Institute of Computational Modelling
Siberian Branch
Russian Academy of Sciences

Accurate Numerical Solution of Convection-Diffusion Problems

Final Report on Grant I/72342
of Volkswagen Foundation

vol. 1

Edited by
Ulrich Rüde,
Vladimir V. Shaidurov

Novosibirsk
Publishing House of Institute of Mathematics

2001

УДК 519.6
ББК 22.19
Б95

Быкова Е.Г., Калпуш Т.В., Карепова Е.Д., Киреев И.В., Пятаев С.Ф., Рюде У., Шайдуров В.В.
Уточнённые численные методы для задач конвекции-диффузии. (на англ. яз.). Том 1. / Под ред. У.Рюде, В.В. Шайдурова. – Новосибирск: Изд-во Ин-та математики, 2001. – 252 с.

Книга на английском языке состоит из двух томов и содержит результаты, полученные в течение выполнения проекта “Уточнённые численные методы для задач конвекции-диффузии” Фонда Фольксвагена. Первый том посвящён результатам, касающимся проекционно-разностных методов аппроксимации уравнений конвекции-диффузии с преобладанием конвекции и проекционно-разностным методам повышенной точности для самосопряжённых эллиптических уравнений второго порядка.

Для специалистов по вычислительной математике.

Bykova E.G., Kalpush T.V., Karepova E.D. Kireev I.V., Pyataev S.F., Rüde U., Shaidurov V.V.
Accurate Numerical Solution of Convection-Diffusion Problems. Ed. by U. Rüde and V.V. Shaidurov. – Novosibirsk: Publishing House of Institute of Mathematics of Siberian Branch of the Russian Academy of Sciences, 2001. – Vol. 1. – 252 p.

This book consists of two volumes and is concerned with the results obtained during carrying out the project ‘Accurate Numerical Solution of Convection-Diffusion Problems’ of the Volkswagen Foundation. The first volume is devoted to the results concerning the projective-difference methods of approximation of the convective-diffusion equations with convection dominated and the projective-difference methods of increasing accuracy for the second-order self-ajoint elliptic equations.

For specialists in computational mathematics.

$Б\frac{1602120000-09}{Я82(03)-01}$ (англ.) Без объявл.

ISBN 5-86134-099-4

Preface

This book in two volumes includes the results obtained in the framework of the Project I/72342 'Accurate Numerical Solution of Convection-Diffusion Problems' of Volkswagen Foundation. The work in accordance with the project started in the end of 1997 and finished in the beginning of 2001.

The final list of russian team includes 7 participants:

V.V.Shaidurov - professor, doktor of physical and mathematical sciences, director of Institute of Computational Modelling of Russian Academy of Sciences; head of Chair on Softwear of Krasnoyarsk State Technical University;

I.V.Kireev - kandidat of physical and mathematical sciences, scientific worker of Institute of Computational Modelling of Russian Academy of Sciences;

E.G.Bykova - kandidat of physical and mathematical sciences, dozent of Krasnoyarsk State Technical University;

L.V.Gilyova - kandidat of physical and mathematical sciences, scientific worker of Institute of Computational Modelling of Russian Academy of Sciences;

E.D.Karepova - kandidat of physical and mathematical sciences, scientific worker of Institute of Computational Modelling of Russian Academy of Sciences;

S.F.Pyataev - diplom. mathematician, scientific worker of Institute of Computational Modelling of Russian Academy of Sciences;

T.V.Kalpush - post-graduate student of Institute of Computational Modelling of Russian Academy of Sciences.

During this period new results were obtained in the following directions:

- increasing accuracy of finite-element schemes for convection-diffusion equations;
- adaptive triangulations in finite-elements and finite-difference methods;
- increasing accuracy and multigrid (cascadic) algorithms for second-order elliptic equations;
- numerical algoritms for time-dependent Navier-Stokes equations.

These results were reported at 7 international congresses and conferences:

- Numerical Methods for Singular Perturbations. Oberwolfach, April, 1998;
- International Congress of Mathematicians. Berlin, August, 1998;

- International Workshop on the Analytical and Computational Methods for Convection-Dominated and Singular Perturbed Problems. Lozenets, Bulgaria, August, 1998;
- International GAMM-Workshop on Multigrid Methods. Bonn, October, 1998;
- International Conference on Numerical Methods for Transport-Dominated and Related Problems. Schloss Wendgrϕben, Germany, September, 1999;
- Sixth European Multigrid Conference. Gent, Belgium, October, 1999;
- Numerical Methods for Singular Perturbation Problems. Oberwolfach, April, 2001.

Several talks and communications were made at 3 russian congresses and conference with foreing participants:

- Third Siberian Congress on Applied and Industrial Mathematics. Novosibirsk, Russia, June, 1998;
- Mathematical Models and the Methods of Their Investigation. Krasnoyarsk, Russia, August, 1999;
- Fourth Siberian Congress on Applied and Industrial Mathematics. Novosibirsk, Russia, June, 2000.

3 young specialists (E.G.Bykova, E.D.Karepova, T.V.Kalpush) made several communications at 5 regional conferences for young scientists.

During this period 9 visits of the russian participants to Germany have been conducted including joint scientific work at

- Erlangen-Nurnberg Friedrich-Alexander University,
- Heidelberg Ruprecht-Karls University,
- Augsburg University,
- Magdeburg Otto-von-Guericke University,
- Dresden Technological University,
- Leipzig Max-Planck Institute for Mathematics in the Sciences,
- Oberwolfach Mathematical Institute.

Two business trips of two german participants to Russia have been conducted including participation in congress at Novosibirsk and joint scientific work in Institute of Computational Modelling of Russian Academy of Sciences in Krasnoyarsk.

Owing to financial support for russian participants, icluding participation in international conferences, and owing to computer up-grade, all russian members had successful progress in scientific level:

- I.V.Kireev defended kandidat thesis [30] in 1997;
- E.G.Bykova defended kandidat thesis [3] in 1998;
- L.V.Gilyova defended kandidat thesis [15] in 2000;
- E.D.Karepova defended kandidat thesis [29] in 2000;

- S.F.Pyataev prepaired kandidat thesis and will defend it in 2001;
- T.V.Kalpush will finish post-graduate course in 2001, will represent thesis, and will defend it in 2001 or 2002.

The most part of the results of this Project was published (see for [1]-[36]). Concerning well-known journals ([34]-[36]), we do not repeat papers from them in present report. Other journals and books, especially in Russian, are not so widely known, therefore we translate the paper from them to English, if necessary, and brought in this report. Some results presented here are only submitted in journals and are publushed here for the first time.

The first volume of this book is devoted to the results concerning the method of approximation of the convection-diffusion equations with convection dominated and the method of increasing accuracy for the second-order self-adjoint elliptic equations. The second volume deals with the multigrid iterative methods for solving the finite-element analogues of the second-order self-adjoint equations and the finite element method for solving the Navier-Stokes time-dependent equations.

In the first part of the present volume new results are presented which are related to the method of fitting and adaptation of grids for approximation of the convection-diffusion equations. The method of fitting for the coefficients of the finite-element grid problem is similar to the difference method of fitting for approximation of the solutions of the boundary layer type. Three different techniques of the adaptation of grids are realized on the basis of a priori or a posteriori estimates of solution derivatives.

In the second part of this volume new results cocerning the nonhomogeneous difference schemes of increased accuracy are presented for the second-order elliptic equations. Besides, using the solution of the Poisson equation as an example, the well-known difference and finite element schemes of the fourth order of accuracy are compared in efficiency.

Russian participations of Project are very grateful to Volkswagen Foundation for the financial support. We tried to use it for most scientific benefit. Many scientists helped us in Russia and Germany, but we would like to thank Prof. L.Tobiska for active participation in joint work, Prof. R.Rannacher for initialization of this work and discussions, and Prof. H.-G.Roos for fruitful discussions. Coordinator of Project, Prof. U.Rüde and his team made many things for effective scientific work. We are very thankful them, but the special thanks to Prof. U.Rüde for his great organizing and scientific work.

References

1. Bykova E.G., Shaidurov V.V.: *A nonuniform difference scheme with fourth order of accuracy in a domain with smooth order boundary.* Siberian Journal of Numerical Mathematics, 1998, vol. 1, № 2, pp. 99–117 (in Russian).
2. Bykova E.G., Shaidurov V.V.: *A two-dimensional nonuniform difference scheme with higher order of accuracy.* Computational Technologies, 1997, vol. 2, № 5, pp. 12–25 (in Russian).
3. Bykova E.G.: *Nonuniform difference schemes of higher-order accuracy for numerical solving some problems of mathematical physics.* Thesis, Krasnoyarsk State University, 1998, 140 pp. (in Russian).
4. Bykova E.G., Shaidurov V.V.: *A nonuniform difference scheme with higher order of accuracy.* In: Mathematical Models and Methoods of Their Investigation, International Conference, Krasnoyarsk State University. 1997, p. 49 (in Russian).
5. Bykova E.G., Shaidurov V.V.: *A nonuniform difference scheme with higher order of accuracy.* In: Abstracts of the Third Siberian Congress on Applied and Industrial Mathematics, part II, Novosibirsk, Institute of Mathematics of Russ. Acad. of Sci., 1998, pp. 7–8 (in Russian).
6. Gilyova L.V.: *A cascadic multigrid algorithm in the finite element method for the three-dimensional Dirichlet problem.* Siberian J. of Numer. Mathematics, 1998, vol. 1, № 3, pp. 217–226 (in Russian).
7. Gilyova L.V., Shaidurov V.V.: *A cascadic multigrid algorithm in the finite element method for the plane elasticity problem.* East-West J. Numer. Math., 1997, vol. 5, № 1, pp. 143–156.
8. Gilyova L.V., Shaidurov V.V.: *A cascadic multigrid algorithm in the finite element method for the elasticity problem.* In: Abstracts of the International Conference on Mathematical Models and Methods of Their Investigation, Krasnoyarsk State University, 1997, pp. 61–62 (in Russian).
9. Gilyova L.V., Shaidurov V.V.: *A cascadic multigrid algorithm in the finite element method for an indefinite-sign elliptic problem.* Report № 402, Augsburg Universität, Institut für Mathematik, 1998.
10. Gilyova L.V., Shaidurov V.V.: *A cascadic multigrid algorithm in the finite element method for an indefinite-sign problem.* In: Abstracts of the Third Siberian Congress on Applied and Industrial Mathematics, part II, Novosibirsk, Institute of Mathematics of Russ. Acad. of Sci., 1998, p. 10 (in Russian).
11. Gilyova L.V., Shaidurov V.V.: *A cascadic multigrid algorithm in the finite element method for an indefinite-sign elliptic problem.* 5-th European Multigrid Conference. Special Topics and Applications, Institut für Computeranwendungen der Universität Stuttgart, 1998, pp. 74–89.
12. Gilyova L.V., Shaidurov V.V.: *A cascade algorithm for solving a discrete analogue of weakly nonlinear elliptic equation.* Technical Report IOKOMO-01-98, Technische Universität Dresden, Fakultät für Mathematik und Naturwissenschaften, 1998.

13. Gilyova L.V., Shaidurov V.V.: *A cascade algorithm for solving a discrete analogue of weakly nonlinear elliptic equation.* Russ. J. Numer. Anal. Math. Modelling, 1999, vol. 14, № 1, pp. 59–69.
14. Gilyova L.V., Shaidurov V.V.: *Justification of asymptotic stability of the triangulation algorithm for the three-dimensional domain.* Siberian J. of Numer. Mathematics, 2000, vol. 3, № 2, pp. 123–136. (in Russian).
15. Gilyova L.V.: *Cascadic iterative algorithms in the finite element method for elliptic boundary value problems.* Thesis, Krasnoyarsk, Institute of Computational Modelling, 2000, 150 pp. (in Russian).
16. Kalpush T.V.: *The construction of orientation grids for the approximation of convection-diffusion problem.* Proceedings of Conference of Young Scientists of KSC'98, Krasnoyarsk, 1998, p. 106 (in Russian).
17. Kalpush T.V.: *An algorithm for the orientation of grids for slolving of finite difference convection-diffusion problem.* In: Abstracts of the International Conference on Mathematical Models and Methods of Their Investigation, Krasnoyarsk State University, 1999, pp. 61–62 (in Russian).
18. Kalpush T.V., Shaidurov V.V.: *The difference scheme for convection-diffusion problem on the oriented grid.* J. of Computational Technologies, 1999, vol. 4, pp. 72–85 (in Russian).
19. Karepova E.D., Shaidurov V.V.: *The Finite Element Method with Fitting Quadrature Rules for Convection-Diffusion Problem.* Preprint № 2, Krasnoyarsk, 1998, 22 pp. (in Russian).
20. Karepova E.D.: *Algebraic Fitting in the Finite Element Method for for Convection- Diffusion Equation with a Small Parameter.* Proceedings of Conference of Young Scientists of KSC'98, Krasnoyarsk, 1998, pp. 24–36 (in Russian).
21. Karepova E.D., Shaidurov V.V.: *Algebraic Fitting in the Finite Element Method for two-dimensionnal Convection-Diffusion Problem.* In: Abstracts of the Third Siberian Congress on Applied and Industrial Mathematics, part II, Novosibirsk, Institute of Mathematics of Russ. Acad. of Sci., 1998, pp. 15–16 (in Russian).
22. Karepova E.D.: *Finite Element Method with Fitted Integration Rules for Convection-Diffusion Equation with Small Diffusion.* In: Workshop'98 on the Analytical and Computational Methods for Convection-Dominated and Singular Perturbed Problems, Bulgaria, 1998, p. 15.
23. Karepova E.D.: *Finite Element Method for the Convection-Diffusion Problem with Regular and Parabolic Boundary Layers.* In: Proceedings of Conference of Young Scientists of KSC'99, Krasnoyarsk, 1999, pp. 29–35 (in Russian).
24. Karepova E.D.: *Numerical Solving of the Convection-Diffusion Problem with Regular and Parabolic Boundary Layers.* In: Proceedings of Conference of Young Scientists of ICM SB RAS'99, Krasnoyarsk, 1999, pp. 32–37 (in Russian).
25. Karepova E.D., Shaidurov V.V.: *Numerical Solving of Two-Dimension Convection-Diffusion Problem with a Small parameter at highest derivative.* In: Proceedings of Conference "Mathematical Models and the Methods of Theirs Investigation, Krasnoyarsk, 1999, pp. 115–116 (in Russian).

26. Karepova E.D., Shaidurov V.V.: *Algebraic fitting in the finite element method for the small parameter reaction-diffusion problem.* In: Advances in Modeling & Analysis, Ser. A: Mathematical Problems, General Mathematical Modeling. A.M.S.E., France, 1999, vol. 36, № 1, pp. 37–54.
27. Karepova E.D., Shaidurov V.V.: *Finite Element Method with Fitted Integration Rule for Convection-Diffusion Problem.* Russian J. of Numerical Analysis, 2000, vol. 15, № 12, pp. 167–182 (in Russian).
28. Karepova E.D., Shaidurov V.V.: *Finite Element Method with Fitted Integration Rule for Convection-Diffusion Problem with Small Parameter.* In: Analytical and Numerical Methods for Convection-Dominated and Singularly Perturbed Problems, Eds. L.G.Vulkov, J.J.K.Miller, G.I.Shishkin, Nova Science, USA, 2000, pp. 51–58.
29. Karepova E.D.: *Finite Element Method for Convection-Diffusion Convection-Dominated Problem.* Thesis, Krasnoyarsk, Institute of Computational Modelling, 2000, 120 pp. (in Russian).
30. Kireev I.V.: *Stressedly-deformed mode in sandwich composite shell of revolution.* Thesis, Novosibirsk, Institute of Hydrodynamics SB RAS, 1997, 138 pp. (in Russian).
31. Nemirovskiy Yu.V., Pyatayev S.F.: *Triangulation of two-dimensional multiply connected domain with concentration and rarefection of a mesh.* In: Applied problems of strength and plasticity. Analysic Methods. Higher School Collection, Moscow, 1998, pp. 146–155 (in Russian).
32. Nemirovskiy Yu.V., Pyatayev S.F.: *Automated triangulation of multiply connected domain with concentration and rarefection of nodes.* J. of Computational Technologies, 2000, vol. 5, № 2, pp. 82–91 (in Russian).
33. Shaidurov V.V.: *Second-order monotone scheme for convection-dominated equations with adaptive triangulations.* In: Tagungsberichte 1998, 1. Halbband, Mathematisches Forschungsinstitut Oberwolfach. Tagungsbericht 15, p. 13.
34. Shaidurov V.V., Timmerman G.: *A cascadic multigrid algorithm for semilinear indefinite elliptic problems.* Computing, 2000, vol. 64, pp. 349–366.
35. Shaidurov V.V., Timmerman G.: *Stable semi-iterative smoother for cascadic and multigrid algorithms.* Lecture Notes in Computational Science and Engineering, 2000, vol. 14, pp. 221–227.
36. Shaidurov V.V., Tobiska L. *The convergence of the cascadic conjugate-gradient method applied to elliptic problems in domains with re-entrant corners.* Math. of Computations, 2000, vol. 69, pp. 501–520.

Table of contents

The finite element method for convection-diffusion convection-dominated problems

Karepova E.D., Shaidurov V.V.

Introduction 13

1 One-dimensional convection–diffusion problem 19

1.1 The differential problem and its properties 19

1.1.1 Boundary layer 19

1.1.2 The asymptotic expansion of the solution 20

1.1.3 The estimates of the remainder term 21

1.1.4 The weak formulation. The Petrov-Galerkin method ... 24

1.2 The finite element method with a linear quadrature rule 26

1.2.1 Construction of the quadrature rule 27

1.2.2 Properties of the discrete problem 32

1.2.3 Convergence result 36

1.3 The finite element method with nonlinear quadrature rule ... 36

1.3.1 Construction of the quadrature rule 37

1.3.2 Properties of the discrete problem 40

1.3.3 Convergence theorem 42

2 Two-dimensional convection-diffusion problem 49

2.1 General remarks 49

2.1.1 Qualitative behaviour of the solution 49

2.1.2 The weak formulation 53

2.2 The scheme with the fitted quadrature rule for a problem without parabolic boundary layers 53

2.2.1 The differential problem 54

2.2.2 Construction of the quadrature rule 61

2.2.3 Properties of the discrete problem. The convergence result 65

2.3 Construction of the method for the problem with regular and parabolic boundary layers 80

2.3.1 Properties of the differential problem. 80

2.3.2 Construction of the fitted quadrature rule. 91

2.3.3 The properties of the discrete problem 95

3 Numerical solution of the discrete problem 100

3.1 Numerical experiments in the one-dimensional case 100

3.2 Test example in the two-dimensional case 105
3.3 The grids 106
3.4 Methods for solving the discrete problem 109
3.5 Discussion of the numerical results 116
References 121

Triangulation of two-dimensional multiply connected domain with concentration and rarefection of grid

Pyataev S.F.

Introduction 130
1 Some recommendations on choice of the function of steps 131
2 Fragmentation of the boundary of multiply connected domain 134
3 Triangulation of a domain 138
4 Conclusion 146
5 Appendix 1 154
6 Appendix 2 154
7 Appendix 3 154
8 Appendix 4 155
9 Appendix 5 155
10 Appendix 6 156
11 Appendix 7 156
12 Appendix 8 158
References 159

A batch of applied programs for numerical solution of convection-diffusion boundary-value problem

Kireev I.V., Pyataev S.F., Shaidurov V.V.

Introduction 161
1 An algorithm of determination of partial derivatives 162
2 Construction of a sequence of embedded grids 164
3 Program realization of the algorithm 167

A difference scheme for convection-diffusion problem on the oriented grid

Kalpush T.V., Shaidurov V.V.

Introduction 177
1 The difference problem statement 178
2 The difference approximation of convective item on an arbitrary trianqular stencil 179
3 Construction of inverse-monotone second-order finite-difference scheme 181
4 The algorithm for the orientation strengthening of the difference grid 188
5 The numerical experiment 190
References 193

A two-dimensional nonuniform difference scheme with higher order of accuracy

Bykova E.G., Shaidurov V.V.

Introduction 196
1 Boundary-value problem and its nonuniform difference approximation 197
2 Stability and solvability of the grid problem 200
3 Convergence of the nonuniform difference scheme 206
4 Numerical examples 210
References 213

A nonuniform difference scheme with fourth order of accuracy in a domain with smooth boundary

Bykova E.G., Shaidurov V.V.

Introduction 214
1 Boundary-value problem 215

2 Construction of the difference grid and classification of its nodes . . 215
3 Interpolation formula 217
4 Construction of difference approximation 218
5 Stability, solvability and convergence of the grid problem 226
6 Numerical examples 235
References 239

Experimental analysis of fourth-order schemes for Poisson's equations

Bykova E.G, Rüde U., Shaidurov V.V.

Introduction 240
1 Formulation of the differential problems 230
2 Tested methods 232
2.1 Five-point scheme and Richardson extrapolation 232
2.2 Nonhomogeneous Bykova-Shaidurov scheme 232
2.3 Khoromskij combination 233
2.4 Nine-point box scheme 233
3 Two ways to compare the computational cost 234
References 241

The finite element method for convection-diffusion convection-dominated problems

Karepova E.D., Shaidurov V.V.

Introduction

The work is devoted to numerical methods for solving singularly perturbed problems for the convection-diffusion equation with the highest derivatives multiplied by a small parameter. In this case the order of the non-perturbed (singular) equation is one less than of the original (perturbed) equation. Therefore the boundary conditions of the perturbed problem are not all fulfilled for the singular one. Some of these conditions are superfluous that leads to the fast variation of the solution in a small vicinity of corresponding parts of a boundary. As a result, the standard finite difference and finite element methods on a uniform grid either are unstable or give poor accuracy for a small parameter of diffusion.

Some data on the asymptotic analysis of the influence of a small parameters in differential equations go back to L.Euler. The modern theoretical and practical investigations have their origin in A.N.Tikhonov's works of 1940s ([49], [50], [51]). The systematic development of methods for solving singularly perturbed problems started in the late 1960s.

In studies of the properties of a differential problem, the methods of the asymptotic expansion with respect to a small parameter were applied (see [14], [33], [16], [43], [17], [38], [42], [44], [88], [18] and the reviews in them) such as the method of the inner and outer expansions ([14] – 1967), the method of M.I.Vishik and L.A.Lusternick ([19] – 1952 and also [45], [19], [53]), and the method of boundary functions being the generalization of the latter one ([15] and [16] - 1960s, and also [11], [17], [13], [18]).

The use of the standard finite difference and finite element methods for solving singularly perturbed problems failed because of poor accuracy and instability of the discrete analogues. Detailed investigations in this field can be found in [110], [74], [102], [63], [23], [4], as well as in the monograph [118] where the present state of numerical methods for solving singularly perturbed problems is covered in considerable detail.

For the problems considered here the constants in the estimates of the convergence of the classical methods, as a rule, depend on a small parameter and increase indefinitely when the parameter approaches zero [4]. Therefore, these methods can not be applied as mentioned above.

There are several approaches to overcome these difficulties. By convention they can be divided into two groups. The first group is made up of various fitted methods in which the coefficients of a difference scheme in the finite difference method or the parameters of a bilinear form and basis functions in the finite element method are chosen with the use of a-priori information on the behavior of the solution of a differential problem (see, e.g., [23]). The second group consists of standard methods on non-uniform grids which are a-priori given or a-posteriori adapted in the process of numerical integration (see, e.g., [5], [58], [37]).

The first attempts to achieve higher-order accuracy are connected with the use of the upwind scheme. The basic idea of this method is to apply an appropriate approximation of the convective term (by the directed differences) and to add artificial viscosity along the streamline direction. It has been proved that this approach leads to the second order convergence for moderate values of the diffusion parameter and to the convergence of only the first order when the value of the parameter is comparable with or less than a mesh size ([103], [128], [125], [59], [70]).

The construction of the methods uniformly convergent with respect to a small parameter is of great importance in numerically solving the problems with a boundary layer. The exponentially fitted methods satisfy this property. They are constructed using the information on a form of the boundary layer component of a solution ([25], [26], [24], [60], [79], [80], [81]). Another way to construct uniformly convergent difference schemes is to use the analytical solution of an equation with constant coefficients. This approach proposed by D.N.Allen and R.V.Southwell [60] is based on the proximity of the original problem to the approximating one with piecewise constant coefficients and gives a discrete problem similar to the exponentially fitted scheme of A.M.Il'in [25]. In the context of this approach, mention should be made of the method of integral identities with special weight functions [39]. This method is constructed in much the same way as the truncated difference schemes of A.A.Samarskii [46].

One more way to achieve higher-order accuracy of the finite difference method outside a boundary layer is connected with increasing the number of nodes in a stencil ([22], [83], [106]). This complicates the stability analysis as well as the two- and three-dimensional generalizations.

As we noted above, the alternative way to construct uniformly convergent methods is to use special grids. First of all, these are the grids proposed by N.S.Bakhvalov [5]. They are logarithmically refined inside the boundary layer. The construction of these grids is based on the estimates of the derivatives of a solution or on the fact that the difference of the values of a solution at any two neighboring nodes of a grid is uniformly bounded with respect to the parameter ([36], [37]). As a rule, this way leads to a nonlinear algebraic equation for some parameters of this function. Therefore various explicit approximations of logarithmic function are used to construct the Bakhvalov grids ([130], [131], [132], [6], [7], [86], [87]).

In [55] and [124] G.I.Shishkin proved that for the problems with a parabolic boundary layer it is impossible to construct a fitted difference scheme with a compact stencil that converges uniformly with respect to a small parameter. Besides, in [55] the nonuniform grid with a piecewise constant mesh size decreasing in a boundary layer was proposed. In this case the upwind scheme is convergent with order $N^{-1}\ln N$ where N is the number of nodes of the grid. For singularly perturbed problems, the general concept of the proof of the uniform convergence of the classical difference schemes on these grids is presented in the monograph [58] by G.I.Shishkin. In [56], [57], [90], [91], [92], [82] this approach is applied to a wide range of singularly perturbed problems in the finite difference framework and in [114], [120], [126], [112] the Shishkin grids are discussed in the context of finite element method.

All these approaches applied to the finite element method together with the specific finite element techniques give a number of tools for numerical solving singularly perturbed problems.

The upwind scheme in the finite element method has several modifications. For example, in the Petrov-Galerkin method [63] the standard piecewise linear trial functions but the piecewise quadratic test functions are used ([75], [93], [94], [95]). K.Morton proposed to construct test functions which yield a simmetric (or nearly simmetric) discrete problem because in this case the Ritz-Galerkin technique is optimal with respect to the energy norm [102]. For one-dimensional problems this method works well but it is difficult to generalize it to higher-dimensional problems ([109], [110], [111]). Mention should be made of the method proposed by M.Tabata in [127] where the convective term is approximated on the upwind elements only ([72], [61], [62]).

T.Hughes and A.Brooks proposed the method using additional viscosity in the streamline direction ([73], [96]). Instead of the standard bilinear form in the Petrov-Galerkin method they considered some its approximation with an additional term introducing additional viscosity in the streamline direction. As a result, the pointwise convergence of the second order can be achieved on the grids oriented in the streamline direction ([108], [99], [100], [101], [61], [133], [134], [135]). This approach is equivalent to the use of the Galerkin method on the special space being the orthogonal product of the space of piecewise linear functions and that of "bubble functions" [71].

We also mention the method of the additive selection of boundary layer functions ([3], [1], [2]). The basic idea of this method is to add one or two exponential functions with a non-local support, that provides a successful approximation of the boundary layer component, to the standard piecewise linear basic.

In the context of adding artificial viscosity, the least squares method can be applied ([97], [98], [84], [85]). A drawback of this method is that when using piecewise polynomial elements, the assumption that the trial and test functions belong to Sobolev's space $H^2(\Omega)$ requires the use of finite elements of $C^1(\Omega)$; but the construction of these elements on an arbitrary triangulation is not easy. Besides, the number of nonzero entries of the stiffness matrix increases.

The application of exponential fitting to the finite element method is represented by two different approaches. In the first approach special piecewise exponential functions are used ([113], [116], [117]). They approximate the smooth component of a solution somewhat worse than piecewise linear ones but give a considerably better approximation of the boundary layer component. This enables to achieve higher-order accuracy in the Galerkin method. We also mention the non-conforming finite element method [119] where discontinuous exponential finite elements are used.

Another approach that extends difference exponential fitting was proposed for the one-dimensional convection-diffusion equation in [122]. The further development of this method is the subject of this work. The basic idea of this approach is to use the standard piecewise linear finite elements on a uniform grid, applying special fitted quadrature rules to approximate the boundary layer component. As a result, the approximate solution converges to the piecewise linear interpolant of the exact one both in the mean square and in the uniform norms.

Recently in the finite difference and finite element methods, adaptive grids are used. They are constructed using a-posteriori information on the approximate solution obtained on a uniform or coarse grid. To estimate the quality of a numerical solution, special functionals named estimators are applied. A number of estimators is proposed in the literature ([64], [65], [68], [66], [76], [67], [77], [78], [129], [89]).

The present work is devoted to the construction and justification of exponentially fitted schemes in the finite element method for the Dirichlet problem for the convection-dominated convection-diffusion equation. Now we outline the basic idea of this approach.

Let Ω be a one- or two-dimensional domain with a piecewise smooth boundary Γ. We consider the Dirichlet problem

$$Lu \equiv -\varepsilon\,\Delta u + \frac{\partial}{\partial x}(b(x)u) = f \qquad \text{in} \quad \Omega, \tag{1}$$

$$u = 0 \qquad \text{on} \quad \Gamma \tag{2}$$

where $\varepsilon \ll 1$ is a positive parameter. The weak formulation of (1) – (2) is given as follows: *find $u \in H_0^1(\Omega)$ such that*

$$a(u,v) = (f,v) \qquad \forall\, v \in H_0^1(\Omega). \tag{3}$$

Here $a(\cdot,\cdot)\colon H_0^1(\Omega) \times H_0^1(\Omega) \to R$ is the bilinear form determined by

$$a(u,v) = \int_\Omega \left(\varepsilon\,\nabla u \nabla v - bu\frac{\partial v}{\partial x}\right)\,d\Omega$$

and $(\cdot,\cdot)$ is the inner product in $L_2(\Omega)$. We represent the solution of (1)–(2) as

$$u = v + \rho \tag{4}$$

where v is the smooth component of the solution which provides a good approximation of u outside the boundary layer and ρ is the boundary layer component which varies fast in a narrow region near some parts of the boundary.

We choose a finite-dimensional space of test functions $T_h \in H_0^1(\Omega)$ with the basis $\{\varphi_j\}_{j=1}^M$. We consider the discrete problem corresponding to (3): *find $u \in T_h$ such that*

$$a^h(u^h,v^h) = f^h(v^h) \qquad \forall\, v^h \in T_h. \tag{5}$$

Here $a^h(\cdot,\cdot) : T_h \times T_h \to R$ is a bilinear form approximating $a(\cdot,\cdot)$ and $f^h : T_h \to R$ is the approximation of the inner product $(f,\cdot)$. In the usual investigation of (5), the following expansion of the error is used:

$$\begin{aligned} a^h(u^h - u^I, w^h) &= a^h(u^h,w^h) - a^h(u^I,w^h) + a(u^I,w^h) \\ &\quad - a(u^I,w^h) + a(u,w^h) - a(u,w^h) \qquad (6) \\ &= f^h(w^h) - f(w^h) + (a - a^h)(u^I,w^h) + a(u - u^I,w^h). \end{aligned}$$

Here u^I is the interpolant of the solution in T_h. In this case, the estimate of the last term in (6) increases indefinitely as ε decreases because the solution contains the boundary layer component ρ. The main point of the presented approach ([122]) is to construct the special approximation of a^h in order to reduce the error $a(\rho, w^h) - a^h(\rho^I, w^h)$ in the estimate

$$\begin{aligned} a^h(u^h - u^I, w^h) &= \big(f^h(w^h) - f(w^h)\big) + \big(a(u, w^h) - a^h(u^I, w^h)\big) \\ &= \big(f^h(w^h) - f(w^h) + a(v, w^h) - a^h(v^I, w^h)\big) \\ &+ \big(a(\rho, w^h) - a^h(\rho^I, w^h)\big) . \end{aligned}$$

The further development of this approach is as follows. Firstly, for the approximation of the boundary layer component we apply the quadrature rules of higher accuracy. Secondly, we use the special approximation of the right-hand side to eliminate the main term of the error of the quadrature rule on the smooth component.

In the first chapter this approach is applied to the one-dime-sional convection-diffusion equation with the highest derivative multiplied by a small parameter. First we construct the discrete problem based on the linear quadrature rule for the approximation of the convection term and use the special quadrature rule for the approximation of the right-hand side. Next we apply the nonlinear quadrature rule. For the obtained grid problems the second order convergence in the uniform norm is proved for small values of ε.

The extension to the two-dimensional case in **the second chapter** complicates the behavior of a solution. Along with a regular boundary layer which is locally described by an ordinary differential equation, a parabolic boundary layer can arise near some parts of the boundary. It satisfies a parabolic differential equation.

In Section 2.1 the general characteristic of the differential problem is given. The comparison principle is proved for the family of differential equations with the boundary conditions of two types. The weak formulation of the problem is presented. **In Section 2.2** the problem free of a parabolic boundary layer of order 0 is considered. Some estimates of the solution and its derivatives are obtained by the comparison principle. On a uniform grid the discrete problem based on the Galerkin method with piecewise linear elements is constructed using the fitted quadrature rules. The first order of convergence is proved.

In Section 2.3 we investigate the problem with regular and parabolic boundary layers. In this case fitting methods fail ([55]). Therefore, together with the fitted quadrature rules for the approximation of the regular boundary layer, we use a special grid refined in the parabolic boundary layer. This grid is similar to that of Bakhvalov type but in the construction of the grid the generating function is not used. Moreover, the distribution of nodes is given by the one-parameter recurrent formula. The stability and convergence results for this problem are obtained on this grid. In this case the first order convergence is also proved.

In the third chapter the numerical results are discussed.

Section 3.1 is devoted to the numerical experiments in the one-dimensional case. The results demonstrate high accuracy and the advantage of the proposed method over well-known ones. Further, some modifications of the Gauss-Seidel method for solving the two-dimensional discrete problem are considered. The calculations were carried out on the grids of three types. In the two-dimensional problem the exact solution was presented in the form of infinite series. All numerical results on stability and convergence are in close agreement with the theoretical ones.

1 One-dimensional convection–diffusion problem

In this chapter the boundary value problem for the ordinary differential convection-dominated convection-diffusion equation is considered. In spite of its simplicity, this problem has the characteristic feature of the convection-dominated problems, namely, a boundary layer. As a result, most of the classical finite difference and finite element methods fail. Thus, we have a simple object to demonstrate in detail all characteristic properties of the problem as well as of the numerical methods proposed.

1.1 The differential problem and its properties

1.1.1 Boundary layer

Consider the ordinary differential equation with the highest derivative multiplied by a small parameter

$$Lu \equiv -\varepsilon u'' + (b(x)u)' = f(x) \quad \text{on} \quad (0,1), \tag{1.1}$$

$$0 < B_0 \le b(x) \le B_1 \quad \text{on} \quad [0,1] \tag{1.2}$$

satisfying the Dirichlet boundary condition

$$u(0) = u_0, \quad u(1) = u_1. \tag{1.3}$$

The functions b and f are assumed to be sufficiently smooth

$$b \in C^2[0,1], \quad f \in C^2(0,1). \tag{1.4}$$

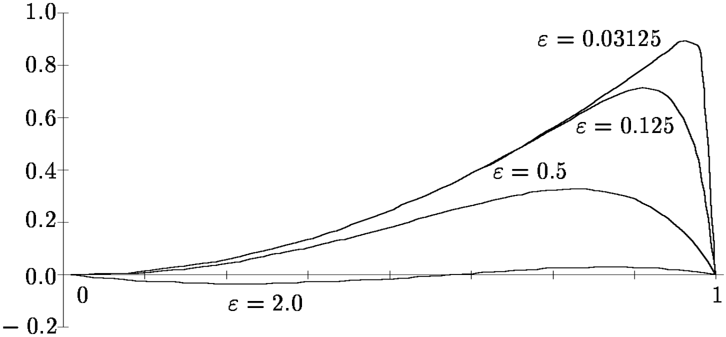


Fig. 1. The appearance of a boundary layer with $\varepsilon \to 0$.

The small coefficient $0 < \varepsilon \ll 1$ of the diffusion term causes the derivatives of the solution to increase exponentially at $x = 1$ ([19], [23]). The appearance of a boundary layer is illustrated in Fig. 1. Here the exact solutions of the problem

$$\begin{aligned} &-\varepsilon u'' + ((1+2x)u)' = 6x^2 + 2x - 2\varepsilon + 2\frac{\exp(-2/\varepsilon)}{1-\exp(-2/\varepsilon)}, \quad x \in (0,1), \\ &\qquad u(0) = u(1) = 0, \end{aligned}$$

are shown for four different values of the diffusion parameter ε.

1.1.2 The asymptotic expansion of the solution

There are many techniques to describe the asymptotic behavior of the solution of the problem (1.1)-(1.3) for small ε. We use the method of expansion in powers of ε proposed by M.I.Vishik and L.A.Lusternik. We introduce the new ('fast') variable $\tau = \dfrac{1-x}{\varepsilon}$ to describe the of boundary layer effects near $x = 1$.

Applying the Vishik - Lusternik technique, we obtain the following expansion of the solution

$$u(x) = v_0(x) + \tilde{\rho}_0(\tau) + \varepsilon\,(v_1(x) + \tilde{\rho}_1(\tau)) + \varepsilon^2 \tilde{z}(x)$$

where v_0 and εv_1 are smooth components which give a good approximation of the solution outside the boundary layer, $\tilde{\rho}_0$ and $\varepsilon\tilde{\rho}_1$ are boundary layer terms, and $\varepsilon^2\tilde{z}(x)$ is a remainder term. Here, $v_0(x)$ is the solution of the reduced problem

$$(bv_0)' = f \quad \text{on } (0,1), \quad v_0(0) = u_0 \tag{1.5}$$

and $v_1(x)$ is the solution of the problem

$$(bv_1)' = v_0'' \quad \text{on } (0,1), \quad v_1(0) = 0. \tag{1.6}$$

The boundary layer functions are described by means of the problems

$$-\tilde{\rho}_0''(\tau) + b(1)\tilde{\rho}_0'(\tau) = 0, \quad \tilde{\rho}_0(0) = u_1 - v_0(1), \quad \lim_{\tau\to\infty} \tilde{\rho}_0(\tau) = 0,$$

and

$$\begin{aligned} &-\tilde{\rho}_1''(\tau) + b(1)\tilde{\rho}_1'(\tau) = b'(1)\tilde{\rho}_0(\tau) - \tau b'(1)\tilde{\rho}_0'(\tau), \\ &\tilde{\rho}_1(0) = -v_1(1), \quad \lim_{\tau\to\infty} \tilde{\rho}_1(\tau) = 0 \end{aligned}$$

with the solutions

$$\tilde{\rho}_0(\tau) = (u_1 - v_0(1)) \exp(-b(1)\tau), \tag{1.7}$$

$$\tilde{\rho}_1(\tau) = \big((u_1 - v_0(1))b'(1)\tau^2/2 - v_1(1)\big) \exp(-b(1)\tau). \tag{1.8}$$

The functions $\tilde{\rho}_k(\tau)$ are defined for $\tau \geq 0$ but for small values of ε they differ from zero only in a small vicinity of the point $\tau = 0$. Therefore we multiply $\tilde{\rho}_0(\tau)$ and $\tilde{\rho}_1(\tau)$ by the cut-off function from $C^2[0,1]$ defined as

$$s(t) = \begin{cases} 0, & t \leq 1/3, \\ \text{monotonically increases on } [1/3, 2/3], \\ 1, & t > 2/3 \end{cases} \tag{1.9}$$

and pass to the variable x:

$$\rho_0(x) = s(x)\tilde{\rho}_0(\tau), \quad \rho_1(x) = s(x)\tilde{\rho}_1(\tau). \tag{1.10}$$

As a result, we get the following expansion of the solution of (1.1)–(1.3) for small ε

$$u(x) = v_0(x) + \rho_0(x) + \varepsilon\,(v_1(x) + \rho_1(x)) + \varepsilon^2 z(x). \tag{1.11}$$

1.1.3 The estimates of the remainder term

We introduce the following norms for the function defined on the segment $[0,1]$

$$||v||_p = \begin{cases} \left(\int\limits_0^1 |v|^p \, dx\right)^{1/p}, & 1 \leq p < \infty, \\ \sup\limits_{[0,1]} \text{vrai}|v|, & p = \infty. \end{cases} \tag{1.12}$$

The following theorem gives the estimate of the remainder term $z(x)$ in the uniform norm.

Theorem 1. *Under the conditions* (1.2), (1.4) *the remainder term* $z(x)$ *of the expansion* (1.11) *obeys the estimate*

$$||z||_\infty \le c_1 \quad ^{*)} \tag{1.13}$$

with a constant c_1 *independent of* ε.

Proof. We express z from (1.11):

$$z(x) = \frac{1}{\varepsilon^2}\left(u(x) - v_0(x) - \rho_0(x) - \varepsilon\left(v_1(x) + \rho_1(x)\right)\right).$$

We substitute this expression in (1.1) and use the expansion of the functions $b(x)$ and $b'(x)$ into the Taylor series at 1. Collecting similar terms, we get

$$Lz(x) = \tilde{f} \equiv a_0 + a_1\frac{1}{\varepsilon}A + a_2\frac{1-x}{\varepsilon}A + a_3\frac{(1-x)^2}{\varepsilon^2}A \tag{1.14}$$

where $a_0(x)$, $a_1(x)$, $a_2(x)$, and $a_3(x)$ are some bounded functions and $A(x) = \exp(-(1-x)b(1)/\varepsilon)$. Since the functions $t\exp(-t)$ and $t^2\exp(-t)$ are bounded on $[0,1]$ by some constants, the last two terms in $\tilde{f}$ are also bounded.

The calculation of $z(0)$ and $z(1)$ by means of boundary conditions for the boundary layer components ρ_0 and ρ_1 and the use of properties of the cut-off function $s(t)$ yield:

$$z(0) = z(1) = 0. \tag{1.15}$$

The problem (1.14)-(1.15) satisfies the comparison principle [122]. Take

$$y(x) = \exp(\sigma x)\left(\gamma_0 + \gamma_1 x + \gamma_2\exp\left(-\frac{(1-x)B_1}{2\varepsilon}\right) + \gamma_3\exp\left(-\frac{(1-x)B_1}{4\varepsilon}\right)\right)$$

as a barrier function where

$$\sigma = 1 + \max_{x\in[0,1]}\frac{|b'| - b'}{2b}.$$

Then

$$Ly(x) \ge |Lz(x)| \quad \text{on} \quad (0,1), \qquad y(0) \ge 0, \quad y(1) \ge 0.$$

Hence by the comparison principle we have

$$|z(x)| \le y(x) \le \max_{x\in[0,1]} y(x) = c_1.$$

*) In what follows, c_i denote constants which are independent of ε, x, and of h at a later time.

This completes the proof. □

Along with the expansion (1.11) consider the asymptotic expansion

$$u(x) = v_0(x) + \rho(x) + \varepsilon z_1(x) \tag{1.16}$$

which will be used to derive the quadrature rule in Section 3. Here $v_0(x)$ is the solution of the reduced problem as before. The boundary layer component is taken like in [122] in the form

$$\rho(x) = s(x)\,(u_1 - v_0(1)) \exp(-(1-x)b(x)/\varepsilon). \tag{1.17}$$

For the remainder term $z_1(x)$ the following estimate is proved in [122].

