alpha)^{-1} f^j. \tag{20.16}$$

Hence,

$$\| \varphi^{j+1} \|_\Phi \leqslant \| T \| \, \| \varphi^j \|_\Phi + \tau \prod_{\alpha=1}^{n} \| (E + \tfrac{1}{2}\tau A_\alpha)^{-1} \| \, \| f^j \|_F$$

or due to inequality (20.12) and the inequality $\| (E + \frac{1}{2}\tau A_\alpha)^{-1} \| \leqslant 1$

$$\| \varphi^{j+1} \|_\Phi \leqslant \| \varphi^j \|_\Phi + \tau \| f^j \|_F. \tag{20.17}$$

Using the recursive connection we obtain the stability condition

$$\| \varphi^j \|_\Phi \leqslant \| g \|_\Phi + \tau j \| f \|, \tag{20.18}$$

where

$$\| f \| = \max_j \| f^j \|_F. \tag{20.19}$$

REMARK 20.2. While checking the stability condition $\| T \| \leqslant 1$ we used the initial condition (20.15). This was not necessary at all. We might have chosen as initial condition any function from the same smoothness class as g and f.

21. A general scheme for the method of approximate factorization of the operator

We will give a general formulation of the approximate factorization method (MARCHUK and YANENKO [1966], YANENKO [1967], p. 155).

We write for problem (18.1) a two-level approximation of the kind:

$$\begin{aligned} &\frac{\varphi^{j+1}-\varphi^j}{\tau}=\Lambda_1\varphi^{j+1}+\Lambda_0\varphi^j+F^j, \quad j=0, 1, 2, \ldots, \\ &\varphi^0=g, \end{aligned} \tag{21.1}$$

where the function F^j approximates f. In the case of multilayer schemes this approximation contains the results of applying the difference operators to φ^{j-1}, φ^{j-2}, etc. Let the operators Λ_1, Λ_0,

$$\Lambda_0=\sum_{\alpha=1}^{q}\Lambda_{0,\alpha}, \qquad \Lambda_1=\sum_{\alpha=1}^{p}\Lambda_{1,\alpha} \tag{21.2}$$

be represented as the sum of the operators, generally speaking, of a more simple structure, then the following *approximative relationships* are valid

$$\begin{aligned} &\prod_{\alpha=1}^{p}(E-\tau\Lambda_{1,\alpha})\sim E-\tau\Lambda_1, \\ &\prod_{\alpha=1}^{q}(E+\tau\Lambda_{0,\alpha})\sim E+\tau\Lambda_0 \end{aligned} \tag{21.3}$$

(i.e. the operators on the left-hand sides are approximations of the operators on the right-hand sides). The relationships (21.3) allow the replacement of (21.1) by the factorized scheme

$$\prod_{\alpha=1}^{p}(E-\tau\Lambda_{1,\alpha})\varphi^{j+1}=\prod_{\alpha=1}^{q}(E+\tau\Lambda_{0,\alpha})\varphi^j+\tau F^j. \tag{21.4}$$

The given scheme approximates problem (18.1) and if $\Lambda_1=\Lambda_0=-\frac{1}{2}A$, and $F^j=f(t_{j+1/2})$, then it has a second-order approximation in τ. The stability analysis of scheme (21.4) is often difficult in comparison with the componentwise splitting schemes. For the sake of simplicity let $f\equiv 0$ in (18.1). We rewrite this relationship in the form

$$\varphi^{j+1}=T\varphi^j, \tag{21.5}$$

where

$$T=(E-\tau\Lambda_{1,p})^{-1}\cdots(E-\tau\Lambda_{1,1})^{-1}(E+\tau\Lambda_{0,1})\cdots(E+\tau\Lambda_{0,q}).$$

If for operator T the estimate

$$\|T\|\leqslant 1+\text{const}\,\tau \tag{21.6}$$

holds, then scheme (21.4) is stable. Estimate (21.6) is rather difficult to obtain. In the

case when $p=q$ and the operators are commutative we have:

$$T=\prod_{\alpha=1}^{p} T_\alpha, \quad T_\alpha=(E-\tau\Lambda_{1,\alpha})^{-1}(E+\tau\Lambda_{0,\alpha}).$$

If at the same time $\Lambda_{1,\alpha}=\Lambda_{0,\alpha}\leqslant 0$, then $\| T_\alpha \| \leqslant 1$ and the scheme is absolutely stable in the norm $\| \varphi \|_\Phi=(\varphi,\varphi)_\Phi^{1/2}$.

From scheme (21.4) we may obtain the scheme of the splitting operator method (DYAKONOV [1971d]).

We suppose that in (18.1) we have

$$A=\sum_{\alpha=1}^{n} A_\alpha, \quad A_\alpha \geqslant 0, \tag{21.7}$$

where the operators A_α are difference approximations of the corresponding one-dimensional differential operators. Let $p=q=n$, and operators $\Lambda_{0,\alpha}$, $\Lambda_{1,\alpha}$ given by

$$\Lambda_{0,\alpha}=-(1-\theta)A_\alpha, \qquad \Lambda_{1,\alpha}=-\theta A_\alpha, \qquad 0\leqslant\theta=\text{const}\leqslant 1. \tag{21.8}$$

Then scheme (21.4) takes the form:

$$\prod_{\alpha=1}^{n}(E+\tau\theta A_\alpha)\varphi^{j+1}=\prod_{\alpha=1}^{n}(E-\tau(1-\theta)A_\alpha)\varphi^j+\tau F^j, \quad j=0,1,2,\ldots, \qquad \varphi^0=g. \tag{21.9}$$

It is easy to notice that the realization of scheme (21.9) consists of solving the sequence of one-dimensional problems (with the sweep method).

REMARK 21.1 Scheme (21.1) and representations (21.2) do not allow a simple determination of the corresponding scheme. For example, the scheme

$$(E-\tau\Lambda_{1,p})\cdots(E-\tau\Lambda_{1,1})\varphi^{j+1}=(E+\tau\Lambda_{0,q})\cdots(E+\tau\Lambda_{0,1})\varphi^j+\tau F^j \tag{21.10}$$

is not equivalent to scheme (21.4) in the general case of noncommutative operators $\Lambda_{1,\alpha}$, $\Lambda_{0,\alpha}$ but only in the case of commutative operators.

REMARK 21.2. In a paper by YANENKO ([1967], pp. 158–159) schemes of approximate factorization arising from the multi-level schemes are presented:

$$A_1\varphi^{j+1}+A_0\varphi^j+A_{-1}\varphi^{j-1}+\cdots+A_{-p+1}\varphi^{j-p+1}+f^j=0. \tag{21.11}$$

These schemes may appear in approximations of equations of the type

$$B_1\frac{\partial\varphi}{\partial t}+B_2\frac{\partial^2\varphi}{\partial t^2}+\cdots+B_p\frac{\partial^p\varphi}{\partial t^p}+A\varphi=f, \tag{21.12}$$

where $B_1, B_2, \ldots, B_p$ are linear operators.

22. The scheme of approximate factorization for the parabolic equation

In papers by DYAKONOV [1962d–f, 1963–1965, 1971b–d, 1972] the method of approximate factorization (splitting operator method) has been developed and grounded for a wide class of equations of parabolic and hyperbolic types. We consider some of these results for the parabolic equation of the form:

$$\frac{\partial \varphi}{\partial t} - \sum_{\alpha=1}^{n} \frac{\partial^2 \varphi}{\partial x_\alpha^2} = f \quad \text{in } \Omega \times \Omega_t \tag{22.1}$$

under conditions

$$\varphi|_{\partial\Omega} = \varphi_{(\Gamma)}, \qquad \varphi = g \quad \text{for } t = 0.$$

Here

$$\Omega = \{0 < x_\alpha < 1 : \alpha = 1, \ldots, n\}, \qquad \Omega_t = \{0 < t < T\}.$$

Let

$$\begin{aligned} &\tau = T/K, \qquad h_\alpha = 1/N_\alpha, \quad \alpha = 1, \ldots, n, \\ &h \equiv (h_1, \ldots, h_n), \qquad |h|^2 = \sum_{\alpha=1}^{n} h_\alpha^2. \end{aligned} \tag{22.2}$$

The spatial grid Ω_h is the set of points $x_l = (l_1 h_1, \ldots, l_n h_n)$, where $1 \leqslant l_\alpha \leqslant N_\alpha - 1$, $l \equiv (l_1, \ldots, l_n)$, $\partial\Omega_h$ is a set of boundary grid points, $\bar{\Omega}_h = \Omega_h + \partial\Omega_h$. The temporal grid is given by $t_j = j\tau$, $j = 0, 1, \ldots, K$.

We introduce the difference operators

$$\Delta_\alpha(\varphi)_l^j = \frac{1}{h_\alpha}(\varphi(t_j, x_l + h_\alpha l_\alpha) - \varphi(t_j, x_l)),$$

$$\nabla_\alpha(\varphi)_l^j = \frac{1}{h_\alpha}(\varphi(t_j, x_l) - \varphi(t_j, x_l - h_\alpha l_\alpha)),$$

where l_α is the n-dimensional unity vector in the direction x_α. We consider a set of difference schemes of the kind

$$\begin{aligned} &\frac{\varphi_l^{j+1} - \varphi_l^j}{\tau} + \sum_{\alpha=1}^{n} A_\alpha(\theta\varphi_l^{j+1} + (1-\theta)\varphi_l^j) + \hat{f}_l^{j+1} \quad \text{in } \Omega_h, \quad 0 \leqslant j \leqslant K, \\ &\varphi_l^0 = g_l, \qquad x_l \in \Omega_h, \\ &\varphi_l^j = \varphi_{(\Gamma)l}^j, \qquad x_l \in \partial\Omega_h, \end{aligned} \tag{22.3}$$

where

$$A_\alpha = -(\Delta_\alpha \nabla_\alpha) \equiv -\Delta_\alpha^h, \qquad 0 \leqslant \theta \leqslant \text{const} \leqslant 1,$$

$$f_l^{j+1} \equiv \begin{cases} f(t_{j+1}, x_l) & \text{for } \theta \neq \frac{1}{2}, \\ f(t_{j+1/2}, x_l) & \text{for } \theta = \frac{1}{2}, \end{cases}$$

and the notation $\varphi_l^j \equiv \varphi(\tau_j, x_l)$ is used. When the solutions are sufficiently smooth,

scheme (22.3) approximates (22.1) with local error of approximation of order $O(\tau+|h|^2)$ for $\theta\neq\frac{1}{2}$ and $O(\tau^2+|h|^2)$ for $\theta=\frac{1}{2}$. For $\theta=0$ scheme (22.3) is explicit; for $\theta=1$ it is purely implicit; for $\theta=\frac{1}{2}$ it is the Crank–Nicolson scheme.

Let Φ and F be the Hilbert spaces of the grid functions $\varphi\in\Phi$ and $f\in F$, defined on Ω_h with inner product

$$(u, v)_\Phi = h_1\cdots h_n \sum_{x_l\in\Omega_h} u_l v_l$$

and norm $\|u\|_\Phi=(u,u)_\Phi^{1/2}$. Then in this norm the schemes (22.3) are absolutely stable independently of τ and h for $\theta\geqslant\frac{1}{2}$, and for $\theta<\frac{1}{2}$ they are stable provided that the following relationship holds:

$$\tau\sum_{\alpha=1}^{n}\frac{1}{h_\alpha^2}\leqslant\frac{1}{2(1-2\theta)}, \tag{22.4}$$

i.e. conditionally stable. From the approximation and stability of these schemes corresponding statements about the convergence follow (see Theorem 3.1).

We now introduce the operator

$$\begin{aligned} B\varphi_l^{j+1} &\equiv \prod_{\alpha=1}^{n}(E-\theta\tau\Delta_\alpha^h)\varphi_l^{j+1} \\ &= \varphi_l^{j+1}-\theta\tau\sum_{\alpha=1}^{n}\Delta_\alpha^h\varphi_l^{j+1} \\ &\quad +(-\theta\tau)^2\sum_{1\leqslant\alpha\leqslant\beta\leqslant n}\Delta_\alpha^h\Delta_\beta^h\varphi_l^{j+1}+\cdots \\ &\quad +(-\theta\tau)^n\Delta_1^h\Delta_2^h\cdots\Delta_n^h\varphi_l^{j+1}, \end{aligned} \tag{22.5}$$

$$x_l\in\Omega_h$$

and consider the scheme

$$\begin{aligned} &B\frac{\varphi_l^{j+1}-\varphi_l^j}{\tau}-\sum_{\alpha=1}^{n}\Delta_\alpha^h\varphi_l^j=\hat{f}_l^{j+1}, \quad x_l\in\Omega_h, \\ &\varphi_l^0=g_l, \qquad x_l\in\Omega_h, \\ &\varphi_l^j=\varphi_{(\Gamma)l}^j, \quad x_l\in\partial\Omega_h. \end{aligned} \tag{22.6}$$

Taking (22.5) into account we may rewrite (22.6) in the form

$$\begin{aligned} &\frac{\varphi_l^{j+1}-\varphi_l^j}{\tau}-\sum_{\alpha=1}^{n}\Delta_\alpha^h(\theta\varphi_l^{j+1}+(1-\theta)\varphi_l^j) \\ &\qquad +Q_n\left(\frac{\varphi_l^{j+1}-\varphi_l^j}{\tau}\right)=\hat{f}_l^{j+1}, \end{aligned} \tag{22.7}$$

where

$$Q_nV_l\equiv\sum_{\alpha=2}^{n}(-\theta\tau)^\alpha\sum_{1\leqslant\alpha_1<\alpha_2<\cdots<\alpha_n\leqslant n}\Delta_{\alpha_1}^h\Delta_{\alpha_2}^h\cdots\Delta_{\alpha_n}^hV_l. \tag{22.8}$$

Thus, we see that (22.6) differs from (22.3) by the presence of the additional term $Q_n(\varphi_l^{j+1}-\varphi_l^j)/\tau$ on the left-hand side. If the solution of the original problem has derivatives (bounded in $\Omega\times\Omega_t$)

$$D_{i_1}^{\beta_1}D_{i_2}^{\beta_2}\cdots D_{i_\alpha}^{\beta_\alpha}\frac{\partial\varphi}{\partial t},\quad i_1<i_2<\cdots<i_\alpha,\quad \alpha\leqslant n,$$

where

$$\beta_\alpha\leqslant 2,\qquad D_l^{\beta_i}\equiv\partial^{\beta_i}/\partial x_l^{\beta_i},$$

then

$$\left\|Q_n\left(\frac{\varphi_l^{j+1}-\varphi_l^j}{\tau}\right)\right\|_\Phi\leqslant M\tau^2,$$

$$\varphi_l^j=\varphi(t_j,x_l),\quad M=\text{const.} \tag{22.9}$$

(Note that the smoothness restriction on, $\varphi(t,x)$ may be weakened; see DYAKONOV [1971d], p. 23.) Thus, if (22.9) holds then scheme (22.6) approximates the original problem with the same order as scheme (22.3). Here, naturally, we have for $\sigma\neq\frac{1}{2}$ instead of (22.9) the inequality

$$\left\|Q_n\left(\frac{\varphi^{j+1}-\varphi^j}{\tau}\right)\right\|_\Phi\leqslant M\tau \tag{22.10}$$

and, also, weakened conditions on the smoothness of $\varphi(t,x)$ would do.

We consider how the transition from the time level t_j to the level t_{j+1} in scheme (22.6) is realized. Note that if we exclude—with the help of boundary conditions—all boundary values of φ from the left-hand side at the expense of the corresponding modifications of $\hat{f}_l^{j+1}$ in the near-boundary grid points, then we may rewrite (22.6) in the form:

$$\prod_{\alpha=1}^{n}(E+\tau\theta A_\alpha)\varphi^{j+1}=\prod_{\alpha=1}^{n}(E+\tau\theta A_\alpha)\varphi^j-\tau A\varphi^j+\tilde{f}^{j+1}, \tag{22.11}$$

where $A=\Sigma_{\alpha=1}^n A_\alpha$, and the vector $\tilde{f}^{j+1}$ is defined through $\hat{f}^{j+1}$, $\varphi_{(\Gamma)}^{j+1}$ and $\varphi_{(\Gamma)}^j$. This is (18.2) with splitting operator $B=\Pi_{\alpha=1}^n(E+\tau\theta A_\alpha)$. The solution of (22.11) on each time step is reduced to solution of (18.4) for $p=n$ and $B_\alpha=(E+\tau\theta A_\alpha)$ by the sweep method.

Stability of scheme (22.6) follows from the a priori estimate

$$\|\varphi^{j+1}\|_\Phi\leqslant\|g\|_\Phi+\tau\sum_{i=1}^{j}\|f^{i+1}\|_F \tag{22.12}$$

(DYAKONOV [1971d], p. 26).

REMARK 22.1. Along with scheme (22.6) one may consider schemes with approximate factorization of the operator, which are obtained by replacing the basic

equation from (22.6) with one of the following equations:

$$\frac{1}{\tau_j}\prod_{\alpha=1}^{n}(E-\tau_j\Delta_\alpha^h)\varphi_l^{j+1}-\frac{\varphi_l^j}{\tau_j}=f_l^{j+1},$$

$$\frac{1}{\tau_j}\prod_{\alpha=1}^{n}(E-\theta_\alpha\tau_j\Delta_\alpha^h)\varphi_l^{j+1}-\frac{\varphi_l^j}{\tau_j}-\sum_{\alpha=1}^{n}(1-\theta_\alpha)\Delta_\alpha^h\varphi_l^j=\hat{f}_l^{j+1},$$

$$\frac{1}{\tau_j}\prod_{\alpha=1}^{n}(E-\theta_\alpha\tau_j\Delta_\alpha^h)\varphi_l^{j+1}-\frac{1}{\tau_j}\prod_{\alpha=1}^{n}(E+(1-\theta_\alpha)\tau_j\Delta_\alpha^h)\varphi_l^j=\hat{f}_l^{j+1},$$

$$\frac{1}{\tau_j}\prod_{\alpha=1}^{n}(E-\tfrac{1}{2}\tau_j\Delta_\alpha^h)\varphi_l^{j+1}-\frac{\varphi_l^j}{\tau_j}-\tfrac{1}{2}\sum_{\alpha=1}^{n}\Delta_\alpha^h\varphi_l^j-\tfrac{1}{4}\tau_j\sum_{\alpha<\beta}\Delta_\alpha^h\Delta_\beta^h\varphi_l^j=f_l^{j+1},$$

where $\tau_j=t_j-t_{j-1}$ is a nonconstant grid step in t and $0\leqslant\theta_\alpha=\text{const}\leqslant 1$ (DYAKONOV [1971d], p. 26).

The Predictor-Corrector Method

The next class of the splitting methods—the predictor-corrector method (approximation correction scheme)—as well as the scheme with a factorization of the operator will be considered in application to the matrix evolutionary equation (18.1) with a time-independent operator.

23. The predictor-corrector method: The case $A = A_1 + A_2$

The idea of the predictor-corrector method is as follows. The whole interval $0 \leqslant t \leqslant T$ is split into a number of partial intervals and within each elementary interval $t_j \leqslant t \leqslant t_{j+1}$ the problem (18.1) is solved in two steps. First, one finds an approximate solution of the problem at time $t_{j+1/2} = t_j + \frac{1}{2}\tau$ using a first-order accuracy scheme which has a considerable "amount" of stability. This step is usually referred to as the predictor. Then, on the whole interval (t_j, t_{j+1}) the original equation is approximated with a second-order scheme which serves as the corrector. It is essential that, while constructing the corrector, the "rough" solution at $t_{j+1/2}$ found in the predictor step is used.

The predictor-corrector scheme may be written in the form:

$$\begin{aligned} &\frac{\varphi^{j+1/4} - \varphi^j}{\frac{1}{2}\tau} + A_1 \varphi^{j+1/4} = 0, \\ &\frac{\varphi^{j+1/2} - \varphi^{j+1/4}}{\frac{1}{2}\tau} + A_2 \varphi^{j+1/2} = 0, \\ &\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^{j+1/2} = 0, \\ &\varphi^0 = g. \end{aligned} \tag{23.1}$$

If we exclude the auxiliary function $\varphi^{j+1/4}$ from the first two equations of (23.1), the system will reduce to the following one:

$$\begin{aligned} &(E + \tfrac{1}{2}\tau A_1)(E + \tfrac{1}{2}\tau A_2)\varphi^{j+1/2} = \varphi^j, \\ &\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^{j+1/2} = 0. \end{aligned} \tag{23.2}$$

Excluding $\varphi^{j+1/2}$ from (23.2) we get

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+A(E+\tfrac{1}{2}\tau A_2)^{-1}(E+\tfrac{1}{2}\tau A_1)^{-1}\varphi^j=0,$$
$$\varphi^0=g. \tag{23.3}$$

We now suppose that the following restriction holds:

$$\tfrac{1}{2}\tau\|A_\alpha\|<1, \quad \alpha=1,2. \tag{23.4}$$

If $A_1\geqslant 0$, $A_2\geqslant 0$, the elements of matrices A_1, A_2 do not depend on t, and the solution φ of (18.1) is sufficiently smooth, then the difference scheme (23.1) is absolutely stable and approximates the original problem with second-order accuracy in τ. At the same time the estimate

$$\|\varphi^j\|_{C_1^{-1}}\leqslant\|g\|_{C_1^{-1}}, \tag{23.5}$$

holds where

$$C_1^{-1}=(E+\tfrac{1}{2}\tau A_1^*)^{-1}(E+\tfrac{1}{2}\tau A_1),$$
$$\|\varphi\|_{C_1^{-1}}=(C_1^{-1}\varphi,\varphi)_\Phi^{1/2}.$$

To prove this statement let us rewrite (23.3) in the form

$$(E+\tfrac{1}{2}\tau A_1)(E+\tfrac{1}{2}\tau A_2)\frac{\varphi^{j+1}-\varphi^j}{\tau}+\Lambda\varphi^j=0,$$

where

$$\Lambda=(E+\tfrac{1}{2}\tau A_1)(E+\tfrac{1}{2}\tau A_2)A(E+\tfrac{1}{2}\tau A_2)^{-1}(E+\tfrac{1}{2}\tau A_1)^{-1}.$$

Expanding the right-hand side of the last relationship into a Taylor series in τ and assuming that

$$\tfrac{1}{2}\|A_\alpha\|<1,$$

we get

$$\Lambda=A+O(\tau^2).$$

With the help of the estimate used in the stabilization method we come to the conclusion that the predictor-corrector method has a second-order of approximation in τ.

Let us consider the stability of this method. To this end we rewrite (23.3) in the form

$$(E+\tfrac{1}{2}\tau A_1)(E+\tfrac{1}{2}\tau A_2)\frac{\Phi^{j+1}-\Phi^j}{\tau}+A\Phi^j=0, \tag{23.6}$$

where

$$\Phi^j=(E+\tfrac{1}{2}\tau A_2)^{-1}(E+\tfrac{1}{2}\tau A_1)^{-1}\varphi^j. \tag{23.7}$$

Difference scheme (23.6) is stable since

$$\|\Phi^{j+1}\|_{C_2}\leqslant\|\Phi^j\|_{C_2}. \tag{23.8}$$

Then taking (23.7) and (23.8) into account we get

$$\|(E+\tfrac{1}{2}\tau A_1)^{-1}\varphi^{j+1}\|_{\Phi}\leqslant\|(E+\tfrac{1}{2}\tau A_1)^{-1}\varphi^{j}\|_{\Phi}$$

or

$$\|\varphi^{j+1}\|_{C_1^{-1}}\leqslant\|\varphi^{j}\|_{C_1^{-1}}\leqslant\cdots\leqslant\|g\|_{C_1^{-1}}.$$

In the case of an inhomogeneous problem we formulate the predictor-corrector method as follows:

$$\begin{aligned}
&\frac{\varphi^{j+1/4}-\varphi^{j}}{\frac{1}{2}\tau}+A_1\varphi^{j+1/4}=f^{j},\\
&\frac{\varphi^{j+1/2}-\varphi^{j+1/4}}{\frac{1}{2}\tau}+A_2\varphi^{j+1/2}=0,\\
&\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+A\varphi^{j+1/2}=f^{j},
\end{aligned}\tag{23.9}$$

where

$$f^{j}=f(t_{j+1/2}).$$

If f^j is chosen in this form, then (23.9) approximates the original problem with second-order accuracy in τ and the estimate

$$\|\varphi^{j}+\tfrac{1}{2}\tau f^{j}\|_{C_1^{-1}}\leqslant\|g+\tfrac{1}{2}\tau f^{0}\|_{C_1^{-1}}+\tau j\|f\|_{C_1^{-1}},\tag{23.10}$$

where

$$\|f\|_{C_1^{-1}}=\max\|f^{j}\|_{C_1^{-1}},$$

is valid, i.e., if $0\leqslant t_j\leqslant T$ we again get stability of the difference scheme (see MARCHUK [1980], pp. 254–255).

Thus if $A_1\geqslant 0$, $A_2\geqslant 0$ and the elements of matrices A_1, A_2 do not depend on time, then the difference scheme (23.9) is absolutely stable and allows to get a solution of second-order accuracy in τ provided that the solution and the right-hand side f of (18.1) are sufficiently smooth.

In conclusion we shall pay attention to the fact that, although difference scheme (23.1) is absolutely stable, the difference scheme for the corrector incorporated here as a part may be absolutely unstable if considered separately. Let us show this. For the sake of simplicity we will consider the case when A is the difference analogue of the two-dimensional Laplace operator, Ω is the unity square, and the solution equals zero on the boundaries of Ω.

In this case the corrector has the form

$$\frac{\varphi_{k,l}^{j+1}-\varphi_{k,l}^{j-1}}{2\tau}-\frac{\varphi_{k+1,l}^{j}-2\varphi_{k,l}^{j}+\varphi_{k-1,l}^{j}}{h^2}-\frac{\varphi_{k,l+1}^{j}-2\varphi_{k,l}^{j}+\varphi_{k,l-1}^{j}}{h^2}=0$$

(since now the predictor is not used in the computations, only integers j are considered).

For this difference problem a solution φ^j will be chosen for which the inequality $\|\varphi^{j}\|_{\Phi}\leqslant\|\varphi^{j-1}\|_{\Phi}$ from the definition of stability does not hold. We will find such

a solution in the form

$$\varphi_{k,l}^{j}=\lambda^{j}\sin(m\pi kh)\sin(\rho\pi lh),$$

where j serves both as an index on the left-hand side and as a power on the right-hand side. Using this expression in the difference equation we obtain the following characteristic equation:

$$\lambda^2+8a_{mp}\lambda-1=0,$$

where

$$a_{mp}=\frac{\tau}{h^2}(\sin^2(\tfrac{1}{2}m\pi h)+\sin^2(-\tfrac{1}{2}\rho\pi h)).$$

Taking as λ the quantity

$$\lambda=-4a_{mp}-\sqrt{1+16a_{mp}^2},$$

we see that

$$\|\varphi^j\|_\Phi/\|\varphi^0\|_\Phi=|\lambda|^j\to\infty \quad \text{as } \tau\to 0$$

(so that $\tau/h^2=\text{const}$), i.e. the scheme is absolutely unstable.

Thus in spite of the fact that the difference scheme used as corrector is absolutely unstable the "amount" of stability the predictor possesses is sufficient for absolute stability of the whole scheme.

24. The predictor-corrector method: The case $A=\Sigma_{\alpha=1}^{n}A_\alpha$

We suppose that in (18.1)

$$A=\sum_{\alpha=1}^{n}A_\alpha, \qquad A_\alpha\geqslant 0, \quad n>2. \tag{24.1}$$

In this case the splitting scheme has a form:

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau A_1)\varphi^{j+1/2n}=\varphi^j+\tfrac{1}{2}\tau f^j,\\
&(E+\tfrac{1}{2}\tau A_2)\varphi^{j+2/2n}=\varphi^{j+1/2n},\\
&\quad\vdots\\
&(E+\tfrac{1}{2}\tau A_n)\varphi^{j+1/2}=\varphi^{j+(n-1)/2n},\\
&\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\varphi^{j+1/2}=f^j,
\end{aligned} \tag{24.2}$$

where we suppose again that $A_\alpha\geqslant 0$ and $f^j=f(t_{j+1/2})$. The system of equations (24.2) is reduced to one equation

$$\begin{aligned}
&\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\prod_{\alpha=n}^{1}(E+\tfrac{1}{2}\tau A_\alpha)^{-1}(\varphi^j+\tfrac{1}{2}\tau f^j)=f^j,\\
&\varphi^0=g.
\end{aligned} \tag{24.3}$$

The predictor-corrector method in this case has second-order accuracy in τ provided that the solution is sufficiently smooth. Write (24.3) in the form:

$$\varphi^{j+1} = T\varphi^j + \tfrac{1}{2}\tau(E+T)f^j, \tag{24.4}$$

where the transition operator T is given by

$$T = E - \tau A \prod_{\alpha=n}^{1} (E + \tfrac{1}{2}\tau A_\alpha)^{-1}. \tag{24.5}$$

The requirements for numerical stability have been reduced, eventually, to estimating the norm of the operator T. Unfortunately in this case the constructive condition $A_\alpha \geqslant 0$ does not allow to prove the stability of the scheme.

In the case when the operators A_α are commutative and have a common basis, the stability of the above schemes follows from the condition $A_\alpha \geqslant 0$ (MARCHUK [1980], p. 265).

REMARK 24.1. The stabilization method as well as the predictor-corrector method for splitting the operator in n-components may also be applied when operator A is time-dependent. However, in this situation the a priori formulation of the stability condition appears to be a more complicated problem. Therefore it is difficult to say to what extent the application of the two above schemes is justified in general situations.

25. The predictor-corrector method for the parabolic equation

Let

$$\Omega = \{0 < x_\alpha < 1 \colon \alpha = 1, 2, 3\}, \qquad \Omega_t = \{0 < t < T\}.$$

For the problem

$$\begin{aligned} &\frac{\partial \varphi}{\partial t} - a^2 \sum_{\alpha=1}^{3} \frac{\partial^2 \varphi}{\partial x_\alpha^2} = 0 \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{for } t = 0, \\ &\varphi = \varphi_{(\Gamma)} \quad \text{in } \partial\Omega \end{aligned} \tag{25.1}$$

we introduce the following difference approximation

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = \varphi_{(\Gamma)} \quad \text{on } \partial\Omega, \\ &\varphi = g \quad \text{for } t = 0. \end{aligned} \tag{25.2}$$

Here Ω and $\partial\Omega$ are already grid domains, and A has the form:

$$A = \sum_{\alpha=1}^{3} A_\alpha, \quad A_\alpha = -a^2(\Delta_{x_\alpha} \nabla_{x_\alpha})/h_{x_\alpha}^2. \tag{25.3}$$

BRAYN [1966] (see also YANENKO [1967], p. 40) suggested the following scheme

based on the idea of the predictor-corrector method:

$$\begin{aligned}
&\frac{\varphi^{j+1/6}-\varphi^{j}}{\frac{1}{2}\tau}+A_1\varphi^{j+1/6}+A_2\varphi^{j}+A_3\varphi^{j}=0,\\
&\frac{\varphi^{j+1/3}-\varphi^{j+1/6}}{\frac{1}{2}\tau}+A_2(\varphi^{j+1/3}-\varphi^{j})=0,\\
&\frac{\varphi^{j+1/2}-\varphi^{j+1/3}}{\frac{1}{2}\tau}+A_3(\varphi^{j+1/2}-\varphi^{j})=0,\\
&\frac{\varphi^{j+1}-\varphi^{j}}{\frac{1}{2}\tau}+A_1\varphi^{j+1/6}+A_2\varphi^{j+1/3}+A_3\varphi^{j+1/2}=0.
\end{aligned} \tag{25.4}$$

Here, the first three equations form the predictor (*stabilizing correction scheme*), bringing φ to time level $t=(j+\frac{1}{2})\tau$; the fourth is the corrector.

Having excluded the fractional steps from (25.4) we get

$$\begin{aligned}
&\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+A\frac{\varphi^{j+1}+\varphi^{j}}{2}+\tfrac{1}{4}\tau^2(A_1A_2+A_1A_3+A_2A_3)\frac{\varphi^{j+1}-\varphi^{j}}{\tau}\\
&\quad+\tfrac{1}{8}\tau^3A_1A_2A_3\frac{\varphi^{j+1}-\varphi^{j}}{\tau}=0.
\end{aligned} \tag{25.5}$$

From this we conclude that scheme (25.4) has a second-order approximation in τ and in h_α. Taking into account that the operators A_α are commutative, we also conclude that this scheme is absolutely stable.

DOUGLAS [1962] suggested the following absolutely stable scheme which has second-order accuracy:

$$\begin{aligned}
&\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau}+\tfrac{1}{2}A_1(\varphi^{j+1/3}+\varphi^{j})+A_2\varphi^{j}+A_3\varphi^{j}=0,\\
&\frac{\varphi^{j+2/3}-\varphi^{j}}{\tau}+\tfrac{1}{2}A_1(\varphi^{j+1/3}+\varphi^{j})\\
&\quad+\tfrac{1}{2}A_2(\varphi^{j+2/3}+\varphi^{j})+A_3\varphi^{j}=0,\\
&\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+\tfrac{1}{2}A_1(\varphi^{j+1/3}+\varphi^{j})\\
&\quad+\tfrac{1}{2}A_2(\varphi^{j+2/3}+\varphi^{j})+\tfrac{1}{2}A_3(\varphi^{j+1}+\varphi^{j})=0,
\end{aligned} \tag{25.6}$$

which may be rewritten in the form:

$$\begin{aligned}
&\frac{\varphi^{j+1/3}-\varphi^{j}}{\tau}+\tfrac{1}{2}A_1(\varphi^{j+1/3}+\varphi^{j})+A_2\varphi^{j}+A_3\varphi^{j}=0,\\
&\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau}+\tfrac{1}{2}A_2(\varphi^{j+2/3}-\varphi^{j})=0,\\
&\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau}+\tfrac{1}{2}A_3(\varphi^{j+1}-\varphi^{j})=0.
\end{aligned} \tag{25.7}$$

Excluding $\varphi^{j+1/3}, \varphi^{j+2/3}$ from (25.7) we again come to (25.5). Therefore, schemes (25.4) and (25.6) are equivalent (YANENKO [1967], p. 42).

We write the predictor-corrector scheme as

$$\begin{aligned}
&\frac{\varphi^{j+1/6}-\varphi^{j}}{\frac{1}{2}\tau}+A_1\varphi^{j+1/6}=0,\\
&\frac{\varphi^{j+1/3}-\varphi^{j+1/6}}{\frac{1}{2}\tau}+A_2\varphi^{j+1/3}=0,\\
&\frac{\varphi^{j+1/2}-\varphi^{j+1/3}}{\frac{1}{2}\tau}+A_3\varphi^{j+1/2}=0,\\
&\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+A\varphi^{j+1/2}=0,
\end{aligned} \tag{25.8}$$

which is special realization of scheme (24.2). Excluding $\varphi^{j+1/6}, \varphi^{j+1/3}, \varphi^{j+1/2}$ from (25.8) we also come to (25.5). Therefore, this scheme is equivalent to the schemes (25.4) and (25.6).

The Alternating Direction and the Stabilizing Correction Methods

26. The alternating direction method

Consider the matrix homogeneous evolutionary problem (18.1):

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega \times \Omega_t, \\ &\varphi = g \quad \text{in } \Omega \qquad \text{for } t = 0, \end{aligned} \tag{26.1}$$

where

$$A = A_1 + A_2.$$

A scheme of the *alternating direction method* for (26.1) has the form

$$\begin{aligned} &\frac{\varphi^{j+1/2} - \varphi^j}{\tau} + \tfrac{1}{2}(A_1 \varphi^{j+1/2} + A_2 \varphi^j) = 0, \\ &\frac{\varphi^{j+1} - \varphi^{j+1/2}}{\tau} + \tfrac{1}{2}(A_1 \varphi^{j+1/2} + A_2 \varphi^{j+1}) = 0, \\ &\quad j = 0, 1, 2, \ldots, \\ &\varphi^0 = g. \end{aligned} \tag{26.2}$$

In this form it was applied by PEACEMAN and RACHFORD [1955] and DOUGLAS [1955] to a parabolic problem with two spatial variables. In this case operator A_α is a difference approximation of the one-dimensional operator $-a^2\partial^2/\partial x_\alpha^2$. Note, that in this case scheme (26.2) is symmetric, i.e. x_1 and x_2 change roles from first fractional step to the second (this fact gives the method its name). It is easy to find the solution of each equation in the parabolic problem by the sweep method (therefore, scheme (26.2) is also called the *alternating direction sweep method* (YANENKO [1967], p. 26).

Eliminating $\varphi^{j+1/2}$ from (26.2) we obtain

$$\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\left(\frac{\varphi^{j+1} + \varphi^j}{2}\right) + \tfrac{1}{4}\tau^2 A_1 A_2 \left(\frac{\varphi^{j+1} - \varphi^j}{\tau}\right) = 0. \tag{26.3}$$

Comparing (26.3) with the Crank–Nicolson scheme we conclude that it has a second-order approximation in τ. Further, if we consider (26.2) when A_α is a three-

point approximation of the operator $-a^2(\partial^2/\partial x_\alpha^2)$,

$$A_\alpha = -a^2(\Delta_{x_\alpha}\nabla_{x_\alpha})/h_\alpha^2,$$

then it is easy to establish that this scheme is absolutely stable (DOUGLAS and RACHFORD [1956], YANENKO [1967], p. 28). If we analyze in this case the behaviour of the error in each direction x_α using Neumann's method, then it appears that in the first semistep the error in the direction x_1 decreases and the error in the direction x_2 increases, then in the second semistep the error in the direction x_1 increases and the error in the direction x_2 decreases. It does not matter how much the error increases in any direction at a given semistep, because it certainly will decrease at the next one, so that on the whole after two semisteps it will not increase in absolute value (YANENKO [1967], pp. 28–29).

YANENKO ([1967], p. 28) noted that a scheme of the alternating direction method is useless for the three-dimensional parabolic problem. It has been shown that in that case the following scheme (with $A_\alpha = -a^2(\Delta_{x_\alpha}\nabla_{x_\alpha})/h_\alpha^2$)

$$\begin{aligned}
&\frac{\varphi^{j+1/3}-\varphi^j}{\tau}+\tfrac{1}{3}(A_1\varphi^{j+1/3}+A_2\varphi^j+A_3\varphi^j)=0,\\
&\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau}+\tfrac{1}{3}(A_1\varphi^{j+1/3}+A_2\varphi^{j+2/3}+A_3\varphi^{j+1/3})=0,\\
&\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau}+\tfrac{1}{3}(A_1\varphi^{j+2/3}+A_2\varphi^{j+2/3}+A_3\varphi^{j+1})=0,
\end{aligned} \tag{26.4}$$

is not absolutely stable. Therefore, for many problems schemes of the stabilizing correction method are preferable (along with schemes of the alternating direction method they are sometimes also called *implicit schemes of alternating direction*).

27. The stabilizing correction method

Stabilizing correction schemes were proposed by DOUGLAS and RACHFORD [1956] for solving the three-dimensional heat conduction equation. So if we have in (26.1)

$$A = A_1 + A_2 + A_3,$$

this scheme has the form

$$\begin{aligned}
&\frac{\varphi^{j+1/3}-\varphi^j}{\tau}+A_1\varphi^{j+1/3}+A_2\varphi^j+A_3\varphi^j=0,\\
&\frac{\varphi^{j+2/3}-\varphi^{j+1/3}}{\tau}+A_2(\varphi^{j+2/3}-\varphi^j)=0,\\
&\frac{\varphi^{j+1}-\varphi^{j+2/3}}{\tau}+A_3(\varphi^{j+1}-\varphi^j)=0,\\
&j=0,1,2,\ldots,\\
&\varphi^0=g.
\end{aligned} \tag{27.1}$$

By eliminating $\varphi^{j+1/3}$ and $\varphi^{j+2/3}$ from (27.1) we obtain the equation

$$\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+A\varphi^{j+1}+\tau^{2}(A_1A_2+A_1A_3+A_2A_3)\left(\frac{\varphi^{j+1}-\varphi^{j}}{\tau}\right)$$
$$+\tau^{3}A_1A_2A_3\left(\frac{\varphi^{j+1}-\varphi^{j}}{\tau}\right)=0. \tag{27.2}$$

It follows that the scheme has first-order accuracy in τ. When we consider it for the heat conduction equation it is easy to establish absolute stability. Besides, the structure of the scheme is as follows: the first fractional step gives the whole approximation of the heat conduction equation, the next fractional steps are corrective and serve to improve the stability. Therefore such schemes are called *stabilizing correction schemes* or *schemes with stability correction.*

DOUGLAS [1962] proposed scheme (25.6) of second-order accuracy, which may also be considered as a stabilizing correction scheme.

28. A general formulation for the stabilizing correction method

The stabilizing correction method was formulated in a general form by DOUGLAS and GUNN [1964] (see also YANENKO [1967], p. 159). So, let us consider the evolutionary problem

$$\begin{aligned}&\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega\times\Omega_t,\\ &\varphi=g \quad \text{in } \Omega \qquad \text{for } t=0,\end{aligned} \tag{28.1}$$

where $A=\Sigma_{\alpha=1}^{n}A_\alpha$. We write the difference multi-layer scheme

$$\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+\Lambda\varphi^{j+1}=F^{j}, \tag{28.2}$$

where

$$\begin{aligned}&F^{j}=B_0\varphi^{j}+B_{-1}\varphi^{j-1}+\cdots+B_{-q+1}\varphi^{j-q+1}+f^{j},\\ &\Lambda=\Lambda_1+\cdots+\Lambda_n,\end{aligned} \tag{28.3}$$

which approximates problem (28.1). Taking (28.3) into account we can write the corresponding scheme in fractional steps:

$$\begin{aligned}&\frac{\varphi^{j+1/n}-\varphi^{j}}{\tau}+\Lambda_1(\varphi^{j+1/n}-\varphi^{j})+\Lambda\varphi^{j}=F^{j},\\ &\frac{\varphi^{j+2/n}-\varphi^{j+1/n}}{\tau}+\Lambda_2(\varphi^{j+2/n}-\varphi^{j})=0,\\ &\quad\vdots\\ &\frac{\varphi^{j+1}-\varphi^{j+(n-1)/n}}{\tau}+\Lambda_n(\varphi^{j+1}-\varphi^{j})=0.\end{aligned} \tag{28.4}$$

By eliminating $\varphi^{j+1/n}, \ldots, \varphi^{j+(n-1)/n}$ from (28.4) we obtain, after simple transformations, an equation "on the whole steps":

$$\frac{\varphi^{j+1}-\varphi^{j}}{\tau}+\Lambda\varphi^{j+1}+\tau^{2}\Phi\left(\frac{\varphi^{j+1}-\varphi^{j}}{\tau}\right)=F^{j}, \tag{28.5}$$

where

$$\Phi=\sum_{i<j}\Lambda_i\Lambda_j+\tau\sum_{i<j<k}\Lambda_i\Lambda_j\Lambda_k+\cdots+\tau^{n-2}\Lambda_1\cdots\Lambda_n. \tag{28.6}$$

By comparing (28.5) and (28.6) we conclude that, with accuracy up to order $O(\tau^2)$, the general stabilizing correction scheme keeps the order of approximation of scheme (28.2)–(28.3). The spectral stability analysis for scheme (28.4) is given by YANENKO ([1967], pp. 160–161).

Note that for two-layer schemes under the condition that the operators are commutative and for $\Lambda_\alpha > 0$ the stability of scheme (28.4) follows from the stability of scheme (28.2) (DOUGLAS and GUNN [1964]).

29. Application of the alternating direction scheme to the parabolic equation

Let us solve the following boundary value problem for the second-order parabolic equation without mixed derivatives

$$\begin{aligned}&\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega\times\Omega_t,\\ &\varphi|_{\partial\Omega}=\varphi_{(\Gamma)}(x,t),\\ &\varphi=g(x) \quad \text{in } \Omega \qquad \text{for } t=0,\end{aligned} \tag{29.1}$$

where

$$A=\sum_{\alpha=1}^{2}A_\alpha, \quad A_\alpha=-\frac{\partial}{\partial x_\alpha}k\frac{\partial}{\partial x_\alpha},$$

$$x=(x_1,x_2)\in\Omega=\{0<x_\alpha<1:\alpha=1,2\},$$

$$k_\alpha>0.$$

We assume that problem (29.1) has a unique and sufficiently smooth solution. THOMPSON [1964] proved the existence and uniqueness of a generalized solution for (29.1) provided that the function $f(x,u)$ satisfies Lipschitz's condition in the variable u. He also showed that the convergence of the solution of difference problem to the solution of differential one follows from the approximation and stability of the corresponding difference operator.

We construct in $\bar{\Omega}=\Omega+\partial\Omega$ grid $\bar{\Omega}^h$ with steps h_α that is uniform in x_α and replace the operators A_α by the difference operators Λ_α:

$$\Lambda_\alpha\varphi=(\Delta x_\alpha k_\alpha^h\nabla_{x_\alpha}\varphi)/h_\alpha^2. \tag{29.2}$$

Contrary to the case of constant coefficients, here, the operators Λ_α are positive and self-adjoint but not commutative. Instead of (29.1) we consider the following

problem approximating (29.1):

$$\begin{aligned} &\partial\varphi/\partial t + \Lambda\varphi = f \quad \text{in } \Omega^h, \quad \Lambda = \Lambda_1 + \Lambda_2, \\ &\varphi|_{\partial\Omega^h} = \varphi^h_{(\Gamma)}, \\ &\varphi = g^h \quad \text{in } \Omega^h \qquad \text{for } t = 0. \end{aligned} \tag{29.3}$$

Now φ and f are vectors and Λ_1 and Λ_2 are matrices. Suppose solution φ belongs to the grid function space Φ.

For solving problem (29.3) in the layer $t_j \leqslant t \leqslant t_{j+1}$ we write the alternating direction scheme in the form (SAMARSKII [1971], pp. 360–365):

$$\begin{aligned} &\frac{\varphi^{j+1/2} - \varphi^j}{\tau} + \tfrac{1}{2}(\Lambda_1^{j+1/2}\varphi^{j+1/2} + \Lambda_2^j \varphi^j) = \tfrac{1}{2} f^j, \\ &\frac{\varphi^{j+1} - \varphi^{j+1/2}}{\tau} + \tfrac{1}{2}(\Lambda_1^{j+1/2}\varphi^{j+1/2} + \Lambda_2^{j+1}\varphi^{j+1}) = \tfrac{1}{2} f^j, \\ &\quad j = 0, 1, 2, \ldots, \end{aligned} \tag{29.4}$$

with initial condition

$$\varphi_1^0 = g. \tag{29.5}$$

Equations (29.4) and (29.5) must be supplemented by difference boundary conditions, which may be written, for example, in the form (see KRYAKVIN [1966]):

$$\varphi^{j+1}|_{\partial\Omega_1^h} = \varphi_{(\Gamma)}(t_{j+1}), \qquad \varphi^{j+1/2}|_{\partial\Omega_2^h} = \bar{\varphi}_{(\Gamma)}, \tag{29.6}$$

where $\partial\Omega_\alpha^h$ is the grid on the sides $x_\alpha = 0$ and $x_\alpha = 1$,

$$\varphi_{(\Gamma)} = \tfrac{1}{2}[\varphi_{(\Gamma)}(t_{j+1}) + \varphi_{(\Gamma)}(t_j)] + \tfrac{1}{4}\tau\Lambda_2[\varphi_{(\Gamma)}(t_{j+1}) - \varphi_{(\Gamma)}(t_j)]. \tag{29.7}$$

We rewrite (29.4) in the form:

$$\begin{aligned} &\frac{2}{\tau}\varphi^{j+1/2} + \Lambda_1^{j+1/2}\varphi^{j+1/2} = F^j, \quad F^j = \frac{2}{\tau}\varphi^j - \Lambda_2^j\varphi^j + f^j, \\ &\frac{2}{\tau}\varphi^{j+1} + \Lambda_2^{j+1}\varphi^{j+1} = F^{j+1/2}, \\ &F^{j+1/2} = \frac{2}{\tau}\varphi^{j+1/2} - \Lambda_2^{j+1/2}\varphi^{j+1/2} + f^j. \end{aligned} \tag{29.8}$$

Each problem in (29.8) with the corresponding boundary conditions from (29.6) is a one-dimensional first boundary value problem and it can be solved by the sweep method.

If the value k_α^h in the grid point i_α is calculated, for example, by the formula

$$(k_\alpha^h)_{i_\alpha} = \tfrac{1}{2}[k_\alpha(x_{i_\alpha}) + k_\alpha(x_{i_\alpha+1})], \quad 1 \leqslant i_\alpha < I_\alpha,$$

then operator Λ_α approximates operator A_α with second-order accuracy, i.e.

$$\Lambda_\alpha\varphi - A_\alpha\varphi = O(h_\alpha^2).$$

Let $k_\alpha = \text{const}$. Operator Λ is self-adjoint and positive in Φ. Introduce the metric

$$\|\varphi\|^2 = \sum_{i_1=1}^{I_1-1} \sum_{i_2=1}^{I_2} (\Delta_{i_1}\varphi)^2 h_1 h_2 + \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2-1} (\Delta_{i_2}\varphi)^2 h_1 h_2. \tag{29.9}$$

In this metric scheme (29.4)–(29.6) is stable on the initial and boundary conditions and on the right-hand side.

Assume that, in $\Omega \times \Omega_t$, (29.1) has a unique solution $\varphi = \varphi(x,t)$ which is continuous and has bounded derivatives

$$\frac{\partial^3 \varphi}{\partial t^3}, \quad \frac{\partial^5 \varphi}{\partial x_\alpha^2 \partial x_\beta^2 \partial t}, \quad \frac{\partial^4 \varphi}{\partial x_\alpha^4}, \qquad 0 \leqslant \alpha, \beta \leqslant 2. \tag{29.10}$$

Then scheme (29.4)–(29.6) converges in the grid norm (29.9) with rate $O(\tau^2 + |h|^2)$ so that

$$\|\varphi^j - \varphi(t_j)\|_\Lambda \leqslant M(|h|^2 + \tau^2),$$

where $M = \text{const}$ does not depend on τ and $|h|$.

REMARK 29.1. The alternating direction scheme with varying k_α has an approximation of order $O(|h|^2 + \tau^2)$ for the solution $\varphi = \varphi(x,t)$, provided that, in addition to (29.10), the obvious requirements of smoothness of $k_\alpha(x,t)$ in x_α and t are satisfied. The distinction from the case of constant coefficients reveals itself in the stability study of the scheme, since though the operators Λ_1, Λ_2 are positive and self-adjoint, they are not commutative. However if, in addition to (29.10), there also exists in $\Omega \times \Omega_t$ a continuous derivative

$$\frac{\partial}{\partial x_1}\left(k_1 \frac{\partial^2}{\partial x_1 \partial x_2}\left(k \frac{\partial \varphi}{\partial x_2}\right)\right),$$

then scheme (29.4)–(29.6) is stable and converges with rate $O(\tau^2 + |h|^2)$, in the case of varying coefficients k_α.

CHAPTER VII

Methods of Splitting with Respect to Physical Processes

30. The method of splitting with respect to physical processes

The method of splitting with respect to physical processes consists of reducing the original evolutionary problem describing complex physical processes to a sequence of problems describing processes of a more simple structure. To explain the idea of the method we will consider the equation describing the process of advection-diffusion of some substance (MARCHUK [1982], pp. 82–90).

The two-dimensional advection-diffusion problem consists of finding the solution of the equation

$$\frac{\partial \varphi}{\partial t}+\frac{\partial u\varphi}{\partial x}+\frac{\partial v\varphi}{\partial y}-\mu\Delta\varphi+\sigma\varphi=f, \tag{30.1}$$

$$\varphi=g \quad \text{for } t=0, \tag{30.2}$$

$$\sigma\geqslant 0, \qquad \mu\geqslant 0.$$

Assume for simplicity that we find the solution (x, y) of (30.1)–(30.2) on the whole plane and that the functions u and v satisfy the continuity equation

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0. \tag{30.3}$$

From the physical point of view problem (30.1)–(30.2) describes, in principle, two different physical processes. The first one is the process of advection of a substance along the trajectory and it is described by

$$\frac{\partial \varphi_1}{\partial t}+\frac{\partial u\varphi_1}{\partial x}+\frac{\partial v\varphi_1}{\partial y}=0, \tag{30.4}$$

$$\varphi_1=g_1 \quad \text{for } t=0. \tag{30.5}$$

The second one is connected with diffusion and absorption of a substance and is given by

$$\frac{\partial \varphi_2}{\partial t}-\mu\Delta\varphi_2+\sigma\varphi_2=f, \tag{30.6}$$

$$\varphi_2=g_2 \quad \text{for } t=0. \tag{30.7}$$

These processes represent two extreme cases of the problem (30.1)–(30.2). Indeed, taking in (30.1)–(30.2) $\mu=0$ and $\sigma=0$ we obtain problem (30.4)–(30.5) and taking $u=v=0$ we obtain problem (30.6)–(30.7).

Let us now try to put (30.4)–(30.5) and (30.6)–(30.7) together. This appears to be possible because of the additivity of the processes of advection and diffusion locally.

Assume that u and v are constants and $g(x, y)$ decreases quickly enough at infinity so that the following representations take place

$$\varphi=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Phi(m,n,t)e^{imx+iny}\,dm\,dn,$$
$$g=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(m,n)e^{imx+iny}\,dm\,dn. \tag{30.8}$$

Then the exact solution of the original problem (30.1)–(30.2) has the form

$$\varphi=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(m,n)\exp\{im(x+ut)+in(y+vt) -[\sigma+\mu(m^2+n^2)]t\}\,dm\,dn. \tag{30.9}$$

Let us now split the problem into two. To that end we consider the time interval $0<t<\tau$ on which we will solve the two problems successively: first (30.4)–(30.5) with the initial condition

$$\varphi_1=g$$

and next (30.6)–(30.7) with the initial condition

$$\varphi_2=\varphi_1.$$

We assume as before that φ_1, φ_2, g are represented as Fourier integrals as in (30.8).

For Fourier components of the solution of (30.4)–(30.5) we find

$$\Phi_1(t)=A\exp\{(imu+inv)t\} \tag{30.10}$$

and for those of (30.6)–(30.7),

$$\Phi_2(t)=\Phi_1(t)\exp\{-[\sigma+\mu(m^2+n^2)]t\}. \tag{30.11}$$

When we take $t=\tau$ in (30.10) and substitute $\Phi_1(\tau)$ in (30.11), we obtain

$$\Phi_2(\tau)=A\exp\{(imu+inv)\tau-[\sigma+\mu(m^2+n^2)]\tau\}.$$

Therefore

$$\varphi_2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}A(m,n)\exp\{im(x+u\tau)+in(y+v\tau)-[\sigma+\mu(m^2+n^2)]\tau\}\,dm\,dn.$$

Thus the solution obtained by the splitting method for $t=\tau$ is identical to the solution of the original problem (30.1)–(30.2). This fact lies at the basis of the method of splitting with respect to physical processes.

Note that in reality u and v are not constant; in this case the splitting algorithm does not give the exact solution for $t=t_j, j=1,2,\ldots$. To correct the result in this case the time interval should be taken small enough to ensure good approximation in cases when the coefficients u and v vary considerably.

The method of splitting with respect to physical processes is widely used for solving problems in hydrodynamics, gas dynamics (KOVENYA and YANENKO [1981], MIGUAL, PINSKY and TAYLOR [1983], GUSHCHIN [1981], TEMAM [1981]), meteorology and oceanology (MARCHUK [1967, 1974a, 1982], MARCHUK, DYMNIKOV et al. [1984]) etc.

CHORIN [1968] proposed the original method for finding the approximate solution of the equations for viscous incompressible liquid. This method is based on Helmholtz's theorem on the decomposition of the vector field. (A vector field may be decomposed in the unique way into a solenoidal part A_f and a potential part A_i, provided that the normal component A_f equals zero on the boundary.) The theoretical grounding of the method is given by TEMAM [1981]. YANENKO, ANUCHINA, PETRENKO and SHOKIN [1970] interpreted this method from the standpoint of splitting with respect to physical processes. A modification of this method was used by KUZNETSOV, MOSHKIN and SMAGULOV [1983] to solve the nontraditional Navier–Stokes problem when the pressure and tangential component of the velocity on some part of boundary are given.

31. The method of particles in a cell

Methods for solving multidimensional problems in mathematical physics are suggested by HARLOW [1967] and HARLOW and WELCH [1965] and named *methods of particles in a cell* or *methods of markers and cells* (MAC method). These methods have been developed intensively in recent years. They are also based on the idea of splitting the physical process into more simple ones and are applied to calculations of multidimensional hydrodynamic currents with strong liquid deformations, large relative displacements and colliding interfaces.

The essence of the method of particles in a cell is as follows. The hydrodynamic equations are reduced using weak approximation on each small time interval to two more simple systems, the first of which describes the adaptation of the hydrodynamic fields omitting advective terms and is integrated by the usual methods in a fixed Euler grid; the second one describes the advection of substances in a Lagrange coordinate system. Only while solving the second system one uses the phenomenological simplification of the continuous medium model replacing it with the system of particles in each cell of the Euler system in such a way that the total mass, momentum and energy of the particles in a cell are identified with the corresponding characteristics for the continuous medium.

When some particle "carrying" certain mass along its trajectory crosses the

boundary of the cell, the mass, momentum and energy of this particle are subtracted from its "old" cell and are added to the "new" cell. Harlow's scheme is based on the explicit methods for solving equations of the first and the second stage and is conditionally stable in the whole scheme.

It is especially fruitful to use implicit schemes in calculations of the first step,. In this case the criterion of stability of the whole scheme coincides with the well-known Courant condition. Some improvements of the MAC method in connection with the introduction of fractional cells and calculations of the locations of markers were suggested by CHEN with co-authors [1970, 1991] and NICOLS [1973].

32. The method of large particles

DYACHENKO [1969], BELOTSERKOVKY and DAVYDOV [1970], YANENKO, ANUCHINA, PETRENKO and SHOKIN [1970, 1971] give various modifications of the method of particles in a cell which diminish substantially the inherent fluctuations of density and pressure and extend the "stability reserve". Various realization schemes are considered. These modifications are currently referred to as the method of large particles.

The main idea of the nonstationary method of large particles consists of splitting the original system of equations into physical processes. We will describe this method "on the physical level" applied to the problem of motion of an ideal compressible gas.

The model medium is replaced by the system of fluid particles coinciding at a given moment with cells of an Euler grid. The stationary solution of the problem (if there exists any) is obtained as a result of positioning; therefore, the whole computational process consists of repeated time steps. The calculations on each time step proceed in three stages:

In the first stage, the variation of momentum and energy of the Lagrange elementary liquid volume (large particle) contained in a Euler cell over time interval is considered (in this stage the boundary of the volume is displaced with respect to the initial position).

In the second stage, the movement of the particles through the boundaries of the Euler cells is modelled and the particles are redistributed in space.

In the third stage, the redistribution of mass, momentum and energy in space takes place (the variation of the flux parameters in the elementary Euler cell obtained by returning the Lagrange volume to the initial position over time interval τ is determined).

Thus the evolution of the whole system over time interval τ is calculated using the following splitting: first the variation of the internal state of the subsystems in the cells/large particles is considered assuming that they are frozen (Euler stage), and next the displacements of all particles proportional to their velocity and time interval τ without variation of the internal state of the subsystem are considered with subsequent recomputation of the computation grid in the initial position (Lagrange and final stages).

Consider the system of equations describing the motion of an ideal compressible gas

$$\frac{\partial \rho u}{\partial t} + \operatorname{div}(\rho u \boldsymbol{u}) + \frac{\partial p}{\partial x} = 0,$$

$$\frac{\partial \rho v}{\partial t} + \operatorname{div}(\rho v \boldsymbol{u}) + \frac{\partial p}{\partial y} = 0,$$

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \boldsymbol{u}) = 0, \tag{32.1}$$

$$\frac{\partial \rho E}{\partial t} + \operatorname{div}(\rho E \boldsymbol{u}) + \operatorname{div}(p \boldsymbol{u}) = 0,$$

$$p = p(\rho, J), \qquad J = E - \tfrac{1}{2}\boldsymbol{u}^2.$$

Here u and v are the components of velocity vector $\boldsymbol{u}$, p is the pressure, ρ is the density, J and E are the specific internal and total energy respectively.

Let the domain where the solution of (32.1) is to be found be covered by an Euler (fixed in space) computation grid consisting of rectangular cells with sides h_1 and h_2. The solution of problem (32.1) consists of the following steps:

In the first (Eulerian) stage, only quantities related to the whole cell are varied, while the liquid is considered to be "instantaneously frozen". We find the solution of the system:

$$\rho \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \qquad \rho \frac{\partial v}{\partial t} + \frac{\partial p}{\partial y} = 0,$$

$$\rho \frac{\partial E}{\partial t} + \operatorname{div}(p \boldsymbol{u}) = 0, \qquad \frac{\partial \rho}{\partial t} = 0, \tag{32.2}$$

$$p = p(\rho, J), \qquad J = E - \tfrac{1}{2}\boldsymbol{u}^2.$$

In the next (Lagrangean) stages, the motion of the mass flux (particles) ΔM through the boundaries of the Euler cells is modelled, the redistribution of mass, momentum and energy in space take place and the final fields of Euler parameters are determined.

The equations of these stages may be combined formally and represented in the form

$$\begin{aligned} &\partial \rho u / \partial t + \operatorname{div}(\rho u \boldsymbol{u}) = 0, \\ &\partial \rho v / \partial t + \operatorname{div}(\rho v \boldsymbol{u}) = 0, \\ &\partial \rho / \partial t + \operatorname{div}(\rho \boldsymbol{u}) = 0, \\ &\partial \rho E / \partial t + \operatorname{div}(\rho E \boldsymbol{u}) = 0. \end{aligned} \tag{32.3}$$

Various versions of the numerical schemes of the large particle method along with some aspects of approximation and stability are presented by BELOTSERKOVSKY and DAVYDOV [1978], BELOTSERKOVSKY [1984], DAVYDOV and SKOTNIKOV [1978] and others.

CHAPTER VIII

The Alternating Triangular Method and the Alternating Operator Method

The alternating direction method proposed by Peaceman, Douglas and Rachford is usually connected with the decomposition of the operator A into "one-dimensional" operators A_α, the sweep method being used for the solution of equations at each fractional step. However (as it was already noted in Section 26) the generalization of this method for problems with three spatial variables faces some difficulties. On the other hand the requirement that the operators A_α are "one-dimensional" is not compulsory. Therefore it is of interest to split the operator A in operators that would allow an effective solution of the problem in each step and would keep the main advantages of the alternating direction method. Such a splitting may be the decomposition of the matrix operator $A=A^*$ into two matrices A_1 and A_2 with $A_1=A_2^*$, $A_1+A_2=A$. Writing these alternating direction schemes formally we obtain schemes of the *alternating triangular method* which was proposed and grounded for some problems of mathematical physics by SAMARSKII [1964a] and IL'IN [1965]. IL'IN [1966] also presented generalizations of the alternating direction methods where A_1 and A_2 are arbitrary matrices (in particular, triangular). These generalizations are called schemes of *the alternating operator method.*

33. The alternating triangular method

Let us consider the evolutionary problem

$$\begin{aligned} &\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega_t, \\ &\varphi=g \quad \text{for } t=0, \end{aligned} \tag{33.1}$$

where $\Omega_t=\{t: 0<t<T\}$ and $A=A(t)$ is a square matrix representing a finite-dimensional approximation of the corresponding operator of the original problem in all variables except for t. Operator A from (33.1) acts in a Hilbert space Φ with inner product $(u,v)_\Phi$ and norm $\|u\|_\Phi=(u,u)_\Phi^{1/2}$. It is assumed that A can be represented as the sum of two triangular positive-definite matrices

$$\begin{aligned} &A=A_1+A_2, \qquad A_1=(a_{ik}^{(1)}), \quad A_2=(a_{ik}^{(2)}), \\ &a_{ik}^{(1)}=0, \quad k>i, \qquad a_{ii}^{(1)}+a_{ii}^{(2)}=a_{ii}, \end{aligned} \tag{33.2}$$

$$(A_\alpha u,u)\geqslant C\cdot\|u\|^2, \qquad C=\text{const}>0, \quad \alpha=1,2. \tag{33.3}$$

If A is a symmetric matrix, then $A_1 = A_2^*, A_1^* = A_2$ and we may take $a_{ii}^{(1)} + a_{ii}^{(2)} = \frac{1}{2}a_{ii}$.

Let us write the scheme of the alternating triangular method (SAMARSKII [1964a] as

$$\begin{aligned}
&\frac{\varphi^{j+1/2}-\varphi^j}{\tau}+\tfrac{1}{2}(A_1^{j+1/2}\varphi^{j+1/2}+A_2^j\varphi^j)=\tfrac{1}{2}f^j,\\
&\frac{\varphi^{j+1}-\varphi^j}{\tau}+\tfrac{1}{2}(A_1^{j+1/2}\varphi^j+A_2^{j+1}\varphi^{j+1})=\tfrac{1}{2}f^j,\\
&\quad j=0,1,\ldots,\\
&\varphi^0=g,
\end{aligned} \tag{33.4}$$

where

$$A_1^{j+1/2}=A_1(t_{j+1/2}), \qquad A_2^j=A_2(t_j),$$
$$f^j=f(t_{j+1/2}),$$
$$\tau=t_{j+1}-t_j.$$

It is easy to notice that to solve the first equation from (33.4) we need to invert the triangular matrix $E+\frac{1}{2}\tau A_1^{j+1/2}$ and to solve the second equation we need to invert the triangular matrix $E+\frac{1}{2}\tau A_2^{j+1}$. SAMARSKII [1964a] noted that the inversion, when the order of the matrix is more than five, will require fewer operations than to solve problem (33.1) by the explicit two-layer scheme.

If conditions (33.3) are satisfied and $A_2(t)\varphi(t)$ and $\mathrm{d}\varphi/\mathrm{d}t \in C^{(1,1)}[0,T]$, then (33.4) has second-order accuracy in τ:

$$\|\varphi^j-\varphi(t_j)\| \leqslant C\cdot\tau^2, \quad j=1,2,\ldots,$$

where C is a positive constant independent of τ (SAMARSKII [1964a].

REMARK 33.1. Scheme (33.4) and the above statement remain valid when A_1 and A_2 are arbitrary linear operators acting in a Hilbert space and satisfying conditions (33.3) (SAMARSKII [1964a]).

REMARK 33.2. If A_α are self-adjoint operators, then scheme (33.4) is a generalization of the alternating direction method proposed by Peaceman, Douglas and Rachford, and it may be considered as the alternating operator scheme.

REMARK 33.3. For the problem

$$\begin{aligned}
&\partial^2\varphi/\partial t^2+A(t)\varphi=f \quad \text{in } \Omega_t,\\
&\varphi=g_0 \quad \text{for } t=0,\\
&\partial\varphi/\partial t=g_1 \quad \text{for } t=0,
\end{aligned} \tag{33.5}$$

SAMARSKII [1964a] formulated also a scheme of the alternating triangular method which has second-order accuracy.

REMARK 33.4. Since the solution of (33.4) may be carried out with explicit formulae, schemes of the alternating triangular method are also called *explicit alternating direction schemes* (IL'IN [1970], p. 109).

34. The alternating operator method

We assume that operator A in (33.1) does not depend on t and can be represented as

$$A=A_1+A_2,$$

where the only requirement for A_α is positive-semidefiniteness

$$(A_\alpha u,u)_\Phi\geqslant 0.$$

We write for (33.1) a scheme of the following kind (IL'IN [1966]):

$$\begin{aligned}&\frac{\varphi^{j+1/2}-\varphi^j}{\tau_j}+\tfrac{1}{2}(A_1\varphi^{j+1/2}+A_2\varphi^j)=\tfrac{1}{2}f^j,\\&\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau_j}+\tfrac{1}{2}A_2(\varphi^{j+1}-\varphi^j)=k_j\frac{\varphi^{j+1/2}-\varphi^j}{\tau_j},\qquad (34.1)\\&j=0,1,\ldots,\\&\varphi^0=g,\end{aligned}$$

where τ_j and k_j are parameters, $f^j=f(t_{j+1/2})$, and operator A and vector f^j are constructed taking into account the boundary conditions of the original problem approximated by (33.1). Formally scheme (34.1) coincides for $k_j=0$ with the Douglas–Rachford scheme and for $k_j=1$ with the Peaceman–Rachford scheme. After excluding $\varphi^{j+1/2}$ (34.1) is transformed to

$$\frac{\varphi^{j+1}-\varphi^j}{\frac{1}{2}\tau_j(1+k_j)}+A\frac{\varphi^{j+1}+k_j\varphi^j}{1+k_j}+\tfrac{1}{4}\tau_jA_1A_2\frac{\varphi^{j+1}-\varphi^j}{\frac{1}{2}\tau_j(1+k_j)}=f^j.\qquad (34.2)$$

If depending on parameters τ_j and k_j the time step for the original problem is chosen as

$$\Delta t_j=\tfrac{1}{2}\tau_j(1+k_j)\qquad (34.3)$$

the order of approximation of (34.1) is equal to $O(\tau_j+k_j\tau_j)$. In the case $k_j=0$, it has a first-order approximation in $\Delta t_j=\frac{1}{2}\tau_j$ and, in the case $k_j=1$, it has a second-order approximation (with respect to $\Delta t_j=\tau_j$).

Let us assume that the approximation error of both equations in (34.1) on the smooth solution of (33.1) tends to zero in the norm $\|\cdot\|_\Phi$ for $\tau=\max\tau_j\to 0$, then the solution φ^j of the difference scheme (34.1) converges to $\varphi(t_j)$ in the norm

$$\|\varphi\|_{A_2}=(\|\varphi\|_\Phi+\|A_2\varphi\|_\Phi^2)^{1/2}\quad\text{for }\tau\to 0$$

(IL'IN [1966].

Note that we still did not make an assumption concerning the structure of the operators A_1 and A_2. Further, the restrictions on A_α in (34.1) are proved to be

satisfied provided that they can be represented as in (33.2)–(33.3) (and so we obtain the schemes of the alternating triangular method).

REMARK 34.1. IL'IN [1966] noted the possibility to generalize (34.1) for $k_j=1$ on quasi-linear problems as the error of approximation reaches the order $O(\tau_j^2)$.

35. The generalized alternating operator method

Let the operator A in (33.1) be represented in the form

$$A=\sum_{\alpha=1}^{n} A_\alpha,$$

where the operators A_α are not necessarily related to differential operators with respect to some single variable (if, of course, they belong to the original problem). The following scheme is a generalization of (34.1) (IL'IN [1966]):

$$\begin{aligned}
&\frac{\varphi^{j+1/n}-\varphi_j}{\tau_j}+A_1\varphi^{j+1/n}+\sum_{\alpha=2}^{n} A_\alpha\varphi^j=f^j,\\
&\frac{\varphi^{j+k/n}-\varphi^{j+(k-1)/n}}{\tau_j}+A_k(\varphi^{j+k/n}-\varphi^j)=\beta_{k-1}^j(\varphi^{j+(n-1)/n}-\varphi^j)/\tau_j, \qquad (35.1)\\
&k=1,\ldots,n,
\end{aligned}$$

where n is the number of fractional steps on the interval (t_j, t_{j+1}). After elimination of the intermediate values $\varphi^{j+k/n}$, we have

$$B\frac{\varphi^{j+1}-\varphi^j}{\tau_j}+A\varphi^j=f^j, \qquad (35.2)$$

where

$$B=\prod_{\alpha=1}^{n}(E+\tau_j A_\alpha)/\omega_j, \quad \omega_j=\prod_{\alpha=1}^{n-1}(1+\beta_\alpha^j).$$

With $\omega_j=2$ scheme (35.1) approximates (33.1) with error $O(\tau^2)$. With $\beta_\alpha^j=0$ it coincides with the Douglas–Rachford stabilizing correction scheme, which has error $O(\tau)$.

36. The scheme of the alternating triangular method for the parabolic equation

Let us consider one of the concrete realizations of the alternating triangular method (IL'IN [1965]) for the two-dimensional heat conduction equation

$$\partial\varphi/\partial t-\operatorname{div}(D(x,y)\operatorname{grad}\varphi)=0. \qquad (36.1)$$

We write the following five-point difference scheme

$$\frac{\varphi_{k,l}^{j+1}-\varphi_{k,l}^{j}}{\Delta t_j}+\Lambda\varphi_{k,l}^{j+1}=0. \tag{36.2}$$

Here and further

$$h_x=h_y=h, \qquad \Delta t_j=t_{j+1}-t_j,$$

$$\Lambda\varphi_{k,l}=\frac{1}{h^2}(p_{k,l}\varphi_{k,l}-a_{k,l}\varphi_{k-1,l}-b_{k,l}\varphi_{k,l-1}-c_{k,l}\varphi_{k+1,l}-d_{k,l}\varphi_{k,l+1}), \tag{36.3}$$

where the coefficients a, b, c, d, p may be variables. We also suppose that the boundary conditions for (36.1) are taken into account for the coefficients of the difference equation.

A scheme of the alternating triangular method has the form

$$\begin{aligned}
&\frac{\varphi_{k,l}^{j+1/2}-\varphi_{k,l}^{j}}{\Delta t_j}+\frac{1}{2h^2}\left(\tfrac{1}{2}p_{k,l}\varphi_{k,l}^{j+1/2}-a_{k,l}\varphi_{k-1,l}^{j+1/2}-b_{k,l}\varphi_{k,l-1}^{j+1/2}\right.\\
&\qquad\left.+\tfrac{1}{2}p_{k,l}\varphi_{k,l}^{j}-c_{k,l}\varphi_{k+1,l}^{j}-d_{k,l}\varphi_{k,l+1}^{j}\right)=0,\\
&\frac{\varphi_{k,l}^{j+1}-\varphi_{k,l}^{j+1/2}}{\Delta t_j}+\frac{1}{2h^2}\left(\tfrac{1}{2}p_{k,l}\varphi_{k,l}^{j+1/2}-a_{k,l}\varphi_{k-1,l}^{j+1/2}-b_{k,l}\varphi_{k,l-1}^{j+1/2}\right.\\
&\qquad\left.+\tfrac{1}{2}p_{k,l}\varphi_{k,l}^{j+1}-c_{k,l}\varphi_{k+1,l}^{j+1}-d_{k,l}\varphi_{k,l+1}^{j+1}\right)=0.
\end{aligned} \tag{36.4}$$

The condition of "spatial" stability with respect to the accumulation of errors on each time step has the form of inequalities,

$$\begin{aligned}
&1+\frac{\Delta t_j}{4h^2}p_{k,l}\geqslant\frac{\Delta t_j}{2h^2}(|a_{k,l}|+|b_{k,l}|),\\
&1+\frac{\Delta t_j}{4h^2}p_{k,l}\geqslant\frac{\Delta t_j}{2h^2}(|c_{k,l}|+|d_{k,l}|),
\end{aligned} \tag{36.5}$$

which must hold for all grid points except, maybe, for some of them. If we introduce the following operators Λ_1 and Λ_2

$$\Lambda_1\varphi_{k,l}=\frac{1}{h^2}(\tfrac{1}{2}p_{k,l}\varphi_{k,l}-a_{k,l}\varphi_{k-1,l}-b_{k,l}\varphi_{k,l-1}),$$

$$\Lambda_2\varphi_{k,l}=\frac{1}{h^2}(\tfrac{1}{2}p_{k,l}\varphi_{k,l}-c_{k,l}\varphi_{k+1,l}-d_{k,l}\varphi_{k,l+1}),$$

then (36.4) can be written as follows:

$$\begin{aligned}
&\frac{\varphi_{k,l}^{j+1/2}-\varphi_{k,l}^{j}}{\Delta t_j}+\tfrac{1}{2}(\Lambda_1\varphi_{k,l}^{j+1/2}+\Lambda_2\varphi_{k,l}^{j+1})=0,\\
&\frac{\varphi_{k,l}^{j+1}-\varphi_{k,l}^{j+1/2}}{\Delta t_j}+\tfrac{1}{2}(\Lambda_1\varphi_{k,l}^{j+1/2}+\Lambda_2\varphi_{k,l}^{j+1})=0,
\end{aligned} \tag{36.6}$$

or

$$(E+\tfrac{1}{2}\Delta t_j \Lambda_1)\varphi^{j+1/2}=(E-\tfrac{1}{2}\Delta t_j \Lambda_2)\varphi^j,$$
$$(E+\tfrac{1}{2}\Delta t_j \Lambda_2)\varphi^{j+1}=(E-\tfrac{1}{2}\Delta t_j \Lambda_1)\varphi^{j+1/2}. \tag{36.7}$$

On "whole steps" scheme (36.6) has the form:

$$(E+\tfrac{1}{2}\Delta t_j \Lambda_1)(E+\tfrac{1}{2}\Delta t_j \Lambda_2)\varphi^{j+1}$$
$$=(E+\tfrac{1}{2}\Delta t_j \Lambda_1)(E-\tfrac{1}{2}\Delta t_j \Lambda_2)\varphi^j \tag{36.8}$$

or

$$\frac{\varphi^{j+1}-\varphi^j}{\Delta t_j}+\tfrac{1}{2}\Lambda(\varphi^{j+1}+\varphi^j)+\tfrac{1}{4}\Delta t_j^2 \Lambda_1 \Lambda_2 \frac{\varphi^{j+1}-\varphi^j}{\Delta t_j}=0. \tag{36.9}$$

From (36.9) we conclude that (36.4) has second-order accuracy in τ (provided that the solution and the initial data are sufficiently smooth). If Λ and $\Lambda_1 \Lambda_2$ are positive-semidefinite operators, then it follows from a theorem by DOUGLAS and PIERCY [1963] that all eigenvalues of the operator

$$T_j=(E+\tfrac{1}{2}\Delta t_j \Lambda+\tfrac{1}{4}\Delta t_j^2 \Lambda_1 \Lambda_2)^{-1}(E-\tfrac{1}{2}\Delta t_j \Lambda+\tfrac{1}{4}\Delta t_j^2 \Lambda_1 \Lambda_2),$$

which is a transition operator from φ^j to φ^{j+1}, are less than 1. This implies the stability and convergence of the scheme.

Note the simplicity of the numerical realization of (36.4): computations on each half-step present the explicit scheme. The number of operations saved on one step is about 50 per cent. Scheme (36.4) may be generalized on multidimensional problems. In this case the operator Λ is split into two suitable operators Λ_1, Λ_2 in such a way that the computations on each half-step can be executed using an explicit scheme. The advantage of scheme (36.4) is more evident when the dimension of the problem increases.

REMARK 36.1. Besides the papers mentioned in this section the alternating triangular schemes (or "explicit alternating direction schemes") are also considered in papers by SAUL'EV [1960], IL'IN [1967], LEES [1961] and others. The alternating triangular method as an iterative method for solving systems of algebraic equations is thoroughly treated by SAMARSKII and NIKOLAEV [1978] and IL'IN [1970].

CHAPTER IX

Splitting Methods and Alternating Direction Methods as Iterative Methods for Stationary Problems

37. The stationing method: General concepts of the theory of iterative methods

The solution of many stationary problems with positive operators may be regarded as a limiting solution of a nonstationary problem for $t \to \infty$. While solving a stationary problem by methods of asymptotic stationing we pay no attention to intermediate values of the solution as they are of no interest. In the case of a nonstationary problem these values do have a physical meaning. In general, this is precisely what unites and divides these classes of problems.

Assume that we have a system of linear algebraic equations (representing, for example, the result of the approximation of some stationary problem in mathematical physics by the finite difference method)

$$A\varphi = f \tag{37.1}$$

where

$$A > 0, \qquad \varphi \in \Phi, \qquad f \in F.$$

We consider instead of (37.1) the nonstationary problem

$$\partial\psi/\partial t + A\psi = f,$$

and represent φ, ψ and f in the form

$$\varphi = \sum_n \varphi_n u_n, \qquad \psi = \sum_n \psi_n u_n, \qquad f = \sum_n f_n u_n,$$

where

$$Au_n = \lambda_n u_n, \qquad A^* u_n^* = \lambda_n u_n^*,$$

$$\varphi_n = (\varphi, u_n^*), \qquad \psi_n = (\psi, u_n^*), \qquad f_n = (f, u_n^*)$$

and $\{u_n\}$ and $\{u_n^*\}$ represent biorthogonal bases. Then using well-known tools we obtain, on the one hand, problems for Fourier coefficients

$$\lambda_n \varphi_n = f_n$$

and, on the other hand, the equations

$$\partial\psi_n/\partial t+\lambda_n\psi_n=f_n, \qquad \psi_n(0)=0.$$

Solving these problems we obtain

$$\varphi=\sum_n (f_n/\lambda_n)u_n,$$

$$\psi=\sum_n (f_n/\lambda_n)(1-e^{-\lambda_n t})u_n.$$

Assuming that the spectrum of the operator A is real we have $\lambda_n>0\,(n=1,2,\ldots)$. This implies that

$$\lim_{t\to\infty}\psi=\varphi.$$

When the operator of the stationary problem has a spectrum of an arbitrary structure such simple and evident relation between the solutions of the problems may not exist.

Of course, a nonstationary problem for a new function ψ may be solved using difference methods in t, for example,

$$\frac{\psi^{j+1}-\psi^j}{\tau}+A\psi^j=f.$$

Then

$$\psi^{j+1}=\psi^j-\tau(A\psi^j-f).$$

If it is our aim to solve the stationary problem, then—with certain relationship between τ and $\beta(A)$—we have

$$\lim_{j\to\infty}\psi^j=\varphi.$$

The parameter τ may either depend on j or not.

While solving the stationary problem it is convenient to regard j not as an index of the time step but as that of the iterative step.

Here we come across one more peculiarity: in nonstationary problems the values of τ must be small enough to ensure the accuracy of the solution, while *in stationary problems the optimal iterative parameters τ are chosen on the condition that the number of iterations should be minimal and they may be relatively large.*

The majority of iterative methods which are applied for solving linear systems may be united by the formula

$$B_j\frac{\varphi^{j+1}-\varphi^j}{\tau_j}=-\alpha(A\varphi^j-f), \tag{37.2}$$

where α is some positive number, $\{B_j\}$ is a sequence of non-singular matrices and $\{\tau_j\}$ is a sequence of real parameters. If we introduce the notion $H_j=\alpha\tau_j B_j^{-1}$, then the process (37.2) may be written in the following equivalent form:

$$\varphi^{j+1}=\varphi^j-H_j(A\varphi^j-f). \tag{37.3}$$

Vectors $\xi^j = A\varphi^j - f$ are usually called the *residual vectors* of the iterative method (37.2) and vectors $\psi^j = \varphi^j - \varphi^*$ ($\varphi^* = A^{-1}f$ is the exact solution of (37.1)) are called the *error vectors* of this method. Subtracting vector φ^* from both sides of relationship (37.3) and replacing f with $A\varphi^*$ we obtain a relationship for the sequence of error vectors:

$$\psi^{j+1} = T_j \psi^j, \tag{37.4}$$

where matrix

$$T_j = E - H_j A \tag{37.5}$$

is called the *jth step operator* of the iterative method (37.2). Multiplying (37.4) by the matrix A we obtain a relationship for the sequence of residual vectors:

$$\xi^{j+1} = (E - AH_j)\xi^j. \tag{37.6}$$

We will call the iterative method (37.2) *convergent* if the sequence $\{\varphi^j\}$ converges to the exact solution φ^* of (37.5) for any initial vector. It is obvious that the convergence of the sequences $\{\psi^j\}$ and $\{\xi^j\}$ to the zero vector for any ψ^0 and ξ^0 ($A\psi^0 = \xi^0$) is a necessary and sufficient convergence condition for iterative method (37.2).

The iterative method (37.2) is called *stationary* if the matrix H_j does not depend on the iteration number (the operator T_j is the constant matrix), otherwise it is called *nonstationary*. We will distinguish specially the class of cyclic iterative methods, which may be regarded both as stationary and nonstationary iterative methods.

A method is called *cyclic* if it has the property $H_j = H_{j+S}$ for any $j \geqslant 0$ and for some fixed $S \geqslant I$. It is easy to see that combining each S successive iterations into one we come to a stationary iterative process of the kind

$$\varphi^{j+1} = \varphi^j - \tilde{H}(A\varphi^j - f), \tag{37.7}$$

where H is determined by the equation

$$E - \tilde{H}A = \prod_{i=0}^{j-1} (E - H_i A). \tag{37.8}$$

On the other hand the cyclic iterative methods, as originally stated, belong to the nonstationary class.

One of our major tasks in studying iterative methods will be the problem of optimization, i.e. the choice of the sequence of matrices $\{H_j\}$ from a given class in order to obtain a more effective computational process.

38. Iterative algorithms

For the effective realization of method (37.2) the operator B_j should have a more simple structure as compared to A. In many cases of practical interest the operator B_j has the form

$$B_j \equiv B_{j,\tau} = \prod_{i=1}^{n} (E + \tau_j \tilde{B}_i), \tag{38.1}$$

where the $\tilde{B}_i$ are some matrices of the same order N as the matrix A. These matrices are chosen in such a way that the matrices $(E+\tau_j \tilde{B}_i)$ might be easily inverted, i.e. the inversion of the whole matrix B_j might be carried out more easily than the inversion of A. The matrices $\tilde{B}_i$ are often chosen with the representation (splitting) of the matrix A as the sum

$$A=\sum_{k=1}^{n} A_k. \tag{38.2}$$

taken in account.

First we take $n=m=2$, $\tau_j=\tau$ and give corresponding iterative algorithms for solving the system (37.1) which are constructed using the splitting and alternating direction schemes of previous sections.

(1) We get the *alternating direction method* (PEACEMAN and RACHFORD [1955]) for $\alpha=2$, $\tilde{B}_i=A_i$. We obtain from (37.2) an algorithm of the kind

$$(E+\tau_j A_1)(E+\tau_j A_2)\frac{\varphi^{j+1}-\varphi^j}{\tau_j}=-2(A\varphi^j-f), \tag{38.3}$$

which may also be written, after simple transformations, in "fractional steps":

$$\begin{aligned}&\frac{\varphi^{j+1/2}-\varphi^j}{\tau_j}+(A_1\varphi^{j+1/2}+A_2\varphi^j)=f,\\ &\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau_j}+(A_1\varphi^{j+1/2}+A_2\varphi^{j+1})=f.\end{aligned} \tag{38.4}$$

(2) We get the *stabilizing correction scheme* when $\alpha=1, B_i=A_i$ (DOUGLAS and RACHFORD [1956]):

$$(E+\tau_j A_1)(E+\tau_j A_2)\frac{\varphi^{j+1}-\varphi^j}{\tau_j}=-(A\varphi^j-f), \tag{38.5}$$

which may be rewritten in the form

$$\begin{aligned}&\frac{\varphi^{j+1/2}-\varphi^j}{\tau_j}+(A_1\varphi^{j+1/2}+A_2\varphi^j)=f,\\ &\frac{\varphi^{j+1}-\varphi^{j+1/2}}{\tau_j}+A_2(\varphi^{j+1}-\varphi^j)=0.\end{aligned} \tag{38.6}$$

(3) The *splitting (fractional steps) method* for an arbitrary value $m=n\geqslant 2$ may be represented in the form (MARCHUK and YANENKO [1966])

$$\begin{aligned}&\frac{\varphi^{j+1/n}-\varphi^i}{\tau_j}+A_1(\varphi^{j+1/n}-\varphi^j)=-\alpha_j\left(\sum_{k=1}^{n} A_k\varphi^j-f\right),\\ &\frac{\varphi^{j+k/n}-\varphi^{j+(k-1)/n}}{\tau_j}+A_k(\varphi^{j+k/n}-\varphi^j)=0,\\ &k=2,\ldots,n\end{aligned} \tag{38.7}$$

or, equivalently,

$$B_j \frac{\varphi^{j+1}-\varphi^j}{\tau_j} = -\alpha_j(A\varphi^j - f), \tag{38.8}$$

where

$$B_j = \prod_{k=1}^{n} (E+\tau_j A_k), \tag{38.9}$$

τ_j and α_j are some iterative parameters.

The other iterative algorithms may be formulated in a similar way.

39. Acceleration of the convergence of iterative methods

The acceleration of the convergence of the formulated iterative algorithms is achieved either by a special choice of the values of the parameters τ_j and α_j or by applying some adaptive accelerating procedure of the type of the conjugate gradient method to these algorithms.

We consider, for example, the case $n=m=2$ under the additional restriction $A_1 = A_2^T$. Let

$$\inf_{z\in\mathbb{R}^N} \frac{(A_i z, z)}{(z,z)} \geqslant S, \quad \sup_{z\in\mathbb{R}^N} \frac{\|A_i z\|^2}{(A_i z, z)} \leqslant \Delta, \qquad i=1,2,$$

where S and Δ are some positive constants. Then the effective value of the parameter τ in the stationary ($\tau_j \equiv \tau$) version of method (38.3) is taken from formula $\tau = (\delta\Delta)^{-1/2}$.

Since for this value of τ the eigenvalues of the matrix

$$B_\tau^{-1} A = (I+\tau A_2)^{-1}(I+\tau A_1)^{-1} A \tag{39.1}$$

are real, nonnegative and belong to the interval $[a,b]$ with the boundaries

$$a = \frac{\delta\sqrt{\Delta}}{\sqrt{\Delta}+\sqrt{\delta}}, \qquad b = \frac{\Delta\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}, \tag{39.2}$$

Chebyshev's semi-iterative method may be applied for the acceleration of method (38.3) (VARGA [1962]):

$$B_\tau(u^{k+1}-u^k) = -\alpha_{k+1}(Au^k - f) + \beta_{k+1} B_\tau(u^k - u^{k-1}),$$

$$\alpha_{k+1} = \frac{4}{b-a}\cdot\frac{C_k(\eta)}{C_{k+1}(\eta)}, \qquad \beta_{k+1} = \begin{cases} 0, & k=0 \\ \dfrac{C_{k-1}(\eta)}{C_{k+1}(\eta)}, & k>0, \end{cases} \tag{39.3}$$

$$k=1,2,\ldots,$$

where $\eta = (b+a)/(b-a)$.

For an arbitrary $\tau > 0$ it is advisable to use the generalized conjugate gradient method to accelerate method (38.3) (MARCHUK and KUZNETSOV [1972], CONCUS,

GOLUB and O'LEARY [1976]):

$$u^{k+1} = u^k - (1/q_k)[B_\tau^{-1}\,\xi^k - l_{k-1}(u^k - u^{k-1})],$$

$$l_{-1} = 0, \qquad q_k = \frac{\|B_\tau^{-1}\,\xi^k\|_A^2}{\|\xi^k\|_{B_\tau^{-1}}^2} - l_{k-1}, \qquad l_k = q_k \frac{\|\xi^{k+1}\|_{B_\tau^{-1}}^2}{\|\xi^k\|_{B_\tau^{-1}}^2}.$$

Other approaches to accelerate splitting methods are possible. For example, various versions of the descent method or the generalized conjugate direction method.

PART 2

Methods for Studying the Convergence of Splitting and Alternating Direction Schemes

Before we discuss the theory of convergence of splitting and alternating direction methods in detail, let us recall once more the sequence of stages we adhere to in this paper in solving approximately the problems of mathematical physics.

Thus, first of all, we approximate the initial problem in all independent variables (to be definite we will call them "spatial") but for time variable t. Finite difference methods, finite elements methods, projection methods, etc. can be used here. In case of a stationary problem we obtain the general system of linear algebraic equations

$$A\varphi = f, \tag{1}$$

and in the nonstationary case we get the following system of equations:

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0. \end{aligned} \tag{2}$$

System (1) can be realized by use of the well-known methods for solving linear algebraic equations including the iterative methods of splitting the operator A. At this stage we must prove the convergence of the applied iterative algorithms.

For the nonstationary problem (2) we can do as follows: either discretize all the problems or preliminarily split the system $\partial\varphi/\partial t + A\varphi = f$ on the system of differential equations of type $\partial\varphi_\alpha/\partial t + A_\alpha\varphi_\alpha = f_\alpha$ $(\alpha = 1, \ldots, n)$, and then obtain a completely discretized problem, using the difference scheme (possibly with additional splitting of the operators A_α). At the next stage of solving the problem we compute the systems of algebraic equations thus obtained (of type (1)). Iterative methods based on the splitting of the problem can be used at this stage. Note that it is possible to approximate the nonstationary problem (2) with the help of the finite element method to get the matrix problem of type (1); to solve it the splitting methods are applicable.

As we see, the first stage of solving the problem belongs to a different part of applied mathematics and it will therefore not be discussed in this paper. We will study the convergence of iterative methods for solving system (1) by splitting the

problem. Besides we are going to consider the aspects of grounding the transitions from a matrix evolutionary problem (2) to the systems of algebraic equations. Suppose, we make a direct transition from this problem to the systems of difference equations by splitting methods. To ground this stage we must study two problems: the approximation of the obtained scheme and its stability. When these are established, the convergence of the difference scheme's solution to the exact solution of problem (2) takes place (see the Introduction for Convergence Theorem 3.1).

If we preliminarily split the matrix evolutionary equation (2) on the system:

$$\partial\varphi_\alpha/\partial t + A_\alpha\varphi_\alpha = f_\alpha, \quad \alpha = 1, \ldots, n, \tag{3}$$

it is necessary to find out in what sense solutions $\{\varphi_\alpha\}$ approximate the exact solution φ, i.e. to ground this stage of solving the problem. The study of further stages of solving each of equations (3) is based as mentioned above on the studies of approximation and stability. Hence, to prove the convergence of splitting methods it is necessary to know how to study the following problems:

(1) the transition from system (2) to the splitting system (3);
(2) the approximation of splitting schemes;
(3) the stability of splitting schemes;
(4) the convergence of iterative methods for solving system (1), using the splitting technique.

Note, first of all, the methods to study approximation. They represent, as a whole, methods which are well developed in the theory of the finite difference methods (expansion in Taylor series, etc.). It should be taken into account here that it is not necessary in the splitting methods on the fractional steps that the solution of the scheme approximates the solution of problem (2). Therefore, the approximation is often checked for "integer" equations of the scheme (i.e. after excluding the "intermediate" equations of the scheme). Note one particular method that is frequently used for approximation in the splitting methods: after the scheme is obtained at the integer steps (and after a possible expansion of the transition operator in step τ of the time grid), it is compared with one of the well-studied schemes, e.g. the implicit scheme of first-order accuracy, the Crank–Nicolson second-order scheme, etc. Frequently, this comparison allows to prove that the order of accuracy of the splitting scheme is equivalent to that of one of these schemes, and that means that the approximation of the scheme under study is established. This method has been used in Part 1 for establishing the approximation orders of the schemes considered there. Taking the above notes into account we come to the conclusion that the first, the third, and the fourth of the abovementioned problems are specific in the theory of splitting and alternating direction methods. In this part, we will focus our attention on the description of the approximations to the solution of these problems.

CHAPTER X

Convergence Studies of the Splitting Schemes by Use of the Fourier Method (Spectral Method)

One of the general approaches to convergence studies of the splitting schemes is the use of the Fourier method (spectral method). Despite the number of restrictions of this method, it is widely used both for analyzing the stability and convergence of various difference schemes and for convergence studies of iterative algorithms. Consider the main concepts of this method.

40. General statement of the Fourier method

Assume that we solve the stationary problem (1) or the nonstationary problem (2) with the help of one of the splitting schemes described in Part 1, which can be reduced to the form:

$$\varphi^{j+1} = T\varphi^j + \tau F^j, \qquad \varphi^0 = g. \tag{40.1}$$

According to Chapter IX, scheme (40.1) should be treated as an iterative solution process for the stationary problem (1), and as a difference scheme for the nonstationary problem (2). Many of the schemes described in Chapters II–IX belong to this class of schemes. (The transition from schemes with fractional steps to schemes of type (40.1) is usually carried out by exclusion of the intermediate equations.)

The most important problem here is the convergence of the approximate solution, obtained by scheme (40.1), to the exact solution of problem (1) or (2). The convergence of φ^j to the exact solution of the stationary problem (1) is considered as the convergence of the iterative process (40.1), while the convergence to the exact solution of the nonstationary problem (2) is to be regarded as the convergence of the approximate solution of the difference scheme (40.1) when the time step τ is being decreased and, in general, may be included in operator T. To prove the convergence in the latter case, it is necessary to study the approximation and stability of scheme (40.1).

Assume that the approximation of (2) by scheme (40.1) is established. As mentioned above, this can be done by classical methods that are developed in the

theory of finite difference methods. To prove the convergence of the approximate solution (40.1) to the exact solution φ it is necessary to establish the stability of scheme (40.1). One of the methods for studying the stability of scheme (40.1) is the Fourier method or the spectral method based on the analysis of the spectrum of the transition operator T. At the same time this method allows to study the convergence of the iterative process (40.1) to the solution of the stationary problem (1). Let us formulate this method here.

Assume that the transition operator T is a linear operator acting in a certain Hilbert space Φ with domain $\mathscr{D}(A)=\Phi$, and $F^j, g, \varphi^j \in \Phi$.

Along with operator T, we introduce the adjoint operator T^* and consider two eigenvalue problems:

$$T\psi=\lambda\psi, \tag{40.2}$$

$$T^*\psi^*=\lambda\psi^*. \tag{40.3}$$

Assume that (40.2) and (40.3) determine two systems of eigenvalues $\{\lambda_n\}$ and $\{\lambda_n^*\}$ and two corresponding systems of eigenfunctions $\{\psi_n\}$, $\{\psi_n^*\}$, complete in Φ and orthonormalized in such a way that

$$(\psi_n, \psi_m^*)_\Phi=\begin{cases}1, & n=m,\\ 0, & n\neq m.\end{cases} \tag{40.4}$$

Here, $(\cdot,\cdot)_\Phi$ denotes the inner product defined in Φ.

Represent the function φ^j, F^j and g in the form of Fourier series:

$$\varphi^j=\sum_n \varphi_n^j\psi_n, \qquad F^j=\sum_n F_n^j\psi_n, \qquad g=\sum_n g_n\psi_n, \tag{40.5}$$

where φ_n^j, F_n^j, g_n are Fourier coefficients determined by

$$\varphi_n^j=(\varphi^j, \psi_n^*)_\Phi, \qquad F_n^j=(F^j, \psi_n^*)_\Phi, \qquad g_n=(g, \psi_n^*)_\Phi.$$

From (40.1) and (40.2) we obtain in the usual way the equations for the Fourier coefficients:

$$\varphi_n^{j+1}=\lambda_n\varphi_n^j+\tau F_n^j, \qquad \varphi_n^0=g_n. \tag{40.6}$$

In order to study the stability of difference schemes of type (40.1) Neumann introduced the so-called spectral criterion, the essence of which is given in the Introduction. If for each Fourier coefficient φ_n^j the following relationship holds:

$$|\varphi_n^j|\leqslant C_{1n}|g_n|+C_{2n}|F_n|, \quad n=1, 2, \ldots, \tag{40.7}$$

where C_{1n}, C_{2n} are positive constants, then the difference scheme (40.1) is declared to be stable (or computationally stable).

Let us see what conditions must be imposed on the parameters of the difference scheme (40.1) to satisfy inequality (40.7). Consequently excluding unknown quantities from (40.6) we obtain

$$\varphi_n^j=\lambda_n^j g_n+\tau\sum_{i=1}^{j}\lambda_n^{j-i}F_n^{i-1}.$$

Hence it follows that

$$|\varphi_n^j| \leqslant |\lambda_n^j|\,|g_n| + \tau \frac{1-|\lambda_n^j|}{1-|\lambda_n|}\,|F_n|, \tag{40.8}$$

where $|F_n| = \max_j |F_n^j|$.

The analysis of relationship (40.8) shows that the stability criterion will be satisfied, if

$$|\lambda_n| \leqslant 1, \quad n = 1, 2, \ldots. \tag{40.9}$$

Obviously, inequality (40.8) transforms, in this case, into

$$|\varphi_n^j| \leqslant |g_n| + \tau_j |F_n|. \tag{40.10}$$

If we solve the nonstationary problem (2) with the help of scheme (40.1) on the finite time interval, then as a rule, $\tau_j \leqslant \text{const}$. Hence, it follows from (40.10) that inequality (40.7) holds.

Thus, under condition (40.9) the difference scheme (40.1) is stable according to von Neumann. It follows from the study of approximation and stability that the approximate solution (40.1) converges to the exact solution φ of the nonstationary problem (2) (see the Introduction for Convergence Theorem 3.1).

It is easily seen that condition $|\lambda_n| < 1$ is also the condition for the convergence of the iterative process (40.1) to the solution of the stationary problem (1).

Note that, in solving nonstationary problems (2) with the help of scheme (40.1), the stability in the sense of (40.7) may also be established under the condition that

$$|\lambda_n| \leqslant 1 + C_0\tau, \quad C_0 = \text{const} > 0, \tag{40.11}$$

which is usually called von Neumann's necessary condition (see GODUNOV and RYABENKII [1967]). This condition allows a growth of the round-off errors, which is exponential in time.

The spectral criterion was already used to establish the splitting schemes in the very first papers on splitting methods (see PEACEMAN and RACHFORD [1955], DOUGLAS [1955, 1962], DOUGLAS and RACHFORD [1956]). It was also used in papers by BIRKHOFF and VARGA [1959], PEARCY [1962], DOUGLAS and PEARCY [1963], DOUGLAS and GUNN [1964], MARCHUK and YANENKO [1966], DYAKONOV [1962e], MARCHUK [1968, 1969] and others.

The spectral criterion is sometimes called harmonic. This term was introduced for studying the stability of the splitting schemes by YANENKO ([1967], p. 16), who formulated it in a somewhat different form though the essence remained the same. Let us illustrate this with an example of a homogeneous equation with constant coefficients. Let

$$\partial u/\partial t = \mathscr{L}(\mathscr{D})u \tag{40.12}$$

be an equation with constant coefficients, where u is a scalar function of $x_1, \ldots, x_m$ and t, and

$$\mathscr{L}(\mathscr{D}) = \sum_{\alpha} a_{\alpha_1 \ldots \alpha_m} \mathscr{D}_1^{\alpha_1} \cdots \mathscr{D}_m^{\alpha_m} \tag{40.13}$$

is a polynomial in the differentiation operators $\mathscr{D}_i = \partial/\partial x_i$.

Consider a harmonic function

$$u = u_0 e^{\omega t + ikx}, \qquad u_0 = \text{const}, \quad k \in \mathbb{Z}. \tag{40.14}$$

For (40.14) to be the solution it is necessary and sufficient that there exists a relationship between ω and k

$$\omega = \mathscr{L}(ik) \tag{40.15}$$

(the so-called dispersion or characteristic equation).

Let

$$\frac{u^{n+1} - u^n}{\tau} = \mathscr{L}\left(\frac{\Delta}{h}\right) u^h \tag{40.16}$$

be an explicit homogeneous approximation to (40.12). Here Δ/h is a certain approximation to $\mathscr{D}$, for example,

$$\frac{\Delta_i}{h_i} = \frac{T_i - E}{h_i} \sim \mathscr{D}_i,$$

where

$$T_i u(x_1, \ldots, x_m) = u(x_1, \ldots, x_i + h_i, \ldots, x_m), \qquad Eu = u.$$

Then (40.14) can be the solution of (40.16) if the difference dispersion equation

$$\frac{e^{\omega\tau} - 1}{\tau} = \mathscr{L}\left(\frac{e^{ikh} - 1}{h}\right)$$

$$\equiv \sum_{\alpha} \alpha_{\alpha_1 \ldots \alpha_m} \left(\frac{e^{ik_1 h_1} - 1}{h_1}\right)^{\alpha_1} \cdots \left(\frac{e^{ik_m h_m} - 1}{h_m}\right)^{\alpha_m} \tag{40.17}$$

holds.

Let $\rho = e^{\omega\tau}$, where ω is determined by the dispersion equation (40.17). Then the condition for the stability of scheme (40.16) will be the following

$$|\rho| \leqslant 1. \tag{40.18}$$

The value $\rho = \rho(k)$ is called the growth coefficient of the harmonic function (40.14). As we can see, this value precisely determines the stability of scheme (40.16).

Such a method for studying the stability has been applied in YANENKO's [1967] monograph not only to (40.16) but also to three-layer schemes and to fractional steps schemes, where the difference dispersion equation is written in terms of the coefficient ρ.

Let us show that this method for analyzing the stability applied to scheme (40.1) for $F^j = 0$ brings us to the earlier formulated spectral criterion.

Let problems (40.2) and (40.3) determine as before two complete systems of eigenfunctions $\{\psi_k\}$, $\{\psi_k^*\}$, orthonormalized according to (40.4). Consider the harmonic function

$$\varphi = \varphi_0 e^{\omega t} \psi_k, \tag{40.19}$$

where k is a fixed integer and $\varphi_0 = \text{const}$. Then (40.19) may be the solution of scheme (40.1) with $F^j = 0$ under conditions

$$g_k = \varphi_0 \psi_k \quad \text{and} \quad e^{\omega\tau} = \lambda_k,$$

where λ_k is the eigenvalue of operator T corresponding to ψ_k.

If $\rho = e^{\omega\tau}$, then the stability condition (40.18) transforms to $|\lambda_k| \leqslant 1$, i.e. coincides with (40.9).

Hence the harmonic stability analysis is nothing but the spectral analysis described above. As YANENKO ([1967], p. 16) notes, the harmonic stability analysis is very convenient in practice.

It may be useful to note that there exist a number of strong restrictions in applying the Fourier method to establish splitting schemes. First of all, it is required that operators T and T^* have a complete system of eigenfunctions. Besides, in using the spectral approach for the convergence studies of the splitting schemes it is necessary to find the eigenvalue of the operator T with a maximum absolute value or its upper estimate as part of the numerical algorithm for solving (40.1). Despite these restrictions, however, the Fourier method is still playing an exceptionally important role in applications

In the following two sections we will demonstrate the method's application for grounding splitting schemes in both stationary and nonstationary problems.

41. The Fourier method and the convergence studies of splitting schemes for stationary problems

Consider the stationary problem in the form:

$$A\varphi = f, \tag{41.1}$$

where

$$A = \sum_{\alpha=1}^{n} \Lambda_\alpha, \tag{41.2}$$

Λ_α are operators of a simpe structure acting in a Hilbert space Φ, $\mathscr{D}(\Lambda_\alpha) = \mathscr{D}(A)$, f is a given element of Φ and φ is the solution to be found.

To solve problem (41.1) consider the method of successive approximations in the form:

$$\prod_{\alpha=1}^{n} (E + \tfrac{1}{2}\tau_\alpha \Lambda_\alpha) \frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^j = f, \tag{41.3}$$

where τ and τ_α are arbitrary relaxation parameters, $\tau_\alpha, \tau > 0$. The difference equation (41.3) was formulated by DYAKONOV [1971d], DOUGLAS and GUNN [1964] and SAMARSKII [1977] (assuming that $\tau_\alpha = \tau$). MARCHUK and YANENKO [1966] considered (41.3) as the scheme for constructing various difference methods.

Obviously, scheme (41.3) is a special case of the general universal algorithm (see MARCHUK et al. [1981]):

$$B \frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^j = f, \tag{41.4}$$

where B is some "stabilizing" operator; in the given case:

$$B=\prod_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau_\alpha \Lambda_\alpha).$$

The realization scheme for each step of the iterative process (41.3) may be represented as

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau_1\Lambda_1)\xi^{j+1/n}=-\tau F^j,\\
&(E+\tfrac{1}{2}\tau_2\Lambda_2)\xi^{j+2/n}=\xi^{j+1/n},\\
&\quad\vdots\\
&(E+\tfrac{1}{2}\tau_n\Lambda_n)\xi^{j+1}=\xi^{j+(n-1)/n},\\
&\varphi^{j+1}=\varphi^j+\xi^{j+1},
\end{aligned}\tag{41.5}$$

where F^j is the residual of the relaxation process, determined by the formula

$$F^j=A\varphi^j-f. \tag{41.6}$$

It follows from the analysis of system (41.5) that each equation has a simple structure which implies the inversion of "simple" operators $(E+\frac{1}{2}\tau_\alpha\Lambda_\alpha)$.

Let us demonstrate the application of the Fourier method to study the convergence of (41.3).

Assume that the operators Λ_α, $\alpha=1,\ldots,n$, commute and generate a common basis of eigenfunctions. Then, we consider n eigenvalue problems:

$$\Lambda_\alpha u_\alpha=\lambda^{(\alpha)}u_\alpha,\quad \alpha=1,2,\ldots,n \tag{41.7}$$

and the corresponding adjoint problems

$$\Lambda_\alpha^* u_\alpha^*=\lambda^{(\alpha)}u_\alpha^*,\quad \alpha=1,2,\ldots,n. \tag{41.8}$$

Let the spectral problems (41.7) and (41.8) constitute n complete orthonormal systems of functions $\{u_{\alpha k}\}$ and $\{u_{\alpha l}^*\}$. Assume that function f in (41.3) can be expanded into a Fourier series in the eigenfunctions of problem (41.7):

$$f=\sum_{k=1}^{\infty} f_k\cdot u_{1k}u_{2k}\cdots u_{nk}. \tag{41.9}$$

We will also find the solution of (41.3) in the form of the series

$$\varphi^j=\sum_{k=1}^{\infty}\varphi_k^j u_{1k}\,u_{2k}\cdots u_{nk}. \tag{41.10}$$

The following statement is valid: The iterative process (41.3) converges if

(1) operators Λ_α, $\alpha=1,\ldots,n$, commute and generate a common basis of eigenfunctions;

(2) the eigenvalues $\{\lambda_k^{(\alpha)}\}$ of problem (41.7) are positive;

(3) $\tau_\alpha\geqslant\tau>0$.

Indeed, substitute (41.9) and (41.10) into (41.3) and take the inner product of the

result with $u_{1l}^*, u_{2l}^*, \ldots, u_{nl}^*$. Then we get the simplest equation for the Fourier coefficients φ_k^j:

$$\prod_{\alpha=1}^{n}(1+\tfrac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})\frac{\varphi_k^{j+1}-\varphi_k^j}{\tau}+\lambda_k\varphi_k^j=f_k, \quad k=1, 2, \ldots, \tag{41.11}$$

where

$$\lambda_k=\sum_{\alpha=1}^{n}\lambda_k^{(\alpha)}, \tag{41.12}$$

and $\lambda_k^{(\alpha)}$ is the eigenvalue of the operator Λ_α, corresponding to $u_{\alpha k}$.

Solving equation (41.11) with respect to the unknown φ_k^{j+1}, we have

$$\varphi_k^{j+1}=\frac{\prod_{\alpha=1}^{n}(1+\frac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})-\tau\lambda_k}{\prod_{\alpha=1}^{n}(1+\frac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})}\varphi_k^j+\frac{\tau f_k}{\prod_{\alpha=1}^{n}(1+\frac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})}. \tag{41.13}$$

Assume that the spectrum $\{\lambda_k^{(\alpha)}\}$ is positive. Then, it is easy to formulate the convergence criterion for the iterative process (41.13). To this end, consider the value

$$q_k=\frac{\prod_{\alpha=1}^{n}(1+\frac{1}{2}\tau\lambda_k^{(\alpha)})-\tau\lambda_k}{\prod_{\alpha=1}^{\alpha}(1+\frac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})}. \tag{41.14}$$

As follows from Section 40 the convergence criterion for the iterative process (41.13) (and therefore for (41.3) as well) is the condition

$$|q_k|<1. \tag{41.15}$$

This means that the following inequality must hold:

$$\prod_{\alpha=1}^{n}(1+\tfrac{1}{2}\tau_\alpha\lambda_k^{(\alpha)})>\tfrac{1}{2}\tau\lambda_k. \tag{41.16}$$

Obviously, the condition

$$\tau_\alpha\geqslant\tau>0 \tag{41.17}$$

is sufficient for (41.16) to hold. Thus, it follows that the iterative process (4.3) converges under the above conditions.

Analogous problems involving the convergence study of splitting schemes by use of the Fourier method have been considered in papers by SAMARSKII and ANDREEV [1964], SAMARSKII and NIKOLAEV ([1978], Chapters X–XI), SAMARSKII ([1977], Chapter X, Sections 3–4).

The problem of choosing parameters τ_α and τ of the relaxation process (41.3) will be considered separately in Chapter XIII.

42. The Fourier method and the grounding of the splitting schemes for nonstationary problems

Consider now nonstationary problems. In choosing a suitable splitting method for such problems we keep in mind, above all, the requirement of effective realization and absolute stability of the schemes.

The approximation problem is likewise important since it is tightly connected with the effectiveness of the numerical algorithm (see SAMARSKII [1977], p. 476). Indeed, if first-order schemes are used to approximate the solution of the nonstationary equation, small steps in time and space should be used in computations to obtain the necessary precision. However, often it is reasonable to apply schemes of second and higher orders. Large steps in time and space may be chosen in this case essentially saving computer time. Although it is clear that the use of second-order schemes leads sometimes to insurmountable difficulties, since difference schemes of second order turn out to be unstable in some cases.

In this section we will demonstrate three applications of the Fourier method for grounding known splitting schemes for the nonstationary problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0. \end{aligned} \tag{42.1}$$

(1) Divide domain Ω_t into subintervals $\tau = \Delta t$ and replace (42.1) with the difference equation:

$$\begin{aligned} &\prod_{\alpha=1}^{n} (E + \tfrac{1}{2}\tau\Lambda_\alpha)\frac{\varphi^{j+1} - \varphi^j}{\tau} + A\varphi^j = f^{j+1/2}, \\ &\varphi^0 = g. \end{aligned} \tag{42.2}$$

Here Λ_α are operators with $\Sigma_{\alpha=1}^{n}\Lambda_\alpha = A$. For the sake of simplicity consider A and Λ_α to be independent of τ.

Scheme (42.2) approximates problem (42.1) with second-order accurary in τ when the solution (41.1) is smooth enough. If operators Λ_α commute and generate a common basis of eigenfunctions, the eigenvalues $\{\lambda_k^{(\alpha)}\}$ of which are nonnegative, then scheme (42.2) is absolutely stable according to von Neumann.

Indeed, let us show that problem (42.1) is approximated by scheme (42.2). Consider the Taylor series

$$\begin{aligned} &\varphi^{j+1} = \varphi^{j+1/2} + \tfrac{1}{2}\tau\varphi_t^{j+1/2} + \tfrac{1}{8}\tau^2\varphi_{tt}^{j+1/2} + \cdots, \\ &\varphi^j = \varphi^{j+1/2} - \tfrac{1}{2}\tau\varphi_t^{j+1/2} + \tfrac{1}{8}\tau^2\varphi_{tt}^{j+1/2} + \cdots. \end{aligned} \tag{42.3}$$

Replacing (42.2) with (42.3), we obtain

$$\partial\varphi/\partial t + A\varphi = f + O(\tau^2) \quad \text{for } t = t_{j+1/2}. \tag{42.4}$$

Hence the difference scheme (42.2) approximates problem (42.1) with order $O(\tau^2)$.

To study its stability, reduce scheme (42.2) to the form:

$$\varphi^{j+1} = T\varphi^j + F^j, \tag{42.5}$$

where

$$T=(E-\tau B^{-1}A),$$

$$B=\prod_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau\Lambda_\alpha),$$

$$F^j=\tau B^{-1}f^{j+1/2}.$$

As follows from Section 40 the following condition is sufficient for the stability of scheme (42.5):

$$|\lambda_k(T)|\leqslant 1, \quad k=1, 2, \ldots, \tag{42.6}$$

where $\lambda_k(T)$ are the eigenvalues of operator T, which, as can easily be seen, under the above conditions coincide with the values q_k of (41.14) for $\tau_\alpha=\tau$. Since $\lambda_k^{(\alpha)}\geqslant 0$ and $\tau>0$, inequality (42.6) holds.

Hence, scheme (42.2) is stable.

The realization scheme of (42.2) coincides with (41.5).

(2) Consider now another scheme of the splitting method, introduced by DYAKONOV [1962f] for the solution of problem (42.1):

$$\sum_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau\Lambda_\alpha)\varphi^{j+1}=\prod_{\alpha=1}^{n}(E-\tfrac{1}{2}\tau\Lambda_\alpha)\varphi^j+\tau f^{j+1/2}. \tag{42.7}$$

The realization scheme of this algorithm is as follows:

$$\varphi^{j+\alpha/2n}=(E-\tfrac{1}{2}\tau\Lambda_{n-\alpha+1})\varphi^{j+(\alpha-1)/2n}, \quad \alpha=1,\ldots,n-1,$$

$$\varphi^{j+1/2}=(E-\tfrac{1}{2}\tau\Lambda_1)\varphi^{j+(n-1)/2n}+\tfrac{1}{2}\tau f^{j+1/2},$$

$$(E+\tfrac{1}{2}\tau\Lambda_1)\varphi^{j+1/n}=\varphi^{j+1/2}+\tfrac{1}{2}\tau f^{j+1/2},$$

$$(E+\tfrac{1}{2}\tau\Lambda_\alpha)\varphi^{j+\alpha/n}=\varphi^{j+(\alpha-1)/n}, \quad \alpha=2,\ldots,n.$$

The Fourier method may be also used to ground this splitting scheme.

Scheme (42.7) is a second-order approximation in τ of problem (42.1) on sufficiently smooth solutions. If operators Λ_α commute and generate a common basis of eigenfunctions, the eigenvalues $\{\lambda_k^{(\alpha)}\}$ of which are nonnegative, then scheme (42.7) is absolutely stable.

To prove this statement consider the Taylor series (42.3) and substitute them for (42.7). As a result we get (42.4). Therefore, scheme (42.7) approximates (42.1) with order $O(\tau^2)$.

To study the stability of (42.7) let us introduce the following notations:

$$B=\prod_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau\Lambda_\alpha), \qquad C=\prod_{\alpha=1}^{n}(E-\tfrac{1}{2}\tau\Lambda_\alpha).$$

Then (42.7) can be reduced to the form (42.1):

$$\varphi^{j+1}=T\varphi^j+F^j, \tag{42.8}$$

where

$$T=B^{-1}C, \qquad F^j=\tau B^{-1}f^{j+1/2}.$$

According to Section 40 the following condition is sufficient for the stability of scheme (42.8):

$$|\lambda_n(T)|\leqslant 1, \tag{42.9}$$

where $\lambda_n(T)$ are the eigenvalues of the transition operator T.

Because of the restrictions imposed on Λ_α it is easy to prove that

$$\lambda_n(T)=\prod_{\alpha=1}^{n}\frac{1-\tau\lambda_n^{(\alpha)}}{1+\tau\lambda_n^{(\alpha)}}. \tag{42.10}$$

Hence, condition (42.9) holds if $\lambda_n^{(\alpha)}\geqslant 0$, i.e., if the splitting scheme (42.7) is absolutely stable.

(3) Let us demonstrate the Fourier method's application for grounding the splitting scheme considered by MARCHUK [1968] for the solution of the non-stationary problem (42.1).

First of all find the solution of (42.1) in the interval $t_j\leqslant t\leqslant t_{j+1/2}$, where $t_{j+1/2}=t_j+\frac{1}{2}\tau$, by the splitting method

$$\begin{aligned}
&\frac{\varphi^{j+1/2n}-\varphi^j}{\frac{1}{2}\tau}+\Lambda_1\varphi^{j+1/2n}=f^{j+1/2},\\
&\frac{\varphi^{j+2/2n}-\varphi^{j+1/2n}}{\frac{1}{2}\tau}+\Lambda_2\varphi^{j+2/2n}=0,\\
&\quad\vdots\\
&\frac{\varphi^{j+1/2}-\varphi^{j+(n-1)/2n}}{\frac{1}{2}\tau}+\Lambda_n\varphi^{j+1/2}=0.
\end{aligned} \tag{42.11}$$

If solution (42.11) is obtained, we find the solution of (42.1) for $t=t_{j+1}$, using a difference scheme of second-order approximation

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\varphi^{j+1/2}=f^{j+1/2}. \tag{42.12}$$

The realization scheme of (42.11)–(42.12) has the form:

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau\Lambda_1)(E+\tfrac{1}{2}\tau\Lambda_2)\cdots(E+\tfrac{1}{2}\tau\Lambda_n)\varphi^{j+1/2}\\
&\quad=\varphi^j+\tfrac{1}{2}\tau f^{j+1/2},\\
&\varphi^{j+1}=\varphi^j+\tau(f^{j+1/2}-A\varphi^{j+1/2}).
\end{aligned} \tag{42.13}$$

Scheme (42.13) approximates (42.1) with second-order accuracy in τ when the solution of (42.1) is smooth enough. If the operators Λ_α commute and generate a common basis of eigenfunctions, the eigenvalues of which are nonnegative, then scheme (42.13) is absolutely stable.

To show this consider the first equation of (42.13). This is a first-order scheme for (42.1) in the interval $t_j \leqslant t \leqslant t_{j+1/2}$ for the class of smooth solutions.

Indeed,

$$(E+\tfrac{1}{2}\tau\Lambda_1)(E+\tfrac{1}{2}\tau\Lambda_2)\cdots(E+\tfrac{1}{2}\tau\Lambda_n)=E+\tfrac{1}{2}\tau A+\tfrac{1}{2}\tau^2 R, \tag{42.14}$$

where

$$R=(\Lambda_1\Lambda_2+\cdots)+\tfrac{1}{2}\tau(\Lambda_1\Lambda_2\Lambda_3+\cdots)+\cdots+(\tfrac{1}{2}\tau)^n\Lambda_1\Lambda_2\cdots\Lambda_n.$$

It follows from (42.13) and (42.14) that

$$\frac{\varphi^{j+1/2}-\varphi^j}{\frac{1}{2}\tau}+A\varphi^{j+1/2}=f^{j+1/2}+O(\tau). \tag{42.15}$$

Correcting the solution with the help of (42.12) we get:

$$\varphi^{j+1}=\varphi^j-\tau(A\varphi^{j+1/2}-f^{j+1/2}). \tag{42.16}$$

Substituting $\varphi^{j+1/2}$ of (42.15) into (42.16) we get the following equation:

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}+A\frac{\varphi^{j+1}+\varphi^j}{2}=f^{j+1/2}+O(\tau^2). \tag{42.17}$$

Hence, scheme (42.13) coincides with the well-known Crank–Nicolson scheme with the accuracy of $O(\tau^2)$. It thus follows that (42.13) approximates (42.1) with second-order accuracy in τ.

To study the stability we exclude the intermediate steps and express φ^{j+1} in φ. Denote

$$B=\prod_{\alpha=1}^{n}(E+\tfrac{1}{2}\tau\Lambda_\alpha).$$

Under the above conditions B is invertible and from (42.13) we obtain

$$\varphi^{j+1/2}=B^{-1}\varphi^j+\tfrac{1}{2}\tau B^{-1}f^{j+1/2}. \tag{42.18}$$

Substituting (42.18) into (42.16) we obtain:

$$\varphi^{j+1}=\Pi\varphi^j+F^j, \tag{42.19}$$

where

$$\Pi=(E-\tau AB^{-1}), \qquad F^j=\tau(E-\tfrac{1}{2}\tau AB^{-1})f^{j+1/2}.$$

It follows from Section 40 that the following condition is sufficient for the stability of scheme (42.19):

$$|\lambda_k(\Pi)|\leqslant 1, \quad k=1,2,\ldots, \tag{42.20}$$

where $\lambda_k(\Pi)$ are the eigenvalues of the operator Π.

When the operators Λ_α commute and generate a common basis of eigenfunctions, operator Π coincides with operator T of (41.15) and the eigenvalues coincide with the values q_k, defined in (41.14) for $\tau_\alpha=\tau$. Then, if $\lambda_k^{(\alpha)}\geqslant 0$, the inequality (42.20) holds and (42.13) is absolutely stable.

It is evident that the Fourier method plays an important role in grounding the splitting schemes. It should be noted, however, that the use of this method not only supposes the general requirements (for example, that eigenfunctions of the operator be complete) which have been considered in Section 40, but also some additional restrictions in each of the cases. For example, in Sections 41 and 42 it was required that the operators Λ_α are commutative and have a nonnegative spectrum. These and other restrictions make it difficult to use the Fourier method for studying the convergence of the splitting schemes.

In the following chapter, we will consider the a priori estimates method which allows, in some cases, to ground a whole class of splitting schemes under weaker restrictions.

CHAPTER XI

The A Priori Estimates Method and the Convergence Studies of the Splitting Schemes

In studying the convergence of the splitting schemes, it is important to know in what way the definition of a scheme's stability is chosen. The a priori estimates method is based on the stability of the scheme's solution in an energetical norm (see the Introduction). The idea of this method is to obtain an a priori estimate for the approximate solution, which helps to prove the scheme's stability and—if there exists an approximation—its convergence as well.

In this chapter we will consider the application of this method for grounding various splitting schemes.

43. The simplest a priori estimates

Let us solve the nonstationary problem:

$$\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t,$$

$$\varphi = g \quad \text{for } t = 0,$$

using one of the splitting schemes (see Part 1). We write such a scheme in the form:

$$\mathscr{L}(\varphi^\tau) = f^\tau, \tag{43.1}$$

where $\mathscr{L}$ is an operator acting in the grid space Φ_τ, $f^\tau \in \Phi_\tau$, and φ^τ is the solution to be found. (In the general case, the right-hand side f^τ in (43.1) may belong to another grid space F^τ.)

We assume that the approximation of (43.1) is established. Following the Introduction, we will regard the difference scheme (43.1) as stable if for any parameter τ characterizing the approximation the following inequality holds:

$$\|\varphi^\tau\|_{\Phi_\tau} \leqslant c\|f^\tau\|_{\Phi_\tau}, \tag{43.2}$$

where c is a positive constant independent of τ and $\|\cdot\|_{\Phi_\tau}$ denotes the norm in Φ_τ.

The definition of stability in this form is in close connection with the notion of correctness of problems with a continuous argument. It may be noted that the

stability establishes the continuous dependence of the solution on the initial data in the case of a problem, (43.1), with a discrete argument. Indeed, as follows from estimate (43.2) small variations of the input data f^τ lead to small variations of the solution φ^τ.

Thus, the definition of stability in the form (43.2) ties the solution itself down to a priori information about the input data of the problem. To analyze the stability of many of the splitting schemes this definition is more convenient and informative than Neumann's definition of stability.

First of all, we consider the simplest a priori estimates obtained on the basis of this definition and show their application for studying the convergence of various splitting schemes.

While solving nonstationary problems by the splitting method (fractional steps method) the difference scheme may often be reduced to the form (40.1):

$$\varphi^{j+1} = T\varphi^j + \tau F^j, \qquad \varphi^0 = g, \tag{43.3}$$

where T is the transition operator from one time layer to another acting in some Hilbert space Φ and g, $F^j \in \Phi$.

To analyze approximation and stability it is convenient to write the scheme in the form (43.3) (especially when approximation and even stability do not occur on each of the intermediate steps).

It was already noted in Section 40 that many splitting schemes are reduced to the form (43.3) by excluding fractional (intermediate) steps.

Assume that the approximation of the exact problem by scheme (43.3) has been established. Let us see what conditions should be imposed on the operator T to obtain the stability estimate (43.2).

The formal solution of (43.3) has the form

$$\varphi^j = T^j g + \tau \sum_{i=1}^{j} T^{j-1} F^{i-1}, \quad 0 \leqslant j\tau \leqslant \text{const}, \tag{43.4}$$

which implies that

$$\begin{aligned} \|\varphi^j\|_\Phi &\leqslant \|T\|_\Phi^j \|g\|_\Phi \\ &\quad + (1 + \|T\|_\Phi + \cdots + \|T\|_\Phi^{j-1}) \cdot \max_j \|F^j\|_\Phi. \end{aligned} \tag{43.5}$$

Assuming that

$$\|T\|_\Phi \leqslant 1 \tag{43.6}$$

we obtain an estimate in the form (43.2) which follows from (43.5):

$$\|\varphi^\tau\|_{\Phi_\tau} \leqslant \|g\|_\Phi + \text{const} \cdot \max_j \|F^j\|_\Phi, \tag{43.7}$$

where

$$\|\varphi^\tau\|_{\Phi_\tau} = \max_j \|\varphi^j\|_\Phi.$$

Hence the following statement is valid.

If (43.3) approximates the evolutionary problem

$$\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t,$$

$$\varphi = g \quad \text{for } t = 0$$

and $\|T\|_{\Phi} \leqslant 1$, then (43.3) is absolutely stable and the approximate solution φ^{τ} converges to the exact solution φ of the problem.

(This follows directly from the above considerations as well as from Convergence Theorem 3.1.)

Naturally, condition (43.6) is a sufficient stability condition. More sophisticated criteria may be found via the norm of powers of the transition operator, $\|T^i\|_{\Phi}$. But a weakened condition makes a constructive procedure of establishing the stability criteria more difficult. It is the sufficient condition (43.6) which is used, as a rule, in practical calculations.

Consider the case that the operator T is self-adjoint, i.e. $T^* = T$ and the problem $T\psi = \lambda\psi$ determines a complete system of eigenfunctions $\{\psi_n\}$ corresponding to eigenvalues $\{\lambda_n\}$. Therefore any element φ of Φ may be represented in a Fourier series.

In this case

$$\|T\|_{\Phi}^2 = \sup_{\varphi \neq 0} \frac{(T\varphi, T\varphi)}{(\varphi, \varphi)} = \max_n |\lambda_n(T)|^2$$

and the stability condition (43.6) converts into:

$$|\lambda_n(T)|^2 \leqslant 1, \quad n = 1, 2, \ldots. \tag{43.8}$$

Thus we get condition (40.9). This means that, in the case of the self-adjoint operator T, the sufficient stability conditions according to Neumann and those according to definition (43.2) coincide.

For example, let us formulate the convergence conditions for the homogeneous two-level splitting scheme:

$$\frac{\varphi^{j+k/p} - \varphi^{j+(k-1)/p}}{\tau} = \Lambda_{k,k-1}\varphi^{j+(k-1)/p} + \Lambda_{k,k}\varphi^{j+k/p}$$

$$k = 1, \ldots, p. \tag{43.9}$$

This scheme was applied by YANENKO ([1967], p. 149) for solving the evolutionary problem for $f = 0$. The operator A is assumed to be split as follows:

$$A = \sum_{k=1}^{p} A^{(k)},$$

$$A^{(k)} = \Lambda_{k,k-1} + \Lambda_{k,k}, \quad k = 1, \ldots, p, \tag{43.10}$$

where $\Lambda_{i,j}$ are operators acting in Φ.

Assume that the operators $\Lambda_{i,j}$ are commutative and that the following conditions

hold:

$$\tau\|\Lambda_{i,i}\|_{\Phi}<1, \quad i=1,\ldots,p,$$

$$\left\|\prod_{i=1}^{p}(E-\tau\Lambda_{i,i})^{-1}(E+\tau\Lambda_{i,i-1})\right\|_{\Phi}<1.$$

Then (43.9) approximates the problem

$$\partial\varphi/\partial t+A\varphi=0 \quad \text{in } \Omega_t,$$

$$\varphi=g \quad \text{for } t=0$$

with first-order accuracy in τ and is absolutely stable. The approximate solution of scheme (43.9) converges to the exact solution of the problem with the order $O(\tau)$.

To prove this we construct the corresponding scheme with integer steps. Equations (43.9) may be rewritten as follows:

$$A_k\varphi^{j+k/p}=B_k\varphi^{j+(k-1)/p}, \tag{43.11}$$

where

$$A_k=(E-\tau\Lambda_{k,k}), \qquad B_k=E+\tau\Lambda_{k,k-1},$$

$$k=1,\ldots,p.$$

Excluding the intermediate equations and using the fact that operators A_k and B_k are commutative we obtain the scheme with integer steps

$$A_0\varphi^{j+1}=B_0\varphi^{j}, \tag{43.12}$$

where

$$A_0=A_1\cdot A_2\cdot\cdots\cdot A_p, \qquad B_0=B_1\cdot B_2\cdot\cdots\cdot B_p.$$

Under conditions $\tau\|\Lambda_{i,i}\|_{\Phi}<1$, $i=1,\ldots,p$, operator A_0 is invertible and it follows from (43.12) that

$$\varphi^{j+1}=A_0^{-1}B_0\varphi^{j}. \tag{43.13}$$

Thus, by excluding fractional steps the splitting scheme (43.9) is reduced to (43.13) with integer steps and has the form (43.3).