Theorem 2. *Assume that the conditions* (1.2), (1.4) *hold and* z_1 *is given by* (1.16) – (1.17). *Then there is a positive constant* c_4 *such that the estimate*

$$|z_1^{(j)}| \le c_4 \quad on \quad [0,1], \quad j = 0, 1, \tag{1.18}$$

holds for sufficiently small ε.

We also evaluate the difference between the functions ρ_0 and ρ.

Lemma 3. *Let* ρ_0 *and* ρ *be the boundary layer components of order* 0 *given by the formulae* (1.10) *and* (1.17) *respectively. Then there is a positive constant* c_5 *such that the estimate*

$$|\rho_0 - \rho| \le c_5\varepsilon \quad on \quad [0,1] \tag{1.19}$$

holds for sufficiently small ε.

Proof. By the mean-value theorem, the following inequality holds for any $x \in [0,1]$:

$$\begin{aligned} &|\exp(-(1-x)b(1)/\varepsilon) - \exp(-(1-x)b(x)/\varepsilon)| \\ &\le (b(1) - b(x))\frac{1-x}{\varepsilon}\exp(-(1-x)\tilde{b}/\varepsilon) \end{aligned}$$

where $\tilde{b} \in [B_0, B_1]$. Since

$$|b(1) - b(x)| \le |1-x|\|b'\|_\infty \le c_6|1-x|$$

and $t^2\exp(-\alpha t) \le c_7$ for all $\alpha \ge 0$ and $t \in [0,\infty)$, the following inequality holds:

$$\begin{aligned} &|b(1) - b(x)|\frac{1-x}{\varepsilon}\exp(-(1-x)\tilde{b}/\varepsilon) \\ &\le c_6\varepsilon\left(\frac{1-x}{\varepsilon}\right)^2\exp(-(1-x)\tilde{b}/\varepsilon) \le c_6c_7\varepsilon. \end{aligned}$$

This completes the proof. □

1.1.4 The weak formulation. The Petrov-Galerkin method

Multiply (1.1) by an arbitrary function $v \in H_0^1(0,1)$. By applying Green's formula, we obtain the weak formulation: *find $u \in H^1(0,1)$ which satisfies the boundary condition* (1.3) *and the equality*

$$a(u,v) = (f,v) \quad \forall\, v \in H_0^1(0,1). \tag{1.20}$$

Here $a(\cdot,\cdot)$: $H^1(0,1) \times H_0^1(0,1) \to R$ is the bilinear form

$$a(u,v) = \int_0^1 (\varepsilon u' - bu)\, v'\, dx, \tag{1.21}$$

and $(\cdot,\cdot)$ is the standard inner product in $L_2(0,1)$.

To solve the problem numerically, we use the Petrov-Galerkin finite element method. To begin with, we describe some spaces and estimates which are necessary for the investigation of convergence.

We introduce a trial space $S_h \in H^1(\Omega)$ with a basis $\{\varphi_j\}_{j=0}^{M+1}$ and a test space $T_h \in H_0^1(\Omega)$ with a basis $\{\psi_j\}_{j=1}^{M}$. Let $a^h(\cdot,\cdot) : S_h \times T_h \to R$ be a bilinear form which approximates the form $a(\cdot,\cdot)$ and $f_h : T_h \to R$ be a functional which approximates the inner product $(f,\cdot)$. Then we have the following formulation of the Petrov-Galerkin method (see, for example, [40]): *find $u^h \in S_h$ satisfying the boundary conditions* (1.3) *and the equality*

$$a^h(u^h,v^h) = f_h(v^h) \quad \forall\, v^h \in T_h. \tag{1.22}$$

Since

$$S_h = span\{\varphi_0,...,\varphi_{M+1}\}, \quad T_h = span\{\psi_1,...,\psi_M\}$$

the formulation (1.22) is equivalent to the linear system of algebraic equations

$$L^h U^h = F^h \tag{1.23}$$

where $U^h = (u_1,...,u_M)^T$ is the vector of unknowns, and $F^h = (f_h(\psi_1)-a(\varphi_0,\varphi_1)u_0, f_h(\psi_2), \ldots, f_h(\psi_{M-1}), f_h(\psi_M)-a(\varphi_{M+1},\varphi_M)u_{M+1})^T$ is the right-hand side vector. L^h is the matrix with the elements

$$L_{ij}^h = a^h(\varphi_j,\psi_i), \qquad i,j = 1,...,M. \tag{1.24}$$

The usual way to investigate the convergence of (1.22) consists in evaluating the difference $u - u^h$ in the energy norm in terms of $u - u^I$ where u^I is the interpolant of the solution in S_h. Then we obtain the estimate in L_p–norm. But for the singularly perturbed problems the estimate of $u-u^I$ may be very poor because of the boundary layer component of the solution. Therefore we study the difference $u - u^h$ directly:

$$\begin{aligned}
|a^h(u^h - u^I, w^h)| &= |a^h(u^h, w^h) - a^h(u^I, w^h) + a(u, w^h) - a(u, w^h)| \\
\leq & \quad |f^h(w^h) - f(w^h)| + |a(u, w^h) - a^h(u^I, w^h)| \qquad (1.25) \\
+ & \quad |a(\rho, w^h) - a^h(\rho^I, w^h)| + |a(z_1, w^h) - a^h(z_1^I, w^h)| \quad \forall\, w^h \in T_h.
\end{aligned}$$

Here u is the solution of the differential problem (1.1), (1.3), u^h is the solution of the discrete problem (1.22), and u^I, v_0^I, ρ^I, $z_1^I \in S_h$ are the interpolants of u, v_0, ρ, z_1 respectively.

The basic idea of the method discussed here is to the construct an approximation of the bilinear form that reduces the error of the boundary layer component. The general analysis of the problem and the construction of the discrete analogue which gives the first order ε–uniform convergence can be found in [122]. We cite some results from this work.

For vectors $V^h = (v_1, \cdots, v_M)^T \in \mathbf{R}^M$ we introduce the discrete p-norms

$$||V^h||_p = \begin{cases} \left(\sum\limits_{i=1}^{M} d_i |v_i|^p \, dx \right)^{1/p}, & 1 \leq p < \infty, \\ \max\limits_{1 \leq i \leq M} |v_i|, & p = \infty \end{cases} \qquad (1.26)$$

and

$$|||V^h|||_p = ||(D^h)^{-1} (L^h)^T V^h||_p, \quad 1 \leq p \leq \infty. \qquad (1.27)$$

Here D^h is the diagonal matrix with the positive elements

$$d_i = \text{meas}(\text{supp } \varphi_i), \quad i = 1, ..., M.$$

Note that (1.27) is a norm in $\mathbf{R}^M$ when the matrix L^h is invertible. Together with the space S_h we consider the space $\overset{\circ}{S}_h = span\{\varphi_1, ..., \varphi_M\}$. The spaces $\overset{\circ}{S}_h$ and T_h are equipped with different norms. In order to introduce these norms we use the isomorphisms $\overset{\circ}{S}_h \leftrightarrow \mathbf{R}^M$ and $T_h \leftrightarrow \mathbf{R}^M$ defined by

$$v^h = \sum_{i=1}^{M} v_i \varphi_i \in \overset{\circ}{S}_h, \quad V^h = (v_1, \cdots, v_M)^T \in \mathbf{R}^M,$$

$$w^h = \sum_{i=1}^{M} w_i \psi_i \in T_h, \quad W^h = (w_1, \cdots, w_M)^T \in \mathbf{R}^M.$$

We introduce for $v^h \in \overset{\circ}{S}_h$ and $w^h \in T_h$ the norms

$$\|v^h\|_{p,h} = \|V^h\|_p \qquad \text{and} \qquad |\!|\!| w^h |\!|\!|_{q,h} = |\!|\!| W^h |\!|\!|_q, \tag{1.28}$$

respectively.

Because of definitions (1.24), (1.28), and the Hőlder inequality the following estimate of the bilinear form a^h holds.

Lemma 4. [122] *Suppose that* $1 \le p \le \infty$, $1/p + 1/q = 1$ *and the matrix* L^h *is nonsingular. Then for all* $v^h \in \overset{\circ}{S}_h$ *and* $w^h \in T_h$ *we have*

$$|a^h(v^h, w^h)| \le \|v^h\|_{p,h} |\!|\!| w^h |\!|\!|_{q,h}. \tag{1.29}$$

Now we formulate the basic convergence result.

Theorem 5. *Let* u *and* u^h *be the solutions of the problems* (1.20) *and* (1.22), *respectively. Then for the interpolant* u^I *of* u *in* S_h *the error estimate*

$$\|u^I - u^h\|_{p,h} \le \sup_{w^h \in T_h/\{0\}} \frac{|(f, w^h) - f_h(w^h) + a^h(u^I, w^h) - a(u, w^h)|}{|\!|\!| w^h |\!|\!|_{q,h}} \tag{1.30}$$

holds where $1/q + 1/p = 1$, $1 \le p \le \infty$.

Remark. The error estimate in the continuous L^p–norm follows from the norm equivalence

$$c_8 \|v\|_{p,h} \le \|v\|_p \le c_9 \|v\|_{p,h} \quad \forall\, v \in \overset{\circ}{S}_h \tag{1.31}$$

where constants c_8, c_9 are independent of h.

The discrete problem which is first-order convergent, uniformly in ε, was constructed in [122]. In the next two sections we will obtain the second-order method.

1.2 The finite element method with a linear quadrature rule

In this section we introduce the restriction

$$b'(x) \ge 0 \quad \text{on} \quad [0, 1] \tag{1.32}$$

which simplifies the proof of stability. This restriction provides the fulfillment of the maximum principle for the problem (1.1) – (1.3). We show that for small ε this restriction can be introduced without loss of generality. Assume that for some $x_0 \in [0,1]$ we have $b'(x) < 0$. Introduce a new unknown function

$$w(x) = u(x)\exp(-\sigma x)$$

with the positive constant

$$\sigma = 1 + \max_{x\in[0,1]} \frac{|b'| - b'}{2b}.$$

Then the problem (1.1) – (1.3) is equivalent to the following one

$$\begin{gathered} -\varepsilon w'' + (b - 2\varepsilon\sigma)w' + (b' + b\sigma - \varepsilon\sigma^2)w = f\exp(-\sigma x) \quad \text{on} \quad (0,1), \\ w(0) = u_0, \qquad w(1) = u_1\exp(-\sigma). \end{gathered}$$

For small ε the coefficient of w is positive on the segment $[0,1]$ since $b' + b\sigma - \varepsilon\sigma^2 \geq b - \varepsilon\sigma^2 \geq 0$. Hence the maximum principle holds.

1.2.1 Construction of the quadrature rule

To approximate the solution u, we use the piecewise linear finite elements on a nonuniform grid

$$\overline{\omega}_h = \{x_i : \ i = 0,1,\ldots,n; \ 0 = x_0 < x_1 < \ldots < x_{n-1} < x_n = 1\} \quad (1.33)$$

with a mesh size $h_i = x_i - x_{i-1}$. For simplicity we consider a quasiuniform grid satisfying the condition

$$c_{10}h \leq h_i \leq h = \max_{1\leq i\leq n} h_i. \qquad (1.34)$$

We denote the set of interior nodes by

$$\omega_h = \{x_i : \ x_i \in \overline{\omega}_h, \quad i = 1,\ldots, \ n-1\}.$$

Introduce the basis functions $\varphi_i(x) \in C[0,1]$ defined by

$$\varphi_i(x) = \begin{cases} (x - x_{i-1})/h_i, & \text{if} \quad x \in [x_{i-1}, x_i] \cap [0,1]; \\ (x_{i+1} - x)/h_{i+1}, & \text{if} \quad x \in (x_i, x_{i+1}] \cap [0,1]; \\ 0 & \text{otherwise} \end{cases}$$

and the spaces of trial and test functions

$$S_h = \text{span}\{\varphi_0,\ldots,\varphi_n\} \quad \text{and} \quad T_h = \text{span}\{\varphi_1,\ldots,\varphi_{n-1}\}.$$

To approximate the problem (1.20), we use the Petrov-Galerkin method: *find* $u^h \in S_h$ *such that* $u^h(0) = u_0$, $u^h(1) = u_1$ *and*

$$a(u^h, w^h) = (f, w^h) \quad \forall\, w^h \in T_h. \tag{1.35}$$

This approach has some disadvantages. With the boundary layer, the solution of the algebraic system has unsatisfactory accuracy. Besides, this system becomes unstable for $\varepsilon \leq h$. Finally, when constructing the algebraic system, one have to integrate functions. Thus, the application of quadrature rules is quite natural. We choose quadrature rules in a special way to ensure stability and to improve the accuracy of the approximate problem obtained.

Therefore we return to the bilinear form (1.21). The first term is integrated exactly for any $u \in S_h$, $v \in T_h$. For the second term we use the following quadrature rule on each interval:

$$\int_{x_{i-1}}^{x_i} bv\,dx \approx (\alpha_i b_{i-1} v_{i-1} + \beta_i b_i v_i)\, h_i \tag{1.36}$$

where $v_i = v(x_i)$ for an arbitrary function $v(x)$. Using this formula for a we obtain the new bilinear form a_h of an algebraic type for $v \in S_h$ and $w^h \in T_h$:

$$a^h(v, w^h) = \sum_{i=1}^{n} (\varepsilon(v_i - v_{i-1})/h_i - \alpha_i b_{i-1} v_{i-1} - \beta_i b_i v_i)\,(w_i^h - w_{i-1}^h). \tag{1.37}$$

The standard way to justify the accuracy of the Galerkin solution is to use Strang's first lemma and the closeness of the bilinear forms a and a^h with arguments from the class of admissible functions. Unfortunately, in our case this method yields poor estimates due to the boundary layer components ρ_0, ρ_1. Therefore we choose the parameters α_i, β_i in such a way as to make these bilinear forms as close as possible just for the functions ρ_0, ρ_1. For example, for the function ρ_0 the exact equality

$$\int_{x_{i-1}}^{x_i} b\rho_0\,dx = (\alpha_i b_{i-1} \rho_{0,i-1} + \beta_i b_i \rho_{0,i})\, h_i$$

should be taken. However, this condition contains the integral in the left-hand side, that does not permit to obtain the explicit expression in the general case. Therefore for convenience we use (1.7), (1.10) for ρ_0 in the right-hand side of this equality and replace $b(x)$ by its linear interpolant. As a result, we arrive at the equality

$$\begin{aligned}\int_{x_{i-1}}^{x_i} & (b_{i-1}(x_i - x)/h_i + b_i(x - x_{i-1})/h_i) \exp(-b(1)(1 - x)/\varepsilon)\,dx \\ &= \alpha_i b_{i-1} \exp(-b(1)(1 - x_{i-1})/\varepsilon) h_i + \beta_i b_i \exp(-b(1)(1 - x_i)/\varepsilon) h_i.\end{aligned} \tag{1.38}$$

Taking the integral in the left-hand side and dividing the obtained equality by $h_i \exp(-b(1)(1-x_i)/\varepsilon)$ we get

$$\begin{aligned}\alpha_i b_{i-1} \exp(-\sigma_i) + \beta_i b_i = b_{i-1} &\left(\frac{1}{\sigma_i^2} - \frac{1}{\sigma_i} \exp(-\sigma_i) - \frac{1}{\sigma_i^2} \exp(-\sigma_i) \right) \\ &+ b_i \left(\frac{1}{\sigma_i} - \frac{1}{\sigma_i^2} + \frac{1}{\sigma_i^2} \exp(-\sigma_i) \right)\end{aligned} \tag{1.39}$$

where $\sigma_i = b(1)h_i/\varepsilon$. To the above equality we add the equation

$$\alpha_i + \beta_i = 1 \tag{1.40}$$

which permits to approximate an integral of a smooth function with the first-order accuracy. Thus, we arrive at the system of linear algebraic equations in two unknowns. Its determinant is given by

$$\xi_i = b_i - b_{i-1} \exp(-\sigma_i). \tag{1.41}$$

Since $b' \geq 0$ and $\exp(-\sigma_i) < 1$, ξ_i is strictly positive. Hence the system (1.39) – (1.40) has an unique solution.

Thus, we can expect that the boundary layer function ρ_0 satisfies the equality

$$\int_{x_{i-1}}^{x_i} b\rho_0 \, dx = (b\rho_0)_i^* h_i + O(h^3). \tag{1.42}$$

The proof of this statement is given later. Here we use the notation

$$(b\rho_0)_i^* = \alpha_i b_{i-1} \rho_{0,i-1} + \beta_i b_i \rho_{0,i}.$$

It is easy to verify that for ρ_1 we have

$$\int_{x_{i-1}}^{x_i} b\rho_1 \, dx = O(\varepsilon).$$

Actually the contribution of this term is still smaller due to the coefficient ε of the function ρ_1 in the expansion (1.11).

Now consider the remaining part of the solution

$$g(x) = v_0(x) + \varepsilon v_1(x) + \varepsilon^2 z(x). \tag{1.43}$$

For $g(x)$ the quadrature rule (1.36) has only the first-order accuracy:

$$\begin{aligned}\int_{xi-1}^{x_i} b(x)g(x) \, dx &= (bv_0)_i^* h_i + (1/2 - \beta_i) h_i^2 (bv_0)'_{i-1} \\ &\quad + (bv_1)_i^* h_i + O(h^3 + \varepsilon h^2 + \varepsilon^2 h) \qquad (1.44) \\ &= (bg)_i^* h_i + (1/2 - \beta_i) h_i^2 (bg)'_{i-1} + O(h^3 + \varepsilon h^2 + \varepsilon^2 h).\end{aligned}$$

Here we use the fact that the functions v_0'', v_1', and z are bounded on $[0,1]$. Take into consideration the equality

$$-\varepsilon(\rho_0+\varepsilon\rho_1)''+(b(\rho_0+\varepsilon\rho_1))'=O(\varepsilon)$$

which results from the definition of ρ_0 and ρ_1. Then we transform the main term of the error to the form

$$(bg)'=(bg)'-\varepsilon g''+O(\varepsilon)=(bu)'-\varepsilon u''+O(\varepsilon)=f+O(\varepsilon).$$

Then instead of (1.44) we get

$$\begin{aligned}\int_{x_{i-1}}^{x_i} b(x)g(x)\,dx &= (bg)_i^* h_i\\ &\quad + (1/2-\beta_i)h_i^2 f_{i-1}+O(h^3+\varepsilon h^2+\varepsilon^2 h).\end{aligned}\tag{1.45}$$

When constructing the bilinear forms a and a^h all the terms are multiplied by $-(w^h)'$. Therefore the main term of the error $a(g,w^h)-a^h(g,w^h)$ on the segment $[x_{i-1},x_i]$ takes the form

$$-(1/2-\beta_i)h_i^2 f_{i-1}(w^h)'_{i-1/2}.\tag{1.46}$$

We construct the quadrature rule for the right-hand side to eliminate this term. We rewrite the functional in the right-hand side as

$$\int_0^1 f(x)w^h(x)\,dx=-\int_0^1 F(x)(w^h(x))'\,dx \quad \forall\, w^h\in T^h\tag{1.47}$$

with the antiderivative $F'(x)=f(x)$. Using the Taylor expansion, in a similar way as (1.44) we obtain

$$\int_{x_{i-1}}^{x_i} F(x)\,dx=F_i^* h_i+(1/2-\beta_i)h_i^2 f_{i-1}+O(h^3).\tag{1.48}$$

Thus, the main term of the error coincides with (1.44). In order to avoid the calculation of the antiderivative, we use the difference analogue of integration by parts taking into account the boundary conditions $w^h(0)=w^h(1)=$

0:

$$\int_0^1 f(x)w^h(x)\,dx = -\int_0^1 F(x)(w^h(x))'\,dx = -\sum_{i=1}^{n}\int_{x_{i-1}}^{x_i} F(x)(w^h)'\,dx$$
$$= -\sum_{i=1}^{n}(w^h)'_{i-1/2}(F_i^* h_i + (1/2-\beta_i)h_i^2 f_{i-1}) + O(h^3)\sum_{i=1}^{n}(w^h)'_{i-1/2}$$
$$= \sum_{i=1}^{n-1} w_i^h (F_{i+1}^* - F_i^*) - \sum_{i=1}^{n}(1/2-\beta_i)h_i^2 f_{i-1}(w^h)'_{i-1/2}$$
$$+O(h^3)\sum_{i=1}^{n}(w^h)'_{i-1/2}.$$

We choose the weights μ_i and ν_i in such a way as to replace the difference between the values of the antiderivative F by the function f with the third-order accuracy:

$$F_{i+1}^* - F_i^* = \mu_i f_{i-1} + \nu_i f_i + O(h^3). \tag{1.49}$$

Then we use the Taylor expansion at the point x_{i-1} and set the coefficients of h_i and h_i^2 to be equal:

$$\begin{aligned} \mu_i + \nu_i &= h_i(1-\beta_i) + h_{i+1}\beta_{i+1}, \\ 2\nu_i h_i &= h_i^2(1-\beta_i) + h_{i+1}\beta_{i+1}(2h_i + h_{i+1}). \end{aligned} \tag{1.50}$$

Hence μ_i and ν_i are uniquely determined. As a result, in the right-hand side we get the approximate functional

$$f_h(w^h) = \sum_{i=1}^{n}(\mu_i f_{i-1} + \nu_i f_i)w_i^h \tag{1.51}$$

with the coefficients μ_i and ν_i from (1.50).

On substitution of the bilinear form $a(\cdot,\cdot)$ and the right-hand side (f, w^h) into (1.35), we obtain the discrete problem: *find* $u^h \in S_h$ *such that* $u^h(0) = u_0$, $u^h(1) = u_1$, *and*

$$a^h(u^h, w^h) = f_h(w^h) \qquad \forall\, w^h \in T_h. \tag{1.52}$$

We rewrite this problem in the equivalent matrix-vector form: *construct the function*

$$u^h = \sum_{i=0}^{n} \gamma_i \varphi_i$$

with the weights γ_i *which satisfying the conditions* $\gamma_0 = u_0$, $\gamma_n = u_1$ *as well as the system of linear algebraic equations*

$$A^h \gamma = F^h \tag{1.53}$$

with the vector of the unknowns $\gamma = (\gamma_1, ..., \gamma_{n-1})^T$ and the given the right-hand side $F^h = (F_1^h, ..., F_{n-1}^h)^T$ where

$$\begin{aligned} F_1^h &= \mu_1 f_0 + \nu_1 f_1 + a_1 u_0, \\ F_i^h &= \mu_i f_{i-1} + \nu_i f_i, \qquad\qquad i = 2, ..., n-2, \\ F_{n-1}^h &= \mu_{n-1} f_{n-2} + \nu_{n-1} f_{n-1} + e_{n-1} u_1. \end{aligned}$$

The matrix A^h has the tridiagonal form

$$A^h = \begin{pmatrix} d_1 & -e_1 & & & & & \\ -a_2 & d_2 & -e_2 & & & 0 & \\ & \dots & \dots & \dots & & & \\ & & \dots & \dots & \dots & & \\ & & & \dots & \dots & \dots & \\ & 0 & & & -a_{n-2} & d_{n-2} & -e_{n-2} \\ & & & & & -a_{n-1} & d_{n-1} \end{pmatrix}$$

and its elements are given by

$$\begin{aligned} a_i &= \varepsilon / h_i + \alpha_i b_{i-1}, \\ d_i &= \varepsilon / h_i + \varepsilon / h_{i+1} + \alpha_{i+1} b_i - \beta_i b_i, \\ e_i &= \varepsilon / h_{i+1} - \beta_{i+1} b_{i+1}, \\ i &= 1, ..., n-1. \end{aligned} \tag{1.54}$$

1.2.2 Properties of the discrete problem

Now we investigate the discrete problem.

Lemma 6. *Under the restrictions* (1.2), (1.4), (1.32) *for any* ε, $h > 0$ *the matrix* A^h *of the system* (1.53) *is an* M*-matrix and hence is nonsingular.*

Proof. First we consider the parameter α_i of the quadrature rule. From the equations (1.39), (1.40) we have

$$\begin{aligned}\alpha_i &= \frac{1}{b_i - b_{i-1}\exp(-\sigma_i)}\Big(b_{i-1}\frac{\sigma_i\exp(-\sigma_i) - 1 + \exp(-\sigma_i)}{\sigma_i^2} \\ &+ b_i\,\frac{\sigma_i^2 - \sigma_i + 1 - \exp(-\sigma_i)}{\sigma_i^2}\Big) \\ &= \frac{1}{b_i - b_{i-1}\exp(-\sigma_i)}\Big(\frac{1-\exp(-\sigma_i)}{\sigma_i^2}\,(b_i - b_{i-1}) + b_i\Big) - \frac{1}{\sigma_i}\end{aligned} \tag{1.55}$$

where $\sigma_i = b(1)h_i/\varepsilon$. Since the inequalities $\sigma_i > 0$, $\exp(-\sigma_i) < 1$ and $b'(x) \geq 0$ hold for arbitrary ε and h, we have

$$\alpha_i > 1/\sigma_i \quad \forall\, \varepsilon, h_i > 0. \tag{1.56}$$

Due to (1.54), (1.56), and inequality $b_{i-1} \leq b(1)$ the estimate

$$0 \leq \frac{\varepsilon}{h_i} - \frac{b_{i-1}\varepsilon}{b(1)h_i} < \frac{\varepsilon}{h_i} + \alpha_i b_{i-1} = a_i$$

holds. Hence the coefficients a_i of the system (1.53) are strongly positive.

The proof of the positiveness of e_i is rather complicated. Because of the definition of β_{i+1} the following equalities hold:

$$\begin{aligned}e_i &= \frac{\varepsilon}{h_i} - \frac{b_{i+1}}{b_{i+1} - b_i\exp(-\sigma_i)}\Bigg(b_i\left(\frac{1}{\sigma_i^2} - \frac{\exp(-\sigma_i)}{\sigma_i} - \frac{\exp(-\sigma_i)}{\sigma_i^2} - \exp(-\sigma_i)\right) \\ &+ b_{i+1}\left(\frac{1}{\sigma_i} - \frac{1}{\sigma_i^2} + \frac{\exp(-\sigma_i)}{\sigma_i^2}\right)\Bigg) \\ &= \frac{\varepsilon}{h_i} - \frac{b_{i+1}^2}{\sigma_i\left(b_{i+1} - b_i\exp(-\sigma_i)\right)} + \frac{b_{i+1}}{b_{i+1} - b_i\exp(-\sigma_i)}\Bigg(\frac{b_i\exp(-\sigma_i)}{\sigma_i} \\ &+ b_i\exp(-\sigma_i) + \frac{b_{i+1} - b_i}{\sigma_i^2}\left(1 - \exp(-\sigma_i)\right)\Bigg).\end{aligned} \tag{1.57}$$

The difference of the two list terms in the right-hand side equals

$$\begin{aligned}&\frac{b(1)b_{i+1} - b_{i+1}^2 + b(1)b_i\exp(-\sigma_i)}{\sigma_i\left(b_{i+1} - b_i\exp(-\sigma_i)\right)} \\ &\qquad = \frac{b_{i+1}(b(1) - b_{i+1}) + b(1)b_i\exp(-\sigma_i)}{\sigma_i\left(b_{i+1} - b_i\exp(-\sigma_i)\right)}.\end{aligned} \tag{1.58}$$

Under the restrictions (1.2) and (1.32) both terms in the numerator are nonnegative and the denominator is positive. For the same reason, in the last term all three quantities in parentheses are nonnegative and the coefficient of the parenthetical expression is positive. Hence we have $e_i > 0 \quad \forall\, i = 1, ..., n-1$.

For the system (1.53) the following relations hold:

$$\begin{gathered} d_1 - a_2 > 0, \\ d_i - a_{i+1} - e_{i-1} = 0, \quad i = 2, ..., n-2, \\ d_{n-1} - e_{n-2} > 0. \end{gathered} \tag{1.59}$$

Because of the positiveness of a_i and e_i for all i, the matrix in (1.53) is diagonal-dominant along columns and strictly diagonal-dominant along the first and the last ones. Taking into account the fact that the matrix A^h is irreducible [21], this leads to the conclusion of Lemma 6. □

Lemma 7. *Let* $W^h = (w_1, ..., w_{n-1})^T$ *be the solution of the problem*

$$\left(A^h\right)^T W^h = Q^h \tag{1.60}$$

with some right-hand side $Q^h = (q_1, ..., q_{n-1})^T$. *Under the restrictions* (1.2), (1.4), (1.32), *and*

$$\varepsilon \le c_{11} h, \qquad c_{11} > 0 \tag{1.61}$$

the estimate

$$\sum_{i=2}^{n-1} |w_i - w_{i-1}| + |w_1| + |w_{n-1}| \le c_{12} \sum_{i=1}^{n-1} |q_i| \tag{1.62}$$

holds with a constant c_{12} *independent of* ε *and* h.

Proof. In a similar way as Lemma 3.3 in [122] we rewrite the system (1.60) in the form

$$\begin{gathered} e_{i-1}(w_i - w_{i-1}) + a_{i+1}(w_i - w_{i+1}) = q_i, \qquad i = 1, ..., n-1, \\ w_0 = w_n = 0. \end{gathered} \tag{1.63}$$

Using the notation

$$v_i = w_i - w_{i-1}$$

we write the difference equation from (1.63) as

$$v_{i+1} = \frac{e_{i-1}}{a_{i+1}} v_i - \frac{q_i}{a_{i+1}}.$$

Let us set

$$\prod_{k=l}^{l-1} \delta_k = 1 \quad \text{and} \quad \sum_{k=l}^{l-1} \delta_k = 0$$

for arbitrary δ_k. Then for all $i = 1, ..., n$ the equality

$$v_i = v_1 \prod_{j=1}^{i-1} \frac{e_{j-1}}{a_{j+1}} - \sum_{j=1}^{i-1} \left(\prod_{k=j+1}^{i-1} \frac{e_{k-1}}{a_{k+1}} \right) \frac{q_j}{a_{j+1}} \tag{1.64}$$

holds. Taking into account the equality

$$\sum_{j=1}^{n} v_j = w_n - w_0 = 0,$$

we obtain the initial value

$$v_1 = \sum_{i=1}^{n} \sum_{j=1}^{i-1} \left(\prod_{k=j+1}^{i-1} \frac{e_{k-1}}{a_{k+1}} \right) \frac{q_j}{a_{j+1}} \Big/ \sum_{i=1}^{n} \prod_{j=1}^{i-1} \frac{e_{j-1}}{a_{j+1}}. \tag{1.65}$$

From (1.64) and (1.65) we get

$$\begin{aligned} \sum_{i=1}^{n} |v_i| &\leq \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} \frac{e_{j-1}}{a_{j+1}} \right) |v_1| + \sum_{i=1}^{n} \sum_{j=1}^{i-1} \left(\prod_{k=j+1}^{i-1} \frac{e_{k-1}}{a_{k+1}} \right) \frac{|q_j|}{a_{j+1}} \\ &\leq 2 \sum_{i=1}^{n} \sum_{j=1}^{i-1} \left(\prod_{k=j+1}^{i-1} \frac{e_{k-1}}{a_{k+1}} \right) \frac{|q_j|}{a_{j+1}}. \end{aligned} \tag{1.66}$$

Taking into consideration the definition of the coefficients a_k and e_k in (1.54), the restriction (1.56), and the fact that b is bounded due to (1.2), we obtain the inequalities

$$\begin{aligned} e_{k-1} &= \varepsilon/h_i - \beta_k b_k \leq \varepsilon/c_{10} h, \\ a_{k+1} &= \varepsilon/h_i + \alpha_{k+1} b_k \geq \varepsilon/c_{10} h_i + B_0/2, \\ 1/a_{j+1} &\leq 1/\left(\varepsilon/h_i + B_0/2\right) \leq 2/B_0. \end{aligned}$$

Applying them to the right-hand side of (1.66), we have the estimate

$$\left(\prod_{k=j+1}^{i-1} \frac{e_{k-1}}{a_{k+1}} \right) \frac{|q_j|}{a_{j+1}} \leq \frac{2}{B_0} \eta^{i-j-1} |q_j| \quad \text{where} \quad \eta = \frac{1}{1 + B_0 c_{10} h / 2\varepsilon}.$$

Using this inequality we change the order of summution in the right-hand side of (1.66):

$$\begin{aligned}\frac{4}{B_0}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\eta^{i-1-j}|q_j| &= \frac{4}{B_0}\sum_{l=1}^{n}\sum_{j=1}^{n-l}\eta^{l-1}|q_j| \\ &\le \frac{4}{B_0}\sum_{l=1}^{n}\eta^{l-1}\sum_{j=1}^{n-1}|q_j| \le \frac{4}{B_0(1-\eta)}\sum_{j=1}^{n-1}|q_j|.\end{aligned} \qquad (1.67)$$

Thus we get (1.62) with the constant $c_{12} = 4(1 + 2c_{11}/B_0c_{10})/B_0$. □

In the terms of the norms in the spaces of trial and test functions (1.26) and (1.28) the estimate (1.62) for functions $w^h \in T_h$ has the form

$$\|(w^h)'\|_1 \le c_{12}|\!|\!|w^h|\!|\!|_{1,h}. \qquad (1.68)$$

1.2.3 Convergence result

Now we consider the main theorem of this section.

Theorem 8. *Let* (1.2), (1.4), (1.32), *and* (1.61) *be valid for the problems* (1.53) *and* (1.1) – (1.3) *with the solutions u^h and u respectively. Then the estimate*

$$\max_{0\le i\le n}|u_i^h - u_i| \le c_{15}(h^2 + \varepsilon^2/h + \varepsilon h + \varepsilon^2) \qquad (1.69)$$

holds.

We will proof the same theorem in more general case in the next section.

Notice that according to (1.69) the approximate solution has the second-order accuracy with respect to h for $\varepsilon << h$, in particular, for $\varepsilon \le h^{3/2}$. The numerical experiments presented in Chapter 3 confirm this result. Thus, for $\varepsilon < h$ the constructed scheme is more accurate in comparison with other well-known methods, for example, with the first-order scheme from [122].

1.3 The finite element method with nonlinear quadrature rule

In this section the monotonicity (1.32) of the function $b(x)$ is not required because of the application of the nonlinear quadrature rule for the approximation of the convective term in the bilinear form. Theoretically this condition is not too restrictive, but in practice it is inconvenient, for example, when the function $b(x)$ is given discretely.

1.3.1 Construction of the quadrature rule

We return to the approximation of the bilinear form (1.21). The first term is integrated exactly for any $u \in S_h$, $v \in T_h$. For the second one we use the following quadrature rule on each interval:

$$\int_{x_{i-1}}^{x_i} bv\,dx \approx (\alpha_i b_{i-1} + \beta_i b_i)\,(\alpha_i v_{i-1} + \beta_i v_i)\,h_i. \tag{1.70}$$

Unlike the similar formula (1.36), in this case in the points for the calculation of the values of the functions b and v are choose individually with the help of the parameters α_i and β_i on each interval $[x_{i-1}, x_i]$. When we use (1.70) for a, we obtain the new bilinear form a^h of an algebraic type for $v, w^h \in S_h$:

$$\begin{aligned} a^h(v, w^h) = \sum_{i=1}^{n} \Big(&\varepsilon(v_i - v_{i-1})/h_i \\ &- (\alpha_i b_{i-1} + \beta_i b_i)\,(\alpha_i v_{i-1} + \beta_i v_i)\Big)(w_i^h - w_{i-1}^h). \end{aligned} \tag{1.71}$$

As before, we choose the parameters α_i, β_i so that the bilinear forms a and a^h are as close as possible just for the function ρ_0. Generally speaking, the exact equality

$$\int_{x_{i-1}}^{x_i} b\rho\,dx = (\alpha_i b_{i-1} + \beta_i b_i)\,(\alpha_i \rho_{0,i-1} + \beta_i \rho_{0,i})\,h_i$$

should be taken. However, this condition contains the integral in the left-hand side, that does not permit to obtain the explicit expression in the general case. Therefore for convenience we replace $b(x)$ by its value $b(x) \approx b_i^* = \alpha_i b_{i-1} + \beta_i b_i$ on $[x_{i-1}, x_i]$. Thus we arrive at the equality

$$\begin{aligned} &\int_{x_{i-1}}^{x_i} b_i^* \exp(-b_i^*(1-x)/\varepsilon)\,dx \\ &= b_i^*\big(\alpha_i \exp(-b_i^*(1-x_{i-1})/\varepsilon) + \beta_i \exp(-b_i^*(1-x_i)/\varepsilon)\big)h_i. \end{aligned} \tag{1.72}$$

Taking the integral in the left-hand side and dividing the obtained equality by $h_i \exp(-b_i^*(1-x_i)/\varepsilon)$, we get

$$\alpha_i \exp(-\sigma_i) + \beta_i = (1 - \exp(-\sigma_i))\,/\sigma_i \tag{1.73}$$

where $\sigma_i = b_i^* h_i/\varepsilon$. To the above equality we add the equation

$$\alpha_i + \beta_i = 1 \tag{1.74}$$

which permits to approximate an integral of a smooth function with the first-order accuracy. Thus, we arrive at the system of linear algebraic equations in two unknowns.

Notice that for the function b with the constant value $b_{const,i}$ on the segment $[x_{i-1}, x_i]$ the system (1.73) – (1.74) becomes linear:

$$\alpha_{const,i} \exp(-\sigma_i) + \beta_{const,i} = \frac{1}{\sigma_i}\left(1 - \exp(-\sigma_i)\right),$$
$$\alpha_{const,i} + \beta_{const,i} = 1.$$

Its solution is obtained in the same way as in [122]:

$$\alpha_{const,i} = \frac{1}{1 - \exp(-\sigma_i)} - \frac{1}{\sigma_i}, \quad \beta_{const,i} = \frac{1}{\sigma_i} - \frac{\exp(-\sigma_i)}{1 - \exp(-\sigma_i)}.$$

In particular, for any positive $b_{const,i}$ this solution satisfies the inequalities

$$1/2 < \alpha_{const,i} < 1, \quad 0 < \beta_{const,i} < 1/2 \quad \forall\, h_i, \varepsilon > 0.$$

Further we consider the case

$$\varepsilon \le h^2 \tag{1.75}$$

which is of practical importance. We express β_i from the system (1.73) – (1.74):

$$\beta_i = \frac{1}{\sigma_i} - \frac{\exp(-\sigma_i)}{1 - \exp(-\sigma_i)}. \tag{1.76}$$

From (1.74) it follows that

$$\alpha_i = 1 - \beta_i. \tag{1.77}$$

Taking into account the definition of b_i^* and (1.77), we can write σ_i as

$$\sigma_i = \left(b_{i-1} + \beta_i \left(b_i - b_{i-1}\right)\right) h_i/\varepsilon.$$

Hence there exists at least one solution α_i, β_i of the system (1.73) – (1.74) with the properties

$$1/2 < \alpha_i < 1, \quad 0 < \beta_i < 1/2.$$

This follows from the fact that for $\beta_i = 0$ the left-hand side of (1.76) is smaller than the right-hand one, and the opposite is true for $\beta_i = 1$. Therefore, the root can be found by the bisection method.

Now consider the remaining part of the solution

$$g(x) = v_0(x) + \varepsilon z_1(x).$$

In a similar way as in (1.44), using the fact that the functions v_0'', v_1', and z_1 are bounded on $[0,1]$, we can show that the quadrature rule (1.70) has only the first-order accuracy:

$$\int_{x_{i-1}}^{x_i} b(x)g(x)\,dx = b_i^* g_i^* h_i + (1/2 - \beta_i)h_i^2 (bg)_{i-1}' + O(h_i^3). \qquad (1.78)$$

We calculate $(b(x)g(x))'$ in (1.78) only at the interior points of the domain Ω. Therefore, considering the inequality (1.75) and the definition of ρ_0 and ρ_1, we obtain the estimate

$$-\varepsilon\left(\rho_0 + \varepsilon\rho_1\right)'' + \left(b\left(\rho_0 + \varepsilon\rho_1\right)\right)' \le c_9\varepsilon \quad \text{for} \quad x \le 1 - h_n.$$

Taking into account this estimate, we transform the main term of the error in (1.78) as follows:

$$(bg)' = (bg)' - \varepsilon g'' + O(\varepsilon) = (bu)' - \varepsilon u'' + O(\varepsilon) = f + O(\varepsilon).$$

Then instead of (1.78) we get

$$\int_{x_{i-1}}^{x_i} b(x)g(x)\,dx = b_i^* g_i^* h_i + (1/2 - \beta_i)h_i^2 f_{i-1} + O(h_i^3 + \varepsilon h_i^2 + \varepsilon^2 h_i).$$

Recall that when constructing the bilinear forms a and a^h all the terms are multiplied by $-(w^h)'$. Therefore the main term of the error on the segment $[x_{i-1}, x_i]$ has the form

$$-(1/2 - \beta_i)h_i^2 f_{i-1}(w^h)_{i-1/2}'.$$

We use the same functional of the right-hand side as in (1.51) with the coefficients (1.50) to eliminate this term.

Substituting the bilinear form $a(\cdot,\cdot)$ and the right-hand side (f, w^h) into (1.35), we obtain the discrete problem: *find* $u^h \in S_h$ *such that* $u^h(0) = u_0$, $u^h(1) = u_1$, *and*

$$a^h(u^h, w^h) = f_h(w^h) \qquad \forall\, w^h \in T_h. \qquad (1.79)$$

We rewrite this problem in the equivalent matrix-vector form: *construct the function*

$$u^h = \sum_{i=0}^{n} \tau_i \varphi_i$$

with the weights τ_i *which satisfy the conditions* $\tau_0 = u_0$, $\tau_n = u_1$ *as well as the system of linear algebraic equations*

$$A^h \tau = F \tag{1.80}$$

with the vector of unknowns

$$\tau = (\tau_1, ..., \tau_{n-1})^T$$

and the given the right-hand side

$$F^h = (F_1^h, ..., F_{n-1}^h)^T$$

where

$$\begin{aligned} F_1^h &= \mu_1 f_0 + \nu_1 f_1 + a_1 u_0, \\ F_i^h &= \mu_i f_{i-1} + \nu_i f_i, \qquad\qquad i = 2, ..., n-2, \\ F_{n-1}^h &= \mu_{n-1} f_{n-2} + \nu_{n-1} f_{n-1} + e_{n-1} u_1. \end{aligned}$$

The matrix A^h has the tridiagonal form

$$A^h = \begin{pmatrix} d_1 & -e_1 & & & & \\ -a_2 & d_2 & -e_2 & & 0 & \\ \cdots & \cdots & \cdots & & & \\ & \cdots & \cdots & \cdots & & \\ & & \cdots & \cdots & \cdots & \\ 0 & & & -a_{n-2} & d_{n-2} & -e_{n-2} \\ & & & & -a_{n-1} & d_{n-1} \end{pmatrix}$$

where

$$\begin{aligned} a_i &= \varepsilon / h_i + \alpha_i b_i^*, \\ d_i &= \varepsilon / h_i + \varepsilon / h_{i+1} + \alpha_{i+1} b_{i+1}^* - \beta_i b_i^*, \\ e_i &= \varepsilon / h_{i+1} - \beta_{i+1} b_{i+1}^*. \end{aligned} \tag{1.81}$$

1.3.2 Properties of the discrete problem

Now we investigate the discrete problem.