Consider the scheme's approximation. Expanding operators A_0 and B_0 in powers of τ we obtain

$$\begin{aligned}A_0&=E-\tau(\Lambda_{1,1}+\Lambda_{2,2}+\cdots+\Lambda_{p,p})\\&\quad+\tau^2(\Lambda_{1,1}+\Lambda_{1,2}+\cdots+\Lambda_{p-1,p-1}\Lambda_{p,p})\\&\quad+\cdots+(-1)^p\tau^p\Lambda_{1,1}\Lambda_{2,2}\cdots\Lambda_{p-1,p-1}\Lambda_{p,p},\end{aligned} \tag{43.14}$$

$$\begin{aligned}B_0&=E+\tau(\Lambda_{1,0}+\Lambda_{2,1}+\cdots+\Lambda_{p,p-1})\\&\quad+\tau^2(\Lambda_{1,0}\Lambda_{2,1}+\cdots+\Lambda_{p-1,p-2}\Lambda_{p,p-1})+\cdots+\tau^p\Lambda_{1,0}\Lambda_{2,1}\cdots\Lambda_{p,p-1}.\end{aligned}$$

In view of (43.14), scheme (43.12) can be transformed into

$$\frac{\varphi^{j+1}-\varphi^j}{\tau}=\sum_{k=1}^{p}(\Lambda_{k,k-1}\varphi^j+\Lambda_{k,k}\varphi^{j+1})+\tau F, \tag{43.15}$$

where

$$F=\sum_{k<l}(\Lambda_{k,k-1}\Lambda_{l,l-1}\varphi^j-\Lambda_{k,k}\Lambda_{l,l}\varphi^{j+1})$$
$$+\tau\sum_{i,k,l}(\Lambda_{i,i-1}\Lambda_{k,k-1}\Lambda_{l,l-1}\varphi^j+\Lambda_{i,i}\Lambda_{k,k}\Lambda_{l,l}\varphi^{j+1})$$
$$+\cdots+\tau^{p-2}(\Lambda_{1,0}\Lambda_{2,1}\cdots\Lambda_{p,p-1}\varphi^j+(-1)^{p-1}\Lambda_{1,1}\Lambda_{2,2}\cdots\Lambda_{p,p}\varphi^{j+1}).$$

Substituting the Taylor series

$$\varphi^{j+1}=\varphi^j+\tau\varphi_t^j+\tfrac{1}{2}\tau\varphi_{tt}^j+\cdots$$

into (43.15) and taking (43.10) into account we obtain:

$$\partial\varphi/\partial t+A\varphi=O(\tau)\quad\text{for } t=t_j.$$

This means that difference scheme (43.9) approximates the problem

$$\partial\varphi/\partial t+A\varphi=0\quad\text{in } \Omega_t,$$

$$\varphi=g\quad\text{for } t=0$$

with the first-order accuracy in τ.

To study the stability of (43.9) we write it in the form (43.13). It follows from (43.6) that the sufficient stability condition for (43.9) will be

$$\|A_0^{-1}B_0\|_\Phi\leqslant 1, \tag{43.16}$$

which holds under the above assumptions.

From the study of approximation and stability the convergence of the two-layer fractional steps scheme (43.9) follows with the order $O(\tau)$.

In the general case of noncommutative operators $\Lambda_{i,j}$ convergence criteria for (43.9) are analogous, although the proof is more complicated. In order to prove the approximation we have to modify the exclusion method in the case of noncommutative operators. The idea of such a modification belongs to YANENKO [1964a] and was worked out in detail by YANENKO [1962] and BOYARINTSEV [1966]. YANENKO ([1967], p. 151) proved the convergence of the splitting method (43.9) in case of commutative operators $\Lambda_{i,j}$ for weak conditions on the operators A_0 and B_0.

44. A priori estimates for the splitting scheme of type $A_j\varphi^{j+1}=B_j\varphi^j+\tau_j f^j$

Consider a more general splitting scheme for the evolutionary problem

$$\partial\varphi/\partial t+A\varphi=f\quad\text{in } \Omega_t,$$

$$\varphi=g\quad\text{for } t=0,$$

which can be written in the form:

$$A_j\varphi^{j+1}=B_j\varphi^j+\tau_j f^j, \qquad \varphi^0=g, \tag{44.1}$$

where A_j and B_j are operators acting in Φ depending, in general, on the number j of the time step τ_j and $g, f^j \in \Phi$.

Excluding fractional steps (see, for example, (43.11)), we often come to schemes of type (44.1).

Assume that (44.1) approximates the evolutionary problem

$$\partial\varphi/\partial t+A\varphi=f \quad \text{in } \Omega_t,$$

$$\varphi=g \quad \text{for } t=0$$

with some order of accuracy in τ, operators A_j are invertible for each j and the following conditions are satisfied:

$$\|A_j^{-1}\|_\Phi \leqslant M, \qquad \|A_j^{-1}B_j\|_\Phi \leqslant 1+N\tau, \tag{44.2}$$

where M and N are positive constants and

$$\tau=\max_j \tau_j, \quad 0 \leqslant j\tau \leqslant T_0 < \infty.$$

Then (44.1) is absolutely stable, and the following a priori estimate holds for its solution φ^j:

$$\|\varphi^j\|_\Phi \leqslant e^{NT_0}\left(\|g\|_\Phi + M\sum_{i=1}^{j}\tau_{i-1}\|f^{i-1}\|_\Phi\right). \tag{44.3}$$

The approximate solution of scheme (44.1) converges to the exact solution of the problem.

Indeed, operators A_j being invertible, (44.1) may be written in the form:

$$\varphi^{j+1}=A_j^{-1}B_j\varphi^j+\tau_j A_j^{-1}f^j, \qquad \varphi^0=g. \tag{44.4}$$

Thus, under conditions (44.2) we obtain successively:

$$\|\varphi^1\|_\Phi \leqslant (1+N\tau)\|g\|_\Phi+\tau_0 M\|f^0\|_\Phi,$$

$$\begin{aligned}\|\varphi^2\|_\Phi &\leqslant (1+N\tau)\|g\|_\Phi+\tau_1 M\|f^1\|_\Phi \\ &\leqslant (1+N\tau)^2\|g\|_\Phi+\tau_0 M(1+N\tau)\|f^0\|_\Phi+M\tau_1\|f'\|_\Phi,\end{aligned}$$

etc., i.e.

$$\|\varphi^j\|_\Phi \leqslant (1+N\tau)^j\|g\|_\Phi+M\sum_{i=1}^{j}\tau_{i-1}(1+N\tau)^{j-i}\|f^{i-1}\|_\Phi.$$

Then for $j=1, 2, \ldots$

$$\|\varphi^j\|_\Phi \leqslant (1+N\tau)^j\left(\|g\|_\Phi+M\sum_{i=1}^{j}\tau_{i-1}\|f^{i-1}\|_\Phi\right).$$

Having $(1+N\tau)^j \leqslant e^{N\tau j} \leqslant e^{NT_0}$, we obtain

$$\|\varphi^j\|_\Phi \leqslant e^{NT_0}\left(\|g\|_\Phi + M\sum_{i=1}^{j}\tau_{i-1}\|f^{i-1}\|_\Phi\right).$$

Scheme (44.1) is, therefore, absolutely stable and the a priori estimate (44.3) is valid.

The scheme's convergence results from its approximation and stability.

It follows that the homogeneous two-layer scheme (43.12) converges under the conditions that

$$\|A_0\|_\Phi \leqslant M, \qquad \|A_0^{-1}B\|_\Phi \leqslant 1+N\tau. \tag{44.5}$$

If in (44.1)

$$A_j = E, \qquad B_j = T, \qquad f^j = F^j, \qquad \tau_j = \tau,$$

we obtain the classical scheme (43.3).

If approximation conditions are satisfied, (43.3) converges when

$$\|T\|_\Phi \leqslant 1+N\tau, \quad N=\text{const}>0. \tag{44.6}$$

This follows from the above considerations.

At the same time the a priori estimate (44.3) for $M=1$, $\tau_{i-1}=\tau$ holds.

The condition

$$\|T^j\|_\Phi \leqslant c, \quad c=\text{const}>0 \tag{44.7}$$

is more general than (44.6).

Replacing (44.6) with (44.7) we easily get the a priori estimate:

$$\|\varphi^j\|_\Phi \leqslant c\left(\|g\|_\Phi + \sum_{i=1}^{j}\tau\|f^{i-1}\|_\Phi\right). \tag{44.8}$$

Note that in the case that self-adjoint operators A_j and $T_j = A_j^{-1}B_j$ have a complete basis of eigenfunctions, the convergence conditions (42.2) are equivalent to the following conditions:

$$\begin{aligned} &\max_n |\lambda_n(A_j^{-1})| \leqslant M, \\ &\max_n |\lambda_n(T_j)| \leqslant 1+N\tau, \end{aligned} \tag{44.9}$$

which are called Neumann's stability conditions (see DYAKONOV [1971d], RICHTMYER and MORTON [1972]). Here $\lambda_n(A)$ denotes, as earlier, the eigenvalue of the operator A.

Estimates of types (43.7), (44.3) and (44.8) are the simplest a priori estimates. They have been constructed using information about the norms of the problem's operators.

Using a priori estimates constructed in the same way, the splitting schemes have been grounded by many authors, for example by DYAKONOV [1962c, 1962f, 1971d],

SAMARSKII [1977] and MARCHUK [1973, 1980]. In particular, the a priori estimates method has been used for grounding the two-cycle componentwise splitting schemes considered in Chapter III.

More sophisticated estimates may be obtained using the energy inequalities method which is considered in the next section.

45. The energy inequalities method for constructing a priori estimates

The energy inequalities method is one of the general and rather effective methods for obtaining a priori estimates. Its essence consists of the following.

Consider scheme (43.1) in a Hilbert space Φ_τ. The inner product of (43.1) with φ^τ in Φ_τ is

$$(\mathscr{L}(\varphi^\tau), \varphi^\tau)_{\Phi_\tau} = (f^\tau, \varphi^\tau)_{\Phi_\tau}. \tag{45.1}$$

Equality (45.1) is called an energy identity (see, for instance, SAMARSKII [1977], p. 359). For linear problems the left-hand side of (45.1) is a quadratic form of φ^τ, and the right-hand side is a bilinear form with respect to f^τ and φ^τ. After some transformations one obtains lower estimates of $(\mathscr{L}(\varphi^\tau), \varphi^\tau)_{\Phi_\tau}$, upper estimates of $(f^\tau, \varphi^\tau)_{\Phi_\tau}$ and, then, a priori estimates of type (43.2). In these transformations the Green difference formulae, formulae of summation by parts, and grid analogues of imbedding theorems are often used. Sometimes one obtains systems of difference inequalities. Estimating their solutions, one obtains (43.2). Estimates of the type (43.2), obtained from (45.1), are called energy inequalities.

Energy inequalities for the simplest difference schemes were first obtained in the well-known article by COURANT, FRIEDRICHS and LEWY [1928]. For a broad class of problems the estimates of this kind were studied by LADYZHENSKAYA [1956, 1968], LEES [1960a–b], SAMARSKII [1961b, 1977], DYAKONOV [1962e, 1971a, 1971d], etc.

Such a method for obtaining energy inequalities has been applied for grounding the splitting and alternating direction methods. Some of the first papers in this direction were by LEES [1960c, 1961], where the energy inequalities method has been applied for grounding the simplest splitting schemes. This approach has been further developed by DYAKONOV [1962c, 1962e] and SAMARSKII [1962c–d, 1963b]. In their papers the splitting schemes have been grounded with the help of a priori estimates for various problems of mathematical physics.

Let us shortly summarize the results obtained in these papers.

LEES [1961] considered the implicit alternating direction methods of Douglas, Peaceman and Rachford for solving the first boundary value problem for the parabolic equation

$$\frac{\partial \varphi}{\partial t} = \sum_{i=1}^{N} \frac{\partial^2 \varphi}{\partial x_i^2}, \qquad x \in \Omega, \quad t > 0 \tag{45.2}$$

with initial condition

$$\varphi(x, 0) = \varphi_0(x), \tag{45.3}$$

The exact solution $\varphi(x, t)$ is assumed to be a sufficiently smooth function, its support for a fixed t lies in the domain Ω. Excluding intermediate steps from the splitting schemes Lees obtains equalities which lead to energy identities of the type (45.1). On the basis of difference Green formulae and grid analogues of the imbedding theorems he comes to the stability estimate for approximate solution φ^h:

$$\| \varphi^h(t) \|_{\Phi_h} \leqslant c \| \varphi^h(0) \|_{\Phi_h}, \tag{45.4}$$

where $\| \cdot \|_{\Phi_h}$ is the usual vector norm in the grid space Φ_h, c is a positive constant depending on N and on the grid coefficient

$$\lambda = \tau \left(\sum_{i=1}^{N} h_i^{-2} \right);$$

here τ and h_i are steps in time and in space respectively.

Besides the estimate (45.4) ("the initial data" stability) LEES [1961] obtained estimates for the convergence speed of the splitting schemes.

As mentioned by LEES [1961] the energy inequalities method allows to ground implicit alternating direction methods for the parabolic equation

$$\frac{\partial \varphi}{\partial t} = \sum_{i=1}^{N} \frac{\partial}{\partial x} \left(a_i(x, t) \frac{\partial \varphi}{\partial x_i} \right), \qquad x \in \Omega, \quad t > 0$$

with variable coefficients $a_i(x, t)$.

In another paper, LEES [1960c] considered two difference methods for solving the mixed problem for the simplest two-dimensional wave equation

$$\frac{\partial^2 \varphi}{\partial t^2} = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} \quad \text{in } \Omega \times [0, T]. \tag{45.5}$$

These methods were obtained by the author from the simplest approximations of problem (45.5) using DOUGLAS and RACHFORD's [1956] splitting procedure. In the same way the energy inequalities method was used for grounding these methods. As mentioned by the author, the results may be generalized for hyperbolic equations of the type:

$$\frac{\partial^2 \varphi}{\partial t^2} = a(x, y, t) \frac{\partial^2 \varphi}{\partial x^2} + b(x, y, t) \frac{\partial^2 \varphi}{\partial y^2} + F\left(x, y, t, \varphi, \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial t} \right).$$

It was shown by DYAKONOV [1962c] that a somewhat modified YANENKO [1959a, 1960] method of fractional steps is—under zero boundary conditions—an absolutely stable and convergent difference scheme for equations of the type:

$$\partial \varphi / \partial t = L\varphi + f, \tag{45.6}$$

where L is a self-adjoint elliptical negatively defined operator with separable variables of order $2m$. Here for $m > 1$ the method's convergence is understood as the convergence to a generalized solution. Difference energy inequalities are the main tool for the proof. The analogous result for the classical solution of problem (45.6)

for $m=1$ was obtained by DYAKONOV [1962e] where a priori estimates like those of LEES [1960c, 1961] were constructed.

SAMARSKII [1962c] considered the alternating direction method (see Section 12) for solving linear and quasi-linear equations of a parabolic type:

$$\partial\varphi/\partial t = L\varphi \equiv \sum_{\alpha=1}^{p} L_\alpha \varphi. \tag{45.7}$$

This method uses any number p of space variables and is applicable for an arbitrary domain G.

Its essence consists of the following. In each layer

$$t_{j+(\alpha-1)/p} \leqslant t \leqslant t_{j+\alpha/p} = t_j + \tau_\alpha/p, \quad \alpha = 1, \ldots, p,$$

a one-dimensional difference equation is solved:

$$\frac{1}{p}\frac{\partial\varphi}{\partial t} - L_\alpha\varphi = 0.$$

To this end homogeneous difference schemes are used:

$$\frac{\varphi^{j+\alpha/p} - \varphi^{j+(\alpha-1)/p}}{\tau} - \Lambda_\alpha \varphi^{j+\alpha/p} = 0. \tag{45.8}$$

These have been studied by TIKHONOV and SAMARSKII [1961], SAMARSKII [1961a, 1962a–b] and SAMARSKII and FRYAZINOV [1961].

In the paper by SAMARSKII [1962c] the uniform stability of the method was proved with respect to the right-hand side, boundary and initial data. It was shown that locally one-dimensional schemes and multidimensional implicit difference schemes (see SAMARSKII [1962b]) have accuracy $O(h^2+\tau)$. In establishing the a priori estimate of stability the author has used the maximum principle for difference parabolic equations.

SAMARSKII [1962d] grounded the local one-dimensional alternating direction method for (45.7) in the case where on each stage (unlike SAMARSKII [1962c]) six-point one-dimensional schemes were considered. The method's convergence is proved for grids with arbitrary steps. Using the energy ("integral") inequalities method and a special summation method for local errors, the a priori estimate for an approximate solution is obtained.

In another paper by SAMARSKII [1963b] two-layer schemes were considered with accuracy of order $O(h^4+\tau^2)$ for the multidimensional heat conduction equation (45.7) applicable for $p \leqslant 3$. For the realization a number of splitting algorithms requiring the same number of operations as the corresponding algorithms of $O(\tau^2+h^4)$ were used (see DOUGLAS and RACHFORD [1956], DYAKONOV [1962e], YANENKO [1959a, 1961]). It was shown that these schemes are absolutely stable and converge in the mean with the rate $O(h^4+\tau^2)$ for any value $\gamma=\tau/h^4$. For stability studies a priori estimates obtained from an energy identity on the basis of Green difference formulae and grid analogues of the imbedding theorems were constructed.

The use of the energy inequalities for grounding the splitting schemes for stationary problems was dealt with in papers by SAMARSKII and ANDREEV [1964], SAMARSKII ([1977], Chapter X), etc.

As we see from these papers the energy inequalities method is widely used for grounding the splitting and alternating direction schemes for various problems of mathematical physics.

To conclude this section let us demonstrate the use of this method for grounding the splitting schemes in case of an implicit difference approximation (see MARCHUK [1980]). To this end consider the problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0. \end{aligned} \tag{45.9}$$

Assume that

$$A = \sum_{\alpha=1}^{n} A_\alpha,$$

where all operators A_α are time-independent and positive-semidefinite in Hilbert space Φ ($A_\alpha \geqslant 0$), i.e. $(A_\alpha\varphi, \varphi)_\Phi \geqslant 0$ for all $\varphi \in D(A_\alpha)$. Consider the splitting algorithm in the form

$$\begin{aligned} &\frac{\varphi^{j+1/2} - \varphi^j}{\tau} + A_1\varphi^{j+1/n} = 0, \\ &\quad\vdots \\ &\frac{\varphi^{j+1} - \varphi^{j+(n-1)/n}}{\tau} + A_n\varphi^{j+1} = f^j. \end{aligned} \tag{45.10}$$

With the help of the energy inequalities method let us prove the following statement.

Scheme (45.10) approximates problem (45.9) with first-order accuracy in τ on sufficiently smooth solutions and is absolutely stable under the condition that $A_\alpha \geqslant 0$. The approximate solution of (45.10) converges to the exact solution of (45.9).

In order to prove this we consider the following. The approximation of scheme (45.10) is obvious; it is proved by excluding fractional steps and using a Taylor series for φ^{j+1}.

Using the energy inequalities method we prove the stability of (45.10). Take the inner product of each of the equations (45.10) with $\varphi^{j+1/n}, \ldots, \varphi^{j+1}$, respectively. Thus using the positive-semidefiniteness of operators A_α, we obtain

$$\|\varphi^{j+\alpha/n}\|_\Phi \leqslant \|\varphi^{j+(\alpha-1)/n}\|_\Phi, \quad \alpha = 1, 2, \ldots, n-1. \tag{45.11}$$

Consider the last equation obtained from (45.10) in more detail. We have

$$\begin{aligned} (\varphi^{j+1}, \varphi^{j+1})_\Phi = &(\varphi^{j+(n-1)/n}, \varphi^{j+1})_\Phi \\ &- \tau(A_n\varphi^{j+1}, \varphi^{j+1})_\Phi + \tau(f^j, \varphi^{j+1})_\Phi. \end{aligned}$$

Taking into account that $A_n \geqslant 0$ we obtain

$$(\varphi^{j+1}, \varphi^{j+1})_\Phi \leqslant (\varphi^{j+(n-1)/n}, \varphi^{j+1})_\Phi + \tau(f^j, \varphi^{j+1})_\Phi.$$

Since

$$|(\varphi^{j+(n-1)/n}, \varphi^{j+1})_\Phi| \leqslant \|\varphi^{j+(n-1)/n}\|_\Phi \|\varphi^{j+1}\|_\Phi,$$

$$|(f^j, \varphi^{j+1})_\Phi| \leqslant \|f^j\|_\Phi \|\varphi^{j+1}\|_\Phi,$$

we have

$$\|\varphi^{j+1}\|_\Phi^2 \leqslant \|\varphi^{j+(n-1)/2}\|_\Phi \|\varphi^{j+1}\|_\Phi + \tau \|f^j\|_\Phi \|\varphi^{j+1}\|_\Phi.$$

Dividing by $\|\varphi^{j+1}\|_\Phi$ we get the following inequality:

$$\|\varphi^{j+1}\|_\Phi \leqslant \|\varphi^{j+(n-1)/n}\|_\Phi + \tau \|f^j\|_\Phi.$$

Excluding the solution with a fractional index with the help of (45.11) we obtain

$$\|\varphi^{j+1}\|_\Phi \leqslant \|\varphi^j\|_\Phi + \tau \|f^j\|_\Phi. \tag{45.12}$$

Taking into account that $\|\varphi^0\|_\Phi = \|g\|_\Phi$ and excluding intermediate values of the solution we obtain the a priori estimate:

$$\|\varphi^{j+1}\|_\Phi \leqslant \|g\|_\Phi + \tau j \|f\|_{\Phi_\tau},$$

where

$$\|f\|_{\Phi_\tau} = \max_j \|f^j\|_\Phi.$$

It thus follows that the difference scheme (45.10) is absolutely stable.

The convergence follows from the approximation and stability. Thus, the statement is proved.

We have come to the conclusion of this chapter that the a priori estimates method is an effective technique for the convergence studies of splitting and alternating direction schemes.

CHAPTER XII

The Splitting of the Evolutionary Problem for a System of Differential Equations

As noted above, for a system of difference equations with properly chosen operators A_α the initial evolutionary problem is preliminarily split in a number of splitting algorithms. Then, each equation of the system is approximated by a suitable difference scheme and additional splitting of the A_α in $A_{\alpha\beta}$ is possible. In this chapter, we consider some of these splitting methods and their grounding for a system of difference equations.

Note that the operator A in the initial problem is not necessarily a matrix. It may be some abstract operator acting in a Banach space Φ. Let us agree to write the partial derivative in t as $\partial\varphi/\partial t$, in all equations under consideration, though, in some occasions it would be more correct to use the conventional notation of the derivative with respect to t.

46. The splitting of problems defined on fractional intervals and the weak approximation method

Assume that in a Banach space Φ the Cauchy abstract problem is considered,

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega_t = \{0 < t < T\}, \\ &\varphi = g \qquad \text{for } t = 0, \quad g \in \Phi, \end{aligned} \tag{46.1}$$

where $A = A(t)$ is a linear operator that has a domain of definition dense in Φ and a range in Φ. The functions $\varphi(t)$ and $f(t)$ are abstract functions of $t \in [0, T]$ with values in Φ.

Let operator A be representable as the sum

$$A = \sum_{\alpha=1}^{n} A_\alpha \tag{46.2}$$

of linear operators $A_\alpha(t)$ with the same domain as A. Then solving problem (46.1) may be approximately reduced to successively solving Cauchy problems of type (46.1), where instead of A we have operators A_α, $\alpha = 1, 2, \ldots, n$. Let us consider the

first of those reduction methods (SAMARSKII [1962c], YANENKO [1964b], YANENKO and DEMIDOV [1966]).

Introduce the grid $t_j = j \times \tau, j = 0, 1, \ldots, J$ with the step τ. Represent f in the form

$$f = \sum_{\alpha=1}^{n} f_\alpha \tag{46.3}$$

and rewrite (46.1) in the form

$$\sum_{\alpha=1}^{n} \left(\frac{1}{n} \frac{\partial \varphi}{\partial t} + A_\alpha \varphi - f_\alpha \right) = 0. \tag{46.4}$$

Introduce the fractional values $t_j + \alpha/n, \alpha = 1, \ldots, n-1$, on the interval $\theta_j = [t_j, t_{j+1}]$ and consider the following system of equations on fractional intervals:

$$\begin{aligned} &\frac{1}{n} \frac{\partial \varphi_\alpha}{\partial t} + A_\alpha \varphi_\alpha = f_\alpha, \quad t \in [t_{j+(\alpha-1)/n}, t_{j+\alpha/n}], \\ &\alpha = 1, \ldots, n \end{aligned} \tag{46.5}$$

with conditions

$$\begin{aligned} &\varphi_1(0) = g, \\ &\varphi_\alpha(t_j) = \varphi_n(t_j), \qquad j = 1, 2, \ldots, \\ &\varphi_\alpha(t_{j+(\alpha-1)/n}) = \varphi_{\alpha-1}(t_{j+(\alpha-1)/n}), \quad j = 0, 1, \ldots, \\ &\alpha = 2, 3, \ldots, n. \end{aligned} \tag{46.6}$$

As a solution of problem (46.5)–(46.6) we will consider the element

$$\varphi^{j+1} = \varphi_n(t_{j+1}). \tag{46.7}$$

The next stage in the solution of the initial problem consists in solving numerically problem (46.5)–(46.6). To that end, the corresponding difference schemes in t (and possibly the additional splitting of A_α in $A_{\alpha\beta}$, $\alpha = 1, 2, \ldots, n$) can be used. If the structure of the operators A_α is simple enough, we obtain an effectively realizable splitting scheme for problem (46.1).

Let us formulate some results to ground the transition from (46.1) to (46.5)–(46.6). To that end, we introduce concepts connected with *the weak approximation method* (YANENKO and DEMIDOV [1966], YANENKO [1964b]). We will say that the family of functions $\Psi_\tau(t)$ weakly approximates the function $\Psi(t)$ in t on the interval $(0, T)$, if

$$\int_{t'}^{t''} (\Psi_\tau(\xi) - \Psi(\xi)) \, d\xi = \delta(t', t'', \tau), \tag{46.8}$$

where $\|\delta\|_\Phi \to 0$ for $\tau \to 0$ and $t', t'' \in (0, T)$. Further, we say that the family of linear differential operators $A_\tau(t)$ weakly approximates the operator $A(t)$ in t, if the weak approximation for coefficients of $A(t)$ takes place. Now, taking (46.2) into account we

introduce the operator

$$A_\tau = \sum_{\alpha=1}^{n} \beta_\alpha(\tau, t) \cdot A_\alpha, \tag{46.9}$$

where the functions $\beta_\alpha(\tau, t)$ are defined in the following way:

$$\beta_\alpha(\tau, t) = n \cdot \delta_{\alpha i}, \quad \text{if } t \in \left(t_j + \frac{\alpha - 1}{n}\tau, t_j + \frac{\alpha\tau}{n} \right).$$

(Here $\delta_{\alpha i}$ is the Kronecker delta.) It is easy to notice that operator A_τ weakly approximates the operator A. Together with (46.1) we consider the following Cauchy problem:

$$\begin{aligned} &\partial\varphi/\partial t + A_\tau \varphi = f_\tau \quad \text{in } \Omega_t, \\ &\varphi = g \quad \text{for } t = 0. \end{aligned} \tag{46.10}$$

Its solution will be defined by $\varphi \equiv \varphi_\tau$. It is easy to see that for $f_\alpha = f_\tau/n$ this problem may also be written in the form (46.5)–(46.6). Assuming $f_\tau = f$ and $f_\alpha = f/n$, then (46.3) and the conditions for weak approximation f by f_τ hold and the problems (46.5)–(46.6) and (46.10) coincide. Thus, problem (46.10) constitutes a splitting system of equations and to ground the transition from (46.1) to (46.5)–(46.6), it is sufficient to study the convergence of the problem's solution φ_τ to $\varphi(t)$ fo r $\tau \to 0$.

Now assume that operator A has a form

$$\mathrm{A} = \sum_{k_1, \ldots, k_m} a_{k_1 \ldots k_m}(x_1, \ldots x_m, t) \frac{\partial^{k_1 + \cdots + k_m}}{\partial x_1^{k_1} \cdots \partial x_m^{k_m}}, \tag{46.11}$$

where $x_1, \ldots, x_m$ are real variables on which the functions in Φ depend. All coefficients $a_{k_1 \ldots k_m}$ are assumed to be real, limited and continuous on t in a uniform metric. The order of the differential operators A_α in representation (46.2) is considered to be finite. Further, if all the derivatives of the function $a(x_1, \ldots, x_m, t)$ exist in A_α and are limited and continuous in a uniform metric, we then say that $a(x_1, \ldots, x_m, t)$ has derivatives up to order "A". When this differentiation procedure can be repeated j times we say that $a(x_1, \ldots x_m, t)$ has derivatives up to order "A^j". The expressions "up to order A" and "up to order A^1" are equivalent.

Assume that $a(x_1, \ldots, x_m, t)$ and f have derivatives in variables $x_1, \ldots, x_m$ up to order "A^j". We formally differentiate (46.1) for vector $\varphi^{(1)}$ with the components

$$v^{(k)} = v^{(k_1, \ldots, k_m)} = \mathrm{D}_1^{k_1} \cdots \mathrm{D}_m^{k_m} u, \quad \mathrm{D}_i^{k_i} \equiv \partial^{k_i}/\partial x_i^{k_i},$$

where the combinations $k = (k_1, \ldots, k_m)$ are taken from the sum (46.11) (necessarily including the combination $(0, \ldots, 0)$, corresponding to $\varphi = v^{(0, \ldots, 0)}$). Then, we get the system

$$\partial\varphi^{(1)}/\partial t + A^{(1)}\varphi^{(1)} = f^{(1)}. \tag{46.12}$$

This system is the first extended system with operator $A^{(1)}$ and vector $f^{(1)}$. In the same way the vectors $\varphi^{(2)} = \{v^{(k+l)}\}$, $\varphi^{(3)} = \{v^{(k+l+q)}\}, \ldots$ are constructed, where the indices $l, q, \ldots$ range over the same combinations as k. For $\varphi^{(2)}, \varphi^{(3)}, \ldots$ we obtain the

second, third,... extended systems, respectively:

$$\partial\varphi^{(j)}/\partial t + A^{(j)}\varphi^{(j)} = f^{(j)}, \quad j=2, 3, \ldots. \tag{46.13}$$

It is easy to verify that presentations

$$A^{(j)} = \sum_{\alpha=1}^{n} A_\alpha^{(j)} \quad \text{and} \quad A_\tau^{(j)} = \sum_{\alpha=1}^{n} \beta_\alpha A_\alpha^{(j)}$$

correspond to the presentation (46.2). Therefore, it is possible to consider the extended systems

$$\frac{\partial\varphi_\tau^{(j)}}{\partial t} + A_\tau^{(j)}\varphi_\tau^{(j)} = f_\tau^{(j)}. \tag{46.14}$$

along with (46.10).

Let us further agree to denote the Cauchy problems (46.1) and (46.10) and their extensions by $I, I_\tau, I^{(j)}, I_\tau^{(j)}$ respectively. The problems I and I_τ are defined in the same space Φ and the problems $I^{(j)}$ and $I_\tau^{(j)}$ in the corresponding Banach space Φ_j.

If φ is the solution of the problem

$$\partial\varphi/\partial t + A\varphi = 0 \quad \text{in } \Omega_t,$$

$$\varphi = g \quad \text{for } t=0$$

and the initial condition is $\varphi(t_1)$, then the transition operator from $\varphi(t_1)$ to $\varphi(t_2)$ will be denoted by $S(t_2, t_1)$, i.e.

$$\varphi(t_2) = S(t_2, t_1)\cdot\varphi(t_1), \quad 0 \leqslant t_1 \leqslant t_2 \leqslant T.$$

System (46.1) is called *correct*, if

$$\|S(t_2, t_1)\|_\Phi \leqslant M(T), \quad 0 \leqslant t_1 \leqslant t_2 \leqslant T$$

and *uniformly correct*, if

$$\|S(t_2, t_1)\|_\Phi \leqslant e^{c(t_2 - t_1)}$$

with constant C only dependent on T. For problems I_τ, $I^{(j)}$ and $I_\tau^{(j)}$ the notions, transition operator, correct and uniformly correct, are defined in a similar way. YANENKO and DEMIDOV [1966] and YANENKO [1964b] formulated the following results. In Statements 46.1 and 46.2 $f \equiv 0$ is assumed.

STATEMENT 46.1. *If the problems* I_τ, $I_\tau^{(1)}$ and $I_\tau^{(2)}$ *are uniformly correct, then* $\varphi_\tau(t)$ *converges uniformly in* t *to* $\varphi(t) = S(t, 0)g$ *for* $\tau \to 0$. *The transition operator* $S(t_2, t_1)$ *satisfies the conditions of uniform correctness for system* (46.1).

STATEMENT 46.2. *If the problems* I_τ and $I_\tau^{(j)}$, $j=1, 2, 3$ *are uniformly correct, then for any* $g \in \Phi_1$ *function* $\varphi_\tau(t)$ *converges uniformly in* t *to* $\varphi(t)$ *for* $\tau \to 0$ *along with its* x_k*-derivatives up to order "*A*". The limiting function* $\varphi(t)$ *has a derivative in* t *and is the solution of problem* (46.1).

STATEMENT 46.3. *If* (a) $g \in \Phi_1$, (b) $A_\alpha \varphi$ *are uniformly continuous in* t, (c) *problems* I_τ *and* $I_\tau^{(1)}$ *are uniformly correct, and* (d) *problems* I_τ *are correct on the right-hand side* (*i.e. their solutions are uniformly dependent on the right-hand sides*), *then* (46.1) *has a unique solution and* $\varphi_\tau(t)$ *converges uniformly to* $\varphi(t)$ *in* t.

STATEMENT 46.4. *Let* $\Phi = L_q(\Omega)$, $q > 0$, $\Omega = \Omega(x_1, \ldots, x_m)$. *If* (a) *problems* I_τ *and* $I_\tau^{(1)}$ *are uniformly correct for* $f \equiv 0$, *and* (b) *problems* I_τ *are correct on the right-hand side, then* $\varphi_\tau(t)$ *is fundamentally uniform in* t *for* $\tau \to 0$, *and any smooth limiting function* $\varphi(t)$ *is a solution of the problem* I. *If* $g^{(1)} = \varphi^{(1)}|_{t=0} \in L_q(\Omega)$, *then* $\varphi(t)$ *is a smooth and unique solution of problem* I.

REMARK 46.1. Here, the function $\varphi(t)$ is called a smooth function if all its derivatives belonging to $L_q(\Omega \times \Omega_t)$ which are present in equation (46.1) exist.

STATEMENT 46.5. *Let* A_α *have the form*

$$A_\alpha \equiv A_\alpha(x, t) \frac{\partial}{\partial x_\alpha}, \qquad \alpha = 1, \ldots, m, \quad x = (x_1, \ldots, x_m),$$

where the $A_\alpha(x, t)$ *are symmetric matrices, continuous in* $\Omega \times \Omega_t$ *along with the first derivatives in space variables. Then the problems* I *and* I_τ *are uniformly correct for* $f \equiv 0$. *The function* $\varphi_\tau(t)$ *converges uniformly in* t *for* $\tau \to 0$ *to the solution of problem* I.

The proofs of Statements 46.1–46.5 are given in the work by YANENKO and DEMIDOV [1966]. Different aspects connected with grounding the transition from (46.1) and (46.10) (and (46.5)–(46.6) in particular) can be found in the papers by YANENKO [1987] and SAMARSKII [1970]. Note that SAMARSKII [1970] estimated the convergence rate of φ_τ as well.

47. The splitting of problems defined on the whole interval

We now consider a splitting method for problem (46.1) that is often used in practice for obtaining splitting schemes (SAMARSKII [1965b, 1970], BAKER and OLIPHANT [1960], MARCHUK [1967]).

We replace problem (46.1) with the following system:

$$\begin{aligned}
&\partial\varphi_1/\partial t + A_1\varphi_1 = f_1(t), \quad \varphi_1(t_j) = v(t_j),\\
&\partial\varphi_2/\partial t + A_2\varphi_2 = f_2(t), \quad \varphi_2(t_j) = \varphi_1(t_{j+1}),\\
&\quad\vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (47.1)\\
&\partial\varphi_n/\partial t + A_n\varphi_n = f_n(t), \quad \varphi_n(t_j) = \varphi_{n-1}(t_{j+1}),\\
&\quad t \in \theta_j = [t_j, t_{j+1}]
\end{aligned}$$

and assume

$$v(t_{j+1}) \equiv \varphi_n(t_{j+1}), \quad j=0, 1, \ldots, J-1,$$
$$v(0)=g. \tag{47.2}$$

We formulate the statements with regard to the proximity of problems (46.1) and (47.1) (SAMARSKII [1970, 1971]).

If $A_\alpha(t')$ and $A_\beta(t'')$, $\alpha \neq \beta$, $\alpha, \beta = 1, \ldots, n$, for any $t', t'' \in [0, T]$ commute, i.e.

$$A_\alpha(t') \cdot A_\beta(t'') = A_\beta(t'') \cdot A_\alpha(t'),$$

then for $f=0$ the following equalities hold

$$v(t_j) = \varphi(t_j), \quad j=0, 1, \ldots, J. \tag{47.3}$$

If $f \neq 0$, then we can choose f_n with $f_1 = f_2 = \cdots = f_{n-1} = 0$ so that (47.3) holds.

If A_α and A_β commute and the exact solution of problem (46.1) satisfies the restriction that

$$\|A_\alpha A_\beta \varphi\| \leqslant M = \text{const} < \infty,$$

then under conditions (46.3) we have:

$$\|v(t_j) - \varphi(t_j)\| \leqslant O(\tau), \quad j=0, 1, \ldots, J. \tag{47.4}$$

Problem (47.1) approximates (46.1) in the *summarized* sense, i.e.

$$\left\| \sum_{\alpha=1}^{n} \psi_\alpha \right\| = O(\tau),$$

where

$$\psi_\alpha = f_\alpha(t) - A_\alpha(t)\varphi(t_{j+1}), \quad \alpha > 1,$$
$$\psi_1 = f_1(t) - A_1(t)\varphi(t) - d\varphi/dt$$

(see SAMARSKII [1970], p. 219).

48. The two-cycle splitting of the problem

Problem (46.1) can be split into a system of Cauchy problems that approximates problem (46.1) with order $O(\tau^2)$, using the *two-cycle method*, i.e. symmetrizing the sequence of problems of type (47.1) (BAKER and OLIPHANT [1960], MARCHUK [1967, 1971], FRYAZINOV [1968]). Let us demonstrate several examples of systems of Cauchy problems obtained by this method, each of these systems having second-order accuracy (with the corresponding additional restrictions on the smoothness of the initial data, of the solution and the appropriate choices of f_α in (46.3)).

EXAMPLE 48.1 (FRYAZINOV [1968]). The system has the form:

$$\begin{aligned}
&\partial\varphi_1/\partial t+\tfrac{1}{2}A_1\varphi_1=f_1, && \varphi_1(t_j)=v(t_j),\\
&\partial\varphi_2/\partial t+\tfrac{1}{2}A_2\varphi_2=f_2, && \varphi_2(t_j)=\varphi_1(t_{j+1}),\\
&\quad\vdots\\
&\partial\varphi_n/\partial t+\tfrac{1}{2}A_n\varphi_n=f_n, && \varphi_n(t_j)=\varphi_{n-1}(t_{j+1}),\\
&\partial\varphi_{n+1}/\partial t+\tfrac{1}{2}A_n\varphi_{n+1}=f_{n+1}, && \varphi_{n+1}(t_j)=\varphi_n(t_{j+1}),\\
&\quad\vdots\\
&\partial\varphi_{2n}/\partial t+\tfrac{1}{2}A_1\varphi_{2n}=f_{2n}, && \varphi_{2n}(t_j)=\varphi_{2n-1}(t_{j+1}),\\
&\quad t\in\theta_j,
\end{aligned}\tag{48.1}$$

where

$$f=\sum_{\alpha=1}^{2n} f_\alpha.$$

It is assumed that

$$v(t_{j+1})=\varphi_{2n}(t_{j+1}).$$

EXAMPLE 48.2 (BAKER and OLIPHANT [1960]). The system has the form:

$$\begin{aligned}
&\partial\varphi_1/\partial t+\tfrac{1}{2}A_1\varphi_1=f_1, && \varphi_1(t_j)=v(t_j),\\
&\quad\vdots\\
&\partial\varphi_{n-1}/\partial t+\tfrac{1}{2}A_{n-1}\varphi_{n-1}=f_{n-1}, && \varphi_{n-1}(t_j)=\varphi_{n-2}(t_{j+1}),\\
&\partial\varphi_n/\partial t+A_n\varphi_n=f_n, && \varphi_n(t_j)=\varphi_{n-1}(t_{j+1}),\\
&\partial\varphi_{n+1}/\partial t+\tfrac{1}{2}A_{n-1}\varphi_{n+1}=f_{n+1}, && \varphi_{n+1}(t_j)=\varphi_n(t_{j+1}),\\
&\quad\vdots\\
&\partial\varphi_{2n}/\partial t+\tfrac{1}{2}A_1\varphi_{2n}=f_{2n}, && \varphi_{2n}(t_j)=\varphi_{2n-1}(t_{j+1}),\\
&\quad t\in\theta_j
\end{aligned}\tag{48.2}$$

for

$$f=\sum_{\alpha=1}^{2n} f_\alpha, \qquad v(t_{j+1})=\varphi_{2n}(t_{j+1}).$$

EXAMPLE 48.3 (MARCHUK [1980], p. 276). The system is assumed to have the following forms. On the interval $t_{j-1}\leqslant t\leqslant t_j$:

$$\begin{aligned}
&\partial\varphi_\alpha/\partial t+A_\alpha\varphi_\alpha=0, \quad \alpha=1,2,\ldots,n-1,\\
&\partial\varphi_n/\partial t+A_n\varphi_n=f+\tfrac{1}{2}\tau A_n f;
\end{aligned}\tag{48.3}$$

and on the interval $t_j \leqslant t \leqslant t_{j+1}$:

$$\begin{aligned} &\partial\varphi_{n+1}/\partial t + A_n\varphi_{n+1} = f - \tfrac{1}{2}\tau A_n f, \\ &\partial\varphi_{n+\alpha}/\partial t + A_{n-\alpha+1}\varphi_{n+\alpha} = 0, \quad \alpha = 2, 3, \ldots, n, \end{aligned} \tag{48.4}$$

under conditions,

$$\begin{aligned} &\varphi_1(t_{j-1}) = v(t_{j-1}), \\ &\varphi_{\alpha+1}(t_{j-1}) = \varphi_\alpha(t_j), \quad \alpha = 1, 2, \ldots, n, \\ &\varphi_{\beta+1}(t_j) = \varphi_\beta(t_{j+1}), \quad \beta = n, \ldots, 2n-1. \end{aligned} \tag{48.5}$$

The approximation of problem (46.1) using (48.3)–(48.5) is considered on the double interval (t_{j-1}, t_{j+1}) under the assumption that

$$v(t_{j+1}) = \varphi_{2n}(\psi_{j+1}).$$

EXAMPLE 48.4 (MARCHUK [1980], p. 276). The system is assumed to have the following forms. On the interval $t_{j-1} \leqslant t \leqslant t_j$:

$$\begin{aligned} &\partial\varphi_1/\partial t + A_1\varphi_1 = 0, \\ &\quad\vdots \\ &\partial\varphi_n/\partial t + A_n\varphi_n = 0, \\ &\varphi_1(t_{j-1}) = v(t_{j-1}), \\ &\varphi_\alpha(t_{j-1}) = \varphi_{\alpha-1}(t_j), \quad \alpha = 2, 3, \ldots, n; \end{aligned} \tag{48.6}$$

on the interval $t_{j-1} \leqslant t \leqslant t_{j+1}$:

$$\partial\varphi_{n+1}/\partial t = f, \quad \varphi_{n+1}(t_{j-1}) = \varphi_n(t_j); \tag{48.7}$$

and on the interval $t_j \leqslant t \leqslant t_{j+1}$:

$$\begin{aligned} &\partial\varphi_{n+2}/\partial t + A_n\varphi_{n+2} = 0, \\ &\quad\vdots \\ &\partial\varphi_{2n+1}/\partial t + A_1\varphi_{2n+1} = 0, \\ &\varphi_\alpha(t_j) = \varphi_{\alpha-1}(t_{j+1}), \quad \alpha = n+2, \ldots, 2n+1. \end{aligned} \tag{48.8}$$

Problem (48.6)–(48.8) approximates (46.1) on (t_{j-1}, t_{j+1}) and under the assumption that

$$v(t_{j+1}) = \varphi_{2n+1}(\psi_{j+1}).$$

All the systems in this section have second-order accuracy and A_α and A_β, $\alpha \neq \beta$, do not have to be commutative. Theoretical aspects concerning these systems are considered in the above-cited literature.

49. Some results on convergence and stability

We assume that appropriate finite difference approximations in all variables (including t) are applied to the systems of Cauchy problems considered in Sections 46–48 and that the possible splitting of the difference approximations of the operators A_α into $A_{\alpha\beta}$ is carried out. Many of the thus obtained splitting schemes can be written in a canonical form:

$$B\frac{\varphi^{j+\alpha/m}-\varphi^{j+(\alpha-1)/m}}{\tau}+\sum_{\beta=0}^{m}A_{\alpha\beta}\varphi^{j+\beta/m}=f^{j+\alpha/m},$$
$$\alpha=1,\dots,m,\quad j=0,1,\dots,J, \tag{49.1}$$
$$\varphi^0=g,$$

where all grid functions are assumed to belong to a Hilbert space $\Phi\equiv H_h$, and $A_{\alpha\beta}$ and B are linear operators from H_h in H_h. The approximation error of (49.1) is defined as the sum of the approximation errors for different equations:

$$\psi=\psi_1+\cdots+\psi_m, \tag{49.2}$$

ψ_α being represented in the form $\psi_\alpha=\psi_\alpha^0+\psi_\alpha^*$ so that $\Sigma_{\alpha=1}^m\psi_\alpha^0=0$ (see SAMARSKII [1970]). We assume further that (49.1) is solvable and that the inverse operator B^{-1} exists.

Let us formulate the statements establishing the connection between stability, approximation and convergence of the schemes of type (49.1) using norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ in the space H_h.

Let the scheme (49.1) be stable, i.e. the following a priori estimate holds:

$$\|\varphi^{j+1}\|_{(1)}\leqslant M_1\|g\|_{(1)}+M_2\max_{0\leqslant j'\leqslant j}\sum_{\alpha=1}^{m}\|f^{j'+\alpha/m}\|_{(2)}, \tag{49.3}$$

where the constants M_1 and M_2 do not depend on τ, the grid parameter h in other variables, g and $f^{j+\alpha/m}$, and the "smoothness" condition for the solution $\varphi(t)$ holds:

$$\sum_{\alpha=1}^{n-1}\left\|\sum_{\beta=1}^{m}A_{\alpha\beta}\left(\sum_{k=\beta+1}^{m}B^{-1}\psi_k^0\right)\right\|_{(2)}\leqslant M_0, \tag{49.4}$$

where $M_0=\text{const}>0$ does not depend on τ and h. Then (49.1) converges and the following estimate is valid:

$$\|\varphi^{j+1}-\varphi(t_{j+1})\|_{(1)}$$
$$\leqslant M_2\max_{0\leqslant j'\leqslant j}\left[\sum_{\alpha=1}^{m}\|\psi_\alpha^*(t_{j'})\|_{(2)}\right.$$
$$\left.+\tau\sum_{\alpha=1}^{m-1}\left\|\sum_{\beta=1}^{m}A_{\alpha\beta}\left(\sum_{k=\beta+1}^{m}B^{-1}\psi_k^0(t_{j'})\right)\right\|_2\right] \tag{49.5}$$

(see SAMARSKII [1970]).

Let us now suppose that $A_{\alpha 0}=0$, $B=B^*>0$ is the constant operator and

$$\sum_{\alpha,\beta=1}^{m}(A_{\alpha\beta}\xi_\beta, \xi_\alpha)\geqslant 0$$

for any $\xi_\alpha \in H_h$, $\alpha=1,\dots m$. Then the following estimate holds for the solution of problem (49.1):

$$\|\varphi^{j+1}\|_B \leqslant \|g\|_B + M_2 \max_{0\leqslant j'\leqslant j}\left[\left\|\sum_{\alpha=1}^{m} f^{j'+\alpha/m}\right\|_{B^{-1}} + \sqrt{\tau}\sum_{\alpha=1}^{m}\|f^{j'+\alpha/m}\|_{B^{-1}}\right], \tag{49.6}$$

where

$$\|\varphi\|_B=(B\varphi,\varphi)^{1/2}, \qquad \|\varphi\|_{B^{-1}}=(B^{-1}\varphi,\varphi)^{1/2}$$

(SAMARSKII [1970]).

Scheme (49.1) converges in the norm $\|\cdot\|_B$, if $\varphi(0)=\varphi^0$ and the conditions of the above statement hold as well as the summarized approximation conditions

$$\max_{0\leqslant j\leqslant J}\left\|\sum_{\alpha=1}^{m}\psi^{j+\alpha/m}\right\|_{B^{-1}}\to 0 \quad \text{for } |h|\to 0,\ \tau\to 0,$$

$$\max_{0\leqslant j\leqslant J}\sum_{\alpha=1}^{m}\|\psi^{j+\alpha/m}\|_{B^{-1}}\leqslant M_0=\text{const}>0,$$

where M_0 does not depend on the grid parameters h and τ (SAMARSKII [1970]).

If $\varphi(0)=\varphi^0$,

$$\left\|\sum_{\alpha=1}^{m}\psi_\alpha\right\|_{B^{-1}}=O(|h|^l+\tau^k),$$

where $k>0$, $l>0$, and the conditions of the above statement hold, then scheme (49.1) converges in the metric $\|\cdot\|_B$ with rate $O(|h|^l+\tau^{k_1})$, where $k_1=\min(k_1,\frac{1}{2})$. If, besides, the restriction (49.4) holds for $\|\cdot\|_{(2)}=\|\cdot\|_{B^{-1}}$, then the scheme converges with rate $O(|h|^l+\tau^{k_2})$, where $k_2=\min(k,1)$ (SAMARSKII [1970], p. 215).

In conclusion, we mention the works dealing with the problems connected with the splitting of the initial problem into a system of simpler problems. General principles of such splitting are given by YANENKO and DEMIDOV [1966], MARCHUK ([1971]; [1980], p. 275), SAMARSKII ([1971], p. 400). Theoretical grounding of such a splitting may be found in works by YANENKO and DEMIDOV [1966], SAMARSKII ([1970]; [1971], p. 403), YANENKO ([1967], p. 170).

The work by MARCHUK ([1980], Chapter V) is dedicated to two-cycle (symmetrized) splitting schemes; theoretical grounding and some modifications are given by SAMARSKII [1971] on the basis of symmetrized splitting. FRYAZINOV [1968, 1969] constructed the schemes for the equations of parabolic type in domains

consisting of p-dimensional parallelepipeds. Somehow a different symmetrization method for the difference schemes is used by GODUNOV and ZABRODIN [1962] for the acoustics equations. A large bibliography on all these problems is available in works by MARCHUK [1980] and SAMARSKII [1970, 1971].

CHAPTER XIII

Convergence Studies and Optimization of Iterative Methods

50. Sufficient conditions for convergence

In this chapter we will consider some variants of the alternating direction method for solving a system of linear algebraic equations

$$Au=f \tag{50.1}$$

with a nonsingular $(N \times N)$ matrix A and vector $f \in \mathbb{R}^N$. Let us take the following formula as the general form of a stationary method

$$B_\tau(u^{k+1}-u^k)=-\alpha\tau(Au^k-f), \quad k=0, 1, \ldots, \tag{50.2}$$

where α and τ are some positive parameters and

$$B_\tau=\prod_{i=1}^{m}(I+\tau B_i) \tag{50.3}$$

is a nonsingular matrix. Here m is a fixed positive integer, and B_i, $i=1, \ldots, m$, is an arbitrary $(N \times N)$ matrix. Here and in the sequel the unit matrix is denoted by E. Further, we will consider nonstationary, iterative methods for which the parameters used depend on the step number, k.

Method (50.2) is called convegent when, for any initial approximation $u^0 \in \mathbb{R}^N$, the sequence of vectors $u^1, u^2, \ldots$ converges to the solution $u^* = A^{-1}f$ of system (50.1)

Let value m and matrices $B_1, \ldots, B_m$ be given. Then the following statement is valid: If the real parts of the eigenvalues of the matrix A are positive, then, for any $\alpha>0$, there exists such a $\hat{\tau}>0$ that method (50.2) converges for any positive $\tau \in (0, \hat{\tau})$.

The inverse result is also valid. If value $\alpha>0$ is given, then a necessary condition for the method to converge for any sufficiently small $\tau>0$ is that the real parts of the eigenvalues of the matrix A must be positive.

The uniform convergence for $\tau \to 0$ of the eigenvalues of the matrix B_τ to the unit matrix is used to prove the above statements. These proofs are given in the book by MARCHUK and KUZNETSOV [1972].

We will consider further only two cases. Let us call the first case commutative. It supposes the existence of a nonsingular matrix Q, which, by a similarity transformation, simultaneously reduces the matrices $A, B_1, \ldots, B_m$ to a diagonal form, i.e.

$$\begin{aligned} &Q^{-1}AQ=\operatorname{diag}\{\lambda_1,\ldots,\lambda_N\},\\ &Q^{-1}B_iQ=\operatorname{diag}\{\lambda_1^{(i)},\ldots,\lambda_N^{(i)}\},\quad i=1,\ldots,m. \end{aligned} \tag{50.4}$$

Obviously, matrices $A, B_1, \ldots, B_m$, as well as matrix B_τ, can commute. In the commutative case we will further assume that the eigenvalues λ_j of the matrix A are real and positive and that the eigenvalues $\lambda_j^{(i)}$ of the matrices $B_i, i=1,\ldots,m$, are real and nonnegative.

The second case, called noncommutative, supposes that $m=2$, the matrix A is determined to be positive in $\mathbb{R}^N$, $A=B_1+B_2$, and matrices B_1 and B_2 are at least positive-semidefinite and, generally speaking, noncommutative. For the sake of simplicity we will use the following notations $B_1=A_1$ and $B_2=A_2$ in the noncommutative case.

Let us, first, determine sufficient conditions for the convergence of the commutative alternating direction method. To this end, we introduce the function

$$g_{\alpha,\tau}(\lambda,\lambda^{(1)},\ldots,\lambda^{(m)})=1-\frac{\alpha\tau\lambda}{\prod_{i=1}^{m}(1+\tau\lambda^{(i)})}. \tag{50.5}$$

Then the spectral radius $\rho(T_{\alpha,\tau})$ of the step matrix

$$T_{\alpha,\tau}=I-\alpha\tau B_\tau^{-1}A \tag{50.6}$$

of method (50.2) can be computed by the formula:

$$\rho(T_{\alpha,\tau})=\max_{1\leqslant j\leqslant N}|g_{\alpha,\tau}(\lambda_j,\lambda_j^{(1)},\ldots\lambda_j^{m})|. \tag{50.7}$$

Since the requirement that the inequality $\rho(T_{\alpha,\tau})<1$ holds is a necessary and sufficient condition for convergence of method (50.2), we have to determine the conditions for which $-1<g_{\alpha,\tau}(\lambda_j,\lambda_j^{(1)},\ldots,\lambda_j^{(m)})<1$ for all j.

Let matrix $B=\Sigma_{i=1}^m B_i$ be nonsingular. Then there exists $\hat{\alpha}>0$ such that $\rho(T_{\alpha,\tau})<1$ for any $\alpha\in(0,\hat{\alpha})$ and $\tau>0$, i.e. the commutative alternating direction method (50.2) for all such values of parameters α and τ converges.

If we assume additionally that $A=B$, then the commutative alternating direction method (50.2) converges for any $\alpha\in(0,2)$ and $\tau>0$.

To prove this statement it is sufficient to note that by assuming $A=\Sigma_{i=1}^m B_i$, we have $\lambda_j=\Sigma_{i=1}^m\lambda_j^{(i)}$. Therefore,

$$0<\frac{\tau\lambda_j}{\prod_{i=1}^{m}(1+\tau\lambda_j^{(i)})}<1,\quad j=1,\ldots N,$$

for any $\tau>0$.

We consider one important example. Let m $(n_i \times n_i)$ matrices S_i of a simple structure be given, i.e. nonsingular $(n_i \times n_i)$ matrices Q_i exist such that

$$Q_i^T S_i Q_i = \operatorname{diag}\{\mu_1^{(i)}, \ldots, \mu_{n_i}^{(i)}\},$$

where the eigenvalues $\mu_1^{(i)}, \ldots, \mu_{n_i}^{(i)}$ are nonnegative for $i=1, \ldots, m$. Define the $(N \times N)$ matrices $(N = n_1 \cdots n_m)$

$$B_i = E_1 \otimes \cdots \otimes E_{i-1} \otimes S_i \otimes E_{i+1} \otimes \cdots \otimes E_m, \quad i=1, \ldots, m, \tag{50.8}$$

where E_s denotes the unit $(n_s \times n_s)$ matrix for $s=1, \ldots, m$, and $\otimes$ denotes the tensor product of matrices. Assume also that

$$A = \sum_{i=1}^{m} B_i. \tag{50.9}$$

Such matrices occur when elliptic boundary problems with separating variables in rectangular domains are approximated by the grid method.

Obviously, with the additional assumption that $\det A \neq 0$, i.e.

$$\mu_j = \sum_{i=1}^{m} \mu_j^{(i)} \neq 0, \quad j=1, \ldots, N,$$

all requirements for the commutative alternating direction method hold. Then, from the above considerations, method (50.2) for solving system (50.1)—with matrix A from (50.9) and matrices $B_1, \ldots, B_m$ from (50.8)—converges for any $\alpha \in (0, 2)$ and for $\tau > 0$.

For a more detailed description of the commutative alternating direction method see Section 51.

Now we consider the noncommutative alternating direction method. Taking the above assumptions into account, we will write the method in a different, more convenient form:

$$(E + \tau A_1)(E + \tau A_2)(u^{k+1} - u^k) = -\alpha\tau(Au^k - f). \tag{50.10}$$

For $\alpha = 1$ we obtain the well-known DOUGLAS-RACHFORD [1956] method and for $\alpha = 2$ the PEACEMAN-RACHFORD [1955] method:

$$(E + \tau A_1)(E + \tau A_2)(u^{k+1} - u^k) = -2\tau(Au^k - f), \quad k=0, 1, \ldots \tag{50.11}$$

We will consider the latter in more detail. Our results will be based on papers by BIRKHOF and VARGA [1959], WACHPRESS and HABETLER [1960], KELLOGG [1963], SAMARSKII [1964a] and others.

We introduce the matrix $D_\tau = (E + \tau A_2)^T (E + \tau A_2)$, error vectors $z^k = u^k - u^*$ of (50.11), and vectors $y_k = (E + \tau A_1)^{-1}(E - \tau A_2) z^k$. Then, after simple transformations,

we immediately obtain the equalities:

$$(E+\tau A_2)z^{k+1} = [(E-\tau A_1)(E+\tau A_1)^{-1}(E-\tau A_2)(E+\tau A_2)^{-1}] \cdot (E+\tau A_2)z^k, \tag{50.12}$$

$$\| z^{k+1} \|_{D_\tau} = \frac{\| (E-\tau A_1)y^k \|}{\| (E+\tau A_1)y^k \|} \cdot \frac{\| (E-\tau A_2)z^k \|}{\| (E+\tau A_2)z^k \|} \cdot \| z^k \|_{D_\tau}$$

and the inequality:

$$\| z^{k+1} \|_{D_\tau} \leqslant \| T_{1,\tau} \| \cdot \| T_{2,\tau} \| \cdot \| z^k \|_{D_\tau}. \tag{50.13}$$

Here $\| v \| = (v, v)^{1/2}$ is the usual Euclidean norm of the vector $v \in \mathbb{R}^N$ and $\| T \|$ is the corresponding norm of the matrix T,

$$\| v \|_{D_\tau} = (D_\tau v, v)^{1/2} = \| (E+\tau A_2)v \|,$$

$$T_{i,\tau} = (E-\tau A_i)(E+\tau A_i)^{-1}, \quad i = 1, 2.$$

It is obvious that

$$\| T_\tau \|_{D_\tau} \leqslant \| T_{1,\tau} \| \cdot \| T_{2,\tau} \|, \tag{50.14}$$

where

$$T_\tau = E - 2\tau B_\tau^{-1} A = (E+\tau A_2)^{-1}(E+\tau A_1)^{-1}(E-\tau A_1)(E-\tau A_2).$$

It follows from BIRKHOF and VARGA [1939] that, if at least one of the matrices A_1, A_2 is positive-definite then the alternating direction method (50.11) converges for any $\tau > 0$.

Assuming in addition that matrices A_1 and A_2 are symmetric (naturally, the conditions of positive-semidefiniteness of matrices A_1 and A_2 and nonsingularity of the matrix $A = A_1 + A_2$ remain) this statement can be reinforced, i.e. the alternating direction method (50.11) converges for any $\tau > 0$. To prove this latter statement it is—according to MARCHUK and KUZNETSOV [1972]—sufficient to show that for any nonzero $z \in \mathbb{R}^N$ the following inequality holds:

$$\| T_\tau z \|_{D_\tau} < \| z \|_{D_\tau}. \tag{50.15}$$

Hence the inequality $\| T_\tau \|_{D_\tau} < 1$ holds, since $\mathbb{R}^N$ is finite-dimensional and T_τ is a linear operator, which means that (50.11) converges. Let us use (50.12) to determine inequality (50.15).

It is easily seen that with the above assumptions the inequality $\| z^{k+1} \|_{D_\tau} \leqslant \| z^k \|_{D_\tau}$ follows from (50.12) and equality holds only when $A_1 y^k = 0$ and $A_2 z^k = 0$. But the latter is possible only, if $z^x \in \operatorname{Ker} A_1 \cap \operatorname{Ker} A_2 = \operatorname{Ker} A$, which contradicts the nonsingularity of matrix A and the condition $z^k \neq 0$. Thus, inequality (50.15) has been proved.

We consider one interesting variant of the noncommutative alternating direction method. Namely, we assume that matrix A is symmetric (we have assumed earlier

that A is positive-definite) and $A_1 = A_2^T$. This means that matrices A_1 and A_2 are positive-definite since the equalities

$$(A_i z, z) = \tfrac{1}{2}(Az, z), \qquad i = 1, 2, \quad z \in \mathbb{R}^N,$$

take place in real space. We will call this variant of the alternating direction method (50.11) symmetric. Since matrices A_1 and A_2 are positive-definite, the method's convergence for any $\tau > 0$ follows from the above considerations.

The important properties of the symmetric alternating direction method are symmetry and positive-definiteness of the matrix

$$\begin{aligned} B_\tau &= (E + \tau A_1)(E + \tau A_1^T) \\ &= (E + \tau A_2^T)(E + \tau A_2), \end{aligned} \tag{50.16}$$

this means that the step matrix T_τ can be symmetrized in the inner product generated by the matrix A. Indeed,

$$\begin{aligned} (T_\tau v, w)_A &= (v, w)_A - 2\tau(B_\tau^{-1} Av, w)_A \\ &= (v, w)_A - 2\tau(B_\tau^{-1} Av, Aw) \\ &= (v, w)_A - 2\tau(v, B_\tau^{-1} Aw)_A \\ &= (v, T_\tau w)_A \end{aligned}$$

for any $v, w \in \mathbb{R}^N$. Moreover, it follows therefore that matrix $B_\tau^{-1} A$ is not only A-self-adjoint but also A-positive-definite. Thus, for any $\tau > 0$ all eigenvalues of the matrix T_τ are real and belong to the interval $(-1, 1)$, and the system of its eigenvectors constitutes the A-orthonormal basis of $\mathbb{R}^N$.

Let us briefly discuss the influence of the parameter α on the convergence of the noncommutative alternating direction method. To this end, we consider the right-hand side of the equality

$$\begin{aligned} \| z^{k+1} \|_{D_\tau}^2 = \| z^k \|_{D_\tau}^2 &- 2\alpha\tau(B_\tau^{-1} Az^k, Az^k)_{D_\tau} \\ &+ \alpha^2 \tau^2 \| B_\tau^{-1} Az^k \|_{D_\tau}^2 \end{aligned}$$

as a square polynomial in α. First, we note that, due to the nonsingularity of the matrix $B_\tau^{-1} A$, we have for any nonzero $z^k \in \mathbb{R}^N$,

$$\| B_\tau^{-1} Az^k \|_{D_\tau} \neq 0.$$

Further for $\alpha = 0$ we have

$$\| z^{k+1} \|_{D_\tau} = \| z^k \|_{D_\tau},$$

and for $\alpha = 2$, according to (50.13), we have the inequality

$$\| z^{k+1} \|_{D_\tau} \leqslant \| z^k \|_{D_\tau}.$$

It thus follows that for any $\tau > 0$

$$\frac{d}{d\alpha} \| z^{k+1} \|_{D_\tau}^2 \Big|_{\alpha=0} = -2\tau(B_\tau^{-1} Az^k, Az^k)_{D_\tau} < 0$$

and, hence, for any $\alpha \in (0, 2)$ the following inequality analogous to inequality (50.15) holds:

$$\| z^{k+1} \|_{D_\tau} < \| z^k \|_{D_\tau}.$$

Therefore, we conclude that

$$\| T_{\alpha,\tau} \|_{D_\tau} < 1$$

for any $\tau > 0$ and $\alpha \in (0, 2)$.

Hence, under the above assumptions the noncommutative alternating direction method converges for any $\tau > 0$ and $\alpha \in (0, 2)$.

From this statement it follows in particular that the Douglas–Rachford method, i.e. (50.10) with parameter $\alpha = 1$, converges for all positive-semidefinite matrices A_1 and A_2.

51. The choice of parameters in the commutative alternating direction method

By optimization of the iterative method we often mean a way of increasing its convergence rate either by a special choice of parameters or with the help of acceleration procedures. For the stationary commutative alternating direction method considered above, the complete optimization presupposes minimizing the spectral radius of the step matrix $T_{\alpha,\tau}$, i.e. the solution of the problem

$$\min_{\alpha,\tau} \rho(T_{\alpha,\tau}). \tag{51.1}$$

This is, however, not an optimal approach since we can only in exceptional cases find $\rho(T_{\alpha,\tau})$ explicitly as a function of variables α and τ. That is why in practice, a set $G \subset \mathbb{R}^{m+1}$ such that all $(m+1)$-dimensional vectors $\hat{\lambda}_j = (\lambda_j, \lambda_j^{(1)}, \ldots \lambda_j^{(m)}) \in G$, $j = 1, \ldots, N$, the components of which are the eigenvalues of the diagonal matrices A, $B_1, \ldots, B_m$ of (50.4), is actually introduced, as well as a function $\Phi_{\alpha,\tau}(\lambda)$ such that for any $\lambda \in G$ the following inequality holds:

$$|g_{\alpha,\tau}(\lambda)| \leqslant \Phi_{\alpha,\tau}(\lambda). \tag{51.2}$$

If the function

$$\Psi(\alpha, \tau) = \max_{\lambda \in G} \Phi_{\alpha,\tau}(\lambda) \tag{51.3}$$

approximates $\rho(T_{\alpha,\tau})$ sufficiently well for values (α, τ) of a set $H \subset \mathbb{R}^2$, in which the best values for the parameters are found, then instead of (51.1) we may consider the approximate optimization problem for the iterative method, i.e. solving the extremum problem

$$\min_{(\alpha,\tau) \in H} \Psi(\alpha, \tau). \tag{51.4}$$

Of course, the required function Ψ can be constructed in several stages subsequently

coming to new majorants $\Phi_{\alpha,\tau}$, by expanding the set G and narrowing the set H, i.e. by imposing additional restrictions on the values of parameters. The analogous transition procedure from extremum problem (51.1) to approximating problem (51.4) can be used in other cases as well, in particular, for multiparametric alternating direction methods.

In considering concrete ways of choosing parameters in the commutative alternating direction method, we assume additionally that $m=\alpha=2$ and $A=A_1+A_2$, where $A_1=B_1$ and $A_2=B_2$. Then

$$g_{\alpha,\tau}(\lambda)\equiv g_\tau(\lambda)=\frac{1-\tau\lambda^{(1)}}{1+\tau\lambda^{(1)}}\cdot\frac{1-\tau\lambda^{(2)}}{1+\tau\lambda^{(2)}},$$

the set $G\subset\mathbb{R}^2$, and the set H consists of positive values of τ. Choose

$$G=[\delta_1,\Delta_1]\times[\delta_2,\Delta_2], \tag{51.5}$$

where

$$\delta_i=\min_{i\leqslant j\leqslant N}\lambda_j^{(i)},\quad \Delta_i=\max_{1\leqslant j\leqslant N}\lambda_j^{(i)},\qquad i=1,2;$$

take

$$\Psi(\alpha,\tau)\equiv\Psi(\tau)=\max_{\lambda=G}|g_\tau(\lambda)|;$$

and assume, in addition, that

$$\delta_1=\delta_2=\delta>0,\qquad \Delta_1=\Delta_2=\Delta>\delta.$$

Then

$$\Psi_{(\tau)}=\left[\max_{\delta\leqslant\lambda\leqslant\Delta}\left|\frac{1-\tau\lambda}{1+\tau\lambda}\right|\right]^2,$$

and the value τ_{opt}, minimizing the function $\Psi(\tau)$, is computed according to the formula,

$$\tau_{\text{opt}}=(\delta\Delta)^{-1/2}. \tag{51.6}$$

Hence, we arrive at the estimate,

$$\rho(T_{\tau_{\text{opt}}})\leqslant\left[\frac{\sqrt{\Delta}-\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}\right]^2. \tag{51.7}$$

However, if $\delta_1\neq\delta_2$ or $\Delta_1\neq\Delta_2$, but $\delta_1\delta_2\neq0$, then with the choice of $G=[\delta,\Delta]\times[\delta,\Delta]$, where $\delta=\min\{\delta_1,\delta_2\}$ and $\Delta=\max\{\Delta_1,\Delta_2\}$, we again obtain (51.6) and the estimate (51.7). But if, for example, $\delta_2=0$ then, choosing

$$\Psi_{(\tau)}=\max_{\delta_1\leqslant\lambda\leqslant\Delta_1}\left|\frac{1-\tau\lambda}{1+\tau\lambda}\right|, \tag{51.8}$$

we obtain $\tau_{opt}=(\Delta_1\delta_1)^{-1/2}$ and come to the estimate

$$\rho(T_{\tau_{opt}})\leqslant\frac{\sqrt{\Delta_1}-\sqrt{\delta_1}}{\sqrt{\Delta_1}+\sqrt{\delta_1}}.$$

We now consider the simplest multiparametric alternating direction method,

$$(E+\tau_k A_1)(E+\tau_k A_2)(u^{k+1}-u^k)=-2\tau_k(Au^k-f),\quad k=0,1,\ldots, \tag{51.9}$$

where the parameters τ_k have period $p>1$, i.e. $\tau_k=\tau_{k+p}$ for all $k\geqslant 0$. If we introduce the matrix $D=Q^TQ$, where the matrix Q is defined by (50.4), then we get the inequality

$$\|z^k\|_D=\max_{1\leqslant j\leqslant N}\left|\prod_{i=0}^{k-1}\frac{1-\tau_1\lambda_j^{(1)}}{1+\tau_1\lambda_j^{(1)}}\cdot\frac{1-\tau_i\lambda_j^{(2)}}{1+\tau_i\lambda_j^{(2)}}\right|\|z^0\|_D.$$

Hence, for all $k=tp$, t a positive integer, we have

$$\|z^k\|_D\leqslant\left[\max_{\delta\leqslant\lambda\leqslant\Delta}\prod_{i=0}^{p}\left|\frac{1-\tau_i\lambda}{1+\tau_i\lambda}\right|\right]^{2t}\|z^0\|_D, \tag{51.10}$$

where

$$\delta=\min_{\substack{1\leqslant j\leqslant N\\ i=1,2}}\lambda_j^{(i)}\quad\text{and}\quad\Delta=\max_{\substack{1\leqslant j\leqslant N\\ i=1,2}}\lambda_j^{(i)}$$

(δ is assumed to be positive).

We introduce the function

$$\Psi(\tau)\equiv\Psi(\tau_0,\ldots,\tau_{p-1})=\max_{0\leqslant i\leqslant p-1}\max_{m_i\leqslant\lambda\leqslant m_{i+1}}\left|\frac{1-\tau_i\lambda}{1+\tau_i\lambda}\right|, \tag{51.11}$$

where $\delta=m_0<m_1<\cdots<m_p=\Delta$ are some positive integers.

Obviously,

$$\|z^k\|_D\leqslant[\Psi(\tau)]^{2t}\|z^0\|_D,$$

with $\Psi(\tau)<1$ for all positive $\tau_0,\tau_1,\ldots,\tau_{p-1}$. Therefore the approximate optimization problem for method (51.9) can be formulated as the problem of minimizing the function $\Psi(\tau)$ on the set $H\subset\mathbb{R}^p$, which consists of vectors with positive components.

Since

$$\min_{\tau\in H}\Psi(\tau)=\max_{0\leqslant i\leqslant p-1}\left[\min_{\tau_i}\max_{m_i\leqslant\lambda\leqslant m_{i+1}}\left|\frac{1-\tau_i\lambda}{1+\tau_i\lambda}\right|\right], \tag{51.12}$$

it is sufficient to solve the minimax problems,

$$\min_{\tau_i>0}\max_{m_i\leqslant\lambda\leqslant m_{i+1}}\left|\frac{1-\tau_i\lambda}{1+\tau_i\lambda}\right|,\quad i=0,1,\ldots,p-1,$$

in order to solve the problem of minimizing the function $\Psi(\tau)$ for τ. According to the above, the solutions of the minimax problems can be found by choosing

$$\tau_i = \tau_{i,\mathrm{opt}} = (m_i m_{i+1})^{-1/2}, \quad i = 0, 1, \ldots, p-1. \tag{51.13}$$

Now we have to choose values $m_1, \ldots, m_{p-1}$, so that the right-hand side of (51.12) is minimal. Obviously, it is enough to take

$$m_i = \delta(\Delta/\delta)^{i/p}, \quad i = 0, 1, \ldots, p.$$

Hence, for $\tau = \tau_{\mathrm{opt}} = (\tau_{0,\mathrm{opt}}, \ldots, \tau_{p-1,\mathrm{opt}})^{\mathrm{T}}$ we obtain the estimate

$$\|z^k\|_D \leqslant \left[\frac{1-h^{1/p}}{1+h^{1/p}}\right]^{2t} \|z^0\|_D, \tag{51.14}$$

where $h = (\delta/\Delta)^{1/2}$.

So, the average term for decreasing the D-norm of the error vector of the optimized cyclic parameter with period p for the alternating direction method is equal to

$$q = \left[\frac{1-h^{1/p}}{1+h^{1/p}}\right]^{2/p}. \tag{51.15}$$

Assume that $h \ll 1$ and find out for what value p (51.15) will be minimal. To find p, taking $h \ll 1$ into account, we get the equation

$$\frac{\mathrm{d}}{\mathrm{d}p}\left(\frac{h^{1/p}}{p}\right) = 0;$$

its solution is the value

$$p = -\ln h.$$

With this value p (51.14) takes the form:

$$\|z^k\|_D \leqslant \left[\frac{\mathrm{e}-1}{\mathrm{e}+1}\right]^{2k/p} \|z^0\|_D. \tag{51.16}$$

Thus, if under the assumption $h \ll 1$, the problem is to decrease the D-norm of the initial error vector in $1/\varepsilon$ times ($\varepsilon < 1$), then with the values of parameters found the method allows to do this in

$$k_\varepsilon \sim \ln(1/h)\ln(1/\varepsilon) \tag{51.17}$$

steps while the one-parameter method requires

$$k_\varepsilon \sim h^{-1}\ln(1/\varepsilon) \tag{51.18}$$

steps.

The above method is one of the simplest for the approximate solution of the optimization problem for the multiparametric alternating direction method. It was used first by PEACEMAN and RACHFORD [1955], DOUGLAS and RACHFORD [1956]

and DYAKONOV [1961, 1962b]. Further studies on the choice of optimum parameters can be found in the works by WACHPRESS [1962], TODD [1967], LEBEDEV [1977] and others. These results have been thoroughly considered by SAMARSKII and NIKOLAEV [1978].

52. The choice of parameters in the noncommutative alternating direction method

Consider, first, the case when matrices A_1 and A_2 are symmetric and one of them, for example, the matrix A_1 is positive-definite. Then, according to (50.14), the estimate

$$\rho(T_\tau) \leqslant \| T_{1,\tau} \| = \max_{\delta_1 \leqslant \lambda \leqslant \Delta_1} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right|, \tag{52.1}$$

occurs, where δ_1 and Δ_1 are minimal and maximal eigenvalues of the matrix A_1 respectively. Now, if we take

$$\Psi(\tau) = \| T_{1,\tau} \| = \max \left\{ \left| \frac{1-\tau\delta}{1+\tau\delta} \right|, \left| \frac{1-\tau\Delta}{1+\tau\Delta} \right| \right\},$$

then with the choice

$$\tau_{\text{opt}} = (\delta_1 \Delta_1)^{-1/2} \tag{52.2}$$

we come to the solution of the approximate optimization problem of the alternating direction method (50.11). As a result we get the estimate

$$\rho(T_{\tau_{\text{opt}}}) \leqslant \frac{\sqrt{\Delta_1} - \sqrt{\delta_1}}{\sqrt{\Delta_1} + \sqrt{\delta_1}}. \tag{52.3}$$

Assume in addition that matrix A_2 is also positive-definite and that δ_2 and Δ_2 are its minimal and maximal eigenvalues respectively. Then we obtain the estimates

$$\| T_{i,\tau} \| = \max_{\delta_i \leqslant \lambda \leqslant \Delta_i} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right| \leqslant \max_{\delta \leqslant \lambda \leqslant \Delta} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right|, \quad i = 1, 2, \tag{52.4}$$

where

$$\delta = \min\{\delta_1, \delta_2\}, \qquad \Delta = \max\{\Delta_1, \Delta_2\}.$$

Now, if we take

$$\Psi_{(\tau)} = \max_{\delta \leqslant \lambda \leqslant \Delta} \left| \frac{1-\tau\lambda}{1+\tau\lambda} \right|^2,$$

then, according to (50.14) and (52.4), we have

$$\rho(T_\tau) \leqslant \Psi(\tau) < 1$$

for any $\tau > 0$. In this case we get the solution of the approximate optimization problem for (50.11) by the choice

$$\tau_{\text{opt}} = (\delta\Delta)^{-1/2}. \tag{52.5}$$

As a result we get the estimate

$$\rho(T_{\tau_{\text{opt}}}) \leqslant \| T_\tau \|_{D_\tau} \leqslant \left[\frac{\sqrt{\Delta}-\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}\right]^2. \tag{52.6}$$

We now consider a non-self-adjoint case. To construct the majorant $\Psi(\tau)$ let us use SAMARSKII's [1964a] method. Since

$$\rho(T_\tau) \leqslant \| T_\tau \|_{D_\tau} \leqslant \| T_{1,\tau} \| \cdot \| T_{2,\tau} \|, \tag{52.7}$$

it will be sufficient to derive the majorant, for example, for $\| T_{1,\tau} \|$, assuming that A_1 is positive-definite. We introduce the values

$$\delta_1 = \inf_{z \in \mathbb{R}^N} \frac{(A_1 z, z)}{(z,z)}, \qquad \Delta_1 = \sup_{z \in \mathbb{R}^N} \frac{\| A_1 z \|^2}{(A_1 z, z)}. \tag{52.8}$$

Obviously, both values exist and are positive, with δ_1 being the minimal eigenvalue of the matrix $S = \frac{1}{2}(A_1 + A_2^{\mathrm{T}})$. It is easy to see that

$$\| T_{1,\tau} \|^2 = \sup_{z \in \mathbb{R}^N} \frac{\| z \|^2 - 2\tau(A_1 z, z) + \tau^2 \| A_1 z \|^2}{\| z \|^2 + 2\tau(A_1 z, z) + \tau^2 \| A_1 z \|^2}$$

$$= \sup_{z \in \mathbb{R}^N} \frac{\| z \|^2 - (2\tau - \tau^2 \Delta_1)(A_1 z, z)}{\| z \|^2 + (2\tau + \tau^2 \Delta_1)(A_1 z, z)}.$$

Hence, using inequality $(A, z, z) \leqslant \Delta_1 \| z \|^2$, we obtain $\| T_{1,\tau} \|^2 \leqslant \Psi^2(\tau)$, where

$$\Psi^2(\tau) = \max_{\delta_1 \leqslant \lambda \leqslant \Delta_1} \left| \frac{1 - (2\tau - \tau^2 \Delta_1)\lambda}{1 + (2\tau + \tau^2 \Delta_1)\lambda} \right|.$$

The analysis of function $\Psi(\tau)$ shows that for $\tau = (\delta_1 \Delta_1)^{-1/2}$ it reaches its minimal value

$$\Psi_{\min} = \left[\frac{\sqrt{\Delta_1}-\sqrt{\delta_1}}{\sqrt{\Delta_1}+\sqrt{\delta_1}}\right]^{1/2}.$$

Thus, if matrix A_1 is positive-definite, then as the solution of the approximate optimization problem we can choose the parameter $\tau_{\text{opt}} = (\delta_1 \Delta_1)^{-1/2}$ for which the estimate

$$\| T_{\tau_{\text{opt}}} \|_{D_\tau} \leqslant \left[\frac{\sqrt{\Delta_1}-\sqrt{\delta_1}}{\sqrt{\Delta_1}+\sqrt{\delta_1}}\right]^{1/2}$$

holds. Assuming that the matrix A_2 is also positive-semidefinite then for $\tau_{\text{opt}} = (\delta\Delta)^{-1/2}$ we obtain the estimate

$$\| T_{\tau_{\text{opt}}} \|_{D_\tau} \leqslant \frac{\sqrt{\Delta}-\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}, \tag{52.9}$$

where $\delta = \min(\delta_1, \delta_2)$ and $\Delta = \max(\Delta_1, \Delta_2)$, and the values δ_2 and Δ_2 are defined as in (52.8) but for matrix A_2.

53. The convergence acceleration procedures for the alternating direction method

We introduce a new matrix

$$D_\tau = A^T[(I+\tau A_1)^{-1}](I+\tau A_1)^{-1}A. \tag{53.1}$$

Then, by simple transformations analogous to those of Section 50, we can show that

$$\| z^{k+1} \|^2_{D_\tau} = \frac{\| x^k \|^2 - 2\tau(A_1 x^k, x^k) + \tau^2 \| A_1 x^k \|^2}{\| x^k \|^2 + 2\tau(A_1 x^k, x^k) + \tau^2 \| A_1 x^k \|^2}$$

$$\cdot \frac{\| y^k \|^2 - 2\tau(A_2 y^k, y^k) + \tau^2 \| A_2 y^k \|^2}{\| y^k \|^2 + 2\tau(A_2 y^k, y^k) + \tau^2 \| A_2 y^k \|^2} \cdot \| z^k \|^2_{D_\tau},$$

where

$$x^k = (E+\tau A_1)^{-1}(E-\tau A_2)y^k,$$
$$y^k = (E+\tau A_2)^{-1}(E+\tau A_1)^{-1}Az^k,$$

and, therefore, for arbitrary positive-semidefinite matrices A_1 and A_2 we have

$$\| z^{k+1} \|_{D_\tau} \leqslant \| T_{1,\tau} \| \cdot \| T_{2,\tau} \| \cdot \| z^k \|_{D_\tau} \leqslant \| z^k \|_{D_\tau}.$$

From this and from equality

$$\| z^{k+1} \|^2_{D_\tau} = \| z^k \|^2_{D_\tau} - 2\tau(B_\tau^{-1}Az^k, z^k)_{D_\tau} + \tau^2 \| B_\tau^{-1}Az^k \|^2_{D_\tau},$$

it follows that matrix $B_\tau^{-1}A$ is D_τ-positive-definite.

In view of this fact we can solve system (50.1) not by (50.10) or (50.11) but by the following iterative method

$$(E+\tau A_1)(E+\tau A_2)(u^{k+1}-u^k) = -\beta_k(Au^k - f), \tag{53.2}$$

for which parameter β_k is chosen according to the formula

$$\beta_k = \frac{(B_\tau^{-1}Az^k, z^k)_{D_\tau}}{\| B_\tau^{-1}Az^k \|^2_{D_\tau}} = \frac{(AB_\tau^{-1}\zeta^k, \zeta^k)_{\tilde{D}_\tau}}{\| AB_\tau^{-1}\zeta^k \|^2_{\tilde{D}_\tau}}, \tag{53.3}$$

where $\zeta^k = Au^k - f$ is the residual vector, and

$$\tilde{D}_\tau = [(E+\tau A_2)^{-1}]^T(E+\tau A_1)^{-1}.$$

The parameter β_k is chosen in correspondence to the descent method from the maximum minimization condition of the D_τ-norm of the error vector z^{k+1}. As is known (see, for instance, MARCHUK and KUZNETSOV [1972]), this method converges with

$$L(T) \leqslant \| T_\tau \|_{D_\tau}, \tag{53.4}$$

where the $L(T)$ are Lipschitz constants of step operator T of method (53.2)–(53.3).

Note that the similar procedure for the inner product generated by matrix D_τ of Section 50 will not be suitable here.

Among different versions of the alternating direction method the case $A_1 = A_2^T$ considered in Section 50 stands out as rather efficient. The transition matrix (50.16) of method (50.10) is A-self-adjoint, which is the reason for using acceleration procedures based on the Chebyshev methods and the generalized conjugate gradient method.

Let it be known that eigenvalues of matrix $B_\tau^{-1}A$ of the symmetric aternating direction method belong to the interval $[a, b)$ with $0 < a < b$. For example, if we use the results from the previous section, then it follows, from the estimate (52.9), that the interval boundaries can be computed according to the formulae

$$a = \frac{\delta\sqrt{\Delta}}{\sqrt{\Delta}+\sqrt{\delta}}, \qquad b = \frac{\Delta\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}.$$

Under this assumption we can use the following Chebyshev semi-iterative method (see VARGA [1962]) with the pre-conditioning matrix B_τ^{-1} for the solution of system (50.1):

$$\begin{aligned} &B_\tau(u^1 - u^0) = -\alpha_1(Au^0 - f), \\ &B_\tau(u^{k+1} - u^k) = -\alpha_{k+1}(Au^k - f) + \beta_{k+1}\beta_\tau(u^k - u^{k-1}), \\ &\quad k = 1, 2, \ldots, \end{aligned} \tag{53.5}$$

where

$$\alpha_{k+1} = \frac{4}{b-a}\frac{C_k(\eta)}{C_{k+1}(\eta)}, \quad \beta_{k+1} = \begin{cases} 0, & k=0, \\ \dfrac{C_{k-1}(\eta)}{C_{k+1}(\eta)}, & k>0, \end{cases} \quad \eta = \frac{b+a}{b-a} \tag{53.6}$$

and $C_k(t)$ is the Chebyshev polynomial of the kth order of the first kind in t.

According to MARCHUK and KUZNETSOV [1972] and GOLUB, CONCUS and O'LEARY [1976] we can also use two-term and three-term generalized conjugate gradient methods with the pre-conditioning matrix B_τ^{-1} for the solution of system (50.1). The two-term formulae of the method have the form:

$$\begin{aligned} &g_k = \begin{cases} B_\tau^{-1}\xi^0, & k=1, \\ B_\tau^{-1}\xi^{k-1} - \alpha_k g_k, & k>1, \end{cases} \\ &u^k = u^{k-1} - \beta_k g_k, \\ &\alpha_k = -\|\xi^{k-1}\|^2_{B_\tau^{-1}} / \|\xi^k\|^2_{B_\tau^{-1}} = \frac{(B_\tau^{-1}\xi^{k-1}, \xi^{k-1})}{(B_\tau^{-1}\zeta^k, \zeta^k)}, \\ &\beta_k = \frac{\|\xi^{k-1}\|^2_{B_\tau^{-1}}}{\|g_k\|^2_A}, \quad k = 1, 2, \ldots; \end{aligned} \tag{53.7}$$

and three-term formulae (see REID [1971]) have the form:

$$u^{k+1}=u^k-\frac{1}{q_k}[B_\tau^{-1}\zeta^k-e_{k-1}(u^k-u^{k-1})],$$

$$e_{-1}=0, \qquad q_k=\frac{\|B_\tau^{-1}\zeta^k\|_A^2}{\|\zeta^k\|_{B_\tau^{-1}}^2}-e_{k-1}, \tag{53.8}$$

$$e_k=q_k\frac{\|\zeta^{k+1}\|_{B_\tau^{-1}}^2}{\|\zeta^k\|_{B_\tau^{-1}}^2}, \quad k=0, 1, 2, \ldots.$$

Here $\zeta^k=Au^k-f$ are residual vectors.

54. Generalizations

The best known generalization of the original alternating direction method is the following one:

$$(R+\tau A_1)R^{-1}(R+\tau A_2)(u^{k+1}-u^k)=-\alpha\tau(Au^k-f), \quad k=0, 1, 2, \ldots, \tag{54.1}$$

where R is, generally speaking, an arbitrary symmetric and positive-definite matrix (see WACHPRESS and HABETLER [1960]). It is easy to see that this method is completely equivalent to the alternating direction method

$$(I+\tau\tilde{A}_1)(I+\tau\tilde{A}_2)(\tilde{u}^{k+1}-\tilde{u}^k)=-\alpha\tau(\tilde{A}\tilde{u}^k-\tilde{f}), \quad k=0, 1, \ldots, \tag{54.2}$$

where

$$\tilde{A}_1=R^{-1/2}A_1R^{-1/2}, \qquad \tilde{A}_2=R^{-1/2}A_2R^{-1/2},$$
$$\tilde{A}=\tilde{A}_1+\tilde{A}_2, \qquad \tilde{u}=R^{1/2}u, \qquad \tilde{f}=R^{-1/2}f. \tag{54.3}$$

Hence, all the abovementioned facts for methods (50.10) and (50.11) refer to method (54.1) as well. Of course, the role of matrix R should be considered, i.e. in deriving the formulae for optimal parameters matrices $\tilde{A}_1$ and $\tilde{A}_2$ of (54.3) should be used instead of A_1 and A_2 of (54.1).

We consider the case $A_1=A_2^T$. To this end, we rewrite, first (54.1) with $\alpha=2$ in the following equivalent form:

$$(R+\tau A_1)u^{k+1/2}=(R-\tau A_2)u^k+f,$$
$$(R+\tau A_2)u^{k+1}=(R-\tau A_1)u^{k+1/2}+f, \tag{54.4}$$

and then, using

$$L=\tfrac{1}{2}R-A_1, \quad L^T=\tfrac{1}{2}R-A_2, \qquad \omega=\frac{2\tau}{2+\tau}, \tag{54.5}$$

we rewrite (54.4) in the form

$$\left(\frac{1}{\omega}R-L\right)(u^{k+1/2}-u^k)=-(Au^k-f),$$
$$\left(\frac{1}{\omega}R-L^{\mathrm{T}}\right)(u^{k+1}-u^{k+1/2})=-(Au^{k+1/2}-f). \tag{54.6}$$

If the above assumptions hold, and

$$A=K-L-L^{\mathrm{T}}, \tag{54.7}$$

$\tau\in(0,+\infty)$, an $\omega\in(0,2)$, then it is easy to see that (54.6) is a generalization of the symmetric upper relaxation method corresponding to splitting (54.7) of matrix A. Hence, the theory developed by LUNN [1964] (see also YOUNG [1971]) can also be used for optimizing method (54.6) by the choice of parameter ω.

CHAPTER XIV

Splitting and Decomposition Methods for Variational Problems

Many problems of economics, optimal control, hydrodynamics, etc. lead to variational problems with restrictions (DUVANT and LIONS [1972], BENSOUSSAN, LIONS and TEMAM [1975], GLOWINSKY, LIONS and TREMOLIER [1979]). To solve these problems BENSOUSSAN et al. [1975] developed methods, some of which (for example, decomposition methods) are based on splitting methods. In this chapter we will consider the general scheme of applying the splitting and decomposition methods to variational problems. We will do this following BENSOUSSAN et al. [1975] and without specifying the type of problems to be solved.

55. Splitting and decomposition methods for classical variational problems

Let the initial stationary problem, after approximation on all its variables, be reduced to the equation

$$A\varphi = f, \tag{55.1}$$

considered in finite-dimensional Hilbert space H with inner product $(\cdot,\cdot)_H$ and norm $\|\cdot\|_H = (\cdot,\cdot)_H^{1/2}$. Matrix A is assumed to be symmetric, positive-definite and representable in the form

$$A = \sum_{\alpha=1}^{n} A_\alpha, \quad n \geqslant 1, \tag{55.2}$$

where A_α are symmetric, positive-definite matrices. It is known that solving equation (55.1) is equivalent to solving the following classical problem of variational calculus

$$J(\varphi) = \inf_{v \in H} J(v),$$

where

$$J(v) = \tfrac{1}{2}(Av, v)_H - (f, v)_H \equiv \sum_{\alpha=1}^{n} J_\alpha(v),$$

$$J_\alpha(v) = \tfrac{1}{2}(A_\alpha v, v)_H - (f\alpha, v)_H$$

and $\{f\alpha\}$ are chosen so that $\Sigma_{\alpha=1}^{n} f_\alpha = f$. According to the method of stationing (see Section 37) we can consider the evolutionary problem

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f, \quad t \in \Omega_t = (0, T), \\ &\varphi = g \qquad \text{for } t = 0, \quad g \in H, \end{aligned} \tag{55.4}$$

and, in case of large values of T, we can approximate the solution of (55.1) by the solution of problem (55.4) $\varphi(T)$. The numerical solution of (55.4), in turn, can be realized by the splitting method. To this end, we represent f as

$$f = \sum_{\alpha=1}^{n} f_\alpha, \quad f_\alpha \in H, \tag{55.5}$$

divide Ω_t in N intervals of length $\tau_0, \tau_1, \ldots, \tau_{N-1}$ such that

$$\tau_0 + \tau_1 + \cdots + \tau_{N-1} = G_N \geqslant T,$$

and use a splitting scheme of type

$$\begin{aligned} &\frac{\varphi^{j+\alpha/n} - \varphi^{j+(\alpha-1)/n}}{\tau_j} + A_\alpha \varphi^{j+\alpha/n} = f_\alpha, \\ &\quad 0 \leqslant j \leqslant N-1, \quad 1 \leqslant \alpha \leqslant n, \\ &\varphi_0 = g. \end{aligned} \tag{55.6}$$

Each realization step of scheme (55.6) is equivalent to the minimization, in H, of the following functional problem,

$$\begin{aligned} &\frac{1}{2\tau} \| v - \varphi^{j+(\alpha-1)/n} \|_H^2 + J_\alpha(v) \\ &\quad 0 \leqslant j \leqslant N-1, \quad 1 \leqslant \alpha \leqslant n, \\ &\varphi_0 = g. \end{aligned} \tag{55.7}$$

This algorithm constructed by using splitting scheme (55.6) is, in fact, the *decomposition method* for solving the variational problem (55.3). This same method reduces (55.3) to a sequence of problems minimizing the functionals (55.7). Note that the convergence of the decomposition algorithm (55.7) follows directly from the convergence theorems for scheme (55.6). These theorems have been considered in foregoing chapters.

56. Decomposition of a general variational problem

Based on algorithm (55.7), we now formulate the decompositon method for the following general variational problem in Hilbert space H:

$$J(\varphi) = \inf_{v \in H} J(v), \tag{56.1}$$

where $J: H \to \mathbb{R} \equiv \mathbb{R}^1$ is a continuous function with a lower bound. We assume that

$J(v)$ can be represented as

$$J = \sum_{\alpha=1}^{n} J_\alpha, \tag{56.2}$$

where the $J_\alpha: H \to \mathbb{R}$ are also continuous functions with a lower bound. We introduce a sequence of parameters τ_j, $j = 0, 1, \ldots, , N-1$ and define elements

$$\varphi^{j+\alpha/n}, \qquad j = 0, 1, \ldots, N-1, \quad \alpha = 1, \ldots, n, \tag{56.3}$$

where $\varphi^0 = g$ is an arbitrary element in H. If $\varphi^{j+(\alpha-1)/n}$ is known, then $\varphi^{j+\alpha/n}$ is determined as a solution of a variational problem of type

$$\inf_{v \in H} \left\{ \frac{1}{2\tau_j} \| v - \varphi^{j+(\alpha-1)/n} \|_H^2 + J_\alpha(v) \right\}. \tag{56.4}$$

This problem has at least one solution. We denote by $\varphi^{j+\alpha/n}$ one of these solutions and compute further $\varphi^{j+(\alpha+1)/n}, \varphi^{j+(\alpha+2)/n}, \ldots$.

It is obvious that the proposed algorithm for solving problem (56.1) is efficient when it is easier to solve a sequence of problems (56.4) than to solve one of (56.1) which depends on the explicit form of J and J_α. It is also easily seen that the given algorithm is of great interest for problems of type (55.3), obtained by discretizing elliptic boundary value problems, etc. Note that the splitting method theory does not allow us to conclude that algorithm (56.4) converges (see Section 58).

57. A variational problem with restrictions

Let K be a closed convex set in H (*the set of restrictions*) and, as in Section 56, let $J: H \to \mathbb{R}$ be a continuous function satisfying the assumptions from Section 56. Instead of (56.1) we consider the problem

$$\inf_{v \in K} J(v), \tag{57.1}$$

that is, a variational problem with restrictions. We also assume that

$$K = \bigcap_{\alpha=1}^{n} K_\alpha, \tag{57.2}$$

where $K_\alpha \subset H$ are closed sets which correspond to perturbations of restrictions. The following algorithm is, in fact, a generalization of algorithm (56.4) for problem (57.1): Starting with some element $\varphi^0 = g \in K_n$, find $\varphi^{j+\alpha/n} \in K_\alpha$ as a solution of problem

$$\inf_{v \in K_\alpha} \left\{ \frac{1}{2\tau_j} \| v - \varphi^{j+(\alpha-1)/n} \|_H^2 + J_\alpha(v) \right\}. \tag{57.3}$$

Algorithm (57.3) is a general form of the decomposition method for a variational problem with restrictions.

58. The convergence of decomposition algorithms

We now consider a decomposition algorithm in an infinite-dimensional space and formulate the conditions for its convergence.

Let Φ be a real reflective Banach space with norm $\|\cdot\|_{\Phi}$ and let K be a convex closed non-empty subset of Φ. Let J be a real functional defined on K such that J is strictly convex; J is lower semicontinuous on K (in a weak or strong topology Φ). We also assume that the set K is either limited or $\lim J(u) = +\infty$, when $\|u\|_{\Phi} \to \infty$, $u \in K$. It is known that under these conditions the functional J has a lower bound on K and there exists a unique element $\varphi \in K$ on which $J(v)$ reaches its minimum:

$$J(\varphi) \leqslant J(v), \quad v \in K. \tag{58.1}$$

Further, let another norm $\|\cdot\|_H$ be defined on Φ. Supplementing Φ with this norm we obtain a Hilbert space H. We assume that

$$\Phi = \bigcap_{\alpha=1}^{n} \Phi_{\alpha},$$

where the Φ_{α} are reflective Banach spaces with norms $\|\cdot\|_{\Phi_{\alpha}}^{\alpha-1}$ and the inclusions $\Phi \subset \Phi_{\alpha} \subset H$ are continuous ($\alpha = 1, \ldots, n$). We introduce the following assumption

$$K = \bigcap_{\alpha=1}^{n} K_{\alpha}, \tag{58.2}$$

where K_{α} is a convex subset of Φ_{α}. We also assume that $J(v)$ can be represented as

$$J(v) = \sum_{\alpha=1}^{n} J_{\alpha}(v), \quad v \in K, \tag{58.3}$$

where $J_{\alpha}(v)$ is a real, convex lower semicontinuous functional defined on K_{α} (in topology Φ_{α}) and $\lim J_{\alpha}(v) = +\infty$ when $\|v\|_{\Phi_{\alpha}} \to \infty, v \in K_{\alpha}$ (if K_{α} is not limited in Φ_{α}).

We formulate the decomposition algorithm for solving (58.1). Let $\tau_0, \tau_1, \ldots$ be a sequence of strictly positive numbers. If φ^0 is an arbitrary element of K_n, then define a family of elements $\varphi^{j+\alpha/n}$, $\alpha = 1, \ldots, n, j = 0, 1, \ldots$, where element $\varphi^{j+\alpha/n}$ is to be found (after $\varphi^0, \ldots, \varphi^{j+(\alpha-1)/n}$ as an element of K_{α} satisfying the inequality

$$\frac{1}{2\tau_j}\|\varphi^{j+\alpha/n} - \varphi^{j+(\alpha-1)/n}\|_H^2 + J_{\alpha}(\varphi^{j+\alpha/n})$$

$$\leqslant \frac{1}{2\tau_j}\|v - \varphi^{j+(\alpha-1)/n}\|_H^2 + J_{\alpha}(v) \tag{58.4}$$

for an element $v \in K_{\alpha}$. Problem (58.2) has a unique solution $\varphi^{j+\alpha/n} \in K_{\alpha}$. From $\varphi^0, \ldots, \varphi^{j+\alpha/n}$ we form the elements

$$\omega^{N+\alpha/n} = \frac{1}{\sigma_N} \sum_{j=0}^{N} \tau_j \cdot \varphi^{j+\alpha/n}, \tag{58.5}$$

where

$$\sigma_N = \sum_{j=0}^{N} \tau_j, \qquad \alpha = 1, \ldots, n, \quad N = 0, 1, \ldots .$$

If

$$\lim_{N \to \infty} \sigma_N = +\infty, \qquad \lim_{N \to \infty} \left\{ \max_{0 \leqslant j \leqslant N} \tau_j \right\} = 0,$$

then

$$\lim_{N \to \infty} \omega^{N + \alpha/n} = \varphi$$

in a weak topology Φ_α for $1 \leqslant \alpha \leqslant n$ where φ is the solution of problem (58.1).

In the work by BENSOUSSAN, LIONS and TEMAM ([1975], p. 200), one can find the proof of this statement as well as a number of concrete realizations of the above decomposition algorithms for solving variational problems of mathematical physics along with corresponding splitting methods.

PART 3

Applications of Splitting Methods to Problems of Mathematical Physics

In this part, we will consider the realization of the general statements of the splitting methods, from the previous chapters for some concrete problems of mathematical physics. We will describe the means to approximate problems in space, angular and other variables earlier accepted as achieved (see Part 1). All this is supposed to give us a complete notion of the methods used for approximating solutions of problems by splitting.

CHAPTER XV

The Heat Conduction Equation

We considered the major algorithms of the splitting and alternating direction methods in Part 1 and used the parabolic equation for illustration (Sections 12, 17, 22, 25, 26, 29, 36). In this chapter, we will consider the schemes for a more general form of the parabolic equation.

59. The two-cycle componentwise splitting scheme for a parabolic equation with three spatial variables

We consider the two-cycle componentwise splitting scheme in application to a diffusion model for propagation of pollutants in the atmosphere from pollution sources (MARCHUK [1982]). We write the problem in the following form:

$$\frac{\partial \varphi}{\partial t} + \sum_{\alpha=1}^{3} A_\alpha \varphi = f \quad \text{in } \Omega \times \Omega_t, \tag{59.1}$$

where

$$A_1\varphi = \frac{\partial u\varphi}{\partial x} - \mu\frac{\partial^2\varphi}{\partial x^2} - \tfrac{1}{2}\varphi\frac{\partial u}{\partial x},$$

$$A_2\varphi = \frac{\partial v\varphi}{\partial y} - \mu\frac{\partial^2\varphi}{\partial y^2} - \tfrac{1}{2}\varphi\frac{\partial v}{\partial y}, \tag{59.2}$$

$$A_3\varphi = \frac{\partial w\varphi}{\partial z} - \frac{\partial v}{\partial z}\frac{\partial \varphi}{\partial z} - \tfrac{1}{2}\varphi\frac{\partial w}{\partial z} - \sigma\varphi.$$

The following notations are used in (59.1) and (59.2): x, y, z are components of the Cartesian system of coordinates (the x-axis is directed to the east, the y-axis to the north, and the z-axis is directed vertically upward); $\Omega = \{0 < x < a,\ 0 < y < b,\ 0 < z < H\}$ is a domain with boundary $\partial\Omega$ consisting of the cylinder's side surface $\partial\Omega_s$ with lower basis $\partial\Omega_0$ (for $z = 0$) and upper basis $\partial\Omega_H$ (for $z = H$); $t \in \Omega_t = (0, T)$ denotes time; φ is the intensity of aerosol substance, migrating along with the flow of air in the atmosphere; u, v, w are components of wind velocity along x-, y- and z-axes respectively; $\mu = \text{const} > 0$, $\nu(z) > 0$ are the horizontal and vertical turbulence coefficients; $\sigma = \text{const} \geqslant 0$ is the coefficient of the substance's absorption by the

environment; $f(x, y, z, t)$ is the function describing the sources of the polluting substance φ under consideration.

It is assumed that for the components of the air flow velocity the mass conservation law represented by the continuity equation holds

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0 \quad \text{in } \Omega\times\Omega_t \tag{59.3}$$

and, besides, the conditions

$$w=0 \quad \text{for } z=0, \qquad z=H \tag{59.4}$$

hold. Values of the functions u, v and w are considered to be known in $\Omega\times\Omega_t$.

The initial condition

$$\varphi=g \quad \text{for } t=0 \tag{59.5}$$

is added to equation (59.1) and the boundary conditions on $\partial\Omega$ are:

$$\begin{aligned} &\varphi=0 && \text{on } \partial\Omega_s,\\ &\partial\varphi/\partial z=\alpha\varphi && \text{on } \partial\Omega_0,\\ &\partial\varphi/\partial z=0 && \text{on } \partial\Omega_H, \end{aligned} \tag{59.6}$$

where g is a given function and $\alpha\geqslant 0$ is some function characterizing the interaction of the pollutants and the Earth's surface. MARCHUK [1982] noted that more general boundary conditions can be used instead of (59.6). The solution φ and right-hand side belong respectively to the Hilbert spaces Φ and F of smooth enough functions.

A direct verification allowing for boundary conditions (59.6) shows that operators A_1, A_2 and A_3 are positive-definite, i.e. $A_\alpha>0$ ($\alpha=1, 2, 3$).

A uniform grid for x, y and z is introduced in Ω with steps Δx, Δy and Δz respectively. The operators A_α are approximated with second-order accuracy by their difference analogues Λ_α having the form (see MARCHUK [1982])

$$\begin{aligned} \Lambda_1\varphi=&\frac{u_{i+1/2,j,k}\varphi_{i+1,j,k}-u_{i-1/2,j,k}\varphi_{i-1,j,k}}{2\Delta x}\\ &-\frac{\mu}{\Delta x^2}(\varphi_{i+1,j,k}-2\varphi_{i,j,k}+\varphi_{i-1,j,k}),\\ \Lambda_2\varphi=&\frac{v_{i,j+1/2,k}\varphi_{i,j+1,k}-v_{i,j-1/2,k}\varphi_{i,j-1,k}}{2\Delta y}\\ &-\frac{\mu}{\Delta y^2}(\varphi_{i,j+1,k}-2\varphi_{i,j,k}+\varphi_{i,j-1,k}),\\ \Lambda_3\varphi=&\frac{w_{i,j,k+1/2}\varphi_{i,j,k+1}-w_{i,j,k-1/2},\varphi_{i,j,k-1}}{2\Delta z}+\sigma\varphi_{i,j,k}\\ &-\frac{1}{\Delta z^2}[v_{k+1/2}(\varphi_{i,j,k+1}-\varphi_{i,j,k})-v_{k-1/2}(\varphi_{i,j,k}-\varphi_{i,j,k-1})], \end{aligned} \tag{59.7}$$

$$i=1, 2, \ldots, I, \quad j=1, 2, \ldots, J, \quad k=1, 2, \ldots, K.$$

It is assumed that the difference analogues of u, v and w satisfy the difference

analogue of the continuity equation,

$$\frac{u_{i+1/2,j,k}-u_{i-1/2,j,k}}{\Delta x}+\frac{v_{i,j+1/2,k}-v_{i,j-1/2,k}}{\Delta y}+\frac{w_{i,j,k+1/2}-w_{i,j,k-1/2}}{\Delta z}=0 \tag{59.8}$$

with

$$w_{1/2}=w_{k+1/2}=0. \tag{59.9}$$

Due to the approximation of boundary conditions (59.6) and by allowing for (59.9), the expressions (59.7) are somewhat changed in the vicinity of boundary points:

$$\begin{aligned}
&\varphi_{i,j,k}=0 \qquad \text{for } i=0, \quad i=I+1,\\
&\varphi_{i,j,k}=0 \qquad \text{for } j=0, \quad j=J+1,\\
&\frac{\varphi_{i,j,k}-\varphi_{i,j,k-1}}{\Delta z}=\alpha\varphi_{i,j,k} \quad \text{for } k=1,\\
&\frac{\varphi_{i,j,k+1}-\varphi_{i,j,k}}{\Delta z}=0 \qquad \text{for } k=K.
\end{aligned} \tag{59.10}$$

It is easily shown that difference operators Λ_1, Λ_2, Λ_3 acting in the Hilbert space $\Phi=H_h$ are positive-definite on the functions $\varphi_{i,j,k}$ that satisfy the boundary conditions (59.10). We write the obtained system of ordinary differential equations with respect to $\varphi=\{\varphi_{i,j,k}\}$ in the form,

$$\begin{aligned}
&\frac{d\varphi}{dt}+\sum_{\alpha=1}^{3}\Lambda_\alpha\varphi=f \quad \text{in } \Omega^h\times\Omega_t,\\
&\varphi=g \quad \text{for } t=0.
\end{aligned} \tag{59.11}$$

To solve (59.11) let us use the two-cycle componentwise splitting method (τ is the time step)

$$\begin{aligned}
&\frac{\varphi^{n+1/6}-\varphi^n}{\tau}+\Lambda_1\frac{\varphi^{n+1/6}+\varphi^n}{2}=0,\\
&\frac{\varphi^{n+2/6}-\varphi^{n+1/6}}{\tau}+\Lambda_2\frac{\varphi^{n+2/6}+\varphi^{n+1/6}}{2}=0,\\
&\frac{\varphi^{n+3/6}-\varphi^{n+2/6}}{\tau}+\Lambda_3\frac{\varphi^{n+3/6}+\varphi^{n+2/6}}{2}=0,\\
&\frac{\varphi^{n+4/6}-\varphi^{n+3/6}}{\tau}+\Lambda_3\frac{\varphi^{n+4/6}+\varphi^{n+3/6}}{2}=0,\\
&\frac{\varphi^{n+5/6}-\varphi^{n+4/6}}{\tau}+\Lambda_2\frac{\varphi^{n+5/6}+\varphi^{n+4/6}}{2}=0,\\
&\frac{\varphi^{n+1}-\varphi^{n+5/6}}{\tau}+\Lambda_1\frac{\varphi^{n+1}+\varphi^{n+5/6}}{2}=0,\\
&n=0,\ 1,\ 2,\ldots\ .
\end{aligned} \tag{59.12}$$

Scheme (59.12) has second-order accuracy in t and is absolutely stable. Each splitting step is realized by the scalar sweep method (GODUNOV and RYABENKII [1977], pp. 51–53).

60. Schemes of second-order accuracy for p-dimensional parabolic equations without mixed derivatives

We consider the p-dimensional heat conduction equation with constant coefficients written in nondimensional form:

$$\begin{aligned} &\partial\varphi/\partial t + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \\ &A \equiv \sum_{\alpha=1}^{p} A_\alpha \equiv \sum_{\alpha=1}^{p} \left(-\frac{\partial^2}{\partial x_\alpha^2} \right), \end{aligned} \tag{60.1}$$

where $f = f(x, t)$ is a known function of coordinates $x = (x_1, x_2, \ldots, x_p)$ and time t. Additionally, the initial condition

$$\varphi = g \quad \text{for } t = 0 \tag{60.2}$$

is set along with the boundary condition,

$$\varphi = 0 \quad \text{in } \partial\Omega \times \Omega_t, \tag{60.3}$$

which does not restrict generality, and the function $g(x)$ is considered to be known.

We assume that Ω is a p-dimensional parallelepiped

$$\{0 \leqslant x_1 \leqslant 1, \ldots, 0 \leqslant x_\alpha \leqslant 1, \ldots, 0 \leqslant x_p \leqslant 1\},$$

and we obtain the grid domain $\Omega^h = \{x_\alpha^h\}_{\alpha=1}^{p}$, where $x_\alpha^h = \{x_{i_\alpha} = i_\alpha h_\alpha\}$ $(i_\alpha = 0, 1, \ldots, I_\alpha + 1)$, and the step h_α for variable x_α does not depend on the number i_α of the grid point. On smooth solutions, we approximate the operator A_α with second-order accuracy by its discrete analogue Λ_α with the help of the relationship

$$\begin{aligned} &\Lambda_\alpha \varphi^h = -\frac{1}{h_\alpha^2}(\varphi_{i_\alpha+1}^h - 2\varphi_{i_\alpha}^h + \varphi_{i_\alpha-1}^h) \equiv -\frac{\Delta_\alpha \Delta_{-\alpha}}{h_\alpha^2}\varphi^h, \\ &1 \leqslant i_\alpha \leqslant I_\alpha. \end{aligned} \tag{60.4}$$

The differential operators A_α and their difference analogues Λ_α on the functions satisfying the condition (56.3) are positive-definite in the corresponding spaces $\Phi = H$ and $\Phi = H_h$. Thus, the initial problem has been reduced to a system of ordinary differential equations (index h is everywhere omitted)

$$\begin{aligned} &\frac{d\varphi}{dt} + \Lambda\varphi = f \quad \text{in } \Omega^h \times \Omega_t, \\ &\Lambda = \sum_{\alpha=1}^{p} \Lambda_\alpha, \qquad \varphi = 0 \quad \text{in } \partial\Omega^h \times \Omega_t \end{aligned} \tag{60.5}$$

with initial condition (60.2).

To solve this problem, we use the scheme with the *splitting (factorized) operator*

(DYAKONOV [1971d]). This scheme has the form

$$B\frac{\varphi^{n+1}-\varphi^n}{\tau}+\sum_{\alpha=1}^{p}\Lambda_\alpha\varphi^n=\hat{f}^{n+1},\qquad \varphi^0=g,$$

$$\varphi^n_{i_\alpha}=0,\qquad i_\alpha=0,\quad i_\alpha=I_\alpha+1, \tag{60.6}$$

where $\hat{f}^{n+1}$ is the second-order approximation in time of the right-hand side f and operator B has the form:

$$B\equiv\prod_{\alpha=1}^{p}(E+\theta\tau\Lambda_\alpha)$$

$$\equiv E+\theta\tau\Lambda+(\theta\tau)^2\sum_{1\leqslant\alpha<\beta\leqslant p}\Lambda_\alpha\Lambda_\beta+\cdots+(\theta\tau)^p\prod_{\alpha=1}^{p}\Lambda_\alpha, \tag{60.7}$$

with $0\leqslant\theta\leqslant 1$ arbitrary. In particular, for $p=2$ we have

$$B=E+\theta\tau\Lambda+\theta^2\tau^2\Lambda_1\Lambda_2. \tag{60.8}$$

In view of (60.7), equation (60.6) can be rewritten in an equivalent form

$$\frac{\varphi^{n+1}-\varphi^n}{\tau}+\Lambda(\theta\varphi^{n+1}+(1-\theta)\varphi^n)+Q_p\frac{\varphi^{n+1}-\varphi^n}{\tau}=\hat{f}^{n+1}, \tag{60.9}$$

where

$$Q_p\equiv\sum_{\alpha=2}^{p}(\theta\tau)^\alpha\sum_{1\leqslant\alpha_1<\alpha_2<\cdots<\alpha_p\leqslant p}\Lambda_{\alpha_1}\Lambda_{\alpha_2}\cdots\Lambda_{\alpha_p}. \tag{60.10}$$

Thus, if the inequality

$$\left\|Q_p\left(\frac{\varphi^{n+1}-\varphi^n}{\tau}\right)\right\|_\Phi\leqslant M\tau^2,\quad M \text{ constant} \tag{60.11}$$

holds, then scheme (60.6) approximates the initial problem (60.5) with the same order of accuracy as the scheme

$$\frac{\varphi^{n+1}-\varphi^n}{\tau}+\Lambda(\theta\varphi^{n+1}+(1-\theta)\varphi^n)=\hat{f}^{n+1}. \tag{60.12}$$

As is known (DYAKONOV [1971d]), scheme (60.11) has a second-order approximation in t for $\theta=\frac{1}{2}$ and a first-order approximation otherwise. For $\theta\neq\frac{1}{2}$ the following inequality is sufficient

$$\left\|Q_p\frac{\varphi^{n+1}-\varphi^n}{\tau}\right\|_\Phi\leqslant M\tau.$$

Scheme (60.6) is absolutely stable for $\theta=\frac{1}{2}$ and for the solution the following estimate (DYAKONOV [1971d]) is valid:

$$\|\varphi^{n+1}\|_\Phi\leqslant\|\varphi^0\|_\Phi+\tau\sum_{j=0}^{n}\|B^{-1}f^{j+1}\|_F$$

$$\leqslant\|g\|_\Phi+\tau\sum_{j=0}^{n}\|f^{j+1}\|_F.$$

Taking into account the value of φ in the boundary points and the form of operator B, the scheme can be rewritten in the form

$$\prod_{\alpha=1}^{p}(E+\tau\theta\Lambda_\alpha)\varphi^{n+1}=\Phi^n, \tag{60.13}$$

where the right-hand side Φ^n is known. The realization of (60.13) consists in consecutively solving the problems

$$\begin{aligned}
&(E+\tau\theta\Lambda_1)\varphi^{n+1/p}=\Phi^n,\\
&(E+\tau\theta\Lambda_2)\varphi^{n+2/p}=\varphi^{n+1/p},\\
&\quad\vdots\\
&(E+\tau\theta\Lambda_p)\varphi^{n+1}=\varphi^{n+(p-1)/p}
\end{aligned} \tag{60.14}$$

by the sweep method.

The DOUGLAS–RACHFORD scheme [1956] is a special case of this scheme for $\theta=1$ and $f\equiv 0$.

61. Schemes for equations with mixed derivatives

We consider the first boundary value problem for a parabolic equation with mixed derivatives in the parallelepiped

$$\Omega=\{0\leqslant x_1\leqslant l_1,\ldots,0\leqslant x_\alpha\leqslant l_\alpha,\ldots,0\leqslant x_p\leqslant l_p\}$$

in the following form (SAMARSKII [1971], p. 372; YANENKO [1967], p. 34):

$$\begin{aligned}
&\frac{\partial\varphi}{\partial t}+\sum_{\alpha,\beta=1}^{p}k_{\alpha\beta}(t)A_{\alpha\beta}\varphi=f(x,t,\varphi),\\
&A_{\alpha\beta}\varphi=-\partial^2\varphi/\partial x_\alpha\partial x_\beta,\\
&0<c_1\sum_{\alpha=1}^{p}\xi_\alpha^2\leqslant\sum_{\alpha,\beta=1}^{p}k_{\alpha\beta}(t)\xi_\alpha\xi_\beta\leqslant c_2\sum_{\alpha=1}^{p}\xi_\alpha^2,\\
&\varphi=g\quad\text{for } t=0,\\
&\varphi=0\quad\text{in } \partial\Omega\times\Omega_t.
\end{aligned} \tag{61.1}$$

As in Section 60 we introduce a uniform grid for variables x_α with the step $h_\alpha=l_\alpha/(I_\alpha+1)$: $x_\alpha^h=i_\alpha h_\alpha, i_\alpha=0,1,\ldots,I_\alpha+1$. We approximate the differential operators $A_{\alpha\beta}$ with second-order accuracy by their finite difference analogues $\Lambda_{\alpha\beta}$. The operator $\Lambda_{\alpha\alpha}$ is written in the traditional form (see (60.4))

$$\Lambda_{\alpha\alpha}=-\Delta_\alpha\Delta_{-\alpha}/h_\alpha^2,\quad 1\leqslant i_\alpha\leqslant I_\alpha, \tag{61.2}$$

and, according to YANENKO [1967], $\Lambda_{\alpha\beta}$ for $\alpha\neq\beta$ can be written in the form

$$\Lambda_{\alpha\beta}\varphi^h=-\frac{1}{4h_\alpha h_\beta}[\varphi^h_{i_\alpha+1,i_\beta+1}-(\varphi^h_{i_\alpha+1,i_\beta-1}+\varphi^h_{i_\alpha-1,i_\beta+1})+\varphi^h_{i_\alpha-1,i_\beta-1}]. \tag{61.3}$$

To solve (61.1) we take the analogous scheme (60.6) with a factorized difference operator $B=\Pi_{\alpha=1}^{p} B_\alpha$:

$$B\frac{\varphi^{n+1}-\varphi^n}{\tau}+\sum_{\alpha,\beta=1}^{p} k_{\alpha\beta}(t_{n+1/2})\Lambda_{\alpha\beta}\varphi^n=\hat{f}^{n+1},$$
$$\varphi^0=g, \tag{61.4}$$
$$\varphi^n_{i_\alpha}=0 \qquad \text{for } i_\alpha=0 \quad \text{and} \quad i_\alpha=I_\alpha+1, \quad \alpha=1, 2, \ldots, p.$$

We choose operators B_α in the form (SAMARSKII [1971])

$$B_\alpha=E+\tau\theta c_2\Lambda_{\alpha\alpha}. \tag{61.5}$$

In $\Phi=H_h$ these operators are self-adjoint, positive (for $\theta>0$) and commutative (because the domain is a parallelepiped). The factorized scheme (61.4) has at least first-order accuracy in τ and is stable for $\theta\geqslant\frac{1}{2}$. With B_α, $\alpha=1, 2, \ldots, p$, being three-point difference operators with constant coefficients, the algorithm for computing φ^{n+1} for a given φ^n coincides with that of Section 60.

For the special case $f\equiv 0$, $k_{\alpha\beta}=\text{const}$, $p=2$, YANENKO [1967] considered the scheme based on the splitting method

$$\frac{\varphi^{n+1/2}-\varphi^n}{\tau}+k_{11}\Lambda_{11}\varphi^{n+1/2}+k_{12}\Lambda_{12}\varphi^n=0,$$
$$\frac{\varphi^{n+1}-\varphi^{n+1/2}}{\tau}+k_{21}\Lambda_{21}\varphi^{n+1/2}+k_{22}\Lambda_{22}\varphi^{n+1}=0. \tag{61.6}$$

This scheme has a first-order approximation in time and is absolutely stable.

For $p=3$ the following scheme has been introduced by YANENKO [1959b, 1961] generalizing (61.6):

$$\frac{\varphi^{n+1/6}-\varphi^n}{\tau}+\tfrac{1}{2}k_{11}\Lambda_{11}\varphi^{n+1/6}+k_{12}\Lambda_{12}\varphi^n=0,$$
$$\frac{\varphi^{n+2/6}-\varphi^{n+1/6}}{\tau}+k_{21}\Lambda_{21}\varphi^{n+1/6}+\tfrac{1}{2}k_{22}\Lambda_{22}\varphi^{n+2/6}=0,$$
$$\frac{\varphi^{n+3/6}-\varphi^{n+2/6}}{\tau}+\tfrac{1}{2}k_{33}\Lambda_{33}\varphi^{n+3/6}+k_{13}\Lambda_{13}\varphi^{n+2/6}=0,$$
$$\frac{\varphi^{n+4/6}-\varphi^{n+3/6}}{\tau}+k_{31}\Lambda_{31}\varphi^{n+3/6}+\tfrac{1}{2}k_{33}\Lambda_{33}\varphi^{n+4/6}=0, \tag{61.7}$$
$$\frac{\varphi^{n+5/6}-\varphi^{n+4/6}}{\tau}+\tfrac{1}{2}k_{22}\Lambda_{22}\varphi^{n+5/6}+k_{23}\Lambda_{23}\varphi^{n+4/6}=0,$$
$$\frac{\varphi^{n+1}-\varphi^{n+5/6}}{\tau}+k_{32}\Lambda_{32}\varphi^{n+5/6}+\tfrac{1}{2}k_{33}\Lambda_{33}\varphi^{n+1}=0.$$

Scheme (61.7) approximates the original problem with first-order accuracy in τ and

is stable if the matrix with elements $\frac{1}{2}k_{\alpha\alpha}$ on the main diagonal and $k_{\alpha\beta}$ on the other diagonals is positive-definite.

Schemes (61.6) and (61.7) are realized on each splitting step by the scalar sweep method.

62. Alternating direction schemes

In the framework of problem (61.1) we demonstrate the alternating direction type scheme proposed by SAMARSKII [1964a] for the case when $k_{\alpha\beta}=k_{\beta\alpha}$. For $p=3$ we have

$$\begin{aligned}
&\frac{\varphi^{n+1/6}-\varphi^n}{\tau}+\tfrac{1}{2}k_{11}\Lambda_{11}\varphi^{n+1/6}=f(x,t_{n+1/6},\varphi^n),\\
&\frac{\varphi^{n+2/6}-\varphi^{n+1/6}}{\tau}+\tfrac{1}{2}k_{22}\Lambda_{22}\varphi^{n+2/6}+k_{21}\Lambda_{21}\varphi^{n+1/6}=f(x,t_{n+2/6},\varphi^{n+1/6}),\\
&\frac{\varphi^{n+3/6}-\varphi^{n+2/6}}{\tau}+\tfrac{1}{2}k_{33}\Lambda_{33}\varphi^{n+3/6}+k_{31}\Lambda_{31}\varphi^{n+2/6}+k_{32}\Lambda_{32}\varphi^{n+2/6}\\
&\quad=f(x,t_{n+3/6},\varphi^{n+2/6}),\\
&\frac{\varphi^{n+4/6}-\varphi^{n+3/6}}{\tau}+\tfrac{1}{2}k_{33}\Lambda_{33}\varphi^{n+3/6}=f(x,t_{n+4/6},\varphi^{n+3/6}),\\
&\frac{\varphi^{n+5/6}-\varphi^{n+4/6}}{\tau}+\tfrac{1}{2}k_{22}\Lambda_{22}\varphi^{n+2/6}+k_{23}\Lambda_{23}\varphi^{n+4/6}=f(x,t_{n+5/6},\varphi^{n+4/6}),\\
&\frac{\varphi^{n+1}-\varphi^{n+5/6}}{\tau}+\tfrac{1}{2}k_{11}\Lambda_{11}\varphi^{n+1/6}+k_{12}\Lambda_{12}\varphi^{n+5/6}+k_{13}\Lambda_{13}\varphi^{n+4/6}\\
&\quad=f(x,t_{n+1},\varphi^{n+5/6}).
\end{aligned}\tag{62.1}$$

On smooth solutions this scheme has a second-order approximation in x_α, a first-order in τ and it converges for any $\tau\to 0$. However, there exists a restriction on $|h|=(\Sigma_{\alpha=1}^{p}h_\alpha^2)^{1/2}$: $|h|\leqslant h_0$, following from the restriction of matrix $(k_{\alpha\beta})$ being positive-definite. Constant h_0 depends on constants γ and γ_1 defined by the relationship

$$\sum_{\alpha,\beta=1}^{p}k_{\alpha\beta}\xi_\alpha\xi_\beta\geqslant\gamma\sum_{\alpha=1}^{p}\xi_\alpha^2 \tag{62.2}$$

in the case of a differential problem, and by the relationship

$$\sum_{\alpha,\beta=1}^{p}k_{\alpha\beta}\Lambda_{\alpha\beta}\xi_\alpha\xi_\beta\geqslant\gamma_1\sum_{\alpha=1}^{p}\xi_\alpha^2 \tag{62.3}$$

in the case of its difference analogue. Scheme (62.1) is realized by the sweep method

on the first three steps and by computations with explicit formulae on the final three stages.

For the case when

$$\begin{aligned} &k_{\alpha\beta} \neq k_{\beta\alpha}, \quad \alpha \neq \beta, \\ &k_{\alpha\beta} = 0, \quad \beta > \alpha, \end{aligned} \tag{62.4}$$

SAMARSKII [1964a] proposed the following difference scheme:

$$\begin{aligned} &\frac{\varphi^{n+1/3} - \varphi^n}{\tau} + k_{11}\Lambda_{11}\varphi^{n+1/3} = f(x, t_{n+1/3}, \varphi^n), \\ &\frac{\varphi^{n+2/3} - \varphi^{n+1/3}}{\tau} + k_{22}\Lambda_{22}\varphi^{n+2/3} + k_{21}\Lambda_{21}\varphi^{n+1/3} = f(x, t_{n+2/3}, \varphi^{n+1/3}), \\ &\frac{\varphi^{n+1} - \varphi^{n+2/3}}{\tau} + k_{33}\Lambda_{33}\varphi^{n+1} + k_{31}\Lambda_{31}\varphi^{n+1/3} + k_{32}\Lambda_{32}\varphi^{n+2/3} \\ &\quad = f(x, t_{n+1}, \varphi^{n+2/3}). \end{aligned} \tag{62.5}$$

This scheme converges for $\tau \to 0$, its realization is analogous to that in Section 61.

63. Schemes of increased order of accuracy

All the above difference schemes for solving parabolic equations did not have more than second-order accuracy in space variables and time. The splitting fractional steps method can be used for obtaining simple *schemes of increased order of accuracy.* Assume here that the grid is uniform and has the same step h in all variables x_α.

The schemes of increased accuracy for the simplest parabolic equations, constructed by factorizing the operator on the upper time layer are considered in works by DOUGLAS and GUNN [1963], SAMARSKII [1963b], SAMARSKII and GULIN [1973], YANENKO [1967], SAMARSKII and ANDREEV [1963]. These schemes are constructed on the basis of an m-layer homogeneous scheme of increased accuracy ($m \geqslant 2$ chosen) in the form:

$$\frac{\varphi^{n+1} - \varphi^n}{\tau} + L\varphi^{n+1} = f^n, \tag{63.1}$$

where f^n is the result of applying difference spatial operators to the functions φ^n, $\varphi^{n-1}, \ldots$.

We consider the method proposed by DOUGLAS and GUNN [1963] and assume that the following two-dimensional heat conduction equation has to be solved

$$\partial\varphi/\partial t + A\varphi = f \tag{63.2}$$

where

$$A = A_1 + A_2, \qquad A_\alpha = -\partial^2/\partial x_\alpha^2, \quad \alpha = 1, 2.$$

To solve (63.2) we use a *scheme of the stabilizing correction type*

$$\frac{\varphi^{n+1/2}-\varphi^n}{\tau}+L_1(\varphi^{n+1/2}-\varphi^n)+L\varphi^n=f^n,$$
$$\frac{\varphi^{n+1}-\varphi^{n+1/2}}{\tau}+L_2(\varphi^{n+1}-\varphi^n)=0. \tag{63.3}$$

We choose the following difference operators L and L_α as approximating operators:

$$L=\tfrac{2}{3}\Lambda-\tfrac{1}{6}(h/\tau)^2E,$$
$$L_1=\tfrac{2}{3}\Lambda_1-\tfrac{1}{6}(h/\tau)^2E, \qquad L_2=\tfrac{2}{3}\Lambda_2,$$

where Λ, Λ_1 and Λ_2 are approximations of the differential operators $-\Delta$, $-\partial^2/\partial x_1^2$, $-\partial^2/\partial x_2^2$. Then, under the condition $\tau/h=\text{const}$, scheme (63.3) has accuracy of order $O(\tau^2+h^2)$ and is absolutely stable.

If we choose for operators L, L_1 and L_2 the following,

$$L=\tfrac{2}{3}(1-\tfrac{1}{8}h^2/\tau)\Lambda,$$
$$L_\alpha=\tfrac{2}{3}(1-\tfrac{1}{8}h^2/\tau)\Lambda_\alpha, \quad \alpha=1, 2,$$

then scheme (63.3) will be absolutely stable and have the accuracy of order $O(\tau^2+h^2)$ if the condition $\tau/h^2=\text{const}$ holds.

Another method of constructing a factorized operator is proposed by SAMARSKII [1963b]. It is based on replacing (63.1) with the scheme,

$$\frac{\varphi^{n+1}-\varphi^n}{\tau}+L\varphi^{n+1}+\frac{1}{\tau}\Phi(\varphi^{n+1}-\varphi^n)=f^n, \tag{63.4}$$

where the operator Φ is chosen in such a way that the operator of the scheme (63.4) is factorized. If we split L according to the expansion,

$$L=L_1+L_2+\cdots+L_p,$$

then

$$E+\tau L+Q=(E+\tau L_1)\cdots(E+\tau L_p). \tag{63.5}$$

Hence, Q can be represented as

$$Q=(E+\tau L_1)\cdots(E+\tau L_p)-(E+\tau L)$$
$$=\tau^2\sum_{\alpha<\beta}L_\alpha L_\beta+\tau^3\sum_{\alpha<\beta<\gamma}L_\alpha L_\beta L_\gamma+\cdots+\tau^pL_1L_2\cdots L_p. \tag{63.6}$$

In the work by SAMARSKII [1963b] the following absolutely stable two-layer scheme of increased order of accuracy,

$$\frac{\varphi^{n+1}-\varphi^n}{\tau}+\theta\Lambda\varphi^{n+1}+(1-\theta)\Lambda\varphi^n+\tfrac{1}{6}h^2\sum_{\alpha<\beta}\Lambda_\alpha\Lambda_\beta\varphi^n=0,$$
$$\theta=\tfrac{1}{2}(1-\tfrac{1}{6}h^2/\tau), \qquad \Lambda=\sum_{\alpha=1}^{p}\Lambda_\alpha, \tag{63.7}$$

approximating the equation

$$\partial\varphi/\partial t + A\varphi = 0, \quad A = -\sum_{\alpha=1}^{p} \partial^2\varphi/\partial x_\alpha^2 \tag{63.8}$$

with accuracy of order $O(\tau^2 + h^4)$ is taken instead of the initial scheme (63.1).