Lemma 9. *When the conditions* (1.2), (1.4), (1.75) *are satisfied for any* h, $\varepsilon > 0$ *the inequalities*

$$1/2 < \alpha_i < 1, \quad 0 < \beta_i < 1/2 \tag{1.82}$$

hold.

The lemma is proved in the same way as Lemma 3.2 from [122].

Lemma 10. *When the conditions* (1.2), (1.4), (1.75) *are satisfied for any* h, $\varepsilon > 0$ *the matrix* A^h *of the system* (1.80) *is an* M*-matrix and hence is nonsingular.*

Proof. From (1.81), (1.82), and (1.2) it follows that $a_i > 0$ for all i. We show that $e_i > 0$ for all i. For e_i we have

$$e_i = \frac{\varepsilon}{h_{i+1}} - \beta_{i+1} b^*_{i+1}.$$

From (1.76) it follows that

$$\beta_{i+1} = \frac{1}{\sigma_{i+1}} - \frac{\exp(-\sigma_{i+1})}{1 - \exp(-\sigma_{i+1})}.$$

Collecting two last equalities and the definition $\sigma_{i+1} = b^*_{i+1} h_{i+1}/\varepsilon$, we get

$$e_i = \frac{b^*_{i+1}}{1 - \exp(-\sigma_{i+1})}.$$

Since the inequalities $\sigma_i > 0, \quad \exp(-\sigma_i) < 1$, and $b(x) \geq B_0 > 0$ hold for any ratio between ε and h, we have

$$e_i > 0 \quad \forall\, \varepsilon, h > 0.$$

For the system (1.80) the relations

$$\begin{aligned} d_1 - a_2 &> 0, \\ d_i - a_{i+1} - e_{i-1} &= 0, \quad i = 2, ..., n-2, \\ d_{n-1} - e_{n-2} &> 0 \end{aligned}$$

hold. Because a_i and e_i are positive for all i, the matrix A^h is diagonal-dominant along columns and strongly diagonal-dominant along the first and last ones. Taking into account the fact that the matrix A^h is irreducible [21], this completes the proof. □

Lemma 11. *Let* $W^h = (w_1, ..., w_{n-1})^T$ *be the solution of the problem*

$$\left(A^h\right)^T W^h = Q^h \tag{1.83}$$

with some right-hand side $Q^h = (q_1, ..., q_{n-1})^T$. *Under the restrictions* (1.2), (1.4), *and* (1.75) *the estimate*

$$\sum_{i=2}^{n-1} |w_i - w_{i-1}| + |w_1| + |w_{n-1}| \leq c_{10} \sum_{i=1}^{n-1} |q_i| \tag{1.84}$$

holds with a constant c_{10} *independent of* ε *and* h.

Proof. We use the inequality (1.66) from Lemma 7:

$$\sum_{i=1}^{n} |v_i| \leq 2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(\prod_{k=j+1}^{i-1}\frac{e_{k-1}}{a_{k+1}}\right)\frac{q_j}{a_{j+1}}. \tag{1.85}$$

Taking into consideration the definition of the coefficients a_k and e_k in (1.81) and the fact that b is bounded, in accordance with (1.2) we get

$$\left(\prod_{k=j+1}^{i-1}\frac{e_{k-1}}{a_{k+1}}\right)\frac{1}{a_{j+1}} = \frac{1}{b_i^*}\left(1-\exp(-\sigma_i)\right)\prod_{k=j+1}^{i-1}\exp(-\sigma_k)$$
$$\leq \frac{1}{B_0}\left(1-\exp(-B_1 h_i/\varepsilon)\right)\exp\left(-\frac{B_0}{\varepsilon}(x_{i-1}-x_j)\right).$$

Using the last inequality, we change the order of summation in (1.85):

$$\sum_{i=1}^{n}|v_i| \leq \frac{2}{B_0}\sum_{i=1}^{n}\left(1-\exp(-B_1 h_i/\varepsilon)\right)\sum_{j=1}^{i-1}\exp\left(-\frac{B_0}{\varepsilon}(x_{i-1}-x_j)\right)q_j$$
$$= \frac{2}{B_0}\sum_{j=1}^{n-1}q_j\sum_{i=j+1}^{n-1}\left(1-\exp(-B_1 h_{i+1}/\varepsilon)\right)\exp\left(-\frac{B_0}{\varepsilon}(x_i-x_j)\right)$$
$$\leq \frac{2}{B_0}\sum_{j=1}^{n-1}q_j\sum_{i=j+1}^{n-1}d\frac{\varepsilon}{x_i-x_j}.$$

Here we applied the inequalities $1-\exp(-t)\leq 1$ and $t\exp(-\alpha t)\leq d$ which are valid for $t\in(0,1)$ and $\alpha\geq 0$ with a constant d. Due to (1.75) and (1.34), the last sum over i can be estimated by a constant c_{11}. Taking into account the definition of v_i, we complete the proof of the estimate (1.84). □

In terms of the norms in the spaces S_h and T_h the estimate (1.84) for functions $w^h \in T_h$ has the form

$$\|(w^h)'\|_1 \leq c_9 |||w^h|||_{1,h}. \tag{1.86}$$

1.3.3 Convergence theorem

Now we consider the main result of this section.

Theorem 12. *Let u be the solution of the problem* (1.1), (1.3) *with the conditions* (1.2), (1.4), *and u^h be the solution of the problem* (1.80) *with the condition* (1.75). *Then the estimate*

$$\max_{0\leq i\leq n}|u_i^h - u_i| \leq c_{15}(h^2+\varepsilon h+\varepsilon+\varepsilon^2+\varepsilon^2/h) \tag{1.87}$$

holds.

Proof. By Theorem 5 for $p = \infty$ the estimate

$$\max_{0\le i\le n} |u_i^h - u_i| = \|u^h - u^I\|_{\infty,h}$$
$$\le \sup_{w^h \in T_h} \frac{|(f,w^h) - f_h(w^h) + a^h(u^h,w^h) - a(u^h,w^h)|}{|\!|\!| w^h |\!|\!|_{1,h}} \tag{1.88}$$

holds.

We denote by $g(x)$ the sum of the smooth component and the remainder term in the expansion (1.11):

$$g(x) = v_0 + \varepsilon v_1 + \varepsilon^2 z(x).$$

Then we can write

$$|(f,w^h) - f_h(w^h) + a^h(u^h,w^h) - a(u^h,w^h)| \le |a^h(\rho_0,w^h) - a(\rho_0,w^h)|$$
$$+\varepsilon|a^h(\rho_1,w^h) - a(\rho_1,w^h)| + |(f,w^h) - f_h(w^h) + a^h(g,w^h) - a(g,w^h)|. \tag{1.89}$$

Consider the first term in the right-hand side

$$\left|\sum_{i=1}^n \left(\varepsilon(\rho_{0,i} - \rho_{0,i-1})/h_i - b_i^* \rho_{0,i}^*\right)\left(w_i^h - w_{i-1}^h\right) - \int_0^1 (\varepsilon\rho_0' - b\rho_0)(w^h)'\,dx\right|$$
$$= \left|\int_0^1 b\rho_0(w^h)'\,dx - \sum_{i=1}^n b_i^*\rho_{0,i}^*\left(w_i^h - w_{i-1}^h\right)\right|.$$

Rewrite the term

$$A_i = \frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\rho_0\,dx - b_i^*\rho_{0,i}^*$$

as

$$A_i = \frac{1}{h_i}\int_{x_{i-1}}^{x_i} \left(b\rho + b\left(\rho_0 - \rho\right)\right)\,dx - b_i^*\rho_i^* - b_i^*\left(\rho_{0,i}^* - \rho_i^*\right)$$
$$\le \frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\rho\,dx - b_i^*\rho_i^* + c_{16}\varepsilon = \tilde{A}_i.$$

Here we use the estimate (1.19) from Lemma 3. Using the identity (1.72), we get

$$b_i^*\left(\alpha_i \exp(-(1-x_{i-1})b_i^*/\varepsilon) + \beta_i \exp(-(1-x_i)b_i^*/\varepsilon)\right)$$
$$= \frac{1}{h_i}\int_{x_{i-1}}^{x_i} b_i^* \exp(-(1-x)b_i^*/\varepsilon)\,dx.$$

Then we transform $\tilde{A}_i$ to the form

$$\begin{aligned}\tilde{A}_i &= b_i^* \beta_i \left(\exp(-(1-x_i)b_i^*/\varepsilon) - \exp(-(1-x_i)b_i/\varepsilon)\right) \\ &+ b_i^* \alpha_i \left(\exp(-(1-x_{i-1})b_i^*/\varepsilon) - \exp(-(1-x_{i-1})b_{i-1}/\varepsilon)\right) \\ &+ \frac{1}{h_i}\int_{x_{i-1}}^{x_i} (b(x) - b_i^*)\,\rho(x)\,dx \\ &+ \frac{b_i^*}{h_i}\int_{x_{i-1}}^{x_i} \left(\exp(-(1-x)b(x)/\varepsilon) - \exp(-(1-x)b_i^*/\varepsilon)\right)\,dx + c_{16}\varepsilon.\end{aligned} \tag{1.90}$$

The first term is estimated by the mean-value theorem:

$$\begin{aligned}&|\exp(-(1-x_i)b_i^*/\varepsilon) - \exp(-(1-x_i)b_i/\varepsilon)| \quad \le \\ &\quad |b_i^* - b_i|\,\frac{1-x_i}{\varepsilon}\exp(-(1-x_i)\tilde{b}/\varepsilon) \quad \text{where} \quad \tilde{b} \in [B_0, B_1].\end{aligned} \tag{1.91}$$

Besides, $|b_i^* - b_i| \le h_i \|b'\|_\infty$ and the function $t^2 \exp(-B_0 t)$ is bounded by a constant c_{17} on $(0,\infty)$. Using (1.2) and (1.82), we obtain

$$\begin{aligned}&b_i^* \beta_i\,|\exp(-(1-x_i)b_i^*/\varepsilon) - \exp(-(1-x_i)b_i/\varepsilon)| \\ &\le c_{17} B_1 \beta_i \|b'\|_\infty \frac{\varepsilon h_i}{1-x_{i-1}} \le c_{18}\varepsilon.\end{aligned}$$

In a similar way we estimate the second term in (1.90):

$$b_i^* \alpha_i\,|\exp(-(1-x_{i-1})b_i^*/\varepsilon) - \exp(-(1-x_{i-1})b_{i-1}/\varepsilon)| \le c_{19}\varepsilon.$$

The third term is also estimated with the help of (1.2) and (1.82):

$$\begin{aligned}&\frac{1}{h_i}\left|\int_{x_{i-1}}^{x_i} (b(x) - b_i^*)\,\rho(x)\,dx\right| \le \|b'\|_\infty \int_{x_{i-1}}^{x_i} \exp(-(1-x)B_0/\varepsilon)\,dx \\ &= \|b'\|_\infty \frac{\varepsilon}{B_0}\left(\exp(-(1-x_i)B_0/\varepsilon) - \exp(-(1-x_{i-1})B_0/\varepsilon)\right) \le c_{20}\varepsilon.\end{aligned} \tag{1.92}$$

The integrand in the fourth term is estimated in the same way as in (1.91):

$$\begin{aligned}&|\exp(-(1-x)b(x)/\varepsilon) - \exp(-(1-x)b_i^*/\varepsilon)| \\ &\le h_i \|b'\|_\infty \frac{1-x}{\varepsilon}\exp(-(1-x)\tilde{\tilde{b}}/\varepsilon), \quad \text{where} \quad \tilde{\tilde{b}} \in [B_0, B_1].\end{aligned} \tag{1.93}$$

Using this inequality, we obtain the estimate of the fourth term in (1.90):

$$\begin{aligned}
&\left|\frac{b_i^*}{h_i}\int_{x_{i-1}}^{x_i} \left(\exp(-(1-x)b(x)/\varepsilon) - \exp(-(1-x)b_i^*/\varepsilon)\right)\, dx\right| \\
&\le c_{22}\int_{x_{i-1}}^{x_i} \frac{1-x}{\varepsilon}\exp(-(1-x)\tilde{\tilde{b}}/\varepsilon)\, dx \\
&\le c_{23}\varepsilon\left(1+\frac{(1-x)\tilde{\tilde{b}}}{\varepsilon}\right)\exp(-(1-x)\tilde{\tilde{b}}/\varepsilon)\Big|_{x_{i-1}}^{x_i} \\
&\le c_{24}\varepsilon\left(\exp(-(1-x_i)\tilde{\tilde{b}}/\varepsilon) - \exp(-(1-x_{i-1})\tilde{\tilde{b}}/\varepsilon)\right) \le c_{24}\varepsilon.
\end{aligned} \tag{1.94}$$

Summarizing the estimates (1.91)–(1.94), we can write

$$|A_i| \le c_{26}\varepsilon. \tag{1.95}$$

Thus, for the first term in (1.89) the following estimate holds:

$$\begin{aligned}
|a^h(\rho_0, w^h) - a(\rho_0, w^h)| &\le \left|\sum_{i=1}^{n} A_i(w_i - w_{i-1})\right| \le c_{27}\varepsilon\sum_{i=1}^{n}|w_i - w_{i-1}| \\
&\le c_{28}\varepsilon\|(w^h)'\|_{1,h} \le c_{29}\varepsilon|\!|\!|(w^h)'|\!|\!|_{1,h}\ .
\end{aligned} \tag{1.96}$$

Estimate the second term in (1.89):

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\left(\varepsilon(\rho_{1,i}-\rho_{1,i-1})/h_i - b_i^*\rho_{1,i}^*\right)\left(w_i^h - w_{i-1}^h\right) - \int_0^1 (\varepsilon\rho_1' - b\rho_1)(w^h)'\, dx\right| \\
&\qquad\qquad = \left|\int_0^1 b\rho_1(w^h)'\, dx - \sum_{i=1}^{n} b_i^*\rho_{1,i}^*\left(w_i^h - w_{i-1}^h\right)\right|.
\end{aligned}$$

Consider the expression

$$B_i = \frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\rho_1\, dx - b_i^*\left(\alpha_i\rho_{1,i-1} + \beta_i\rho_{1,i}\right).$$

Taking into consideration the form of the function ρ_1, we can estimate the first term in the above expression:

$$\left|\frac{1}{h_i}\int_{x_{i-1}}^{x_i} b\rho_1\, dx\right| \le c_{30}\varepsilon/h_i.$$

The second term is bounded due to the estimates (1.2) and (1.82). Thus, we have

$$B_i \le c_{31}\varepsilon/h_i.$$

Hence, we obtain

$$\begin{aligned}|a^h(\rho_1, w^h) - a(\rho_1, w^h)| &\le \left|\sum_{i=1}^{n} B_i(w_i - w_{i-1})\right| \\ &\le \varepsilon c_{32}\sum_{i=1}^{n}\frac{1}{h_i}|w_i - w_{i-1}| \le \frac{\varepsilon}{h}c_{33}\|(w^h)'\|_{1,h} \le \frac{\varepsilon}{h}c_{34}\vert\vert\vert(w^h)'\vert\vert\vert_{1,h}.\end{aligned} \tag{1.97}$$

This estimate is worse than (1.96). However, the boundary layer component ρ_1 and the estimate (1.97) have to be multiplied by ε, that gives the same order of convergence.

Finally, consider the last term in (1.89):

$$\begin{aligned}&\Big|\int_0^1 fw^h\,dx - \sum_{i=1}^{n}(\mu_i f_{i-1} + \nu_i f_i)w_i^h - \int_0^1(\varepsilon g' - bg)(w^h)'\,dx \\ &\quad - \sum_{i=1}^{n}\Big(\varepsilon(g_i - g_{i-1})/h_i - b_i^* g_i^*\Big)(w_i^h - w_{i-1}^h)\Big| \\ &\le \Big| -\int_0^1 F(x)(w^h)'\,dx + \int_0^1 bg(w^h)' - \\ &\quad - \sum_{i=0}^{n-1}\left(\left((F_{i+1}^* - F_i^*) + \eta_i h_i^3\right)w_i^h + b_i^* g_i^*(w_i^h - w_{i-1}^h)\right)\Big| \\ &= \Big|\int_0^1 (bg - F)(w^h)' + \sum_{i=1}^{n}\left(F_i^*(w_i^h - w_{i-1}^h) + \eta_i h^3 w_i^h - b_i^* g_i^*(w_i^h - w_{i-1}^h)\right)\Big|\end{aligned}$$

where $F(x)$ is the antiderivative of $f(x)$ and η_i are the values of a function bounded on $[0,1]$. Consider the term

$$\begin{aligned}C_i = \frac{1}{h_i}\int_{x_{i-1}}^{x_i}&(bg - F)\,dx + \alpha_i F_{i-1} + \beta_i F_i \\ &- (\alpha_i b_{i-1} - \beta_i b_i)\,(\alpha_i g_{i-1} - \beta_i g_i)\,.\end{aligned}$$

Using the expansion of the functions bv_0, bv_1 and $b^*v_0^*$, $b^*v_1^*$ into the Taylor series at the point x_{i-1} and taking into account the fact that bz, b^*z^* are bounded, we get

$$\frac{1}{h_i}\int_{x_{i-1}}^{x_i} bg\,dx - b^*g^* = (1/2 - \beta_i)\,h_i(bg)'_{i-1} + O(h^2 + \varepsilon h + \varepsilon^2). \tag{1.98}$$

In view of the definitions of the boundary layer components ρ_0 and ρ_1 the estimate

$$-\varepsilon\,(\rho_0+\varepsilon\rho_1)''+(b\,(\rho_0+\varepsilon\rho_1))'\le c_{35}\varepsilon\quad\forall\, x\in[0,1]$$

holds. With the last inequality we have

$$(bg)'=(bg)'-\varepsilon g''+O(\varepsilon)=(bu)'-\varepsilon u''+O(\varepsilon)=f+O(\varepsilon).$$

Than the main term of the error in (1.98) has the form

$$-(1/2-\beta_i)h_if_{i-1}(w^h)'_{i-1/2}+O(h^2+\varepsilon h).$$

From the identity (1.48) we obtain

$$-\frac{1}{h_i}\int_{xi-1}^{x_i}F(x)\,dx+F_i^*=-(1/2-\beta_i)h_if_{i-1}+O(h^2).$$

Thus, the estimate

$$|C_i|\le c_{36}(h^2+\varepsilon h+\varepsilon^2)$$

holds and we have

$$\begin{aligned}&|(f,w^h)-f_h(w^h)+a^h(g,w^h)-a(g,w^h)|\le\left|\sum_{i=1}^{n}C_i(w_i^h-w_{i-1^h})+h_i^3\eta_iw_i\right|\\&\le c_{37}(h^2+\varepsilon h+\varepsilon^2)\sum_{i=1}^{n}|w_i^h-w_{i-1^h}|\le c_{38}(h^2+\varepsilon h+\varepsilon^2)\|(w^h)'\|_{1,h}\qquad(1.99)\\&\le c_{39}(h^2+\varepsilon h+\varepsilon^2)|\!|\!|w^h|\!|\!|_{1,h}.\end{aligned}$$

Finally, the estimate (1.87) follows from the relations (1.88), (1.89) and estimates (1.96), (1.97), and (1.99). □

Notice that, as before, for $\varepsilon\ll 1$ and even for $\varepsilon\le h$ in the case of practical importance, the accuracy of the obtained solution in accordance with the estimate (1.87) is of the second order with respect to h. This is confirmed by numerical experiments in Chapter 3.

As a result, for $\varepsilon<h$ the constructed scheme is more accurate than the similar one of the first-order accuracy from [122]. Therefore we concentrate our efforts upon this case. It should be noted that the convergence is proved for a non-uniform grid.

Now we discuss the question connecting the calculation of the coefficients α_i and β_i of the nonlinear system (1.73) – (1.74) on each interval $[x_{i-1},x_i]$.

Taking into account (1.76), we have the following formula for the coefficient β_i:

$$\beta_i = \frac{1}{\sigma_i} - \frac{\exp(-\sigma_i)}{1-\exp(-\sigma_i)}, \text{ where } \sigma_i = (b_{i-1} + \beta_i\,(b_i - b_{i-1}))\,h_i/\varepsilon.$$

Since the derivative b' is bounded, we have

$$b_i = b_{i-1} + h_{i-1}\delta_{i-1} \quad \text{where} \quad |\delta_{i-1}| \le \|b'\|_\infty.$$

Using this notation, from the system (1.73) – (1.74) we get

$$\begin{aligned} \beta_i = &-\frac{\exp(-\sigma_i)}{1-\exp(-\sigma_i)} \\ &+ \frac{1}{b_{i-1} + h_{i-1}\delta_{i-1}\beta_i}\left(\frac{b_{i-1}}{\sigma_{i-1}} + \frac{h_{i-1}\delta_{i-1}}{\sigma_{i-1}(1-\exp(-\sigma_{i-1}))} - \frac{h_{i-1}\delta_{i-1}}{\sigma_{i-1}^2}\right). \end{aligned} \tag{1.100}$$

Taking into account (1.2) and (1.75), we obtain the following estimate of the derivative of the right-hand side of (1.100) with respect to β_i:

$$\left|\frac{h_{i-1}\delta_{i-1}}{(b_{i-1} + h_{i-1}\delta_{i-1}\beta_i)^2}\left(\frac{b_{i-1}}{\sigma_{i-1}} + \frac{h_{i-1}\delta_{i-1}}{\sigma_{i-1}(1-\exp(-\sigma_{i-1}))} - \frac{h_{i-1}\delta_{i-1}}{\sigma_{i-1}^2}\right)\right| \le c_{40}\varepsilon$$

with a constant c_{40} independent of ε and h. Thus, for $\varepsilon \ll 1$ the right-hand side of (1.100) is a contraction operator on $[0,1]$ with a sufficiently small contraction coefficient of order ε. Therefore we define β_i as the limit of the iterative process

$$\beta_i = \lim_{j\to\infty} s_j$$

where

$$\begin{aligned} s_0 &= \beta_{const,i}, \\ s_{j+1} &= \frac{\varepsilon}{h_i(b_{i-1} + s_j(b_i - b_{i-1}))} \\ &- \frac{\exp(-(b_{i-1} + s_j(b_i - b_{i-1}))h_i/\varepsilon)}{1-\exp(-(b_{i-1} + s_j(b_i - b_{i-1}))h_i/\varepsilon)}. \end{aligned} \tag{1.101}$$

Then $\alpha_i = 1 - \beta_i$ is determined from (1.74).

The numerical experiments confirm the fast of convergence of the iterative process (1.101). When calculations were performed for the model problem, 2-4 iterations were need to obtain an accuracy of 10^{-7}.

2 Two-dimensional convection-diffusion problem

2.1 General remarks

2.1.1 Qualitative behaviour of the solution

Let Ω be the unit square $(0,1) \times (0,1)$ with boundary Γ. Consider the Dirichlet problem

$$Lu \equiv -\varepsilon \Delta u + \frac{\partial}{\partial x}(b(x)u) = f \qquad \text{in} \quad \Omega, \tag{2.1}$$

$$u = 0 \qquad \text{on} \quad \Gamma. \tag{2.2}$$

Here, as usually, $\varepsilon \ll 1$ is a positive small parameter. The functions $b(x)$ and $f(x,y)$ are sufficiently smooth:

$$b \in C^3[0,1], \quad f(x,y) \in C^3(\bar{\Omega}). \tag{2.3}$$

Under these assumptions the problem (2.1), (2.2) has a unique solution in $C^2(\Omega)$ (see, e.g., [88]).

The behaviour of the solution in the two-dimensional case is more complicated than in the one-dimensional case. In addition to the exponentional (regular) boundary layer, as in Chapter 1, there is a parabolic boundary layer that arises near some parts of the boundary. The boundary layer of this type is formed due to the fact that the characteristics of the reduced problem (for $\varepsilon = 0$) is tangent to the boundary. Besides, corner boundary layers can arise at the vertices of square.

Let the conditions

$$0 < B_1 \le b(x) \le B_2 < \infty, \quad x \in [0,1]; \tag{2.4}$$

$$f(0,0) = f(1,0) = f(0,1) = f(1,1) = 0 \tag{2.5}$$

be fulfilled. Then the solution of the problem (2.1) – (2.2) belongs to $C^3(\bar{\Omega})$ ([41]). Notice that the derivatives up to the third order are continuous and hence are bounded on $\bar{\Omega}$. But the constants in the estimates of this derivatives depend on ε and increase indefinitely as ε tends to zero. As for the fourth derivatives, they belong to $C(\bar{\Omega}')$ for any subdomain $\Omega' \subset \Omega$ with positive distance from 4 corners. Therefore fourth derivatives are continuous everywhere in $\bar{\Omega}$ except 4 corners.

Let us introduce the notations

$$\Gamma_{in} = \{(x,y) : x = 0, \quad y \in (0,1)\},$$
$$\Gamma_{out} = \{(x,y) : x = 1, \quad y \in (0,1)\},$$
$$\Gamma_{tg} = \{(x,y) : x \in (0,1), \quad y = 0,1\}.$$

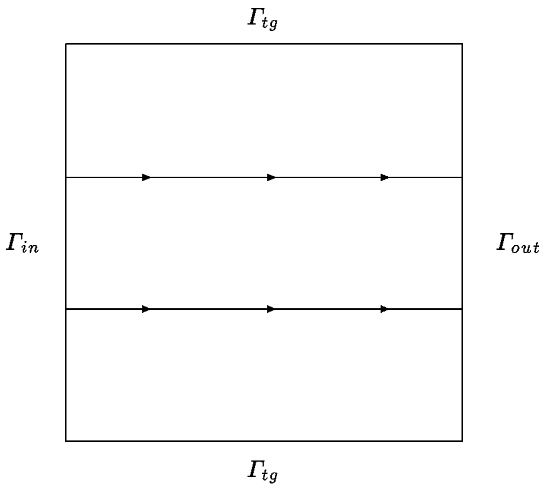


Fig. 2. Domain Ω.

Here the regular boundary layer arises near Γ_{out} and the parabolic boundary layer arises along Γ_{tg} (see Fig. 2).

Notice that in general the operator corresponding to the left-hand side of (2.1) with the mixed boundary conditions does not satisfy the maximum principle (for example, for $b' < 0$), however the comparison principle still holds. Later the comparison principle is applied to the differential operator of the form

$$\mathcal{L}u \equiv -\varepsilon \Delta u + b\frac{\partial u}{\partial x} + du \tag{2.6}$$

where $b(x)$ satisfies the assumptions (2.3), (2.4) and $d(x)$ is a bounded function on $[0,1]$ that is defined in each individual case.

Lemma 13. *Let $\varepsilon > 0$ be small enough. Assume that* (2.3), (2.4) *hold and $u, w \in C^2(\Omega) \bigcap C(\bar{\Omega})$ satisfy*

$$|\mathcal{L}u| \le \mathcal{L}w \quad in \quad \Omega, \qquad |u| \le w \quad on \quad \Gamma. \tag{2.7}$$

Then the estimate

$$|u| \le w \qquad on \quad \bar{\Omega} \tag{2.8}$$

is valid.

Proof. Introduce the functions

$$v(x,y) = u(x,y)\exp(-\sigma x) \qquad \text{and} \qquad z(x,y) = w(x,y)\exp(-\sigma x) \tag{2.9}$$

with the constant

$$\sigma = 1 + \max_{x\in[0,1]} \frac{|d(x)|}{b(x)}. \tag{2.10}$$

Transform the differential operator $\mathcal{L}$ into

$$\left(\widetilde{\mathcal{L}}\Phi(x,y)\right) \equiv \exp(-\sigma x)\mathcal{L}\left(\Phi(x,y)\exp(\sigma x)\right), \tag{2.11}$$

that gives

$$\widetilde{\mathcal{L}}\Phi(x,y) = -\varepsilon\,\Delta\Phi(x,y) + (b(x) - 2\varepsilon\sigma)\frac{\partial\Phi(x,y)}{\partial x} + (d(x) + \sigma b(x) - \varepsilon\sigma^2)\Phi(x,y).$$

Assume that

$$\varepsilon \in \left(0, B_1/(4\sigma^2)\right]. \tag{2.12}$$

Taking into consideration the definition of σ and the smallness of ε, we obtain

$$\begin{aligned} d + \sigma b - \varepsilon\sigma^2 &= -|d| + b + |d| - B_1/2 \geq B_1/2 \geq 0 \quad \text{on} \quad [0,1], \\ b - 2\varepsilon\sigma \geq b - 2B_1\sigma/4\sigma^2 &= b - B_1/2\sigma \geq B_1/2 \quad \text{on} \quad [0,1]. \end{aligned} \tag{2.13}$$

From (2.7) we have

$$\begin{aligned} &|\widetilde{\mathcal{L}}v| \leq \widetilde{\mathcal{L}}z \qquad \text{in} \quad \Omega, \\ &|v| \leq z \qquad \text{on} \quad \Gamma. \end{aligned}$$

The operator $\widetilde{\mathcal{L}}$ satisfies the maximum principle (see [115]). As a consequence we obtain

$$|v| \leq z \qquad \text{on} \quad \bar{\Omega}.$$

Multiplying the last inequality by $\exp(\sigma x)$, we get (2.8). □

When on Γ_{tg} we can estimate not a function u but its normal derivative only, the comparison principle also holds.

Lemma 14. *Let* $\varepsilon > 0$ *be small enough. Assume that* (2.3), (2.4) *hold, and* $u, w \in C^2(\Omega \cup \Gamma_{tg}) \bigcap C(\bar{\Omega})$ *satisfy*

$$\begin{aligned} &|\mathcal{L}u| \leq \mathcal{L}w \qquad \text{in} \quad \Omega, \\ &|u| \leq w \qquad \text{on} \quad \Gamma \setminus \Gamma_{tg}, \qquad \left|\frac{\partial u}{\partial n}\right| \leq \frac{\partial w}{\partial n} \quad \text{on} \quad \Gamma_{tg}. \end{aligned} \tag{2.14}$$

Then the estimate

$$|u| \leq w \qquad \text{on} \quad \bar{\Omega} \tag{2.15}$$

be valid.

Proof. We introduce the constant σ by (2.10) and assume that ε satisfies (2.12). We use (2.9) and consider the operator $\tilde{\mathcal{L}}$ from (2.11). Then $\tilde{\mathcal{L}}$ satisfies the maximum principle again. Notice that on Γ_{tg} we have $\partial/\partial n = \pm\partial/\partial y$, therefore

$$|\tilde{\mathcal{L}}v| \le \tilde{\mathcal{L}}z \qquad \text{in} \quad \Omega,$$

$$|v| \le z \qquad \text{on} \quad \Gamma \setminus \Gamma_{tg}, \qquad \left|\frac{\partial v}{\partial n}\right| \le \frac{\partial z}{\partial n} \quad \text{on} \quad \Gamma_{tg}.$$

First we prove (by contradiction) the statement of the lemma in the case of $v = 0$ and, consequently, $z \ge 0$ on $\bar{\Omega}$. For this purpose we suppose that there exists a point $(x, y) \in \bar{\Omega}$ where $z(x, y) < 0$. Assume that at a point $(x_0, y_0) \in \bar{\Omega}$ we have

$$z(x_0, y_0) = \min_{\bar{\Omega}} z(x, y) < 0. \tag{2.16}$$

Since $\tilde{\mathcal{L}}$ satisfies the maximum principle, (x_0, y_0) does not belong to Ω. Because of the condition on $\Gamma \setminus \Gamma_{tg}$ the point (x_0, y_0) does not belong to this part of the boundary. It remains that $(x_0, y_0) \in \Gamma_{tg}$. Assume that, for definiteness, $y_0 = 1$. Because of the condition on Γ_{tg} we have

$$\frac{\partial z}{\partial n}(x_0, 1) = \frac{\partial z}{\partial y}(x_0, 1) \ge 0.$$

If $\partial z/\partial y(x_0, 1) > 0$ then due to continuity there exists an interval $[1-\delta, 1]$ in y on which this inequality holds. Use the Taylor expansion

$$z(x_0, 1-\delta) = z(x_0, 1) - \delta\frac{\partial z}{\partial y}(x_0, \eta), \quad \eta \in [1-\delta, 1].$$

It implies $z(x_0, 1-\delta) < z(x_0, 1)$ that is in contradiction with (2.16). Therefore

$$\frac{\partial z}{\partial y}(x_0, 1) = 0.$$

Applying this reasoning to the second derivative, we obtain

$$\frac{\partial^2 z}{\partial y^2}(x_0, 1) \ge 0.$$

In a similar way we get

$$\frac{\partial z}{\partial x}(x_0, 1) = 0 \quad \text{and} \quad \frac{\partial^2 z}{\partial x^2}(x_0, 1) \ge 0.$$

Using the above four relations in the expression $(\widetilde{\mathcal{L}}z)(x_0, 1)$, we obtain

$$(\widetilde{\mathcal{L}}z)(x_0, 1) \leq z(x_0, 1)B_1/2 < 0.$$

This is in contradiction with the condition $\widetilde{\mathcal{L}}z \geq 0$ on $\bar{\Omega}$ which follows from the same condition on Ω and from the continuity of v and its first and second derivatives. Thus, our assumption that v can take negative values is wrong. Hence, $z \geq 0$ on $\bar{\Omega}$.

Finally, using the last statement for the functions $z - v$ and $z + v$, we obtain $z - v \geq 0, \;\; z + v \geq 0$. Hence $|v| \leq z$ on $\bar{\Omega}$ that implies

$$|u| \leq w \quad \text{on} \quad \bar{\Omega}. \qquad \square$$

2.1.2 The weak formulation

Multiply (2.1) by an arbitrary function $v \in H_0^1(\Omega)$. By applying Green's formula we obtain the weak formulation: *find* $u \in H_0^1(\Omega)$ *such that for all* $v \in H_0^1(\Omega)$

$$a(u, v) = (f, v) \tag{2.17}$$

with the bilinear form

$$a(u, v) = \int_\Omega \left(\varepsilon \, \nabla u \nabla v - bu \frac{\partial v}{\partial x} \right) \, d\Omega \tag{2.18}$$

and the inner product

$$(f, v) = \int_\Omega fv \, d\Omega. \tag{2.19}$$

Let us introduce the norm

$$\|v\|_\infty = \sup_{\bar{\Omega}} \text{vrai} \, |v|.$$

We use the notations ∂_1 for a partial derivative $\partial/\partial x$ and ∂_2 for a partial derivative $\partial/\partial y$. Similarly we denote the second derivatives by $\partial_{22} = \partial_2(\partial_2)$ and so on.

2.2 The scheme with the fitted quadrature rule for a problem without parabolic boundary layers

In this section we consider the method for the problem (2.1) – (2.2) with a solution free of a parabolic boundary layer near Γ_{tg}.

2.2.1 The differential problem

Assume that

$$f(x,y)=0 \qquad \text{on} \quad \varGamma_{tg}. \tag{2.20}$$

Then under conditions (2.4), (2.20) the first and second partial derivatives of the solution of the problem (2.1) – (2.2) with respect to y are bounded. Namely, the following estimates hold.

Lemma 15. *Assume that* $0<\varepsilon\ll 1$ *and* (2.3), (2.4), (2.20) *are valid for the problem* (2.1)–(2.2). *Then we have*

$$\|u\|_\infty+\|\partial_2 u\|_\infty+\|\partial_{22}u\|_\infty\le c_1. \tag{2.21}$$

Proof. Assume that $d=b'(x)$ and σ is given by (2.10). Take the barrier function

$$w(x,y)=c_2\exp(\sigma x) \quad \text{where} \quad c_2=2\|f\|_\infty/B_1.$$

Taking into consideration (2.13), (2.2), and (2.4) we have

$$\mathcal{L}w(x,y)\ge\|f\|_\infty\ge|Lu(x,y)| \quad \text{in} \quad \varOmega,$$

$$w(x,y)\ge|u(x,y)| \quad \text{on} \quad \varGamma.$$

Thus, applying Lemma 13 and using the upper bound of the function w, we conclude that the solution u is bounded uniformly with respect to ε.

To prove the estimate for the first derivative on $\varOmega$, we differentiate the equation (2.1) with respect to y and introduce the notation $v_1=\partial_2 u$. Then we get

$$\mathcal{L}v_1=\partial_2 f \quad \text{in} \quad \varOmega.$$

Since $u(0,y)=u(1,y)=0$, we obtain

$$v_1=0 \quad \text{on} \quad \varGamma\setminus\varGamma_{tg}. \tag{2.22}$$

From (2.1), (2.2), (2.20) we have

$$\partial_2 v_1=\partial_{22}u=0 \quad \text{on} \quad \varGamma_{tg}. \tag{2.23}$$

Now, setting $c_2=2\|\partial_2 f\|_\infty/B_1$ and taking into account (2.13), (2.4), (2.22), and (2.23), we see that the barrier function $w(x,y)=c_2\exp(\sigma x)$ satisfies the relations

$$|\mathcal{L}v_1|\le\mathcal{L}w \quad \text{in} \quad \varOmega,$$

$$|v_1|\le w \quad \text{on} \quad \varGamma\setminus\varGamma_{tg}, \qquad \left|\frac{\partial v_1}{\partial n}\right|\le\frac{\partial w}{\partial n} \quad \text{on} \quad \varGamma_{tg}.$$

Thus, by Lemma 14 v_1 is bounded on $\bar{\Omega}$ by the function w which satisfies the estimate $w \leq c_2 \exp(\sigma)$ on $\bar{\Omega}$. Hence, $\partial_2 u$ is uniformly bounded on $\bar{\Omega}$ by a constant independent of ε.

It remains to show that the second derivative $\partial_{22}u$ is bounded. To do this, we twice differentiate (2.1) with respect to y, put $v_2 = \partial_{22}u$, and to use Lemma 13 with the same barrier function $w(x,y)$ and with the constant $c_2 = 2\|\partial_{22}f\|_\infty / B_1$. □

Let us consider the following expansion of the solution

$$u = v_0 + \rho_0 + \varepsilon\eta. \tag{2.24}$$

Here v_0 is the solution of the reduced problem

$$\partial_1 \left(b(x)v_0\right) = f(x,y) \quad \text{in} \quad \Omega, \tag{2.25}$$

$$v_0 = 0 \quad \text{on} \quad \Gamma_{in}. \tag{2.26}$$

The function ρ_0 is the regular boundary layer component

$$\rho_0(x,y) = g(y)s(x)\exp\left(-(1-x)b(x)/\varepsilon\right) \tag{2.27}$$

where $g(y) = -v_0(1,y)$ and $s(t)$ is the cut-off function $s \in C^3([0,1])$ satisfying (1.9). The solution of the problem (2.25), (2.26) has the form

$$v_0(x,y) = \frac{1}{b(x)} \int_0^x f(t,y)\,dt. \tag{2.28}$$

Due to (2.3), (2.4) we have $v_0 \in C^3(\bar{\Omega})$. Because of (2.20)

$$v_0(x,y) = 0 \quad \text{on} \quad \Gamma_{tg}.$$

In view of the definitions of ρ_0, g, s and with (2.20) we get

$$\rho_0(x,y) = 0 \quad \text{on} \quad \Gamma \setminus \Gamma_{out}, \quad \rho_0(1,y) = -v_0(1,y) \quad \text{on} \quad \Gamma_{out}. \tag{2.29}$$

To estimate the remainder term in the expansion (2.24) we need the following lemma.

Lemma 16. *Assume that* $\varepsilon > 0$ *is small enough and* (2.4) *hold. Let* $s(x) \in C^3[0,1]$ *be the cut-off function* (1.9). *Then the function*

$$\psi_1(x) = \frac{1-x}{\varepsilon} \exp(-(1-x)b(x)/\varkappa\varepsilon)s(x) \tag{2.30}$$

with a fixed constant $\varkappa \in (1,2)$ *satisfies the inequality*

$$\mathcal{L}\psi_1 \geq c_1 \left(\frac{1}{\varepsilon} + \frac{1-x}{\varepsilon^2}\right) \exp(-(1-x)b(x)/\varkappa\varepsilon) - c_2 \tag{2.31}$$

for the operator $\mathcal{L}$ *from* (2.6) *with the function* $d = b'(x)$ *and some positive constants* c_1 *and* c_2 *independent of* ε.

Proof. We introduce the notation

$$E_1(x,\varepsilon) = \exp\left(-\frac{(1-x)b(x)}{\varkappa\varepsilon}\right).$$

Then we obtain the relation

$$\begin{aligned}\mathcal{L}\psi_1 = {} & \frac{1}{\varepsilon}b(x)\left(\frac{2}{\varkappa}-1\right)s(x)E_1(x,\varepsilon) + \frac{1-x}{\varepsilon^2}b^2(x)\left(\frac{1}{\varkappa}-\frac{1}{\varkappa^2}\right)s(x)E_1(x,\varepsilon) \\ & + \frac{1-x}{\varepsilon}a_1(x)s(x)E_1(x,\varepsilon) + \left(\frac{1-x}{\varepsilon}\right)^2 a_2(x)s(x)E_1(x,\varepsilon) \qquad (2.32) \\ & + \left(2+\frac{1-x}{\varepsilon}\right)a_3(x)s'(x)E_1(x,\varepsilon) + (1-x)s''(x)E_1(x,\varepsilon)\end{aligned}$$

with bounded functions a_1, a_2, a_3. First we consider the right-hand side of (2.32) on the segment $[2/3,1]$. Remember that $s=1$ and $s'=s''=0$ on this segment. Due to the definition of $\varkappa$ the coefficients $2/\varkappa-1$ and $1/\varkappa-1/\varkappa^2$ are positive. Since $b(x)\ge B_1>0$, the sum of the first and second terms in the right-hand side of (2.32) has the lower bound

$$c_3\left(\frac{1}{\varepsilon}+\frac{1-x}{\varepsilon^2}\right)E_1(x,\varepsilon).$$

To estimate the remaining two nonzero terms, we use the inequality

$$x^\alpha\exp(-\beta x)\le(\alpha/\beta)^\alpha\exp(-\alpha),\quad x\in[0,\infty) \qquad (2.33)$$

which holds for each $\alpha\ge 0,\ \ \beta>0$ (see [4]). Setting $t=(1-x)/\varepsilon$ and $t=(1-x)^2/\varepsilon^2$, and using the fact that a_1 and a_2 are bounded, we estimate these two terms from below by a negative constant $-c_4$. Hence we have

$$\mathcal{L}\psi_1\ge c_3\left(\frac{1}{\varepsilon}+\frac{1-x}{\varepsilon^2}\right)E_1(x,\varepsilon)-c_4 \quad \text{on} \quad [2/3,1]. \qquad (2.34)$$

Now we consider (2.32) on the segment $x\in[0,2/3]$. The right-hand side can be expressed as

$$\mathcal{L}\psi_1=\left(a_4(x)+\frac{1}{\varepsilon}a_5(x)+\frac{1}{\varepsilon^2}a_6(x)\right)E_1(x,\varepsilon)$$

where the functions a_4,a_5,a_6 are bounded on $[0,2/3]\times[0,1]$. Ones, we use (2.33) for $t=1/\varepsilon$ and $\alpha=0,1,2$. This gives

$$\mathcal{L}\psi_1\ge -c_5 \quad \text{on} \quad [0,2/3]. \qquad (2.35)$$

In a similar way for the expression in the right-hand side of (2.31) we obtain the upper bound

$$\left(\frac{1}{\varepsilon}+\frac{1-x}{\varepsilon^2}\right)E_1(x,\varepsilon)\le c_6 \quad \text{on} \quad [0,2/3]. \tag{2.36}$$

Let us set

$$c_1=c_3 \quad \text{and} \quad c_2=\max\{c_4, c_3c_6+c_5\}. \tag{2.37}$$

Thus, the estimate (2.34) involves (2.31) on the segment $[2/3,1]$. On the remaining segment $[0,2/3]$ from (2.35)–(2.37) we get

$$\mathcal{L}\psi_1 \ge -c_5 \ge c_3c_6-c_2 \ge c_1\left(\frac{1}{\varepsilon}+\frac{1-x}{\varepsilon^2}\right)E_1(x,\varepsilon)-c_2.$$

This estimate together with (2.34) completes the proof. □

Lemma 17. *Assume that $\varepsilon>0$ is small enough and* (2.4) *hold. Then the function*

$$\psi_2(x)=(1-\exp(-(1-x)B_2/\varepsilon))\,(\exp(\sigma x)-1) \tag{2.38}$$

with the constant σ from (2.10) *satisfies the inequality*

$$(\mathcal{L}\psi_2)\,(x,y)\ge\frac{1}{4}B_1\exp\sigma x \quad \textit{on} \quad \Omega \tag{2.39}$$

where the operator $\mathcal{L}$ is given by (2.6) *with $d=b'(x)$.*

Proof. Introduce the notation

$$E_2(x,\varepsilon)=\exp\left(-\frac{(1-x)B_2}{\varepsilon}\right).$$

Then we obtain the relation

$$\begin{aligned}\mathcal{L}\psi_2 =\ & \frac{B_2^2}{\varepsilon}E_2(x,\varepsilon)\,(\exp(\sigma x)-1)+2\sigma B_2E_2(x,\varepsilon)\exp(\sigma x)\\ & -\varepsilon\sigma^2\,(1-E_2(x,\varepsilon)))\exp(\sigma x)-b\frac{B_2}{\varepsilon}E_2(x,\varepsilon)(\exp(\sigma x)-1)\\ & +b\sigma\,(1-E_2(x,\varepsilon))\exp(\sigma x)+b'\,(1-E_2(x,\varepsilon))\,(\exp(\sigma x)-1).\end{aligned}$$

Due to the upper estimate of $b(x)$ in (2.4) the sum of the first and fourth terms is nonnegative. We discard it and use simple transformations:

$$\begin{aligned}\mathcal{L}\psi_2 \ge\ & (-\varepsilon\sigma^2+b\sigma-|b'|)\exp(\sigma x)\\ & +(2\sigma B_2+\varepsilon\sigma^2-b\sigma-|b'|)\exp(-(1-x)B_2/\varepsilon)\exp(\sigma x).\end{aligned}$$

For $\varepsilon \le B_1/(2\sigma^2)$ in view of (2.10) the inequality

$$|b'(x)| \le b(x)\sigma/2.$$

Then we have

$$-\varepsilon\sigma^2+b(x)\sigma-|b'(x)| \ge \frac{1}{4}B_1, \quad 2\sigma B_2+\varepsilon\sigma^2-b(x)\sigma-|b'(x)| \ge \varepsilon\sigma^2+\frac{1}{2}B_1 \ge 0.$$

Hence we obtain

$$\mathcal{L}\psi_2 \ge \frac{1}{4}B_1\exp(\sigma x).$$

That completes the proof of the lemma.□

Lemma 18. *Let* $\varepsilon > 0$ *be small enough and the operator* $\mathcal{L}$ *be defined by* (2.6) *with a bounded function* $d(x)$. *Then the function*

$$\psi_3(x) = \left(1+\frac{1}{\varepsilon}\exp(-(1-x)B_1/2\varepsilon)\right)\exp(\sigma x) \tag{2.40}$$

with the constant σ *from* (2.10) *satisfies the inequality*

$$\begin{aligned}\mathcal{L}\psi_3 \ge {} & \frac{B_1^2}{8\varepsilon^2}\exp(-(1-x)B_1/2\varepsilon)\exp\sigma x \\ & +\frac{B_1}{2}\left(1+\exp(-(1-x)B_1/2\varepsilon)\right)\exp\sigma x.\end{aligned} \tag{2.41}$$

Proof. Introduce the notation

$$E_3(x,\varepsilon) = \exp\left(-\frac{(1-x)B_1}{2\varepsilon}\right)$$

and assume that

$$\varepsilon \le \min\left\{\frac{B_1}{2\sigma^2}, \frac{B_1}{8\sigma}\right\}. \tag{2.42}$$

Then we obtain the relation

$$\begin{aligned}(\mathcal{L}\psi_3)(x,y) = {} & \left(\frac{B_1b_1}{2\varepsilon^2}-\frac{B_1^2}{4\varepsilon^2}-\frac{B_1\sigma}{\varepsilon}\right)E_3(x,\varepsilon)\exp(\sigma x) \\ & +\left(-\varepsilon\sigma^2+b_1\sigma+d\right)\left(1+\varepsilon^{-1}E_3(x,\varepsilon)\right)\exp(\sigma x).\end{aligned}$$

Because of (2.4) and (2.42) the factor in the first term is estimated from below:

$$\frac{B_1b_1}{2\varepsilon^2}-\frac{B_1^2}{4\varepsilon^2}-\frac{B_1\sigma}{\varepsilon} \ge \frac{B_1^2}{4\varepsilon^2}-\frac{B_1\sigma}{\varepsilon} \ge \frac{B_1^2}{8\varepsilon^2}.$$

The factor in the second term is evaluated from below due to (2.13):

$$-\varepsilon\sigma^2 + b_1\sigma + d \geq B_1/2.$$

These three inequalities involve (2.41). □

The following Lemma describes the behaviour of the remainder term in the expansion (2.24) and of its derivatives.

Lemma 19. *Let* $\varepsilon > 0$ *be small enough and* (2.3), (2.4), (2.20) *be valid for the problem* (2.1)–(2.2). *Then the remainder term* η *in* (2.24) *satisfies the estimates*

$$\|\eta\|_\infty \leq c_7, \tag{2.43}$$
$$|\partial_1\eta(x,y)| \leq c_8(1+\varepsilon^{-1}\exp(-B_1(1-x)/2\varepsilon)), \quad (x,y)\in\bar{\Omega}, \tag{2.44}$$
$$\|\partial_{22}\eta\|_\infty \leq c_9\varepsilon^{-1}. \tag{2.45}$$

Proof. First we set $d = b'(x)$ in (2.6). Then we get $\mathcal{L}u = f$. Simple calculations show that η in (2.24) satisfies

$$L\eta = \tilde{f} \equiv a_0(x,y) + \frac{1}{\varepsilon}a_1(x,y)A(x,\varepsilon) + \frac{1-x}{\varepsilon^2}a_2(x,y)A(x,\varepsilon) \text{ on } \bar{\Omega} \tag{2.46}$$

where

$$A(x) = \exp(-(1-x)b(x)/\varepsilon)$$

and a_0, a_1, a_2 are bounded functions on $\bar{\Omega}$. Therefore the right-hand side of (2.46) is estimated in the following way:

$$|L\eta| \leq c_{10} + \left(c_{11}\frac{1}{\varepsilon} + c_{12}\frac{1-x}{\varepsilon^2}\right)A(x,\varepsilon) \tag{2.47}$$

with appropriate constants c_{10}, c_{11}, and c_{12}. Let us use the barrier function

$$w(x,y) = c_{13}\psi_1(x,y) + c_{14}\psi_2(x,y)$$

with constants

$$c_{13} = \max\{c_{11}, c_{12}\}/c_1 \quad \text{and} \quad c_{14} = 4(c_{10}+c_2c_3)/B_1$$

where the functions ψ_1, ψ_2 are given in Lemmata 16 and 17. This yields

$$|(\mathcal{L}\eta)(x,y)| \leq (\mathcal{L}w)(x,y) \quad \text{in } \Omega.$$

Moreover, we have

$$w \geq 0 = |\eta| \quad \text{on} \quad \Gamma.$$

Thus, all the assumptions of Lemma 16 are satisfied and consequently (2.43) holds. Together with the estimates for the functions ψ_1, ψ_2 this implies the inequality

$$|\eta(x,y)| \leq w(x,y). \tag{2.48}$$

Moreover, since

$$w(0,y) = w(1,y) = 0 \quad \forall\, y \in \ [0,1],$$

from (2.48) we have

$$|\partial_1 \eta(0,y)| \leq c_{15} \quad \text{and} \quad |\partial_1 \eta(1,y)| \leq c_{16}\varepsilon^{-1}. \tag{2.49}$$

In order to prove (2.44) we differentiate (2.46) with respect to x. Introduce the notation $\zeta = \partial_1 \eta$ and set $d = 2b'$ in (2.6). Then we obtain

$$\begin{aligned}(\mathcal{L}\zeta)(x,y) &= a_3(x,y)\\ &+ \frac{1}{\varepsilon^2} a_4(x,y)A(x,\varepsilon) + \frac{1-x}{\varepsilon^3} a_5(x,y)A(x,\varepsilon)\end{aligned} \tag{2.50}$$

where the functions a_3, a_4, and a_5 are bounded on $\bar{\Omega}$. The right-hand side of (2.50) is estimated in the following way:

$$|(\mathcal{L}\zeta)(x,y)| \leq c_{17} + c_{18}\varepsilon^{-2}\exp(-B_1(1-x)/(2\varepsilon)). \tag{2.51}$$

Now we take the barrier function $w(x,y) = c_{19}\psi_3(x,y)$ from Lemma 18 with the constant $c_{19} = \max\{8c_{17}/B_1^2,\ 2c_{19}/B_1\}$. We get

$$|(\mathcal{L}\zeta)(x,y)| \leq (\mathcal{L}w)(x,y) \quad \text{in } \Omega,$$

$$w \geq 0 = |\zeta| \quad \text{on} \quad \Gamma_{tg}, \quad w \geq |\zeta| \quad \text{on} \quad \Gamma_{in} \cup \Gamma_{out}.$$

The last inequality follows from (2.49). Thus, due to Lemma 13 we obtain

$$|\zeta| \leq w \quad \text{on} \quad \bar{\Omega}$$

that involves (2.44).

In order to prove (2.45) we consider the equality

$$\partial_{22}\eta = \frac{1}{\varepsilon}(\partial_{22}u - \partial_{22}v_0 - \partial_{22}\rho_0) \tag{2.52}$$

which follows from (2.24). Taking into consideration (2.21), (2.27), and (2.28), we obtain (2.45). □

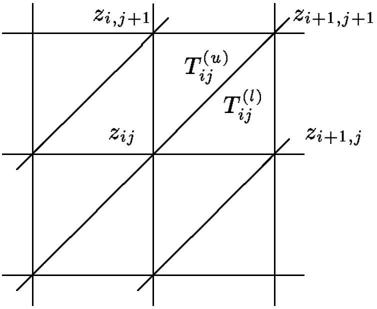


Fig. 3. The fragment of the triangulation $\mathcal{T}_h$.

2.2.2 Construction of the quadrature rule

For the implementation of the Galerkin method we construct an uniform triangulation $\mathcal{T}_h$. To do this, we consider the grid

$$x_i = ih, \qquad y_j = jh, \qquad i,j = 0,1,\dots,n,$$

with the mesh size $h = 1/n$ for integer $n \geq 2$. We denote the set of nodes by

$$\bar{\Omega}_h = \{z_{ij} = (x_i, y_j),\ i,j = 0,1,\dots,n\},$$

the set of interior nodes by

$$\Omega_h = \{z_{ij} = (x_i, y_j),\ i,j = 1,2,\dots,n-1\},$$

and the set of boundary nodes by

$$\Gamma_h = \{z_{ij} = (x_i, y_j),\ i = 0,1,\ j = 0,1,\dots,n;\ i = 0,1,\dots,n,\ j = 0,1\}.$$

Then the triangulation $\mathcal{T}_h$ is constructed by dividing each elementary rectangle $\Omega_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ into two *elementary* triangles by the diagonal passing from (x_i, y_j) to (x_{i+1}, y_{j+1}) (see Fig. 3).

At each node $z_{ij} \in \Omega_h$ we introduce the basis function φ_{ij} which equals 1 at the node z_{ij}, equals 0 at any other node of $\bar{\Omega}_h$, and is linear on each elementary triangle of $\mathcal{T}_h$. Denote the linear span of these functions by

$$H^h = \operatorname{span}\{\varphi_{ij}\}_{i,j=1}^{n-1}.$$

With these notations, we arrive at the Galerkin problem: *find $u^h \in H^h$ such that*

$$a(u^h, v^h) = (f, v^h) \qquad \forall\, v^h\ \in H^h. \tag{2.53}$$

But the solution of this problem is unstable and has poor accuracy because of the boundary layer component ([9]). In the same way as in one-dimensional case, we provide the stability and improve the accuracy by the special approximation of the bilinear form a with the fitted quadrature rule.

Let $T^{(l)}_{ij}$ (or $T^{(u)}_{ij}$, respectively) be an arbitrary triangle of $\mathcal{T}_h$ with the vertices $z_{i,j}$, $z_{i+1,j+1} = (x_{i+1}, y_{j+1})$, and $z_{i+1,j} = (x_{i+1}, y_j)$ ($z_{i+1,j} = (x_i, y_{j+1})$, respectively) as in Fig. 3. We denote the elementary part of the bilinear form (2.18) on an arbitrary triangle $T = T^{(l)}_{ij}$ or $T = T^{(u)}_{ij}$ by

$$a_T(u,v) = \int_T ((\varepsilon\partial_1 u - bu)\,\partial_1 v + \varepsilon\partial_2 u\partial_2 v)\,d\Omega. \tag{2.54}$$

In principle, freezing the coefficient b on a triangle T is enough to perform the integration exactly. But the accuracy of this formula is unsatisfactory because of the boundary layer function ρ_0. Therefore, we try to get another quadrature rule.

Thus, we apply the three-point quadrature rule on a triangle $T = T^{(l)}_{ij}$ for the approximation of the bilinear form (2.54):

$$\int_T g(x,y)d\Omega \approx \frac{h^2}{2}\left(\alpha_{1i}g(z_{ij}) + \alpha_{2i}g(z_{i+1,j}) + \alpha_{3i}g(z_{i+1,j+1})\right).$$

Then an elementary contribution of the algebraic bilinear form can be expressed as

$$\begin{aligned} a^h_{T^{(l)}_{ij}}(w^h,v^h) = \frac{h^2}{2}\Big(\big(\varepsilon\partial_1 w^h - b_i(\alpha_{1i}w^h(z_{i,j}) + \alpha_{2i}w^h(z_{i+1,j}) \\ + \alpha_{3i}w^h(z_{i+1,j+1}))\big)\partial_1 v^h + \varepsilon\partial_2 w^h\partial_2 v^h\Big). \end{aligned} \tag{2.55}$$

From here on we use the notation $b_i = b(x_i)$. We choose the weights $\alpha_{k,i}$ from the following two requirements. Firstly, in order to guarantee the first order accuracy for smooth functions, the quadrature rule have to be exact for constant functions. This immediately gives the equation

$$\alpha_{1i} + \alpha_{2i} + \alpha_{3i} = 1. \tag{2.56}$$

Secondly, we try to minimize the difference

$$a_T(\rho_0, v^h) - a^h_T(\rho^I_0, v^h), \qquad v^h \in H^h, \tag{2.57}$$

for the regular boundary layer function ρ_0 and its piecewise linear interpolant $\rho_0^I \in H^h$. For this purpose we put

$$\begin{aligned}&\int_T (\varepsilon\partial_x\zeta_i - b_i\zeta_i)\,\partial_1 v^h d\Omega \\ &= \frac{h^2}{2}\Big(\varepsilon\partial_1\zeta_i^I - b_i\left(\alpha_{1i}\zeta_i^I(z_{i,j}) + \alpha_{2i}\zeta_i^I(z_{i+1,j})\alpha_{3i}\zeta_i^I(z_{i+1,j+1})\right)\Big)\partial_1 v^h\end{aligned} \tag{2.58}$$

for the function

$$\zeta_i(x) = \exp\left(-(1-x)b_i/\varepsilon\right)$$

and its piecewise linear interpolant $\zeta_i^I(x,y)$ on $T_{ij}^{(l)}$.

To diminish the difference stencil, we put $\alpha_{3i} = 0$. Thus, for the parameters of the quadrature rule we have the system of linear algebraic equations

$$\begin{aligned}\alpha_{1i} + \exp(\sigma_i)\alpha_{2i} + \exp(\sigma_i)\alpha_{3i} &= \frac{1}{\sigma_i}\left(\exp\sigma_i - 1\right), \\ \alpha_{1i} + \alpha_{2i} + \alpha_{3i} &= 1, \\ \alpha_{3i} &= 0\end{aligned} \tag{2.59}$$

where $\sigma_i = b_i h/\varepsilon$. It has the unique solution

$$\alpha_{1i} = \frac{\exp\sigma_i}{(\exp\sigma_i - 1)} - \frac{1}{\sigma_i}, \quad \alpha_{2i} = \frac{1}{\sigma_i} - \frac{1}{\exp\sigma_i - 1}, \quad \alpha_{3i} = 0. \tag{2.60}$$

With the weights obtained we rewrite (2.55) in the following form:

$$\begin{aligned}a^h_{T^{(l)}_{ij}}(w^h, v^h) = \frac{h^2}{2}\Big(&\frac{b_i}{\exp\sigma_i - 1}(w^h_{i+1,j} - w^h_{ij}\exp\sigma_i)\partial_1 v^h \\ &+ \varepsilon\partial_2 w^h\partial_2 v^h\Big).\end{aligned} \tag{2.61}$$

From here on we use the notation $v_{ij} = v(z_{ij})$ for any function $v(x,y)$.

In a similar way on the triangle $T_{ij}^{(u)}$ we obtain the following approximation of the bilinear form (2.54):

$$\begin{aligned}a^h_{T^{(u)}_{ij}}(w^h, v^h) = \frac{h^2}{2}\Big(&\frac{b_i}{\exp\sigma_i - 1}(w^h_{i+1,j+1} - w^h_{i,j+1}\exp\sigma_i)\frac{\partial v^h}{\partial x} \\ &+ \varepsilon\frac{\partial w^h}{\partial y}\frac{\partial v^h}{\partial y}\Big).\end{aligned} \tag{2.62}$$

To integrate the right-hand side, we use the simple quadrature rules

$$\int_{T_{ij}^{(l)}} fv\,d\Omega \approx \frac{1}{6}h^2(f_{ij}v_{ij} + f_{i+1,j}v_{i+1,j} + f_{i+1,j+1}v_{i+1,j+1}),$$
$$\int_{T_{ij}^{(u)}} fv\,d\Omega \approx \frac{1}{6}h^2(f_{i,j+1}v_{i,j+1} + f_{i+1,j}v_{i+1,j} + f_{i+1,j+1}v_{i+1,j+1}).$$

This gives the elementary terms of the approximation of the right-hand side on an element $T \in \mathcal{T}_h$:

$$\begin{aligned} f^h_{T_{ij}^{(l)}}(v^h) &= \frac{1}{6}h^2(f_{ij}v_{ij} + f_{i+1,j}v_{i+1,j} + f_{i+1,j+1}v_{i+1,j+1}), \\ f^h_{T_{ij}^{(u)}}(v^h) &= \frac{1}{6}h^2(f_{i,j}v_{i,j} + f_{i,j+1}v_{i,j+1} + f_{i+1,j+1}v_{i+1,j+1}). \end{aligned} \tag{2.63}$$

Summing the elementary terms like (2.61), (2.62), and (2.63) over all $T \in \mathcal{T}_h$, we obtain the approximations of the bilinear and linear forms

$$a^h(w^h, v^h) = \sum_{T\in\mathcal{T}_h} a^h_T(w^h, v^h), \tag{2.64}$$

$$f^h(v^h) = \sum_{T\in\mathcal{T}_h} f^h_T(v^h).$$

Now we come to the 'fitted' Galerkin problem: *find* $u^h \in H^h$ *such that*

$$a^h(u^h, v^h) = f^h(v^h) \qquad \forall\, v^h \in H^h. \tag{2.65}$$

This problem is equivalent to the system of linear algebraic equations

$$\begin{aligned} (L^h u^h)_{ij} \equiv u^h_{ij}\left(h\left(\frac{b_i \exp\sigma_i}{\exp\sigma_i - 1} + \frac{b_{i-1}}{\exp\sigma_{i-1} - 1}\right) + 2\varepsilon\right) - u^h_{i+1,j}\frac{b_i h}{\exp\sigma_i - 1} \\ - u^h_{i-1,j}\frac{b_{i-1}h\exp\sigma_{i-1}}{\exp\sigma_{i-1} - 1} - \varepsilon u^h_{i,j-1} - \varepsilon u^h_{i,j+1} \\ = f_{ij}h^2, \quad i,j = 1,2,...,n-1, \end{aligned} \tag{2.66}$$

where $u^h_{ij} = 0$ for $i = 1,...,n-1$ and $j = 0,n$ or for $j = 1,...,n-1$ and $i = 0,n$. The parameters $\{u^h_{ij}\}_{i,j=1}^{n-1}$ give the solution of the problem (2.65)

$$u^h = \sum_{i,j=1}^{n-1} u_{ij}\varphi_{ij}. \tag{2.67}$$

Enumerate the remaining unknowns and the equations in (2.66) from 1 to $(n-1)^2$ in the same way (for example, in the lexicographic order) and rewrite the system (2.66) in the vector-matrix form

$$A^h U = F \tag{2.68}$$

where

$$\begin{aligned} U &= (u^h_{1,1}, ..., u^h_{1,n-1}, u^h_{2,1}, ..., u^h_{n-1,n-1})^T, \\ F &= (f^h(\varphi_{1,1}), ..., f^h(\varphi_{1,n-1}), ..., f^h(\varphi_{n-1,n-1}))^T. \end{aligned} \tag{2.69}$$

Notice that the matrix A^h is irreducible [21], diagonal-dominant along columns and strongly diagonal-dominant along columns for $i = 0, n$. Consequently, A^h is an M-matrix and the system (2.66) satisfies the difference comparison principle and has a unique solution [21].

2.2.3 Properties of the discrete problem. The convergence result

Now we investigate the approximating properties of the discrete problem (2.65).

Lemma 20. *Let u be a solution of the problem* (2.1), (2.2) *with the conditions* (2.3), (2.4), (2.20), *and u^h be a solution of the discrete problem* (2.65). *Assume also that*

$$\varepsilon \le h. \tag{2.70}$$

Then the estimate

$$\begin{aligned} &|a^h(u^h - u^I, \varphi_{ij})| \\ &\qquad \le ch^2(\varepsilon + h + \exp(-B_1(1 - x_{i+1})/2\varepsilon)) \quad \forall\, i, j = 1, ..., n-1 \end{aligned} \tag{2.71}$$

holds.

Proof. Using the expansion (2.24) we have

$$\begin{aligned} |a^h(u^h - u^I, \varphi_{ij})| &\le |f^h(\varphi_{ij}) - f(\varphi_{ij})| + |a(u, \varphi_{ij}) - a^h(u^I, \varphi_{ij})| \\ &\le |f^h(\varphi_{ij}) - f(\varphi_{ij})| + |a(v_0, \varphi_{ij}) - a^h(v_0^I, \varphi_{ij})| \\ &+ |a(\rho_0, \varphi_{ij}) - a^h(\rho_0^I, \varphi_{ij})| + \varepsilon |a(\eta, \varphi_{ij}) - a^h(\eta^I, \varphi_{ij})|. \end{aligned} \tag{2.72}$$

Here $v_0^I, \rho_0^I, \eta^I \in H^h$ are the piecewise linear interpolants of the functions v_0, ρ_0, η and $i, j = 1, ..., n-1$.

We evaluate each term in the right-hand side of (2.72). First we estimate each expression on an elementary triangle $T \in \mathcal{T}_h$ and then we get the estimate over the whole support of φ_{ij}. It is equivalent to the estimate over $\bar{\Omega}$.

Let us take $T = T_{ij}^{(l)}$. Consider the expression

$$F_T = \frac{h^2}{6} f_{ij} - \int_T f\varphi_{ij}\, d\Omega.$$

Since $f \in C^2(\bar{\Omega})$ we use the Taylor formula

$$f(x,y) = f_{ij} + h\pi_1(x,y), \qquad |\pi_1| \le c_1 \quad \text{on} \quad T.$$

This gives

$$|F_T| = |h \int_T \pi_1 \varphi_{ij}\, d\Omega| \le \frac{1}{6} c_1 h^3.$$

The same estimate is valid on any other elementary triangle $T \in \operatorname{supp} \varphi_{ij}$. Taking the sum over the whole support of φ_{ij}, we obtain

$$|f^s(\varphi_{ij}) - f(\varphi_{ij})| \le c_1 h^3. \tag{2.73}$$

Because of different smoothness of the solution in the x- and y-directions, we expand the bilinear forms (2.18) and (2.64) as a sum in the x- and y-directions:

$$\begin{aligned} a(u,v) &= a_1(u,v) + a_2(u,v), \\ a^h(u^h,v^h) &= a_1^h(u^h,v^h) + a_2^h(u^h,v^h). \end{aligned}$$

Then for the elementary bilinear forms (2.54) and (2.55) we get

$$\begin{aligned} a_T(u,v) &= a_{1T}(u,v) + a_{2T}(u,v), \\ a_T^h(u^h,v^h) &= a_{1T}^h(u^h,v^h) + a_{2T}^h(u^h,v^h) \end{aligned}$$

where

$$\begin{aligned} a_{1T}(u,v) &= \int_T (\varepsilon \partial_1 u - bu)\, \partial_1 v\, d\Omega, \\ a_{2T}(u,v) &= \varepsilon \int_T \partial_2 u \partial_2 v\, d\Omega, \end{aligned}$$

and

$$\begin{aligned} a_{1T}^h(w^h,v^h) &= \frac{h^2}{2}\big(\varepsilon \partial_1 w^h - b_i \left(\alpha_{1i} w^h(z_{ij}) + \alpha_{2i} w^h(z_{i+1,j})\right)\big)\partial_1 v^h, \\ a_{2T}^h(w^h,v^h) &= \frac{h^2}{2} \varepsilon \partial_2 w^h \partial_2 v^h. \end{aligned}$$

According to Lemma 15 the solution is sufficiently smooth in the y-direction, then it is easy to get the estimate of the difference

$$|a_2(u, \varphi_{ij}) - a_2^h(u^I, \varphi_{ij})|. \tag{2.74}$$

Consider the inequality

$$\begin{aligned} |a_2(u, \varphi_{ij}) - a_2^h(u^I, \varphi_{ij})| \le |a_2(u, \varphi_{ij}) - a_2(u^I, \varphi_{ij})| \\ + |a_2(u^I, \varphi_{ij}) - a_2^h(u^I, \varphi_{ij})| \end{aligned} \tag{2.75}$$

and estimate both terms in its right-hand side.

On $T = T_{ij}^{(u)}$ we have

$$A_{ij}^{(u)} = -\frac{1}{h} \int_T \varepsilon \left(\partial_2 u - \partial_2 u^I \right) d\Omega.$$

Use the Taylor expansion at z_{ij}

$$\begin{aligned} u(x_i, y_{j+1}) &= u(x_i, y_j) + h\partial_2 u(x_i, y_j) + h^2 \pi_1(x_i, y), \\ \partial_2 u(x, y) &= \partial_2 u(x_i, y_j) + h\pi_2(x_i, y) \end{aligned}$$

where $|\pi_1| \le c_8$ and $|\pi_2| \le c_8$ on T due to the estimate (2.21). Since u^I is the piecewise linear interpolant of u, T the equality

$$\partial_2 u^I(x, y) = \frac{u(x_i, y_{j+1}) - u(x_i, y_j)}{h} = \partial_2 u(x_i, y_j) + h\pi_1(x_i, y)$$

holds. Hence we have $|\partial_2 u - \partial_2 u^I| \le c_9 h$. Thus we obtain

$$|A_{ij}^{(u)}| \le c_{10} \varepsilon h^2.$$

The same contribution into the error comes from the triangles $T_{ij}^{(u)}$, $T_{i,j-1}^{(u)}$, $T_{i-1,j-1}^{(l)}$, and $T_{i-1,j}^{(l)}$. On the triangles $T_{ij}^{(l)}$ and $T_{i-1,j-1}^{(u)}$ the derivative $\partial_2 \varphi_{ij}$ equals zero and therefore these triangles do not make a contribution into the error. As a result, we have the following estimate of the first term in the right-hand side of (2.75):

$$|a_2(u, \varphi_{ij}) - a_2(u^I, \varphi_{ij})| \le c_{10} \varepsilon h^2.$$

Since the approximation of the second derivative gives the exact expression for linear functions, the second term in the right-hand side of (2.75) equals zero on any $T \in \mathcal{T}_h$. Finally, in the y-direction we have

$$|a_2(u, \varphi_{ij}) - a_2^h(u^I, \varphi_{ij})| \le c_{10} \varepsilon h^2. \tag{2.76}$$

To obtain the estimates in the x-direction, on a triangle $T \in \mathcal{T}_h$ we introduce the intermediate bilinear form $a^f_{1T}(u,v)$ obtained from $a_{1T}(u,v)$ by freezing the function $b(x)$ at the point x_i:

$$a^f_{1T}(u,v) = \int_T (\varepsilon\partial_1 u - b_i u)\,\partial_1 v d\Omega. \tag{2.77}$$

Then for an arbitrary function v we have the estimate

$$\begin{aligned}|a_1(v,\varphi_{ij}) - a^h_1(v^I,\varphi_{ij})| &\le |a_1(v,\varphi_{ij}) - a^f_1(v,\varphi_{ij})|\\ &+|a^f_1(v,\varphi_{ij}) - a^f_1(v^I,\varphi_{ij})| + |a^f_1(v^I,\varphi_{ij}) - a^h_1(v^I,\varphi_{ij})|.\end{aligned} \tag{2.78}$$

First with the help of (2.78) we obtain the estimate for $v = v_0$. On $T = T^{(l)}_{ij}$ we consider

$$B^{(l)}_{ij} = a_{1T}(v_0,\varphi_{ij}) - a^f_{1T}(v_0,\varphi_{ij}) = -\frac{1}{h}\int_T (b_i - b(x))v_0\,d\Omega.$$

Expand b and v_0 in the Tailor series at x_i:

$$\begin{aligned}&b(x) = b_i + (x - x_i)b'(x_i) + h^2\pi_3(x),\\ &v_0(x,y) = v_{0,ij} + h\pi_4(x,y).\end{aligned}$$

Due to the smoothness of b and v_0, the functions $\pi_3(x)$ and $\pi_4(x,y)$ are bounded on T. As a result, we have

$$B^{(l)}_{ij} = -\frac{h^2}{3}v_{0,ij}b'(x_i) + h^3\pi_5, \qquad |\pi_5| \le c_{11}.$$

In a similar way on $T^{(l)}_{i-1,j}$ we obtain the equality

$$B^{(l)}_{i-1,j} = \frac{h^2}{3}v_{0,i-1,j}b'(x_{i-1}) + h^3\pi_6, \qquad |\pi_6| \le c_{11}$$

with the same constant c_{11} independent of ε, h, i, j. Therefore due to the smoothness of b and v_0 we have

$$|B^{(l)}_{ij} + B^{(l)}_{i-1,j}| \le 2c_{11}h^3.$$

The same contribution comes from the triangles $T^{(u)}_{i,j-1}$ and $T^{(u)}_{i-1,j-1}$. On the triangles $T^{(u)}_{ij}$ and $T^{(l)}_{i-1,j-1}$ we have $\partial_1\varphi_{ij} = 0$ and consequently these triangles do not make a contribution into the error. As a result, we get

$$|a_1(v_0,\varphi_{ij}) - a^f_1(v_0,\varphi_{ij})| \le 4c_{11}h^3. \tag{2.79}$$

Next according to (2.78) we evaluate the difference

$$|a_1^f(v_0, \varphi_{ij}) - a_1^f(v_0^I, \varphi_{ij})|. \tag{2.80}$$

On $T = T_{ij}^{(l)}$ we introduce the notation

$$C_{ij}^{(l)} = -\frac{1}{h}\int_T \left(\varepsilon(\partial_1 v_0 - \partial_1 v_0^I) - b_i(v_0 - v_0^I)\right)\, d\Omega.$$

Due to the smoothness of v_0 as well as the definition of a piecewise linear interpolant, the equality

$$\partial_1 v_0^I(x,y) = \frac{v_{0,i+1,j} - v_{0,i,j}}{h} = \partial_1 v_0(x_i, y_j) + h\pi_7(x, y_j)$$

holds. Here $|\pi_7(x,y_j)| \leq c_{12}$ on T. Moreover, we have

$$\partial_1 v_0(x,y) = \partial_1 v_0(x_i, y_j) + h\pi_8(x,y)$$

where π_8 is bounded on T. Hence the estimate

$$|\partial_1 v_0 - \partial_1 v_0^I| \leq c_{13} h \tag{2.81}$$

is valid. For functions v_0 and v_0^I on T use Tailor expansion in the form

$$\begin{aligned} v_0(x,y) &= v_0(x_i,y_j) + (x - x_i)\partial_1 v_0(x_i,y_j) \\ &\quad + (y - y_j)\partial_2 v_0(x_i,y_j) + h^2\pi_9(x,y) \quad \text{where} \quad |\pi_9| \leq c_{14} \quad \text{on} \quad T, \end{aligned} \tag{2.82}$$

$$v_0^I(x,y) = v_0(x_i,y_j) + (x - x_i)\partial_1 v_0^I(x_i,y_j) + (y - y_j)\partial_2 v_0^I(x_i,y_j).$$

Due to the inequality (2.81) on T, the similar inequality in the y-direction

$$|\partial_2 v_0 - \partial_2 v_0^I| \leq c_{15} h,$$

and (2.82) we get the estimate

$$|v_0 - v_0^I| \leq c_{16} h^2.$$

Then we have

$$|C_{ij}^{(l)}| \leq c_{17} h^2(\varepsilon + h). \tag{2.83}$$

The same contribution comes from the triangles $T_{i-1,j}^{(l)}$, $T_{i,j-1}^{(u)}$, and $T_{i-1,j-1}^{(u)}$. The triangles $T_{ij}^{(u)}$ and $T_{i-1,j-1}^{(l)}$ do not make a the contribution into the error because of equality $\partial_1 \varphi_{ij} = 0$.

From (2.83) we obtain

$$|a_1^f(v_0,\varphi_{ij}) - a_1^f(v_0^I,\varphi_{ij})| \le 4c_{18}(\varepsilon + h)h^2. \qquad (2.84)$$

To complete the proof of the estimate (2.78) for $v = v_0$, we evaluate the term $|a_1^f(v_0^I,\varphi_{ij}) - a_1^h(v_0^I,\varphi_{ij})|$. On $T = T_{ij}^{(l)}$ we have

$$\begin{aligned} D_{ij}^{(l)} &= \frac{b_i}{h}(\int_T v_0^I \, d\Omega - \frac{h^2}{2}(\alpha_{1i}v_{0,i,j} + \alpha_{2i}v_{0,i+1,j})) \\ &= \frac{b_i h}{6}(v_{0,i,j} + v_{0,i+1,j} + v_{0,i+1,j+1} - 3\alpha_{1i}v_{0,i,j} - 3\alpha_{2i}v_{0,i+1,j}). \end{aligned}$$

Use the Taylor expansion (2.82) for $v_{0,i+1,j}$ and $v_{0,i+1,j+1}$ near (x_i, y_j). Since $\alpha_{1i} + \alpha_{2i} = 1$, we get

$$D_{ij}^{(l)} = \frac{b_i h^2}{6}((2 - 3\alpha_{2i})\partial_1 v_{0,i,j} + \partial_2 v_{0,i,j}) + h^3\pi_{9,ij}$$

where $\pi_{9,ij}$ is a function bounded on T. In a similar way on $T_{i-1,j}^{(l)}$ we obtain the expression

$$D_{i-1,j}^{(l)} = -\frac{b_i h^2}{6}((2 - 3\alpha_{2,i-1})\partial_1 v_{0,i-1,j} + \partial_2 v_{0,i-1,j}) + h^3\pi_{9,i-1,j}$$

with the value $\pi_{9,i-1,j}$ of the function π_9 bounded on T.

Consider the function

$$\widetilde{\alpha}(t) = t\left(\frac{\exp(t)}{(\exp(t) - 1)^2} - \frac{1}{t^2}\right).$$

This function approaches zero as $t \to 0$ or $t \to +\infty$. Since it is continuous on the interval $(0, \infty)$, it is bounded

$$|\widetilde{\alpha}(t)| \le c_{19} \qquad \text{on} \quad (0, \infty)$$

with a constant c_{19} independent of h, ε, x, t. Taking into account

$$\partial_1 \alpha_2(x) = \frac{b_1'(x)}{b_1(x)}\widetilde{\alpha}(hb_1(x)/\varepsilon),$$

(2.4), and the smoothness of b and its derivative, we get

$$|\partial_1 \alpha_2(x)| \le c_{19}\|b'\|_\infty / B_1.$$

Since the functions b, $\alpha_2(x)$, $\partial_1 v_0$, $\partial_2 v_0$ have bounded derivatives with respect to x, by the mean value theorem we get

$$|D_{ij}^{(l)} + D_{i-1,j}^{(l)}| \le c_{20} h^3.$$

The same contribution comes from the triangles $T_{i,j-1}^{(u)}$, $T_{i-1,j-1}^{(u)}$. The triangles $T_{ij}^{(u)}$ and $T_{i-1,j-1}^{(l)}$ do not make a contribution into the error because of the equality $\partial_1 \varphi_{ij} = 0$. As a result, we get

$$|a_1^f(v_0^I, \varphi_{ij}) - a_1^h(v_0^I, \varphi_{ij})| \le 2c_{20} h^3. \tag{2.85}$$

From (2.78) together with (2.79), (2.84), and (2.85) we obtain the estimate of the second term in (2.72)

$$|a_1(v_0, \varphi_{ij}) - a_1^h(v_0^I, \varphi_{ij})| \le 2c_{21} h^2 (h + \varepsilon). \tag{2.86}$$

Now let us obtain the similar estimates for the pair ρ_0 and ρ_0^I. To do this, we consider the third term in (2.72). On $T = T_{ij}^{(l)}$ we have

$$E_{ij}^{(l)} = a_T(\rho_0, \varphi_{ij}) - a_T^f(\rho_0, \varphi_{ij}) = -\frac{1}{h} \int_T (b_i - b(x)) \rho_0 \, d\Omega.$$

Since $|b_i - b(x)| \le h \|b'\|_\infty$ on $[x_i, x_{i+1}]$, we get

$$|E_{ij}^{(l)}| \le c_{22} \int_T |\rho_0| \, d\Omega. \tag{2.87}$$

Let us examine two variants of the behavior of the function $g(y) = -v_0(1, y)$. First assume that $g(y)$ changes its sign on the segment $[y_j, y_{j+1}]$. It involves $|g(y)| \le h \|g'\|_\infty$ on $[y_j, y_{j+1}]$. Therefore $|\rho_0| \le c_{23} h$ on T and we have

$$|E_{ij}^{(l)}| \le c_{24} h^3. \tag{2.88}$$

Next we assume that $g(y)$ does not change its sign on the segment $[y_j, y_{j+1}]$ and is, for example, nonnegative. Then by the mean value theorem we get

$$\int_T |\rho_0| \, d\Omega = \int_T \rho_0 \, d\Omega = \frac{h^2}{2} \rho_0(\tau^*, y^*) \tag{2.89}$$

where $(\tau^*, y^*) \in T$. Because of (2.4) we obtain

$$\rho_0(\tau^*, y^*) \le c_{25} \exp(-(1 - x_{i+1}) B_1 / \varepsilon). \tag{2.90}$$

Combining (2.87), (2.88)–(2.90), we get the estimate

$$|E_{ij}^{(l)}| \leq c_{26} h^2 (h + \exp(-(1 - x_{i+1}) B_1/\varepsilon)).$$

The same contribution comes from the triangles $T_{i-1,j}^{(l)}$, $T_{i,j-1}^{(u)}$, and $T_{i-1,j-1}^{(u)}$. On the triangles $T_{ij}^{(u)}$, $T_{i-1,j-1}^{(l)}$ we have $\partial_1 \varphi_{ij} = 0$ and consequently the contribution of these triangles equals zero. As a result, we have

$$|a(\rho_0, \varphi_{ij}) - a^f(\rho_0, \varphi_{ij})| \leq c_{27} h^2 (h + \exp(-(1 - x_{i+1}) B_1/\varepsilon)). \qquad (2.91)$$

Now we evaluate the difference $a_1^f(\rho_0, \varphi_{ij}) - a_1^h(\rho_0^I, \varphi_{ij})$. Let us take $T = T_{ij}^{(l)}$ and introduce the function

$$\hat{\rho}(x, y) = g^I(y) \exp(-(1 - x) b_i/\varepsilon)$$

where g^I is the piecewise linear interpolant of g. According to the construction of the quadrature rule, we have

$$a_1^f(\hat{\rho}, \varphi_{ij}) = a_1^h(\rho_0^I, \varphi_{ij}).$$

Thus, on $T_{ij}^{(l)}$ we obtain the representation

$$\begin{aligned} G_{ij}^{(l)} &= a_{1T}^f(\rho_0, \varphi_{ij}) - a_{1T}^h(\rho_0^I, \varphi_{ij}) = a_{1T}^f(\rho_0, \varphi_{ij}) - a_{1T}^f(\hat{\rho}, \varphi_{ij}) \\ &= -\frac{1}{h} \int_T (\varepsilon \partial_1 (\rho_0 - \hat{\rho}) - b_i(\rho_0 - \hat{\rho}))\, d\Omega. \end{aligned} \qquad (2.92)$$

Taking into account the behavior of the functions ρ_0 and $\hat{\rho}$ we conclude that the estimate of G_{ij} is worst near $x = 1$. Consider this case in detail. Assume that $x_i \geq 2/3$. Then we get

$$\begin{aligned} |\rho_0 - \hat{\rho}| &= |g(y) \exp(-(1 - x) b(x)/\varepsilon) - g^I(y) \exp(-(1 - x) b_i/\varepsilon)| \\ &\leq |g(y) \left(\exp(-(1 - x) b(x)/\varepsilon) - \exp(-(1 - x) b_i/\varepsilon)\right)| \\ &\quad + |g(y) - g^I(y)| \exp\left(-(1 - x) b_i/\varepsilon\right). \end{aligned} \qquad (2.93)$$

To estimate the first term we take into account the fact that the function $t \exp(-t)$ is bounded for $t \in [0, \infty)$ and use the mean value theorem. Then we obtain the upper bound

$$c_{28} h \exp\left(-(1 - x_{i+1}) B_1/2\varepsilon\right).$$

Since the interpolant g^I approximates g with the second order accuracy, we have the estimate of the second term in (2.93):

$$|\rho_0 - \hat{\rho}| \leq c_{29} h \left(h + \exp\left(-(1 - x_{i+1}) B_1/2\varepsilon\right)\right). \qquad (2.94)$$

Using similar reasoning for the first derivative, we get

$$\begin{aligned}|\partial_1\rho_0-\partial_1\hat{\rho}| &= \Big|g(y)\left(\frac{b(x)}{\varepsilon}-b'(x)\frac{1-x}{\varepsilon}\right)\exp\left(-(1-x)b(x)/\varepsilon\right)\\ &\quad - g^I(y)\frac{b_i}{\varepsilon}\exp\left(-(1-x)b_i/\varepsilon\right)\Big|\\ &\le \frac{b(x)}{\varepsilon}g(y)\big|\exp\left(-(1-x)b(x)/\varepsilon\right)-\exp\left(-(1-x)b_i/\varepsilon\right)\big|\\ &\quad + \left|\frac{1}{\varepsilon}\left(g(y)b(x)-g^I(y)b_i\right)\right|\exp\left(-(1-x)b_i/\varepsilon\right)\\ &\quad + \left|g(y)b'(x)\frac{1-x}{\varepsilon}\right|\exp\left(-(1-x)b(x)/\varepsilon\right)\\ &\le c_{30}\frac{h}{\varepsilon}\exp\left(-(1-x_{i+1})B_1/2\varepsilon\right).\end{aligned} \tag{2.95}$$

Combining (2.94) and (2.95), we have

$$|G_{ij}^{(l)}|\le c_{31}h^2\left(h+\exp(-(1-x_{i+1})B_1/2\varepsilon)\right)\quad \text{for}\quad x_i\ge 2/3. \tag{2.96}$$

Now assume that $x_i<2/3$. Then we have

$$\exp(-(1-x)b_1(x^*)/\varepsilon)\le c_{32}\varepsilon^2\le c_{33}h^2\quad\text{for all}\quad x^*\in[x_i,x_{i+1}].$$

This gives

$$\begin{aligned}|\rho_0-\hat{\rho}| &\le |g(y)|s(x)\exp(-(1-x)b(x)/\varepsilon)\\ &\quad+|g^I(y)|s(x)\exp(-(1-x)b_i/\varepsilon)\le c_{34}h^2,\\ |\partial_1\rho_0-\partial_1\hat{\rho}| &\le c_{35}\varepsilon^{-1}\big(|g(y)|s(x)\exp(-(1-x)b(x)/\varepsilon)\\ &\quad+|g^I(y)|s(x)\exp(-(1-x)b_i/\varepsilon)\big)\le c_{36}h.\end{aligned}$$

Therefore for $x_i<2/3$ the following estimate is valid:

$$|G_{ij}^{(l)}|\le c_{37}h^3.$$

Thus, (2.96) holds for all $x_i\in(0,1)$. The same estimates are valid for the triangles $T_{i-1,j}^{(l)}$, $T_{i,j-1}^{(u)}$, $T_{i-1,j-1}^{(u)}$. The contribution from the triangles $T_{ij}^{(u)}$ and $T_{i-1,j-1}^{(l)}$ is zero since the derivative $\partial_1\varphi_{ij}$ equals zero on these triangles.