In this case the factorized scheme has the form:

$$(E+\theta\tau\Lambda_1)(E+\theta\tau\Lambda_2)\cdots(E+\theta\tau\Lambda_p)\varphi^{n+1} = \Omega\varphi^n + f^n \tag{63.9}$$

where

$$\Omega = E - \tau\left[\tfrac{1}{6}h^2 \sum_{\alpha<\beta} \Lambda_\alpha\Lambda_\beta + (1-\theta)\Lambda\right]$$
$$-(\theta\tau)^2 \sum_{\alpha<\beta} \Lambda_\alpha\Lambda_\beta + (\theta\tau)^3 \sum_{\alpha<\beta<\gamma} \Lambda_\alpha\Lambda_\beta\Lambda_\gamma$$
$$+\cdots+(-1)^{p-1}(\theta\tau)^{p-1}\Lambda_1\Lambda_2\cdots\Lambda_p. \tag{63.10}$$

A scheme of increased order of accuracy is likewise constructed for the equation

$$\frac{\partial\varphi}{\partial t} - \sum_{\alpha,\beta=1}^{2} k_{\alpha\beta}\frac{\partial^2\varphi}{\partial x_\alpha \partial x_\beta} = 0,$$
$$k_{\alpha\beta} = \text{const}, \quad k_{11} = k_{22} = 1. \tag{63.11}$$

The initial homogeneous scheme has the form

$$\frac{\varphi^{n+1}-\varphi^{n-1}}{2\tau} + \frac{\Lambda_{11}+\Lambda_{22}}{2}(\varphi^{n+1}-\varphi^{n-1}) - 2k_{12}\Lambda_{12}\varphi^n$$
$$+\frac{h^2}{12}\frac{\varphi^{n+1}-2\varphi^n+\varphi^{n-1}}{\tau^2} + \frac{h^2}{6}b\Lambda_{11}\Lambda_{22}\varphi^n = 0, \tag{63.12}$$

where $\Lambda_{11}, \Lambda_{22}$ and Λ_{12} are given by formulae (61.2) and (61.3) and the parameter b is defined by

$$b = 1 + 2k_{12}^2 - 3|k_{12}|.$$

The factorized scheme has the form

$$L_1L_2\frac{\varphi^{n+1}-\varphi^n}{\tau} = (1-2\theta)\frac{\varphi^n-\varphi^{n-1}}{\tau} - \theta\Lambda(\varphi^n+\varphi^{n-1})$$
$$-4\theta k_{12}\Lambda_{12}\varphi^n + 2(1-\theta)\tau b\Lambda_1\Lambda_2\varphi^n, \tag{63.13}$$

where

$$L_i = E + \theta\tau\Lambda_{ii}, \quad \theta = \frac{1}{1+\frac{1}{6}h^2/\tau}.$$

All schemes from this section are realized on the basis of the scalar sweep method in all directions which is fairly simple.

64. Finite element method schemes and splitting schemes for two-dimensional parabolic equations

We consider the diffusion equation

$$\frac{\partial \varphi}{\partial t} - \sum_{\alpha,\beta=1}^{2} \frac{\partial}{\partial x_\alpha} k_{\alpha\beta} \frac{\partial \varphi}{\partial x_\beta} = f(x_1, x_2, t) \tag{64.1}$$

in the two-dimensional domain Ω with boundary $\partial\Omega$ and choose homogeneous boundary and initial conditions:

$$\begin{aligned} &\varphi = 0 \quad \text{for } \partial\Omega, \\ &\varphi = 0 \quad \text{for } t = 0. \end{aligned} \tag{64.2}$$

We assume that coefficients $k_{\alpha\beta}(x_1, x_2)$ satisfy the uniform ellipticity condition in Ω and that $f(x_1, x_2, t)$ is a sufficiently smooth function. We introduce a rectangular uniform grid with steps h_1 and h_2 in the domain Ω and triangulate each grid cell by diagonals of positive or negative direction depending on the sign of the integral

$$\delta_{mn} = \tfrac{1}{2} \int_{x_{1,m}}^{x_{1,m}+h_1} \int_{x_{2,n}}^{x_{2,n}+h_2} (k_{12} + k_{21}) \, dx_1 \, dx_2.$$

We denote by the domain Ω^h the largest conjunction of triangles belonging to Ω and we denote subdomains of Ω^h with positive and negative triangulations of cells by Ω_1^h and Ω_2^h respectively.

We introduce a piecewise linear prolongation of the grid functions on the triangulated domain Ω^h with the help of coordinate functions, $\omega_{m,n}(x_1, x_2)$:

$$\omega_{m,n}(x_{1,k}, x_{2,l}) = \begin{cases} 1 & \text{for } (m,n) = (k,l), \\ 0 & \text{for } (m,n) \neq (k,l), \end{cases} \tag{64.3}$$

$$\varphi^h = \sum \varphi_{m,n}(t) \omega_{m,n}(x_1, x_2), \quad (x_{1,m}, x_{2,n}) \in \Omega^h \setminus \partial\Omega^h.$$

The semidiscrete Galerkin equation for $\varphi_{m,n}(t)$ in using the "lumping" method for an evolutionary term can be written in the form:

$$\begin{aligned} &\frac{\partial \varphi_{m,n}}{\partial t}(1, \omega_{m,n}) + I(\varphi^h, \omega_{m,n}) = (f, \omega_{m,n}), \\ &(\varphi^h, \omega_{m,n}) = 0 \quad \text{for } t = 0, \\ &I(\varphi^h, \omega) = \int_\Omega \sum_{\alpha,\beta=1}^{2} k_{\alpha\beta} \frac{\partial \varphi^h}{\partial x_\beta} \frac{\partial \omega}{\partial x_\alpha} \, d\Omega, \\ &(\varphi^h, \omega) = \int_\Omega \varphi^h \omega \, d\Omega. \end{aligned} \tag{64.4}$$

In operator form equation (64.4) is given by

$$\theta \partial\varphi/\partial t + \Lambda\varphi = \tilde{f}. \tag{64.5}$$

Here $\theta > 0$ and $\Lambda \geqslant 0$ are grid operators, $\varphi = \{\varphi_{m,n}(t)\}$ and $\tilde{f} = \{(f, \omega_{m,n})\}$ are grid functions.

It is shown by MARCHUK and KUZIN [1983] that the grid operator Λ can be represented as the sum of one-dimensional three-point operators

$$\Lambda = \Lambda_{11} + \Lambda_{22} + \Lambda_{12} + \Lambda_{21},$$

where operator Λ_{11} acts on the grid lines which are parallel to the axis x_1, operator Λ_{22} on the lines parallel to the axis x_2, operator Λ_{12} on diagonals of positive direction in the subdomain Ω_1^h and Λ_{21} on negative diagonals in the subdomain Ω_2^h. It is proved that the operators are positive-semidefinite ($\Lambda_{\alpha\beta} \geqslant 0, \alpha,\beta = 1,2$) if the following conditions hold:

$$k_{11} \geqslant \frac{h_1}{2h_2}|k_{12} + k_{21}|, \qquad k_{22} \geqslant \frac{h_2}{2h_1}|k_{12} + k_{21}|.$$

Basing on this we can use one of the splitting methods, in particular the two-cycle method of type (59.12) with Crank–Nicolson schemes on each elementary step for solving (64.5) (MARCHUK [1980]).

Similar results are obtained for the generalized von Neumann boundary conditions.

The application of piecewise linear prolongation to equation (64.1) under the condition that $k_{12} + k_{21} = 0$ on a nonuniform triangulated grid which is topologically equivalent to a rectangular one also leads to the grid operator splitting in four one-dimensional operators. In this case the conditions of positive-semidefiniteness of the one-dimensional operators are reduced to some geometrical conditions on the grid parameters.

This method has been used in the ocean dynamics models for computing the stream function (KUZIN [1980]) and for solving the three-dimensional equation of heat flux in the ocean, which is reduced by special splitting along the sections to a series of equations of type (64.1) (KUZIN [1984]).

CHAPTER XVI

Equations of Hyperbolic Type

At present there exists a number of effective algorithms for solving problems that are connected with equations of a hyperbolic type. This chapter features methods based on special splitting algorithms for solving such problems.

65. The stabilization scheme for the multidimensional equation of oscillations

We consider an N-dimensional equation of oscillations

$$\partial^2\varphi/\partial t^2 + A\varphi = f \quad \text{in } \Omega \times \Omega_t, \tag{65.1}$$

$$\varphi = 0 \quad \text{on } \partial\Omega, \tag{65.2}$$

$$\partial\varphi/\partial t = q \text{ and } \varphi = p \qquad \text{in } \Omega \qquad \text{for } t = 0, \tag{65.3}$$

where Ω is an N-dimensional parallelepiped and $A = -\Sigma_{j=1}^{N} \partial^2/\partial x_j^2$. We approximate (65.1)–(65.2) in space by a finite difference problem on a uniform grid Ω^h with step h_j along the axis x_j:

$$\mathrm{d}^2\varphi^h/\mathrm{d}t^2 + \Lambda\varphi^h = f^h \quad \text{in } \Omega^h \times \Omega_t, \tag{65.4}$$

$$\varphi^h = p^h \text{ and } \mathrm{d}\varphi^h/\mathrm{d}t = q^h \quad \text{in } \Omega^h \qquad \text{for } t = 0. \tag{65.5}$$

Here

$$\Lambda = \sum_{j=1}^{N} \Lambda_j$$

$$\Lambda_1\varphi^h = \begin{cases} \dfrac{-\varphi_{i_1+1,\cdot} + 2\varphi_{i_1,\cdot}}{h_1^2}, & i_1 = 1, \\ \dfrac{-\varphi_{i_1+1,\cdot} + 2\varphi_{i_1,\cdot} - \varphi_{i_1-1,\cdot}}{h_1^2}, & 2 \leqslant i_1 \leqslant M_1 - 1, \\ \dfrac{2\varphi_{i_1,\cdot} - \varphi_{i_1-1,\cdot}}{h_1^2}, & i_1 = M_1, \end{cases}$$

i_1 and M_1 are the index and the number of points along the axis x_1. $\Lambda_2, \Lambda_3, \ldots, \Lambda_N$ are defined in the same way. Note, that all operators $\Lambda_j > 0$.

For the approximate solution of problem (65.4)–(65.5) in time we use the following scheme (from now on the index h is omitted)

$$B\frac{\varphi^{n+1}-2\varphi^n+\varphi^{n-1}}{\tau^2}+\Lambda\varphi^n=f^n, \tag{65.6}$$

$$B=\prod_{j=1}^{N}(E+\tfrac{1}{2}\tau^2\Lambda_j). \tag{65.7}$$

Equation (65.6) approximates (65.4) with order $O(\tau^2)$. Equation (65.6) can be written in the form

$$\frac{\varphi^{n+1}-2\varphi^n+\varphi^{n-1}}{\tau^2}+B^{-1}\Lambda\varphi^n=B^{-1}f^n. \tag{65.8}$$

A Fourier spectrum analysis of the components of the solution vector (65.8) leads to the following necessary stability condition of scheme (65.6)–(65.7)

$$\tau\leqslant 2\sqrt{\beta(B^{-1}\Lambda)},$$

where $\beta(B^{-1}\Lambda)$ is the upper bound of the operator $B^{-1}\Lambda$ spectrum. Thus, the problem of choosing the parameter τ satisfying the stability condition is reduced to computing the maximal eigenvalue of the problem

$$\Lambda u=\lambda Bu,$$

in the assumption that all eigenvalues of $B^{-1}\Lambda$ are positive. This problem is solved, for example, with the help of the Lusternik iterative process (see MARCHUK [1964a]). The realization of difference scheme (65.6) can be written as:

$$\begin{aligned}
&(E+\tfrac{1}{2}\tau^2\Lambda_1)\xi^{n+1/N}=-\Lambda\varphi^n+f^n,\\
&(E+\tfrac{1}{2}\tau^2\Lambda_2)\xi^{n+2/N}=\xi^{n+1/N},\\
&\quad\vdots\\
&(E+\tfrac{1}{2}\tau^2\Lambda_N)\xi^{n+1}=\xi^{n+(N-1)/N},\\
&\varphi^{n+1}=2\varphi^n-\varphi^{n-1}+\tau^2\xi^{n+1}.
\end{aligned} \tag{65.9}$$

This problem is solved consecutively for $n=2, 3, \ldots$ using the initial data (65.3)

66. Approximate factorization schemes for equations of oscillations

We consider the following problem for the two-dimensional homogeneous equation of oscillations

$$\frac{\partial^2\varphi}{\partial t^2}-a^2\left(\frac{\partial^2\varphi}{\partial x^2}+\frac{\partial^2\varphi}{\partial y^2}\right)=0, \quad -\infty<x,y<\infty, \tag{66.1}$$

$$\varphi=p \text{ and } \partial\varphi/\partial t=q \quad \text{for } t=0, \qquad a^2=\text{const}. \tag{66.2}$$

We use a scheme of second-order accuracy for an approximate solution of the problem

$$\frac{\varphi^{n+1}-2\varphi^n+\varphi^{n-1}}{\tau^2}+\Lambda\frac{\varphi^{n+1}+\varphi^{n-1}}{2}=0,$$

$$\Lambda=\Lambda_1+\Lambda_2, \tag{66.3}$$

$$\Lambda_1\varphi=-a^2\frac{\varphi_{i+1,j}-2\varphi_{i,j}+\varphi_{i-1,j}}{h_1^2}, \qquad \Lambda_2\varphi=-a^2\frac{\varphi_{i,j+1}-2\varphi_{i,j}+\varphi_{i,j-1}}{h_2^2}.$$

We rewrite (66.3) in the form

$$(E+\tfrac{1}{2}\tau^2\Lambda)\frac{\varphi^{n+1}+\varphi^{n-1}}{2}=\varphi^n \tag{66.4}$$

and approximately factorize the operator on the left-hand side of (66.4):

$$E+\tfrac{1}{2}\tau^2\Lambda\simeq(E+\tfrac{1}{2}\tau^2\Lambda_1)(E+\tfrac{1}{2}\tau^2\Lambda_2).$$

Finally, we replace scheme (66.3) by the following factorized one:

$$(E+\tfrac{1}{2}\tau^2\Lambda_1)(E+\tfrac{1}{2}\tau^2\Lambda_2)\frac{\varphi^{n+1}+\varphi^{n-1}}{2}=\varphi^n. \tag{66.5}$$

Scheme (66.5) is stable and has second-order accuracy. Scheme (66.5) is realized as follows

$$(E+\tfrac{1}{2}\tau^2\Lambda_1)\varphi^{n+1/2}=\varphi^n,$$

$$(E+\tfrac{1}{2}\tau^2\Lambda_2)\frac{\varphi^{n+1}+\varphi^{n-1}}{2}=\varphi^{n+1/2}, \quad n=1,2,\ldots, \tag{66.6}$$

$$\varphi^0=p^h, \qquad \varphi^1=p^h+\tau q^h.$$

Now we consider the scheme of approximate factorization of the operator on the upper level for solving the multidimensional equation of oscillations (65.1) where

$$A=-\sum_{\alpha=1}^{N}\frac{\partial^2\varphi}{\partial x_\alpha^2}, \qquad f=0,$$

$$\Omega=\{x: -\infty<x<\infty\}, \quad x=(x_1,x_2,\ldots,x_N)$$

(KONOVALOV [1962]). To solve it we use an implicit difference scheme,

$$\frac{(\varphi^{n+1}-2\varphi^n+\varphi^{n-1})}{\tau^2+\Lambda\varphi^{n+1}}=0,$$

and rewrite it in the form

$$\left(E+\tau^2\sum_{i=1}^{N}\Lambda_i\right)\varphi^{n+1}=2\varphi^n-\varphi^{n-1}. \tag{66.7}$$

Using the relationship

$$\left(E+\tau^2\sum_{i=1}^{N}\Lambda_i\right)\varphi^{n+1}=\prod_{i=1}^{N}(E+\tau^2\Lambda_i)\varphi^{n+1}+O(\tau^2),$$

we can replace scheme (66.7) by the following factorized one

$$\prod_{i=1}^{N}(E+\tau^2\Lambda_i)\varphi^{n+1}=2\varphi^n-\varphi^{n-1}. \tag{66.8}$$

The realization of scheme (66.8) has the form:

$$\begin{aligned}
&\frac{\varphi^{n+1/N}-2\varphi^n+\varphi^{n-1}}{\tau^2}+\Lambda_1\varphi^{n+1/N}=0,\\
&\frac{\varphi^{n+i/N}-\varphi^{n+(i-1)/N}}{\tau^2}+\Lambda_i\varphi^{n+i/N}=0,\quad i=2,\ldots,N,\\
&\varphi^0=p^h,\qquad \varphi^1=p^h+\tau q^h.
\end{aligned} \tag{66.9}$$

A scheme similar to the DOUGLAS–RACHFORD [1956] scheme suggested for solving a multidimensional heat conduction equation can be used for solving a multidimensional equation of oscillations. It has the form

$$\begin{aligned}
&\frac{\varphi^{n+1/N}-2\varphi^n+\varphi^{n-1}}{\tau^2}+\Lambda_1\varphi^{n+1/N}+\sum_{i=2}^{N}\Lambda_i\varphi^n=0,\\
&\frac{\varphi^{n+i/N}-\varphi^{n+(i-1)/N}}{\tau^2}+\Lambda_i(\varphi^{n+i/N}-\varphi^n)=0,\\
&\quad i=2,\ldots,N,\qquad n=1,2,\ldots,\\
&\varphi^0=p^h,\qquad \varphi^1=p^h+\tau q^h.
\end{aligned} \tag{66.10}$$

Schemes (66.9) and (66.10) approximate the differential problem (65.1)–(65.3) for the N-dimensional heat conduction equation. The approximations are consistent for $\tau/h_i=\text{const}$, $i=1,\ldots,N$. A Fourier analysis shows that schemes (66.9) and (66.10) are unconditionally stable independent of the value $\tau/h_i=\text{const}$.

Note also that, in the case $N=2$, scheme (66.10) can be used for a more general equation as well (KONOVALOV [1962])

$$\begin{aligned}
&\frac{\partial^2\varphi}{\partial t^2}=a_{11}\frac{\partial^2\varphi}{\partial x_1^2}+2a_{12}\frac{\partial^2\varphi}{\partial x_1\partial x_2}+a_{22}\frac{\partial^2\varphi}{\partial x_2^2},\\
&\quad a_{ij}=\text{const},\qquad a_{11}>0,\qquad a_{11}a_{22}-a_{12}^2>0.
\end{aligned}$$

It has the form

$$\begin{aligned}
&\frac{\varphi^{n+1/2}+\varphi^{n-1}-2\varphi^n}{\tau^2}+a_{11}\Lambda_{11}\varphi^{n+1/2}+2a_{12}\Lambda_{12}\varphi^n+a_{22}\Lambda_{22}\varphi^n=0.\\
&\frac{\varphi^n-\varphi^{n+1/2}}{\tau^2}+a_{22}\Lambda_{22}(\varphi^{n+1}-\varphi^n)=0.
\end{aligned} \tag{66.11}$$

67. Local one-dimensional schemes for the multidimensional hyperbolic equations

Consider the following problem for an N-dimensional hyperbolic equation with variable coefficients

$$\frac{\partial^2\varphi}{\partial t^2}+\sum_{i=1}^{N} A_i\varphi=f(x,t) \quad \text{in } \Omega_T,$$
$$A_i\varphi=-\frac{\partial}{\partial x_i}\left(k_i(x,t)\frac{\partial\varphi}{\partial x_i}\right), \qquad \varphi=\mu(x,t) \quad \text{on } \partial\Omega, \tag{67.1}$$

$$\varphi=p \text{ and } \partial\varphi/\partial t=q \quad \text{in } \Omega \qquad \text{for } t=0. \tag{67.2}$$

Here

$$k_i(x,t)\geqslant\delta>0, \quad x=(x_1,x_2,\ldots,x_N),$$

Ω is an arbitrary N-dimensional domain with the boundary $\partial\Omega$,

$$\bar{\Omega}=\Omega+\partial\Omega, \qquad \Omega_T=\Omega\times(0<t\leqslant T],$$
$$\bar{\Omega}_T=\bar{\Omega}\times[0\leqslant t\leqslant T].$$

It is assumed that this problem has a unique solution $\varphi=\varphi(x,t)$ which is continuous in $\bar{\Omega}_T$ and has all required derivatives (i.e, is sufficiently smooth).

Two assumptions are used with respect to the domain $\bar{\Omega}$:

(1) A section of domain Ω by any straight line parallel to the axis Ox_i may consist only of a finite number of intervals.

(2) It is possible to construct in the domain $\bar{\Omega}$ a bounded grid $\bar{\Omega}^h$ with steps h_i, $i=1,2,\ldots,N$.

In order to locally construct a one-dimensional scheme for solving (67.1)–(67.3) we do as follows (SAMARSKII [1964b]). We approximate with step τ/N the operators

$$\Phi_i\varphi=\frac{1}{N}\frac{\partial^2\varphi}{\partial t^2}-(A_i\varphi+f_i), \qquad i=1,2,\ldots,N$$

successively, where f_i satisfies the condition

$$\sum_{i=1}^{N} f_i=f.$$

In order to approximate the derivative $\partial^2\varphi/\partial t^2$ with step τ/N we use for $N=2$ the expression,

$$\varphi_{\bar{t}_i t_i}=\frac{\varphi_i-2\varphi_{i-1}+\check{\varphi}_i}{\tau^2}\sim\frac{1}{4}\left(\frac{\partial^2\varphi}{\partial t^2}\right)^{n+(i-1)/2}, \quad i=1,2, \tag{67.3a}$$

where

$$\varphi_i=\varphi^{n+i/2}, \qquad \check{\varphi}_i=\varphi^{n-1+i/2},$$
$$\varphi_0=\check{\varphi}=\varphi^{n-1}, \qquad \varphi_2=\varphi^n,$$

and for $N=3$,

$$\varphi_{\bar{t}_i t_i}=\frac{\varphi_i-\varphi_{i-1}-\varphi_{i-2}+\check{\varphi}_i}{\tau^2}\sim\frac{2}{g}\frac{\partial^2\varphi}{\partial t^2},\quad i=1,\ 2,\ 3, \tag{67.3b}$$

where

$$\varphi_{-1}=\check{\varphi}_2=\varphi^{n-1+2/3},\qquad \varphi_{-2}=\check{\varphi}_1=\varphi^{n-2/3}.$$

To approximate $A_i\varphi+f_i$ on the grid Ω^h we use the one-dimensional difference operator of second-order accuracy,

$$\Lambda_i\varphi^h+f_i^h.$$

The coefficients of the operator A_i and f_i^h are fixed at the moment

$$t_i'=0.5\ (t_{n+i/N}+t_{n-1+i/N})=t_{n+i/N-1/2}=t_n+(i/N-\tfrac{1}{2})\tau,$$

so that

$$\Lambda_i=\Lambda_i(t_i'),\qquad \varphi_i=\varphi_i(x,\ t_i').$$

Local one-dimensional schemes for hyperbolic equations have the form (from now on h is omitted)

$$\begin{aligned}&\varphi_{\bar{t}_i t_i}=-\sigma_N\Lambda_i(\varphi_i+\check{\varphi}_i)+2\sigma_N\varphi_i,\\ &\quad i=1,\ldots,N,\quad N=2,\ 3,\end{aligned} \tag{67.4}$$

where

$$\sigma_N=\begin{cases}\frac{1}{4} & \text{for } N=2,\\ \frac{1}{3} & \text{for } N=3.\end{cases}$$

If $N=2$ we obtain a three-level additive scheme and if $N=3$ a four-layer one. Here lies the difference from the parabolic equations, where the form of local one-dimensional schemes does not depend on the number of dimensions N. Equation (67.4) can be written in the form

$$(E+\sigma_N\tau^2\Lambda_i)(\varphi_i+\check{\varphi}_i)=\begin{cases}2(\varphi_{i-1}+\sigma_2\tau^2\varphi_i) & \text{for } N=2,\\ \varphi_{i-1}+\varphi_{i-2}+2\sigma_3\tau^2 f_i^h & \text{for } N=3.\end{cases} \tag{67.5}$$

Finding φ_i is reduced to solving a three-point equation

$$(E+\sigma_N\tau^2\Lambda_i)\varphi_i=F_i \tag{67.6}$$

along circles parallel to the axis Ox_i by the sweep method using a boundary condition

$$\varphi_i=\mu(x,\ t_{n+i/N})\quad \text{for } x\in\partial\Omega_h^i. \tag{67.7}$$

If the operator $A_i\varphi$ has junior terms then in using the sweep method the requirement that the step sizes are sufficiently small on the grid Ω^h steps arises. To get rid of these restrictions on the grid Ω^h junior terms should be taken on intermediate layers. The first of the initial conditions, $\varphi(x,0)=p(x)$, is approximated

exactly

$$\varphi^h(x, 0)=p(x). \tag{67.8}$$

To compute the intermediate values

$$\varphi^{1/2}=\varphi(x, \tfrac{1}{2}\tau) \quad \text{for } N=2$$

and

$$\varphi^{1/3}=\varphi(x, \tfrac{1}{3}\tau), \quad \varphi^{2/3}=\varphi(x, \tfrac{2}{3}\tau) \qquad \text{for } N=3$$

we use the following equations: for $N=2$,

$$\begin{aligned} &(E+\tfrac{1}{4}\tau^2\Lambda_1)\varphi^{1/2}=F_1, \\ &F_1=p+\tfrac{1}{2}\tau q-\tfrac{1}{4}\tau^2\Lambda_1 p+\tau^2(f_{1/2}-\tfrac{1}{8}(-\Lambda\varphi+f))_{t=0}; \end{aligned} \tag{67.9}$$

for $N=3$,

$$\begin{aligned} &(E+\tfrac{1}{4}\tau^2\Lambda_1)\varphi^{1/3}=F_1, \\ &(E+\tfrac{1}{4}\tau^2\Lambda_2)(\varphi^{2/3}+p)=2\varphi^{1/3}+F_2, \\ &F_1=p+\tfrac{1}{3}\tau q-\tau^2\Lambda_1 p+q^2(\tfrac{2}{3}f_1-\tfrac{1}{6}(-\Lambda\varphi+f))_{t=0}, \\ &F_2=\tau^2(\tfrac{2}{3}f_2-\tfrac{1}{9}(-\Lambda\varphi+f))_{t=0}. \end{aligned} \tag{67.10}$$

The local one-dimensional schemes (67.4) with the corresponding approximations of the boundary and initial conditions converge with rate $O(\tau+|h|^2)$ if certain smoothness conditions for $\varphi(x, t)$ and $f(x, t)$ hold (SAMARSKII [1964b]).

68. The splitting scheme for multidimensional hyperbolic systems of equations

The splitting method for multidimensional hyperbolic systems was first suggested by BAGRINOVSKY and GODUNOV [1957]. These authors considered only the explicit approximation for which the splitting scheme has few advantages as compared to conventional explicit schemes.

ANUCHINA and YANENKO [1959] suggested an implicit splitting scheme for a multidimensional hyperbolic system. We demonstrate the idea of this scheme for the case of a two-dimensional system of acoustics equations.

We consider the following system of equations

$$\begin{aligned} &\frac{\partial u_1}{\partial t}-a^2\frac{\partial v}{\partial x_1}=0, \qquad \frac{\partial u_2}{\partial t}-a^2\frac{\partial v}{\partial x_2}=0, \\ &\frac{\partial v}{\partial t}-\left(\frac{\partial u_1}{\partial x_1}+\frac{\partial u_2}{\partial x_2}\right)=0. \end{aligned} \tag{68.1}$$

In order to solve (68.1) we use a weighted splitting scheme along the coordinates

x_1 and x_2. On the first stage of splitting we have:

$$\frac{u_1^{n+1/2}-u_1^n}{\tau}-a^2\frac{\Delta_{-1}}{h_1}(\alpha v^{n+1/2}+\beta v^n)=0,$$

$$\frac{u_2^{n+1/2}-u_2^n}{\tau}=0, \tag{68.2}$$

$$\frac{v^{n+1/2}-v^n}{\tau}-\frac{\Delta_1}{h_1}(\alpha u_1^{n+1/2}+\beta u_1^n),$$

and on the second stage we have

$$\frac{u_1^{n+1}-u_1^{n+1/2}}{\tau}=0,$$

$$\frac{u_2^{n+1}-u_2^{n+1/2}}{\tau}-a^2\frac{\Delta_{-2}}{h_2}(\alpha v^{n+1}+\beta v^{n+1/2})=0. \tag{68.3}$$

$$\frac{v^{n+1}-v^{n+1/2}}{\tau}-\frac{\Delta_2}{h_2}(\alpha u_2^{n+1}+\beta u_2^{n+1/2})=0,$$

$$\alpha+\beta=1,$$

where

$$\Delta_i\varphi=\frac{\varphi_{1,\ldots,i+1,\ldots}-\varphi_{1,\ldots,i,\ldots}}{h_i}, \qquad \Delta_{-i}\varphi=\frac{\varphi_{1,\ldots,i,\ldots}-\varphi_{1,\ldots,i-1,\ldots}}{h_i}.$$

Scheme (68.2)–(68.3) is absolutely stable for $\alpha\geqslant\frac{1}{2}$ and approximates the original system (68.1) with first-order accuracy in τ. For $\alpha=\beta=\frac{1}{2}$ the approximation accuracy is $O(\tau^2+h^2)$ (YANENKO [1967], p. 63).

Scheme (68.2)–(68.3) is realized in each stage of splitting by simple three-point sweeps.

The approximate factorization method (66.5), used for solving the two-dimensional wave equation, is generalized for the case of N-dimensional hyperbolic systems of second order (DYAKONOV [1971d]).

We assume that it is necessary to solve the following problem

$$\partial^2\varphi/\partial t^2+A\varphi=f \quad \text{in } \Omega\times\Omega_t, \tag{68.4}$$

and

$$\varphi=p \text{ and } \partial\varphi/\partial t=q \quad \text{in } \Omega \qquad \text{for } t=0,$$
$$\varphi=0 \quad \text{on } \partial\Omega. \tag{68.5}$$

Here

$$\varphi=(\varphi_1(x), \varphi_2(x), \ldots, \varphi_N(x)),$$

$$A\varphi=A_1\varphi+A_2\varphi,$$

$$A_1\varphi\equiv-\sum_{r,l=1}^{d}\frac{\partial}{\partial x_r}\left(a_{r,l}\frac{\partial\varphi}{\partial x_l}\right)+\sum_{r=1}^{d}b_r\frac{\partial}{\partial x_r}\varphi+C\varphi,$$

$$A_2\varphi\equiv a_0(x, D\varphi),\quad D\varphi\equiv\left(\varphi_1, \varphi_2, \ldots, \varphi_N, \frac{\partial\varphi_1}{\partial x_1}, \frac{\partial\varphi_1}{\partial x_2}, \ldots, \frac{\partial\varphi_N}{\partial x_d}\right),$$

$a_{r,l}$, b_r and C are $(N\times N)$ square matrices depending on x, f, p and q; $a_0(x, 0)=0$.

Let Λ_1^n, Λ_2^n and $\Lambda^n=\Lambda_1^n+\Lambda_2^n$ be the difference analogues of operators A_1, A_2 and A at time $t_n=n\tau$ on a rectangular grid. To solve (68.4)–(68.5) we use the following difference scheme (DYAKONOV [1965a, 1967b, 1971d])

$$\prod_{r=1}^{d}(E-\tau^2\theta\Lambda_r)\frac{\varphi_i^{n+1}-\varphi_i^{n-1}}{\tau^2}-\frac{2(\varphi_i^n-\varphi_i^{n-1})}{\tau^2}$$

$$+\Lambda^n(\varphi^n)_i+2\theta\sum_{r=1}^{d}\Lambda_r(\varphi^n-\varphi^{n-1})_i=f_i^n$$

$$\text{for } x_i\in\Omega^h, \tag{68.6}$$

$$\varphi_i^0=p_i,\qquad \varphi_i^1=p_i^1=p_i+\tau q_i+\tfrac{1}{2}\tau^2\left[\frac{\partial^2\varphi}{\partial t^2}\bigg|_{t=0}\right],$$

$$\varphi_i^n=0\quad\text{for } x_i\in\partial\Omega^h,$$

where $\partial^2\varphi/\partial t^2$ for $t=0$ is defined by (68.4)–(68.5). Scheme (68.6) approximates (68.4)–(68.5) with the error $O(\tau^2)$ in time. If the conditions

$$\Lambda_1^{(n)}=(\Lambda_1^{(n)})^*,$$

$$|((\Lambda_1^{(n)}-\Lambda_1^{(n-1)})u, v)|\leqslant\tau\sigma_1\|u\|_{R_1}\|v\|_{R_1},$$

$$\mu_0R_1\leqslant\tfrac{1}{2}\Lambda_1^{(n)}\leqslant\mu_1R_1,\quad \mu_0\equiv\delta_{1/2}>0,$$

$$|(\Lambda_2^{(n)}(u), v)|\leqslant\sigma_2\|u\|_{R_1}^2+\sigma_3\|v\|_{R_1}^2,$$

$$\theta\geqslant\tfrac{1}{2}(\mu_0+\mu_1),\qquad \tau\leqslant\tau_*<(4\sigma_3)^{-1},$$

$$R_1u_i=-\sum_{r=1}^{d}\Lambda_ru_i,$$

$$R_\alpha u_i=(-1)^d\sum_{r_1<r_2<\cdots<r_\alpha}\Lambda_{r_1}\Lambda_{r_2}\cdots\Lambda_{r_\alpha}u_i,\quad \sigma_i=\text{const}>0,$$

hold, then for the solution of (68.6) the following a priori estimate is valid:

$$\|\varphi\|_H^2 \leqslant \text{const}\left\{\left\|\frac{\partial\varphi}{\partial t}\right\|^2 + \|\varphi^0\|_{R_1}^2 + \|\varphi^1\|_{R_1}^2 + \|\varphi_1^1\|_{R_1^{-1}}^2 + \|f_1^2\|_{R_1^{-1}}^2 + \tau\sum_{n=2}^{\hat{k}-1}\left\|\frac{\partial}{\partial t}f_1^n\right\|_{R_1^{-1}}^2 + \tau\sum_{n=1}^{\hat{k}}\sum_{\alpha=2}^{d}\tau^{2\alpha-1}\|f_\alpha^n\|_{R_\alpha^{-1}}^2\right\}.$$

Applied to multidimensional hyperbolic systems the splitting method allows to obtain majorant approximations. Let

$$\frac{\partial\varphi}{\partial t} + \sum_{i=1}^{N} A_i\frac{\partial\varphi}{\partial x_i} + B\varphi = 0 \tag{68.7}$$

be a linear hyperbolic system which is symmetric according to Friedrichs. Here $\varphi = \{\varphi_1, \ldots \varphi_N\}$ is a vector function and A_i and B are symmetric $(N \times N)$ matrices. YANENKO [1964a] showed the possibility of constructing the majorant schemes by the splitting method. Let us show a possible way of constructing majorant approximations for equations of type (68.7) (see YANENKO [1967]). If the A_i are positive matrices, then, using the approximation

$$\frac{\varphi^{n+1}-\varphi^n}{\tau} + \sum_{i=1}^{N} A_i\frac{E-T_{-i}}{h_i}\varphi^n + B\varphi^n = 0, \tag{68.8}$$

we obtain the scheme

$$\varphi^{n+1} = C_0\varphi^n + \sum_{i=1}^{N} C_{-i}T_{-i}\varphi^n, \tag{68.9}$$

where

$$C_0 = E - \sum_{i=1}^{N}(\tau/h_i)A_i - \tau B,$$

$$C_{-1} = (\tau/h_1)A_1, \qquad C_{-2} = (\tau/h_2)A_2, \quad \ldots, \quad C_{-N} = (\tau/h_N)A_N,$$

$$T_{-i}f(x_1, \ldots, x_N) = f(x_1, \ldots, x_i - h_i, \ldots, x_N).$$

If τ/h_i is sufficiently small, then matrix C_0 is positive, the matrices C_{-i} are identically positive.

If the A_i are negative matrices, then $E - T_{-i}$ should be replaced by $T_i - E$ in (68.8). If the A_i are the matrices with an arbitrary sign then the following representation is possible

$$A_i = A_i^1 + A_i^2,$$

where the A_i^1 are nonnegative and the A_i^2 are nonpositive. In this case the scheme has the form:

$$\varphi^{n+1} = C_0\varphi^n + \sum_{i=1}^{N} C_{-i}T_{-i}\varphi^n + \sum_{i=1}^{N} C_iT_i\varphi^n,$$

where

$$C_0 = E - \sum_{i=1}^{N} (\tau/h_i)\, A_i^1 + \sum_{i=1}^{N} (\tau/h_i) A_i^2 - \tau B,$$

$$C_{-i} = (\tau/h_i) A_i^1, \quad C_i = -(\tau/h_i) A_i^2, \qquad i = 1, \ldots, N.$$

Matrices C_{-i} and C_i are always positive and matrix C_0 is positive if τ/h_i is sufficiently small.

The splitting method allows to construct easily realizable positive schemes. We consider an explicit majorant splitting scheme

$$\frac{\varphi^{n+s/N} - \varphi^{n+(s-1)/N}}{\tau} + \left[A_s^1 \frac{E - T_{-s}}{h_s} + A_s^2 \frac{T_s - E}{h_s} + \frac{B}{N} \right] \varphi^{n+(s-1)/N} = 0,$$

$$s = 1, \ldots, N.$$

Scheme (68.10) is reduced to

$$\varphi^{n+s/N} = C_s \varphi^{n+(s-1)/N},$$

where

$$C_s = C_{0s} + C_{-s} T_{-s} + C_s T_s,$$

$$C_{0s} = E - (\tau/h_s) A_s^1 + (\tau/h_s) A_s^2 - \tau B/N,$$

$$C_{-s} = (\tau/h_s) A_s^1, \qquad C_s = -(\tau/h_s) A_s^2,$$

$$s = 1, \ldots, N.$$

Operators C_{-s} and C_s are always positive and operators C_{0s} are positive if τ/h_s is small.

Hence, the scheme

$$\varphi^{n+1} = C_m C_{m-1} \cdots C_1 \varphi^n$$

is positive.

Majorant schemes with low accuracy are, as a rule, maximally stable. They often lead to convergence in the space C^1 and are easily realizable.

The splitting method for multidimensional hyperbolic systems has first been suggested by BAGRINOVSKI and GODUNOV [1957].

The general concepts of splitting schemes (their approximation and convergence) and their use for solving hyperbolic equations are given by MARCHUK [1970, 1980], DYAKONOV [1963b], SAMARSKII [1946b], LEES [1960c], KONOVALOV [1962], YANENKO [1967], ANUCHINA and YANENKO [1959] and WANG [1984].

Numerical splitting algorithms developed for applied problems (nonlinear problems included) described by hyperbolic equations are considered by BELOTSERKOVSKY [1984], SCOTT and ROGER [1983] and LAVAL [1983].

CHAPTER XVII

Integro-Differential Transport Equations

69. Statement of the problem and the scheme of incomplete splitting

Let Ω be a convex limited domain in $\mathbb{R}^3$ with piecewise smooth boundary $\partial\Omega$ and variable point $x=(x_1, x_2, x_3)$. Let ω be a sphere of unit radius in $\mathbb{R}^3$ with the center at the origin, $s=(s_1, s_2, s_3)$ is the unit vector $(\Sigma_{i=1}^3 s_i^2=1)$, taking values on ω. $S_i, i=1,2,3$, can be represented by

$$s_1=\sin\theta\cdot\cos\psi, \qquad s_2=\sin\theta\cdot\sin\psi, \qquad s_3=\cos\theta,$$
$$0\leqslant\theta\leqslant\pi, \quad 0\leqslant\psi\leqslant 2\pi.$$

We denote the integral over ω by

$$S\varphi=\int_\omega \varphi\,ds\equiv\int_0^\pi \sin\theta\,d\theta\int_0^{2\pi}\varphi\,d\psi$$

and the time variable by $t\in\Omega_t=(0,T)$, $T<\infty$.

The integro-differential transport equation for particles in a medium with isotropic scattering has the form:

$$\frac{1}{v_c}\frac{\partial\varphi}{\partial t}+(s,\nabla)\varphi+\sigma\varphi=\frac{\sigma_s}{4\pi}S\varphi+f(s,x,t), \tag{69.1}$$

where $v_c=\text{const}>0$ is the velocity of particles,

$$(s,\nabla)\varphi=\sum_{i=1}^3 s_i\,\partial\varphi/\partial x_i,$$

$f(s,x,t)\equiv f(\theta,\psi,x,t)$ is the function of sources, $\sigma=\sigma(x)$, $\sigma_s=\sigma_s(x)$ and the restrictions

$$0<\sigma_0\leqslant\sigma(x)\leqslant\sigma_1<\infty, \qquad 0\leqslant\sigma_s(x)\leqslant\sigma_{s_1},$$
$$0<\sigma_{c0}\leqslant\sigma_c(x)=\sigma(x)-\sigma_s(x), \tag{69.2}$$

for constant $\sigma_0, \sigma_1, \sigma_{c0}, \sigma_{s_1}$ are assumed to hold.

For (69.1) we will take boundary and initial conditions of the type

$$\varphi(s, x, t)=0$$
$$\text{for } (s, n)<0, \quad x\in\partial\Omega, \quad t\in\Omega_t, \quad s\in\omega, \tag{69.3}$$
$$\varphi(s, x, 0)=g(s, x) \quad \text{for } x\in\Omega, \quad s\in\omega, \tag{69.4}$$

where $n=(n_1, n_2, n_3)$ is the unit vector of the external normal to $\partial\Omega$, $(s, n)=\Sigma_{i=1}^{3} s_i n_i$, and $g(s, x)$ is the function of the particles' initial distribution.

If the spatial domain in which the process of particles' transport is considered is the slab where $x_s\equiv z\in\Omega=(0, H)$, $H<\infty$, $-\infty<x_1, x_2<\infty$ and the properties of the medium and those of f and g do not depend on x_1, x_2 and ψ, then problems (69.1), (69.3)–(69.4) take the form

$$\frac{1}{v_c}\frac{\partial\varphi}{\partial t}+\mu\frac{\partial\varphi}{\partial z}+\sigma(z)\varphi(\mu, z, t)$$
$$=\tfrac{1}{2}\sigma_s(z)\int_{-1}^{1}\varphi(\mu', z, t)\,d\mu'+f,$$
$$\varphi(\mu, 0, t)=0 \quad \text{for } \mu>0,$$
$$\varphi(\mu, H, t)=0 \quad \text{for } \mu<0, \tag{69.5}$$
$$\varphi(\mu, z, 0)=g(\mu, z),$$

where $\mu=\cos\theta\in[-1, 1]$.

We will consider the splitting schemes on examples of the above-stated problems. Note that the schemes are simple to write for problems with anisotropic scattering (when $S\varphi=\int_\Omega\theta(s, s')\varphi(s', x, t)\,ds'$) and also for the more complex problems of the transport theory (in particular, in curvilinear geometries).

For the approximation of the problems, let us introduce grids on all variables, denote the grid step in t by Δt and define difference approximations for the operators in (69.1): S_h, $A_1=\sigma_h-(1/4\pi)\sigma_{sh}S_h$ and $A_2=(s, \nabla)_h$ are difference approximations of the operators S, $\sigma-(\sigma_s/4\pi)S$ and (s, ∇) respectively. Here σ_h, $\sigma_{s,h}$ are diagonal matrices of corresponding dimension. We assume that the approximating operators S_h, A_1 and A_2 are constructed so that the relationships

$$(A_1\varphi_h, \varphi_h)\geqslant\gamma(\varphi_h, \varphi_h), \quad \gamma=\text{const}>0,$$
$$(A_2\varphi_h, \varphi_h)\geqslant 0, \qquad S_h e=e \tag{69.6}$$

hold, where $e=(1, 1, \ldots, 1)$ and the operators act in the Hilbert space $\Phi\equiv H_n$ with the inner product $(\cdot, \cdot)$. Now we formulate the *scheme of incomplete splitting* for the solution of these problems (MARCHUK and YANENKO [1964]):

$$\frac{\varphi^{n+1/2}-\varphi^n}{\tau}+A_1(\alpha\varphi^{n+1/2}+\beta\varphi^n)=f_h^n, \tag{69.7a}$$

$$\frac{\varphi^{n+1}-\varphi^{n+1/2}}{\tau}+A_2(\alpha\varphi^{n+1}+\beta\varphi^{n+1/2})=0, \tag{69.7b}$$

where $\alpha\geqslant 0$, $\beta\geqslant 0$, $\alpha+\beta=1$, $\tau=v_c\Delta t$.

For $\alpha=\beta=\frac{1}{2}$ it has the form

$$\begin{aligned}&\frac{\varphi^{n+1/2}-\varphi^{n}}{\tau}+A_1\frac{(\varphi^{n+1/2}+\varphi^{n})}{2}=f_h^n,\\ &\frac{\varphi^{n+1}-\varphi^{n+1/2}}{\tau}+A_2\frac{(\varphi^{h+1}+\varphi^{n+1/2})}{2}=0.\end{aligned} \tag{69.8}$$

Scheme (69.8) is absolutely stable, but since the operators A_1 and A_2 do not commute it has only a first-order approximation in τ.

Let us describe the realization of scheme (69.7) which is valid for the scheme (69.8) as well. Assuming vector φ^n to be known, let us apply operator S_h to (69.7a). As a result we obtain the following equation for functions $\varphi_0^{n+1/2}=S_h\varphi^{n+1/2}$, $\varphi_0^n=S_h\varphi^n$:

$$\frac{\varphi_0^{n+1/2}-\varphi_0^n}{\tau}+\sigma_{ch}(\alpha\varphi_0^{n+1/2}+\beta\varphi_0^n)=S_hf_h^n,$$

where $\sigma_{ch}=\sigma_h-\sigma_{sh}$. From this result we obtain

$$\varphi_0^{n+1/2}=(E+\alpha\tau\sigma_{ch})^{-1}[(E-\beta\tau\sigma_{ch})\varphi_0^n+S_hf_h^n].$$

(Note that here $(E+\alpha\tau\sigma_{ch})^{-1}$ and $(E-\beta\tau\sigma_{ch})$ are diagonal matrices and hence, that the scheme is easily realizable at this stage.) Then we rewrite (69.7a) in the form:

$$\frac{\varphi^{n+1/2}-\varphi^n}{\tau}+\sigma_h(\alpha\varphi^{n+1/2}+\beta\varphi^n)=\frac{\sigma_s h}{4\pi}(\alpha\varphi_0^{n+1/2}+\beta\varphi_0^n)+f_h.$$

Since $\varphi_0^{n+1/2}$, φ_0^n and φ^n are already known, we invert the diagonal matrix and find

$$\varphi^{n+1/2}=(E+\alpha\tau\sigma_h)^{-1}\left[\frac{\sigma_{sh}}{4\pi}(\alpha\varphi_0^{n-1/2}+\beta\varphi_0^n)+f_h^n-(E-\beta\tau\sigma_h)\varphi^n\right].$$

Realization of (69.7b) depends on the approximation chosen for the operator (s,∇). So does realization of the boundary condition (69.3).

Assuming $\alpha=\beta=\frac{1}{2}$ in (69.7) and writing the *two-cycle scheme of incomplete splitting* in the form

$$\begin{aligned}&(E+\tfrac{1}{2}\tau A_1)\varphi^{n-1/2}=(E-\tfrac{1}{2}\tau A_1)\varphi^{n-1},\\ &(E+\tfrac{1}{2}\tau A_2)(\varphi^n-\tau f^n)=(E-\tfrac{1}{2}\tau A_2)\varphi^{n-1/2},\\ &(E+\tfrac{1}{2}\tau A_2)\varphi^{n+1/2}=(E-\tfrac{1}{2}\tau A_2)(\varphi^n+\tau f^n),\\ &(E+\tfrac{1}{2}\tau A_1)\varphi^{n+1}=(E-\tfrac{1}{2}\tau A_1)\varphi^{n+1/2},\\ &\quad n=1,2,\ldots,\end{aligned} \tag{69.9}$$

where $f^n=f_h|_{t=t^n}$, we obtain an absolutely stable scheme with a second-order approximation in τ (on smooth solutions!) on the interval $t_{n-1}\leqslant t\leqslant t_{n-1}$. Equations (69.9) are solved in the same way as in the case of scheme (69.7).

70. The scheme of complete splitting

We represent operator A_2 as a sum $A_2 = \Sigma_{i=1}^{3} A_{2,i}$, where $A_{2,i}$ is an approximation of the one-dimensional operator $s_i \partial/\partial x_i$. By splitting operator A_2 of (69.7b) we obtain the *scheme of complete splitting* (MARCHUK and YANENKO [1964])

$$\frac{\varphi^{n+1/4} - \varphi^n}{\tau} + A_1(\alpha\varphi^{n+1/4} + \beta\varphi^n) = f_n, \tag{70.1a}$$

$$\frac{\varphi^{n+(i+1)/4} - \varphi^{n+i/4}}{\tau} + A_{2,i}(\alpha\varphi^{n+(i+1)/4} + \beta\varphi^{n+i/4}) = 0, \tag{70.1b}$$

$$i = 1, 2, \ldots .$$

This scheme is absolutely stable but it has only first-order accuracy in τ (since operators A_1 and $A_{2,i}$ do not commute). Equation (70.1a) is realized like in the scheme of incomplete splitting.

We now consider the solution of equations (70.1b). Let Ω be a parallelepiped. Then these equations are easily solvable if the operators $s_i \partial/\partial x_i$ are approximated by schemes of the running count. Realization of the boundary conditions is obvious in this case. In the case of an arbitrary convex domain Ω we can do as follows. Let $\tilde{\Omega}$ be a parallelepiped including Ω. We extend the definitions of σ_s, g and f by zeros in $\tilde{\Omega} \setminus \Omega$ and we consider $\sigma(x)$ in $\tilde{\Omega} \setminus \Omega$, an arbitrary positive constant. Hence, we obtain new functions $\tilde{\sigma}$, $\tilde{\sigma}_s$, $\tilde{f}$ and $\tilde{g}$ and can consider the problem

$$\begin{aligned} &\frac{1}{v_c}\frac{\partial\tilde{\varphi}}{\partial t} + (s, \nabla)\tilde{\varphi} + \tilde{\sigma}\tilde{\varphi} = \frac{\tilde{\sigma}_s}{4\pi} S\tilde{\varphi} + \tilde{f}, \\ &\quad x \in \tilde{\Omega}, \quad t \in \Omega_t, \quad s \in \omega, \\ &\tilde{\varphi} = 0 \quad \text{for } x \in \partial\tilde{\Omega}, \quad (s, \tilde{n}) < 0, \quad s \in \omega, \\ &\tilde{\varphi} = \tilde{g} \quad \text{for } t = 0. \end{aligned} \tag{70.2}$$

The solution of this problem will coincide with that of the initial problem in Ω. Solving (70.2) by means of the splitting schemes considered above we will also obtain the approximate solution of the initial problem.

71. The scheme of approximate factorization of the operator

The scheme has the form

$$(E + \tau A_1)(E + \tau A_2)(\varphi^{n+1} - \varphi^n) = -2\tau(A\varphi^n - f^n), \tag{71.1}$$

where

$$f^n = f_h|_{t = t_{n+1/2}}, \qquad A = A_1 + A_2.$$

It approximates the initial problem (69.1), (69.3)–(69.4) or (69.5) with second-order accuracy in τ and is absolutely stable. It can be realized through the solution of the

following equations (like the scheme of incomplete splitting)

$$\begin{aligned}&(E+\tau A_1)\varphi^{n+1/3}=2\tau F_n,\\&(E+\tau A_2)\varphi^{n+2/3}=\varphi^{n+1/3},\\&\varphi^{n+1}=\varphi^n+\varphi^{n+2/3},\end{aligned}\tag{71.2}$$

where

$$F_n=-(A\varphi^n-f^n),\quad n=0,\ 1,\ 2,\ldots .$$

We consider scheme (71.1) in more detail by applying it to problem (69.5) for concrete difference approximations of the operators (see SMELOV [1978]). To this end divide the segment $[-1, 1]$ in $2M$ equal intervals and regard the center of each one as the grid point along the angular variable. Let us number these grid points as follows:

$$\mu_{-M}<\mu_{-M+1}<\cdots<\mu_{-1}<\mu_1<\cdots<\mu_{M-1}<\mu_M.$$

Hence $\mu_j=-\mu_{-j}$. On the segment $[0, H]$ we introduce an arbitrary system of grid points

$$0=z_0<z_1<\cdots<z_N=H$$

choosing them in such a way that points of discontinuity in the equations' coefficients coincide with some of these grid points. We assume that the indices k, j of any grid function a_{k_j} will denote the following correspondence to the grid points:

$$\begin{aligned}&(k,j)\to(z_k,\mu_j)\qquad \text{for } j>0,\\&(k,j)\to(z_{k-1},\mu_j)\quad \text{for } j<0.\end{aligned}$$

We define difference approximations A_1 and A_2 as follows:

$$\begin{aligned}&(A_1\varphi)_{kj}=\sigma_k\varphi_{k,j}-\tfrac{1}{2}\Delta\mu\sigma_{s,k}\sum_{m=-M}^{M}\varphi_{k,m},\\&(A_2\varphi)_{kj}=\begin{cases}\mu_j(\varphi_{k,j}-\varphi_{k-1,j})/h_k, & j>0,\\ \mu_j(\varphi_{k+1,j}-\varphi_{k,j})/h_k, & j<0,\end{cases}\end{aligned}\tag{71.3}$$

where

$$\begin{aligned}&k=1,\ldots,N,\qquad j=\pm1,\ldots,\pm M,\\&h_k=z_k=z_k-z_{k-1},\\&\Delta\mu=1/M,\\&\sigma_k=\sigma(z_{k-1/2}),\qquad \sigma_{s,k}=\sigma_s(z_{k-1/2}),\\&z_{k-1/2}=\tfrac{1}{2}(z_k+z_{k-1}).\end{aligned}$$

We assume that in (71.3)

$$\varphi_{0,j}=\varphi_{N+1,j}=0$$

(i.e., the boundary conditions of the problem are satisfied). With such difference

approximations of the transport equation operators system (7.12) can be written in component form as follows:

$$F^n_{kj} = -(A_1\varphi^n)_{kj} - (A_2\varphi^n)_{kj} + f^n_{kj},$$

$$F^n_k = \Delta\mu \sum_{m=-M}^{M} F^n_{km},$$

$$\varphi^{n+1/3}_{k,j} = \frac{2\tau}{1+\tau\sigma_k}\left[\frac{\sigma_{sk}\tau}{2(1+\sigma_{ck}\tau)}F^n_k + F^n_{kj}\right], \tag{71.4}$$

$$\varphi^{n+2/3}_{k,j} = \frac{1}{1+|\mu_j|\tau/h_x}\left(\frac{|\mu_j|\cdot\tau}{h_k}\varphi^{n+2/3}_{k-|j|/j,j} + \varphi^{n+1/3}_{k,j}\right),$$

$$\varphi^{n+1}_{k,j} = \varphi^n_{k,j} + \varphi^{n+2/3}_{k,j},$$

$$k = 1, 2, \ldots, N, \qquad j = \pm 1, \ldots, \pm M,$$

where

$$f^n_{kj} = f(z_{k-1/2}, \mu_j, t_{n+1/2}), \quad \sigma_{ck} = \sigma_k - \sigma_{sk},$$
$$\varphi^{n+2/1}_{0,j} = \varphi^{n+2/3}_{N+1,j} = 0.$$

Scheme (71.4) has an approximation of order $O(\tau^2 + h + \Delta\mu^2)$, $h = \max_k h_k$, on smooth solutions and is absolutely stable. It also preserves the major properties of the operators of the original problem.

72. The method of integral identities and the splitting method

Let us now construct the scheme for a problem in a slab on the basis of the method of integral identities and the two-cycle splitting method (MARCHUK [1980], p. 464). We denote by φ^+ the solution of the problem for $\mu > 0$, and by φ^- the solution for $\mu < 0$. Then the transport equation can be written as a system of two equations for $\mu > 0$:

$$\frac{1}{v_c}\frac{\partial\varphi^+}{\partial t} + \mu\frac{\partial\varphi^+}{\partial z} + \sigma\varphi^+ = \tfrac{1}{2}\sigma_s\int_0^1(\varphi^+ + \varphi^-)\,d\mu' + f^+, \tag{72.1}$$

$$\frac{1}{v_c}\frac{\partial\varphi^-}{\partial t} - \mu\frac{\partial\varphi^-}{\partial z} + \sigma\varphi^- = \tfrac{1}{2}\sigma_s\int_0^1(\varphi^+ + \varphi^-)\,d\mu' + f^-,$$

with boundary conditions

$$\varphi^+(0, \mu, t) = 0, \qquad \varphi^-(H, \mu, t) = 0. \tag{72.2}$$

We sum and subtract the two equations (72.1). As a result we obtain two new

equations,

$$\frac{1}{v_c}\frac{\partial u}{\partial t}+\mu\frac{\partial u}{\partial z}+\sigma u=\sigma_s\int_0^1 u\,d\mu'+g, \tag{72.3a}$$

$$\frac{1}{v_c}\frac{\partial u}{\partial t}+\mu\frac{\partial u}{\partial z}+\sigma v=r, \tag{72.3b}$$

where

$$u=\tfrac{1}{2}(\varphi^+ +\varphi^-), \quad v=\tfrac{1}{2}(\varphi^+ -\varphi^-),$$
$$g=\tfrac{1}{2}(f^+ +f^-), \quad r=\tfrac{1}{2}(f^+ -f^-).$$

The boundary conditions (72.2) transform into

$$\begin{aligned} &u+v=0 \quad \text{for } z=0,\\ &u-v=0 \quad \text{for } z=H, \end{aligned} \tag{72.4}$$

and the initial conditions are

$$u=u^0, \qquad v=v^0 \qquad \text{for } t=0. \tag{72.5}$$

We write (72.3)–(72.5) in operator form. To this end we introduce vector functions w, w^0 and F and operator A:

$$w=\begin{pmatrix} u\\ v\end{pmatrix}, \qquad w^0=\begin{pmatrix} u^0\\ v^0\end{pmatrix}, \qquad F=\begin{pmatrix} g\\ r\end{pmatrix},$$

$$A=\begin{bmatrix} \sigma-\sigma_s\displaystyle\int_0^1 d\mu' & \mu\dfrac{\partial}{\partial z}\\[2ex] \mu\dfrac{\partial}{\partial z} & \sigma \end{bmatrix} \tag{72.6}$$

We consider them in a Hilbert space $L_2(D)$, $D=(0,H)\times(0,1)$ of vector functions with the inner product

$$(a,b)=\sum_{i=1}^{2}\int_0^1 d\mu\int_0^M a^i b^i\,dz,$$

where a^i and b^i are components of vector functions a and b. Now we can write the problem under consideration as follows:

$$\begin{aligned} &\frac{1}{v_c}\frac{\partial w}{\partial t}+Aw=F \quad \text{in } D\times(0,T),\\ &w=w^0 \quad \text{in } D \qquad \text{for } t=0, \end{aligned} \tag{72.7}$$

where function w for any t belongs to the domain of operator A (in particular, it satisfies conditions (72.4)).

We construct a difference approximation of the problem along the variable z. To this end we introduce two systems of grid points: the main system $\{z_k\}_{k=0}^{N}$, $z_0=0$, $z_N=H$ and the auxiliary system $\{z_{k+1/2}\}_{k=0}^{N-1}$. The points of these two systems alternate, i.e. $z_{k-1/2}<z_k<z_{k+1/2}$. We integrate (72.3a) in z with limits $(z_0, z_{1/2})$, $(z_{k-1/2}, z_{k+1/2})$, $k=1,\ldots,N-1$ and $(z_{N-1/2}, z_N)$ and (72.3b) with limits (z_{k-1}, z_k), $k=1,\ldots,N$. In the elementary integrals we approximately replace the function u by its value at the corresponding grid point z_k and the function v by its value at $z_{k-1/2}$, take boundary conditions into account and omit the approximation errors. Then we obtain the following system of equations:

$$\begin{aligned} &\frac{1}{v_c}\frac{\partial\varphi}{\partial t}+A\varphi=F, \quad t\in\Omega_t, \\ &\varphi=\varphi^{(0)}, \end{aligned} \tag{72.8}$$

where

$$\varphi=(u_0, v_{1/2}, u_1, \ldots, u_{N-1}, v_{N-1/2}, u_N),$$

$$F=(g_0, r_{1/2}, g_1, \ldots, g_{N-1}, r_{N-1/2}, g_N),$$

$$\varphi^{(0)}=(u_0^{(0)}, v_{1/2}^{(0)}, u_1^{(0)}, \ldots, u_{N-1}^{(0)}, v_{N-1/2}^{(0)}, u_N^{(0)}),$$

$$\sigma_k=\frac{1}{\Delta z_k}\int_{z_{k-1/2}}^{z_{k+1/2}}\sigma\,dz, \qquad \sigma_{sk}=\frac{1}{\Delta z_k}\int_{z_{k-1/2}}^{z_{k+1/2}}\sigma_s\,dz,$$

$$\sigma_{k+1/2}=\frac{1}{\Delta z_{k+1/2}}\int_{z_k}^{z_{k+1}}\sigma\,dz, \qquad g_k=\frac{1}{\Delta z_k}\int_{z_{k-1/2}}^{z_{k+1/2}}g\,dz,$$

$$r_{k+1/2}=\frac{1}{\Delta z_{k+1/2}}\int_{z_k}^{z_{k+1}}r\,dz,$$

$$\Delta z_0=z_{1/2}-z_0,$$

$$\Delta z_k=z_{k+1/2}-z_{k-1/2}, \quad k=1,\ldots,N-1,$$

$$\Delta z_N=z_N-z_{N-1/2},$$

$$\Delta z_{k-1/2}=z_k-z_{k-1}, \quad k=1,\ldots,N,$$

$$h=\max(\Delta z_0, \Delta z_N, \Delta z_k, \Delta z_{k-1/2}),$$

$$A=A_2+A_1,$$

$$A_1=\operatorname{diag}\left(\sigma_{i/2}-\sigma_{s,1/2}\int_0^1\gamma_{1/2}\,d\mu'\right),$$

$$A_2 = \begin{bmatrix} \mu/\Delta z_0 & \mu/\Delta z_0 & 0 & 0 \\ -\mu/\Delta z_{1/2} & 0 & \mu/\Delta z_{1/2} & 0 \\ 0 & -\mu/\Delta z_{N-1/2} & 0 & \mu/\Delta z_{N-1/2} \\ 0 & 0 & -\mu/\Delta z_N & \mu/\Delta z_N \end{bmatrix},$$

$$\gamma_{i/2} = \begin{cases} 1 & \text{for } i \text{ even}, \\ 0 & \text{for } i \text{ odd}. \end{cases}$$

Let $H_h(0,2N)$ be the Hilbert space of vector functions $a=(a_0, a_{1/2}, a_1, \ldots, a_{N-1}, a_{N-1/2}, a_N)$ with inner product and norm

$$(a,b) = \sum_{i=0}^{2N} \int_0^1 \Delta z_{i/2} a_{i/2} b_{i/2} \, d\mu, \qquad \|a\| = (a,a)^{1/2}, \quad a, b \in H_h(0,2N).$$

In obtaining (72.8), the approximation errors in case of a uniform grid in the first and the last equations are first-order in h and in other cases $O(h^2)$ on smooth solutions. Basing on this fact we can prove that the estimate (MARCHUK [1980], p. 472)

$$\max_t \|\varphi_T - \varphi\|(t) \leqslant o(h^2 \ln^{1/2}(1/h))$$

holds, where φ_T is the vector of the values of the exact solution of the problem:

$$\varphi_T = (u(z_0, \mu, t), v(z_{1/2}, \mu, t), \ldots, v(z_{N-1/2}, \mu, t), u(z_N, \mu, t)).$$

Note, also, that operators A_1 and A_2 satisfy the relationships

$$(A_2 a, a) \geqslant 0,$$

$$(A_1 a, a) \geqslant \gamma \|a\|^2, \quad \gamma = \text{const} > 0.$$

This allows us to formulate the algorithm for solving (72.8) using the two-cycle splitting method:

$$\begin{aligned} &(E + \tfrac{1}{2}\tau A_2)\varphi^{j-2/3} = (E - \tfrac{1}{2}\tau A_2)\varphi^{j-1}, \\ &(E + \tfrac{1}{2}\tau A_1)\varphi^{j-1/3} = (E - \tfrac{1}{2}\tau A_1)\varphi^{j-2/3}, \\ &\varphi^{j+1/3} = \varphi^{j-1/3} + 2\tau F^j, \\ &(E + \tfrac{1}{2}\tau A_1)\varphi^{j+2/3} = (E - \tfrac{1}{2}\tau A_1)\varphi^{j+1/3}, \\ &(E + \tfrac{1}{2}\tau A_2)\varphi^{j+1} = (E - \tfrac{1}{2}\tau A_2)\varphi^{j+2/3}, \end{aligned} \tag{72.9}$$

where F^j is the vector with components

$$F_i^j = \frac{1}{(t_{j+1} - t_{j-1})} \int_{t_{j-1}}^{t_{j+1}} F_i \, dt, \quad t_j = j\Delta t, \qquad \tau = v_c \Delta t.$$

It follows from the properties of A_1 and A_2 that, if the solution is sufficiently smooth,

scheme (72.9) approximates (72.8) with order $O(\Delta t^2)$ and is stable:

$$\max_j \| \varphi^j \| \leqslant C(\| \varphi^{(0)} \| + \max_j \| F^j \|).$$

It follows from this inequality and approximation that

$$\max_j \| \varphi(t_j) - \varphi^j \| \leqslant O(\tau^2).$$

Hence

$$\max_j \| \varphi_T(t_j) - \varphi^j \| \leqslant o(\tau^2 + h^2 |\ln^{1/2}(1/h)|). \tag{72.10}$$

We now consider the solution of system of equations (72.9). Since matrix $(I + \frac{1}{2}\tau A_2)$ is three-diagonal, the first and the last equations of (72.9) are easily solvable for any fixed μ. Since matrix operator $(I + \frac{1}{2}\tau A_1)$ is diagonal, the solution of the second equation of (72.9) is reduced to the following computations:

$$\varphi_i^{j-1/3} = \frac{1}{1 + \frac{1}{2}\tau\sigma_i} \left[(1 - \tfrac{1}{2}\tau\sigma_i)\varphi_i^{j-2/3} + \frac{\tau\sigma_{s,i}}{1 + \frac{1}{2}\tau\sigma_{ci}} \int_0^1 \varphi_i^{j-2/3} \, d\mu \right], \tag{72.11}$$

$$\varphi_{i-1/2}^{j-1/3} = \frac{1 - \frac{1}{2}\tau\sigma_{i-1/2}}{1 + \frac{1}{2}\tau\sigma_{i-1/2}} \varphi_{i-1/2}^{j-2/3}, \quad i = 1, \ldots, N.$$

Analogous formulae can be written for solving the fourth equation of (72.9).