Therefore we obtain

$$|a_1^f(\rho_0,\varphi_{ij})-a_1^h(\rho_0^I,\varphi_{ij})|\le c_{38}h^2(h+\exp(-(1-x_{i+1})B_1/2\varepsilon). \tag{2.97}$$

Finally, we need to obtain the estimate for the last term in the right-hand side of (2.72) for the functions η and η^I:

$$\left|a_1(\eta,\varphi_{ij}) - a_1^h(\eta^I,\varphi_{ij})\right|.$$

On $T = T_{ij}^{(l)}$ we consider

$$H_{ij}^{(l)} = a_{1T}(\eta,\varphi_{ij}) - a_{1T}^f(\eta,\varphi_{ij}) = -\frac{1}{h}\int_T (b_i - b(x))\eta \, d\Omega.$$

Because η is bounded on T according to Lemma 19 and b is sufficiently smooth, we get

$$\left|H_{ij}^{(l)}\right| \le c_{39}h^2.$$

The contribution from the other triangles with the vertex $z_{ij} = (x_i, y_j)$ has the same order. Therefore we have

$$\left|a_1(\eta,\varphi_{ij}) - a_1^f(\eta,\varphi_{ij})\right| \le 4c_{39}h^2.$$

Now we estimate the difference $\left|a_1^f(\eta,\varphi_{ij}) - a_1^f(\eta^I,\varphi_{ij})\right|$. On $T = T_{ij}^{(l)}$ we have

$$I_{ij}^{(l)} = -\frac{1}{h}\int_T \varepsilon(\partial_1\eta - \partial_1\eta^I) - b_i(\eta - \eta^I))\, d\Omega.$$

Due to (2.44) we get

$$|\partial_1\eta| \le c_{40}\left(1 + \varepsilon^{-1}\exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right) \qquad \text{on} \quad T. \tag{2.98}$$

Because of the Lagrange theorem $|\partial_1\eta^I(x,y)| = |\partial_1\eta(t,y_j)|$ on T. It involves the same estimate as (2.98) for both expressions. Under (2.70) we obtain

$$\varepsilon|\partial_1\eta - \partial_1\eta^I| \le c_{41}\varepsilon\left(1 + \varepsilon^{-1}\exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right). \tag{2.99}$$

Further, in order to estimate the difference $\eta - \eta^I$ on T we use the Taylor formula for η in the form

$$\eta(x,y) = \eta_{ij} + (y - y_j)\partial_2\eta(x_i,y_j) + h\pi_{10}(x,y_j). \tag{2.100}$$

Here because of (2.44) and (2.45) the function π_{10} satisfies the inequality

$$|\pi_{10}| \le c_{42}\left(1 + h\varepsilon^{-1} + \varepsilon^{-1}\exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right) \tag{2.101}$$

on T. Moreover, we have

$$\eta^I(x,y) = \eta_{ij} + (x - x_i)\frac{\eta_{i+1,j} - \eta_{ij}}{h} + (y - y_j)\frac{\eta_{i,j+1} - \eta_{ij}}{h} \quad \text{on} \quad T.$$

Using (2.100) for $\eta_{i+1,j}$ and $\eta_{i,j+1}$, we obtain

$$|\eta - \eta^I| \le c_{43} h \left(1 + \varepsilon^{-1} h + \varepsilon^{-1} \exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right). \tag{2.102}$$

Combining (2.99) and (2.102) we get on T

$$|I_{ij}^{(l)}| \le c_{44} \frac{h^2}{\varepsilon} \left(\varepsilon + h + \exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right). \tag{2.103}$$

The same estimate is valid for the triangles $T_{i-1,j}^{(l)}$, $T_{i,j-1}^{(u)}$, $T_{i-1,j-1}^{(u)}$. The contribution from the triangles $T_{ij}^{(u)}$ and $T_{i-1,j-1}^{(l)}$ equals zero. Therefore we have

$$|a_1^f(\eta, \varphi_{ij}) - a_1^f(\eta^I, \varphi_{ij})| \le c_{45} \frac{h^2}{\varepsilon} \left(\varepsilon + h + \exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right). \tag{2.104}$$

It remains to evaluate the difference

$$|a_1^f(\eta^I, \varphi_{ij}) - a_1^h(\eta^I, \varphi_{ij})|. \tag{2.105}$$

On $T = T_{ij}^{(l)}$ we have

$$\begin{aligned} J_{ij}^{(l)} &= -\frac{b_i}{h}\left(\int_T \eta^I \, d\Omega - \frac{h^2}{2}(\alpha_{1i}\eta_{ij} + \alpha_{2i}\eta_{i+1,j})\right) \\ &= -\frac{b_{1i} h}{6}(\eta_{ij} + \eta_{i+1,j} + \eta_{i+1,j+1} - 3\alpha_{1i}\eta_{ij} - 3\alpha_{2i}\eta_{i+1,j}). \end{aligned}$$

Use the Taylor formula in the form (2.100) for $\eta_{i+1,j}$ and $\eta_{i+1,j+1}$ and the equality $\alpha_{1i} + \alpha_{2i} = 1$. Since α_{1i}, α_{2i}, and $b(x)$ are bounded, we have

$$J_{ij}^{(l)} = \frac{b_i h^2}{6} \partial_2 \eta_{ij} + h^2 \pi_{11}(x, y)$$

with the estimate of $\pi_{11}(x, y)$ similar to (2.101). Considering (2.105) on the triangle $T_{i,j-1}^{(l)}$, we get

$$|J_{ij}^{(u)} + J_{ij-1}^{(u)}| \le c_{46} \frac{h^2}{\varepsilon} \left(\varepsilon + h + \exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right).$$

The triangles $T_{i-1,j-1}^{(u)}$ and $T_{i,j-1}^{(u)}$ make the same contribution. The contribution of the triangles $T_{ij}^{(u)}$ and $T_{i-1,j-1}^{(l)}$ equals zero.

Therefore for (2.105) we get

$$\begin{aligned} &\left| a_{1T}^f(\eta^I, \varphi_{ij}) - a_{1T}^h(\eta^I, \varphi_{ij}) \right| \\ &\qquad \le c_{47} \frac{h^2}{\varepsilon} \left(\varepsilon + h + \exp\left(-B_1(1 - x_{i+1})/2\varepsilon\right)\right). \end{aligned} \tag{2.106}$$

Combining (2.73), (2.76), (2.86), (2.97), and (2.106), due to (2.70) we obtain the estimate (2.71).□

The next result gives the barrier functions to estimate the right-hand side of (2.71).

Lemma 21. *Let us assume that*

$$\varepsilon \le c_1 h \tag{2.107}$$

with a constant c_1. *There exist the mesh functions* φ^h *and* ψ^h *on* $\bar{\Omega}_h$ *with the properties*

$$|\varphi^h| \le c_2 \quad \text{in} \quad \Omega_h, \tag{2.108}$$

$$|\psi^h| \le c_3 h \quad \text{in} \quad \Omega_h \tag{2.109}$$

such that

$$L^h \varphi^h \ge h^2 \quad \text{in} \quad \Omega_h, \tag{2.110}$$

$$\varphi^h \ge 0 \quad \text{on} \quad \Gamma_h \tag{2.111}$$

and

$$L^h \psi^h \ge h^2 \exp\left(-B_1(1-x_{i+1})/2\varepsilon\right) \quad \text{in} \quad \Omega_h, \tag{2.112}$$

$$\psi^h \ge 0 \quad \text{on} \quad \Gamma_h. \tag{2.113}$$

Proof. Consider the expression

$$\frac{1}{h}\left(\frac{b_i \exp(s_i)}{\exp(s_i)-1} - \frac{b_{i-1}\exp(s_{i-1})}{\exp(s_{i-1})-1}\right) \tag{2.114}$$

where $s_i = b_i h/\varepsilon$. It can be thought as the difference of the values of the function $f(s) = (s\exp(s))/(\exp(s)-1)$ at neighboring nodes of the grid. Then by the mean value theorem we have

$$\begin{gathered} \frac{\varepsilon}{h^2}\left|\frac{s_i \exp(s_i)}{\exp(s_i)-1} - \frac{s_{i-1}\exp(s_{i-1})}{\exp(s_{i-1})-1}\right| \le \frac{\varepsilon}{h^2} f'(s^*)|s_i - s_{i-1}| \\ \le f'(s^*) b'(x^*) \le c_5 \quad \text{where} \quad s^* \in \left[\frac{B_1 h}{\varepsilon}, \frac{B_2 h}{\varepsilon}\right]. \end{gathered} \tag{2.115}$$

We take into account the estimate

$$|s_i - s_{i-1}| = \frac{h}{\varepsilon}|b_i - b_{i-1}| \le \frac{h}{\varepsilon} b'(x^*) h, \quad x^* \in [x_{i-1}, x_i].$$

Thus, the difference (2.114) is bounded.

Now we consider the expression

$$\frac{1}{h}\left(\frac{b_i}{\exp(s_i)-1}-\frac{b_{i-1}}{\exp(s_{i-1})-1}\right). \tag{2.116}$$

In a similar way we introduce the function $f(s) = s/(\exp(s) - 1)$. Using the mean value theorem we get the estimate for (2.116)

$$\begin{aligned}&\frac{\varepsilon}{h^2}\left|\frac{s_i}{\exp(s_i)-1}-\frac{s_{i-1}}{\exp(s_{i-1})-1}\right| \leq \frac{\varepsilon}{h^2}f'(s^{**})|s_i - s_{i-1}| \\ &\leq f'(s^{**})b'(x^{**}) \leq c_6 \quad \text{where} \quad s^{**} \in \left[\frac{B_1 h}{\varepsilon}, \frac{B_2 h}{\varepsilon}\right], \quad x^{**} \in [x_{i-1}, x_i].\end{aligned} \tag{2.117}$$

Thus, the difference in (2.116) is also bounded.

Put $\sigma = 4(c_5 + c_6)/B_1$ and introduce the function φ^h by

$$\varphi^h_{0,j} = 0, \quad \frac{\varphi^h_{i,j} - \varphi^h_{i-1,j}}{h} = \sigma \exp(\sigma x_i), \quad i = 1, ..., n; \;\; \forall\, j = 1, ..., n-1.$$

We want to show that φ^h satisfies the conditions (2.110), (2.111). Rewrite $\dfrac{1}{h^2}L^h\varphi^h$ in the form

$$\begin{aligned}&-\frac{b_{i-1}\exp(s_{i-1})}{h\,(\exp(s_{i-1})-1)}\varphi^h_{i-1,j} + \left(\frac{b_{i-1}}{h\,(\exp(s_{i-1})-1)}\right. \\ &\qquad \left. + \frac{b_i \exp(s_i)}{h\,(\exp(s_i)-1)}\right)\varphi^h_{ij} - \frac{b_i}{h\,(\exp(s_i)-1)}\varphi^h_{i+1,j}.\end{aligned} \tag{2.118}$$

Rearranging the terms in (2.118), we have

$$\begin{aligned}&\frac{b_i \exp(s_i)}{h\,(\exp(s_i)-1)}\left(\varphi^h_{ij} - \varphi^h_{i-1,j}\right) - \frac{b_i}{h\,(\exp(s_i)-1)}\left(\varphi^h_{i+1,j} - \varphi^h_{ij}\right) \\ &+\frac{1}{h}\left(\frac{b_i\exp(s_i)}{\exp(s_i)-1} - \frac{b_{i-1}\exp(s_{i-1})}{\exp(s_{i-1})-1}\right)\varphi^h_{i-1,j} \\ &-\frac{1}{h}\left(\frac{b_i}{\exp(s_i)-1} - \frac{b_{i-1}}{\exp(s_{i-1})-1}\right)\varphi^h_{i0}.\end{aligned} \tag{2.119}$$

Take into consideration the inequalities $\exp(-b_i h/\varepsilon) \leq \exp(-B_1 h/\varepsilon) \leq 1/2$ for $c_1 = B_1/\ln(1/2)$ in (2.107) and $\exp(\sigma h) \geq 1$. Then the difference of two first terms in (2.119) has the lower estimate

$$\begin{aligned}
&\frac{b_i \exp(s_i)}{\exp(s_i)-1}\sigma \exp(\sigma x_i) - \frac{b_i}{\exp(s_i)-1}\sigma \exp(\sigma x_{i+1}) \\
&= \frac{b_i \exp(s_i)}{\exp(s_i)-1}\sigma \left(\exp(\sigma x_i) - \exp\left(\sigma x_i + h\sigma - b_i h/\varepsilon\right)\right) \\
&\geq \frac{1}{2}\sigma b_i \exp(\sigma x_i) \geq \frac{1}{2}\sigma B_1 \exp(\sigma x_i).
\end{aligned}$$

In view of the definition of σ the first term in (2.119) is twice as much as the remaining two terms, hence the conditions (2.110), (2.111) are valid.

Show that the condition (2.108) holds for the function φ^h. From the definition of φ^h we have

$$\frac{\varphi^h_{i,j} - \varphi^h_{0,j}}{h} = \sigma \sum_{k=0}^{i} \exp(\sigma x_k) = \sigma \exp(\sigma x_i) \sum_{k=0}^{i} \exp(-k\sigma h).$$

Then φ^h has the following representation

$$\varphi^h_{ij} = h\sigma \exp(\sigma x_i) \sum_{k=0}^{i} \exp(-k\sigma h).$$

The sum in this expression is the partial sum of a geometric progression, hence we get the estimate

$$\varphi^h_{ij} \leq \frac{h\sigma \exp(\sigma x_i)}{1-\exp(-\sigma h)}.$$

For sufficiently small h the following inequality holds:

$$\frac{\sigma h}{1-\exp(-\sigma h)} \leq 2.$$

It can be proved using the Taylor expansion of the function $f(t) = 1 - \exp(-t)$ at zero. Then we have $\varphi^h_{ij} \leq 2\exp(\sigma x_i)$ on Ω_h. The proof of the properties of φ^h_{ij} is complete.

Now consider the function ψ^h defined by

$$\frac{\psi^h_{ij} - \psi^h_{i-1,j}}{h} = \exp(-B_1(1-x_{i+1})/2\varepsilon) \quad \forall\, j = 0,1,...,n; \quad \psi^h_{i,0} = 0.$$

In order to prove the properties (2.112), (2.113) we consider

$$\begin{aligned}
\frac{1}{h^2}L^h\psi^h &= -\frac{b_{i-1}\exp(s_{i-1})}{h\left(\exp(s_{i-1})-1\right)}\psi^h_{i-1,j} \\
&+ \left(\frac{b_{i-1}}{h\left(\exp(s_{i-1})-1\right)} + \frac{b_i \exp(s_i)}{h\left(\exp(s_i)-1\right)}\right)\psi^h_{ij} - \frac{b_i}{h\left(\exp(s_i)-1\right)}\psi^h_{i+1,j}.
\end{aligned} \tag{2.120}$$

For convenience we rearrange terms:

$$\begin{aligned}
&\frac{b_i \exp(s_i)}{h\,(\exp(s_i)-1)}\left(\psi^h_{ij}-\psi^h_{i-1,j}\right)-\frac{b_i}{h\,(\exp(s_i)-1)}\left(\psi^h_{i+1,j}-\psi^h_{ij}\right)\\
&+\frac{1}{h}\left(\frac{b_i \exp(s_i)}{\exp(s_i)-1}-\frac{b_{i-1}\exp(s_{i-1})}{\exp(s_{i-1})-1}\right)\psi^h_{i-1,j}\\
&-\frac{1}{h}\left(\frac{b_i}{\exp(s_i)-1}-\frac{b_i}{\exp(s_i)-1}\right)\psi^h_i .
\end{aligned}\tag{2.121}$$

Take into consideration the inequality $\exp\left(-B_1 h/2\varepsilon\right) \le 1/2$ which holds for $\varepsilon \le h\,B_1/\ln 4$. Then we estimate the difference of the first two terms in (2.121):

$$\begin{aligned}
&\frac{b_i \exp(s_i)}{\exp(s_i)-1}\exp(-B_1(1-x_{i+1})/2\varepsilon)-\frac{b_i}{\exp(s_i)-1}\exp(-B_1(1-x_{i+2})/2\varepsilon)\\
&=\frac{b_i \exp(s_i)}{\exp(s_i)-1}\exp(-B_1(1-x_{i+1})/2\varepsilon)\left(1-\exp\left(\frac{B_1 h}{2\varepsilon}-\frac{b_i h}{\varepsilon}\right)\right)\\
&\ge \frac{1}{2} b_i \exp(-B_1(1-x_{i+1})/2\varepsilon) \ge c_{10}\exp(-B_1(1-x_{i+1})/2\varepsilon)
\end{aligned}$$

with the constant $c_{10} = B_1/2$.

The coefficients of $\psi^h_{i-1,j}$ and $\psi^h_{i+1,j}$ are bounded on Ω_h due to the estimates (2.115) and (2.117), respectively. Since the functions $\psi^h_{i-1,j}$ and $\psi^h_{i+1,j}$ themselves are of the first order with respect to h, we obtain the estimate (2.112).

Next we examine the condition (2.109) for ψ^h. In the same way as for the function φ^h, we have

$$\begin{aligned}
\psi^h_{ij} &= \psi^h_{0j} + h\exp(-B_1(1-x_{i+1})/2\varepsilon)\sum_{k=0}^{i}\exp(-kB_1 h)/2\varepsilon)\\
&\le h\exp(-B_1(1-x_{i+1})/2\varepsilon)\frac{1}{1-\exp(-B_1 h/2\varepsilon)}\\
&\le c_8 h\exp(-B_1(1-x_{i+1})/2\varepsilon) \le c_9 h \quad \text{on} \quad \Omega_h .
\end{aligned}$$

Thus, the function ψ^h satisfies (2.109). This completes the proof. □

Finally, using Lemmata 20 and 21, we formulate the main result.

Theorem 22. *Assume that* (2.4), (2.20) *hold. Then there exist constants* h_0 *and* c_1 *independent of* h *and* ε *such that* $\forall\ h \le h_0$ *and for* $\varepsilon \le h$ *the solution* u^h *of the problem* (2.65) *satisfies the estimate*

$$\max_{\Omega_h}|u-u^h| = \|u^I-u^h\|_{\infty,h} \le c_1 h \tag{2.122}$$

where u is the solution of the problem (2.18), (2.19).

Proof. Introduce the function

$$\phi^h = c_2 h \varphi^h + c_3 \psi^h$$

with φ^h and ψ^h from Lemma 21. From (2.110) – (2.113) we get

$$\begin{aligned} |L^h \phi^h| \ &\geq \ h^2 \left(h + \exp\left(-B_1(1-x_{i+1})/2\varepsilon\right)\right) \quad \text{in} \quad \Omega_h, \\ |\phi^h| &\geq 0 \quad \text{on} \quad \Gamma_h. \end{aligned}$$

Then by Lemma 21 the inequality

$$|\phi^h| \leq c_4 h \quad \text{on} \quad \bar{\Omega}_h$$

holds. In view of the definition (2.66) of the operator L^h and by Lemma 20 we have

$$\begin{aligned} |L^h(u^h - u^I)_{ij}| &\leq c_5 h^2 (\varepsilon + h + \exp(-(1-x_{i+1})B_1/2\varepsilon)) \\ &\leq L^h(\phi^h)_{ij} \quad \forall\, i,j = 1,...,n-1. \end{aligned}$$

Therefore

$$|u^h(x_i,y_j) - u(x_i,y_j)| \leq c_6 h$$

for $1 \leq i \leq n-1$ and $1 \leq j \leq n-1$. For $i = 0, n$ and $j = 0,...,n$ or for $j = 0, n$ and $i = 0,...,n$ the difference $u^h(x_i,y_j) - u(x_i,y_j)$ equals zero. Since $u^I(x_i,y_j) = u(x_i,y_j)$, we have (2.122). □

2.3 Construction of the method for the problem with regular and parabolic boundary layers

In this section we reject the restriction (2.20) and consider the convection-diffusion problem whose solution has regular and parabolic boundary layers.

2.3.1 Properties of the differential problem.

Consider the problem (2.1), (2.2) under the conditions which are stronger than (2.3) and (2.5) in order to simplify the proofs. Let

$$b \in C^{7+\alpha}[0,1], \quad f \in C^{6+\alpha}(\Omega), \qquad \alpha \in (0,1), \tag{2.123}$$

and

$$f = 0 \qquad \textit{in vicinities of four corners of } \ \Omega. \tag{2.124}$$

To describe the behaviour of the solution for small ε, we use the following expansion of the solution

$$u = u_0 + \rho_0 + \varepsilon\eta \quad \text{on} \quad \bar{\Omega}. \tag{2.125}$$

Here u_0 is the solution of the 'partially reduced' problem

$$L_{par}u_0 \equiv -\varepsilon\partial_{22}u_0 + \partial_1(bu_0) = f \quad \text{in} \quad \Omega, \tag{2.126}$$

$$u_0 = 0 \quad \text{on} \quad \Gamma \setminus \Gamma_{out}. \tag{2.127}$$

Five consistency conditions of orders $0 - 4$ are fulfilled for the data in this parabolic-type problem [34]. It guarantees that $\partial_1^{k_1}\partial_2^{k_2}u_0$ are Hölder-continuous for $2k_1 + k_2 \leq 8$. As for the derivatives $\partial_1^{k_1}\partial_2^{k_2}u$, then five conditions of orders $0 - 4$ are fulfilled for the data in initial elliptic-type problem [35] and these derivatives are Hölder-continuous for $k_1 + k_2 \leq 8$. The function ρ_0 is the regular boundary layer component

$$\rho_0(x, y) = g(y)s(x)\exp\left(-(1-x)b(x)/\varepsilon\right) \tag{2.128}$$

where

$$g(y) = -u_0(1, y) \qquad \text{on} \quad [0, 1]. \tag{2.129}$$

The cut-off function $s(x) \in C^4[0, 1]$ was introduced in (1.9). The function u_0 in (2.125), unlike the analogous component in (2.24), is not smooth in the y-direction. But it still is sufficiently smooth in the x-direction.

The operator L_{par} satisfies the comparison principle. We formulate it for the family of differential operators

$$\mathcal{L}_{par}v \equiv -\varepsilon\partial_{22}v + b\partial_1 v + dv \tag{2.130}$$

where $b(x)$ satisfies the conditions (2.3), (2.4) and $d(x)$ is a sufficiently smooth bounded function which will be specified further in each individual case.

Lemma 23. *Assume that $\varepsilon > 0$ and (2.4) holds. Assume also that $u, w \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfy the inequalities*

$$|\mathcal{L}_{par}u| \leq \mathcal{L}_{par}w \quad \textit{in} \quad \Omega, \qquad |u| \leq w \quad \textit{on} \quad \Gamma \setminus \Gamma_{out}. \tag{2.131}$$

Then we have

$$|u| \leq w \qquad \textit{on} \quad \bar{\Omega}. \tag{2.132}$$

The lemma can be proved in the same way as Lemma 13 for the operator $\mathcal{L}$.

The following lemmata give some estimates of the functions from (2.125) and of its derivatives, that are required to construct and to investigate the discrete problem.

Lemma 24. *Assume that* $\varepsilon > 0$ *is sufficiently small and* (2.123), (2.124), (2.4) *hold. Then the estimate*

$$\left|\frac{\partial^j u(x,y)}{\partial y^j}\right| \le c_1 \left(1 + \varepsilon^{-j/2} B(y)\right), \quad j = 1, 2, 3, \quad on \quad \bar{\Omega} \qquad (2.133)$$

is valid where $B(y) = \exp\left(-\gamma y/\sqrt{\varepsilon}\right) + \exp\left(-\gamma(1-y)/\sqrt{\varepsilon}\right)$ *with a constant* $\gamma > 0$.

Proof. Set $d(x) = b'(x)$ for the operator $\mathcal{L}$ introduced by (2.6). Let σ be defined by (2.10). Assume that ε satisfies the condition (2.12). For the barrier function

$$w(x,y) = c_2 \left(1 - \exp\left(-\gamma y/\sqrt{\varepsilon}\right)\right) \exp(\sigma x)$$

we put $\gamma = \sqrt{B_1/2}$ and $c_2 = 2\|f\|_\infty / B_1$. According to (2.13) we have

$$\begin{aligned} (\mathcal{L}w)(x,y) &= c_2 \gamma^2 \exp\left(-\gamma y/\sqrt{\varepsilon}\right) \exp(\sigma x) + \left(-\varepsilon\sigma^2 + \sigma b(x) + d(x)\right) w(x,y) \\ &\ge |(\mathcal{L}u)(x,y)| \quad \text{for} \quad (x,y) \in \Omega, \\ w(x,y) &\ge |u(x,y)| = 0 \quad \text{for} \quad (x,y) \in \Gamma. \end{aligned}$$

Thus, applying Lemma 13 we see that the solution u is bounded on $\bar{\Omega}$.

Besides, due to the equality $w(x,0) = 0$ on the boundary $y = 0$ the estimate

$$|\partial_2 u(x,0)| \le \partial_2 w(x,0) \le c_3 \varepsilon^{-1/2}. \qquad (2.134)$$

holds. To prove the same estimate on the boundary $y = 1$ we take

$$w(x,y) = c_4 \left(1 - \exp\left(-\gamma(1-y)/\sqrt{\varepsilon}\right)\right) \exp(\sigma x)$$

as the barrier function and use Lemma 13. In a similar way as (2.134), we get

$$|\partial_2 u(x,y)| \le \partial_2 w(x,y) \le c_5 \varepsilon^{-1/2} \quad \text{for} \quad (x,y) \in \Gamma_{tg}. \qquad (2.135)$$

In addition, because of (2.2) we have

$$|\partial_2 u(x,y)| = 0 \quad \text{for} \quad (x,y) \in \Gamma_{in} \cup \Gamma_{out}. \qquad (2.136)$$

To prove the estimate (2.133) for $j = 1$ we differentiate the equation (2.1) with respect to y. We introduce the notation $v_1 = \partial_2 u$. Then we get

$$\mathcal{L} v_1 = \partial_2 f \quad \text{in} \quad \Omega.$$

Choosing a sufficiently large positive constant c_6 we see that the barrier function

$$w(x,y) = c_6 \left(1 + \varepsilon^{-1/2} B(y)\right) \exp(\sigma x)$$

satisfies the relations

$$\begin{aligned}(\mathcal{L}w)\,(x,y) &= c_6\gamma^2\varepsilon^{-1/2}B(y)\exp(\sigma x) + \left(-\varepsilon\sigma^2 + \sigma b(x) + d(x)\right) w(x,y)\\ &\geq |(\mathcal{L}v_1)(x,y)| \quad \text{for} \quad (x,y) \in \Omega,\\ w(x,y) &\geq |v_1(x,y)| \quad \text{for} \quad (x,y) \in \Gamma.\end{aligned}$$

Thus, for $\partial_2 u$ Lemma 13 yields the estimate similar to (2.133) for $j = 1$ on $\bar{\Omega}$.

To prove the estimate (2.133) for $j = 2$ we twice differentiate the equation (2.1) with respect to y. Set $v_2 = \partial_{22}u$ and get $\mathcal{L}v_2 = \partial_{22}f$ in Ω.

Due to (2.2) $v_2 = 0$ on the boundary $\Gamma \setminus \Gamma_{tg}$. Since $\partial_{11}u = \partial_1 u = u = 0$ on Γ_{tg}, because (2.1) and (2.3) we have

$$|v_2| = |\partial_{22}u| = |-\partial_{11}u + \varepsilon^{-1}(b\partial_1 u + ub' - f)| \leq c_7\varepsilon^{-1} \text{ on } \Gamma_{tg}. \qquad (2.137)$$

As before, the barrier function

$$w(x,y) = c_8\left(1 + \varepsilon^{-1}B(y)\right)\exp(\sigma x)$$

satisfies the assumptions of Lemma 13 with constant c_8 large enough:

$$\begin{aligned}(\mathcal{L}w)(x,y) &\geq |(\mathcal{L}v_2)(x,y)| \quad \text{for} \quad (x,y) \in \Omega,\\ w(x,y) &\geq |v_2(x,y)| \quad \text{for} \quad (x,y) \in \Gamma.\end{aligned}$$

Consequently, for $j = 2$ (2.133) holds on $\bar{\Omega}$.

It remains to prove (2.133) for $j = 3$. We differentiate the equation (2.1) with respect to y three times and set $v_3 = \partial_{222}u$. As a result, we have

$$\mathcal{L}v_3 = \partial_{222}f \qquad \text{in} \quad \Omega.$$

From (2.2) we get $v_3 = 0$ on $\Gamma \setminus \Gamma_{tg}$. Now let us evaluate the normal derivative $\partial v_2/\partial n$ on Γ_{tg}. For this, act on equation (2.1) by operator $\partial_{22} - \partial_{11}$:

$$-\varepsilon\partial_{2222}u + \varepsilon\partial_{1111}u + b'\partial_{22}u + b\partial_{122}u - \partial_{111}(bu) = \partial_{22}f - \partial_{11}f \quad \text{in} \quad \Omega.$$

Let us take the limit of this expression in point $(x,y) \in \Gamma_{tg}$. Due to boundary condition (2.2) and the equality $\partial_{22}u = -f/\varepsilon$ on Γ_{tg} following (2.2), we get

$$-\varepsilon\partial_2 v_3 - \frac{1}{\varepsilon}\partial_1(bv_3) = \partial_{22}f - \partial_{11}f \quad \text{on} \quad \Gamma_{tg}.$$

Therefore

$$|\partial_2 v_3| \leq c_9\varepsilon^{-2} \quad \text{on} \quad \Gamma_{tg}.$$

The barrier function

$$w(x,y) = c_{10}\left(1+\varepsilon^{-3/2}B(y)\right)\exp(\sigma x)$$

with constant c_{10} large enough satisfies

$$\begin{aligned}(\mathcal{L}w)(x,y) &\geq |(\mathcal{L}v_3)(x,y)| \quad \text{for} \quad (x,y)\in\Omega,\\ w(x,y) &\geq |v_3(x,y)| \quad \text{for} \quad (x,y)\in\Gamma\setminus\Gamma_{tg},\\ \frac{\partial w(x,y)}{\partial n} &\geq \left|\frac{\partial v_3(x,y)}{\partial n}\right| \quad \text{for} \quad (x,y)\in\Gamma_{tg}.\end{aligned}$$

Applying Lemma 14 we complete the proof of the estimate (2.133) for $j=3$ on $\bar{\Omega}$. The lemma is proved. □

Lemma 25. *Assume that* $\varepsilon>0$ *is small enough and* (2.123), (2.124), (2.4) *hold. Then the estimates*

$$\left|\frac{\partial^k u_0}{\partial x^k}\right| \leq c_1, \quad k=0,1,2,3,4, \tag{2.138}$$

hold.

Proof. Assume that σ is given by (2.10) and ε satisfies (2.12). We derive (2.138) by the comparison principle for the operator $\mathcal{L}_{par}$ (Lemma 23).

Set $d=b'$, then the barrier function

$$w(x,y) = c_2 x\exp(\sigma x) \tag{2.139}$$

with the constant $c_2 = 2\|f\|_\infty$ satisfies the relations

$$\begin{aligned}(\mathcal{L}_{par}w)(x,y) &= c_2 b(x)\exp(\sigma x) + (d(x)+\sigma b(x))\,w(x)\\ &\geq (\mathcal{L}_{par}u_0)(x,y) \quad \text{for} \quad (x,y)\in\Omega,\\ w(x,y) &\geq |u_0(x,y)| \quad \text{for} \quad (x,y)\in\Gamma\setminus\Gamma_{out}.\end{aligned}$$

Using Lemma 23, we see that u_0 is bounded on $\bar{\Omega}$.

From (2.127) it follows that $\partial_1 u_0 = 0$ on Γ_{tg} and $\partial_{22}u_0 = 0$ on Γ_{in}. Due to this together with (2.3), (2.4), and (2.126) the derivative $\partial_1 u_0$ is bounded on $\Gamma\setminus\Gamma_{out}$:

$$\partial_1 u_0(x,y) \leq c_3 \quad \text{for} \quad (x,y)\in\Gamma\setminus\Gamma_{out}. \tag{2.140}$$

Now we derive (2.138) for $k=1$. Differentiate the equation (2.126) with respect to x and introduce the notation $v_1 = \partial_1 u_0$. Setting $d=2b'$ we have

$$(\mathcal{L}_{par}v_1)(x,y) = \partial_1 f - u_0 b'' \quad \text{in} \quad \Omega.$$

In view of (2.140) for v_1, the barrier function w from (2.139) with the constant $c_2 = 2(\|\partial_1 u_0\|_\infty + \|u_0 b''\|_\infty)$ satisfies the assumptions of Lemma 23. Consequently, $\partial_1 u_0$ is bounded on $\bar{\Omega}$.

To prove (2.138) for $k = 2$ we differentiate the equation (2.126) with respect to x twice and introduce the notation $v_2 = \partial_{11} u_0$. Setting $d = 3b'$ we get

$$(\mathcal{L}_{par} v_2)(x, y) = \partial_{11} f - 3b'' \partial_1 u_0 - u_0 b''' \quad \text{in} \quad \Omega.$$

From (2.127) we get

$$v_2 = 0 \qquad \text{on} \quad \bar{\Gamma}_{tg}.$$

Now let us evaluate v_2 on Γ_{in}. For this, act on (2.126) by operator $b\partial_1 + \varepsilon\partial_{22}$:

$$-\varepsilon^2 \partial_{2222} u_0 + b^2 \partial_{11} u_0 + \varepsilon b' \partial_{22} u_0 + 2b' b \partial_1 u_0 + b'' b u_0 = b \partial_1 f + \varepsilon \partial_{22} f \qquad \text{in} \quad \Omega.$$

Let us take the limit of this expression in point $(0, y) \in \Gamma_{in}$, $0 < y < 1$. Due to boundary condition (2.127) and the equality $\partial_1 u_0 = f/b$ on Γ_{in} following (2.126), we get

$$b^2 v_2 + 2b' f = b \partial_1 f + \varepsilon \partial_{22} f \qquad \text{on} \quad \Gamma_{in}.$$

Therefore due to (2.3), (2.4) we have

$$|v_2| \leq c_4 \qquad \text{on} \quad \Gamma_{in}.$$

Then the barrier function w with an appropriate constant c_2 satisfies the assumptions of Lemma 23. Hence $\partial_{11} u_0$ is bounded on $\bar{\Omega}$.

Now let us prove (2.138) for $k = 3$. Differentiate (2.126) 3 times with respect to x and denote $v_3 = \partial_{111} u_0$. Taking $d = 4b'$ in (2.130), we get

$$\mathcal{L}_{par} v_3 = \partial_{111} f - 6b'' \partial_{11} u_0 - 4b''' \partial_1 u_0 - b^{(IV)} u_0 \qquad \text{in} \quad \Omega. \qquad (2.141)$$

Boundary condition (2.127) implies

$$v_3 = 0 \qquad \text{on} \quad \bar{\Gamma}_{tg}. \qquad (2.142)$$

In order to evaluate v_3 on Γ_{in} act on (2.126) by operator

$$L_2 \equiv b^2 \partial_{11} - 2\varepsilon b' \partial_{22} + \varepsilon b \partial_{122} + \varepsilon^2 \partial_{2222}.$$

Coefficients of this operator are chosen to eliminate mixed partial derivatives $\partial_{1122} u_0$, $\partial_{122} u_0$, $\partial_{12222} u_0$. As a result, we get the equality

$$\begin{aligned} -\varepsilon^3 \partial_{222222} u_0 + 3\varepsilon^2 b' \partial_{2222} u_0 + b^3 \partial_{111} u_0 + 3b^2 b' \partial_{11} u_0 + \varepsilon b^2 b'' \partial_{22} u_0 \\ -2\varepsilon (b')^2 \partial_{22} u_0 = L_2 f - 3b^2 b'' \partial_1 u_0 - b^2 b''' u_0 \qquad \text{in} \quad \Omega. \end{aligned}$$

Let us take limit of this expression on Γ_{in}. Due to boundary conditions (2.127) on Γ_{in} we have

$$b^3\partial_{111}u_0 = L_2 f - 3b^2b'\partial_{11}u_0 - 3b^2b''\partial_1 u_0 \qquad \text{on} \quad \Gamma_{in}.$$

Because of smoothness of f, b and estimates (2.4) and (2.138) proved yet for $k = 1, 2$, we get

$$|v_3| = |\partial_{111}u_0| \le c_5 \qquad \text{on} \quad \Gamma_{in}. \tag{2.143}$$

Then taking barrier function w in (2.139) with corresponding constant c_2, we satisfy conditions of Lemma 23. It implies (2.138) for $k = 3$.

Finally, let us prove (2.138) for $k = 4$. Differentiate (2.126) 4 times with respect to x and denote $v_4 = \partial_{1111}u_0$. Taking $d = 5b'$ in (2.130), we get

$$\begin{aligned} \mathcal{L}_{par}v_4 = \partial_{1111}f - 10b''\partial_{111}u_0 - 10b'''\partial_{11}u_0 \\ - 5b^{(IV)}\partial_1 u_0 - b^{(V)}u_0 \qquad \text{in} \quad \Omega. \end{aligned} \tag{2.144}$$

Boundary condition (2.127) implies

$$v_4 = 0 \qquad \text{on} \quad \bar{\Gamma}_{tg}. \tag{2.145}$$

In order to evaluate v_4 on Γ_{in}, act on (2.126) by operator

$$\begin{aligned} L_3 \equiv b^3\partial_{111} + \varepsilon b^2\partial_{1122} + \varepsilon^2 b\partial_{12222} + \varepsilon^3 b\partial_{222222} \\ +3\varepsilon\left(2(b')^2 - bb''\right)\partial_{22} - 3\varepsilon bb'\partial_{122} + 5\varepsilon^2 b'\partial_{2222}. \end{aligned}$$

Coefficients of this operator are chosen to eliminate mixed partial derivatives $\partial_{1222222}u_0$, $\partial_{112222}u_0$, $\partial_{12222}u_0$, $\partial_{11122}u_0$, $\partial_{122}u_0$, $\partial_{1122}u_0$. As a result, we get the equality

$$b^4\partial_{1111}u_0 + L_4 u_0 = L_3 f - 4b^3b'\partial_{111}u_0 - 6b^3b''\partial_{11}u_0 - 4b^3b'''\partial_1 u_0 \qquad \text{in} \quad \Omega$$

where operator L_4 includes u_0 and partial derivatives of u_0 with respect to y. All terms of this equality are continuous in $\bar{\Omega}$ except two corners $(0,0)$ end $(0,1)$. Let us take limit of this expression on Γ_{in}. Due to boundary conditions (2.127) on Γ_{in} we have

$$b^4\partial_{1111}u_0 = L_3 f - 4b^3b'\partial_{111}u_0 - 3b^3b''\partial_{11}u_0 - 4b^3b''\partial_1 u_0 \qquad \text{in} \quad \Omega.$$

Because of smoothness of f, b and estimates (2.4) and (2.138) proved yet for $k = 1, 2, 3$, we get

$$|v_4| = |\partial_{1111}u_0| \le c_6 \qquad \text{on} \quad \Gamma_{in}. \tag{2.146}$$

Then taking barrier function w in (2.139) with corresponding constant c_2, we satisfy conditions of Lemma 23. It implies (2.138) for $k = 4$. The proof is complete. □

Lemma 26. *Let* $\varepsilon > 0$ *be sufficiently small and* (2.123), (2.124), (2.4) *be valid for the problem* (2.1)–(2.2). *Then the remainder term in* (2.125) *satisfies*

$$\|\eta\|_\infty \leq c_1, \tag{2.147}$$
$$|\partial_1 \eta(x,y)| \leq c_2(1+\varepsilon^{-1}\exp(-B_1(1-x)/2\varepsilon)) \quad in \quad \Omega, \tag{2.148}$$
$$|\partial_{11} \eta(x,y)| \leq c_3\varepsilon^{-1}(1+\varepsilon^{-1}\exp(-B_1(1-x)/2\varepsilon)) \quad in \quad \Omega. \tag{2.149}$$

Proof. First we set $d = b'(x)$ in (2.6) so that $\mathcal{L}u = f$. In view of (2.126), (2.129), and the estimate (2.138) for $j = 2$, by an elementary calculation we show that η in (2.125) satisfies

$$\begin{aligned}(\mathcal{L}\eta)(x,y) = a_0(x,y) &+ \frac{1}{\varepsilon}a_1(x,y)A(x,\varepsilon)\\ &+\frac{1-x}{\varepsilon^2}a_2(x,y)A(x,\varepsilon) \text{ on } \bar{\Omega}\end{aligned} \tag{2.150}$$

where $A(x,\varepsilon) = \exp(-(1-x)b(x)/\varepsilon)$ and a_0, a_1, a_2 are bounded functions on $\bar{\Omega}$. In a similar way as in Lemma 19, the right-hand side of (2.150) is evaluated by

$$|\mathcal{L}\eta| \leq c_3 + \left(c_4\frac{1}{\varepsilon} + c_5\frac{1-x}{\varepsilon^2}\right)A(x,\varepsilon). \tag{2.151}$$

Let us use the barrier function

$$w(x,y) = c_{13}\psi_1(x,y) + c_{14}\psi_2(x,y)$$

where the functions ψ_1, ψ_2 are taken from Lemmata 16, 17 with the constants from Lemma 19. This gives

$$|(\mathcal{L}\eta)(x,y)| \leq (\mathcal{L}w)(x,y) \quad \text{in } \Omega.$$

Moreover, we have

$$w \geq 0 = |\eta| \quad \text{on} \quad \Gamma.$$

Thus, all the assumptions of Lemma 16 are satisfied. By this lemma, η is bounded on $\bar{\Omega}$.

Besides, since

$$w(0,y) = w(1,y) = 0 \quad \forall\, y \in [0,1],$$

we have

$$|\partial_1\eta(0,y)| \leq c_6 \quad \text{and} \quad |\partial_1\eta(1,y)| \leq c_7\varepsilon^{-1}. \tag{2.152}$$

In order to show (2.148), we first differentiate (2.150) with respect to x. Denoting $v = \partial_1 \eta$ and setting $d = 2b'$ in (2.6), from (2.138) for $j = 3$ we obtain the representation

$$\begin{aligned}(\mathcal{L}v)(x,y) &= a_3(x,y) \\ &+ \frac{1}{\varepsilon^2} a_4(x,y) A(x,\varepsilon) + \frac{1-x}{\varepsilon^3} a_5(x,y) A(x,\varepsilon)\end{aligned} \tag{2.153}$$

with functions a_3, a_4, and a_5, bounded on $\bar{\Omega}$. The right-hand side of (2.153) can be estimated in the following way:

$$|(\mathcal{L}v)(x,y)| \le c_8 + c_9 \varepsilon^{-2} \exp(-B_1(1-x)/(2\varepsilon)).$$

Now choosing the barrier function $w = c_{10}\psi_3$ from Lemma 18 with an appropriate constant c_{10}, we get

$$|(\mathcal{L}v)(x,y)| \le (\mathcal{L}w)(x,y) \quad \text{in } \Omega,$$

$$w \ge 0 = |v| \quad \text{on} \quad \bar{\Gamma}_{tg}, \quad w \ge |v| \quad \text{on} \quad \Gamma_{in} \cup \Gamma_{out}.$$

The last inequality follows from (2.152). Thus, due to Lemma 13 we obtain

$$|v| \le w \quad \text{on} \quad \bar{\Omega}$$

that implies (2.148).

In order to prove (2.149), differentiate (2.150) twice with respect to x and denote $v_2 = \partial_{11}\eta$. Taking $d = 3b'$ in (2.6) we get

$$\varepsilon \mathcal{L} v_2 = \partial_{11} f - 3\varepsilon b'' \partial_1 \eta - \varepsilon b''' \eta - L_5 u_0 - L_6 \rho_0 \tag{2.154}$$

where

$$L_5 u_0 = -\varepsilon \partial_{1111} u_0 - \varepsilon \partial_{1122} u_0 + \partial_{111}(b u_0), \tag{2.155}$$

$$L_6 \rho_0 = -\varepsilon \partial_{1111} \rho_0 - \varepsilon \partial_{1122} \rho_0 + \partial_{111}(b \rho_0), \tag{2.156}$$

Due to smoothness of b, f and estimates (2.147), (2.148) three first terms in the right-hand side or (2.154) are bounded by constant (independent of ε).

Next examine $L_5 u_0$. Due to the equation (2.126) we get

$$L_5 u_0 = -\varepsilon \partial_{1111} u_0 + \partial_{11} \left(\mathcal{L}_{par} u_0\right) = -\varepsilon \partial_{1111} u_0 + \partial_{11} f.$$

Take into consideration smoothness of f and the estimate (2.138) for $k = 4$. Then last expression is bounded by constant (independent of ε).

Finally, examine $L_6\rho_0$. Introduce one more operator

$$\mathcal{L}_1\rho_0 = -\varepsilon\partial_{11}\rho_0 + b\partial_1\rho_0. \tag{2.157}$$

Taking this operator into consideration, we rewrite $L_6\rho_0$ in the following form:

$$L_6\rho_0 = \partial_{11}\mathcal{L}_1\rho_0 + b'\partial_{11}\rho_0 + 2b''\partial_1\rho_0 + b'''\rho_0 - \varepsilon\partial_{1122}\rho_0. \tag{2.158}$$

Direct computation of $\partial_{11}\mathcal{L}_1\rho_0$ gives equalities

$$\begin{aligned}\partial_{11}\mathcal{L}_1\rho_0 &= -\varepsilon\partial_{1111}\rho_0 + b\partial_{111}\rho_0 + 2b'\partial_{11}\rho_0 + b''\partial_1\rho_0 \\ &= A(x,\varepsilon)\sum_{i=0}^{4}\sum_{j=0}^{\min\{i+2,4\}} p_{ij}(x,y)\frac{(1-x)^i}{\varepsilon^j}\end{aligned} \tag{2.159}$$

where functions p_{ij} is bounded in modulus due to smoothness of b. Because of boundedness of $t^k\exp(-t)$ on $(0,\infty)$, we have

$$\frac{(1-x)^i}{\varepsilon^j}A^{1/2}(x,\varepsilon) \leq c_{11}. \tag{2.160}$$

It implies the following estimate for the right-hand side of (2.159):

$$|\partial_{11}\mathcal{L}_1\rho_0| \leq c_{12}\left(1+\varepsilon^{-2}\exp\left(-B_1(1-x)/(2\varepsilon)\right)\right). \tag{2.161}$$

Next let us evaluate $|\partial_{11}\rho_0|$. From its definition we have

$$\partial_{11}\rho_0 = \left(q_1 + \frac{1}{\varepsilon}q_2 + \frac{1}{\varepsilon^2}q_3 + \frac{1-x}{\varepsilon}q_4 + \frac{1-x}{\varepsilon^2}q_5 + \frac{(1-x)^2}{\varepsilon^2}q_6\right)A(x,\varepsilon)$$

with functions q_i bounded because of smoothness of b. Due to (2.160) we have

$$|\partial_{11}\rho_0| \leq c_{13}\left(1+\varepsilon^{-2}\exp\left(-B_1(1-x)/(2\varepsilon)\right)\right) \quad \text{in} \quad \Omega. \tag{2.162}$$

Terms $b''\partial_1\rho_0$ and $b'''\rho_0$ are evaluated by same way that gives

$$\begin{aligned}&|b'\partial_{11}\rho_0 + 2b''\partial_1\rho_0 + b'''\rho_0| \\ &\quad\leq c_{14}\left(1+\varepsilon^{-2}\exp\left(-B_1(1-x)/(2\varepsilon)\right)\right) \quad \text{in} \quad \Omega.\end{aligned} \tag{2.163}$$

Finally, we examine term $\partial_{1122}\rho_0$. Due to definition of ρ_0

$$\partial_{1122}\rho_0 = \partial_{1122}\left(g(y)s(x)A(x,\varepsilon)\right) = g''(y)\left(s(x)A(x,\varepsilon)\right)''. \tag{2.164}$$

Since $g(y) = -u_0(1, y)$, we need to estimate $\partial_{22}u_0$ on Γ_{out}. Take the equation (2.126) in the form

$$\partial_{22}u_0 = -\varepsilon^{-1}\left(f - \partial_1(bu_0)\right) \qquad \text{in} \qquad \Omega.$$

Then come to the limit in a point of Γ_{out} and take into consideration (2.138) with $k = 0, 1$. As a result, we get

$$|\partial_{22}u_0| \le c_{15}\varepsilon^{-1} \qquad \text{on} \qquad \Gamma_{out}.$$

Thus, combine this inequality with (2.164) and (2.162), we have

$$\begin{aligned} \varepsilon|\partial_{1122}\rho_0| &\le c_{16}|\partial_{11}\rho_0| \\ &\le c_{17}\left(1 + \varepsilon^{-2}\exp\left(-B_1(1-x)/(2\varepsilon)\right)\right) \quad \text{in} \quad \Omega. \end{aligned} \tag{2.165}$$

So, combine estimate (2.161) – (2.163), (2.165) to obtain the estimate

$$\varepsilon|\mathcal{L}v_2| \le c_{18} + c_{19}\varepsilon^{-2}\exp\left(-B_1(1-x)/(2\varepsilon)\right) \quad \text{in} \quad \Omega. \tag{2.166}$$

Boundary condition (2.127) implies

$$|\partial_{11}\eta| = 0 \qquad \text{on} \qquad \bar{\Gamma}_{tg}. \tag{2.167}$$

Now let us evaluate the boundary value of $\partial_{11}\eta$ on Γ_{in}. Act on (2.125) by operator L:

$$Lu = Lu_0 + L\rho_0 + \varepsilon L\eta \quad \text{in} \quad \Omega. \tag{2.168}$$

Then use equation (2.1), (2.126) and come to the limit in a point of Γ_{in}:

$$f = f - \varepsilon\partial_{11}u + L\rho_0 + \varepsilon L\eta \qquad \text{on} \qquad \Gamma_{in}.$$

Since cut-off function equals 0 in the vicinity of Γ_{in}, then $L\rho_0 = 0$ on Γ_{in}. Function η equals 0 on Γ_{in}. Therefore

$$\mathcal{L}_1\eta = \partial_{11}u,$$

from which

$$\varepsilon|\partial_{11}\eta| \le B_1|\partial_1\eta| + |\partial_{11}u| \qquad \text{on} \qquad \Gamma_{in}.$$

Applying (2.148) and (2.138), we get the estimate

$$\varepsilon|\partial_{11}\eta| \le c_{20} \qquad \text{on} \qquad \Gamma_{in}. \tag{2.169}$$

Next, consider the boundary value of $\partial_{11}\eta$ on Γ_{out}. For this, act on (2.125) by operator $\mathcal{L}_1$:

$$\mathcal{L}_1 u = \mathcal{L}_1 u_0 + \mathcal{L}_1 \rho_0 + \varepsilon \mathcal{L}_1 \eta \quad \text{in} \quad \Omega.$$

Then come to the limit in a point of Γ_{out} and use the boundary condition $u = 0$ on Γ_{out}:

$$f = Lu = \mathcal{L}_1 u_0 + \mathcal{L}_1 \rho_0 + \varepsilon \mathcal{L}_1 \eta \quad \text{on} \quad \Gamma_{out}. \tag{2.170}$$

Direct computation gives the equality

$$\mathcal{L}_1 \rho_0 = -b' \quad \text{on} \quad \Gamma_{out}.$$

Applying it with (2.148) and (2.138) for $k = 1, 2$ to (2.170), we get

$$\varepsilon |\partial_{11}\eta| \le c_{21}\varepsilon^{-1} \qquad \text{on} \qquad \Gamma_{out}. \tag{2.171}$$

Thus, choosing the barrier function $w = c_{22}\psi_3$ from Lemma 18 with an appropriate constant c_{22}, we get

$$\varepsilon |(\mathcal{L}v_2)(x, y)| \le (\mathcal{L}w)(x, y) \quad \text{in } \Omega,$$

$$\varepsilon |v_2| = 0 \le w \quad \text{on} \quad \bar{\Gamma}_{tg}, \qquad \varepsilon |v_2| \le w \quad \text{on} \quad \Gamma_{in} \cup \Gamma_{out}$$

because of (2.166), (2.167), (2.169) and (2.171) respectively. Due to Lemma 13 we obtain

$$\varepsilon |v_2| \le w \quad \text{on} \quad \bar{\Omega}$$

that implies (2.149). □

2.3.2 Construction of the fitted quadrature rule.

For the approximation of the regular boundary layer we use the technique considered in the previous section. For the approximation of the parabolic layer we construct the special grid based on the extension method (see [36], [37], [5]).

First we put $h = 1/n$ with even integer $n \ge 2$ and take the uniform grid in the x-direction:

$$x_i := ih, \quad i = 0, 1, ..., n.$$

Next, in the y-direction we algorithmically introduce the graded grid in the y-direction:

$$y_j := \begin{cases} 0, \quad \text{for} \quad j = 0, \\ y_{j-1} + \dfrac{c_0 h}{1 + \varepsilon^{-1/2}\exp(-\gamma y_{j-1}/\sqrt{\varepsilon})}, \quad \text{for} \quad j = 1, ..., n/2, \\ 1 - y_{n-j}, \quad \text{for} \quad j = n/2 + 1, ..., n. \end{cases} \tag{2.172}$$

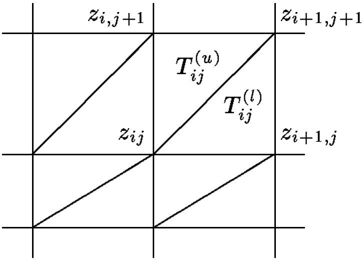


Fig. 4. Fragment the trangulation $\mathcal{T}_h$.

The constant c_0 satisfies the condition $y_{n/2} = 1/2$. Unfortunately, this leads to a nonlinear equation in c_0.

We define the mesh size in the y-direction by

$$h_j = y_j - y_{j-1}, \quad j = 1, \dots, n. \tag{2.173}$$

We denote the set of nodes by

$$\bar{\Omega}_h = \{z_{ij} = (x_i, y_j), \quad i, j = 0, 1, \dots, n\},$$

the set of interior nodes by

$$\Omega_h = \{z_{ij} = (x_i, y_j), \quad i, j = 1, 2, \dots, n-1\},$$

and the set of boundary nodes by

$$\Gamma_h = \{z_{ij} = (x_i, y_j),\ i = 0, 1 \text{ and } j = 0, 1, \dots, n; i = 0, 1, \dots, n, \text{ and } j = 0, 1\}.$$

Then the triangulation $\mathcal{T}_h$ is constructed by dividing each *elementary* rectangle $\Omega_{ij} = [x_i, x_{i+1}] \times [y_j, y_{j+1}]$ into two *elementary* triangles by the diagonal passing from (x_i, y_j) to (x_{i+1}, y_{j+1}) (see Fig. 4).

For each interior node $z_{ij} \in \Omega_h$ we introduce the basis function φ_{ij} which equals 1 at the node z_{ij}, equals 0 at any other node of $\bar{\Omega}_h$, and is linear on each elementary triangle of $\mathcal{T}_h$. Denote the linear span of these functions by

$$H^h = \text{span}\{\varphi_{ij}\}_{i,j=1}^{n-1}.$$

With this notations, we formulate the Galerkin problem: *find* $u^h \in H^h$ *such that*

$$a(u^h, v) = (f, v) \qquad \forall\, v \in H^h. \tag{2.174}$$

As before, in order to ensure the stability and to improve the accuracy of the method, we construct the special quadrature rule that provides good approximation on the smooth and boundary layer components of the solution. Since this technique was described in detail in Section 2.2.2, now we sketch the broad outlines of the construction of the quadrature rule on the nonuniform grid.

Let $T_{ij}^{(l)}$ (or $T_{ij}^{(u)}$, respectively) be an arbitrary elementary triangle of $\mathcal{T}_h$ with the vertices $z_{i,j}$, $z_{i+1,j+1} = (x_{i+1}, y_{j+1})$, and $z_{i+1,j} = (x_{i+1}, y_j)$ (or $z_{i+1,j} = (x_i, y_{j+1})$, respectively) as in Fig. 4. We consider the bilinear form

$$a_T(u,v) = \int_T \left(\left(\varepsilon \frac{\partial u}{\partial x} - bu \right) \frac{\partial v}{\partial x} + \varepsilon \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} \right) d\Omega. \tag{2.175}$$

Its approximation by piecewise linear functions $w^h, v^h \in H^h$, for example, on the triangle $T_{ij}^{(l)}$ has the form

$$\begin{aligned} a^h_{T^{(l)}_{ij}}(w^h, v^h) = \frac{hh_{j+1}}{2} \Big(& \left(\varepsilon b_i (\alpha_{1i} w^h(z_{ij}) + \alpha_{2i} w^h(z_{i+1,j}))\right) \frac{\partial v^h}{\partial x} \\ & + \varepsilon \frac{\partial w^h}{\partial y} \frac{\partial v^h}{\partial y} \Big). \end{aligned} \tag{2.176}$$

As before, the weights α_{1i} and α_{2i} are chosen in such a way as to satisfy two requirements, namely, to guarantee the first order accuracy for a smooth solution, and to reduce the error of approximation of the difference $a_{T_{ij}}(\rho_0, v^h) - a^h_{T_{ij}}(\rho_0^I, v^h)$ for the regular boundary layer component ρ_0 and its piecewise linear interpolant $\rho_0^I \in H^h$ on each element $T_{ij} \in \mathcal{T}_h$. The first requirement involves the equation

$$\alpha_{1i} + \alpha_{2i} = 1. \tag{2.177}$$

To satisfy the second one, we demand that the equality

$$a_{T_{ij}}(\zeta_i, v^h) = a^h_{T_{ij}}(\zeta_i^I, v^h) \tag{2.178}$$

be valid for the function $\zeta_i(x) = \exp\left(-(1-x)b_i/\varepsilon\right)$ and its piecewise linear interpolant $\zeta_i^I(x,y)$ on T_{ij}. Solving the system (2.177), (2.178), we get the unique solution

$$\alpha_{1i} = \frac{\exp \sigma_i}{(\exp \sigma_i - 1)} - \frac{1}{\sigma_i}, \quad \alpha_{2i} = \frac{1}{\sigma_i} - \frac{1}{\exp \sigma_i - 1} \tag{2.179}$$

where $\sigma_i = b_i h/\varepsilon$. With this weights we obtain the following approximation of the elementary bilinear form (2.176):

$$\begin{aligned} a^h_{T^{(l)}_{ij}}(w^h, v^h) = \frac{1}{2} h h_{j+1} \Big(& \frac{b_i}{\exp\sigma_i - 1} (w^h_{i+1,j} - w^h_{ij} \exp\sigma_i) \frac{\partial v^h}{\partial x} \\ & + \varepsilon \frac{\partial w^h}{\partial y} \frac{\partial v^h}{\partial y} \Big). \end{aligned} \tag{2.180}$$

In a similar way, on the triangle $T^{(u)}_{ij}$ we have

$$\begin{aligned} a^h_{T^{(u)}_{ij}}(w^h, v^h) = \frac{1}{2} h h_{j+1} \Big(& \frac{b_i}{\exp\sigma_i - 1} (w^h_{i+1,j+1} - w^h_{i,j+1} \exp\sigma_i) \frac{\partial v^h}{\partial x} \\ & + \varepsilon \frac{\partial w^h}{\partial y} \frac{\partial v^h}{\partial y} \Big). \end{aligned} \tag{2.181}$$

To integrate the right-hand side, we use the simple piecewise constant approximation. This gives the following approximation of the linear form

$$\begin{aligned} f^h_{T^{(l)}_{ij}}(v^h) &= \frac{1}{6} h h_{j+1} (f_{ij} v_{ij} + f_{i+1,j} v_{i+1,j} + f_{i+1,j+1} v_{i+1,j+1}), \\ f^h_{T^{(u)}_{ij}}(v^h) &= \frac{1}{6} h h_{j+1} (f_{ij} v_{ij} + f_{i,j+1} v_{i,j+1} + f_{i+1,j+1} v_{i+1,j+1}). \end{aligned} \tag{2.182}$$

Summing (2.180), (2.181), and (2.182) over all the triangles $T \in \mathcal{T}_h$, we obtain the approximations of the bilinear and linear forms

$$a^h(w^h, v^h) = \sum_{T \in \mathcal{T}_h} a^h_T(w^h, v^h), \tag{2.183}$$

$$f^h(v^h) = \sum_{T \in \mathcal{T}_h} f^h_T(v^h). \tag{2.184}$$

Now we come to the fitted Galerkin problem: *find $u^h \in H^h$ such that*

$$a^h(u^h, v^h) = f^h(v^h) \qquad \forall\, v^h \in H^h. \tag{2.185}$$

This problem is equivalent to the system of linear algebraic equations

$$\begin{aligned}
&(L^h u^h)_{ij} \equiv \\
&u^h_{ij}\left(\frac{h_j+h_{j+1}}{2}\left(\frac{b_i \exp\sigma_i}{\exp\sigma_i - 1} + \frac{b_{i-1}}{\exp\sigma_{i-1}-1}\right) + \varepsilon h\left(\frac{1}{h_j}+\frac{1}{h_{j+1}}\right)\right) \\
&- u^h_{i+1,j}\frac{h_j+h_{j+1}}{2}\,\frac{b_i}{\exp\sigma_i - 1} - u^h_{i-1,j}\frac{h_j+h_{j+1}}{2}\,\frac{b_{i-1}\exp\sigma_{i-1}}{\exp\sigma_{i-1}-1} \qquad (2.186)\\
&- u^h_{i,j-1}\varepsilon\frac{h}{h_j} - u^h_{i,j+1}\varepsilon\frac{h}{h_{j+1}} = f_{ij}\frac{h_j+h_{j+1}}{2}h, \quad i,j = 1,2,...,n-1;
\end{aligned}$$

$u^h_{ij} = 0$ for $i = 1,...,n-1$ and $j = 0, n$ or for $j = 1,...,n-1$ and $i = 0, n$.

The parameters $\{u^h_{ij}\}_{i,j=1}^{n-1}$ give the solution of the problem (2.185)

$$u^h = \sum_{i,j=1}^{n-1} u_{ij}\varphi_{ij}. \qquad (2.187)$$

Eliminate the boundary unknowns and enumerate the remaining unknowns and the equations from 1 to $(n-1)^2$ in the same way (for example, in the lexicographic order). We obtain the shortened system

$$A^h U = F \qquad (2.188)$$

where

$$\begin{aligned}
U &= (u^h_{1,1}, ..., u^h_{1,n-1}, u^h_{2,1}, ..., u^h_{n-1,n-1})^T, \\
F &= (f^h(\varphi_{1,1}), ..., f^h(\varphi_{1,n-1}), ..., f^h(\varphi_{n-1,n-1}))^T.
\end{aligned}$$

Note that the matrix A^h is irreducible [21], has positive diagonal elements and non-positive off-diagonal ones. Then this matrix is diagonal-dominant along columns and strictly diagonal-dominant along some of them. Therefore A^h is an M-matrix. Hence, the system (2.186) satisfies the comparison principle and has unique solution [21].

2.3.3 The properties of the discrete problem

Now we investigate the discrete problem. The following lemma establishes the error of the approximation of the problem (2.1), (2.2) by the grid problem (2.186).

Lemma 27. *Let* u *be a solution of the problem* (2.1), (2.2) *under the conditions* (2.4), (2.3), (2.70), *and* u^h *be a solution of the discrete problem* (2.186). *Then the estimate*

$$\begin{aligned}&|\left(L^h(u^h - u^I)\right)_{ij}| \\ &\le c_1 h(h_j + h_{j+1})\left(\varepsilon + h + \exp\left(-(1 - x_{i+1})B_1/2\varepsilon\right)\right) \quad \forall\, i, j = 1, ..., n-1\end{aligned} \tag{2.189}$$

holds.

Proof. Consider the operator L^h as the sum of two operators of difference differentiation with respect to x and y

$$L^h v = L_1^h v + L_2^h v \tag{2.190}$$

where

$$\begin{aligned}\left(L_1^h v\right)_{ij} = &\left(\left(\frac{b_i \exp\sigma_i}{\exp\sigma_i - 1} + \frac{b_{i-1}}{\exp\sigma_{i-1} - 1}\right) v_{ij} - \frac{b_i}{\exp\sigma_i - 1} v_{i+1,j}\right. \\ &- \left.\frac{b_{i-1}\exp\sigma_{i-1}}{\exp\sigma_{i-1} - 1} v_{i-1,j}\right) \frac{h_j + h_{j+1}}{2},\end{aligned} \tag{2.191}$$

$$\begin{aligned}\left(L_2^h v\right)_{ij} = &\,\varepsilon h\left(\frac{1}{h_j} + \frac{1}{h_{j+1}}\right) v_{ij} - \varepsilon\frac{h}{h_j} v_{i,j-1} - \varepsilon\frac{h}{h_{j+1}} v_{i,j+1}, \\ &i, j = 1, ..., n-1.\end{aligned} \tag{2.192}$$

Using the Tailor expansion at $(x_i, y_j) \in \bar{\Omega}_h$, we have

$$\begin{aligned}\left(L_2^h u\right)_{ij} = &\frac{1}{2} - \varepsilon h(h_j + h_{j+1})\partial_{22} u(x_i, y_j) \\ &- \varepsilon h\left(h_j^2 \pi_1(x_i, y) + h_{j+1}^2 \pi_2(x_i, y)\right).\end{aligned}$$

Because of (2.133) the inequalities

$$\begin{aligned}&|\pi_1(x_i, y)| \le c_2\left(1 + \varepsilon^{-3/2} B(y_{j-1})\right), \ |\pi_2(x_i, y)| \le c_3\left(1 + \varepsilon^{-3/2} B(y_j)\right), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } y_j \le 1/2, \\ &|\pi_1(x_i, y)| \le c_4\left(1 + \varepsilon^{-3/2} B(y_j)\right), \ |\pi_2(x_i, y)| \le c_5\left(1 + \varepsilon^{-3/2} B(y_{j+1})\right), \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } y_j > 1/2\end{aligned}$$

hold. In view of the definition (2.173) of the mesh size in the y direction, we get

$$|h_j \pi_1(x_i, y)|, \ |h_{j+1}\pi_2(x_i, y)| \le c_6 h\varepsilon^{-1}.$$

Then we obtain

$$\left(L_2^h u\right)_{ij} = -\varepsilon h \frac{h_j + h_{j+1}}{2} \partial_{22} u(x_i, y_j) + c_7 h^2 (h_j + h_{j+1}) \pi_3(x_i, y) \quad (2.193)$$

where $\pi_3(x_i, y)$ is bounded on $[y_{j-1}, y_{j+1}]$.

Using the expansion (2.125), we can write

$$L_1^h u = L_1^h u_0 + L_1^h \rho_0 + \varepsilon L_1^h \eta. \quad (2.194)$$

Now we consider each term in detail.

Using the Taylor expansion of u_0 at (x_i, y_j), we have

$$\begin{aligned} \left(L_1^h u_0\right)_{ij} = \frac{h_j + h_{j+1}}{2} \Bigg(& (b_i - b_{i-1}) u_{0ij} \\ & + \left(\frac{b_i}{\exp \sigma_i - 1} - \frac{b_{i-1} \exp \sigma_{i-1}}{\exp \sigma_{i-1} - 1} \right) h \partial_1 u_0(x_i, y_j) \Bigg) \\ & + h^2 (h_j + h_{j+1}) \pi_4(x, y_j). \end{aligned} \quad (2.195)$$

Since $b(x)$ is smooth, by the mean value theorem we can write (2.195) in the form

$$\begin{aligned} \left(L_1^h u_0\right)_{ij} = {} & h \frac{h_j + h_{j+1}}{2} \left(b'(x_i) u_{0ij} + b_i \partial_1 u_0(x_i, y_j) \right) \\ & + h^2 (h_j + h_{j+1}) \pi_5(x, y_j) \end{aligned} \quad (2.196)$$

where π_5 is bounded on $[x_{i-1}, x_{i+1}]$.

Further, using the explicit form of the boundary layer function $\rho_0(x, y)$ in (2.191), we obtain

$$\begin{aligned} \left(L_1^h \rho_0\right)_{ij} = \Bigg(& \left(\frac{b_i \exp \sigma_i}{\exp \sigma_i - 1} + \frac{b_{i-1}}{\exp \sigma_{i-1} - 1} \right) u_0(1, y_j) s_i \exp\left(-(1 - x_i) b_i / \varepsilon\right) \\ & - \frac{b_i}{\exp \sigma_i - 1} u_0(1, y_j) s_{i+1} \exp\left(-(1 - x_{i+1}) b_{i+1} / \varepsilon\right) \\ & - \frac{b_{i-1} \exp \sigma_{i-1}}{\exp \sigma_{i-1} - 1} u_0(1, y_j) s_{i-1} \exp\left(-(1 - x_{i-1}) b_{i-1} / \varepsilon\right) \Bigg) \frac{h_j + h_{j+1}}{2}. \end{aligned} \quad (2.197)$$

Using the mean value theorem, the smoothness of $b(x)$, and (2.33), we have the estimate

$$\begin{aligned} & \left| \exp\left(-(1 - x_{i+1}) b_{i+1} / \varepsilon\right) - \exp\left(-(1 - x_{i+1}) b_i / \varepsilon\right) \right| \\ & \leq |b_i - b_{i-1}| \frac{1 - x_{i-1}}{\varepsilon} \exp\left(-(1 - x_{i-1}) b_i^* / \varepsilon\right) \leq c_8 h \exp\left(-(1 - x_{i+1}) B_1 / 2\varepsilon\right) \end{aligned}$$

where $b_i \in [B_1, B_2]$. Rearranging the terms in (2.197) and taking into consideration the smoothness of $s(x)$, we get

$$\begin{aligned}
\left(L_1^h \rho_0\right)_{ij} = \Bigg(\Bigg(& \frac{b_i \exp \sigma_i}{\exp \sigma_i - 1} \exp\left(-(1-x_i)b_i/\varepsilon\right) \\
&\qquad - \frac{b_i}{\exp \sigma_i - 1} \exp\left(-(1-x_{i+1})b_i/\varepsilon\right)\Bigg) \\
+ \Bigg(& \frac{b_{i-1}}{\exp \sigma_{i-1} - 1} \exp\left(-(1-x_i)b_{i-1}/\varepsilon\right) \\
&\qquad - \frac{b_{i-1} \exp \sigma_{i-1}}{\exp \sigma_{i-1} - 1} \exp\left(-(1-x_i)b_i/\varepsilon\right)\Bigg) \\
+ & \, h\pi_6(x, y_j)\Bigg) \frac{h_j + h_{j+1}}{2} u_0(1, y_j) = h(h_j + h_{j+1})\pi_6(x, y_j)
\end{aligned} \tag{2.198}$$

where

$$|\pi_6(x, y_j)| \le c_9 \exp\left(-(1-x_{i+1})B_1/2\varepsilon\right), \quad x \in [x_{i-1}, x_{i+1}].$$

By the mean value theorem, from (2.191) the equality

$$\varepsilon\left(L_1^h \eta\right)_{ij} = \frac{1}{2}\varepsilon(h_j + h_{j+1})(b_i - b_{i-1})\eta_{ij} + h(h_j + h_{j+1})\pi_7(x, y_j) \tag{2.199}$$

follows where due to (2.148) we have

$$|\pi_7(x, y_j)| \le c_{10}\left(\varepsilon + \exp\left(-(1-x_{i+1})B_1/2\varepsilon\right)\right), \quad x \in [x_{i-1}, x_{i+1}]. \tag{2.200}$$

Taking into consideration (2.147) and the smoothness of $b(x)$, the equality (2.199) can be rewritten as

$$\varepsilon\left(L_1^h \eta\right)_{ij} = h(h_j + h_{j+1})\pi_8(x, y_j) \tag{2.201}$$

with the estimate of π_8 similar to (2.200).

Thus, combining (2.193), (2.196), (2.198), and (2.201), we obtain

$$\begin{aligned}
\left|\left(L^h(u^h - u)\right)_{ij}\right| = \Big|\left(L^h u^h\right)_{ij} + \Big(&\varepsilon\partial_{22}u(x_i, y_j) - b_i\partial_1 u_0(x_i, y_j) \\
&- u_{0ij}b'(x_i)\Big) h\frac{h_j + h_{j+1}}{2} + h(h_j + h_{j+1})\pi_{10}(x, y)\Big|
\end{aligned} \tag{2.202}$$

where

$$|\pi_{10}(x,y)| \leq c_{10}\left(h+\varepsilon+\exp\left(-(1-x_{i+1})B_1/2\varepsilon\right)\right),$$
$$x \in [x_{i-1},x_{i+1}],\ y \in [x_{j-1},y_{j+1}].$$

At the nodes of the grid the following equality holds:

$$\left(L^h u^h\right)_{ij} = h\frac{h_j+h_{j+1}}{2} f_{ij} = h\frac{h_j+h_{j+1}}{2}(Lu)_{ij}.$$

Now consider the expression

$$L_1 v = -\varepsilon\partial_{11}v + \partial_1(b(x)v).$$

Use the expansion (2.125) and write

$$L_1 u = L_1 u_0 + L_1\rho_0 + \varepsilon L_1\eta.$$

Because of (2.33), for $L_1\rho_0$ the following estimate holds:

$$\begin{aligned} |L_1\rho_0| = \Big| & g(y)\exp\left(-(1-x)b(x)/\varepsilon\right)\Big((1-x)b'' \\ & + \varepsilon^{-1}(1-x)bb' - b' + \varepsilon^{-1}(1-x)^2\left(b'\right)^2\Big)\Big| \\ & \leq c_{11}\exp\left(-(1-x)B_1/2\varepsilon\right) \quad \text{for} \quad x \in [x_i,x_{i+1}]. \end{aligned}$$

Taking into account (2.147), (2.148) and (2.149), we get

$$\varepsilon\,|L_1\eta| \leq c_{12}\big(\varepsilon+\exp\left(-(1-x)B_1/2\varepsilon\right)\big) \quad \text{for} \quad x \in [x_i,x_{i+1}].$$

Thus, with (2.138) for $\partial_{11}u_0$, we obtain

$$(L_1 u)_{ij} = b_i\partial_1 u_0(x_i,y_j) + u_{0ij}b'(x_i) + \pi_{11}(x,y) \tag{2.203}$$

where π_{11} is estimated similarly to (2.200).

By substituting (2.203) into (2.202) we complete the proof. □

In order to prove the convergence of the numerical solution to the exact one, we define the barrier function for the right-hand side of (2.189).

Lemma 28. *There exist grid functions φ^h and ψ^h on $\bar{\Omega}_h$ with the properties*

$$|\varphi^h| \leq c_1 \quad \text{on} \quad \Omega_h, \tag{2.204}$$
$$|\psi^h| \leq c_2 h \quad \text{on} \quad \Omega_h, \tag{2.205}$$

such that

$$L^h\varphi^h \geq hh_{j+1} \quad in \quad \Omega_h, \tag{2.206}$$

$$\varphi^h \geq 0 \quad on \quad \Gamma_h; \tag{2.207}$$

$$L^h\psi^h \geq hh_{j+1}\exp\left(-B_1(1-x_{i+1})/2\varepsilon\right) \quad in \quad \Omega_h, \tag{2.208}$$

$$\psi^h \geq 0 \quad on \quad \Gamma_h. \tag{2.209}$$

This lemma is proved in much the same way as Lemma 21.

Finally, the following convergence result is valid.

Theorem 29. *Assume that* (2.4), (2.3) *hold. Then there exist constants* h_0 *and* c *independent of* h *and* ε *such that* $\forall\ h \leq h_0$ *and for* $\varepsilon \leq h$ *the solution* u^h *of the problem* (2.186) *satisfies the estimate*

$$\max_{\hat{\Omega}_h} |u - u^h| \leq ch \tag{2.210}$$

where u *is the solution of the problem* (2.18), (2.19).

The proof follows from Lemmata 27 and 28 as in the previous case.

Thus, we constructed the grid problem for the convection-diffusion problem with regular and parabolic boundary layers. Its solution converges to the exact one with the first order in the uniform discrete norm. The numerical experiments described in Chapter 3 confirm this.

3 Numerical solution of the discrete problem

3.1 Numerical experiments in the one-dimensional case

As a test example we considered the problem

$$-\varepsilon u'' + ((1+2x)u)' = f, \quad x \in (0,1),$$
$$u(0) = u(1) = 0$$

where

$$f(x) = 6x^2 + 2x - 2\varepsilon + 2d, \quad d = \frac{\exp(-2/\varepsilon)}{1-\exp(-2/\varepsilon)}.$$

The parameter ε was taken in the range from 1/10 to 1/5120. The exact solution of this problem is given by

$$u(x) = x^2 + d - (d+1)\exp\left(\frac{x^2+x-2}{\varepsilon}\right).$$

We compared the numerical results obtained by the stable upwind scheme

$$\begin{aligned} -\frac{\varepsilon}{h^2}(u_{i+1} - 2u_i + u_{i-1}) + b_i \frac{u_i - u_{i-1}}{h} + b_i' u_i = f_i, \\ i = 1, ..., n-1, \quad u_0 = u_n = 0, \end{aligned} \tag{3.1}$$

by the difference scheme with exponential fitting (see [23])

$$\begin{aligned} -\frac{\varepsilon\sigma_i}{h^2}\,(u_{i+1} - 2u_i + u_{i-1}) + \frac{b_i}{2h}\,(u_{i+1} - u_{i-1}) + b_i' u_i = f_i, \\ u_0 = u_n = 0 \end{aligned} \tag{3.2}$$

with the variable fitting coefficient $\sigma_i = \frac{b_i h}{2\varepsilon} \mathrm{cth} \left(\frac{b_i h}{2\varepsilon} \right)$; by the proposed two schemes (1.53)–(1.54) and (1.80)–(1.81); and by the first-order scheme from [122]. The number n of the nodes of the grid varied from 10 to 320 and the mesh size was difined as $h = 1/n$. The error of the numerical solution was calculated exactly:

$$\delta(n) = \|u - u^h\|_{\infty,h} = \max_{\omega^h} |u_{ij} - u_{ij}^h|.$$

Table 1. The error of the simple upwind scheme (3.1).

ε	h					
	1/10	1/20	1/40	1/80	1/160	1/320
1/10	1.51$_{10}$-1	1.53$_{10}$-1	8.98$_{10}$-2	5.33$_{10}$-2	2.86$_{10}$-2	1.48$_{10}$-2
1/20	8.96$_{10}$-2	1.74$_{10}$-1	1.65$_{10}$-1	9.64$_{10}$-2	5.66$_{10}$-2	3.02$_{10}$-2
1/40	4.76$_{10}$-2	1.13$_{10}$-1	1.87$_{10}$-1	1.71$_{10}$-1	9.98$_{10}$-2	5.83$_{10}$-2
1/80	5.23$_{10}$-2	4.79$_{10}$-2	1.26$_{10}$-1	1.93$_{10}$-1	1.74$_{10}$-1	1.02$_{10}$-1
1/160	5.37$_{10}$-2	2.87$_{10}$-2	6.20$_{10}$-2	1.33$_{10}$-1	1.97$_{10}$-1	1.75$_{10}$-1
1/320	5.41$_{10}$-2	2.44$_{10}$-2	2.44$_{10}$-2	6.93$_{10}$-2	1.37$_{10}$-1	1.99$_{10}$-1
1/640	5.40$_{10}$-2	3.01$_{10}$-2	1.55$_{10}$-2	3.21$_{10}$-2	7.30$_{10}$-2	1.39$_{10}$-1
1/1280	5.72$_{10}$-2	3.02$_{10}$-2	1.58$_{10}$-2	1.23$_{10}$-2	3.60$_{10}$-2	7.50$_{10}$-2
1/2560	5.80$_{10}$-2	3.02$_{10}$-2	1.59$_{10}$-2	8.05$_{10}$-3	1.64$_{10}$-2	3.80$_{10}$-2
1/5120	5.93$_{10}$-2	3.03$_{10}$-2	1.59$_{10}$-2	8.11$_{10}$-3	6.21$_{10}$-3	1.84$_{10}$-2

The numerical results are given in Tables 1–5 and in Figures 5–7. In the figures the error of the simple upwind scheme is marked by the number 2, the error of the first-order scheme from [122] with the fitted quadrature rule is marked by 3, the error of the difference scheme with exponential fitting

Table 2. The error of the fitted first-order finite element scheme from [128].

ε	h					
	1/10	1/20	1/40	1/80	1/160	1/320
1/10	1.71_{10}-2	3.01_{10}-3	3.83_{10}-3	2.54_{10}-3	1.43_{10}-3	7.55_{10}-4
1/20	3.49_{10}-2	8.04_{10}-3	1.90_{10}-3	2.24_{10}-3	1.46_{10}-3	8.16_{10}-4
1/40	4.95_{10}-2	1.73_{10}-2	3.88_{10}-3	1.05_{10}-3	1.21_{10}-3	7.82_{10}-4
1/80	5.70_{10}-2	2.49_{10}-2	8.62_{10}-3	1.90_{10}-3	5.51_{10}-4	6.27_{10}-4
1/160	6.07_{10}-2	2.89_{10}-2	1.25_{10}-2	4.30_{10}-3	9.43_{10}-4	2.82_{10}-4
1/320	6.26_{10}-2	3.09_{10}-2	1.45_{10}-2	6.25_{10}-3	2.14_{10}-3	4.69_{10}-4
1/640	6.36_{10}-2	3.19_{10}-2	1.55_{10}-2	7.28_{10}-3	3.12_{10}-3	1.07_{10}-3
1/1280	6.40_{10}-2	3.24_{10}-2	1.60_{10}-2	7.79_{10}-3	3.64_{10}-3	1.56_{10}-3
1/2560	6.43_{10}-2	3.26_{10}-2	1.63_{10}-2	8.05_{10}-3	3.90_{10}-3	1.82_{10}-3
1/5120	6.44_{10}-2	3.27_{10}-2	1.64_{10}-2	8.18_{10}-3	4.03_{10}-3	1.95_{10}-3

Table 3. The error of the difference scheme (3.2) with exponential fitting.

ε	h					
	1/10	1/20	1/40	1/80	1/160	1/320
1/10	1.58_{10}-2	6.23_{10}-3	2.79_{10}-3	6.37_{10}-4	1.08_{10}-4	1.77_{10}-5
1/20	3.07_{10}-2	9.29_{10}-3	2.93_{10}-3	1.37_{10}-3	3.15_{10}-4	5.31_{10}-5
1/40	4.51_{10}-2	1.65_{10}-2	5.02_{10}-3	1.41_{10}-3	6.80_{10}-4	1.56_{10}-4
1/80	5.25_{10}-2	2.38_{10}-2	8.69_{10}-3	2.63_{10}-3	7.01_{10}-4	3.38_{10}-4
1/160	5.63_{10}-2	2.77_{10}-2	1.22_{10}-2	4.47_{10}-3	1.35_{10}-3	3.58_{10}-4
1/320	5.81_{10}-2	2.97_{10}-2	1.42_{10}-2	6.19_{10}-3	2.27_{10}-3	6.82_{10}-4
1/640	5.91_{10}-2	3.07_{10}-2	1.52_{10}-2	7.20_{10}-3	3.12_{10}-3	1.14_{10}-3
1/1280	5.95_{10}-2	3.12_{10}-2	1.57_{10}-2	7.71_{10}-3	3.62_{10}-3	1.56_{10}-3
1/2560	5.98_{10}-2	3.14_{10}-2	1.60_{10}-2	7.97_{10}-3	3.88_{10}-3	1.82_{10}-3
1/5120	5.99_{10}-2	3.15_{10}-2	1.61_{10}-2	8.10_{10}-3	4.01_{10}-3	1.95_{10}-3

Table 4. The error of the fitted finite element scheme (1.53)–(1.54) with the linear quadrature rule.