Let us now formulate the difference approximation of equations in μ. To this end we divide the interval $0 \leqslant \mu \leqslant 1$ in partial intervals $\Delta\mu_l$ by grid points so that the best approximation of integrals in (72.11) on the given class of solutions is ensured:

$$\int_0^1 \psi(\mu) \, d\mu \cong \sum_{l=1}^m s_l \Psi_l, \quad \Psi_l = \Psi(\mu_l),$$

where s_l is the weight of the chosen square formula. Replacing integrals in (72.9) with this square formula and considering the system when $\mu = \mu_l$, $l = 1, \ldots, m$, we come to the system of linear algebraic equations approximating the original problem

$$\begin{aligned}
&(E + \tfrac{1}{2}\tau A_{2,l})\varphi_l^{j-2/3} = (E - \tfrac{1}{2}\tau A_{2,l})\varphi_l^{j-1}, \\
&(E + \tfrac{1}{2}\tau A_{1,l})\varphi_l^{j-1/3} = (E - \tfrac{1}{2}\tau A_{1,l})\varphi_l^{j-2/3}, \\
&\varphi_l^{j+1/3} = \varphi_l^{j-1/3} + 2\tau F_l^j, \\
&(E + \tfrac{1}{2}\tau A_{1,l})\varphi_l^{j+2/3} = (E - \tfrac{1}{2}\tau A_{1,l})\varphi_l^{j+1/3}, \\
&(E + \tfrac{1}{2}\tau A_{2,l})\varphi_l^{j+1} = (E - \tfrac{1}{2}\tau A_{2,l})\varphi_l^{j+2/3}, \\
&\varphi_l^0 = \varphi_l^{(0)},
\end{aligned} \tag{72.12}$$

where $\varphi_l^{j-1}, \ldots, \varphi_l^{j+1}$ are vectors of dimension $(2N+1)$, $l = 1, \ldots, m$, $\varphi_l^{(0)} = \varphi^{(0)}(\mu_l)$,

$A_{2,l}=A_2(\mu_l)$, and operator $A_{1,l}$ acts according to the formula

$$(A_{1,l}\varphi_l^j)_{i/2}=\sigma_{i/2}\varphi_{l,i/2}^j-\sigma_{s,i/2}\sum_{k=1}^{m}s_k\varphi_{k,i/2}^j,$$

$$i=0,\ldots,2N.$$

As was noted above, the first and the last equations of (72.12) are easily solvable, for example, by the sweep method. The second equation is solved by the formulae

$$\varphi_{l,i}^{j-1/3}=\frac{1}{1+\frac{1}{2}\tau\sigma_i}\left[(1-\tfrac{1}{2}\tau\sigma_i)\varphi_{l,i}^{j-2/3}+\frac{\tau\sigma_{s,i}}{1+\frac{1}{2}\tau\sigma_{ci}}\sum_{k=1}^{m}s_k\varphi_k^{j-2/3}\right],$$

$$i=0,\ 1,\ldots,N,$$

$$\varphi_{l,i-1/2}^{j-1/3}=\frac{1-\frac{1}{2}\tau\sigma_{i-1/2}}{1+\frac{1}{2}\tau\sigma_{i-1/2}}\varphi_{l_j,i-1/2}^{j-2/3},\quad i=1,\ldots,N. \tag{72.13}$$

The fourth equation is solved using these same formulae with $j+1$ instead of j.

Thus, the algorithm for the numerical solution of the nonstationary transport equation is completely determined. Hence, we come to the absolutely stable scheme with accuracy of order $O(\tau^2+h^2\ln^{1/2}(1/h))$ on smooth solutions. The order of accuracy in μ depends on the choice of the square formula. (Note that $\mu=0$ is also allowed as a grid point in the scheme under consideration.)

We did not impose restrictions on the grid steps yet. Note, however, that if the sweep method is used in the scheme for solving three-diagonal systems, then for its stability it is sufficient to require that $\tau<\max(\Delta z_{i/2})$. For solving practical problems this condition imposes some restrictions on the choice of the time step.

By using the method of integral identities, and the splitting and quadrature methods in the same way we can construct difference approximations for problems in the theory of particles' transport that depend on several spatial variables (see MARCHUK and AGOSHKOV [1981], pp. 405–410).

73. The numerical scheme for the nonstationary transport equation in (x,y) geometry

We consider the nonstationary problem for the transport equation in (x,y) geometry:

$$\frac{1}{v_c}\frac{\partial\varphi}{\partial t}+\mu\frac{\partial\varphi}{\partial x}+\eta\frac{\partial\varphi}{\partial y}+\sigma\varphi=\frac{\sigma_s}{2\pi}\int_0^1 d\gamma'\int_0^{2\pi}\varphi\,d\psi'+f, \tag{73.1}$$

$$\varphi=0\quad\text{for } (x,y)\in\partial\Omega,\quad(\mu n_x+\eta n_y)<0, \tag{73.2}$$

$$\varphi=g\quad\text{for } t=0, \tag{73.3}$$

where

$$\Omega=\{x, y: 0<x<a, 0<y<b\},$$
$$\varphi=\varphi(\gamma, \psi, x, y, t), \qquad \sigma=\sigma(x, y), \qquad \sigma_s=\sigma_s(x, y),$$
$$s=(\mu, \eta)\equiv(s_1, s_2),$$
$$s_1=\sqrt{1-\gamma^2}\cos\psi, \qquad s_2=\sqrt{1-\gamma^2}\sin\psi,$$
$$\gamma=s_s=\cos\theta, \qquad 0\leqslant\psi\leqslant 2\pi, \qquad 0\leqslant\theta\leqslant\tfrac{1}{2}\pi,$$

$n=(n_x, n_y)$ is the external normal vector to Ω and the restrictions (69.2) are assumed to hold.

We attempt in this section to consider the algorithm for the numerical solution of problem (73.1)–(73.3), which is absolutely stable and excludes any iterative processes. We will construct such an algorithm using the following methods of applied mathematics: the method of integral identities (to approximate the problem in spatial variables); the splitting method (to approximate the problem in time variable); and the quadrature method (to approximate in angular variables). We assume further that the solution of the problem (73.1)–(73.3) is sufficiently smooth and focus on the algorithmic side of the numerical solution of the problem. We will give the major statements without proof since most of them are discussed in corresponding sections of the book by MARCHUK and AGOSHKOV [1981].

73.1. Approximation in spatial variables

Let us first perform some transformations of the problem. Since its solution not infrequently has some peculiarities where the vector $s=(\mu, \eta)$ approaches directions parallel to the coordinate axes, it is expedient to transform the original problem so that the vector $s=(\mu, \eta)$ gets a value between these directions. Such transformations allow in many cases to increase the approximation accuracy of the problem. Let us perform such transformations for problem (73.1)–(73.3).

We write the integral term in (73.1) in the form

$$\frac{\sigma_s}{2\pi}\int_0^1 d\gamma'\int_0^{2\pi}\varphi\, d\psi'=\sum_{k=1}^{4}\sigma_s S\varphi^{(k)}, \tag{73.4}$$

where

$$S\varphi^{(k)}=\frac{1}{2\pi}\int_0^1 d\gamma'\int_0^{\pi/2}\varphi^{(k)}(\gamma', \psi', x, y, t)\, d\psi',$$
$$\varphi^{(k)}=\varphi(\gamma', \psi'+\tfrac{1}{2}(k-1)\pi, x, y, t),$$

and consider problem (73.1)–(73.3) consecutively on the subdomains $\Omega\times\omega_k$, where

$$\omega_k=\{\gamma, \psi: 0<\gamma<1, \tfrac{1}{2}(k-1)\pi<\psi<\tfrac{1}{2}k\pi, k=1, \ldots 4\}.$$

Then replacing ψ by $\psi+\frac{1}{2}(k-1)\pi$, we obtain the following system of four equations:

$$\frac{1}{v_c}\frac{\partial\varphi^{(k)}}{\partial t}+\mu^{(k)}\frac{\partial\varphi^{(k)}}{\partial x}+\eta^{(k)}\frac{\partial\varphi^{(k)}}{\partial y}+\sigma\varphi^{(k)}=\sigma_s\sum_{k'=1}^{4}S\varphi^{(k')}+f^{(k)} \tag{73.5}$$

$$\varphi^{k}=0 \quad \text{on } \partial\Omega \qquad \text{for } (\mu^{(k)}n_x+\eta^{(k)}n_y)<0, \tag{73.6}$$

$$\varphi^{(k)}=g^{(k)} \quad \text{for } t=0,$$
$$k=1,\ 2,\ 3,\ 4, \tag{73.7}$$

where

$$0<\psi<\tfrac{1}{2}\pi, \quad 0<\gamma<1,$$
$$\mu^{(k)}=\sqrt{1-\gamma^2}\cos(\psi+\tfrac{1}{2}(k-1)\pi),$$
$$\eta^{(k)}=\sqrt{1-\gamma^2}\sin(\psi+\tfrac{1}{2}(k-1)\pi),$$
$$f^{(k)}=f(\gamma,\psi+\tfrac{1}{2}(k-1)\pi,x,y,t),$$
$$g^{(k)}=g(\gamma,\psi+\tfrac{1}{2}(k-1)\pi,x,y).$$

As is easily seen, directions s parallel to the coordinate axes are already excluded from problem (73.5)–(73.7).

We take the inner product in $L_2(\Omega)$ of (73.5) and (73.7) and an arbitrary function $v(x,y)\in\mathring{W}_2^1(\Omega)$ (i.e., having first derivatives from $L_2(\Omega)$ and being equal to zero on $\partial\Omega$). Then after integrating by parts we get the equalities,

$$\left(\frac{1}{v_c}\frac{\partial\varphi^{(k)}}{\partial t},v\right)+\left(\varphi^{k},-\mu^{(k)}\frac{\partial v}{\partial x}-\eta^{(k)}\frac{\partial v}{\partial y}\right)+(\sigma\varphi^{(k)},v)$$
$$=\sum_{k'=1}^{4}S(\sigma_s\varphi^{(k')},v)+(f^{(k)},v), \tag{73.8}$$

$$(\varphi^{(k)},v)=(g^{(k)},v) \quad \text{for } t=0,$$
$$k=1,2,3,4, \tag{73.9}$$

$$(u,v)=\int_{\Omega}uv\,dx\,dy.$$

We write conditions (73.6) in a generalized form

$$\int_{\partial\Omega^{(k)}_{-}}|\mu^{(k)}n_x+\eta^{(k)}n_y|\varphi^{(k)}v^{(k)}\,dx\,dy=0,$$
$$k=1,\ 2,\ 3,\ 4, \tag{73.10}$$

where $v^{(k)}\in W_2^1(\Omega)$ and $v^{(k)}|_{\partial\Omega^{(k)}_{-}}=0$, $\partial\Omega^{(k)}$ is that part of the boundary $\partial\Omega$ for which $(\mu^{(k)}n_x+\eta^{(k)}n_y)>0$ for a given value k. Thus the exact solution of the original problem satisfies equalities (73.8)–(73.10) that are used further for constructing

integral identities and for approximating problems in spatial variables.

We introduce on Ω a uniform grid

$$x_i = ih, \qquad y_j = jh, \qquad a = b, \qquad N = M, \qquad h = a/N$$

(only for the sake of simplicity; the algorithm for a nonuniform grid both in x and in y has been considered by MARCHUK and AGOSHKOV [1981], p. 306) and define the functions,

$$\varphi_i(x) = \frac{1}{\gamma_{x,i}} \cdot \begin{cases} (x - x_{i-1})/h, & x \in (x_{i-1}, x_i), \\ (x_{i+1} - x)/h, & x \in (x_i, x_{i+1}), \\ 0, & x \notin (x_{i-1}, x_{i+1}); \end{cases}$$

$$\varphi_j(y) = \frac{1}{\gamma_{y,j}} \cdot \begin{cases} (y - y_{j-1})/h, & y \in (y_{j-1}, y_j), \\ (y_{j+1} - y)/h, & y \in (y_j, y_{j+1}), \\ 0, & y \notin (y_{j-1}, y_{j+1}); \end{cases}$$

where $\gamma_{x,i}$ and $\gamma_{y,j}$ are normalizing coefficients, which can be further defined from the conditions

$$\int_0^a \varphi_i(x)\,dx = 1, \qquad \int_0^b \varphi_j(y)\,dy = 1; \tag{73.11}$$

in the case of a uniform grid (73.11) gives the following values $\gamma_{x,i} = \gamma_{y,j} = h$.

To construct systems of integral identities we assume in (73.8) and (73.9)

$$v(x, y) = \varphi_i(x)\varphi_j(y), \quad i, j = 1, \ldots, N-1.$$

As a result we come to systems of integral identities the elements of which are arranged in vectors of dimension $(N-1)^2$ and matrices of order $(N-1)^2$ and numbered as follows: for $k=1$ we start from $x=0$, $y=0$; for $k=2$ from $x=a$, $y=0$; for $k=3$ from $x=a$, $y=b$; for $k=4$ from $x=0$, $y=b$. By assuming simple approximations in these systems we come to a problem of the type:

$$\frac{1}{v_c}\frac{\partial \varphi^{(k)}}{\partial t} + (|\mu^{(k)}|(\hat{I} \otimes \hat{A}) + |\eta^{(k)}|(\hat{A} \otimes \hat{I}) + \hat{\sigma}^{(k)})\varphi^{(k)}$$

$$= \hat{\sigma}_s^{(k)} \sum_{k'=1}^{4} S\hat{p}^{(k,k')}\varphi^{(k')} + f^{(k)}, \tag{73.12}$$

$$\varphi^{(k)} = g^{(k)} \quad \text{for } t = 0, \qquad k = 1, 2, 3, 4, \tag{73.13}$$

where

$$G^{(k)}=(\underbrace{G^{(k)}_{1,1}, G^{(k)}_{2,1}, \ldots, G^{(k)}_{N-1,1}}_{j=1}, \ldots, \underbrace{G^{(k)}_{1,M-1}, G^{(k)}_{2,M-1}, \ldots, G^{(k)}_{N-1,M-1}}_{j=M-1}),$$

$$G^{(k)}_{n,m}=\begin{cases}(G^{(k)}, \varphi_n(x)\varphi_m(y)), & k=1,\\ (G^{(k)}, \varphi_{N-m}(x)\varphi_m(y)), & k=2,\\ (G^{(k)}, \varphi_{N-m}(x)\varphi_{M-m}(y)), & k=3,\\ (G^{(k)}, \varphi_n(x)\varphi_{M-m}(y)), & k=4,\end{cases}$$

$G=\varphi,\ g,\ f,$

$$\hat{A}=\frac{1}{h}\underbrace{\begin{bmatrix} 1 & & & 0\\ -1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1\end{bmatrix}}_{N-1}, \qquad \hat{I}=\underbrace{\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix}}_{N-1},$$

$$\hat{A}\otimes\hat{B}=\begin{bmatrix} A_{11}\hat{B} & \cdots & A_{1,N-1}\hat{B}\\ \vdots & & \vdots\\ A_{N-1,1}\hat{B} & \cdots & A_{N-1,N-1}\hat{B}\end{bmatrix},$$

$\hat{\sigma}^{(k)}$ and $\hat{\sigma}^{(k)}_s$ are diagonal matrices of order $(N-1)\times(M-1)$ with elements

$$\sigma^{(k)}_{n,m}=\begin{cases}\sigma_{n,m}, & k=1,\\ \sigma_{N-n,m}, & k=2,\\ \sigma_{N-n,M-m}, & k=3,\\ \sigma_{n,M-m}, & k=4,\end{cases} \qquad \sigma^{(k)}_{s,n,m}=\begin{cases}\sigma_{s,n,m}, & k=1,\\ \sigma_{s,N-n,m}, & k=2,\\ \sigma_{s,N-n,M-m}, & k=3,\\ \sigma_{s,n,M-m}, & k=4,\end{cases}$$

$\sigma_{n,m}=(\sigma, \varphi_n(x)\varphi_m(y)), \qquad \sigma_{s,n,m}=(\sigma_s, \varphi_n(x)\varphi_m(y)),$

$\hat{p}^{(\cdot,\cdot)}=\hat{I}\otimes\hat{I}, \quad k=1, 2, 3, 4,$

$\hat{p}^{(1,2)}=\hat{p}^{(2,1)}=\hat{p}^{(3,4)}=\hat{p}^{(4,3)}=\hat{I}\otimes\hat{p},$

$\hat{p}^{(1,3)}=\hat{p}^{(3,1)}=\hat{p}^{(3,4)}=\hat{p}^{(4,2)}=\hat{p}\otimes\hat{p},$

$\hat{p}^{(1,4)}=\hat{p}^{(4,1)}=\hat{p}^{(2,3)}=\hat{p}^{(3,2)}=\hat{p}\otimes\hat{I},$

$$\hat{p}=\underbrace{\begin{bmatrix}0 & & 1\\ & \iddots & \\ 1 & & 0\end{bmatrix}}_{N-1}.$$

Note that the numbering of vector and matrix elements is chosen in such a way that the values $\{\varphi^{(1)}_{n.m}, \varphi^{(2)}_{N-n,m}, \varphi^{(3)}_{N-n,M-m}, \varphi^{(4)}_{n,M-m}\}$, i.e., the approximations of the

averaged values of the exact solution

$$\{\varphi^{(1)}_{T,n,m}, \varphi^{(2)}_{T,N-n,m}, \varphi^{(3)}_{T,N-n,M-m}, \varphi^{(4)}_{T,n,M-m}\}\varphi^{(k)}_{T,n,m} = (\varphi^{(k)}, \varphi_n(x)\varphi_m(y)),$$

are in correspondence with the grid point (x_n, y_m). Analogous relationships exist for the grid points and elements of the vectors $\{g^{(k)}\}$ and $\{f^{(k)}\}$.

Let H_h be a Hilbert space of vector functions with inner product and norm defined by

$$(u, v)_h = \sum_{k=1}^{4} \sum_{n=1}^{N-1} \sum_{m=1}^{M-1} \int_0^1 d\gamma \int_0^{\pi/2} u^{(k)}_{n,m}(\gamma, \psi) v^{(k)}_{n,m}(\gamma, \psi)\, d\psi,$$

$$\| u \|_h = (u, u)_h^{1/2}.$$

We define in H_h the vector functions

$$G = (G^{(1)}, G^{(2)}, G^{(3)}, G^{(4)})^T, \quad G = \varphi, f, g$$

and the matrix operators

$$\hat{\Lambda}_1 = \operatorname{diag}(\Lambda_{1,kk}),$$

$$\hat{\sigma} = \operatorname{diag}(\sigma_{kk}), \qquad \hat{\sigma}_s = \operatorname{diag}(\sigma_{s,kk}),$$

$$\Lambda_{1,kk} = |\mu^{(k)}| = (\hat{I} \otimes \hat{A}) + |\eta^{(k)}|(\hat{A} \otimes \hat{I}),$$

$$\sigma_{kk} = \hat{\sigma}^{(k)}, \quad \sigma_{s.kk} = \hat{\sigma}^{(k)}_s, \qquad k = 1, 2, 3, 4,$$

$$\hat{\Lambda}_2 = \hat{\sigma} - \hat{\sigma}_s \cdot \begin{bmatrix} \hat{p}^{(1,1)} & \dots & \hat{p}^{(1,4)} \\ \vdots & & \vdots \\ \hat{p}^{(4,1)} & \dots & \hat{p}^{(4,4)} \end{bmatrix} \cdot S.$$

Then problem (73.12)–(73.13) may take the form

$$\frac{1}{v_c} \frac{\partial \varphi}{\partial t} + (\hat{\Lambda}_1 + \hat{\Lambda}_2)\varphi = f \quad \text{in } \Omega_t, \tag{73.14}$$

$$\varphi = g \quad \text{for } t = 0. \tag{73.15}$$

Note the following properties of the operators $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ (see MARCHUK and AGOSHKOV [1981]):

(a) Matrix $\hat{\Lambda}_1$ is a lower triangular matrix with positive elements on the diagonal and nonpositive elements outside the diagonal.

(b) $\hat{\Lambda}_1$ is positive-semidefinite with

$$(\hat{\Lambda}_1 v, v)_h \geq \tfrac{1}{2} dh \sum_{k=1}^{4} \int_0^1 d\gamma \int_0^{\pi/2} d\psi \sum_{n,m=1}^{N-1} (|\mu^{(k)}| + |\eta^{(k)}|)|v_{n,m}|^2, \tag{73.16}$$

where $d \approx \pi^2$,

(c) Operator $\hat{\Lambda}_2$ is symmetric in H_h and positive-definite

$$(\hat{\Lambda}_2 v, v)_h \geqslant \tilde{\sigma}_{c0} \| v \|_h^2, \tag{73.17}$$

where

$$\tilde{\sigma}_{c0} = \min_{n,m} (\sigma - \sigma_s, \varphi_n(x)\varphi_m(y)) \geqslant \sigma_{c0} > 0.$$

These properties guarantee the existence and uniqueness of the solution $\varphi(\gamma, \psi, t)$ of problem (73.14)–(73.15) with

$$\max_{t\in[0,T]} \| \varphi \|_h(t) + \left(\int_0^T \| \varphi \|_h^2 \, dt \right)^{1/2}$$

$$\leqslant C\left(\| g \|_h + \left(\int_0^T \| f \|_h^2 \, dt \right)^{1/2} \right). \tag{73.18}$$

Besides, according to MARCHUK and AGOSHKOV [1981], the following error estimate holds:

$$\max_{t\in[0,T]} \| \varphi_T - \varphi \|_h(t) + \left(\int_0^T \| \varphi_T - \varphi \|_h^2 \, dt \right)^{1/2} \leqslant \text{const} \cdot h, \tag{73.19}$$

where vector $\varphi_T = \varphi_T(\gamma, \psi, t)$ has a structure similar to that of vector $\varphi(\gamma, \psi, t)$ but with elements $(\varphi, \varphi_n(x)\varphi_m(y))$ equal to the averaged values of the exact solution of problem (73.1)–(73.3).

73.2. Approximation in the time variable

The properties of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ allow us to formulate the following algorithm for solving (73.14)–(73.15) based on the two-cycle componentwise splitting method

$$\begin{aligned}
&(E + \tfrac{1}{2}\tau\hat{\Lambda}_1)\varphi^{j-2/3} = (E - \tfrac{1}{2}\tau\hat{\Lambda}_1)\varphi^{j-1}, \\
&(E + \tfrac{1}{2}\tau\hat{\Lambda}_2)\varphi^{j-1/3} = (E - \tfrac{1}{2}\tau\hat{\Lambda}_2)\varphi^{j-2/3}, \\
&\varphi^{j+1/3} = \varphi^{j-1/3} + 2\tau f^j, \\
&(E + \tfrac{1}{2}\tau\hat{\Lambda}_2)\varphi^{j+2/3} = (E - \tfrac{1}{2}\tau\hat{\Lambda}_2)\varphi^{j+1/3}, \\
&(E + \tfrac{1}{2}\tau\hat{\Lambda}_1)\varphi^{j+1} = (E - \tfrac{1}{2}\tau\hat{\Lambda}_1)\varphi^{j+2/3},
\end{aligned} \tag{73.20}$$

where E is the unit matrix with the same dimension as $\hat{\Lambda}_1$,

$$\tau = v_c \Delta t, \quad \Delta t = T/y,$$

$$t_j = j\Delta t, \qquad \varphi^0 = g, \qquad f^j = \int_{t_{j-1}}^{t_{j+1}} f \, dt/(t_{j+1} - t_{j-1}).$$

It also follows from the properties of Λ_1 and Λ_2 that scheme (73.20) is absolutely stable with

$$\max_j \|\varphi^j\|_h \leqslant C(\|g\|_h + \max \|f^j\|_h), \quad j=2k, \quad k=0, 1, 2, \ldots, \quad C=\text{const}. \tag{73.21}$$

When the solutions of the approximated problems are sufficiently smooth, scheme (73.20) has a second-order approximation in τ in the points $t_{zj}, j=1, 2, \ldots, \frac{1}{2}J$.

Consequently,

$$\max \|\varphi_T(t_{2j}) - \varphi^{2j}\|_h \leqslant \text{const}(\tau^2 + h). \tag{73.22}$$

73.3. Approximation in angular variables

We approximate the equations in the angular variables γ and ψ. To that end we introduce an approximation of the integral $s\varphi$ by means of a suitable quadrature formula,

$$S\varphi \cong \sum_{l=1}^{L} \sum_{d=1}^{D} A_{ld}\varphi(\gamma_l, \psi_d),$$

where A_{ld} are weights, (γ_l, ψ_d) are the values at the grid points. By replacing integrals in (73.20) with quadrature formulae and considering equations from (73.20) with $\gamma=\gamma_l$, $l=1, \ldots L$, $\psi=\psi_d$, $d=1, \ldots, D$, we come to a system of linear algebraic equations that approximates the original problem

$$(E+\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j-2/3} = (E-\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j-1}, \tag{73.23}$$

$$(E+\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j-1/3} = (E-\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j-2/3}, \tag{73.24}$$

$$\varphi_{ld}^{j+1/3} = \varphi_{ld}^{j-1/3} + 2\tau f_{ld}^{j}, \tag{73.25}$$

$$(E+\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j+2/3} = (E-\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j+1/3}, \tag{73.26}$$

$$(E+\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j+1} = (E-\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j+2/3}, \tag{73.27}$$

where vectors φ_{ld}^j have the same structure as φ^j in (73.20) (they are regarded as approximations to $\varphi^j(\gamma_l, \psi_d)$), $\varphi_{ld}^0 = g(\gamma_l, \psi_d)$ and the operator $\hat{B}_{2,ld}$ is defined as follows:

$$\hat{\Lambda}_{2,ld}\varphi_{ld}^j = \hat{\gamma}\varphi_{ld}^j - \hat{\sigma}_s \cdot \begin{bmatrix} \hat{p}^{(1,1)} & \ldots & \hat{p}^{(1,4)} \\ \vdots & & \vdots \\ \hat{p}^{(4,1)} & \ldots & \hat{p}^{(4,4)} \end{bmatrix} \cdot \sum_{l=1}^{L} \sum_{d=1}^{D} A_{ld}\varphi_{ld}^j.$$

Thus the algorithm for the numerical solution of the nonstationary transport equation has been defined. As a result we have obtained an absolutely stable scheme of first-order accuracy in h and second-order accuracy in τ on smooth solutions. The order of approximation in the angular variables is defined by the quadrature formula chosen (LEBEDEV [1971, 1976]).

73.4. Numerical realization of the algorithm

Let

$$\xi_{ld}^{j-1}=(E-\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j-1}, \qquad \xi_{ld}^{j-2/3}=(E-\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j-2/3},$$
$$\xi_{ld}^{j+1/3}=(E-\tfrac{1}{2}\tau\hat{\Lambda}_{2,ld})\varphi_{ld}^{j+1/3}, \qquad \xi_{ld}^{j+2/3}=(E-\tfrac{1}{2}\tau\hat{\Lambda}_{1,ld})\varphi_{ld}^{j+2/3}.$$

Then (73.23)–(73.27) can be written in a component form as follows:

(a) Equations (73.23):

$$\left[1+\frac{\tau}{2h}(|\mu_{ld}^{(k)}|+|\eta_{ld}^{(k)}|)\right]\varphi_{l,d,n,m}^{k,j-2/3} - \frac{\tau}{2h}|\mu_{ld}^{(k)}|\,\varphi_{l,d,n-1,m}^{k,j-2/3}-\frac{\tau}{2h}|\eta_{l,d,n,m-1}^{(k)}|=\xi_{l,d,n,m}^{k,j-1}, \tag{73.23'}$$

where

$$\mu_{ld}^{(k)}=\mu^{(k)}(\gamma_l,\psi_d), \qquad \eta_{ld}^{(k)}=\eta^{(k)}(\gamma_l,\psi_d),$$
$$\varphi_{l,d,0,m}^{k,j-2/3}\equiv 0, \qquad \varphi_{l,d,n,0}^{k,j-2/3}\equiv 0,$$
$$\xi_{l,d,n,m}^{k,j-1}=\left[1-\frac{\tau}{2h}(|\mu_{ld}^{(k)}|+|\eta_{ld}^{(k)}|)\right]\varphi_{l,d,n,m}^{k,j-1} + \frac{\tau}{2h}|\mu_{ld}^{(k)}|\varphi_{l,d,n-1,m}^{k,j-1}+\frac{\tau}{2h}|\eta_{ld}^{(k)}|\varphi_{l,d,n,m-1}^{k,j-1}$$
$$\varphi_{l,d,0,m}^{k,j-1}\equiv 0, \qquad \varphi_{l,d,n,0}^{k,j-1}\equiv 0,$$
$$k=1,2,3,4, \qquad n=1,\ldots,N-1, \qquad m=1,\ldots,M-1,$$
$$l=1,\ldots,L, \qquad d=1,\ldots,D.$$

Equations (73.23) are solved according to the formulae

$$\varphi_{l,d,n,m}^{k,j-2/3}=\frac{\xi_{l,d,n,m}^{k,j-1}+\frac{\tau}{2h}|\mu_{ld}^{(k)}|\varphi_{l,d,n-1,m}^{k,j-2/3}+\frac{\tau}{2h}|\eta_{l,d}^{(k)}|\varphi_{l,d,n,m-1}^{k,j-2/3}}{1+\frac{\tau}{2h}(|\mu_{ld}^{(k)}|+|\eta_{ld}^{(k)}|)} \tag{73.28}$$

(b) Equations (73.24):

$$(1+\tfrac{1}{2}\tau\sigma_{n,m}^{(1)})\varphi_{l,d,n,m}^{1,j-1/3} - \tfrac{1}{2}\tau\sigma_{s,n,m}^{(1)}\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}\cdot(\varphi_{l',d',N-n,m}^{2,j-1/3}+\varphi_{l',d',n,m}^{1,j-1/3} + \varphi_{l',d',N-n,M-m}^{3,j-1/3}+\varphi_{l',d',n,M-m}^{4,j-1/3})=\xi_{l,d,n,m}^{1,j-2/3},$$
$$\vdots \tag{73.24'}$$

$$(1+\tfrac{1}{2}\tau\sigma^{(4)}_{n,m})\varphi^{4,j-1/3}_{l,d,n,m}$$

$$-\tfrac{1}{2}\tau\sigma^{(4)}_{s,n,m}\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}\cdot(\varphi^{4,j-1/3}_{l',d',n,m}+\varphi^{1,j-1/3}_{l',d',N-n,m}$$

$$+\varphi^{2,j-1/3}_{l',d',N-n,M-m}+\varphi^{3,j-1/3}_{l',d',n,M-m})=\xi^{4,j-1/3}_{l,d,n,m},$$

$l=1,\ldots,L,\qquad d=1,\ldots,D,$

$n=1,\ldots,N-1,\qquad m=1,\ldots,M-1,$

where

$$\xi^{1,j-2/3}_{l,d,n,m}=(1-\tfrac{1}{2}\tau\sigma^{(1)}_{n,m})\varphi^{1,j-2/3}_{l,d,n,m}$$

$$+\tfrac{1}{2}\tau\sigma^{(1)}_{s,n,m}\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}\cdot(\varphi^{1,j-2/3}_{l',d',n,m}+\varphi^{2,j-2/3}_{l',d',N-n,m}$$

$$+\varphi^{3,j-2/3}_{l',d',N-n,M-m}+\varphi^{4,j-2/3}_{l',d',n,M-m}),$$

$$\vdots \tag{73.29}$$

$$\xi^{4,j-2/3}_{l,d,n,m}=(1-\tfrac{1}{2}\tau\sigma^{(4)}_{n,m})\varphi^{4,j-2/3}_{l,d,n,m}$$

$$+\tfrac{1}{2}\tau\sigma^{(4)}_{s,n,m}\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}\cdot(\varphi^{4,j-2/3}_{l',d',n,m}+\varphi^{1,j-2/3}_{l',d',N-n,m}$$

$$+\varphi^{2,j-2/3}_{l',d',N-n,M-m}+\varphi^{3,j-2/3}_{l',d',n,M-m}).$$

The solution of (73.24′) is reduced to solving a succession of systems of order $4LD$, each of which is defined by equations (73.24′) with fixed n and m. In the case of isotropic dispersion, i.e. in considering problems (73.1)–(73.3) (and also in case of an anisotropic indicatrix of type

$$\theta=\sum_{p=1}^{P}\theta_p\cdot P_p(\mu_0),\quad \mu_0=(s,s'),$$

P is finite, $\{P_p(\mu)\}$ is the Legendre polynomial) equations (73.24) are solvable in an explicit form and their solutions are:

$$\varphi^{1,j-1/3}_{l,d,n,m}=\frac{2}{\tau+\tau\sigma^{(1)}_{n.m}}\Bigg[\xi^{1,j-2/3}_{l,d,n,m}+\frac{\tau\sigma^{(1)}_{s,n,m}}{2+\tau(\sigma^{(1)}_{nm}-\sigma^{(1)}_{s,n,m})}$$

$$\cdot\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}(\xi^{1,j-2/3}_{l',d',n,m}+\xi^{2,j-2/3}_{l',d',N-n,m}+\xi^{3,j-2/3}_{l',d',N-n,M-m}$$

$$+\xi^{4,j-2/3}_{l',d',n,M-m})\Bigg],$$

$$\vdots \tag{73.30}$$

$$\varphi_{l,d,n,m}^{4,j-1/3}=\frac{2}{2+\tau\sigma_{nm}^{(1)}}\Bigg[\xi_{l,d,n,m}^{4,j-2/3}+\frac{\tau\sigma_{s,n,m}^{(4)}}{2+\tau(\sigma_{nm}^{(4)}-\sigma_{s,n,m}^{(4)})}$$

$$\cdot\sum_{l'=1}^{L}\sum_{d'=1}^{D}A_{l'd'}(\xi_{l',d',n,m}^{4,j-2/3}+\xi_{l',d',N-n,m}^{1,j-2/3}+\xi_{l',d',N-n,M-m}^{2,j-2/3}$$

$$+\xi_{l',d',n,M-m}^{3,j-2/3})\Bigg],$$

$$l=1,\ldots,L,\qquad d=1,\ldots,D.$$

(c) Equations (73.25):

$$\varphi_{l,d,nm}^{k,j+1/3}=\varphi_{l,d,nm}^{k,j+1/3}+2\tau f_{l,d,n,m}^{k},$$

$$k=1,2,3,4,\qquad l=1,\ldots,L,\qquad d=1,\ldots,D,\qquad (73.31)$$

$$n=1,\ldots,N-1,\qquad m=1,\ldots,M-1.$$

(d) Equations (73.26) have the same form as (73.24) and are likewise solved. Elements of vector $\varphi_{ld}^{j+2/3}$, $\xi_{ld}^{j+1/3}$ are found correspondingly according to formulae (73.30), or (73.29) if j is replaced by $j+1$.

(e) Equations (73.27) are solved according to

$$\varphi_{l,d,n,m}^{k,j+1}=\frac{\xi_{l,d,n,m}^{k,j+2/3}+\frac{\tau}{2h}|\mu_{ld}^{(k)}|\varphi_{l,d,n-1,m}^{k,j+1}+\frac{\tau}{2h}|\eta_{ld}^{(k)}|\varphi_{l,d,n,m-1}^{k,j+1}}{1+\frac{\tau}{2h}(|\mu_{ld}^{(k)}|+|\eta_{ld}^{(k)}|)},$$

$$k=1,\ldots,4,\qquad l=1,\ldots,L,\qquad d=1,\ldots,D,$$

$$n=1,\ldots,N-1,\qquad m=1,\ldots,M-1,\qquad (73.32)$$

$$\varphi_{l,d,0,m}^{k,j+1}=0,\qquad \varphi_{l,d,n,0}^{k,j+1}=0,$$

where

$$\xi_{l,d,n,m}^{k,j+2/3}=\left[1+\frac{\tau}{2h}(|\mu_{ld}^{(k)}|+|\eta_{ld}^{(k)}|)\right]\varphi_{l,d,n,m}^{k,j+2/3}$$

$$+\frac{\tau}{2h}|\mu_{ld}^{(k)}|\varphi_{l,d,n-1,m}^{k,j+2/3}+\frac{\tau}{2h}|\eta_{ld}^{(k)}|\varphi_{l,d,n,m-1}^{k,j+2/3},$$

$$\varphi_{l,d,0,m}^{k,j+2/3}=0,\qquad \varphi_{l,d,n,0}^{k,j+2/3}=0.$$

Thus, the algorithm of numerical solution for problem (73.1)–(73.3) has been completely defined. Note in conclusion that a scheme with a second-order of approximation in the spatial variables is given in MARCHUK and AGOSHKOV ([1981], p. 317). Further, we remark that, formally, δ-sources in x, y and t are possible in the above algorithm and that the algorithm itself can be generalized to the problem of anisotropic dispersion.

74. Splitting methods as iterative algorithms for stationary transport equations

We consider scheme (71.1) as an iterative algorithm for solving the stationary problem of the transport equation in the slab:

$$\mu\frac{\partial\varphi}{\partial z}+\sigma\varphi=\tfrac{1}{2}\sigma_s\int_{-1}^{1}\varphi(\mu',z)\,d\mu'+f(\mu,z),$$

$$\varphi(\mu,0)=0,\quad \mu>0,\qquad (74.1)$$

$$\varphi(\mu,H)=0,\quad \mu<0,$$

assuming that the parameter τ is positive and depends on the step number of the iterative process. Operators A_1 and A_2 are defined according to (71.3). We have, therefore, the following iterative process

$$(E+\tau_nA_1)(E+\tau_nA_2)(\varphi^{n+1}-\varphi^n)=-2\tau_n(A\varphi^n-f),\qquad (74.2)$$

which is realized according to formulae (71.4). To estimate the rate of convergence of the process we write (74.2) for errors $\varepsilon^n=\varphi-\varphi^n$ as

$$(E+\tau_nA_1)(E+\tau_nA_2)(\varepsilon^{n+1}-\varepsilon^n)=-2\tau_nA\varepsilon^n$$

or

$$(E+\tau_nA_1)(E+\tau_nA_2)\varepsilon^{n+1}=(E-\tau_nA_1)(E-\tau_nA_2)\varepsilon^n$$

or

$$(E+\tau_nA_2)\varepsilon^{n+1}=T_n(E+\tau_nA_2)\varepsilon^n,\qquad (74.3)$$

where

$$T_n=(E+\tau_nA_1)^{-1}(E-\tau_nA_1)\cdot(E-\tau_nA_2)\cdot(E+\tau_nA_2)^{-1}.$$

When we consider equation (74.3) in Euclidean space, we have

$$\|(E+\tau_nA_2)\varepsilon^{n+1}\|\leqslant\|T_n\|\cdot\|(E+\tau_nA_2)\varepsilon^n\|.$$

Since

$$(A_1\varphi,\varphi)\geqslant\sigma_{c0}\|\varphi\|^2,\qquad (A_2\varphi,\varphi)\geqslant 0,$$

we have

$$\|(E-\tau_nA_2)(E+\tau_nA_2)^{-1}\|^2$$

$$=\sup_{\varphi\neq 0}\frac{\|(E-\tau_nA_2)\varphi\|^2}{\|(E+\tau_nA_2)\varphi\|^2}$$

$$=\sup_{\varphi\neq 0}\frac{\|\varphi\|^2-2\tau_n(A_2\varphi,\varphi)+\tau_n^2\|A_2\varphi\|^2}{\|\varphi\|^2+2\tau_n(A_2\varphi,\varphi)+\tau_n^2\|A_2\varphi\|^2}\leqslant 1,$$

$$\|(E+\tau_n A_1)^{-1}(E-\tau_n A_1)\|^2$$
$$=\|(E-\tau_n A_1)(E+\tau_n A_1)^{-1}\|^2$$
$$=\sup_{\varphi\neq 0}\frac{1-2\tau_n(A_1\varphi,\varphi)/\|\varphi\|^2+\tau_n^2\|A_1\varphi\|^2/\|\varphi\|^2}{1+2\tau_n(A_1\varphi,\varphi)/\|\varphi\|^2+\tau_n^2\|A_1\varphi\|^2/\|\varphi\|^2},$$
$$\leqslant\frac{1-2\tau_a\sigma_{c0}+\tau_b^2\|A_1\|^2}{1+2\tau_a\sigma_{c0}+\tau_b^2\|A_1\|^2}$$
$$\leqslant\frac{1-2\tau_a\sigma_{c0}+\tau_b^2\sigma_1^2}{1+2\tau_a\sigma_{c0}+\tau_b^2\sigma_1^2}\equiv\theta^2<1,$$

where $0<\tau_a,\tau_b<\infty$ are the boundaries of parameter τ_n. Hence, $\|\tau_n\|\leqslant\theta<1$ and for (74.2) we obtain the convergence and estimates of the kind

$$\|\varphi^n-\varphi\|\leqslant\|(E+\tau_n A_2)(\varphi^n-\varphi)\|$$
$$\leqslant\theta^n\|(E+\tau_n A_2)(\varphi^0-\varphi)\|\to 0,\quad n\to\infty. \tag{74.4}$$

We now take a parameter τ_n independent of the step size of the iterative process. Then

$$\theta=\theta(\tau)=\frac{1-2\tau\sigma_{c0}+\tau^2\sigma_1^2}{1+2\tau\sigma_{c0}+\tau^2\sigma_1^2}. \tag{74.5}$$

From the condition, $\theta'(\tau)=0$, we find the value τ providing the minimum value of the function $\theta(\tau)$:

$$\tau_{\text{opt}}=1/\sigma_1 \tag{74.6}$$

(note also that $\theta''(\tau_{\text{opt}})\geqslant 0$).

The choice of τ_n can also be optimized from the simplified approximate problem, for example, from the diffusion problem (PENENKO, SULTANGAZIN and BALASH [1969]). Namely, by using this simplification of the problem we have $\tau_{\text{opt}}=1/\sigma_{c1}$ where

$$\sigma_{c1}=\sup_z \text{vrai}\ \sigma_c.$$

Thus, we have considered the iterative process with an (approximately) factorized operator for solving a stationary problem in the slab. It is easy to see that the above conclusions are valid for multidimensional problems. Besides, other splitting schemes (see MARCHUK and YANENKO [1964], MARCHUK [1980], etc.) can be used for solving stationary transport equations.

In conclusion here is a concise review of the literature dealing with different aspects of the splitting methods theory for integro-differential transport equations.

Various splitting schemes for nonstationary transport equations are described by MARCHUK and YANENKO [1964], MARCHUK [1980], MARCHUK and AGOSHKOV [1981] and YANENKO [1967]; concrete ways of approximating operators in all variables are given by MARCHUK [1980] and MARCHUK and AGOSHKOV [1981].

Splitting methods as iterative algorithms for solving stationary transport equations are considered by MARCHUK and YANENKO [1964], SMELOV [1978], MARCHUK and SULTANGAZIN [1965], KOCHERGIN and KUZNETSOV [1969], MARCHUK, PENENKO and SULTANGAZIN [1969]. Iterative splitting schemes for systems of spheric harmonics are discussed by BOYARINTSEV and UZNADZE [1967] and SULTANGAZIN [1979] and the problem of optimizing the parameters of iterative processes in problems of the transport theory are discussed by SMELOV [1978], PENENKO, SULTANGAZIN and BALASH [1969], MARCHUK and SULTANGAZIN [1965].

Some aspects of realizing boundary conditions in splitting schemes for a transport equation are considered by MARCHUK and YANENKO [1964], MARCHUK and AGOSHKOV [1981], YANENKO [1967], KOROBITSINA [1969], MARCHUK, PENENKO and SULTANGAZIN [1969]. Note also that this last work is in fact a survey of a certain period of research into splitting algorithms and their theoretical grounding for solving stationary problems for transport equations.

CHAPTER XVIII

The Splitting Method for Problems of Hydrodynamics

The approximation of equations of hydro-, aero- and gas dynamics leads to systems of nonlinear algebraic equations. The effective solution of these equations is a complex problem, especially when solving multidimensional problems by means of implicit schemes in time. On the other hand, the implicit schemes are particularly convenient for gas dynamics and Navier–Stokes equations since they essentially weaken the restrictions on the value of a time step. In this case, like in many applications, the splitting method is a constructive approach.

This chapter discusses various approaches to construct splitting schemes for equations of hydro- and gas dynamics. For the sake of simplicity we demonstrate these methods for splitting differential systems of equations on the same level as in a weak approximation method (see Section 46).

75. Splitting schemes for Navier–Stokes equations with ε-perturbation of incompressible fluid equations

The artificial compressibility method for solving problems of incompressible fluid currents has been described by YANENKO [1967].

We consider the problem of a symmetric stationary flux with an incompressible fluid obstacle in a two-dimensional plane.

The process is described by

$$\sum_{i=1}^{2} u_i D_i \boldsymbol{u} + (1/\rho)\,\mathrm{grad}\, p = \nu \Delta \boldsymbol{u}, \tag{75.1}$$

$$\mathrm{div}\, \boldsymbol{u} = 0. \tag{75.2}$$

Here u_1 and u_2 are velocity components in directions x_1 and x_2 respectively, ρ is the density, p is the pressure, $\nu = \mu/\rho$ is the kinematic viscosity coefficient, $D_i = \partial/\partial x_i$ and $\mathrm{grad} = (\partial/\partial x_1, \partial/\partial x_2)$. To construct a relaxation process, the nonstationary system obtained from the weak compressibility assumption is formulated in corres-

pondence with (75.1) and (75.2). This system has the form

$$\frac{\partial \boldsymbol{u}^\varepsilon}{\partial t}+\sum_{i=1}^{2} u_i^\varepsilon \mathscr{D}_i \boldsymbol{u}^\varepsilon+(1/\rho)\operatorname{grad} p^\varepsilon=\nu\Delta \boldsymbol{u}, \tag{75.3}$$

$$\varepsilon\frac{\partial p^\varepsilon}{\partial t}+\varepsilon\sum_{i=1}^{\varepsilon} u_i^\varepsilon \mathscr{D}_i p^\varepsilon+p^\varepsilon \operatorname{div} \boldsymbol{u}^\varepsilon=0. \tag{75.4}$$

In (75.3) and (75.4) $\varepsilon>0$ is considered to be sufficiently small.

Approximating the system of equations

$$\rho=1, \qquad t_n=\tau n, \qquad t_{n+1/2}=\tau(n+\tfrac{1}{2}),$$

on the first subinterval $[t_n, t_{n+1/2}]$ we can use a splitting scheme on coordinates x_1 and x_2 for equations (75.3), and (75.4)

$$\begin{aligned}
&\frac{1}{2}\frac{\partial u_1^\varepsilon}{\partial t}+u_1^\varepsilon\frac{\partial u_1^\varepsilon}{\partial x_1}+\frac{\partial p^\varepsilon}{\partial x_1}=\nu\frac{\partial^2 u_1^\varepsilon}{\partial x_1^2},\\
&\frac{1}{2}\frac{\partial u_2^\varepsilon}{\partial t}+u_1^\varepsilon\frac{\partial u_2^\varepsilon}{\partial x_1}=\nu\frac{\partial^2 u_2^\varepsilon}{\partial x_2},\\
&\tfrac{1}{2}\varepsilon\frac{\partial p^\varepsilon}{\partial t}+\varepsilon u_1^\varepsilon\frac{\partial p^\varepsilon}{\partial x_1}+p^\varepsilon\frac{\partial u_1^\varepsilon}{\partial x_1}=0;
\end{aligned} \tag{75.5}$$

and on the subinterval $[t_{n+1/2}, t_{n+1}]$ we can use

$$\begin{aligned}
&\frac{1}{2}\frac{\partial u_1^\varepsilon}{\partial t}+u_2^\varepsilon\frac{\partial u_1^\varepsilon}{\partial x_2}=\nu\frac{\partial^2 u_1^\varepsilon}{\partial x_2^2},\\
&\frac{1}{2}\frac{\partial u_2^\varepsilon}{\partial t}+u_2^\varepsilon\frac{\partial u_2^\varepsilon}{\partial x_2}=\nu\frac{\partial^2 u_2^\varepsilon}{\partial x_2^2},\\
&\tfrac{1}{2}\varepsilon\frac{\partial p^\varepsilon}{\partial t}+\varepsilon u_2^\varepsilon\frac{\partial p^\varepsilon}{\partial x_2}+p^\varepsilon\frac{\partial u_2^\varepsilon}{\partial x_2}=0.
\end{aligned} \tag{75.6}$$

By approximating systems (75.5) and (75.6) by finite differences on a rectangular grid and by implicit schemes in time, we can solve them by three-point sweeps along the grid lines. To solve problem (75.1)–(75.2) we can use a nonstationary system which is different from (75.3)–(75.4). It has the following form

$$\frac{\partial \boldsymbol{u}}{\partial t}+\sum_{i=1}^{2} u_i D_i \boldsymbol{u}+\operatorname{grad} p=\nu\Delta \boldsymbol{u}, \tag{75.7}$$

$$\frac{\partial q}{\partial t}+\operatorname{div} \boldsymbol{u}=0. \tag{75.8}$$

Here

$$q=\tfrac{1}{2}(q_1+q_2), \quad q_i=p+\tfrac{1}{2}u_i^2, \quad i=1,2.$$

The corresponding splitting schemes have the form:

$$\begin{aligned}
&\frac{1}{2}\frac{\partial u_1}{\partial t}+\frac{\partial q_1}{\partial x_1}=\nu\frac{\partial^2 u_1}{\partial x_1^2},\\
&\frac{1}{2}\frac{\partial u_2}{\partial t}+u_1\frac{\partial u_2}{\partial x_1}=\nu\frac{\partial^2 u_2}{\partial x_1^2},\\
&\frac{1}{2}\frac{\partial q_1}{\partial t}+\frac{\partial u_1}{\partial x_1}=0,\\
&\quad \text{on } [t_n, t_{n+1/2}],\\
&\frac{1}{2}\frac{\partial u_1}{\partial t}+u_2\frac{\partial u_1}{\partial x_2}=\nu\frac{\partial^2 u_1}{\partial x_2^2},\\
&\frac{1}{2}\frac{\partial u_2}{\partial t}+\frac{\partial q_2}{\partial x_2}=\nu\frac{\partial^2 u_2}{\partial x_2^2},\\
&\frac{1}{2}\frac{\partial q_2}{\partial t}+\frac{\partial u_2}{\partial x_2}=0,\\
&\quad \text{on } [t_{n+1/2}, t_n].
\end{aligned} \tag{75.9}$$

(75.10)

In difference approximations of systems (75.5), (75.6), (75.9) and (75.10) τ is a relaxation parameter of the iterative process. It is worth noting that boundary conditions for the system (75.1)–(75.2) have intentionally been omitted in this discussion to focus on major aspects of splitting. Splitting boundary conditions requires a separate study in each particular case.

The ε-perturbation introduced in (75.2) can also be used for solving nonstationary equations. TEMAM [1981] uses the following perturbed equations:

$$\frac{\partial \boldsymbol{u}^\varepsilon}{\partial t}+\sum_{i=1}^{n} u_i^\varepsilon \mathscr{D}_i \boldsymbol{u}^\varepsilon+\tfrac{1}{2}(\operatorname{div} \boldsymbol{u}^\varepsilon)+\operatorname{grad} p^\varepsilon=\nu\Delta \boldsymbol{u}^\varepsilon, \tag{75.11}$$

$$\varepsilon\frac{\partial p^\varepsilon}{\partial t}+\operatorname{div} \boldsymbol{u}^\varepsilon=0. \tag{75.12}$$

In this case, it is easier to approximate (75.11) and (75.12) than the original equations, because the condition div $\boldsymbol{u}=0$ is replaced by an evolutionary equation. Here the following problems arise:

- existence and uniqueness of solutions of perturbed equations;
- convergence of $\boldsymbol{u}^\varepsilon$ and p^ε to $\boldsymbol{u}$ and p for $\varepsilon\to 0$;
- discretization of (75.11) and (75.12) and convergence of discrete approximations to a solution of the original problem.

The existence of solutions of perturbed problems and the convergence of exact solutions for $\varepsilon\to 0$ are discussed by TEMAM [1981], where an alternating direction scheme for approximation in time was used and where unconditional stability in certain spaces and convergence of the discrete approximation to a solution of the Navier–Stokes equations when $\varepsilon,\tau,h\to 0$ was proved.

76. Splitting schemes restoring divergence for incompressible fluid equations

A fractional steps scheme requiring no perturbed equations and based on the projection method has been suggested by CHORIN [1968] and further studied by TEMAM [1981]. The classical statement consists of finding a vector function $\boldsymbol{u}$ and a scalar function p such that

$$\frac{\partial \boldsymbol{u}}{\partial t}+\sum_{i=1}^{n} u_i \frac{\partial}{\partial x_i}\boldsymbol{u}+\operatorname{grad} p=\nu\Delta\boldsymbol{u}+f \quad \text{in } \Omega\times(0,T], \tag{76.1}$$

$$\operatorname{div}\boldsymbol{u}=0, \tag{76.2}$$

$$\boldsymbol{u}=0 \quad \text{on } \partial\Omega\times[0,T], \tag{76.3}$$

$$\boldsymbol{u}(x,0)=\boldsymbol{u}_0(x) \quad \text{in } \Omega. \tag{76.4}$$

First, we introduce a function space $H_0^1(\Omega)=\mathring{W}_2^1(\Omega)$, H—the closure of W in $L_2(\Omega)$ and V—the closure of v in $L_2(\Omega)$,

$$W_2^1=\{\boldsymbol{u}\in D(\Omega)\colon \operatorname{div}\boldsymbol{u}=0\}.$$

Then, in a weak statement, it is required for given $\boldsymbol{u}_0\in H$ and $f\in L_2(0,T;H)$ to find a function $\boldsymbol{u}\in L_2(0,T;V)$ that satisfies the conditions

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{u},\boldsymbol{v})+b(\boldsymbol{u},\boldsymbol{u},\boldsymbol{v})+\nu[\boldsymbol{u},\boldsymbol{v}]=(f,\boldsymbol{v}) \quad \forall v\in V, \tag{76.5}$$

$$\boldsymbol{u}(0)=\boldsymbol{u}_0, \qquad [\boldsymbol{u},\boldsymbol{v}]=\sum_{i=1}^{n}\left(\frac{\partial \boldsymbol{u}}{\partial x_i},\frac{\partial \boldsymbol{v}}{\partial x_i}\right). \tag{76.6}$$

Here $(\boldsymbol{u},s)=\int_\Omega \boldsymbol{u}v\,\mathrm{d}\Omega$, and

$$b(\boldsymbol{u},s,w)=\int_\Omega \sum_{ij=1}^{n} u_i\left(\frac{\partial}{\partial x_i}v_j\right)w_j\,\mathrm{d}\Omega$$

is a three-linear form, $\boldsymbol{u}_0\in H$, $f\in L_2(0,T;H)$.

Let the interval $[0,T]$ be divided into subintervals of length τ. We assume that

$$f^n=\frac{1}{\tau}\int_{(n-1)\tau}^{n\tau} f(t)\,\mathrm{d}t, \quad n=1,\ldots,N.$$

We introduce elements $\boldsymbol{u}^{n+1/2}$, $i=0,1$, $n=0,\ldots,N-1$, and, beginning with $\boldsymbol{u}^0=\boldsymbol{u}_0$, compute all $\boldsymbol{u}^{n+1/2}$ consecutively using the formulae

$$\boldsymbol{u}^{n+1/2}\in H_0^1(\Omega),$$

$$\left(\frac{\boldsymbol{u}^{n+1/2}-\boldsymbol{u}^n}{\tau},\boldsymbol{v}\right)+\hat{b}(\boldsymbol{u}^{n+1/2},\boldsymbol{u}^{n+1/2},\boldsymbol{v})+\nu[\boldsymbol{u}^{n+1/2},\boldsymbol{v}]=(f^n,\boldsymbol{v}) \tag{76.7}$$

$$\forall \boldsymbol{v}\in H_0^1(\Omega),$$

$$\begin{aligned} &u^{n+1} \in H, \\ &(u^{n+1}, v) = (u^{n+1/2}, v) \quad \forall v \in H. \end{aligned} \tag{76.8}$$

Here

$$\hat{b}(u, v, w) = \tfrac{1}{2}(b(u, v, w) - b(u, w, v))$$

is a continuous three-linear form on $H_0^1(\Omega)$ and

$$\hat{b}(u, s, v) = 0 \quad \forall u, v \in H_0^1(\Omega). \tag{76.9}$$

Equation (76.8) means that u^{n+1} is an orthogonal projection of $u^{n+1/2}$ on H in $L_2(\Omega)$. That is why u^{n+1} can be defined by the equality

$$u^{n+1} = p_H u^{n+1/2},$$

where p_H is an orthogonal projector in $L_2(\Omega)$ on the subspace H. TEMAM ([1981], Theorem 1.1.4) shows that, in the orthogonal complement H^1, the difference $u^{n+1} - u^{n+1/2}$ is a gradient of some function of $H^1(\Omega)$. We denote this by τp^{n+1} and obtain

$$\frac{u^{n+1} - u^{n+1/2}}{\tau} + \operatorname{grad} p^{n+1} = 0, \quad p^{n+1} \in H^1(\Omega). \tag{76.10}$$

Relationship (76.10) is one of the consequences of (76.8). The second relationship, equivalent to (76.8) and (76.10), is given by

$$\begin{aligned} &\operatorname{div} u^{n+1} = 0 \quad \text{in } \Omega, \\ &u^{n+1} \cdot n = 0 \quad \text{on } \partial\Omega; \\ &u^{n+1} \in L_2(\Omega). \end{aligned}$$

These expressions are equivalent to a certain Neumann problem with respect to p^{n+1}. We apply operator div to (76.10) and obtain

$$\begin{aligned} &\Delta p^{n+1} = \frac{1}{\tau} \operatorname{div} u^{n+1/2} \quad (\operatorname{div} u^{n+1} = 0), \\ &\frac{\partial p^{n+1}}{\partial n} = 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{76.11}$$

Relationship (76.7) is a nonlinear Dirichlet problem. Writing (76.7) for $s \in D(\Omega)$, we obtain the classical form

$$\begin{aligned} &\frac{u^{n+1/2} - u^n}{\tau} + \sum_{i=1}^{n} u_i^{n+1/2} \frac{\partial u^{n+1/2}}{\partial x_i} + \tfrac{1}{2}(\operatorname{div} u^{n+1/2}) u^{n+1/2} \\ &\quad = f^{n+} \nu \Delta u^{n+1/2}. \end{aligned} \tag{76.12}$$

Using a weak form of equations in this case helps to perform a separation of operators which is not obvious in the classical statement. The last term on the left-hand side of (76.12) is a stabilizing addition arising as a result of replacing $\hat{b}(u, u, v)$ by $\hat{b}(u, u, v)$ in (76.7). A further solution of the obtained problems has been described by TEMAM [1981]. It is also noteworthy that the generated function p^{n+1} is not a "real" pressure.

The splitting idea connected with recalculation of the pressure given in a projection statement is, in fact, a consequence of ideas underlying the Harlow method for particles in cells. According to this scheme let us, first, compute an intermediate velocity field from

$$\frac{\boldsymbol{u}^{n+1/2}-\boldsymbol{u}^n}{\tau}+\sum_{i=1}^{n} u_i^n \frac{\partial \boldsymbol{u}^n}{\partial x_i}=\nu\Delta \boldsymbol{u}^n+f. \tag{76.13}$$

Then we correct this field by taking into account the pressure gradient

$$\boldsymbol{u}^{n+1}=\boldsymbol{u}^{n+1/2}-\tau \operatorname{grad} p, \tag{76.14}$$

where p is a stationary solution of

$$\frac{\partial p}{\partial t}+\operatorname{div} \boldsymbol{u}^{n+1/2}=\tau\Delta p. \tag{76.15}$$

As a result of the above stages both (76.1) and (76.2) are satisfied.

BELOTSERKOVSKY, GUSHCHIN and SHCHENNIKOV [1975] used another splitting scheme, namely, the explicit scheme of splitting along the physical factors. We introduce the notation $\operatorname{rot} \boldsymbol{u}=\boldsymbol{w}$ and consider the velocity field $\boldsymbol{u}$ and pressure field p at the time $t_n=n\tau$ to be known, then we can write the scheme for finding unknown functions at time t_{n+1} as the three-stage splitting scheme (76.16)–(76.18):

$$\frac{\boldsymbol{u}^{n+1/2}-\boldsymbol{u}^n}{\tau}=-\sum_{i=1}^{n} u_i^n \frac{\partial \boldsymbol{u}^n}{\partial x_i}+\nu\Delta \boldsymbol{u}^n, \tag{76.16}$$

$$\Delta p=-\frac{1}{\tau}\operatorname{div} \boldsymbol{u}^{n+1/2}, \tag{76.17}$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\tau}=-\operatorname{grad} p. \tag{76.18}$$

As noted earlier, equation (76.17) is obtained by applying operation div to both sides of (76.18) and taking the continuity equation into account (vector $\boldsymbol{u}^{n+1}$ is solenoidal). BELOTSERKOVSKY [1984] suggested the following physical interpretation of the above scheme. It is assumed at the first stage (76.16) that the momentum transport (a mass unit impulse) is realized only via convection and diffusion. The field $\boldsymbol{u}^{n+1/2}$ thereby obtained does not satisfy the incompressibility condition giving nevertheless a correct description of the turbulent characteristics since it is possible to show that $\operatorname{rot} \boldsymbol{u}^{n+1/2}=w^{n+1}$.

The obtained intermediate velocity field helps to find the field of pressure on the second stage (76.17) taking into account the fact that the field $\boldsymbol{u}^{n+1}$ is solenoidal. In this approach the function p^{n+1} is a physical pressure and boundary conditions for it are not defined. On the contrary, in the projection method, p^{n+1} satisfies Neumann boundary conditions.

It is assumed at the third stage (76.18) that the transport is realized due to pressure gradients only.

The following variant (76.19)–(76.21) of realization has been considered for

increasing the reserve of stability by using an implicit scheme at the first stage (BELOTSERKOVSKY [1984]).

$$\frac{\boldsymbol{u}^{n+1/2}-\boldsymbol{u}^n}{\tau} = -\sum_{i=1}^{n} u_i^n \frac{\partial \boldsymbol{u}^{n+1/2}}{\partial x_i} + \nu\Delta \boldsymbol{u}^{n+1/2} - \operatorname{grad} p^n + f, \tag{76.19}$$

$$\Delta\delta p = \frac{1}{\tau}\operatorname{div} \boldsymbol{u}^{n+1/2}, \qquad \delta p = p^{n+1}-p^n, \tag{76.20}$$

$$\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^{n+1/2}}{\tau} = -\operatorname{grad}(\delta p). \tag{76.21}$$

The proposed modification allows us to release the hard restrictions on the time step inherent to the first scheme. Another advantage of this approach is that the pressure increment in time and not the pressure itself is found in the second stage.