ε	h					
	1/10	1/20	1/40	1/80	1/160	1/320
1/10	2.53$_{10}$-3	1.19$_{10}$-3	3.74$_{10}$-4	1.02$_{10}$-4	2.66$_{10}$-5	6.79$_{10}$-6
1/20	1.29$_{10}$-3	8.30$_{10}$-4	3.80$_{10}$-4	1.18$_{10}$-4	3.18$_{10}$-5	8.25$_{10}$-6
1/40	1.51$_{10}$-3	6.55$_{10}$-4	2.94$_{10}$-4	1.32$_{10}$-4	4.95$_{10}$-5	1.20$_{10}$-5
1/80	2.29$_{10}$-3	3.98$_{10}$-4	3.29$_{10}$-4	1.15$_{10}$-4	5.09$_{10}$-5	2.75$_{10}$-5
1/160	2.77$_{10}$-3	6.00$_{10}$-4	1.02$_{10}$-4	1.65$_{10}$-4	4.95$_{10}$-5	2.74$_{10}$-5
1/320	3.00$_{10}$-3	7.01$_{10}$-4	1.53$_{10}$-4	3.21$_{10}$-5	8.25$_{10}$-6	2.27$_{10}$-5
1/640	3.11$_{10}$-3	7.63$_{10}$-4	1.79$_{10}$-4	3.87$_{10}$-5	2.24$_{10}$-5	4.12$_{10}$-5
1/1280	3.16$_{10}$-3	7.92$_{10}$-4	1.92$_{10}$-4	4.52$_{10}$-5	9.72$_{10}$-6	1.28$_{10}$-5
1/2560	3.19$_{10}$-3	8.06$_{10}$-4	2.00$_{10}$-4	4.84$_{10}$-5	1.13$_{10}$-5	2.44$_{10}$-5
1/5120	3.20$_{10}$-3	8.12$_{10}$-4	2.03$_{10}$-4	5.00$_{10}$-5	1.22$_{10}$-5	2.84$_{10}$-6

Table 5. The error of the fitted finite element scheme (1.80)–(1.81) with the nonlinear quadrature rule.

ε	h					
	1/10	1/20	1/40	1/80	1/160	1/320
1/10	2.27$_{10}$-3	1.50$_{10}$-3	4.11$_{10}$-4	1.12$_{10}$-4	2.86$_{10}$-5	7.22$_{10}$-6
1/20	9.24$_{10}$-4	1.12$_{10}$-3	7.13$_{10}$-4	1.94$_{10}$-4	5.20$_{10}$-5	1.32$_{10}$-5
1/40	1.84$_{10}$-3	2.39$_{10}$-4	5.54$_{10}$-4	3.47$_{10}$-4	9.43$_{10}$-5	2.49$_{10}$-5
1/80	2.46$_{10}$-3	4.67$_{10}$-4	6.02$_{10}$-5	2.75$_{10}$-4	1.71$_{10}$-4	4.64$_{10}$-5
1/160	2.82$_{10}$-3	6.29$_{10}$-4	1.17$_{10}$-4	2.75$_{10}$-5	1.37$_{10}$-4	8.47$_{10}$-5
1/320	3.01$_{10}$-3	7.20$_{10}$-4	1.59$_{10}$-4	2.94$_{10}$-5	1.73$_{10}$-5	6.83$_{10}$-5
1/640	3.11$_{10}$-3	7.69$_{10}$-4	1.81$_{10}$-4	3.99$_{10}$-5	7.38$_{10}$-6	9.53$_{10}$-6
1/1280	3.16$_{10}$-3	7.94$_{10}$-4	1.94$_{10}$-4	4.57$_{10}$-5	9.99$_{10}$-6	1.85$_{10}$-6
1/2560	3.19$_{10}$-3	8.06$_{10}$-4	2.00$_{10}$-4	4.84$_{10}$-5	1.14$_{10}$-5	2.50$_{10}$-6
1/5120	3.20$_{10}$-3	8.13$_{10}$-4	2.03$_{10}$-4	5.03$_{10}$-5	1.22$_{10}$-5	2.86$_{10}$-6

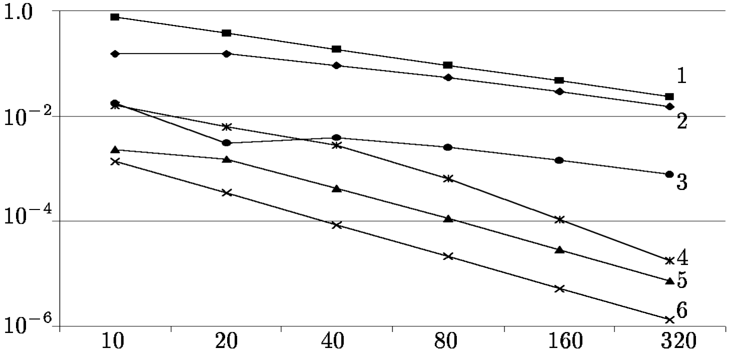


Fig. 5. The maximum error $\delta(n)$ in the one-dimensional case for $\varepsilon = 1/10$.

is marked by 4, and the error of the presented finite element scheme with the nonlinear quadrature rule is marked by 5. For comparison we show the straight lines with slopes tg$\varphi = 1$ and tg$\varphi = 2$ which are marked by 1 and 6 respectively. The numerical results for the presented scheme with the linear qudrature rule does not differ visually from the polygonal line 5 and are not shown in the figures.

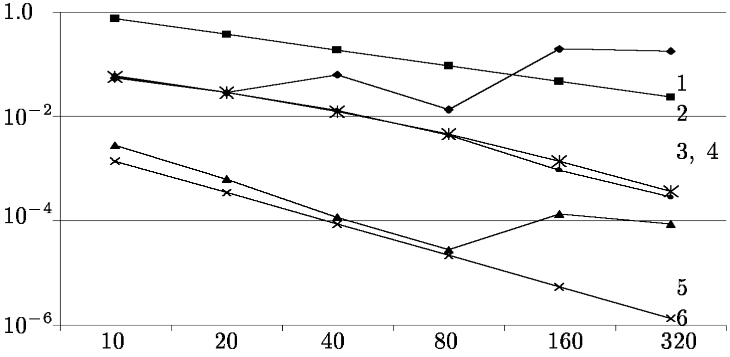


Fig. 6. The maximum error $\delta(n)$ in the one-dimensional case, for $\varepsilon = 1/160$.

The study of the behaviour of the error of the simple upwind scheme (3.1) shows that for $\varepsilon > h$ (Fig. 5) this scheme has the first-order accuracy, but with ε decreasing the order of accuracy decreases. In [103] it was shown that the scheme (3.2) with exponential fitting has the second-order accuracy for $\varepsilon > h$ (Fig. 5) and only the first-order accuracy for small value of ε. This is seen in Figs. 5–7, where the slope of the polygonal line 4 is changed

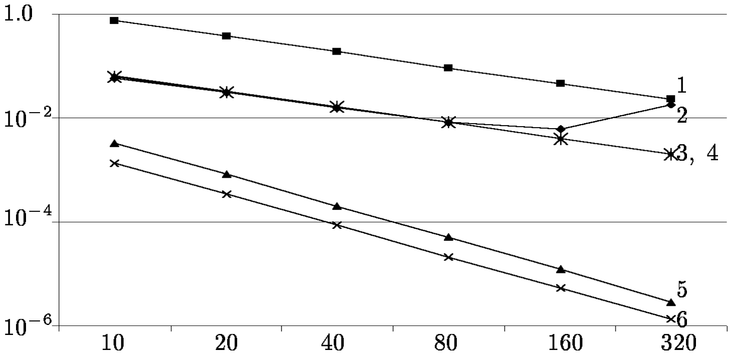


Fig. 7. The maximum error $\delta(n)$ in the one-dimensional case for $\varepsilon = 1/5120$.

at around $\varepsilon = h$. Calculations for the scheme from [122] confirm the first-order convergence for small values of the diffusion parameter. The presented schemes have the second-order convergence not only for $\varepsilon < h/2$ that the theoretically proved but for $\varepsilon > 2h$ as well. The results (see Tables 4, 5 and Figs. 5–7, polygonal line 5) show that for all values of ε these scheme are more accurate than those consider here.

3.2 Test example in the two-dimensional case

Let Ω be the square $(0,1) \times (0,1)$ with the boundary Γ. As a test example we consider the problem

$$-\varepsilon\Delta u + \partial_1 u = 1 \quad \text{in} \quad \Omega, \tag{3.3}$$

$$u = 0 \quad \text{on} \quad \Gamma. \tag{3.4}$$

The solution of this problem has a parabolic boundary layer along the boundary Γ_{tg} and a regular boundary layer near Γ_{out}.

The calculations were done on grids uniform in the x-direction. To refine the grid in the y-direction in the parabolic boundary layer, two approaches were considered. The first approach has been proposed by N.S.Bakhvalov in [5]. This approach uses the estimates of the normal derivative of the solution. We consider two kinds of these grids. The second approach has been considered by G.I.Shishkin ([112], [58]). He use the grid with a piecewise constant mesh size that is refined in the parabolic boundary layer.

To solve the discrete problem we applied the pointwise and block Gauss-Seidel methods. We also use the cascadic multigrid algorithm where the interpolation on a coarser grid is taken as the initial guess on a finer one.

3.3 The grids

First we construct the grid in the y-direction according to the works by N.S.Bakhvalov [5] and V.D.Liseikin [36], [37].

Define the nodes of the grid on the segment $[0, 1]$ by a non-singular transformation $\lambda(q) : [0, 1] \to [0, 1]$ in the following way:

$$y_j = \lambda(jh), \quad j = 0, 1, ..., N, \quad h = 1/N. \tag{3.5}$$

The generating function $\lambda(q)$ is taken so that the difference of the values of the solution at the neighboring nodes in the y-direction is uniformly founded:

$$|u(x, y_{j+1}) - u(x, y_j)| \le c_1 h, \quad j = 0, 1, ..., N - 1.$$

This condition is satisfied if $\lambda(q)$ is a piecewise smooth function and

$$\left| \frac{\partial u(x, \lambda(q))}{\partial q} \right| \le c_2, \quad q \in (0, 1). \tag{3.6}$$

According to [36], the use of the estimates (2.133) of the derivative $\partial_2 u(x, y)$ instead of the derivative itself leads to the stronger condition

$$\left| \frac{\partial^k u(x, \lambda(q))}{\partial q^k} \right| \le c_3, \quad 0 \le q \le 1, \quad k > 1. \tag{3.7}$$

Since the solution has two parabolic boundary layers in Ω near the boundaries $\Gamma^0_{tg} = \{(x, y) : x \in [0, 1], \quad y = 0\}$ and $\Gamma^1_{tg} = \{(x, y) : x \in [0, 1], \quad y = 1\}$, we consider the function $\lambda(q)$ which is symmetric about the point $q = 0.5$. The explicit form of the local transformation $\lambda(q)$ in the vicinity of a parabolic boundary layer, for example, near Γ^0_{tg}, can be found as the solution of the problem

$$\frac{dq}{dy} = c_4 \exp\left(-\gamma y/\sqrt{\varepsilon}\right), \quad q(0) = 0, \quad c_4 = 1/\int_0^{q_*} \exp\left(-\gamma t/\sqrt{\varepsilon}\right)\, dt$$

where $q_* > q$ is the thickness of a boundary layer.

Then on $[0, q_*]$ the generating function has the form

$$\lambda(q) = \sqrt{\varepsilon} \ln\left(1 + 4(1 - \sqrt{\varepsilon})q\right), \quad 0 \le q \le q_*. \tag{3.8}$$

On the segment $[q_*, 0.5]$ the function $\lambda(q)$ is the tangent $\gamma q + \delta$ to the curve (3.8) at the point q_*. The point q_* of sewing and the parameters γ, δ are obtained by the following iterative process:

1. the point $q_*^0 = h[n/4]$ is taken as an initial guess;
2. with the value q_*^k we construct the straight line $\gamma^k q + \delta^k$ passing through the points $(q_*^k, \lambda(q_*^k))$ and $(0.5,\ 0.5)$;
3. we determine q_*^{k+1} from the equation

$$\frac{\partial\lambda(q_*^{k+1})}{\partial q} = \gamma^k;$$

4. if $|q_*^{k+1} - q_*^k| > \delta_{step}$ with the a priori chosen error δ_{step} then go to step (2) else q_*^k is chosen as the point of sewing, γ^k and δ^k are taken as the parameters of the straight line, and the iterative process is terminated.

Thus, the generating function for the Bakhvalov grid has the form

$$\lambda(q) = \begin{cases} \sqrt{\varepsilon}\ln\left(1 + 4(1-\sqrt{\varepsilon})q\right), & 0 \le q \le q_*, \\ \gamma q + \delta, \quad q_* \le q \le 0.5, \\ 1 - \lambda(1-q), & 0.5 \le q \le 1. \end{cases}$$

We can consider the grids presented in Chapter 2 as grids of the Bakhvalov type, since they are constructed using the information on the behaviour of the normal derivative of the solution.

In this case unlike (3.5) the generating function can not be defined exactly. The nodes of the grid are determined by the equalities

$$\begin{aligned} &y_0 = 0, \\ &y_j = y_{j-1} + \frac{c_0 h}{1 + \varepsilon^{-1/2}\exp\left(-\gamma y_{j-1}/\sqrt{\varepsilon}\right)}, \quad j = 1, 2, ..., \frac{n}{2}, \\ &y_{n/2} = 0.5, \\ &y_j = 1 - y_{n-j}, \quad j = \frac{n}{2} + 1, ..., n. \end{aligned} \tag{3.9}$$

Here $h = 1/n$. The grid parameter c_0 is determined from the nonlinear equation

$$y_{n/2} = y_{n/2-1} + \frac{c_0 h}{1 + \varepsilon^{-1/2}\exp\left(-\gamma y_{n/2-1}/\sqrt{\varepsilon}\right)}.$$

In the numerical experiments we used the Jacobi-type iterative process.

Another way of grid refinement that we used in the numerical experiments has been proposed by G.I.Shishkin (see, for example, [112], [58]).

Let $n+1$ be the number of nodes in the y-direction. The thickness of the numerical parabolic boundary layer is determined by

$$\tau = \min\{1/4, \sqrt{\varepsilon}\ln n\}.$$

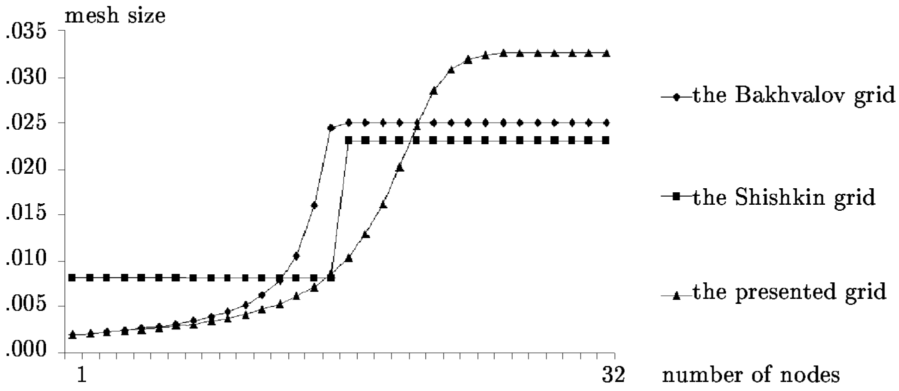


Fig. 8. The mesh-size functions.

The mesh-size is piecewise constant. In the vicinity of the parabolic boundary layer $y \in [0, \tau] \cup [1 - \tau, 1]$ it is taken by

$$h_p = \frac{\tau}{[n/4]}$$

and in the remaining part $y \in [\tau, 1 - \tau]$ it is determined by

$$h_r = \frac{1 - 2\tau}{[n/2]}.$$

Table 6. The characteristics of the grids.

n	number of nodes in bound. layer		thickness of bound. layer	
	Bakhvalov	Shishkin	Bakhvalov	Shishkin
32	8	8	5.943_{10}-2	1.096_{10}-1
64	16	16	7.540_{10}-2	1.315_{10}-1
128	31	32	7.540_{10}-2	1.554_{10}-1
256	62	64	8.107_{10}-2	1.750_{10}-1
512	123	128	8.107_{10}-2	1.972_{10}-1
1024	245	256	8.107_{10}-2	2.191_{10}-1
2048	489	512	8.107_{10}-2	2.411_{10}-1

The mesh size functions for each grid are demonstrated in Fig. 8. The number of nodes in the vicinity of the parabolic boundary layer and the thickness of the layer for the Bakhvalov and Shishkin grids given in Table 6 for various values n for $\varepsilon = 10^{-3}$. On the presented grid (3.9) the thickness of the boundary layer is not clearly defined.

3.4 Methods for solving the discrete problem

In Chapter 2 the discrete problem (2.188) was obtained. To solve it we apply the pointwise and block Gauss-Seidel methods. Now we briefly describe these methods according to [47].

We represent the matrix A^h in (2.188) as the sum of the lower triangular matrix B^h with a nonzero diagonal and the upper triangular matrix C^h with the zero diagonal

$$A^h = B^h + C^h \tag{3.10}$$

where

$$B^h = \begin{pmatrix} a^h_{11} & 0 & 0 & \cdots & 0 & 0 \\ a^h_{21} & a^h_{22} & 0 & \cdots & 0 & 0 \\ a^h_{31} & a_{32} & a^h_{33} & \cdots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a^h_{M1} & a^h_{M2} & a^h_{M3} & \cdots & a^h_{MM-1} & a^h_{MM} \end{pmatrix},$$

$$C^h = \begin{pmatrix} 0 & a^h_{12} & a^h_{13} & \cdots & a^h_{1M-1} & a^h_{1M} \\ 0 & 0 & a^h_{23} & \cdots & a^h_{2M-1} & a^h_{2M} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \cdots & 0 & a^h_{M-1,M} \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

$M = (n-1) \times (n-1)$.

Using these notations we rewrite the Gauss-Seidel method as

$$B^h u^{(k+1)} + C^h u^{(k)} = F, \; k = 0, 1, ...; \quad u^{(0)} = 0. \tag{3.11}$$

From here on, k is the number of iteration steps.

The iterative process (3.11) can be rewritten in the canonical form

$$B^h \left(u^{(k+1)} - u^{(k)} \right) + A^h u^{(k)} = F. \tag{3.12}$$

The operator B^h is a triangular matrix, hence it is not self-adjoint.

Taking into account (2.186), we can write the system (2.188) in the form

$$-a_{ij}u_{i-1,j} + b_{ij}u_{ij} - c_{ij}u_{i+1,j} - d_{ij}u_{i,j-1} - e_{ij}u_{i,j+1} = f_{ij}, \tag{3.13}$$
$$i, j = 1, ..., n-1$$

where

$$\begin{aligned} a_{ij} &= \frac{h_j + h_{j+1}}{2} \frac{b_{i-1} \exp \sigma_{i-1}}{\exp \sigma_{i-1} - 1}, \qquad & d_{ij} &= \varepsilon \frac{h}{h_j}, \\ c_{ij} &= \frac{h_j + h_{j+1}}{2} \frac{b_i}{\exp \sigma_i - 1}, \qquad & e_{ij} &= \varepsilon \frac{h}{h_{j+1}}, \\ b_{ij} &= a_{i+1,j} + c_{i-1,j} + d_{ij} + e_{ij}. \end{aligned} \tag{3.14}$$

Then the pointwise Gauss-Seidel method (3.11) can be rewritten in the form

$$u_{ij}^{(k+1)} = \frac{1}{b_{ij}} \left(f_{ij} + a_{ij} u_{i-1,j}^{(k+1)} + c_{ij} u_{i+1,j}^{(k)} + d_{ij} u_{i,j-1}^{(k+1)} + e_{ij} u_{i,j+1}^{(k)} \right), \tag{3.15}$$
$$i, j = 1, ..., n-1, \quad k = 0, 1, \ldots .$$

The numerical experiment demonstrated that this method failes, especially on the Bakhvalov grids. This method is sensitive to the grid refinement in a parabolic boundary layer. We can see this in the example given below.

At the same time A^h has a certain block structure. We use this property and consider the block Gauss-Seidel method. We denote by $\mathbf{U}_i = (u_{i,1}, u_{i,2}, ..., u_{i,n-1})^T$ the vector whose components are the values u_{ij} of the grid function for fixed i. Then the grid equations (3.13) can be rewritten as the system of three-level vector equations

$$-A_i \mathbf{U}_{i-1} + B_i \mathbf{U}_i - C_i \mathbf{U}_{i+1} = \mathbf{F}_i, \quad j = 1, 2, ..., n-1, \tag{3.16}$$

where A_i and C_i are diagonal $(n-1) \times (n-1)$ matrices. Here

$$diag\,(A_i) = (a_{i,1}, a_{i,2}, ..., a_{i,n-1})^T,$$

$$diag\,(C_i) = (c_{i,1}, c_{i,2}, ..., c_{i,n-1})^T,$$

$$\mathbf{F}_i = (f_{i,1}, f_{i,2}, ..., f_{i,n-1})^T,$$

and B_i is a tridiagonal $(n-1) \times (n-1)$ matrix

$$B_i = \begin{pmatrix} b_{i1} & -e_{i1} & 0 & 0 & \cdots & 0 & 0 & 0 \\ -d_{i2} & b_{i2} & -e_{i2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & -d_{i3} & b_{i3} & -e_{i3} & \cdots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \cdots & -d_{i,n-2} & b_{i,n-2} & -e_{i,n-2} \\ 0 & 0 & 0 & 0 & \cdots & 0 & b_{i,n-1} & -e_{i,n-1} \end{pmatrix}.$$

The block Gauss-Seidel method for the system (3.16) has the form

$$\begin{gathered} B_i \mathbf{U}_i^{(k+1)} = \mathbf{F}_i + A_i \mathbf{U}_{i-1}^{(k+1)} + C_i \mathbf{U}_{i+1}^{(k)}, \\ i = 1, 2, ..., n-1, \quad k = 0, 1.... \end{gathered} \tag{3.17}$$

To determin $\mathbf{U}_i^{(k+1)}$, one have to invert the tridiagonal matrix B_i. To do this, the sweep method can be applied.

The pointwise representation of (3.17) has the form

$$\begin{gathered} -d_{ij} u_{i,j-1}^{(k+1)} + b_{ij} u_{ij}^{(k+1)} - e_{ij} u_{i,j+1}^{(k+1)} = f_{ij} + a_{ij} u_{i-1,j}^{(k+1)} + c_{ij} u_{i+1,j}^{(k)}, \\ i = 1, ..., n-1, \quad j = 1, ..., n-1, \quad k = 0, 1.... \end{gathered} \tag{3.18}$$

The numerical experiments showed the advantage of the block Gauss-Seidel method over the pointwise one.

Moreover, the convergence of the block Gauss-Seidel method is independent of the grid refinement in the vicinity of a parabolic boundary layer. Here we prove this theoretically. Denote the error of the block iterative method after k iteration steps by

$$r_{ij}^{(k)} = u_{ij}^{(k)} - u_{ij}, \qquad i = 1, 2, \dots, n-1.$$

We fix i and take the maximum of the modulus of $r_{ij}^{(k)}$ which is achieved at some j_0:

$$\left| r_{i,j_0}^{(k)} \right| = \max_{1 \le j \le n-1} \left| r_{ij}^{(k)} \right|. \tag{3.19}$$

We subtract (3.13) from (3.18) and obtain

$$-d_{i,j_0} r_{i,j_0-1}^{(k+1)} + b_{i,j_0} r_{i,j_0}^{(k+1)} - e_{ij_0} r_{i,j_0+1}^{(k+1)} = a_{i,j_0} r_{i-1,j_0}^{(k+1)} + c_{i,j_0} r_{i+1,j_0}^{(k)}.$$

Rearranging some terms to the right-hand side and taking modulus of both sides, we have

$$b_{i,j_0} \left| r_{i,j_0}^{(k+1)} \right| \le d_{i,j_0} \left| r_{i,j_0-1}^{(k+1)} \right| + e_{i,j_0} \left| r_{i,j_0+1}^{(k+1)} \right| + a_{i,j_0} \left| r_{i-1,j_0}^{(k+1)} \right| + c_{i,j_0} \left| r_{i+1,j_0}^{(k)} \right|.$$

Using (3.19) we rewrite the last inequality in the form

$$(b_{i,j_0} - d_{i,j_0} - e_{i,j_0}) \left| r_{i,j_0}^{(k+1)} \right| \le a_{i,j_0} \left| r_{i-1,j_0}^{(k+1)} \right| + c_{i,j_0} \left| r_{i+1,j_0}^{(k)} \right|. \tag{3.20}$$

Thus, we get

$$-a_{ij_0} \left| r_{i-1,j_0}^{(k+1)} \right| + s_i \left| r_{i,j_0}^{(k+1)} \right| - c_{ij_0} \left| r_{i+1,j_0}^{(k)} \right| \le 0 \tag{3.21}$$

where $s_i = b_{i,j_0} - d_{i,j_0} - e_{i,j_0}$. Let us introduce the notation

$$\rho_i = \frac{b_i}{\exp\sigma_i - 1}$$

where $\sigma_i = b_i h/\varepsilon$. Then $\dfrac{b_i \exp\sigma_i}{\exp\sigma_i - 1} = b_i + \rho_i$. With this notation, dividing the inequality (3.21) by $\dfrac{h_{j_0} + h_{j_0+1}}{2}$ we get

$$-(b_{i-1} + \rho_{i-1}) \left| r^{(k+1)}_{i-1,j_0} \right| + (b_i + \rho_{i-1} + \rho_i) \left| r^{(k+1)}_{i,j_0} \right| - \rho_i \left| r^{(k)}_{i+1,j_0} \right| \le 0. \quad (3.22)$$

Then we consider the k-th iteration step of the majorized Gauss-Seidel process

$$-(b_{i-1} + \rho_{i-1}) t^{(k+1)}_{i-1} + (b_i + \rho_{i-1} + \rho_i) t^{(k+1)}_i - \rho_i t^{(k)}_{i+1} = 0,\ i = 1, 2, \ldots, n-1,$$
$$t^{(k+1)}_0 = t^{(k+1)}_n = 0. \quad (3.23)$$

Lemma 30. *Let the inequality*

$$\left| r^{(k)}_{i,j_0} \right| \le t^{(k)}_i$$

be valid for all $i = 1, 2, \ldots, n-1$. Then the estimate

$$\left| r^{(k+1)}_{i,j_0} \right| \le t^{(k+1)}_i \quad \forall\, i = 1, 2, \ldots, n-1 \quad (3.24)$$

holds.

Proof. Because of (3.21) we have

$$\left| r^{(k+1)}_{i,j_0} \right| \le \frac{b_{i-1} + \rho_{i-1}}{b_i + \rho_{i-1} + \rho_i} \left| r^{(k+1)}_{i-1,j_0} \right| + \frac{\rho_i}{b_i + \rho_{i-1} + \rho_i} \left| r^{(k)}_{i+1,j_0} \right|, \quad i = 1, 2, \ldots, n-1.$$

Taking into account (3.23) we get

$$t^{(k+1)}_i = \frac{b_{i-1} + \rho_{i-1}}{b_i + \rho_{i-1} + \rho_i} t^{(k+1)}_{i-1} + \frac{\rho_i}{b_i + \rho_{i-1} + \rho_i} t^{(k)}_{i+1} \quad i = 1, 2, \ldots, n-1.$$

Now we use induction on i.

1. For $i = 1$ we have
$$t^{(k+1)}_1 = \frac{\rho_1}{b_1 + \rho_0 + \rho_1} t^{(k)}_2$$
and
$$\left| r^{(k+1)}_{1,j_0} \right| = \frac{\rho_1}{b_1 + \rho_0 + \rho_1} \left| r^{(k)}_{2,j_0} \right| \le \frac{\rho_1}{b_1 + \rho_0 + \rho_1} t^{(k)}_2 = t^{(k+1)}_1.$$

2. Let the statement (3.24) be valid for $i \leq m-1$. Then we obtain

$$\begin{aligned} \left|r_{m,j_0}^{(k+1)}\right| &= \frac{b_{m-1}+\rho_{m-1}}{b_m+\rho_{m-1}+\rho_m}\left|r_{m-1,j_0}^{(k+1)}\right| + \frac{\rho_m}{b_m+\rho_{m-1}+\rho_m}\left|r_{m+1,j_0}^{(k)}\right| \\ &\leq \frac{b_{m-1}+\rho_{m-1}}{b_m+\rho_{m-1}+\rho_m}t_{m-1}^{(k+1)} + \frac{\rho_m}{b_m+\rho_{m-1}+\rho_m}t_{m+1}^{(k)} = t_m^{(k+1)}. \end{aligned}$$

The proof of the lemma is completed. □

Thus, the convergence estimate of the block Gauss-Seidel method coincides with that of the pointwise Gauss-Seidel method for an ordinary differential equation and is independent of the grid refinement in the y-direction.

First, we investigate numerically the convergence of the pointwise and block Gauss-Seidel methods for a model problem free of a boundary layer on a uniform grid.

We consider the problem

$$-\varepsilon \Delta u + \partial_1 u = 0 \quad \text{in} \quad \Omega, \qquad u = 1 \quad \text{on} \quad \Gamma.$$

It has the exact solution $u \equiv 1$.

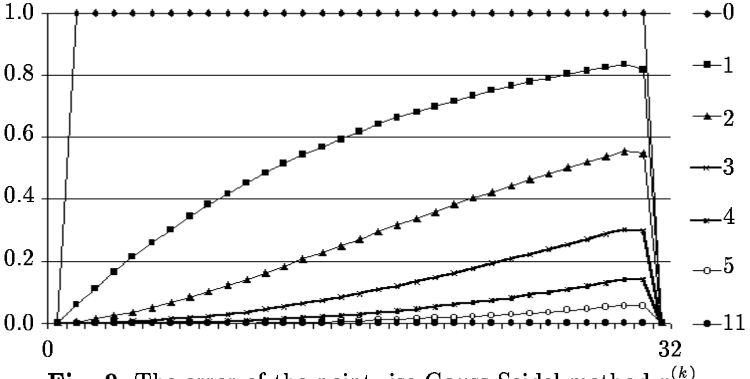


Fig. 9. The error of the pointwise Gauss-Seidel method $r_{i,0.5}^{(k)}$.

In the Figures 9 and 10 the behaviour of the error

$$s_{i,0.5}^{(k)} = \left|u(x_i, 0.5) - u^{(k)}(x_i, 0.5)\right|$$

along the middle line $y = 0.5$ after k iteration steps is demonstrated for the pointwise and block Gauss-Seidel methods respectively. As the initial guess we take

$$u_{ij}^{(0)} = 0, \quad i,j = 1,...,n-1 \qquad \text{and} \qquad r_{ij}^{(0)} = 1, \quad i,j = 1,...,n-1.$$

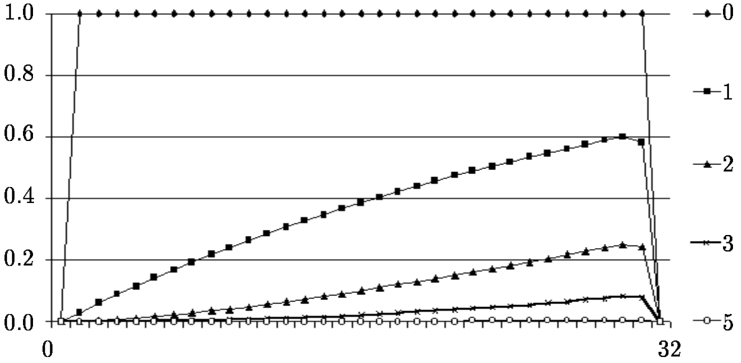


Fig. 10. The error of the block Gauss-Seidel method $r^{(k)}_{i,0.5}$.

The use of the cascadic algorithm allows to improve further the convergence. With this approach we take the interpolation of the solution on a coarse grid as the initial guess on the finer grid with the halved mesh size.

Now we consider the construction of the interpolation from a coarse grid to a finer one.

In the numerical experiments we applied the linear interpolation in the y-direction

$$u^I(x_i, y_k^*) = \frac{y_j - y_k^*}{h_j} u_{i,j-1} + \frac{y_k^* - y_{j-1}}{h_j} u_{i,j} \tag{3.25}$$

where $y_k^* \in [y_{j-1}, y_j]$, $h_j = y_j - y_{j-1}$. Let us show that with this interpolation the order of accuracy holds when the nodes of the grid are defined by (3.9).

To do this, we rewrite (3.25) in the form

$$u^I(x_i, y_k^*) = \alpha u_{i,j-1} + (1-\alpha) u_{i,j}, \quad \alpha = (y_j - y_k^*)/h_j.$$

Using the Taylor expansion with the second-order reminder term for $u_{i,j-1}$ and $u_{i,j}$ about (x_i, y_k^*), we have

$$\begin{aligned} \left|u^I(x_i, y_k^*) - u(x_i, y_k^*)\right| &\le |\alpha u_{i,j-1} + (1-\alpha) u_{i,j} - u(x_i, y_k^*)| \\ &\le \frac{1}{2}\left(\alpha (y_k^* - y_{j-1})^2 + (1-\alpha)(y_j - y_k^*)^2\right) |\partial_{22} u(x_i, y_k^*)| \\ &\le c_1 h_j^2 |\partial_{22} u(x_i, y_k^*)| \le c_2 h^2. \end{aligned} \tag{3.26}$$

Here we used the estimate (2.133) from Lemma 24 and the definition (3.9) of the grid.

Note that the estimate (3.26) for the Shishkin grid has the form [58]

$$\left|u^I(x_i, y_k^*) - u(x_i, y_k^*)\right| \le c_3 h^2 \ln^2 (1/h).$$

In the x-direction the grid is uniform. With decreasing the mesh size from $2h$ to h, we transform the grid equations (3.13) to determine the values of $u(x_i, y_j)$ for $i = 2m - 1$, $m = 1, 2, ..., [n/2]$, $j = 1, 2, ..., n - 1$:

$$-d_{ij}u^I_{i,j-1} + b_{ij}u^I_{ij} - c_{ij}u^I_{i,j+1} = f_{ij} + a_{ij}u_{i-1,j} + c_{ij}u_{i+1,j}.$$

The values of $u_{i-1,j}$ and $u_{i+1,j}$ at each level $i = 2k - 1$ are known from the previous grid. Therefore to determine the values of $u(x_{2m-1}, y_j)$ at each level m one has to solve the system of linear algebraic equations with the tridiagonal matrix B_{2k-1}, $m = 1, 2, ..., [n/2]$. We show that in this case the order of convergence also holds.

Table 7. The number of iteration step in the Gauss-Seidel method.

n	Gauss-Seidel method					
	pointwise		block		cascadic algorithm	
	Bakhvalov	Shishkin	Bakhvalov	Shishkin	Bakhvalov	Shishkin
32	28	15	2	2	2	2
64	97	31	2	2	2	2
128	355	70	3	3	2	2
256	1307	178	5	5	4	4
512	*	896	19	19	14	15
1024	*	*	151	152	124	125
2048	*	*	1052	1049	877	891
* – convergence was not achieved after 2500 iteration steps						

Consider the error

$$\delta_{ij} = \left|u_{ij} - u^I_{ij}\right|, \quad i = 2m - 1, \quad m = 1, 2, ..., [n/2], \quad j = 1, 2, ..., n - 1.$$

It satisfies the system of equations

$$-d_{ij}\delta_{i,j-1} + b_{ij}\delta_{ij} - e_{ij}\delta_{i,j+1} = \theta_{ij}, \quad |\theta_{ij}| \le c_5 h^2$$

for $i = 2m - 1$, $m = 1, 2, ..., [n/2]$, $j = 1, 2, ..., n - 1$. Then we have

$$(a_{ij} + c_{ij}) \max_{i,j=1,2,\dots,n-1} |\delta_{ij}| \le \max_{i,j=1,2,\dots,n-1} |\theta_{ij}|.$$

Taking into account the definitions of a_{ij} and c_{ij} we get

$$\max_{i,j=1,2,\dots,n-1} |\delta_{ij}| \le c_6 h^2.$$

The number of iteration step that is required to achieve the convergence criterion is shown in Table 7 for $\varepsilon = 1/2560$ on $n \times n$ grids. As the convergence criterion we used the following restriction on the residual after k

iteration steps:

$$\max_{i,j=1,2,\dots,n-1} \left| \left(L^h u^{(k)} \right)_{ij} - f_{ij} \right| \leq \Delta^h .$$

We put

$$\Delta^h = 10^{-5} H^2 \cdot 2^{1-H/h}$$

where $1/H$ is the nodes of the coarser grid.

3.5 Discussion of the numerical results

We write the solution of the problem (3.3)–(3.4) as the series

$$u = \sum_{n=1}^{\infty} \gamma_n \psi_n(x) \sin(\pi n y) = S_{sol} \tag{3.27}$$

where

$$\psi_n(x) = C_{1n} \exp(\lambda_1^n x) + C_{2n} \exp(\lambda_2^n x) - 1, \tag{3.28}$$

$$C_{1n} = \frac{\exp(\lambda_2^n) - 1}{\exp(\lambda_2^n) - \exp(\lambda_1^n)}, \quad C_{2n} = \frac{1 - \exp(\lambda_1^n)}{\exp(\lambda_2^n) - \exp(\lambda_1^n)}, \tag{3.29}$$

$$\lambda_1^n = \frac{1 + \sqrt{1 + (2\varepsilon\pi n)^2}}{2\varepsilon}, \quad \lambda_2^n = \frac{1 - \sqrt{1 + (2\varepsilon\pi n)^2}}{2\varepsilon}, \tag{3.30}$$

$$\gamma_n = \begin{cases} 0, & \text{if } n \text{ is even}, \\ -\dfrac{4}{\varepsilon(\pi n)^3}, & \text{if } n \text{ is odd}. \end{cases} \tag{3.31}$$

Lemma 31. *The series* (3.27) *converges uniformly for* $x \in [0,1]$.

Proof. Consider the sequence of the functions $\{\psi_n(x)\}_{n=1}^{\infty}$ and show that it is uniformly bounded on $x \in [0,1]$.

Let us calculate the derivatives $\psi_n'(x)$, $\psi_n''(x)$:

$$\psi_n'(x) = \lambda_1^n C_{1n} \exp(\lambda_1^n x) + \lambda_2^n C_{2n} \exp(\lambda_2^n x),$$

$$\psi_n''(x) = \lambda_1^{n2} C_{1n} \exp(\lambda_1^n x) + \lambda_2^{n2} C_{2n} \exp(\lambda_2^n x).$$

Because of (3.29) and (3.30) the following inequalities hold:

$$\begin{array}{lll} \lambda_1^n \geq 0, & \lambda_2^n \leq 0 & \forall\, n = 1, 2, \dots , \\ C_{1n} \leq 0, & C_{2n} \leq 0 & \forall\, n = 1, 2, \dots . \end{array}$$

Since

$$\psi''_n(x) \geq 0 \qquad \forall\, x \in [0,1] \quad \forall\, n = 1, 2, \dots\,,$$

$\psi_n(x)$ is convex function on $[0,1]$. At the point of maximum the equality

$$\psi'_n(x_0) = \lambda_1^n C_{1n} \exp(\lambda_1^n x_0) + \lambda_2^n C_{2n} \exp(\lambda_2^n x_0) = 0$$

is valid. Then we have

$$\exp(x_0) = \left(\frac{\lambda_2^n}{\lambda_1^n} \frac{\exp(\lambda_1^n) - 1}{\exp(\lambda_2^n) - 1} \right)^{1/(\lambda_1^n - \lambda_2^n)}.$$

Calculate $\lim\limits_{n\to\infty} |\psi_n(x_0)|$. Let n be sufficiently large, for example, $2\varepsilon\pi n >> 1$, then $\lambda_1^n \approx \pi n$, and $\lambda_2^n \approx -\pi n$. It is easy to calculate that

$$\exp(x_0) = \left(-\frac{\exp(\pi n) - 1}{\exp(-\pi n) - 1} \right)^{1/2\pi n} = \exp\left(1/2\right).$$

Then we get

$$\begin{aligned} \psi_n(x_0) &= \frac{1}{\exp(\pi n) + 1} \exp(\pi n/2) + \frac{\exp(\pi n)}{\exp(\pi n) + 1} \exp(-\pi n/2) - 1 \\ &= \frac{2}{\exp(\pi n/2) + \exp(-\pi n/2)} - 1. \end{aligned}$$

This yields

$$\lim_{n\to\infty} |\psi_n(x_0)| = 1. \tag{3.32}$$

Thus, the sequence $\{\psi_n(x)\}_{n=1}^{\infty}$ is uniformly bounded on $[0,1]$, in other words, there exists such a constant M that

$$|\psi_n(x)| \leq M \qquad \forall\, x \in [0,1] \quad \forall\, n = 1, 2, \dots\,.$$

The sequence of the functions $\{\sin(\pi n y)\}_{n=1}^{\infty}$ is uniformly bounded on $[0,1]$ by 1.

Therefore, the terms of the series (3.27) satisfy the inequality

$$|\gamma_n \psi_n(x) \sin(\pi n y)| \leq M\gamma_n, \quad \forall\, n = 1, 2, \dots \tag{3.33}$$

where γ_n are the terms of the convergent series

$$S = -\frac{4M}{\varepsilon\pi^3} \sum_{k=1}^{\infty} \frac{1}{(2k-1)^3} \tag{3.34}$$

Table 8. The error $R^{n,K}_{abs}$ for $\varepsilon = 10^{-3}$.

n	Bakhvalov grids			Shishkin grids		
	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$
32	7.160$_{10}$-3	7.160$_{10}$-3	7.160$_{10}$-3	7.224$_{10}$-3	7.223$_{10}$-3	7.223$_{10}$-3
64	1.545$_{10}$-3	1.546$_{10}$-3	1.546$_{10}$-3	2.993$_{10}$-3	2.992$_{10}$-3	2.992$_{10}$-3
128	6.824$_{10}$-4	6.812$_{10}$-4	6.812$_{10}$-4	1.165$_{10}$-3	1.164$_{10}$-3	1.164$_{10}$-3
256	2.471$_{10}$-4	2.460$_{10}$-4	2.459$_{10}$-4	4.013$_{10}$-4	4.008$_{10}$-4	4.009$_{10}$-4
512	7.389$_{10}$-5	7.200$_{10}$-5	7.188$_{10}$-5	1.209$_{10}$-4	1.203$_{10}$-4	1.202$_{10}$-4
1024	2.819$_{10}$-5	1.909$_{10}$-5	1.889$_{10}$-5	3.634$_{10}$-5	3.379$_{10}$-5	3.374$_{10}$-5

Table 9. The error $R^{n,K}_{abs}$ for $\varepsilon = 10^{-2}$.

n	Bakhvalov grids			Shishkin grids		
	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$
32	1.755$_{10}$-3	1.755$_{10}$-3	1.755$_{10}$-3	3.629$_{10}$-3	3.629$_{10}$-3	3.629$_{10}$-3
64	4.896$_{10}$-4	4.896$_{10}$-4	4.896$_{10}$-4	9.579$_{10}$-4	9.579$_{10}$-4	9.579$_{10}$-4
128	1.254$_{10}$-4	1.254$_{10}$-4	1.254$_{10}$-4	2.435$_{10}$-4	2.435$_{10}$-4	2.435$_{10}$-4
256	3.158$_{10}$-5	3.159$_{10}$-5	3.159$_{10}$-5	6.113$_{10}$-5	6.111$_{10}$-5	6.111$_{10}$-5
512	7.916$_{10}$-6	7.919$_{10}$-6	7.920$_{10}$-6	1.531$_{10}$-5	1.530$_{10}$-5	1.530$_{10}$-5
1024	2.614$_{10}$-6	2.120$_{10}$-6	2.120$_{10}$-6	3.987$_{10}$-6	3.977$_{10}$-6	3.977$_{10}$-6

Table 10. The error $R^{n,K}_{abs}$ on the grid (3.9).

n	$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-3}$		
	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$	$R^{n,1000}_{abs}$	$R^{n,2000}_{abs}$	$R^{n,3000}_{abs}$
32	1.54$_{10}$-3	1.54$_{10}$-3	1.54$_{10}$-3	3.22$_{10}$-3	3.22$_{10}$-3	3.22$_{10}$-3
64	4.46$_{10}$-4	4.46$_{10}$-4	4.46$_{10}$-4	1.57$_{10}$-3	1.57$_{10}$-3	1.57$_{10}$-3
128	1.16$_{10}$-4	1.16$_{10}$-4	1.16$_{10}$-4	6.85$_{10}$-4	6.85$_{10}$-4	6.85$_{10}$-4
256	2.95$_{10}$-5	2.95$_{10}$-5	2.95$_{10}$-5	2.48$_{10}$-4	2.46$_{10}$-4	2.46$_{10}$-4
512	7.43$_{10}$-6	7.39$_{10}$-6	7.39$_{10}$-6	7.37$_{10}$-5	7.20$_{10}$-5	7.19$_{10}$-5
1024	2.55$_{10}$-6	1.85$_{10}$-6	1.85$_{10}$-6	2.83$_{10}$-5	1.90$_{10}$-5	1.89$_{10}$-5

of numbers. Then according to the Weierstrass criterion of the uniform convergence of functional series, the series (3.27) uniformly converges.□

The estimate (3.33) shows that the series (3.27) converges at least as (3.34). We denote by S_K the partial sum of (3.34). The following estimate holds (see [54]):

$$|S - S_K| \leq \frac{4M}{\varepsilon\pi^3}\,\frac{1}{K^2}.$$

Therefore to achieve the given accuracy δ it is necessary to take at most K terms where

$$K = 2\sqrt{\frac{M}{\varepsilon\pi^3}\delta}.$$

From (3.32) we have that the constant M is close to 1.