77. The general principle for constructing splitting schemes for Navier–Stokes equations

We consider the original system of Navier–Stokes equations in a Cartesian coordinate system, x_α ($\alpha = 1, 2, 3$). In the absence of external forces this system can be written in a divergent form

$$\frac{\partial \boldsymbol{U}}{\partial t} + \frac{\partial \boldsymbol{w}_\alpha}{\partial x_\alpha} = 0 \tag{77.1}$$

Here $\boldsymbol{U}$ is a vector of a flow's state, $\boldsymbol{w}_\alpha$ is a vector of hydrodynamic fluxes

$$\boldsymbol{U} = \begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix} \equiv \begin{bmatrix} \rho \\ \rho u_1 \\ \rho u_2 \\ \rho u_3 \\ E \end{bmatrix},$$

$$\boldsymbol{w}_\alpha = \boldsymbol{w}_\alpha\left(\boldsymbol{U}, \frac{\partial \boldsymbol{U}}{\partial x}\right) = \begin{bmatrix} w_{\alpha 0} \\ w_{\alpha 1} \\ w_{\alpha 2} \\ w_{\alpha 3} \\ w_{\alpha 4} \end{bmatrix} = \begin{bmatrix} \rho u_\alpha \\ \rho u_1 u_\alpha + p_{1\alpha} \\ \rho u_2 u_\alpha + p_{2\alpha} \\ \rho u_3 u_\alpha + p_{3\alpha} \\ \rho u_\alpha E + u_\beta p_{\alpha\beta} - \kappa \dfrac{\partial T}{\partial x_\alpha} \end{bmatrix} \tag{77.2}$$

In representation (77.2) we sum over β, $p_{\alpha\beta} = p\delta_{\alpha\beta} - G_{\alpha\beta}$, p is the pressure, ρ is the

density, $\delta_{\alpha\beta}$ is the Kronecker symbol, $E=e+\frac{1}{2}|\boldsymbol{u}|^2$ is the mass density of full energy,

$$G_{\alpha\beta}=\lambda \operatorname{div} \boldsymbol{u}\delta^{\alpha\beta}+\mu\left(\frac{\partial u_\alpha}{\partial x_\beta}+\frac{\partial u_\beta}{\partial x_\alpha}\right),$$
$$\lambda=\mu'-\tfrac{2}{3}\mu, \tag{77.3}$$

μ and μ' are the coefficients of volume and dissipative viscosity. System (77.1) may be represented in a nondivergent form as well. We introduce a new vector of a state f

$$f=f(U). \tag{77.4}$$

We consider (77.4) as reversible, equivalent to $U=U(f)$, and obtain nonsingular matrices,

$$A=\frac{\partial U}{\partial f}=\{A_{ij}\}\equiv\left\{\frac{\partial U_i}{\partial f_j}\right\}, \tag{77.5}$$

$$a=\frac{\partial f}{\partial U}=\{a_{ij}\}\equiv\left\{\frac{\partial f_i}{\partial U_j}\right\}. \tag{77.6}$$

Thus, taking this into account, system (77.1) can be rewritten in the form

$$\frac{\partial f}{\partial t}+\alpha\frac{\partial w_\alpha}{\partial x_\alpha}=0. \tag{77.7}$$

It follows from relationships (77.2) and (77.3) that $w_\alpha=w_\alpha(f, \partial f/\partial x)$. System (77.7) can be rewritten in operator form:

$$(\partial/\partial t+\Omega)f=0,$$
$$\Omega=B_\alpha D_\alpha+C_{\alpha\beta}D_\alpha D_\beta, \tag{77.8}$$
$$\beta_\alpha=a\,\partial w_\alpha/\partial f, \qquad c_{\alpha\beta}=a\,\partial w_\alpha/\partial f_\beta.$$

Then, equation (77.8) takes the form

$$\frac{\partial f}{\partial t}+B_\alpha\frac{\partial f}{\partial x_\alpha}+C_{\alpha\beta}\frac{\partial^2 f}{\partial x_\alpha\partial x_\beta}=0. \tag{77.9}$$

We introduce, further, operators

$$\Omega_\alpha=B_\alpha D_\alpha, \qquad R_\alpha=-C_{\alpha\beta}D_\alpha D_\beta$$

(no summation over α), then

$$\Omega=\sum_{\alpha=1}^{3}(\Omega_\alpha-R_\alpha). \tag{77.10}$$

System (77.9) is nondivergent in comparison with the equivalent system (77.1). Various vectors of state, e.g., (ρ, u_1, u_2, u_3, T), (ρ, u_1, u_2, u_3, e), and $f=(\rho, u_1, u_2, u_3, p)$ are

possible. In the first case

$$\Omega_\alpha=\begin{bmatrix} u_\alpha\frac{\partial}{\partial x_\alpha} & \delta_{1\alpha}\rho\frac{\partial}{\partial x_\alpha} & \delta_{2\alpha}\rho\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}\rho\frac{\partial}{\partial x_\alpha} & 0 \\ \delta_{1\alpha}a^2\frac{\partial}{\partial x_\alpha} & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & 0 & \delta_{1\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ \delta_{2\alpha}a^2\frac{\partial}{\partial x_\alpha} & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & \delta_{2\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ \delta_{3\alpha}a^2\frac{\partial}{\partial x_\alpha} & 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ 0 & \delta_{1\alpha}s^2\frac{\partial}{\partial x_\alpha} & \delta_{2\alpha}s^2\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}s^2\frac{\partial}{\partial x_\alpha} & u_\alpha\frac{\partial}{\partial x_\alpha} \end{bmatrix},$$

$$a^2=\frac{1}{\rho}\frac{\partial p}{\partial\rho},\qquad b^2=\frac{1}{\rho}\frac{\partial p}{\partial T},\qquad s^2=\left(\frac{p}{\rho}-\rho\frac{\partial l}{\partial\rho}\right)\Big/\frac{\partial l}{\partial T}.$$

Considering vector $f=(\rho, u_\alpha, p)$ and a continuity equation in divergent form simplifies the form of operators Ω_α to

$$\Omega_\alpha=\begin{bmatrix} \frac{\partial}{\partial x_\alpha} & 0 & 0 & 0 & 0 \\ 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & 0 & \delta_{\alpha1}\frac{1}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & \delta_{\alpha2}\frac{1}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & \delta_{\alpha3}\frac{1}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & \delta_{1\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & \delta_{2\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & u_\alpha\frac{\partial}{\partial x_\alpha} \end{bmatrix},$$

where $c^2=\mathrm{d}p/\mathrm{d}\rho$ is the square of the speed of sound.

The choice of vector f for the Navier–Stokes equation determines the form of representation R_α. The simplest form can be written for $f=(\rho, u_\alpha, T)$:

$$R_\alpha=\begin{bmatrix} \frac{\partial}{\partial x_\alpha}G_{1\alpha} \\ \frac{\partial}{\partial x_\alpha}G_{2\alpha} \\ \frac{\partial}{\partial x_\alpha}G_{3\alpha} \\ \frac{\partial}{\partial x_\alpha}\kappa\frac{\partial T}{\partial x_\alpha}+\Phi \end{bmatrix},$$

$$\Phi = 2\mu\left[\left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + \left(\frac{\partial u_3}{\partial x_3}\right)^2\right] + \lambda(\operatorname{div} \boldsymbol{u})^2$$

$$+\mu\left[\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right)^2 + \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2}\right)^2\right].$$

Now we consider splitting of the Navier–Stokes equations in a divergent and nondivergent form. Let us do this in Cartesian coordinates. The generalization of the splitting method for equations of gas dynamics and Navier–Stokes equations for the cases of curvilinear coordinates is carried out like in the above case (KOVENYA and YANENKO [1981]). We represent the vector w_α of fluxes defined in equations (77.1) and (77.2) as a sum of three terms

$$w_\alpha = w_{1\alpha} + w_{2\alpha} + w_{3\alpha}, \tag{77.11}$$

where $w_{1\alpha}$ is the vector of convective fluxes, $w_{2\alpha}$ the vector of flux due to pressure forces, and $w_{3\alpha}$ the vector of dissipative forces:

$$w_{1\alpha} = \begin{bmatrix} w^1_{\alpha 0} \\ w^1_{\alpha 1} \\ w^1_{\alpha 2} \\ w^1_{\alpha 3} \\ w^1_{\alpha 4} \end{bmatrix} \equiv \begin{bmatrix} \rho u_\alpha \\ \rho u_1 u_\alpha \\ \rho u_2 u_\alpha \\ \rho u_3 u_\alpha \\ \rho u_\alpha E \end{bmatrix}, \qquad w_{2\alpha} = \begin{bmatrix} w^2_{\alpha 0} \\ w^2_{\alpha 1} \\ w^2_{\alpha 2} \\ w^2_{\alpha 3} \\ w^2_{\alpha 4} \end{bmatrix} \equiv \begin{bmatrix} 0 \\ \delta_{1\alpha} p \\ \delta_{2\alpha} p \\ \delta_{3\alpha} p \\ u_\alpha p \end{bmatrix},$$

$$w_{3\alpha} = \begin{bmatrix} w^3_{\alpha 0} \\ w^3_{\alpha 1} \\ w^3_{\alpha 2} \\ w^3_{\alpha 3} \\ w^3_{\alpha 4} \end{bmatrix} \equiv \begin{bmatrix} 0 \\ -G_{1\alpha} \\ -G_{2\alpha} \\ -G_{3\alpha} \\ -\kappa \dfrac{\partial T}{\partial x_\alpha} - \displaystyle\sum_{\beta=1}^{3} u_\beta G_{\alpha\beta} \end{bmatrix}.$$

When we consider each vector w_α as a sum (77.11) for every direction x_α, we can write the system of equations in the form of the splitting on physical processes and spatial directions,

$$\frac{\partial U}{\partial t} + \sum_{\alpha=1}^{3} \left(\sum_{k=1}^{3} \frac{\partial w_{\alpha k}}{\partial x_\alpha} \right) = 0. \tag{77.12}$$

Representing system (77.12) as

$$\frac{\partial \boldsymbol{U}}{\partial t}+\sum_{m=1}^{g} \mathscr{L}_m \boldsymbol{U}=0, \tag{77.13}$$

where

$$\sum_{m=1}^{g} \mathscr{L}_m \boldsymbol{U}=\sum_{\alpha=1}^{3}\left(\sum_{k=1}^{3} \frac{\partial}{\partial x_\alpha} w_{\alpha k}\right),$$

we have a system split on a differential level (approximating weakly) in the form

$$\frac{\partial \boldsymbol{U}}{\partial t}+\sum_{j=1}^{N} \alpha_i \mathscr{L}_j \boldsymbol{U}=0,$$

where

$$\begin{aligned}
&\alpha_1=N, \quad \alpha_l=0 \quad \text{for } l>1, \quad t_n \leqslant t \leqslant t_n+\tau/N,\\
&\alpha_2=N, \quad \alpha_l=0 \quad \text{for } l \neq 2, \quad t_n \leqslant t \leqslant t_n+2\tau/N,\\
&\quad \vdots\\
&\alpha_N=N, \quad \alpha_l=0 \quad \text{for } l \neq N, \quad t_n+\tau(N-1) \leqslant t \leqslant t_n+\tau.
\end{aligned}$$

Here N is the number of steps, the $\mathscr{L}_i$ are differential operators taking into account the splitting in physical processes and spatial directions.

When we write the system in the nondivergent form (77.7), splitting in physical processes and spatial directions is done as in (77.12) and (77.13): one separates the convective flux matrices and the matrices determined by the pressure and dissipative forces with the subsequent splitting along the spatial directions so that operator Ω is represented as

$$\Omega=\sum_{\alpha=1}^{3}\left(\sum_{k=1}^{3} \Omega_k^\alpha\right). \tag{77.14}$$

Here index α is related to splitting in directions x_α, and k to splitting in physical processes. Depending on the choice of a vector f of unknown functions, the matrix differential operator Ω has various structures. Hence, the form of operators $\Omega_{\alpha k}$ depends on the choice of f. The arbitrary character of the choice of unknown functions allows us to consider various forms of splitting and to choose those that have the largest reserve of stability in a difference form or the simplest realization.

We illustrate some of the splitting forms with examples. First, we only consider a system of equations of gas dynamics ($\kappa=0$, $G_{ij}=0$) resolved with respect to the vector $f=(\rho, u_1, u_2, u_3, p)^{\mathrm{T}}$. Since $\kappa=0$ and $G_{ij}=0$, the dissipative forces matrix $\Omega_{3\alpha}$ is absent, and consequently $k=1$ or 2. In considering continuity equations in a divergent form and other equations in a nondivergent form the corresponding

matrices $\Omega_{k\alpha}$ have the form,

$$\Omega_{1\alpha}=\begin{bmatrix} \frac{\partial}{\partial x_\alpha}u_\alpha & 0 & 0 & 0 & 0 \\ 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & 0 & 0 \\ 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 & 0 \\ 0 & 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} & 0 \\ 0 & 0 & 0 & 0 & u_\alpha\frac{\partial}{\partial x_\alpha} \end{bmatrix}, \tag{77.15}$$

$$\Omega_{2\alpha}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\delta_{1\alpha}}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & 0 & 0 & 0 & \frac{\delta_{2\alpha}}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & 0 & 0 & 0 & \frac{\delta_{3\alpha}}{\rho}\frac{\partial}{\partial x_\alpha} \\ 0 & \delta_{1\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & \delta_{2\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}\rho c^2\frac{\partial}{\partial x_\alpha} & 0 \end{bmatrix}, \tag{77.16}$$

where c is the speed of the sound.

We consider splitting of equations (77.7) in physical processes and spatial variables for vector $f=(\rho, u_1, u_2, u_3, T)^{\mathrm{T}}$. The convective fluxes matrix and the matrix of pressure gradient forces are written as follows:

$$\Omega_{1\alpha}=u_\alpha\frac{\partial}{\partial x_\alpha}I, \tag{77.17}$$

$$\Omega_{2\alpha}=\begin{bmatrix} 0 & \delta_{1\alpha}\rho\frac{\partial}{\partial x_\alpha} & \delta_{2\alpha}\rho\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}\rho\frac{\partial}{\partial x_\alpha} & 0 \\ \delta_{1\alpha}a^2\frac{\partial}{\partial x_\alpha} & 0 & 0 & 0 & \delta_{1\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ \delta_{2\alpha}a^2\frac{\partial}{\partial x_\alpha} & 0 & 0 & 0 & \delta_{2\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ \delta_{3\alpha}a^2\frac{\partial}{\partial x_\alpha} & 0 & 0 & 0 & \delta_{3\alpha}b^2\frac{\partial}{\partial x_\alpha} \\ 0 & \delta_{1\alpha}s^2\frac{\partial}{\partial x_2} & \delta_{2\alpha}s^2\frac{\partial}{\partial x_\alpha} & \delta_{3\alpha}s^2\frac{\partial}{\partial x_\alpha} & 0 \end{bmatrix}, \tag{77.18}$$

where

$$a^2=\frac{1}{\rho}\frac{\partial p}{\partial \rho}, \qquad b^2=\frac{1}{\rho}\frac{\partial p}{\partial t}, \qquad s^2=\frac{p}{\rho}.$$

It is obvious that realizations of schemes based on splitting (77.15)–(77.16) or (77.17)–(77.18) are different. The above examples can be continued if we choose gas dynamics functions or some combinations of them as the functions to be found. In considering difference splitting schemes for boundary value problems the choice of unknown functions is also determined by the form of the boundary conditions.

For Navier–Stokes equations written in the nondivergent form (77.7) splitting of equations in physical processes and spatial directions is performed in the same way as for equations of gas dynamics. The dissipative forces matrix is added in this case separate from the convective transport matrix and the matrix of flows is determined by pressure forces. For vector $f=(\rho, u_\alpha, T)^{\mathrm{T}}$ matrix $\Omega_{3\alpha}$ has the form

$$\Omega_{3\alpha}f=-\frac{1}{\rho}\frac{\partial}{\partial x_\alpha}\begin{bmatrix} 0 \\ G_{1\alpha} \\ G_{2\alpha} \\ G_{3\alpha} \\ \kappa\dfrac{\partial}{\partial x_\alpha} \end{bmatrix}. \tag{77.19}$$

Some other examples of splitting gas dynamics and Navier–Stokes equations are possible. These examples will belong to various realizations differing in the reserve of stability, effectiveness of realization and convenience for calculating boundary conditions. So, the approximation of equations in a divergent form leads to conservative difference schemes satisfying conservation laws. This yields a physically more grounded result. For nonstationary solutions this way seems to be most justified. At the same time the numerical schemes of this type are nonlinear and require either internal iterations or a complex linearization.

Although equations written in a divergent form are more efficient in realization, they yet have the conservation property only asymptotically, that is, in the process of stationing. In this case, it is efficient to consider the predictor-corrector method that permits the construction of conservative difference schemes suitable for solving nonstationary problems (see KOVENYA and LEBEDEV [1984]). At the predictor stage in the approximation of equations in nondivergent form a scheme of splitting in physical processes and coordinate directions is constructed. At the corrector stage the equation is approximated in a divergent form.

The above schemes have been used in either way for performing a whole range of computations for problems of gas dynamics and viscous gas flows in both two-dimensional and three-dimensional cases.

Papers by KOVENYA and YANENKO [1981], TEMAM [1981], BELOTSERKOVSKY [1984], YANENKO [1967] and GUSHCHIN [1981] discuss the construction and the convergence studies of splitting schemes for Navier–Stokes equations. Splitting

schemes and the results of computing the viscous fluid fluxes by using the notion of "artificial compressibility" are given by YANENKO [1967] and CHORIN [1968]. YANENKO [1967] also considered the coordinatewise splitting scheme. Splitting methods for solving problems of the stratified fluid dynamics are described by LYTKIN and CHERNYKH [1975], YOUNG and HORT [1972], GUSHCHIN [1981], VASILYEV, KUZNETSOV, LYTKIN and CHERNYKH [1974]. Splitting methods for solving a wide range of problems of gas and aerodynamics along with some computational results are discussed by KUZNETSOV and STRELETS [1983], KOVENYA and LEBEDEV [1984], KOVENYA and YANENKO [1981] and BELOTSERKOVSKY [1984]. Monographs by MARCHUK [1974], KOVENYA and YANENKO [1981] and BELOTSERKOVSKY [1984] are actually surveys of the results of constructing, studying and realizing splitting schemes for problems of geophysical hydrodynamics, gas dynamics, hydrodynamics and mechanics of a continuous medium.

CHAPTER XIX

Problems in Meteorology

A number of implicit schemes for integrating equations of dynamic meteorology have been developed and grounded by MARCHUK [1964b, 1965, 1967, 1974a]. These schemes are widely applied for solving problems of weather forecast, theory of climate and environment protection (see the related publications by DYMNIKOV and ISHIMOVA [1979], MARCHUK [1980], MARCHUK, DYMNIKOV et al. [1975, 1979, 1984], BURRIDGE and HAYES [1974], LEPAS et al. [1974], ZENG and ZHANG [1982], DYMNIKOV and LYKOSOV [1983], DYMNIKOV and FILIN [1985], MARCHUK, DYMNIKOV et al. [1981].

78. Equations of atmosphere hydrothermodynamics

Equations of atmosphere hydrothermodynamics with a quasistatic approximation in spherical σ-coordinates have the following form (MARCHUK, DYMNIKOV et al. [1984]):

$$\frac{du}{dt}-(l+\frac{u}{a}\operatorname{tg}\varphi)v+\frac{1}{a\cos\varphi}\left(\frac{\partial\Phi}{\partial\lambda}+\frac{RT}{p_s}\frac{\partial p_s}{\partial\lambda}\right)=F_u, \tag{78.1}$$

$$\frac{dv}{dt}+\left(l+\frac{u}{a}\operatorname{tg}\varphi\right)u+\frac{1}{a}\left(\frac{\partial\Phi}{\partial\varphi}+\frac{T}{p_s}\frac{\partial p_s}{\partial\varphi}\right)=F_v, \tag{78.2}$$

$$\frac{dT}{dt}-\frac{RT}{c_p\sigma p_s}\tau=F_T+\varepsilon_T, \tag{78.3}$$

$$\frac{dq}{dt}=F_q-\varepsilon_q, \tag{78.4}$$

$$\frac{\partial p_s}{\partial t}+\frac{1}{a\cos\varphi}\left(\frac{\partial p_s u}{\partial\lambda}+\frac{\partial p_s v\cos\varphi}{\partial\varphi}\right)+\frac{\partial p_s\dot{\sigma}}{\partial\sigma}=0, \tag{78.5}$$

$$\frac{\partial\Phi}{\partial\sigma}=-\frac{RT}{\sigma}, \tag{78.6}$$

$$\tau=p_s\dot{\sigma}+\sigma\left(\frac{\partial p_s}{\partial t}+\frac{u}{a\cos\varphi}\frac{\partial p_s}{\partial\lambda}+\frac{v}{a}\frac{\partial p_s}{\partial\varphi}\right), \tag{78.7}$$

where

$$\frac{d}{dt}=\frac{\partial}{\partial t}+K, \quad K=\frac{u}{a\cos\varphi}\frac{\partial}{\partial\lambda}+\frac{v}{a}\frac{\partial}{\partial\varphi}+\dot{\sigma}\frac{\partial}{\partial\sigma};$$

t is the time variable; λ is the longitude and φ the latitude; $\sigma=p/p_s$ is a vertical coordinate (p is the pressure, p_s is the pressure at the surface of the Earth); u, v and $\dot{\sigma}$ are wind velocity components in the longitudinal, latitudinal and vertical direction, respectively; $\Phi=gz$ is a geopotential on a constant σ-surface (where g is the gravitational acceleration and z is the altitude above the sea level); T is the temperature (K); q is the specific humidity; F_u and F_v are the rates of impulse moment changes due to Reynolds stresses; F_T and F_q are the rates of temperature and specific humidity changes, respectively, caused by small-scale diffusion and meso-scale convection; $\varepsilon_T=\varepsilon_r+\varepsilon_f$ are diabatic heat fluxes (ε_r is a radiation heat flux and ε_f a heat flux due to condensation); $\varepsilon_q=c-E$ is the rate of humidity transformations (C and E are terms describing condensation and evaporation processes); $l=2\Omega\sin\varphi$ is a Coriolis parameter (Ω is the angular speed of rotation of the Earth); a is the radius of the Earth; R and C_p are gas constants for dry air and specific heat of air at constant pressure.

The following initial conditions are considered to be given for system (78.1)–(78.6)

$$\begin{aligned} &u=u_0(\lambda,\varphi,\sigma), \qquad v=v_0(\lambda,\varphi,\sigma),\\ &p_s=p_{s0}(\lambda,\varphi),\\ &T=T_0(\lambda,\varphi,\sigma), \qquad q=q_0(\lambda,\varphi,\sigma) \end{aligned} \tag{78.8}$$

as well as the boundary conditions which assume that a solution is periodic in longitude and restricted on the North and South Poles. As far as the underlying surface being a solid body is a σ-coordinate surface ($\sigma=1$), the corresponding kinematic condition of type

$$\dot{\sigma}=0 \quad \text{for } \sigma=1 \tag{78.9}$$

is used. An analogous condition can be stated for the upper boundary of the atmosphere

$$\dot{\sigma}=0 \quad \text{for } \sigma=0, \tag{78.10}$$

and the distribution of a geopotential on the lower boundary is also given:

$$\Phi=gz_s\equiv\Phi_s \tag{78.11}$$

where z_s is the Earth's surface altitude above sea level.

Equations (78.1)–(78.4) can take a different form. At least three possible forms are known. By multiplying any of these equations by p_s and using the continuity equation (78.5) we obtain a divergent form

$$\begin{aligned} &\frac{\partial p_s u}{\partial t}+K_d u-\left(l+\frac{u}{a}\operatorname{tg}\varphi\right)p_s v\\ &\quad+\frac{p_s}{a\cos\varphi}\left(\frac{\partial\Phi}{\partial\lambda}+RT\frac{\partial}{\partial\lambda}\ln p_s\right)=p_s F_u, \end{aligned} \tag{78.12}$$

$$\frac{\partial p_s v}{\partial t}+K_d v+\left(l+\frac{u}{a}\operatorname{tg}\varphi\right)p_S u$$
$$+\frac{p_s}{a}\left(\frac{\partial \Phi}{\partial \varphi}+RT\frac{\partial}{\partial \varphi}\ln p_s\right)=p_s F_v, \tag{78.13}$$

$$\frac{\partial p_s T}{\partial t}+K_d T-\frac{RT}{C_p \sigma}\tau=p_s(F_T+\varepsilon_T), \tag{78.14}$$

$$\frac{\partial p_s q}{\partial t}+K_d q=p_s(F_q-\varepsilon_q), \tag{78.15}$$

where

$$K_d=\frac{1}{a\cos\varphi}\left(\frac{\partial}{\partial \lambda}p_s u\cdot+\frac{\partial}{\partial \varphi}p_s v\cos\varphi\cdot\right)+\frac{\partial}{\partial \sigma}p_s\dot{\sigma}\cdot.$$

We use the following substitutions

$$\sqrt{p_s}=\pi, \qquad \sqrt{p_s}u=U, \qquad \sqrt{p_s}v=V,$$
$$\sqrt{p_s}T=\theta, \qquad \sqrt{p_s}q=Q, \tag{78.16}$$

and we obtain a symmetrized form of equations

$$\frac{\partial U}{\partial t}+K_s U-\left(l+\frac{U}{a\pi}\operatorname{tg}\varphi\right)V$$
$$+\frac{1}{a\cos\varphi}\left(\pi\frac{\partial \Phi}{\partial \lambda}+2R\theta\frac{\partial}{\partial \lambda}\ln\pi\right)=\pi F_u, \tag{78.17}$$

$$\frac{\partial V}{\partial t}+K_s V+\left(l+\frac{U}{a\pi}\operatorname{tg}\varphi\right)U$$
$$+\frac{1}{a}\left(\pi\frac{\partial \Phi}{\partial \varphi}+2R\theta\frac{\partial}{\partial \varphi}\ln\pi\right)=\pi F_v, \tag{78.18}$$

$$\frac{\partial \theta}{\partial t}+K_s\theta-\frac{R\theta}{C_p\pi^2}\tau=\pi(F_T+\varepsilon_T), \tag{78.19}$$

$$\frac{\partial Q}{\partial t}+K_s Q=\pi(F_q-\varepsilon_q), \tag{78.20}$$

where

$$K_s=\frac{1}{2a\cos\varphi}\left[\left(\frac{U}{\pi}\frac{\partial}{\partial \lambda}+\frac{\partial}{\partial \lambda}\frac{U}{\pi}\cdot\right)+\left(\frac{V\cos\varphi}{\pi}\frac{\partial}{\partial \varphi}+\frac{\partial}{\partial \varphi}\frac{V\cos\varphi}{\pi}\cdot\right)\right]$$
$$+\tfrac{1}{2}\left(\dot{\sigma}\frac{\partial}{\partial \sigma}+\frac{\partial}{\partial \sigma}\dot{\sigma}\cdot\right). \tag{78.21}$$

We introduce the following notation for an absolute eddy Z and energy E

$$Z=l+\frac{1}{a\cos\varphi}\left(\frac{\partial v}{\partial\lambda}-\frac{\partial u\cos\varphi}{\partial\varphi}\right),$$

$$E=\tfrac{1}{2}(u^2+v^2).$$

Now we can write the motion equations (78.1) and (78.2) in the so-called Lamb form

$$\frac{\partial u}{\partial t}+\dot\sigma\frac{\partial u}{\partial\sigma}-Z_v+\frac{1}{a\cos\varphi}\left[\frac{\partial}{\partial\lambda}(\Phi+E)+RT\frac{\partial}{\partial\lambda}\ln p_s\right]=F_u, \tag{78.22}$$

$$\frac{\partial v}{\partial t}+\dot\sigma\frac{\partial v}{\partial\sigma}+Z_u+\frac{1}{a}\left[\frac{\partial}{\partial\varphi}(\Phi+E)+RT\frac{\partial}{\partial\varphi}\ln p_s\right]=F_v. \tag{78.23}$$

The system of equations of atmosphere hydrothermodynamics with adiabatic approximation has a number of integral properties. The conservation laws for total energy and enstrophy $(=Z^2)$ are the most interesting among them. Let the right-hand sides in equations (78.1)–(78.4) be equal to zero, and let z_s be absent for some time. Multiplying (78.1) by u, (78.2) by v and adding these expressions we obtain, after some transformations using equations of continuity (78.5) and quasi-statics (78.6), the following equation for the changes of kinetic energy in time:

$$\frac{\partial p_s E}{\partial t}+K_d E+\frac{1}{a\cos\varphi}\left(\frac{\partial p_s u\Phi}{\partial\lambda}+\frac{\partial p_s v\cos\varphi\Phi}{\partial\varphi}\right)+\frac{\partial}{\partial\sigma}\left(\sigma\frac{\partial p_s}{\partial t}+p_s\dot\sigma\right)\Phi=-\frac{RT}{\sigma}\tau. \tag{78.24}$$

Adding equations (78.24) and (78.3) and integrating the obtained expression in σ from 0 to 1 and over the Earth's surface G using the corresponding boundary conditions, we obtain

$$\frac{\partial}{\partial t}\int_G p_s\int_0^1(E+C_p T)\,d\sigma\,dG=0,\quad dG=a^2\cos\varphi\,d\varphi\,d\lambda. \tag{78.25}$$

Since it is known (LORENZ [1955]) that the total potential energy of a vertical air column in hydrostatic systems (being the sum of two kinds of energy—potential gz and internal $C_v T$, where C_v is specific heat of air at constant volume) coincides with its enthalpy $C_p T$, equation (78.25) is the conservation law for the total energy of an atmospheric system.

If the term RT is replaced in (78.3) by $R\tilde T$ where $\tilde T$ is the average temperature (a value independent of spatial coordinates and time), then the total energy conservation law takes the quadrature form

$$\frac{\partial}{\partial t}\int_G p_s\int_0^1\left(E+\frac{C_p}{2\tilde T}T^2\right)dG\,d\sigma=0. \tag{78.26}$$

The quadrature conservation law is a fundamental property of system (78.1)–(78.6), which allows the construction of a stable computation process for obtaining its solution.

In the case of nondivergent barotropic flow ($p_s \equiv \text{const}$, $\dot{\sigma} \equiv 0$) the Lamb form for the motion equations (78.22) and (78.23) along with the continuity equation in the form

$$\frac{1}{a\cos\varphi}\left(\frac{\partial u}{\partial \lambda}+\frac{\partial v\cos\varphi}{\partial \varphi}\right)=0 \tag{78.27}$$

allows to show the existence of the enstrophy conservation law

$$\frac{\partial}{\partial t}\int_G p_s \int_0^1 Z^2\,d\sigma\,dG=0. \tag{78.28}$$

This law has also a fundamental meaning because the real atmosphere is quasi-barotropic and quasi-geostrophic, and therefore has a specific energy cascade in the direction of longitudinal waves and enstrophy cascade in the direction of the high wave numbers.

We assume that $p_s \equiv 1$ in relationships (78.1)–(78.8) and obtain, as a special case, a system of equations in p-coordinates where, besides the quasi-statics equation

$$\frac{\partial \Phi}{\partial p}=-\frac{RT}{p}, \tag{78.29}$$

the continuity equation

$$\frac{1}{a\cos\varphi}\left(\frac{\partial u}{\partial \lambda}+\frac{\partial v\cos\varphi}{\partial \varphi}\right)+\frac{\partial \tau}{\partial p}=0 \tag{78.30}$$

turns out to be non-evolutionary.

The system of equations of atmosphere hydrothermodynamics with quasi-static approximation is irregular, i.e. it does not belong to the Cauchy–Kovalevskaya type systems but can be reduced to the evolutionary type by integrating, for example, the quasi-statics equation (78.6) in σ from σ to 1 and using condition (78.11):

$$\Phi=\Phi_s+\int_\sigma^1 \frac{RT}{\sigma}\,d\sigma; \tag{78.31}$$

and by integrating the continuity equation (78.5) in σ from 0 to 1 and using conditions (78.9) and (78.10):

$$\frac{\partial p_s}{\partial t}+\int_0^1 D\,d\sigma=0 \tag{78.32}$$

with the vertical velocity analogue $\dot{\sigma}$ being represented as follows

$$\dot{\sigma}=\frac{1}{p_s}\left(\sigma\int_0^1 D\,d\sigma-\int_0^\sigma D\,d\sigma\right),$$
$$D=\frac{1}{a\cos\varphi}\left(\frac{\partial p_s u}{\partial\lambda}+\frac{\partial p_s v\cos\varphi}{\partial\varphi}\right). \tag{78.33}$$

79. The general splitting method with respect to physical processes based on the separation of characteristic times

Systems of equations of atmosphere hydrothermodynamics appear to be fairly difficult to integrate (especially for long times) owing to the simultaneous presence of a great many physical processes with various characteristic time scales varying from a few seconds (gravity waves) to a few days and more (synoptic and planetary waves). It is natural, therefore, to use various computational schemes to describe terms related to the various physical processes.

Following MESINGER and ARAKAWA [1976] we consider, first, a linearized system ("shallow water") including advective terms and terms responsible for the description of gravity waves

$$\frac{\partial u}{\partial t}+c\frac{\partial u}{\partial x}+g\frac{\partial h}{\partial x}=0,$$
$$\frac{\partial h}{\partial t}+c\frac{\partial h}{\partial x}+H\frac{\partial u}{\partial x}=0, \tag{79.1}$$

where h is the altitude of the free surface, and H is its average value. The ratio of the velocity of the gravity waves propagation $\sqrt{gH}$ to the velocity of the advective transport c constitutes a value of order ~ 10. It is expedient, therefore, to solve the system of advection equations

$$\frac{\partial u}{\partial t}+c\frac{\partial u}{\partial x}=0,$$
$$\frac{\partial h}{\partial t}+c\frac{\partial h}{\partial x}=0 \tag{79.2}$$

first for system (79.1) in the limits of a given time step.

We denote the thereby obtained intermediate values of u^{n+1} and h^{n+1} by $u^{n+1/2}$ and $h^{n+1/2}$ and use them as initial conditions for solving the remaining subsystem

$$\frac{\partial u}{\partial t}+g\frac{\partial h}{\partial x}=0,$$
$$\frac{\partial h}{\partial t}+H\frac{\partial u}{\partial x}=0. \tag{79.3}$$

The values u^{n+1} and h^{n+1}, obtained after solving this subsystem, are taken to be the final approximate values of the variables at the time level $n+1$. This procedure is repeated on each successive time step.

We consider a vector function $\psi=\{u,h\}$ and a matrix

$$A=\begin{pmatrix} c & g \\ H & c \end{pmatrix}$$

and write (79.1) in the form

$$\frac{\partial\psi}{\partial t}+A\frac{\partial\psi}{\partial x}=0. \tag{79.4}$$

By using the following representation of the matrix A,

$$A=A_1+A_2, \qquad A_1=cE, \quad A_2=\begin{pmatrix} 0 & g \\ H & 0 \end{pmatrix},$$

we can represent the above procedure for solving systems of "shallow water" equations as a splitting scheme,

$$\begin{aligned} &\frac{\psi^{n+1/2}-\psi^n}{\Delta t}+A_1\Lambda\psi^{n+1/2}=F_1^n, \\ &\frac{\psi^{n+1}-\psi^{n+1/2}}{\Delta t}+A_2\Lambda\psi^{n+1}=F_2^{n+1/2}, \end{aligned} \tag{79.5}$$

where the difference operator Λ approximates a differential operator $\partial/\partial x$; F_1^n and $F_2^{n+1/2}$ are the result of applying difference operators to the function ψ on previous time steps.

The solution of a full diabatic problem in a symmetrized form, which is based on the method of splitting with respect to physical processes and uses weak approximation method symbolics, may be represented in various forms (MARCHUK, DYMNIKOV et al. [1984]):

(a) for the computation of radiative heat fluxes,

$$\begin{aligned} &\frac{\partial\theta_1}{\partial t}=\pi\varepsilon_r, \quad \pi=\pi(t_n), \\ &\theta_1(t_n)=\theta(t_n); \end{aligned} \tag{79.6}$$

(b) for the computation of heat fluxes due to condensation, dry and moist convection,

$$\begin{aligned} &\frac{\partial\theta_2}{\partial t}=\pi\varepsilon_f, \quad \frac{\partial Q_2}{\partial t}=-\pi\varepsilon_q, \qquad \pi\equiv\pi(t_n), \\ &\theta_2(t_n)=\theta_1(t_{n+r}), \qquad Q_2(t_n)=Q(t_n); \end{aligned} \tag{79.7}$$

(c) for the solution of diffusion equations,

$$\begin{aligned}
&\frac{\partial U_3}{\partial t}=\pi F_u, \qquad \frac{\partial V_3}{\partial t}=\pi F_v, \\
&\qquad\qquad\qquad\qquad\qquad\qquad \pi\equiv\pi(t_n) \\
&\frac{\partial \theta_3}{\partial t}=\pi F_T, \qquad \frac{\partial Q_3}{\partial t}=\pi F_q, \\
&V_3(t_n)=V(t_n), \qquad V_3(t_n)=V(t_n), \\
&\theta_3(t_n)=\theta_2(t_{n+1}), \qquad Q_3(t_n)=Q_2(t_{n+1});
\end{aligned} \tag{79.8}$$

(d) for the solution of transport equations,

$$\begin{aligned}
&\frac{\partial U_4}{\partial t}+K_s U_4=0, \qquad \frac{\partial V_4}{\partial t}+K_s V_4=0, \\
&\qquad\qquad\qquad\qquad\qquad\qquad \pi\equiv\pi(t_n), \\
&\frac{\partial \theta_4}{\partial t}+K_s\theta_4=0, \qquad \frac{\partial Q_4}{\partial t}+K_s Q_4=0, \\
&U_4(t_n)=U_3(t_{n+1}), \qquad V_4(t_n)=V_3(t_{n+1}), \\
&\theta_4(t_n)=\theta_3(t_{n+1}), \qquad Q_4(t_n)=Q_3(t_{n+1});
\end{aligned} \tag{79.9}$$

(e) for the solution of adaptation equations,

$$\begin{aligned}
&\frac{\partial U_5}{\partial t}-\left(l+\frac{U_5}{a\pi_5}\operatorname{tg}\varphi\right)V_5+\frac{1}{a\cos\varphi}\left(\pi_5\frac{\partial\Phi}{\partial\lambda}+2R\theta_5\frac{\partial}{\partial\lambda}\ln\pi_5\right)=0, \\
&\frac{\partial V_5}{\partial t}+\left(l+\frac{U_5}{a\pi_5}\operatorname{tg}\varphi\right)U_5+\frac{1}{a}\left(\pi_5\frac{\partial\Phi}{\partial\varphi}+2R\theta_5\frac{\partial}{\partial\varphi}\ln\pi_5\right)=0, \\
&\frac{\partial\theta_5}{\partial t}-\frac{R\theta_5}{C_p\pi_5^2}\tau=0, \\
&\frac{\partial\pi_5}{\partial t}+\frac{1}{2\pi_5}\int_0^1 D\,d\sigma=0, \qquad Q_5\equiv Q_4(t_{n+1}), \\
&U_5(t_n)=U_4(t_{n+1}), \qquad V_5(t_n)=V_4(t_{n+1}), \qquad \theta_5(t_n)=\theta_4(t_{n+1}), \\
&\pi_5(t_n)=\pi(t_n).
\end{aligned} \tag{79.10}$$

Formulae (78.31), (78.7), (78.33) are used for computing Φ, τ, D and $\dot{\sigma}$.

A formal analysis of the above splitting scheme yields a first-order approximation in time even in the case when the Crank–Nicolson scheme is used as a basic scheme on each splitting stage. We shall consider stability questions in the following sections.

80. The approximation of equations in spatial variables and the discrete analogues of conservation laws

In Section 78 we have described the equations of atmosphere hydrothermodynamics in symmetrized form. The purpose of such a symmetrization is, first of all, the possibility to construct absolutely stable difference schemes, whose stability is proved in a quadratic norm equivalent to the system's square energy. For illustration, consider a one-dimensional transport problem

$$\frac{\partial\theta}{\partial t}+\frac{1}{2}\left(u\frac{\partial\theta}{\partial x}+\frac{\partial u\theta}{\partial x}\right)=0. \tag{80.1}$$

It is easily seen that for (80.1) the relationship

$$\frac{\partial}{\partial t}\int_G \theta^2\,dG=0$$

holds if u becomes zero at the boundary of domain G or if θ is periodic. This follows from the skew-symmetricity of the operator

$$K_s\theta\equiv\frac{1}{2}\left(u\frac{\partial\theta}{\partial x}+\frac{\partial u\theta}{\partial x}\right).$$

Finite difference approximations of operator K_s will also retain the skew-symmetricity condition (skew-symmetric matrices will appear) if the same symmetric uniform scheme is used for each of the terms $u\partial\theta/\partial x$ and $\partial u\theta/\partial x$. Thus we can write

$$\frac{\partial\theta^h}{\partial t}+K_s^h\theta^h=0,$$

where

$$(K_s^h\theta^h,\theta^h)_\Phi=0, \tag{80.2}$$

and the inner product

$$(\varphi^h,\psi^h)_\Phi=\sum_i \varphi_i^h\psi_i^h$$

is chosen in a corresponding grid space Φ.

The operator K_s^h retains its skew-symmetric form irrespective of the choice of u^h, i.e. it is independent of the type of projector transforming u into u^h. The independence of the skew-symmetricity property of the type of projection operator allows the construction of a family of schemes satisfying condition (80.2) for problem (80.1). Using the Crank–Nicolson scheme in time we obtain (index h is omitted)

$$\frac{\theta^{n+1}-\theta^n}{\Delta t}+K_s\frac{\theta^{n+1}+\theta^n}{2}=0. \tag{80.3}$$

Taking the inner product of (80.3) and $\frac{1}{2}(\theta^{n+1}+\theta^n)$ we obtain

$$(\theta^{n+1},\theta^{n+1})_\Phi=(\theta^n,\theta^n)_\Phi,$$

which means that problem (80.1) is stable.

If we choose staggered grids where function θ is determined on integer grid points and the function u on fractional grid points, then a scheme of second-order accuracy in space for (80.1) can be represented as follows

$$\frac{\theta_i^{n+1}-\theta_i^n}{\Delta t}+\frac{u_{i+1/2}\,\theta_{i+1}^{n+1/2}-u_{i-1/2}\,\theta_{i-1}^{n+1/2}}{2\Delta x}=0 \tag{80.4}$$

and will have the property that

$$\sum_i(\theta_i^{n+1})^2=\sum_i(\theta_i^n)^2. \tag{80.5}$$

It is fairly simple to write the approximation in space variables and the scheme for the three-dimensional transport equation in the general case:

$$\begin{aligned}&\frac{\theta_{ijk}^{n+1}-\theta_{ijk}^n}{\Delta t}+\frac{1}{2a\cos\varphi_j\Delta\lambda}\left[u_{i+1/2,j,k}^n\,\theta_{i+1,j,k}^{n+1/2}-u_{i-1/2,j,k}^n\,\theta_{i-1,j,k}^{n+1/2}\right]\\&+\frac{1}{2a\cos\varphi_j\Delta\varphi}\left[(v\cos\varphi)_{i,j+1/2,k}^n\,\theta_{i,j+1,k}^{n+1/2}-(v\cos\varphi)_{i,j-1/2,k}^n\,\theta_{i,j-1,k}^{n+1/2}\right]\\&+\frac{1}{2\Delta\sigma}\left(\dot\sigma_{i,j,k+1/2}^n\,\theta_{i,j,k+1}^{n+1/2}-\dot\sigma_{i,j,k-1/2}^n\,\theta_{i,j,k-1}^{n+1/2}\right)=0,\end{aligned} \tag{80.6}$$

where

$$u_{i+1/2,j,k}=\tfrac{1}{2}(u_{i+1/2,j+1/2,k}+u_{i+1/2,j-1/2,k}),$$

$$(v\cos\varphi)_{i,j+1/2,k}=\tfrac{1}{2}((v\cos\varphi)_{i+1/2,j+1/2,k}+(v\cos\varphi)_{i-1/2,j+1/2,k}).$$

The grid points where the functions u and v are determined have indices $i+\frac{1}{2},j+\frac{1}{2}$ and k; the grid points where a function $\dot\sigma$ is determined have indices i,j and $k+\frac{1}{2}$.

It is easily seen that the matrices for each difference operator approximating the derivatives along λ,φ and σ are skew-symmetric, i.e. (80.6) can be written in vector form,

$$\frac{\theta^{n+1}-\theta^n}{\Delta t}+(K_s^\lambda+K_s^\varphi+K_s^\sigma)\theta^{n+1/2}=0, \tag{80.7}$$

where

$$(K_s^\mu\theta,\theta)_\Phi=0,\quad \mu=\lambda,\varphi,\sigma. \tag{80.8}$$

Transport problems can be represented in form (80.7)–(80.8) for other substances as well: specific humidity and horizontal velocity components.

At the adaptation stage the scheme has the form:

$$\begin{aligned}&\frac{U_{i+1/2,j+1/2,k}^{n+1}-U_{i+1/2,j+1/2,k}^n}{\Delta t}-\left(l_{j+1/2}+\frac{U_{i+1/2,j+1/2,k}^{n+1/2}}{a\pi_{i+1/2,j+1/2}^{n+1/2}}\operatorname{tg}\varphi_{j+1/2}\right)\\&\times V_{i+1/2,j+1/2,k}^{n+1/2}+\frac{1}{a\cos\varphi_j\Delta\lambda}\left[\pi_{i+1/2,j+1/2}^{n+1/2}\nabla_i\Phi_{j+1/2,k}^{n+1/2}\right.\\&\left.+2R\theta_{i+1/2,j+1/2,k}^{n+1/2}\times\nabla_i(\ln\pi)_{j+1/2}^{n+1/2}\right]=0,\end{aligned} \tag{80.9a}$$

$$\frac{V^{n+1}_{i+1/2,j+1/2,k}-V^{n}_{i+1/2,j+1/2,k}}{\Delta t}+\left(l_{j+1/2}+\frac{U^{n+1/2}_{i+1/2,j+1/2,k}}{a\pi^{n+1/2}_{i+1/2,j+1/2}}\operatorname{tg}\varphi_{j+1/2}\right)$$

$$\times U^{n+1/2}_{i+1/2,j+1/2,k}+\frac{1}{a\Delta\varphi}[\pi^{n+1/2}_{i+1/2,j+1/2}\nabla_i\Phi^{n+1/2}_{j+1/2,k} \tag{80.9b}$$

$$+2R\theta^{n+1/2}_{i+1/2,j+1/2,k}\times\nabla_j(\ln\pi)^{n+1/2}_{i+1/2}]=0,$$

$$\frac{\pi^{n+1}_{ij}-\pi^{n}_{ij}}{\Delta t}+\frac{1}{2\pi^{n+1/2}_{ij}}\sum_{k=1}^{K}D^{n+1/2}_{ijk}\Delta\sigma=0, \tag{80.9c}$$

where

$$D_{ijk}=\frac{1}{a\cos\varphi_j\Delta\lambda}\left[\nabla_{i-1/2}\pi_j U_{jk}+\frac{\Delta\lambda}{\Delta\varphi}\nabla_{j-1/2}\pi_i\cos\varphi V_{ik}\right], \tag{80.9d}$$

$$\frac{\theta^{n+1}_{ijk}-\theta^{n}_{ijk}}{\Delta t}-\frac{R}{2C_p a\cos\varphi_j\pi^{n+1/2}_{ij}}\sum_{\nu=-1,0}\left[\frac{1}{\Delta\lambda}\theta^{n+1/2}_{i+1/2+\nu,j,k}U^{n+1/2}_{i+1/2+\nu,j,k}\right.$$

$$\times\nabla_{i+\nu}(\ln\pi)^{n+1/2}_{j}+\frac{1}{\Delta\varphi}\theta^{n+1/2}_{i,j+1/2+\nu,k}V^{n+1/2}_{i,j+1/2+\nu,k}\cos\varphi_{j+1/2+\nu}\nabla_{j+\nu}$$

$$\left.\times(\ln\pi)^{n+1/2}_{i}\right]-\frac{1}{C_p\pi^{n+1/2}_{ij}\Delta\sigma}\sum_{\nu=-1,0}\left[(\Phi^{n+1/2}_{i,j,k+1/2+\nu/2}-\right.$$

$$\left.\Phi^{n+1/2}_{i,j,k+\nu/2})\sum_{\mu+1}^{k+\nu}D^{n+1/2}_{ij\mu}\Delta\sigma\right]=0, \tag{80.9e}$$

$$\Phi_{ijk}-\frac{1}{\Delta\sigma}\nabla_{k-1/2}(\Phi_{ij}\sigma)=\frac{R\theta_{ijk}}{\pi_{ij}}, \tag{80.9f}$$

$$\nabla_r=()_{r+1}-()_r,\qquad ()_{r+1/2}=\tfrac{1}{2}[()_{r+1}+()_r],$$

r is any of the indices i,j,k, or n.

Let $z_s\equiv 0$ and let the term RT in (78.3) be equal to $R\tilde{T}$. Hence, multiplying (80.9a) by $U^{n+1/2}_{i+1/2,j+1/2,k}$, and (80.9b) by $V^{n+1/2}_{i+1/2,j+1/2,k}$, (80.9e) by $\theta^{n+1/2}_{ijk}$ and summing over all i, j and k, we obtain the relationship

$$\sum_{i,j,k}\cos\varphi_j\left(U^2_{i+1/2,j+1/2,k}+V^2_{i+1/2,j+1/2,k}+\frac{C_p\theta^2_{ijk}}{\tilde{T}}\right)^{n+1}$$

$$=\sum_{i,j,k}\cos\varphi_j\left(U^2_{i+1/2,j+1/2,k}+V^2_{i+1/2,j+1/2,k}+\frac{C_p\theta^2_{ijk}}{\tilde{T}}\right)^{n}, \tag{80.10}$$

which is, in fact, the difference analogue of the conservation law (78.26).

81. The method of splitting with respect to geometric variables and numerical realization

Since the transport operator K_s is the sum of three operators (transport in each geometric direction), each being skew-symmetric, it is quite natural to perform

a further splitting of the problem. Hence, we obtain the following algorithm:

$$\begin{aligned}
&\frac{\psi^{n+1/3}-\psi^n}{\Delta t}+K_s^{\lambda}\frac{\psi^{n+1/3}+\psi^n}{2}=0,\\
&\frac{\psi^{n+2/3}-\psi^{n+1/3}}{\Delta t}+K_s^{\varphi}\frac{\psi^{n+2/3}+\psi^{n+1/3}}{2}=0,\\
&\frac{\psi^{n+1}-\psi^{n+2/3}}{\Delta t}+K_s^{\sigma}\frac{\psi^{n+1}+\psi^{n+2/3}}{2}=0,
\end{aligned} \tag{81.1}$$

where ψ is any of the functions U, V, θ or Q.

Due to (80.8) it is obvious that the above difference scheme is absolutely stable and quadratic functionals conserve their value exactly. In the general case, scheme (81.1) has a first-order approximation in time. The cyclic commutation method allows, however, the construction of a second-order approximation scheme in the general case of noncommuting operators as well (MARCHUK [1980]).

Thus, a three-dimensional transport problem is reduced to a set of one-dimensional problems.

Now we consider methods for inverting these operators. Transport in the λ-direction (along latitude circles) is realized by the cyclic sweep method. It is the same for all unknown functions (the only difference is in determining the "transport" velocities, which are parts of the expression for operator K_s). The algorithm for transport in the σ-direction is also the same for all unknown functions (scalar sweep). Transport in the φ-direction (along longitudes) is realized in a more complex way. A method for constructing cyclically closed rings of meridians shifted by 180° has been used by MARCHUK, DYMNIKOV et al. [1984]. In this case components of the vector values (U and V) in the transition over the Pole must change their sign while the scalar values (θ and Q) keep the same sign. Computation is realized via a cyclic sweep. The skew-symmetric form of the transport operator guarantees, in this case, conservation of the quadratic values.

To solve the system of algebraic equations at the adaptation stage (80.9) the following iterative process with respect to values $U^{n+1/2}, V^{n+1/2}, \theta^{n+1/2}$ and π^{n+1} is used (some aspects of its convergence and the choice of optimal parameters are discussed by MARCHUK, DYMNIKOV et al. [1984]):

$$\Phi_{ijk}^{(s)}-\frac{1}{\Delta\sigma}\nabla_{k-1/2}(\sigma\Phi_{ij}^{(s)})=\frac{R\theta_{ijk}^{(s)}}{\pi_{ij}^{(s)}}, \tag{81.2}$$

$$A_{i+1/2,j+1/2,k}^{(s)}=\tfrac{1}{2}\Delta t\left[l_{j+1/2}+\frac{U_{i+1/2,j+1/2,k}^{(s)}\operatorname{tg}\varphi_{j+1/2}}{a\pi_{ij}^{(s)}}\right],$$

$$\begin{aligned}
(F_1)_{i+1/2,j+1/2,k}^{n,(s)}=U_{i+1/2,j+1/2,k}^{n}-\frac{\Delta t}{2a\cos\varphi_{j+1/2}\Delta\lambda}[&\pi_{i+1/2,j+1/2}^{(s)}\\
&\times\nabla_i\Phi_{j+1/2,k}^{(s)}+2R\theta_{i+1/2,j+1/2,k}^{(s)}\nabla_i(\ln\pi)_{j+1/2}^{(s)}],\\
(F_2)_{i+1/2,j+1/2,k}^{n,(s)}=V_{i+1/2,j+1/2,k}^{n}-\frac{\Delta t}{2a\Delta\varphi}[&\pi_{i+1/2,j+1/2}^{(s)}\nabla_j\Phi_{i+1/2,k}^{(s)}\\
&+2R\theta_{i+1/2,j+1/2,k}^{(s)}\nabla_j(\ln\pi)_{i+1/2}^{(s)}],
\end{aligned} \tag{81.3}$$

$$U^{(s+1)}_{i+1/2,j+1/2,k} - A^{(s)}_{i+1/2,j+1/2,k} V^{(s+1)}_{i+1/2,j+1/2,k} = (F_1)^{n,(s)}_{i+1/2,j+1/2,k},$$
$$A^{(s)}_{i+1/2,j+1/2,k} U^{(s+1)}_{i+1/2,j+1/2,k} + V^{(s+1)}_{i+1/2,j+1/2,k} = (F_2)^{n,(s)}_{i+1/2,j+1/2,k}, \tag{81.4}$$

$$D^{(s+1)}_{ijk} = \frac{1}{a\cos\varphi_j \Delta\lambda}\left[\nabla_{i-1/2}(\pi^{(s)}_j U^{(s+1)}_{j,k}) + \frac{\Delta\lambda}{\Delta\varphi}\nabla_{j-1/2}(\pi^{(s)}_i \cos\varphi\, V^{(s+1)}_{i,k})\right], \tag{81.5}$$

$$\theta^{(s+1)}_{ijk} = (1-\tau_\theta)\theta^{(s)}_{ijk} + \tau_\theta\left\{\theta^n_{ijk} + \frac{\Delta t R}{4C_p \pi^{(s)}_{ij} a\cos\varphi_j} \cdot \sum_{\nu=-1,0}\left[\frac{1}{\Delta\lambda}\theta^{(s)}_{i+1/2+\nu,j,k} U^{(s+1)}_{i+1/2+\nu,j,k}\nabla_{i+\nu}(\ln\pi)^{(s)}_j + \frac{1}{\Delta\varphi}\cos\varphi_{j+1/2+\nu}\theta^{(s)}_{i,j+1/2+\nu,k} V^{(s+1)}_{i,j+1/2+\nu,k}\nabla_{j+\nu}(\ln\pi)^{(s)}_i\right] + \frac{\Delta t}{2C_p\pi^{(s)}_{ij}}\sum_{\nu=-1,0}\left[(\Phi^{(s)}_{i,j,k+(\nu+1)/2} - \Phi^{(s)}_{i,j,k+\nu/2})\sum_{\mu=1}^{k+\nu} D^{(s+1)}_{ij\mu}\right]\right\}, \tag{81.6}$$

$$\pi^{(s+1)}_{ij} = (1-\tau_\pi)\pi^{(s)}_{ij} + \tau_\pi\left(\pi^n_{ij} - \frac{\Delta t\Delta\sigma}{2\pi^{(s)}_{ij}}\sum_{k=1}^{K} D^{(s+1)}_{ijk}\right). \tag{81.7}$$

The following notations are introduced in (81.2)–(81.7): s is the iteration number (indices $(n+1)$ and $(n+\frac{1}{2})$ are omitted); τ_θ and τ_π are relaxation parameters in θ and π (at the resolution $\Delta\lambda = \Delta\varphi = 5^\circ$ the optimal value of the parameters τ_θ and τ_π is the value ≈ 0.25).

The computational process is organized as follows. The distributions of U, V, θ and π on the preceding step in time are taken as the initial approximation (when $s=0$). Distribution of the geopotential Φ is computed using the known "old" values θ and π and equation (81.2). Auxiliary values A, F_1 and F_2 are computed using formulae (81.3). Here, in computing F_1 in polar regions the short waves in zonal gradients of functions $\ln\pi$ and Φ are filtered by means of a fast Fourier transformation. By solving system (81.4) we find values U and V by explicit formulae. Then, substituting them into (81.5) we compute the flat divergence D. Finally, using relationships (81.6) and (81.7), we compute "new" values θ and π (in this case, as above, we filter out short waves in zonal gradients $\sqrt{\pi}U$ and $\ln\pi$).

Note that the accepted procedure of filtering short waves does not break conservation laws. We perform iterations at each step in time until either the following relationship holds

$$\max_{i,j,k} |\theta^{(s+1)}_{ijk} - \theta^{(s)}_{ijk}| < \varepsilon,$$

where ε is a given value of absolute error, or the number of iterations exceeds a given limit value.

CHAPTER XX

Problems in Oceanology

The problem of mathematically modeling the ocean's general circulation attracts many scientists. The progress in the research into the physics of the ocean stimulated mathematicians to construct numerical models of ocean circulations, tide waves, currents, etc. These models influenced significantly the development of research into the central problem of modern geophysical hydrodynamics, the interaction of the atmosphere and the ocean.

This chapter describes the major aspects of the construction of numerical algorithms for the solution of ocean dynamics problems based on the splitting method (MARCHUK [1974a], MARCHUK, DYMNIKOV et al. [1984]).

82. Statement of the problem and the splitting of equations with respect to physical processes

The problem of the general circulation of the ocean is described by the following system of equations written in Cartesian coordinates for the sake of simplicity:

$$
\begin{aligned}
&\frac{du}{dt} + lu = -\frac{1}{\rho_0}\frac{\partial \rho}{\partial y} + \mu \Delta v + \frac{\partial}{\partial z}\kappa \frac{\partial v}{\partial z},\\
&\frac{dV}{dt} - lv = -\frac{1}{\rho_0}\frac{\partial \rho}{\partial x} + \mu \Delta u + \frac{\partial}{\partial z}\kappa \frac{\partial u}{\partial z},\\
&\frac{\partial \rho}{\partial z} = g\rho, \qquad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0,\\
&\frac{dT}{dt} + \gamma_T w = \mu_T \Delta T + \frac{\partial}{\partial z}\kappa_T \frac{\partial T}{\partial z},\\
&\frac{dS}{dt} + \gamma_S w = \mu_S \Delta S + \frac{\partial}{\partial z}\kappa_S \frac{\partial S}{\partial z},\\
&\rho = \rho(T, S), \qquad \frac{d}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}.
\end{aligned}
\tag{82.1}
$$

System (82.1) is obtained from the general equations of the hydrodynamics of a rotating fluid by using traditional approximations for problems of the large-scale

dynamics of the ocean, namely, Bousinesq, hydrostatics, incompressibility of the water, linear mechanism of macroturbulent viscosity and diffusion.

Equations (82.1) are written for deviations $\bar{p}(z)$, $\bar{\rho}(z)$, $\bar{T}(z)$, $\bar{S}(z)$ of pressure p, density ρ, temperature T, and salinity S respectively from their average values on the vertical with

$$\gamma_T=\frac{d\bar{T}}{dz}, \quad \gamma_S=\frac{d\bar{S}}{dz}.$$

We assume that the solution domain Ω of the problem represents a closed basin, bounded by the unperturbed surface of the ocean $z=0$, bottom $H=H(x,y)$ and cylindric lateral surface σ. We add the corresponding boundary and initial conditions to (82.1):

$$\begin{aligned}
&u=v=0, \quad \frac{\partial T}{\partial v}=\frac{\partial S}{\partial v}=0 \qquad \text{on } \sigma,\\
&\kappa\frac{\partial u}{\partial z}=f_1, \qquad \kappa\frac{\partial v}{\partial z}=f_2, \qquad w=0,\\
&a_1\frac{\partial T}{\partial z}+b_1 T=C_1,\\
&\qquad\qquad \text{for } z=0;\\
&a_2\frac{\partial S}{\partial z}+b_2 S=C_2
\end{aligned} \tag{82.2}$$

$$\begin{aligned}
&\kappa\frac{\partial u}{\partial z}=0, \qquad \kappa\frac{\partial v}{\partial z}=0,\\
&\qquad\qquad\qquad\qquad \text{for } z=H,\\
&w=u\frac{\partial H}{\partial z}+v\frac{\partial H}{\partial y}, \qquad \frac{\partial T}{\partial v}=\frac{\partial S}{\partial v}=0
\end{aligned} \tag{82.3}$$

$$u=u^0, \quad v=v^0, \quad T=T^0, \quad S=S^0 \qquad \text{for } t=0,$$

where v is an external normal vector to σ, and $a_i, b_i, c_i, f_i, u^0, v^0, T^0$ and S^0 are given coefficients and functions. The set of processes described by system (82.1) can conditionally be divided in three main types: transport of momentum, heat and salinity along the trajectories, their diffusion and adaptation of the mass fields and currents. To solve approximately problem (82.1)–(82.3) the splitting method with respect to the corresponding physical process can be used.

We pay no attention to methods for approximating (82.1) in spatial variables (see MARCHUK [1974a], MARCHUK, DYMNIKOV et al. [1984] and ZALESNY [1984, 1985]) and write a splitting system on time intervals $t_n \leqslant t \leqslant t_{n+1}$ in terms of differential equations. We have to solve three subsystems consecutively:

$$\begin{aligned}
&\frac{1}{3}\frac{\partial u}{\partial t}+u^n\frac{\partial u}{\partial x}+V^n\frac{\partial u}{\partial y}+w^n\frac{\partial u}{\partial z}=0,\\
&\frac{1}{3}\frac{\partial v}{\partial t}+u^n\frac{\partial v}{\partial x}+v^n\frac{\partial v}{\partial y}+w^n\frac{\partial v}{\partial z}=0,\\
&\frac{1}{3}\frac{\partial T}{\partial t}+u^n\frac{\partial T}{\partial x}+v^n\frac{\partial T}{\partial y}+w^n\frac{\partial T}{\partial z}=0,
\end{aligned} \tag{82.4}$$

$$\frac{1}{3}\frac{\partial S}{\partial t}+u^n\frac{\partial S}{\partial x}+v^n\frac{\partial S}{\partial y}+w^n\frac{\partial S}{\partial z}=0,$$

$$\frac{1}{3}\frac{\partial u}{\partial t}=\mu\Delta u+\frac{\partial}{\partial z}\kappa\frac{\partial u}{\partial z},$$

$$\frac{1}{3}\frac{\partial v}{\partial t}=\mu\Delta v+\frac{\partial}{\partial z}\kappa\frac{\partial v}{\partial z},$$

$$\frac{1}{3}\frac{\partial T}{\partial t}=\mu_T\Delta T+\frac{\partial}{\partial z}\kappa_T\frac{\partial T}{\partial z},\tag{82.5}$$

$$\frac{1}{3}\frac{\partial S}{\partial t}=\mu_S\Delta S+\frac{\partial}{\partial z}\kappa_S\frac{\partial S}{\partial z},$$

$$\frac{1}{3}\frac{\partial u}{\partial t}-lv=-\frac{1}{\rho_0}\frac{\partial \rho}{\partial x},$$

$$\frac{1}{3}\frac{\partial v}{\partial t}+lu=-\frac{1}{\rho_0}\frac{\partial \rho}{\partial y},$$

$$\frac{1}{3}\frac{\partial p}{\partial z}=g\rho,\qquad \frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0,\tag{82.6}$$

$$\frac{1}{3}\frac{\partial T}{\partial t}+\gamma_T w=0,\qquad \frac{1}{3}\frac{\partial S}{\partial t}+\gamma_S w=0,$$

$$\rho=\alpha_T T+\alpha_S S+\rho_1(T^n,S^n).$$

Note that in this case it is assumed, in splitting system (82.1), that equations (82.1) are linearized at the interval $t_j\leqslant t\leqslant t_{j+1}$. This is indicated by the indices n in (82.4) and in the equation of state in system (82.6). We add to the subsystems (82.4)–(82.6) the corresponding initial conditions and to (82.5)–(82.6) the boundary conditions following from (82.2).

For system (82.5) we have the following boundary conditions

$$u=v=0,\quad \frac{\partial T}{\partial \nu}=\frac{\partial S}{\partial \nu}=0\qquad \text{on } \sigma,$$

$$\kappa\frac{\partial u}{\partial z}=f_1,\qquad \kappa\frac{\partial v}{\partial z}=f_2,$$

$$a_1\frac{\partial T}{\partial z}+b_1 T=c_1,$$

$$\text{for } z=0,\tag{82.7}$$

$$a_2\frac{\partial S}{\partial z}+b_2 S=c_2,$$

$$\kappa\frac{\partial u}{\partial z}=0,\quad \kappa\frac{\partial v}{\partial z}=0,$$

$$\text{for } z=H;$$

$$\frac{\partial T}{\partial \nu}=\frac{\partial S}{\partial \nu}=0$$

and for system (82.6)

$$\begin{aligned} &(u,v)=0 \quad \text{on } \sigma, \\ &w=0 \quad \text{for } z=0, \\ &w=uH_x+vH_y \qquad \text{for } z=H, \end{aligned} \tag{82.8}$$

where

$$u=(u,v).$$

83. The splitting of adaptation equations in planes and generalization for the nonhydrostatic case

The equations solved by the methods considered above have been obtained on the first two stages of splitting with respect to physical processes, described by systems (82.4) and (82.5).

Now we consider a method for solving system (82.6) based on further splitting in the coordinate planes (x,z) and (y,z) (MARCHUK [1974a]). Because this method allows generalization, we assume that a more complete evolutionary equation is used instead of the hydrostatics equation. Hence, we arrive at the following two systems of equations, assuming for the sake of simplicity that the state equation is linear ($\rho_1=0$):

$$\begin{aligned} &\frac{1}{6}\frac{\partial u}{\partial t}-\tfrac{1}{2}lv=-\frac{1}{\rho_0}\frac{\partial \rho}{\partial x}, \\ &\frac{1}{6}\frac{\partial v}{\partial t}+\tfrac{1}{2}lu=0, \\ &\frac{1}{6}\frac{\partial w}{\partial t}-\tfrac{1}{2}g=-\frac{1}{2\rho}\frac{\partial p}{\partial z}, \\ &\frac{\partial u}{\partial x}+\frac{1}{2}\frac{\partial w}{\partial z}=0, \\ &\frac{1}{6}\frac{\partial \rho}{\partial t}+\tfrac{1}{2}\gamma w=0; \end{aligned} \tag{83.1}$$

$$\begin{aligned} &\frac{1}{6}\frac{\partial u}{\partial t}-\tfrac{1}{2}lv=0, \\ &\frac{1}{6}\frac{\partial v}{\partial t}+\tfrac{1}{2}lu=-\frac{1}{\rho_0}\frac{\partial p}{\partial y}, \\ &\frac{1}{6}\frac{\partial w}{\partial t}-\tfrac{1}{2}g=-\frac{1}{2\rho}\frac{\partial p}{\partial z}, \end{aligned} \tag{83.2}$$

$$\frac{\partial v}{\partial y}+\frac{1}{2}\frac{\partial w}{\partial z}=0,$$

$$\frac{1}{6}\frac{\partial \rho}{\partial t}+\tfrac{1}{2}\gamma w=0;$$

where $\gamma=\alpha_T\gamma_T+\alpha_S\gamma_S$.

By using continuity equations we can reduce each of systems (83.1) and (83.2) to problems for stream functions.

84. The splitting of adaptation equations with respect to "topography"

The above system of ocean dynamics equations (82.1) is non-evolutionary. That is why its splitting with respect to physical processes has been performed on a "physical level". To be able to use the results of the general theory of the splitting method given in the previous chapters, we reduce (82.1) to an evolutionary form. We will, however, not consider here the transformation of the full system (82.1) to its evolutionary form (see MARCHUK, DYMNIKOV et al. [1984] and ZALESNY [1984]). Let us illustrate this technique by transforming adaptation equations (82.6) under the assumption that the state condition ($\rho_1=0$) is linear.

We transform the solution domain Ω of the problem into a cylinder with unit base by substitution of $z_1=z/H$ and represent the horizontal components of the velocity vector in the form

$$u=\bar{u}+u', \qquad v=\bar{v}+v', \qquad \bar{a}=\int_0^1 a\,dz.$$

Hence, we obtain the following system of equations

$$\tfrac{1}{3}H\frac{\partial u}{\partial t}-Hiv=-\frac{1}{\rho_0}\left(H\frac{\partial \rho}{\partial x}-\frac{\partial H}{\partial x}z_1\frac{\partial p}{\partial z_1}\right),$$

$$\frac{1}{3}\frac{\partial v}{\partial t}+Hlu=-\frac{1}{\rho_0}\left(H\frac{\partial p}{\partial y}-\frac{\partial H}{\partial y}z_1\frac{\partial p}{\partial z_1}\right),$$

$$\frac{\partial p}{\partial z_1}=\rho Hg, \qquad \frac{\partial uH}{\partial x}+\frac{\partial vH}{\partial y}+\frac{\partial w_1}{\partial z_1}=0, \tag{84.1}$$

$$\frac{1}{3}\frac{\partial \rho}{\partial t}+\gamma\left[w_1+z_1\left(\frac{\partial H}{\partial x}u+\frac{\partial H}{\partial y}v\right)\right]=0,$$

$$\gamma=\alpha_T\gamma_T+\alpha_S\gamma_S, \qquad w_1=w-z_1\left(\frac{\partial H}{\partial x}u+\frac{\partial H}{\partial y}v\right).$$

Excluding functions ρ and w_1 from (84.1) we obtain, assuming $\gamma \geqslant \gamma_0 > 0$,

$$\begin{aligned}
&\tfrac{1}{3}\rho_0 H\frac{\partial u}{\partial t}-\rho_0 Hlv+\left(H\frac{\partial p}{\partial x}-\frac{\partial H}{\partial x}z_1\frac{\partial p}{\partial z_1}\right)=0,\\
&\tfrac{1}{3}\rho_0 H\frac{\partial v}{\partial t}+\rho_0 Hlu+\left(H\frac{\partial p}{\partial y}-\frac{\partial H}{\partial y}z_1\frac{\partial p}{\partial z_1}\right)=0,\\
&-\frac{1}{3}\frac{\partial}{\partial t}\frac{\partial}{\partial z_1}\frac{1}{gH_j}\frac{\partial p}{\partial z_1}-\frac{\partial}{\partial z_1}z_1\left(u\frac{\partial H}{\partial x}+v\frac{\partial H}{\partial y}\right)+\frac{\partial uH}{\partial x}+\frac{\partial vH}{\partial y}=0.
\end{aligned} \tag{84.2}$$

We rewrite system (84.2) in operator form

$$B\frac{\partial\varphi}{\partial t}+A\varphi=0, \quad A=A_1+A_2, \tag{84.3}$$

where

$$\varphi=(u,v,p),$$

$$B=\begin{pmatrix} \tfrac{1}{3}\rho_0 H & 0 & 0\\ 0 & \tfrac{1}{3}\rho_0 H & 0\\ 0 & 0 & -\frac{1}{3}\frac{\partial}{\partial z_1}\frac{1}{gH\gamma}\frac{\partial}{\partial z_1}\end{pmatrix},$$

$$A_1=\begin{pmatrix} 0 & -\rho_0 Hl & H\frac{\partial}{\partial x}\\ \rho_0 Hl & 0 & H\frac{\partial}{\partial y}\\ \frac{\partial}{\partial x}H & \frac{\partial}{\partial y}H & 0\end{pmatrix}, \quad A_2=\begin{pmatrix} 0 & 0 & -\frac{\partial H}{\partial x}z_1\frac{\partial}{\partial z_1}\\ 0 & 0 & -\frac{\partial H}{\partial y}z_1\frac{\partial}{\partial z_1}\\ -\frac{\partial}{\partial z_1}z_1\frac{\partial H}{\partial x} & -\frac{\partial}{\partial z_1}z_1\frac{\partial H}{\partial y} & 0\end{pmatrix},$$

For the approximate solution of (84.3) we use a splitting scheme with respect to "topography" (ZALESNY [1984, 1985])

$$\begin{aligned}
&\tfrac{1}{2}B\frac{\partial\varphi_1}{\partial t}+A_1\varphi_1=0,\\
&\tfrac{1}{2}B\frac{\partial\varphi_2}{\partial t}+A_2\varphi_2=0.
\end{aligned} \tag{84.4}$$

Hence, at the first stage of splitting we come to the solution of

$$\begin{aligned}
&\tfrac{1}{6}\rho_0 H\frac{\partial u}{\partial t}-\rho_0 Hlv+H\frac{\partial p}{\partial x}=0,\\
&\tfrac{1}{6}\rho_0 H\frac{\partial v}{\partial t}+\rho_0 Hlu+H\frac{\partial p}{\partial y}=0,
\end{aligned} \tag{84.5}$$

$$-\frac{1}{6}\frac{\partial}{\partial t}\frac{\partial}{\partial z_1}\frac{1}{gH_\gamma}\frac{\partial p}{\partial z_1}+\frac{\partial uH}{\partial x}+\frac{\partial vH}{\partial y}=0.$$

We add to system (84.5) the boundary conditions following from (82.8):

$$w=0 \quad \text{for } z_1=0,1;$$

or in terms of function p:

$$\begin{aligned} &-\frac{\partial}{\partial t}\frac{1}{gH_\gamma}\frac{\partial p}{\partial z_1}=0 \quad \text{for } z_1=0,1,\\ &(u,v)=0 \quad \text{on } \sigma. \end{aligned} \tag{84.6}$$

At the second stage of splitting we have

$$\begin{aligned} &\tfrac{1}{6}\rho_0 H\frac{\partial u}{\partial t}-\frac{\partial H}{\partial x_1}z_1\frac{\partial p}{\partial z_1}=0,\\ &\tfrac{1}{6}\rho_0 H\frac{\partial v}{\partial t}-\frac{\partial H}{\partial y_1}z_1\frac{\partial p}{\partial z_1}=0,\\ &-\frac{1}{6}\frac{\partial}{\partial t}\frac{\partial}{\partial z_1}\frac{1}{gH_\gamma}\frac{\partial p}{\partial z_1}-\frac{\partial}{\partial z_1}z_1\left(u\frac{\partial H}{\partial x}+v\frac{\partial H}{\partial y}\right)=0 \end{aligned} \tag{84.7}$$

with boundary conditions

$$\begin{aligned} &-\frac{\partial}{\partial t}\frac{1}{gH_\gamma}\frac{\partial p}{\partial z_1}=0 \quad \text{for } z=0,\\ &-\frac{\partial}{\partial t}\frac{1}{gH_\gamma}\frac{\partial p}{\partial z_1}=uH_x+vH_y \quad \text{for } z=1. \end{aligned} \tag{84.8}$$

It is easily shown that relationships

$$B\geqslant 0,$$

$$(A_i\varphi,\varphi)=0, \quad i=1,2 \qquad \text{for } \gamma\neq 0$$

hold for operators B, A_1 and A_2 (with the corresponding boundary conditions).

85. The splitting of "shallow water" equations with respect to coordinates

System (84.5) obtained at the first splitting stage with respect to "topography" may be transformed to a linear system of the "shallow water" equations:

$$\begin{aligned} &\tfrac{1}{6}\rho_0 H\frac{\partial u_n}{\partial t}-\rho_0 Hlv_n+H\frac{\partial p_n}{\partial x}=0,\\ &\tfrac{1}{6}\rho_0 H\frac{\partial v_n}{\partial t}+\rho_0 Hlu_n+H\frac{\partial p_n}{\partial y}=0,\\ &\tfrac{1}{6}\lambda_n\frac{\partial p_n}{\partial t}+gH\left(\frac{\partial u_nH}{\partial x}+\frac{\partial v_nH}{\partial y}\right)=0. \end{aligned} \tag{85.1}$$

Here λ_n are eigenvalues of a self-adjoint spectral problem,

$$-\frac{\partial}{\partial z}\frac{1}{\gamma}\frac{\partial \varphi_n}{\partial z}=\lambda_n \varphi_n,$$

$$\frac{\partial \varphi_n}{\partial z}=0 \quad \text{for } z=0,1$$

with $\lambda_0=0, \lambda_i>0, i=1,2,\ldots$.

System (85.1) can also be split along the coordinates x and y into two subsystems:

$$\begin{aligned}
&\tfrac{1}{12}\rho_0 H\frac{\partial u_n}{\partial t}-\tfrac{1}{2}\rho_0 H l v_n+H\frac{\partial p_n}{\partial x}=0,\\
&\tfrac{1}{12}\rho_0 H\frac{\partial v_n}{\partial t}+\tfrac{1}{2}\rho_0 H l u_n=0,\\
&\tfrac{1}{12}\lambda_n\frac{\partial p_n}{\partial t}+gH\frac{\partial u_n H}{\partial x}=0;
\end{aligned} \tag{85.2}$$

$$\begin{aligned}
&\tfrac{1}{12}\rho_0 H\frac{\partial u_n}{\partial t}-\tfrac{1}{2}\rho_0 H l v_n=0,\\
&\tfrac{1}{12}\rho_0 H\frac{\partial v_n}{\partial t}+\tfrac{1}{2}\rho_0 H l u_n+H\frac{\partial p_n}{\partial y}=0,\\
&\tfrac{1}{12}\lambda_n\frac{\partial p_n}{\partial t}+gH\frac{\partial v_n H}{\partial y}=0.
\end{aligned} \tag{85.3}$$

Systems (85.2) and (85.3), or to be exact, their discrete analogues are realized by simple three-point sweeps.

Note that in the absence of rotation effect ($l=0$) the system coincides with acoustics equations in the two-dimensional case with the accuracy up to the notations. A method of splitting with respect to coordinates analogous to (85.2)–(85.3) with concrete discretization of the problem in spatial variables has been considered by YANENKO [1967].

One can find various versions of splitting methods for ocean dynamics problems as well as algorithms for their realizations and results of numerical experiments in the works by MARCHUK [1974a], MARCHUK, BUBNOV, ZALESNY and KORDZADZE [1983], MARCHUK, KORDZADZE and SKIBA [1974], MARCHUK et al. [1980], MARCHUK and ZALESNY [1974], KUZIN [1980, 1984] and ZALESNY [1984, 1985].

References

ABRAMOV, A.A. and V.B. ANDREEV (1963), An application of the sweep method for finding periodic solutions of differential and difference equations, *Zh. Vychisl. Mat. i Mat. Fiz.* 3 (2) (in Russian).

ALBRECHT, I. (1965), Eine Verallgemeinerung des Verfahrens der Alternierenden, *Z. Angew. Math. Mech.* **45** (3–6) (Sonderheft).

AMSDEN, A.A. and F.H. HARLOW (1970), A simplified MAC technique for incompressible fluid flow calculations, *J. Comput. Phys.* **6** (2), 322–325.

ANDERSON, D.A. (1974), A comparison of numerical solutions to the inviscid equations of fluid motion, *J. Comput. Phys.* **15** (1), 1–20.

ANDREEV, V.B. (1967), On the splitting difference schemes for general p-dimensional parabolic equations of second order with mixed derivatives, *Zh. Vychisl. Mat. i Mat. Fiz.* **7** (2) (in Russian).

ANDREEV, V.B. (1969), On the convergence of the splitting difference schemes approximating the third boundary value problem for parabolic equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (2) (in Russian).

ANUCHINA, N.N. (1970), On the methods for computing compressible liquid flows with large deformations, *Chisl. Metody Mekh. Sploshn. Sredy Inform. Bul.* **1** (4) (in Russian).

ANUCHINA, N.N. and N.N. YANENKO (1959), The implicit splitting schemes for hyperbolic equations and systems, *Dokl. Acad. Sci. USSR* **128** (6) (in Russian).

BABUŠKA, I. (1970), Approximation by hill functions, Tech. Note BN-648, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, MD.

BAGRINOVSKY, K.A. and S.K. GODUNOV (1957), Difference schemes for multidimensional problems, *Dokl. Acad. Sci. USSR* **115** (3) (in Russian).

BAKER, G.A. (1960), An implicit numerical method for solving the n-dimensional heat equation, *Quart. Appl. Math.* **17** (4), 440–443.

BAKER, G.A. and T.A. OLIPHANT (1960), An implicit numerical method for solving the two-dimensional heat equation, *Quart. Appl. Math.* **17** (4).

BALDWIN, B.S. and R.W. MACCORMACK (1974), Numerical solution of the interaction of a strong shock wave with a hypersonic turbulent boundary layer, AIAA Paper 558, New York.

BAUM, E. and E. NDEFO (1973), A temporal ADI computational technique, in: *Proceedings AIAA Computers and Fluid Dynamics Conference*, Springs, CA (AIAA, New York) 133–140.

BELOTSERKOVSKY, O.M. (1984), *Numerical Modelling in the Mechanics of Continuous Media* (Nauka, Moscow) (in Russian).

BELOTSERKOVSKY, O.M. and Yu.M. DAVYDOV (1970), The method of "large particles" for the problems in dynamics, *Chysl. Metody Mekh. Sploshn. Sredy Inform. Bul.* 1 (3) (in Russian).

BELOTSERKOVSKY, O.M. and Yu.M. DAVYDOV (1978), The method of "large particles" (schemes and applications), Moscow Physical Technical Institute, Moscow, USSR.

BELOTSERKOVSKY, O.M., Yu.P. GOLOVACHEV, V.G. GRUDNITSKY, Yu.M. DAVYDOV, V.K. DUSHIN, Yu.P. LUNKIN, K.M. MAGOMEDOV, V.K. MOLODTSOV, F.D. POPOV, A.I. TOLSTYKH, V.N. FOMIN and A.S. KHOLODOV (1974), *Numerical Study of Modern Problems in Gas Dynamics* (Nauka, Moscow).

BELOTSERKOVSKY, O.M., V.A. GUSHCHIN and V.V. SHCHENNIKOV (1975), The splitting method in application to solving problems in the dynamics of viscous incompressible liquid, *Zh. Vychisl. Mat. i Mat. Fiz.* **15** (1) (in Russian).

BELOTSERKOVSKY, O.M., F.D. POPOV, A.I. TOLSTYKH, V.N. FOMIN and A.S. KHOLODOV (1970), Numerical solution of some problems in gas dynamics, *Zh. Vychisl. Mat. i Mat. Fiz.* **10** (2) (in Russian).

BELUKHINA, I.G. (1969), The difference schemes for solving two-dimensional dynamic elasticity problem with mixed boundary conditions, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (2) (in Russian).

BENSOUSSAN, A. (1971), Pure decentralization for interrelated payoffs, in: *Proceedings Symposium on Optimization*, Los Angeles, CA.

BENSOUSSAN, A., J.-L. LIONS and R. TEMAM (1975), *Methods of Decomposition, Decentralization, Coordination and Their Applications*, Metody Vychisl. Mat. (Nauka, Novosibirsk).

BEREZIN, Yu.A., V.M. KOVENYA and N.N. YANENKO (1976), *The Difference Method for Solving Round Flow Problems, Aeromechanics* (Nauka, Moscow) (in Russian).

BEREZIN, Yu.A. and V.A. VSHIVKOV (1980), *The Particle Method in the Dynamics of Rarefied Plasma* (Nauka, Novosibirsk) (in Russian).

BEREZIN, Yu.A. and N.N. YANENKO (1984), The splitting method for the problem in physics of semiconductors, *Dokl. Acad. Sci. USSR* **274** (6) (in Russian).

BIRKHOF, G. and R. VARGA (1959), Implicit alternating direction methods, *Trans. Amer. Soc.* **92** (1), 13–24.

BIRKHOF, G., R. VARGA and D. YOUNG (1962), *Alternating Direction Implicit Methods*, Advances in Computers **3** (Academic Press, New York).

BOYARINTSEV, Yu.Yu. (1966), *On the Convergence of the Splitting Method and Local Correctness Criterion for Difference Equations with Varying Coefficients*, Nekotorye Voprosy Prikl. i Vychisl. Mat. (Nauka, Novosibirsk) (in Russian).

BOYARINTSEV, Yu.Yu. and O.P. UZNADZE (1967). On the convergence of splitting the system of spherical harmonics, *Zh. Vychisl. Mat. i Mat Fiz.* **7** (6) (in Russian).

BRAILOVSKAYA, I.Yu., T.V. KUSKOVA and L.A. CHUDOV (1968), *The Difference Methods for Solving Navier–Stokes Equations (Survey)*, Vychisl. Metody i Programmirovanie **XI** (Moskov. Gos. Univ., Moscow) (in Russian).

BRAYN, K. (1966), A scheme for numerical integration of the equations for motion on an irregular grid free of nonlinear instability, *Monthly Weather Rev.* **94** (1).

BRIAN, P.L.I. (1961), A finite difference method of high order accuracy for the solution of three-dimensional heat conduction problems, *AIChE J.* **7**, 367–370.

BULEYEV, N.I. (1960), The numerical method for solving two- and three-dimensional diffusion equations, *Mat. Sb.* **5** (2) (in Russian).

BULEYEV, N.I. (1970), The method of incomplete factorization for solving two- and three-dimensional equation of diffusion type, *Zh. Vychisl. Mat. i Mat. Fiz.* **10** (4) (in Russian).

BURRIDGE, D.M. and F.R. HAYES (1974), Development of the British operational model, The GARP Programme on Numerical Experimentation, Rept. 4, 102–104.

BUTLER, T.D. (1974), *Recent Advances in Computational Fluid Dynamics*, Lecture Notes in Computer Science **11** (Springer, Berlin) 1–21.

CHAN, R.K.-C. and R.L. STREET (1970), A computer study of finite amplitude water waves, *J. Comput. Phys.* **6**, 68–97.

CHAN, R.K.-C., R.L. STREET and J. FROMM (1971), Numerical modelling of the water waves: The development of Summac method, in: *Proceedings Second International Conference on Numerical Methods in Fluid Dynamics*, Lecture Notes in Physics (Springer, Berlin).

CHORIN, A.J. (1968), Numerical solution of the Navier–Stokes equations, *Math. Comp.* **22** (104), 745–762.

CLINE, M.C. (1974), Computation of steady nozzle flow by a time dependent method, *AIAA J.* **12** (4), 419–420.

CONCUS, P., G.H. GOLUB and D.P. O'LEARY (1976), A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, in: J.R. BUNCH and D.J. ROSE, eds., *Sparse Matrix Computations* (Academic Press, New York) 309–332.

COURANT, R., K.O. FRIEDRICHS and H. LEWY (1928), Über die partiellen Differenzengleichungen der mathematischen Physik, *Math. Ann.* **100**, 32–74.

CRANE, C.M. (1974), A new method for the numerical solution of time dependent viscous flow, *Appl. Sci. Res.* **30** (4), 47–77.

DAVIS, R.T., U. GHIA and K.N. GHIA (1974), Laminar incompressible flow past a class of blunted wedges using the Navier–Stokes equations, *Comput. & Fluids* **2** (2), 211–223.

DAVYDOV, Yu. and V.P. SKOTNIKOV (1978), *The Method of "Large Particles": Aspects of Approximation, Scheme Viscosity and Stability* (VC Acad. Sci. USSR, Moscow) (in Russian).

DOUGLAS, Jr, J. (1955), On the numerical integration of $\partial^2 u/\partial x^2 + \partial^2 u/\partial x^2 = \partial u/\partial t$ by implicit methods, *J. SIAM* **3**, 42–65.

DOUGLAS Jr, J. (1961), Alternating direction iteration for mildly nonlinear elliptic difference equations, *Numer. Math.* **3**, 92–99.

DOUGLAS Jr, J. (1962), Alternating direction methods for three space variables, *Numer. Math.* **4**, 41–63.

DOUGLAS Jr, J. and T. DUPONT (1971), Alternating direction Galerkin methods on rectangles, in: *Numerical Solution of Partial Equations*, II, *SYNSPADE* (Academic Press, New York).

DOUGLAS Jr, J. and J.E. GUNN (1962), Alternating direction methods for parabolic system in m-space variables, *J. Assoc. Comput. Mach.* **9**, 450–456.

DOUGLAS Jr, J. and J.E. GUNN (1963), Two high-order correct difference analogues for the equation of multidimensional heat flow, *Math. Comp.* **17**, 71–80.

DOUGLAS Jr, J. and J.E. GUNN (1964), A general formulation of alternating direction methods, Part I. Parabolic and hyperbolic problems, *Numer. Math.* **6** (5), 428–453.

DOUGLAS Jr, J. and B.F. JONES (1963), On predictor-corrector methods for nonlinear parabolic differential equations, *J. SIAM* **11**, 195–204.

DOUGLAS Jr, J., R. KELLOGG and R. VARGA (1963), Alternating direction methods for n-space variables, *Math. Comp.* **17**.

DOUGLAS Jr, J. and C.M. PEARCY (1963), On convergence of alternating direction procedures in the presence of singular operators, *Numer. Math.* **5**, 175–184.

DOUGLAS Jr, J. and H.H. RACHFORD Jr (1956), On the numerical solution of heat conduction problems in two- and three-space variables, *Trans. Amer. Math. Soc.* **82**, 421–439.

DRYA, M. (1967), On the stability of splitting schemes in C, *Zh. Vychisl. Mat. i Mat. Fiz.* **7** (2) (in Russian).

DRYA, M. (1971a), The splitting difference schemes for systems of hyperbolic equations of the first order, *Zh. Vychisl. Mat. i Mat. Fiz.* **11** (2) (in Russian).

DRYA, M. (1971b), On the convergence of the splitting difference schemes for parabolic systems in C in the internal points of the domain, *Zh. Vychisl. Mat. i Mat. Fiz.* **11** (3) (in Russian).

DUPONT, T. (1968), A factorization procedure for the solution of elliptic difference equations, *SIAM J. Numer. Anal.* **5** (4).

DUVANT, G. and J. LIONS (1972), *Les Inéquations en Mécanique et en Physique* (Dunod, Paris).

DYACHENKO, V.F. (1965), On the new method for numerical solution of the non-stationary problems with two spatial variables in gas dynamics, *Zh. Vychisl. Mat. i Mat. Fiz.* **5** (4) (in Russian).

DYAKONOV, Ye.G. (1961), The alternating direction method for solving systems of finite difference equations, *Dokl. Acad. Sci. USSR* **138** (2) (in Russian).

DYAKONOV, Ye.G. (1962a), The grid method for the parabolic equations of the second order with separable variables, *Dokl. Acad. Sci. USSR* **142** (6) (in Russian).

DYAKONOV, Ye.G. (1962b), On the one way to solve Poisson equation, *Dokl. Acad. Sci. USSR* **143** (1) (in Russian).

DYAKONOV, Ye.G. (1962c), On some difference schemes for solving boundary value problems, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (1) (in Russian).

DYAKONOV, Ye.G. (1962d), On the constructing of the splitting difference schemes for multidimensional non-stationary problems, *Uspekhi Mat. Nauk* **17** (4) (in Russian).

DYAKONOV, Ye.G. (1962e), The splitting difference schemes for multidimensional stationary problems, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (4) (in Russian).

DYAKONOV, Ye.G. (1962f), The splitting difference schemes for non-stationary equations, *Dokl. Acad. Sci. USSR* **144** (1) (in Russian).

DYAKONOV, Ye.G. (1963a), On the application of the splitting difference operators, *Zh. Vychisl. Mat. i Mat. Fiz.* **3** (2) (in Russian).

DYAKONOV, Ye.G. (1963b), On the application of the splitting difference schemes for hyperbolic equations with varying coefficients, *Dokl. Acad. Sci. USSR* **151** (4) (in Russian).

DYAKONOV, Ye.G. (1964a), On the application of the splitting difference schemes for some systems partial equations, *Uspekhi Mat. Nauk* **14** (1) (in Russian).

DYAKONOV, Ye.G. (1964b), The splitting difference schemes of the second order of accuracy for parabolic

equations without mixed derivatives, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (5) (in Russian).

DYAKONOV, Ye.G. (1964c), The splitting difference schemes for general parabolic equations of the second order with varying coefficients, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (2) (in Russian).

DYAKONOV, Ye.G. (1965a), On the application of the splitting difference schemes for some systems of parabolic and hyperbolic equations, *Sibirsk. Mat. Zh.* **6** (3) (in Russian).

DYAKONOV, Ye.G. (1965b), *The Splitting Difference Scheme of the Second Order of Accuracy for Multi-dimensional Parabolic Equations with Varying Coefficients*, Numerical Methods and Programming **3** (Moskov. Gos. Univ., Moscow) (in Russian).

DYAKONOV, Ye.G. (1966), On the application of the splitting difference schemes for some systems of integro-differential equations, *Vestnik Moskov. Univ. Mat.* **5** (in Russian).

DYAKONOV, Ye.G. (1967a), *On the Class of Partial Equations Arising in Closing the Difference Methods Based on Splitting the Operator*, Vychisl. Metody i Programmirovanie. **IV** (Moskov. Gos. Univ., Moscow) (in Russian).

DYAKONOV, Ye.G. (1967b), The splitting difference schemes for the systems of equations of the kind $L_0 \partial u/\partial t + L_1 u = f$, *Dokl. Acad. Sci. USSR* **176** (2) (in Russian).

DYAKONOV, Ye.G. (1967c), *Economical Difference Methods Based on the Splitting Difference Operator for Some Systems of Partial Equations*, Numerical Methods and Programming **6** (Moskov. Gos. Univ., Moscow) (in Russian).

DYAKONOV, Ye.G. (1970), The iterative methods for solving difference analogues of the boundary value problems for elliptic equations, Institute of Cybernetics Acad. Sci. USSR, Kiev (in Russian).

DYAKONOV, Ye.G. (1971a), On some operator inequalities and their applications, *Dokl. Acad Sci. USSR* **198** (5) (in Russian).

DYAKONOV, Ye.G. (1971b), *On the Difference Methods for Solving Some Nonstationary Systems*, Applied Mathematics and Programming **6** (Shtiintsa, Kishinev) (in Russian).

DYAKONOV, Ye.G. (1971c), *The Difference Schemes of Increased Accuracy for Systems of Mixed Equations*, Applied Mathematics and Programming (Shtiintsa, Kishinev) (in Russian).

DYAKONOV, Ye.G. (1971d), *The Difference Methods for Solving Boundary Value Problems*, I: *Stationary Problems* (Moskov. Gos. Univ., Moscow) (in Russian).

DYAKONOV, Ye.G. (1972), *The Difference Methods for Solving Boundary Value Problems*, II: *Non-stationary Problems* (Moskov. Gos. Univ., Moscow) (in Russian).

DYAKONOV, Ye.G. and V.I. LEBEDEV (1967), *The Splitting Method for the Third Boundary Value Problem*, Numerical Methods and Programming **4** (Moskov. Gos. Univ., Moscow) (in Russian).

DYMNIKOV, V.P. and S.K. FILIN (1985), The numerical modelling of the atmospheric circulation response to the temperature anomaly on the surface of the ocean in North Atlantic, Preprint No. 100, OVM Acad. Sci. USSR, Moscow (in Russian).

DYMNIKOV, V.P. and A.V. ISHIMOVA (1979), Unadiabatic model for the short-term weather forecasting, *Meteorologia i Gidrologia* **6** (in Russian).

DYMNIKOV, V.P. and V.N. LYKOSOV (1983), Spectral analysis of the quasi-stationary atmospheric circulation response to the temperature anomaly on the surface of the ocean, Preprint No. 61, OVM Acad. Sci. USSR, Moscow (in Russian).

FAIRWEATHER, G. and A.R. MITCHELL (1966), Some computational results of an improved A.D.I. method for the Dirichlet problem, *Comput. J.* **9** (3).

FAIRWEATHER, G. and A.R. MITCHELL (1967), A new computational procedure for A.D.I. methods, *SIAM J. Numer. Anal.* **4**, 163–170.

FAIRWEATHER, G., A.R. GOURLAY and A.R. MITCHELL (1967), Some high accuracy difference schemes with a splitting operator for equations of parabolic and elliptic type, *Numer. Math.* **10**, 56–66.

FILIPPOV, A.F. (1955), On the stability of the difference schemes, *Dokl. Acad. Sci. USSR* **100** (6) (in Russian).

FORESTER, C. and A. EMERY (1972), A computational method for low Mach number unsteady compressible free convective flows, *J. Comput. Phys.* **10**, 487–502.

FRIEDRICHS, K.O. (1954), Symmetric hyperbolic linear differential equations, *Comm. Pure Appl. Math.* **7**, 345–392.

FRYAZINOV, I.V. (1964), On the difference approximation of the boundary conditions for the third boundary value problem, *Zh. Vychisl. Mat. i. Mat. Fiz.* **4** (6) (in Russian).

FRYAZINOV, I.V. (1966), On the solution of the third boundary value problem for two-dimensional heat conduction equation in arbitrary domain by local one-dimensional method, *Zh. Vychisl. Mat. i Mat. Fiz.* **6** (3) (in Russian).

FRYAZINOV, I.V. (1968), The economical symmetrized schemes for the solution of boundary value problems for multidimensional parabolic equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **8** (2) (in Russian).

FRYAZINOV, I.V. (1969a), The a priori estimates for some family of economical schemes, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (3) (in Russian).

FRYAZINOV, I.V. (1969b), The economical schemes of increased order of accuracy for solving multidimensional parabolic equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (6) (in Russian).

GELFAND, I.M. and O.V. LOKUTSIEVSKY (1962), The sweep method for solving difference equations, in: S.K. GODUNOV and V.S. RYABENKII, eds., *Introduction to the Theory of Difference Schemes* (Fizmatgiz, Moscow) (in Russian).

GLOWINSKY, R., J.-L. LIONS and R. TREMOLIER (1979), *Numerical Study of Variational Inequalities* (Mir, Moscow) (in Russian).

GODUNOV, S.K. (1959), The difference method for numerical computation of discontinuous solutions of equations in hydrodynamics, *Mat. Sb.* **47** (3) (in Russian).

GODUNOV, S.K. (1962), The method of the orthogonal sweep for solving system of difference equations, *Zh. Vychisl. Mat. i Mat. Fiz* **2** (6) (in Russian).

GODUNOV, S.K. and V.S. RYABENKII (1977), *The Difference Schemes: Introduction to the Theory*, (Nauka, Moscow) (in Russian).

GODUNOV, S.K. and K.A. SEMENDYAEV (1962), The difference methods for numerical solution of the problems in gas dynamics, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (7) (in Russian).

GODUNOV, S.K. and A.V. ZABRODIN (1962), On the difference schemes of the second order of accuracy for multidimensional problems, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (4) (in Russian).

GODUNOV, S.K., A.V. ZABRODIN, M.Y. IVANOV et al. (1976), *Numerical solution of Multidimensional Problems in Gas Dynamics* (Nauka, Moscow) (in Russian).

GOURLAY, A. and A. MITCHELL (1961), Split operator methods for hyperbolic system in p-space variables, *Math. Comp.* **21**, 351–354.

GOURLAY, A. and A MITCHELL (1966), Alternating direction methods for hyperbolic systems, *Numer. Math.* **8**, 137–149.

GOURLAY, A. and A. MITCHELL (1967), Intermediate boundary corrections for split operator methods in three dimensions, *BIT* **7**, 31–38.

GOURLAY, A. and A. MITCHELL (1968), High accuracy A.D.I. methods for parabolic equations with variable coefficients, *Numer. Math.* **12**, 180–185.

GOURLAY, A. and A. MITCHELL (1969), A classification of split difference methods for hyperbolic equations in several space dimensions, *SIAM J. Numer. Anal.* **6**, 62–71.

GUITTET, J. (1967), Une nouvelle de directions alternées à q variables, *J. Math. Anal. Appl.* **17**, 199–213.

GUNN, J. (1965), The solution of elliptic difference equations by semi-explicit iterative techniques, *SIAM J. Numer. Anal.* **2** (1).

GUSHCHIN, V.A. (1981), The splittig method for solving the problem in dynamics of inhomogeneous viscous incompressible liquid, *Zh. Vyschisl. Mat. i Mat. Fiz.* **21** (4) (in Russian).

HABBARD, B. (1966), Alternating direction schemes for the heat equation in a general domain, *SIAM J. Numer. Anal.* **2** (3).

HARLOW, F. (1967), *Numerical Method of Particles in Cells for the Problems in Hydrodynamics*, Numerical Methods in Hydrodynamics (Mir, Moscow) (in Russian).

HARLOW, F. and J. WELCH (1965), Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface, *Phys. Fluids* **8**, 2182–2189.

IL'IN, V.P. (1965), On the splitting of the parabolic and elliptic difference equations, *Sibirsk. Mat. Zh.* **6** (1) (in Russian).

IL'IN, V.P. (1966), On the application of the alternating direction method for solving quasi-linear parabolic and elliptic equations, in: *Some Aspects of Applied and Numerical Mathematics* (Nauka, Novosibirsk) (in Russian).

IL'IN, V.P. (1967), On the explicit alternating direction schemes, *Izv. Sib. Otd. Acad. Sci. USSR Ser. Tekhn. Nauk* **13** (3) (in Russian).

IL'IN, V.P. (1970), *The Difference Methods for Solving Elliptic Equations* (Novosibirsk Gos. Univ., Novosibirsk) (in Russian).

KARCHEVSKY, M.M., A.V. LAPIN and A.D. LYASHKO (1972), Economical difference schemes for quasi-linear parabolic equations, *Izv. Vuzov Mat.* **3** (118) (in Russian).

KELDYSH, M.V. (1942), On Galerkin's method for solving boundary value problems, *Izv. Acad. Sci. USSR, Mat.* **6** (in Russian).

KELLOGG, R. (1963), Another alternating direction implicit method, *J. SIAM* **11**, 976–979.

KELLOGG, R. (1964), An alternating direction method for operator equations, *J. SIAM* **12**, (4).

KELLOGG, R. and J. SPANIER (1965), On optimal alternating direction parameters for singular matrices, *Math. Comp.* **19** (1).

KOCHERGIN, V.P. and Yu.A. KUZNETSOV (1969), On the solution of the system of linear equation by the splitting method, in: *Numerical Methods in Transport Theory* (Atomizdat, Moscow) (in Russian).

KONOVALOV, A.N. (1962), The method of fractional steps for solving Cauchy problem for multidimensional oscillation equation, *Dokl. Acad. Sci. USSR* **147** (1) (in Russian).

KONOVALOV, A.N. (1964), The application of the splitting method to the numerical solution of dynamic problems in the theory of elasticity, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (4) (in Russian).

KONOVALOV, A.N. (1972), The problem of the filtration of multiphase incompressible liquid (Novosibirsk Gos. Univ., Novosibirsk) (in Russian).

KOROBITSINA, J.L. (1969), On the boundary conditions in the splitting scheme for kinetic equation, in: *Numerical Methods in Transport Theory* (Atomizdat, Moscow) (in Russian).

KOVENYA, V.M. and A.S. LEBEDEV (1984), The application of the splitting method in predictor-corrector schemes for solving problems in gas dynamics, *Chysl. Metody Mekh. Sploshn. Sredy* **15** (2) (in Russian).

KOVENYA, V.M. and N.N. YANENKO (1981), *The Splitting Methods for Problems in Gas Dynamics* (Nauka, Novosibirsk) (in Russian).

KRYAKVIN, S.A. (1966), On the accuracy of the alternating direction schemes for heat conduction equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **6** (in Russian).

KUTLER, P. and H. LOMAX (1971), Shock-capturing, finite-difference approach to supersonic flows, *J. Spacecraft and Rockets* **8**, 1175–1182.

KUTLER, P., W. REINHARDT and R. WARMING (1973), Multishocked, three-dimensional supersonic flowfields with real gas effects, *AIAA J.* **11**, 657–664.

KUTLER, P., L. SAKELL and G. AIELLO (1974), On the shock-on-shock interaction problem, AIAA Paper 524, New York, 10 pp.

KUZIN, V.I. (1980), On the solution of the equation for barotropic Rossby waves by the finite element method with splitting, in: *Mathematical Modelling of Ocean Dynamics* (VC Sib. Otd. Acad. Sci. USSR, Novosibirsk) (in Russian).

KUZIN, V.I. (1984), The numerical model of global ocean circulation based on finite element method with splitting, *Mathematical Modelling of the Dynamics of Ocean and Internal Reservoirs* (VC Sib. Otd. Acad. Sci. USSR, Novosibirsk) (in Russian).

KUZNETSOV, A.Yu. and M.K. STRELETS (1983), Numerical modelling of sufficiently undersonic stationary non-isothermal flows of homogeneous viscous gas in channels, *Chysl. Metody Mekh. Sploshn. Sredy* **14** (6) (in Russian).

KUZNETSOV, B.P., N.P. MOSHKIN and S. SMAGULOV (1983), Numerical study of viscous incompressible liquid flows in channels of complex geometry under given pressure jumps, *Chysl. Metody Mekh. Sploshn. Sredy* **14** (5) (in Russian).

LADYZHENSKAYA, O.A. (1956), On the solution of non-stationary operator equations, *Mat. Sb.* **39** (4) (in Russian).

LADYZHENSKAYA, O.A. (1958), On the non-stationary operator equations and their applications to linear problems in mathematical physics, *Mat. Sb.* **45** (in Russian).

LADYZHENSKAYA, O.A. and V.Y. RIVKIND (1971), On the convergent difference schemes for Navier–Stokes equations, *Chysl. Metody Mekh. Sploshn. Sredy Inform. Bul.* **2** (1) (in Russian).

LAVAL, P. (1983), Nouveaux schèmas de dèsintègration pour la rèsolution des problèmes hyperboliques et paraboliques non-linéaires applications aux équations d'Euler et de Navier–Stokes, *Rech. Aérospat.* **4**.

LAX, P. (1962), On the stability of finite difference approximations to the solutions of hyperbolic equations with varying coefficients, *Mathematics* (Translations coll.) **6** (3) (in Russian).

LAX, and B. WENDROFF (1964), Difference schemes for hyperbolic equations with high order of accuracy, *Comm. Pure Appl. Math.* **17**, 381–398.

LEBEDEV, V.I. (1971), On the type of quadrature formulae of increased algebraic accuracy for sphere, *Dokl. Acad. Sci. USSR* **231** (1) (in Russian).

LEBEDEV, V.I. (1976), On the quadratures on the sphere, *Zh. Vychisl. Mat. i Mat. Fiz.* **16** (2) (in Russian).

LEBEDEV, V.I. (1977), On Zolotarev's problem in alternating direction method, *Zh. Vychisl. Mat. i Mat. Fiz.* **17** (2) (in Russian).

LEES, M. (1960a), A priori estimates for the solutions of difference approximations to parabolic partial differential equations, *Duke Math. J.* **27**, 297–311.

LEES, M. (1960b), Energy inequalities for the solution of differential equations, *Trans. Amer. Math. Soc.* **94**, 58–73.

LEES, M. (1960c), Alternating direction methods for hyperbolic differential equations, *J. SIAM* **10**, 610–616.

LEES, M. (1961), Alternating direction and semi-explicit difference methods for parabolic partial differential equations, *Numer. Math.* **3**, 398–412.

LEES, M. (1966), A linear three-level difference scheme for quasi-linear parabolic equations, *Math. Comp.* **20**, 516–522.

LEPAS, J., M. BEKLARECHE, J. CAIFFIER, L. FINKE and A. TAGNIT-HAAON (1974), Primitive equations model-implicit method for numerical integration, The GARP Programme on Numerical Experimentation, Rept. 4, 65.

LERAT, A. and R. PEYRET (1975), Propriétés dispersives et dissipatives, d'une classe des schémas aux différences pour les systèmes hyperboliques non-linéaires, *Rech. Aérospat.* **2**, 61–79.

LIONS, P. and B. MERCIER (1978), Splitting algorithms for the sum of two non-linear operators, Rapport Interne 29, Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, France.

LIONS, J.-L. and R. TEMAM (1966), Une méthode d'éclatement des opérateurs et des contraintes en calcul des variations, *C.R. Acad. Sci. Paris* **263**.

LORENZ, E. (1955), Available potential energy and the maintenance of the general circulation, *Tellus* **7**, 157–167.

LUNN, M. (1964), On the equivalence of SOR, SSOR and USSOR as applied to ordered systems of linear equations, *Comput. J.* **7**, 72–75.

LYTKIN, Yu.M. and G.G. CHERNYKH (1975), On the internal waves induced by the collapse of displacement zone in stratified liquid, *Dynamics of Continuous Media* **22** (in Russian).

MACCORMACK, R. and B. BALDWIN (1975), A numerical method for solving Navier–Stokes equations with application to shock-boundary layer interactions, AIAA Paper 1, New York, 8 pp.

MACCORMACK, R. and A. PAULIAY (1972), Computational efficiency achieved by time splitting of finite difference operators, AIAA Paper 154, New York, 7 pp.

MARCHUK, G.I. (1964a), The new approach to numerical solving the equations of weather forecasting, in: *Proceedings Symposium on Long-Term Forecasting Methods*, USA.

MARCHUK, G.I. (1964b), The numerical algorithm for solving the equations of weather forecasting, *Dokl. Acad. Sci. USSR* **156** (2), 308–311 (in Russian).

MARCHUK, G.I. (1965), Numerical methods for the problems in weather forecasting and theory of climate, in: *Lectures on the Numerical Methods for Short-Term Weather Forecasting* (Gidrometeoizdat, Leningrad) (in Russian).

MARCHUK, G.I. (1967), *Numerical Methods for Weather Forecasting* (Gidrometeoizdat, Leningrad) (in Russian).

MARCHUK, G.I. (1968), Some application of splitting methods of the solution of mathematical physics problems, *Apl. Mat.* **13**, 103–132.

MARCHUK, G.I. (1969), The splitting method for problems in mathematical physics, in: *Numerical Methods for Problems in Mechanics of Continuous Media* (in Russian).

MARCHUK, G.I. (1970), Methods and problems in numerical mathematics, *Internat. Math. Congr. Nizza*; also: (1972), *Reports of Soviet Mathematicians* (Nauka, Moscow) (in Russian).

MARCHUK, G.I. (1971), On the theory of splitting method, in: *Numerical Solution of Partial Differential Equations*, II, *SYNSPADE* (Academic Press, New York).

MARCHUK, G.I. (1973), *Introduction into the Methods of Numerical Analysis* (Cremonese, Rome).

MARCHUK, G.I. (1974a), *Numerical Solution of the Problems in Atmosphere and Ocean Dynamics* (Gidrometeoizdat, Leningrad) (in Russian).

MARCHUK, G.I. (1974b), *Numerical Methods in the Computation of Ocean Currencies* (Vychisl. Centr. Sib. Otd. Acad. Sci. USSR, Novosibirsk) (in Russian).

MARCHUK, G.I. (1980), *Methods of Numerical Mathematics* (Nauka, Moscow) (in Russian).

MARCHUK, G.I. (1982), *Mathematical Modelling in the Problem of Environment* (Nauka, Moscow) (in Russian).

MARCHUK, G.I. and V.I. AGOSHKOV (1981), *Introduction to the Projection Grid Methods* (Nauka, Moscow) (in Russian).

MARCHUK, G.I., M.A. BUBNOV, V.B. ZALESNY and A.A. KORDZADZE (1983), Mathematical modelling of the sea currents, tide waves and elaboration of numerical algorithms, in: *Actual Problems in Numerical and Applied Mathematics* (Nauka, Novosibirsk) (in Russian).

MARCHUK, G.I. and V.P. DYMNIKOV, eds. (1974), *Dynamic Meteorology and Numerical Weather Forecasting* (Gidrometeoizdat, Moscow) (in Russian).

MARCHUK, G.I., V.P. DYMNIKOV, V.Ya. GALIN et al. (1975), The hydrodynamic model of the global atmosphere and ocean circulation (in Russian).

MARCHUK, G.I., V.P. DYMNIKOV and V.B. LYKOSOV (1981), On the relation between index cycles of the atmosphere circulation and spatial spectrum of the kinetic energy in the model of the general circulation of the atmosphere, Tech. Memo. 31, ECMWF, 33 pp.

MARCHUK, G.I., V.P. DYMNIKOV, V.N. LYKOSOV, V.Ya. GALIN, I.M. BOBYLEVA and V.L. PEROV (1979), The global atmosphere circulation model, *Izv. Acad. Sci. USSR Ser. Fiz. Atm. Okeana* **15** (in Russian).

MARCHUK, G.I., V.P. DYMNIKOV, V.B. ZALESNY, V.N. LYKOSOV, and V.Ya. GALIN (1984), *Mathematical Modelling of the Global Atmosphere and Ocean Circulation* (Gidrometeoizdat, Lemingrad) (in Russian).

MARCHUK, G.I., G.R. KONTAREV and G.S. RIVIN (1967), The short-term weather forecasting based on the whole equations in bounded territory, *Izv. Acad. Sci. USSR Ser. Fiz. Atm. Okeana* **3** (in Russian).

MARCHUK, G.I., A.A. KORDZADZE and Yu.N. SKIBA (1975), The computation of main hydrological fields of the Black Sea, *Izv. Acad. Sci. USSR Ser. Fiz. Atm. Okeana* **11** (4) (in Russian).

MARCHUK, G.I. and V. KUZIN (1983), On the combination of finite element and splitting methods in the solution of parabolic equations, *J. Comput. Phys.* **52**, 237–272.

MARCHUK, G.I. and Yu.A. KUZNETSOV (1972), *Iterative Methods and Quadratic Functionals* (Nauka, Novosibirsk) (in Russian).

MARCHUK, G.I. and Yu.A. KUZNETSOV (1974), Méthodes itératives et fonctionnelles quadratiques, in: *Sur les Méthodes Numériques en Sciences Physiques et Economiques*, Méthodes Mathématiques de l'Informatique **4** (Dunod, Paris) 3–132.

MARCHUK, G.I. and V.I. LEBEDEV (1981), *Numerical Methods in the Theory of Neutron Transport* (Atomizdat, Moscow) (in Russian).

MARCHUK, G.I. et al. (1980), *Mathematical Models of Ocean Circulation* (Nauka, Novosibirsk) (in Russian).

MARCHUK, G.I., V.V. PENENKO and U.M. SULTANGAZIN (1969), On the solution of the kinetic equation by the splitting method, in: *Numerical Methods in Transport Theory* (Atomizdat, Moscow) (in Russian).

MARCHUK, G.I. and A.S. SARKISYAN, eds. (1980), *Mathematical Models of Ocean Circulation* (Nauka, Novosibirsk) (in Russian).

MARCHUK, G.I. and Yu.N. SKIBA (1976), The numerical computation of the conjugate problem in the model of theoretical interaction between atmosphere, oceans and continents, *Izv. Acad. Sci. USSR Ser. Fiz. Atm. Okeana* **12** (5) (in Russian).

MARCHUK, G.I. and U.M. SULTANGAZIN (1965a), On the convergence of the splitting method for the equations of radiation transport, *Dokl. Acad. Sci. USSR* **161** (1) (in Russian).

MARCHUK, G.I. and U.M. SULTANGAZIN (1965b), On the solution of the kinetic transport equation by the splitting method, *Dokl. Acad. Sci. USSR* **163** (4) (in Russian).

MARCHUK, G.I. and U.M. SULTANGAZIN (1965c). On the grounding of the splitting method for the equation of radiation transport, *Zh. Vychisl. Mat. i Mat. Fiz.* **5** (5) (in Russian).

MARCHUK, G.I. and N.N. YANENKO (1964), The solution of the multidimensional kinetic equation by the splitting method, *Dokl. Acad. Sci. USSR* **157** (6) (in Russian).

MARCHUK, G.I. and N.N. YANENKO (1966), Application of the splitting (fractional steps) method to the

solving of problems in mathematical physics, in: *Some Aspects of Numerical and Applied Mathematics* (Nauka, Novosibirsk) (in Russian).

MARCHUK, G.I. and V.B. ZALESNY (1974), The numerical model of the large-scale circulation in the World Ocean, in: *Numerical Methods for Computation of Ocean Currents* (in Russian).

MELADZE, G. (1970), The schemes of increased order of accuracy for systems of elliptic and parabolic equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **10** (2) (in Russian).

MESINGER, F. and A. ARAKAWA (1976), *Numerical Methods Used in Atmospheric Models*, GARP Publ. Series **17**, 135.

MIGUAL, O., P. PINSKY and R. TAYLOR (1983), Operator split methods for the numerical solution of the elastoplastic dynamic problem, *Comput. Methods Appl. Engrg.* **39**, 137–157.

MITCHELL, A. and G. FAIRWEATHER (1964), Improved forms of the alternating direction methods of Douglas, Peaceman and Rachford for solving parabolic and elliptic equations, *Numer. Math.* **6**, 285–292.

MORRIS, J. (1970), On the numerical solution of a heat equation associated with a thermal print-head, *J. Comput. Phys.* **5**, 208–228.

MOSKOLKOV, M.N. (1969), On the family of the difference schemes for quasi-linear parabolic equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (6) (in Russian).

NICOLS, B. (1971), Recent extensions to the MAC method for incompressible fluid flows, in: *Proceedings Second International Conference on Numerical Methods in Fluid Dynamics*, Lecture Notes on Physics (Springer, Berlin).

NICOLS, B. (1973), Further development of the method of markers and cells for incompressible fluid flows, in: *Numerical Methods in Fluid Mechanics* (Mir, Moscow) (in Russian).

OLIPHANT, T. (1961), An implicit, numerical method for solving two-dimensional time-dependent diffusion problems, *Quart. Appl. Math.* **19** (3).

PEACEMAN, D.W. and H.H. RACHFORD Jr (1955), The numerical solution of parabolic and elliptic differential equations, *J. SIAM* **3**, 28–42.

PEARCY, C. (1962), On convergence of alternating direction procedures, *Numer. Math.* **4**, 172–176.

PENENKO, V.V. and A.Ye. ALOYAN (1975), Some applications of the splitting method for the problems in mesometeorology, *Tr. Zap. Sib. Reg. N. I. Gidrometeorol. In-t* **14** (in Russian).

PENENKO, V.V., U.M. SULTANGAZIN and B.A. BALASH (1969), The solution of the kinetic equation by the splitting method, in: *Numerical Methods in Transport Theory* (Atomizdat, Moscow) (in Russian).

POPOV, Yu.P. and A.A. SAMARSKII (1970), Completely conservative difference schemes for the equations of magnetic hydrodynamics, *Zh. Vychisl. Mat. i Mat. Fiz.* **10** (4) (in Russian).

PRACHT, W. (1973), The implicit method for the computation of creeping with application to the problem of continental drift, in: *Numerical Methods in Fluid Mechanics* (Moscow, Mir) (in Russian).

PYSTA, S. (1969), The splitting difference schemes for systems of mixed differential equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **9** (4) (in Russian).

RACHFORD Jr, H.H. (1968), Rounding errors in alternating direction methods for parabolic problems, *Apl. Mat.* **13**, 177–180.

RAVIART, P. (1967), Sur l'approximation de certaines équations d'évolution linéaires et non-linéaires, *J. Math. Pures Appl.* **46**, 11–107; **46**, 109–183.

REID, J. (1971), On the method of conjugate gradients for the solution of large sparse systems of linear equations, in: *Large Sparse Sets of Linear Equations* (Academic Press, London) 231–251.

RICHTMYER, R.D. and K.W. MORTON (1972), *Difference Methods for Boundary Value Problems* (Mir, Moscow) (in Russian).

RIZZI, A.W. and M. INOUYE (1973a), A time-split finite volume technique for three-dimensional blunt-body flow, AIAA Paper 133, New York, 14 pp.

RIZZI, A.W. and M. INOUYE (1973b), Time finite volume method for three-dimensional blunt-body, *AIAA J.* **11**, 1478–1485.

RIZZI, A.W., A. KLAVINS and R. MACCORMACK (1975), A generalized hyperbolic marching technique for three-dimensional supersonic flow with shocks, in: Lecture Notes in Physics **35** (Springer, Berlin) 341–346.

ROZHDESTVENSKY, B.L. and N.N. YANENKO (1968), *Systems of Quasi-Linear Equations* (Nauka, Moscow) (in Russian).

RUSANOV, V.V. (1960), On the stability of the matrix sweep method, in: *Numerical Mathematics* **6** (Mir, Moscow) (in Russian).

RYABENKII, V.S. and A.F. FILIPPOV (1956), *On the Stability of Difference Equations* (Gostekhizdat, Moscow) (in Russian).

SAMARSKII, A.A. (1961a), A priori estimates for the solution of the difference analogue of a parabolic differential equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **1**, 441–460 (in Russian).

SAMARSKII, A.A. (1961b), A priori estimates for difference equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **1**, 972–1000 (in Russian).

SAMARSKII, A.A. (1962a), The homogeneous difference schemes for nonlinear parabolic equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (1) (in Russian).

SAMARSKII, A.A. (1962b), On the convergence and accuracy of homogeneous difference schemes for one- and multidimensional parabolic equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **2**, 603–634 (in Russian).

SAMARSKII, A.A. (1962c), On an economical difference method for the solution of a multidimensional parabolic equation in an arbitrary region, *Zh. Vychisl. Mat. i Mat. Fiz.* **2**, 787–811 (in Russian).

SAMARSKII, A.A. (1962d), On the convergence of the method of fractional steps for heat conduction problems, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (6) (in Russian).

SAMARSKII, A.A. (1963a), Local one-dimensional difference schemes on non-uniform grids, *Zh. Vychisl. Mat. i Mat. Fiz.* **3** (3) (in Russian).

SAMARSKII, A.A. (1963b), Schemes of increased order of accuracy for the multi-dimensional heat conduction equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **3** (5) (in Russian).

SAMARSKII, A.A. (1964a), On the economical algorithm for the numerical solution of a system of differential and algebraic equations, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (3) (in Russian).

SAMARSKII, A.A. (1964b), Local one-dimensional difference schemes for multidimensional hyperbolic equations in an arbitrary domain, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (4) (in Russian).

SAMARSKII, A.A. (1965a), Economical difference schemes for a hyperbolic system of equations with mixed derivatives and their application to the equations of the theory of elasticity, *Zh. Vychisl. Mat. i Mat. Fiz.* **5** (1) (in Russian).

SAMARSKII, A.A. (1965b), On the additivity principle in constructing economical difference schemes, *Dokl. Acad. Sci. USSR* **165** (6) (in Russian).

SAMARSKII, A.A. (1966), The additive schemes, Thesis of the Reports on the International Mathematicians Congress in Moscow (in Russian).

SAMARSKII, A.A. (1967), The classes of stable schemes, *Zh. Vychisl. Mat. i Mat. Fiz.* **7** (5) (in Russian).

SAMARSKII, A.A. (1970), *Some Aspects of the General Theory of Difference Schemes: Partial Differential Equations* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. (1971), *Introduction to the Theory of Difference Schemes* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. (1977), *The Theory of Difference Schemes* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. and V.B. ANDREEV (1963), On the difference schemes of increased order of accuracy for the elliptic equation with several spatial variables, *Zh. Vychisl. Mat. i Mat. Fiz.* **3** (6) (in Russian).

SAMARSKII, A.A. and V.B. ANDREEV (1964), Iterative alternating direction schemes for the numerical solution of the Dirichlet problem, *Zh. Vychisl. Mat. i Mat. Fiz.* **4** (6) (in Russian).

SAMARSKII, A.A. and V.B. ANDREEV (1976), *The Difference Schemes for Elliptic Equations* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. and I.V. FRYAZINOV (1961), On the convergence of homogeneous difference schemes for a heat conduction equation with discontinuous coefficients, *Zh. Vychisl. Mat. i Mat. Fiz.* **1**, 806–824 (in Russian).

SAMARSKII, A.A. and I.V. FRYAZINOV (1971), On the convergence of the local one-dimensional scheme for the multidimensional heat conduction equation on non-uniform grids, *Zh. Vychisl. Mat. i Mat. Fiz.* **11** (3) (in Russian).

SAMARSKII, A.A. and A.V. GULIN (1973), *The Stability of Difference Schemes* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. and E.S. NIKOLAEV (1978), *Methods for Solving Grid Equations* (Nauka, Moscow) (in Russian).

SAMARSKII, A.A. and Yu.P. POPOV (1975), *Difference Schemes for Gas Dynamics* (Nauka, Moscow) (in Russian).

SAUL'EV, V.K. (1960), *Integration of Parabolic Equations by the Grid Method* (Fizmatgiz, Moscow) (in Russian).

SCOTT, W. and P. ROGER (1983), A more accurate method for the numerical solution of non-linear partial differential equations, *J. Comput. Phys.* **49**, 342–348.

SMELOV, V.V. (1978), *Lectures on the Neutron Transport Theory* (Atomizdat, Moscow) (in Russian).

SOFRONOV, I.D. (1965), The difference scheme with diagonal directions of sweeps for a heat conduction equation, *Zh. Vychisl. Mat. i Mat. Fiz.* **5** (in Russian).

STRANG, G. and G. FIX (1977), *The Theory of the Finite Element Method* (Mir, Moscow) (in Russian).

SULTANGAZIN, U.M. (1971), On the foundation of the method of weak approximation for the spherical harmonics equation, Preprint, Sib. Otd. Acad. Sci. USSR, Novosibirsk (in Russian).

SULTANGAZIN, U.M. (1979), *The Methods of Spherical Harmonics and Discrete Ordinates in Problems in Kinetic Transport Theory* (Nauka, Alma-Ata) (in Russian).

TEMAM, R. (1968), Sur la stabilité et la convergence de la méthode des pas fractionnaires, *Ann. Mat. Pura Appl.* (4) **79**.

TEMAM, R. (1969), Sur l'approximation de la solution des équations de Navier–Stokes par la méthode des pas fractionnaires, *Arch. Rational Mech. Anal.* **32**, 135–153.

TEMAM, R. (1970), Quelques méthodes de décomposition en analyse numérique, *Acte du Congrès International des Mathématiques* **3**.

TEMAM, R. (1981), *Navier–Stokes Equation and Numerical Analysis* (Mir, Moscow) (in Russian).

THOMPSON, R.J. (1964), Difference approximations for inhomogeneous and quasi-linear equations, *J. SIAM* **12**, 189–199.

TIKHONOV, A.N. and A.A. SAMARSKII (1961), Homogeneous difference schemes, *Zh. Vychisl. Mat. i Mat. Fiz.* **1**, 5–63 (in Russian).

TODD, J. (1967), Inequalities of Chebyshev, Zolotareff, Caur and W.B. Jordan, in: *Proceedings Symposium on Inequalities*, Wright-Patterson Air Force Base, Dayton, OH (Academic Press, New York), 321–328.

VALIULIN, A.N. and N.N. YANENKO (1967), Economical difference schemes of increased accuracy for a polyharmonic equation, *Izv. Sib. Otd. Acad. Sci. USSR Ser. Tekhn. Nauk* **13** (3) (in Russian).

VARGA, R.S. (1962), *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, NJ).

VASILYEV, O.R., B.G. KUZNETSOV, Yu.M. LYTKIN and G.P. CHERNYKH (1974), The evolution of a turbulent fluid domain in a stratified medium, *Izv. Acad. Sci. USSR Mekh. Zhidk. Gasa* **3** (in Russian).

VIECELLI, J. (1971), A computing method for incompressible flows bounded by moving walls, *J. Comput. Phys.* **8**, 119–143.

VOROBYEV, Yu.V. (1968), The stochastic iterative process in the alternating direction method, *Zh. Vychisl. Mat. i Mat. Fiz.* **8** (3) (in Russian).

WACHPRESS, E. (1962), Optimum alternating direction implicit iteration parameters for a model problem, *J. SIAM* **10**, 339–350.

WACHPRESS, E. (1963), Extended application of alternating direction implicit iteration model theory, *J. SIAM* **11**, 991–1016.

WACHPRESS, E. (1966), *Iterative Solution of Elliptic Systems and Applications to the Neutron Diffusion Equations of Reactor Physics* (Prentice-Hall, Englewood Cliffs, NJ).

WACHPRESS, E. and G. HABETLER (1960), An alternating direction implicit iteration technique, *J. SIAM* **8**, 403–424.

WANG, G. (1984), The splitting schemes for solving the initial and boundary value problems of hyperbolic partial differential equation, *J. Nanjing Univ. Natur. Sci. Ed.* **1**, 21–38.

WARMING, R., P. KUTLER and H. LOMAX (1973), Second- and third-order non-centred difference schemes for non-linear hyperbolic equations, *AIAA J.* **11**, 189–196.

WASOW, W. and G. FORSYTHE (1963), *Finite Difference Methods for Partial Differential Equations* (in Russian).

WIDLUND, O. (1966), On the rate of an alternating direction implicit method in a non-commutative case, *Math. Comp.* **20**.

WIDLUND, O. (1971), On the effects of scaling of the Peaceman–Rachford method, *Math. Comp.* **25**.

WILKE, H. (1974), Zur Anwendung der SMAS-Methode bei der Behandlung hydrodynamischer insbesondere grenzflachendynamischer Probleme, *Chem. Tech.* **26**, 456–457.

YANENKO, N.N. (1959a), On the difference method for multidimensional heat conduction equation, *Dokl. Acad. Sci. USSR* **125** (6) (in Russian).

YANENKO, N.N. (1959b), Simple implicit schemes for multidimensional problems, *Dokl. Vsesoyuzn. Soveshch. Vychisl. Mat. Vychisl. Tekhn.*, Moscow (in Russian).

YANENKO, N.N. (1960), On the economical implicit schemes (the method of fractional steps), *Dokl. Acad. Sci. USSR* **134** (5) (in Russian).

YANENKO, N.N. (1961), On the implicit difference schemes for a multidimensional heat conduction equation, *Izv. Vuzov Mat.* **4** (23) (in Russian).

YANENKO, N.N. (1962), On the convergence of the splitting method for a heat conduction equation with varying coefficients, *Zh. Vychisl. Mat. i Mat. Fiz.* **2** (5) (in Russian).

YANENKO, N.N. (1964a), *Some Aspects of the Theory of Convergence of Difference Schemes with Constant and Varying Coefficients*, Trudy IV Vsesoyuzn. Mat. S'ezda **2** (Nauka, Moscow) (in Russian).

YANENKO, N.N. (1964b), On the weak approximation to systems of differential equations, *Sibirsk Mat. Zh.* **5** (6) (in Russian).

YANENKO, N.N. (1966), *The Method of Fractional Steps for Multidimensional Problems in Mathematical Physics* (Lectures for students of NGU) (Novosibirsk Gos. Univ., Novosibirsk) (in Russian).

YANENKO, N.N. (1967), *The Method of Fractional Steps for Multidimensional Problems in Mathematical Physics* (Nauka, Novosibirsk) (in Russian).

YANENKO, N.N. (1972), Modern numerical methods for problems in mechanics of continuous media, in: *Proceedings International Mathematics Congress* (Nauka, Moscow) (in Russian).

YANENKO, N.N. (1973), *Difference Methods for Problems in Mathematical Physics*, Trudy MIAN SSSR **22** (Nauka, Moscow) (in Russian).

YANENKO, N.N., N.N. ANUCHINA, V.Ye. PETRENKO and Yu.I. SHOKIN (1970), On the methods for gas dynamics problems with large deformations, in: *Numerical Methods in Mechanics of Continuous Media* (Nauka, Novosibirsk) (in Russian).

YANENKO, N.N., N.N. ANUCHINA, V.Ye. PETRENKO and Yu.I. SHOKIN (1971), On the methods for computation of the fluid flows with large deformations and the approximation viscosity of difference schemes, in: *Proceedings 2nd International Colloquium on Gas Dynamics of Explosion and Reacting Systems* (in Russian).

YANENKO, N.N. and Yu.Ye. BOYARINTSEV (1961), On the convergence of the difference schemes for a heat conduction equation with varying coefficients, *Dokl. Acad. Sci. USSR* **139** (6) (in Russian).

YANENKO, N.N. and G.V. DEMIDOV (1966), The method of weak approximation as the constructive method for solving a Cauchy problem, in: *Some Aspects of Numerical and Applied Mathematics* (Nauka, Novosibirsk) (in Russian).

YANENKO, N.N., V.A. SUCHKOV a ·d Yu. Ya. POGODIN (1959), On the difference solution of the heat conduction equation in curvilinear coordinates, *Dokl. Acad. Sci. USSR* **128** (5) (in Russian).

YANENKO, N.N. and I.K. YAUSHEV (1966), On the absolutely stable scheme for integrating hydrothermodynamic equations, in: *Difference Methods for Problems in Mathematical Physics*, Trudy MIAN **74** (Nauka, Moscow) (in Russian).

YOUNG, D. (1971), *Iterative Solution of Large Linear Systems* (Academic Press, New York).

YOUNG, J.A. and C.W. HORT (1972), Numerical calculation of internal wave motions, *J. Fluid Mech.* **56**, 265–276.

ZALESNY, V.B. (1984), *Modelling of Large-Scale Motions in World Ocean* (OVM Acad. Sci. USSR, Moscow) (in Russian).

ZALESNY, V.B. (1985), Numerical model of ocean dynamics based on the splitting method, *Sov. J. Numer. Anal. Math. Modelling* **1** (2), 1–22.

ZENG, Q. and X. ZHANG (1982), Perfectly energy-conservative time-space finite difference schemes and the consistent split method to solve the dynamical equations of compressible fluid, *Sci. Sinica Ser. B* **25**, 866–880.

Subject Index

see also the contents for further details

Adjoint operator T^*, 304
Advection-diffusion, 283
Alternating direction method, 277, 298
Alternating direction sweep method, 277
Alternating operator method, 289
Alternating triangular method, 289, 292
Absolute stability, 218, 221
Approximation, 206
Approximative relationship, 263
Auxiliary system, 396

Cauchy problem, 212, 327
Computational stability, 217
Conditional stability, 221
Convergence theorem, 222
Convergent iterative method, 297
Correct Cauchy system, 330
Crank–Nicolson scheme, 212, 224
Cyclic, 297
Cyclic sweep method, 438

Decomposition method, 356

Energy identity, 322
Energy inequalities method, 322
Error vector, 297
Evolutionary equation, 211
Explicit alternating direction scheme, 291
Factorization or factorized scheme, 254
Finite-dimensional approximation, 209
Fourier method, 304

Green formula, 211
Grid, 207
Grid function, 207
Grid point, 207

Heat conduction equation, 239, 240, 252

Implicit scheme of alternating direction, 278
Incomplete factorization scheme, 258

jth step operator, 297

Kronecker delta, 329

Lipschitz constant, 350

Main system, 396
Method of large particles, 286
Method of markers and cells, 285
Method of particles in a cell, 285

Neumann's stability condition, 321
Nonstationary, 297, 310
Norm $\|\varphi\|$, 239
Norm of operator T, 214

One-dimensional scheme, 242
Order, 207

Projection, 207

Residual vector, 297, 352

Scheme of complete splitting, 392
Scheme of incomplete splitting, 390
Scheme of increased order of accuracy, 371

Set of restrictions, 357
Source operator S, 212
Space Φ, 210
Spatial variable, 301
Stationary problem, 307
Spectral criterion of stability, 217
Spectral method, 304
Splitting method, 298
Splitting operator, 367
Stability, 216, 219
Stabilizing correction scheme, 274, 279, 298
Stationary iterative method, 297
Step of the grid, 207
Step operator T, 212
Symmetric alternating direction method, 343

Two-level scheme, 212, 332

Uniformly correct Cauchy system, 330

Weak approximation method, 328