In the numerical experiments the series was calculated within an accuracy $\delta = 10^{-5}$. The exact solution was calculated as the partial sums S_{sol}^{1000}, S_{sol}^{2000}, and S_{sol}^{3000}. In Tables 8, 9, and Figures 11, 12 the numerical results are presented on the sequence of grids for $\varepsilon = 10^{-3},\ 10^{-2}$. We use the notations

$$R_{abs}^{n,K} = \max_{i,j=0,\dots,n} \left| u_{ij}^n - S_{sol}^K(x_i, y_j) \right|.$$

Here u_{ij}^n is the solution of the discrete problem at the node (x_i, y_j) of the $(n+1) \times (n+1)$ grid, K is the number of the terms of the series. In the Figures 11, 12 the values of $R_{abs}^{n,3000}(n)$ are marked by the numbers 2, 3, and 4 for the Shishkin, the Bakhvalov grids and the grid (3.9) respectively. For comparison the straight lines with slapes $\mathrm{tg}\varphi = 1$ and $\mathrm{tg}\varphi = 2$ marked by 1 are shown in Figures 12 and 11 respectively. For $\varepsilon = 10^{-2}$ the method has the second-order convergence. When ε decreases to 10^{-3}, the method becomes first-order convergent.

Finally, we discuss the results obtained in the two-dimensional case with the special approximation of the right-hand side similar to that considered in Chapter 1. We considered the Dirichlet problem

$$\begin{aligned} -\varepsilon\Delta u + \partial_1((1+2x)u) &= f \quad \text{in} \quad \Omega, \\ u &= 0 \quad \text{on} \quad \Gamma \end{aligned}$$

where

$$f = 6x^2 + 2x - 2\varepsilon + 2d, \quad d = \frac{\exp(-2/\varepsilon)}{1 - \exp(-2/\varepsilon)}.$$

The solution of this problem has the parabolic boundary layer near the boundary Γ_{tg} and the regular one near Γ_{out}.

Table 11 contains the results obtained on the Bakhvalov grid with the fitted quadrature rule with the special and standard approximations of the

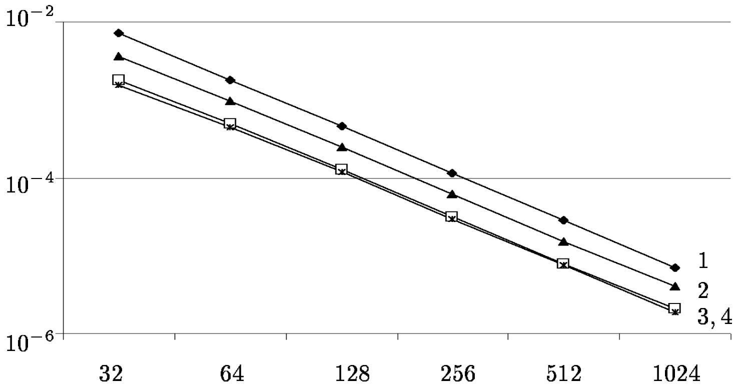


Fig. 11. The error $R_{abs}^{n,3000}(n)$ for $\varepsilon = 10^{-2}$.

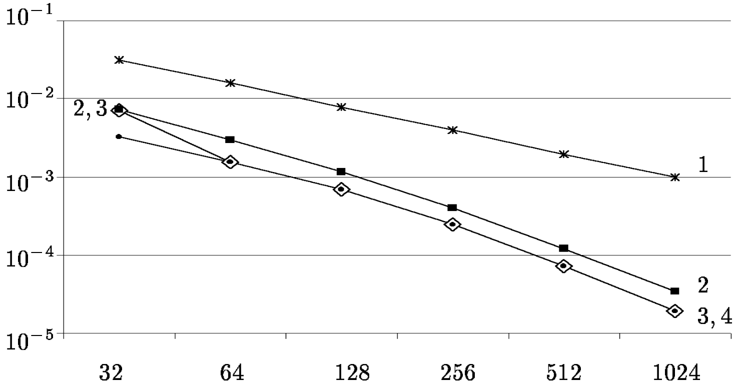


Fig. 12. The error $R_{abs}^{n,3000}(n)$ for $\varepsilon = 10^{-3}$.

Table 11. The error R_{abs}^{n} for standard and special approximations of the right-hand side.

approximation of the right-hand side	n					
	32	64	128	256	512	1024
standard	3.99_{10}-2	1.99_{10}-2	9.68_{10}-3	4.52_{10}-3	1.93_{10}-3	6.27_{10}-4
special	4.32_{10}-3	2.65_{10}-3	1.36_{10}-3	6.26_{10}-4	2.33_{10}-4	1.40_{10}-4

right-hand side for $\varepsilon = 1/2560$. The results demonstrate that the application of the special quadrature rule for the approximation of the right-hand side improves the accuracy.

References

1. Bagaev B.M.: *The Galerkin Method for Ordinary Differential Equation with a Small Parameter.* In: Numerical Methods of Mechanics of Continua, Novosibirsk: Computing Center of Siberian Section of Academy of Science USSR, 1979, vol. 10, № 1, pp. 1–16 (In Russian).
2. Bagaev B.M.: *The Method of Galerkin for Equation with a Small Parameter Affecting the Highest Derivative.* In: The Methods of Approximation and Interpolation, Novosibirsk: Computing Center of Siberian Section of Academy of Science USSR, 1981, pp. 4–13 (In Russian).
3. Bagaev B.M., Shaidurov V.V.: *The Variation - Difference Solution of Equation with a Small Parameter.* In: The proceeding of Computing Center of Siberian Section of Academy of Science USSR, Novosibirsk, 1977, pp. 89–99 (In Russian).
4. Bagaev B.M., Shaidurov V.V.: *The Grid Methods for Solving of Problems with Boundary Layer: In 5 parts.* Novosibirsk: Nauka, 1973, Part 1, 199 p. (In Russian).
5. Bakhvalov N.S.: *On the Optimization of Methods for Solving Boundary Value Problems with Boundary Layers.* U.S.S.R. Comput. Maths. Maths. Phys., 1969, vol. 9, № 4, pp. 841–859 (In Russian).
6. Boglaev I.P.: *Approach Solution of Non-linear Boundary Value Problem with a Small Parameter Affecting the Highest Derivative.* U.S.S.R. Comput. Maths. Maths. Phys., 1984, vol. 24, № 11 pp. 1649–1656 (In Russian).
7. Boglaev I.P.: *On the numerical integration of Kauchy Singularly Perturbed Problem for for Ordinary Differential Equation.* U.S.S.R. Comput. Maths. Maths. Phys., 1985, vol. 25, № 7, pp. 1009–1022 (In Russian).
8. Boglaev U.P.: *Iterative Sweep Method of Approximate Solution of Singularly Perturbed Non-linear Boundary Value Problems.* Dokl. of Academy of Science USSR, 1977, vol. 228, № 6, pp. 1241–1244 (In Russian).
9. Boglaev U.P.: *Iterative Method of Approximate Solution of Singularly Perturbed Problems.* Dokl. of Academy of USSR, 1976, vol. 227, № 5, pp. 1009–1022 (In Russian).
10. Boglaev U.P.: *About Numerical Methods for Solution of Singularly Perturbed Problems.* Diff. Eq., 1985, vol. XXI, № 10, pp. 1804–1806 (In Russian).
11. Butuzov V.F.: *On Asymptotic of Solutions of Singularly Perturbed Equations of Elliptic in rectangle.* Diff. Eq., 1975, vol. XI, № 6, pp. 1030–1041 (In Russian).
12. Butuzov V.F.: *On Obtaining of Boundary Functions for Some Singularly Perturbed Problems of Elliptic.* Diff. Eq., 1977, vol. XIII, № 10, pp. 1829–1835 (In Russian).

13. Butuzov V.F., Nikitin A.G.: *The Singularly Perturbed Boundary Value Problems for Elliptic Equations in rectangle in the critical case.* J. Numer. Meth. Math. Phis., 1984, vol.24, № 9, pp. 1320–1330 (In Russian).
14. Van Dyke M.: *Perturbation Methods in Fluid Mechanics.* Moskow: Mir, 1967 (In Russian).
15. Vasil'eva A.B.: *The Asymptotic of Solution for an Ordinary Nonlinear Differential Problems with a Small Parameter Affecting the Highest Derivatives.* Uspehi Math. Nauk, 1963, vol. 18, № 3, pp. 15–86 (In Russian).
16. Vasil'eva A.B., Butuzov V.F.: *The Asymptotic expansions of Solutions for the Singularly Perturbed Equations.* Moskow: Nauka, 1973 (In Russian).
17. Vasil'eva A.B., Butuzov V.F.: *The Singularly Perturbed Equations in the critical case.* Moskow: Moskow State University Press, 1978 (In Russian).
18. Vasil'eva A.B., Butuzov V.F.: *The Asymptotic Methods in the Theory of Singularly Perturbations.* Moskow: Vysshaya shkola, 1990(In Russian).
19. Vishik M.I., Lusternick L.A.: *Regular Degeneration and Boundary Layer for Linear Differential Equations with a Small Parameter.* Uspehi Math. Nauk, 1957, vol. 12, № 5(77), pp. 3–122 (In Russian).
20. Vladimirov V.S.: *The Equations of Mathematical Physic.* Moskow: Nauka, 1976 (In Russian).
21. Voevodin V.V., Kuznetsov U.A.: *Matrixes and Calculations.* Moskow: Nauka, 1984 (In Russian).
22. Gushchin V.A., Shchennikov V.V.: *On a Monotone Difference Scheme of second order.* U.S.S.R. Comput. Maths. Maths. Phys., 1974, vol. 14, № 3, pp. 789–792 (In Russian).
23. Doolan E.P., Miller J.J.H., Schilders W.H.A.: *Uniform numerical methods for problems with initial and boundary layers.* Moscow, Mir, 1983, (In Russian).
24. Emel'janov K.V.: *On a Difference Scheme for a Differential Equation with a Small Parameter Affecting the Highest Derivatives.* Numer. Meth. of Continuum Mechanics, 1970, vol. 1, № 5, pp. 20–30 (In Russian).
25. Il'in A.M.: *Difference Scheme for Differential Equation with a Small Parameter at the Highest Derivatives.* Mat. Zametki, 1969, vol. 6, № 2, pp. 237–248 (In Russian).
26. Il'in A.M.: *Agreement of Asymptotic Expansions of Solution of Boundary Value Problems.* Moskow: Nauka, 1989 (In Russian),
27. Karepova E.D., Shaidurov V.V.: *The Algebraic Fitted in the Finite Elements Method for Reaction-Diffusion Problems with a Small Parameter.* Dep in VINITI, Krasnoyarsk, 1996, № 2951-B96, 21 p. (In Russian).
28. Karepova E.D., Shaidurov V.V.: *The Finite Elements Method for Ordinary Differential Equation with a Small Parameter.* Dep. in VINITI, Krasnoyarsk, 1997, № 1252-B97, 20 p. (In Russian).
29. Karepova E.D., Shaidurov V.V.: *The Finite Elements Method with Fitted Quadrature Formula for the Convection-Diffusion Equation.* Dep. in VINITI, Krasnoyarsk, 1998, № 415-B98, 22 p. (In Russian).
30. Karepova E.D., Shaidurov V.V.: *Algebraic fitting in the finite element method for the small parameter reaction-diffusion problem.* Advances in Modeling &

Analysis, Ser. A: Mathematical Problems General Mathematical Modeling, A.M.S.E., France, 1999, vol. 36, pp. 37–54.

31. Karepova E.D., Shaidurov V.V.: *Finite Element Method with Fitted Integration Rule for Convection-Diffusion Problem.* Russian J. of Numerical Analysis, 2000, vol. 15, № 12, pp. 167–182 (In Russian).
32. Karepova E.D.: *Finite Element Method with Fitted Integration Rules for Convection-Diffusion Equation with Small Diffusion.* Workshop'98 on the Analitical and Computational Methods for Convection-Dominated and Singular Pertrubed Problems, Bulgaria, 1998, p.15.
33. Koul G.: *The Perturbation Method in Applied Mathematics.* Moscow, Mir, 1972 (In Russian).
34. Ladyzhenskaja O.A., Solonnikov V.A., Ural'ceva N.N.: *Linear and Qusilinear Equations of Parabolic Type.* American Mathematical Society, Providence, 1968.
35. Ladyzhenskaja O.A., Ural'ceva N.N.: *Linear and Qusilinear Equations of Elliptic Type.* Moscow, Nauka, 1973 (In Russian).
36. Liseikin V.D., Petrenko V.E.: *An Adaptive-Invariant Method for the Numerical Solution of Problems with Boundary and interior layers.* Novosibirsk, Computer Center of the Russian Academy of Science, 1989 (In Russian).
37. Liseikin V.D.: *The review of the methods of construction of structural adaptive grids.* U.S.S.R. Comput. Maths. Maths. Phys., 1996, vol. 36, № 1, pp. 3–42 (In Russian).
38. Lomov S.A.: *The Introduction into the Theory of Singularly Perturbations.* Moscow, Nauka, 1981 (In Russian).
39. Marchuk G.I.: *The Methods of Numerical Mathematics.* Moscow, Nauka, 1977 (In Russian).
40. Marchuk G.I., Agoshkov V.I.: *The Introduction into Projective-Discrete Methods.* Moscow, Nauka, 1981 (In Russian).
41. Marchuk G.I., Shaidurov V.V.: *The Increase of Accuracy of Difference Schemes.* Moscow, Nauka, 1979 (In Russian).
42. Moiseev N.N.: *The Asymptotic Methods of Nonlinear Mechanics.* Moscow, Nauka, 1981 (In Russian),
43. Naife A.H.: *The Perturbation Methods.* Moscow, Mir, 1976 (In Russian)
44. Naife A.H.: *The Introduction into Perturbation Methods.* Moscow, Mir, 1984 (In Russian).
45. Oleinik O.A.: *On Elliptic Equations with a Small Parameter in the Highest Derivatives.* Mat. Sbornik, 1952, vol. 31, № 1, pp. 104-117 (In Russian).
46. Samarski A.A.: *The Introduction into the Theory of Defference Schemes.* Moscow, Nauka, 1971 (In Russian).
47. Samarski A.A., Nickolaev E.S.: *The Methods of Resolving of the Greed Equations.* Moscow, Nauka, 1978 (In Russian).
48. Ciarlet F.: *The Finite Elements Method for Elliptic Problems.* Amsterdam–New York–Oxford, North-Holland Publishing Company, 1978
49. Tihonov A.N.: *On the Dependence of Solutions of Differential Equations on a Small Parameter.* Mat. Sbornik, 1948, vol. 22(64), № 2, pp. 193–204 (In Russian).

50. Tihonov A.N.: *On the systems of Differential Equations With Parameters.* Mat. Sbornik, 1950, vol. 27(69), № 1, pp. 147–156 (In Russian).
51. Tihonov A.N.: *The systems of Differential Equations With Parameters.* Mat. Sbornik, 1952, vol. 31(73), № 3, pp. 575–586 (In Russian).
52. Tihonov A.N., Samarski A.A.: *The Equation Of Mathematical Physics.* Moscow, Nauka, 1972 (In Russian).
53. Trenogin V.A.: *The Development and Aplication of Asymptotic Method by Vishik and Lusternik.* Uspehi Mat. Nauk, 1970, vol. 25, № 4, pp. 123–156 (In Russian).
54. Fihtengol'tc G.M.: *The Course of Differential and Integration Calculus.* Moscow, Nauka, 1961, vol. 2 (In Russian).
55. Shishkin G.I.: *The Numerical Solution of Elliptic Equations with a Small Parameter in the Highest Derivatives.* Doklady Acad. Nauk SSSR, 1979, vol.245, № 4, pp. 804–808 (In Russian).
56. Shishkin G.I.: *A Difference Scheme on a Non-Uniform Mesh for a Differential Equation with a Small Parameter in the Highest Derivative.* U.S.S.R. Comput. Maths. Maths. Phys., 1983, vol. 23, № 1, pp. 59–66 (In Russian).
57. Shishkin G.I.: *Approximation of the Solutions of Singularly Perturbed Boundary Value Problem with Parabolic Boundary Layer.* U.S.S.R. Comput. Maths. Maths. Phys., 1989, vol. 29, № 7, pp. 1–10 (In Russian).
58. Shishkin G.I.: *Discrete Approximation of Singularly Perturbed Elliptic and Parabolic Equations.* Russian Academy of Sciences, Ural Section, Ekaterinburg, 1992 (In Russian).
59. Abrahamsson L., Keller H.B. Kreiss H. O.: *Difference approximations for singular perturbations of systems of ordinary differential equations.* Numer. Math., 1974, № 22, pp. 367–391 (In Russian).
60. Allen D.N. de G., Southwell R.V.: *Relaxation methods applied to determine the motion, in 2D, of a viscous fluid post a fixed cylinder.* Quart J. Mech. Appl. Math., 1955, vol. VIII, № 2, pp. 129–145.
61. Angermann L.: *Numerical solution of second order elliptic equations on plane domains.* Math. Modelling Anal. Numer., 1991, vol. 25, № 2, pp. 169–191.
62. Angermann L.: *Pseudouniform in ε-convergence of an elliptic Singularly Perturbed problem.* Uni. Erlagen, Inst. f. Angew. Math., 1991.
63. Babuska I., Aziz A.K.: *Survey lectures on the mathematical foundation of the finite element method.* In: The mathematical foundation of the finite element method with applications to partial differential equations, New York, Academic Press, 1972, pp. 1–362.
64. Babuska I., Rheinboldt W.C.: *Error estimates for adaptive finite element computation.* SIAM J. Numer. Anal., 1978, vol. 15, pp. 736–754.
65. Babuska I., Szymezak W.G.: *Adaptivity and error estimation for the finite element method applied to convection-diffusion problems.* SIAM J. Numer. Anal., 1984, vol. 21, pp. 910–946.
66. Babuska I., Miller A.: *A feedback finite element method with a posteori error estimation.* J. Comp. Meth. Engrg., 1987, vol. 61, pp. 1–40.
67. Babuska I., Suri M.: *On lacking and robustness in the finite element method.* SIAM J. Numer. Anal., 1992, vol. 44, pp. 283–301.

68. Bank R.E., Weiser A.: *Some a posteriori error estimators for elliptic partial differential equations.* Math. Comp., 1985, vol. 44, pp. 283–301.
69. Barret J.W., Morton K.W.: *Optimal finite element solutions to diffusion-convection problems in one dimension.* Int. J. Numer. Meth. Eng., 1980, vol. 15, pp. 1457–1474.
70. Berger A.E., Solomon J.M., Ciment M.: *Analysis of a uniformly accurate difference method for a singular perturbation problem.* Math. Comp., 1980, № 151, pp. 695–731.
71. Brezzi F., Russo A.: *Choosing bubbles for advection-diffusion problems.* Math. Models and Meth. in Appl. Sciences, 1994, vol. 4, pp. 571–587.
72. Bristeau M.O., Glovinski R., Periaux J., Pironneau O.: *On the numerical solution of nonlinear problems in fluid dynamics least squares and finite element methods.* In: Proseedings of FENOMECH78, Comp. Math. Appl. Mech. Eng., 1978, vol. 3, № 17/18, pp. 619–657.
73. Brooks A.N., Hughes T.J.R.: *Streamline upwind Petrov-Galerkin methods for advection dominaated flowss.* In Proceedings: Third International Conference on Finite Element Methods in Fluid Flow, Canada, Banff, 1980, pp. 645–680.
74. Ciarlet P.: *The finite element method for elliptic problem.* North-Holland, Amsterdam, 1978.
75. Christic I., Griffiths D.F., Mitchell A.R., Zienkiewicz O.C.: *Finite element methods for second order differential equations with significan first derivatives.* Int. J. Numer. Meth. Eng., 1976, vol. 10, pp. 1389–1396.
76. Duran R., Muschietti M.A., Rodriguez R.: *On the asymptotic exactness of error estimators for linear triangular elements.* Numer. Math., 1991, V.59, pp.107–127.
77. Duran R., Muschietti M.A., Rodriguez R.: *Asymptotically exact error estimators for rectangular finite elements.* SIAM J. Numer. Anal., 1992, vol.29, pp.78–88.
78. Eriksson K., Johnson C.: *Adaptive streamline diffusion finite element methods for stationary convection diffusion problems.* Math. Comp., 1993, vol. 60, pp. 167–188.
79. Farrell P.A.: *Uniformly convergent difference schemes for singularly perturbed turning and non-turning point problems.* Ph. D. thesis, Dublin, Trinity College, 1983.
80. Farrell P.A.: *Sufficient conditions for the uniform convergence of difference schemes for singularly perturbed turning point problem.* SIAM J. Numer. Anal., 1988, № 25, pp. 618–643.
81. Farrell P.A., Hegarty A.F.: *Numerical rezalts for singularly perturbed linear and quasilinear differential equations using a coarse grid.* In Proc. BAILI Conf., Boundary and Inter. Layers Comput. and Asympt. Meth., 1980, pp. 275–280.
82. Farrel P.A., Miller J.J.H., O'Riordan E., Shishkin G.I.: *On the non-existence of ε-uniform finite difference methods on uniform meshes for semilinear two point boundary value problems.* Math. of Comp., 1998, vol. 67, № 222, pp. 603–617.

83. Fornberg B.: *Generation of finite difference formulas.* Math. Comp., 1988, № 51, pp. 702–705.
84. Franca L.P., Frey S.L., Hughes T.J.R.: *Stabilized finite element method: I. Application to the advective-diffusive model.* Comp. Meth. Appl. Mech. Engrg., 1992, vol. 95. pp. 253–276.
85. Franca L.P., Madureira A.L.: *Element free stability parameters for stabilized methods applied to fluids.* Comput. Meth. Appl. Mech. Eng., 1993, vol. 105, pp. 395–403.
86. Gartland E.C.: *Graded-mesh difference schemes for singularly perturbed two-point boundary value problems.* Math. Comp., 1988, vol. 51, № 184, pp. 631–657.
87. Gartland E.C.: *Strong uniform stability and exact discretizations of a model singular perturbation problem and its finite difference approximations.* Appl. Math. Comput., 1989, vol. 31, pp. 473–485.
88. H. Goering, A. Felgenhauer, G. Lube, H.-G. Roos, L. Tobiska.: *Singularly Perturbed Differential Equation.* Berlin, Acad.-Verlag., 1983.
89. Hangleiter R., Lube G.: *Boundary layer-adapted grids and domain decomposition in stabilized Galerkin methods for elliptic problems.* Amsterdam, CWI Quarterly, 1997, vol. 10, № 3& 4, pp. 215–238.
90. Hegarty A.F., Miller J.J.H., O'Riordan E., Shishkin G.I.: *Spetial meshes for finite difference approximations to an advection-diffusion equation with parabolic layers.* J. of Comp. Physics., 1995, vol. 117, pp. 47–54.
91. Hegarty A.F., Miller J.J.H., O'Riordan E., Shishkin G.I.: *Numerical Results for advection-dominated heat transfer in a moving fluid with a non-slip boundary condition. – Int. J. Num. Meth. Heat Fluid Flow.* 1995, vol. 5, pp. 131–140.
92. Hegarty A.F., Miller J.J.H., O'Riordan E., Shishkin G.I.: *On a novel mesh the regular boundary layers arising in advection-dominanted transport in two dimansions.* Communication in Num. Meth. in Engineering., 1995, vol. 11, pp. 435–441.
93. Hemker P.W.: *A numerical study of stiff two-point boundary value problems.* Amsterdam, Mathematical Center, 1977.
94. Heinrich J.C., Zienkiewicz O.C.: *Quadratic finite elements schemes for two-dimensional convective transport problems.* Int. J. Numer. Meth. Engng., 1977, vol. 11, pp. 1831–1844.
95. Heinrich J.C., Huyokarh P.S., Zienkiewicz O.C., Mitchell A.R.: *An upwind finite element scheme for two-dimensional convective ransport problems.* Int. J. Numer. Meth. Engng., 1977, vol. 11, pp. 131–143.
96. Hughes T.J.R., Brooks A.N.: *A multidimansional upwind scheme with no crosswind diffusion.* In : Finite Element Methods Convection Dominated Flows, New York, American Society of Mechanical Engineers Press, 1980, vol. 34.
97. Hughes T.J.R., Shakib F.: *Computational Aerodynamics and the finite element method.* In : AIAA/AAS Astrodynemics Conf., Reno, 1988, pp. 35–48.

98. Hughes T.J.R., Franca L.P., Hulbart G.M.: *A new finite element formulation for computational fluid dynamics: the Galerkin/least-squares method for advective diffusive equations.* Comp. Meth. Appl. Mech. Engrg., 1989, vol. 73, pp. 173–189.
99. Johnson C.: *Streamline diffusion methods for problems in fluid mechanics.* In: Finite Element in Fluids, Chichester, 1986, pp. 251–261.
100. Johnson C.: *Numerical Solution of Partial Differential Equations by the Finite Element Method.* Cembridge, Cembridge Univ. Press, 1987.
101. Johnson C., Schatz A.H., Wahlbin L.B.: *Crosswind smear and poinwise errors in streamline diffusion finite element methods.* Math. Comp., 1987, vol. 49, pp. 25–38.
102. Kellogg R.B, Stynes M.: *Optimal approximatility of solutions of singular perturbed two-points boundary value problems.* SIAM J. Numer. Anal., 1997, vol. 34, № 5, pp. 1808–1816.
103. Kellogg R.B, Tsan A.: *Analysis of some difference approximations for a singular perturbation problem without turning points.* Math. Comp., 1978, vol. 32, pp. 1025–1039.
104. Kratsch F., Roos H.-G.: *Monotonieerhabtende upwind-schemata in zweidimensionolen fall.* ZAMM, 1992, vol. 72, pp. 201–208.
105. Ladyzhenskaya O.A., Ural'tseva N.N.: *Linear and Quasilinear Elliptic Equations.* New York, Academ. Press, 1968.
106. Leonard B.P.: *A stable and accurate convective modelling procedure based on quadratic upstream integration* Comp. Meth. in Appl. Mech. Eng., 1979, № 19, pp. 59–98.
107. Levinson N.: *The first boundary value problem for $\varepsilon\Delta u + A(x,y)u_x + B(x,y)u_y + C(x,y)u = D(x,y)$ for small ε.* Annal. of Math., 1950, vol. 51, № 2, pp. 428–445.
108. Mizukami A., Hughes T.J.R.: *Petrov-Galerkin finite element methods for convecting-dominated flows : an accurate upwinding Technique for Satisfying the Maximum Principle.* Comput. Meths. Appl. Mech. Engrg., 1985, vol. 50, pp. 181–193.
109. Morton K.W. Scotney B.W.: *Petrov-Galerkin methods and diffusion-convection problem in 2D.* In: The mathematics of finite elements and applications, ed. Whitheman, New York, Academic Press, 1985. pp. 343–366.
110. Morton K.W.: *Galerkin finite element methods and their generalizations.* In: The State of the Art in Numerical Analysis, Oxford, Clarendon Press, 1987, pp. 645–680.
111. Morton K.W. Murdoch T., Süli E.: *Optimal error estimation for Petrov-Galerkin methods and diffusion-convection problems in two dimansion.* Numer. Math., 1992, vol. 61, pp. 359–372.
112. Miller J.J.H., O'Riordan E., Shishkin G.I.: *Fitted Numerical Methods for Singular Perturbation Problems* Error Estimates in the Maximum Norm for Linear Problems in One and Two Dimensions.
113. Miller J.J.H., Wong S.: *A new non-conforming Petrov-Galerkin finite element methods with triangular elements for a singularly perturbed advection-diffusion problem.* IMA J. Numer. Anal., 1994, vol. 14, pp. 257–276.

114. Nikolova M., Axelsson O.: *Uniform in ε convergence of finite element method for convection-diffusion equation using a priori chosen meshes.* Amsterdam, CWI Quarterly, 1997, vol. 10, № 3& 4, pp. 253–276.
115. Protter M.H., Weinberger H.F.: *Maximum Principles in Differential Equations.* Berlin, Springer-Vertag, 1984.
116. O'Riordan E., Stynes M.: *A uniformly convergent diffrence scheme for elliptic singular perturbation problem.* In: Discretization Method of Singular Perturbation and Flow Problems, Magdeburg, Technical Univ. Otto von Guericke, 1989, pp. 48–55.
117. O'Riordan E., Stynes M.: *A globally uniformly convergent finite element method for a singular perturbation elliptic problem in two dimansionsn.* Math. Comp., 1991, vol. 57, pp. 47–62.
118. Roos H.-G., Stynes M., Tobiska L.: *Numerical methods for singular perturbed differential equations.* Berlin, Springer-Verlag, 1996.
119. Roos H.-G., Adam D., Felgenhauer A.: *A novel nonconforming Uniformly convergent finite element method in two dimensions.* J. of Math. Anal. and Appl., 1996, № 201, pp. 715–755.
120. Roos H.-G., Skalicky T.: *A comparison of the finite element mrthod on Shishkin and Gartland-type meshes for convection-diffusion problems.* Amsterdam, CWI Quarterly, 1997, vol. 10, № 3& 4, pp. 277–300.
121. Roos H.-G., Linβ T.: *Sufficient conditions for uniform convergence on the layer-adapted grids.* Techn. Univ. Dresden, 1998, Prepr. MATH-NM-13-1998.
122. Shaidurov V., Tobiska L.: *Special integration formulae for a convection–diffusion problem.* East–West J.Numer.Math., 1995, vol. 3, № 4, pp. 281–299.
123. Shaidurov V., Karepova E.: *Finite Element Method with Fitted In tegration Rules for Singulary Perturbed Problem.* ENUMATH-1997, Second European Conference on Numerical Mathematics and Advanced Applications, Germany, Heidelberg, 1997, pp. 223–224.
124. Shishkin G.I.: *On finite difference fitted schemes for singularly perturbed boundary value problems with parabolic boundary layer.* J. of Math. Anal. and Appl., 1997, № 208, pp. 181–204.
125. Stoyan G.: *Monotone difference schemes for convection-diffusion problems.* ZAMM, 1979, № 59, pp. 361–372.
126. Stynes M., Tobiska L.: *Error estimates and numerical experiments for streamline-diffusion type methods on arfitiory and Shiahkin meshes.* Amsterdam, CWI Quarterly, 1997, vol. 10, № 3& 4, pp. 337–352.
127. Tabata M.: *A finite element approximation corresponding to the upwind differencing.* Memories Numerical Mathematics, 1977, vol. 1, pp. 47–63.
128. Tobiska L.: *Diskretisierungsverfahren zur Lösung singulär gestörter Randwertprobleme.* ZAMM, 1983, № 63, pp. 115–123.
129. Verfürth R.: *A posteriori error estimation and adaptive mesh-refinement techniques.* J. Comput. Appl. Math., 1994, vol. 50, pp. 67–83.
130. Vulanovic R.: *On a numerical solution of a type of singularly perturbed boundary value problem by using a spetial discretization mesh.* Univ. u Novom Sadu Zb. Rad. Prirod., Math. Fak., Ser. Math., 1983, V. 13, pp. 187–201.

131. Vulanovic R.: *Non-equidistant generations of the Gushchin-Shennikov scheme.* ZAMM, 1987, vol. 67, pp. 625–632.

132. Vulanovic R.: *Non-equidistant finite difference methods for elliptic singular perturbation methods.* In: Computational methods for boundary and interior layers in several dimansions, Dublin, Boole Press, 1991, pp. 203–223.

133. Zhou G.: *Local L^2-error analysis of streamline diffusion FE-method for non-stationary hiperbolic systems.* Prepr. 94-07(SEB359), Univ. of Heidelberg, 1994.

134. Zhou G., Rannacher R.: *Pointwise Superconvergence of the streamline diffusion finite element method.* Prepr. 94-72(SEB359), Univ. of Heidelberg, 1994.

135. Zhou G.: *How accurate is the streamline diffusion finite element method.* Math. Comp., 1997, vol. 66, pp. 31–44.

Triangulation of two-dimensional multiply connected domain with concentration and rarefection of grid

Pyataev S.F.

Introduction

Wide use of the finite element method for solution of various kinds of problems raises the requirements to the level of automation of domain fragmentation. There are algorithms and programs allowing to construct uniform grids on simply connected domains [1-4]. The advantages of the algorithms are their universality with respect to the shape of boundary of the domain as well as the possibility to triangulate simply and multiply connected domains with concentration and rarefaction of grid; the latter is attained by division of the initial domain into a number of simply connected domains and fragmentation or consolidation of one-dimensional final elements on boundaries of some subdomains. An obvious disadvantage of the triangulation algorithms for simple connected domains when applying to multiply connected domains or concentration of the grid is a great amount of handwork: division of the domain into subdomains, fragmentation of each boundary, input of information, etc. The idea of triangulation algorithm for multiply connected domain with concentration of grid described in [5]. It avoids the necessity of division of the domain into a collection of subdomains and at the same time retains the disadvantage connected with hand fragmentation of each contour (with the exception of the simplest elements of the contours: linear regions and arcs). Except that, indistinctness of the introduced in the paper requirements with respect to the properties of a new node being constructed (proximity to previously constructed node, proximity to one-dimensional finite element, simultaneous proximity to the node and the element, etc.) makes the programming considerably more difficult and forces the user either do develop conditions for a new node being constructed or quite reject the algorithm.

For the purpose of constructing a completely automated process of triangulation of arbitrary two-dimensional multiply connected domains, an algorithm of fragmentation for arbitrary piecewise smooth closed boundary contours is developed in the present paper.

In the third section, on the basis of the scheme proposed in [4, 5] and representing a consecutive filling of domain with triangular elements, the process of triangulation of the domain is constructed. The process of filling starts from the boundary which is preliminarily fragmented into one-dimensional finite elements. In the course of construction of triangular elements the boundary of the domain being not yet triangulated (following [4], we will call it *current grid boundary*, CGB) represents a number of continuous closed piecewise curves with possible self-intersections. A detailed description of construction of new nodes and elements is given in this section; in particular, the criteria of selection of previously constructed node (or construction of a new one) are given. And as a consequence, the criteria of construction of an element are given in the case when some regions of CGB close in. In the course of fragmentation of the boundary of domain and its triangulation a function of steps is used which adjusts the sizes of one-dimensional and triangular finite elements according to their position in the domain. Any positive function can appear as the function of steps; the principles of its construction are given in section 2.

Presentation of both the algorithms is given in a form convenient for programming. In appendices some auxiliary procedures are given, which are necessary for the work of the program and which, apparently, should be designed as subroutines.

1 Some recommendations on choice of the function of steps

In many problems one can beforehand make certain assumptions on subdomains of large gradients of the sought functions, appearing, as a rule, in the locations of concentrators of different kind, on lines of jump of coefficients of the problem, due to singularities in some points of boundary conditions, in the points of sharp change in the character of the boundary, etc. For concentration of the grid in such subdomains a necessity appears to construct finite elements with a step less than the basic step h_0 used for larger part of the domain Ω. Since during triangulation the triangular elements are constructed successively, their size can be determined according to their locations, by means of certain positive function of steps $h(x, y)$ with parameters responsible for the "centers" and "sizes" of the subdomains of concentration. These parameters should be chosen so that on leaving the subdomain the sizes of triangle elements would be of the order h_0.

Apparently, exact recommendations on construction of the function of steps cannot be given owing to the absence of exact definition of the notion of the domain of concentration. Therefore let restrict ourselves to formulation of general principles of construction of these functions, extending descriptive ideas of one-dimensional case to two-dimensional one. Let in one-dimensional case a qualitative graph of the function of steps is represented on Fig. 1.

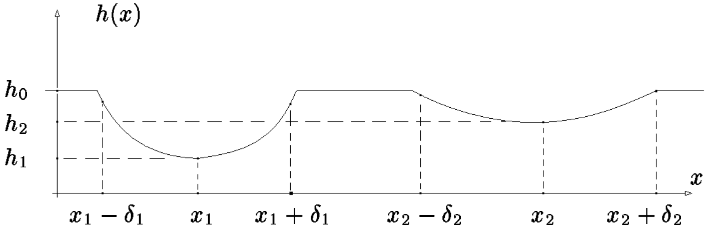


Fig. 1. An example of graph of the function of steps.

It is convenient to represent the function of steps in the form of a sum

$$h(x) = h_0 + \sum_{i=1}^{n} (h_i - h_0) f_i(x, x_i, \delta_i) ,$$

where δ_i is "characteristic size" of the i–th domain of concentration; x_i is center of domain of concentration; f_i is equal to 1 in the point x_i is of zero order outside its domain of concentration. In this case an approximate graph of the function f_i can be represented like on Fig. 2.

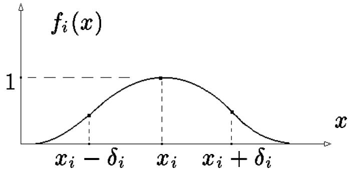


Fig. 2. An approximate graph of the function $f_i(x)$.

So, one can take $f_i(x)$ as one of the variants:

$$\mathrm{ch}^{-n_i}\left(\frac{x - x_i}{\delta_i}\right) \; ; \quad \left(1 + \left|\frac{x - x_i}{\delta_i}\right|^{n_i}\right)^{-1} \; ; \quad \exp\left\{-\left|\frac{x - x_i}{\delta_i}\right|^{n_i}\right\} ;$$

or

$$f_i(x) = \begin{cases} 1 - \left| \dfrac{x - x_i}{\delta_i} \right|^{n_i}, & \text{if } x \in (x_i - \delta_i, x_i + \delta_i)\ , \\ 0, & \text{if } x \notin (x_i - \delta_i, x_i + \delta_i). \end{cases}$$

Here the degree n_i is positive and characterizes the value of gradient of the function f_i.

However, an expansion of these variants over two-dimensional case by direct introduction of the second coordinate would not embrace the cases when the domain of concentration is stretched not along one of the axes but along some direction determined by a vector $(\cos\alpha, \sin\alpha)$. To eliminate this shortcoming, equip every such domain with a local coordinate system, in which the direction of stretching of the domain coincides with one of the new axes. Evidently, this transformation of coordinate system should take into account transfer and rotation, i.e.,

$$\tilde{x}_i = (x - x_i)\cos\alpha_i + (y - y_i)\sin\alpha_i,$$

$$\tilde{y}_i = -(x - x_i)\sin\alpha_i + (y - y_i)\cos\alpha_i,$$

where $(x_i,\ y_i)$ are coordinates of the center of i-th concentration; α_i is angle of rotation of the axes of i-th concentration.

Then the functions of steps in two-dimensional case by analogy with one-dimensional case can be taken in the form

$$h(x,y) = h_0 + \sum_{i=1}^{n} (h_i - h_0) f_i(x, y, x_i, y_i, \alpha_i, \beta_i, \delta_i),$$

where β_i, δ_i are "characteristic sizes" of the domain of concentration, and f_i can be presented, for instance, as

$$2\left(\mathrm{ch}^{n_i}\left(\frac{\tilde{x}_i}{\beta_i}\right) + \mathrm{ch}^{m_i}\left(\frac{\tilde{y}_i}{\delta_i}\right)\right)^{-1}; \qquad \left(1 + \left(\frac{\tilde{x}_i}{\beta_i}\right)^{n_i} + \left(\frac{\tilde{y}_i}{\delta_i}\right)^{m_i}\right)^{-1};$$

$$\exp\left\{-\left(\frac{\tilde{x}_i}{\beta_i}\right)^{n_i} - \left(\frac{\tilde{y}_i}{\delta_i}\right)^{m_i}\right\}; \qquad \begin{cases} 1 - \left|\dfrac{\tilde{x}_i}{\beta_i}\right|^{n_i} \cdot \left|\dfrac{\tilde{y}_i}{\delta_i}\right|^{m_i}, & \text{if } (x,y) \in \Omega_i, \\ 0\ , & \text{if } (x,y) \notin \Omega_i. \end{cases}$$

where $\Omega_i = (x_i - \beta_i, x_i + \beta_i) \times (y_i - \delta_i, y_i + \delta_i)$, the degrees n_i, m_i as before are positive and characterize the gradients of the functions f_i along the direction $(\cos\alpha_i\ , \sin\alpha_i)$ and orthogonal to it.

2 Fragmentation of the boundary of multiply connected domain

Let the boundary of a multiply connected domain be formed by N piecewise smooth closed contours given in some Cartesian coordinate system $0xy$ in parametric form.

Consider a piecewise smooth contour $\mathit{\Gamma}$ (its index is omitted) formed by L smooth curves $\gamma_n, \quad n = 1, \dots, L$, whose parametric equations are

$$(x(t),\ y(t)) \equiv \boldsymbol{x}(t) = \boldsymbol{x}_n(t), \quad t_n^- \le t \le t_n^+, \tag{2.1}$$

where $t_n^-,\ t_n^+$ are the limits of variation of the parameter t for γ_n. From the conditions of continuity and closeness of the contour $\mathit{\Gamma}$ it follows that

$$\boldsymbol{x}_n(t_n^+) = \boldsymbol{x}_{n+1}(t_{n+1}^-), \quad n = 1, \dots, L-1,$$

$$\boldsymbol{x}_1(t_1^-) = \boldsymbol{x}_L(t_L^+).$$

Parametrization (2.1) must be such that for the inner contour $\mathit{\Gamma}$ the direction of encircling under increase of the parameter t would be clockwise, and for the external contour would be counterclockwise.

Fragmentation of $\mathit{\Gamma}$ is performed successively, starting from the first smooth curve $\gamma_1 : \ \boldsymbol{x} = \boldsymbol{x}_1(t)$. The first node on $\mathit{\Gamma}$ is $\boldsymbol{y}_1 = \boldsymbol{x}_1(t_1^-)$. Assume that $l-1$ first curves of the contour $\mathit{\Gamma}$ are fragmented already, the last constructed node on these curves is $\boldsymbol{y}_{n_{l-1}} = \boldsymbol{x}_{l-1}(t_{l-1}^+) = \boldsymbol{x}_l(t_l)$ and a part of the curve γ_l is fragmented, with the last node $\boldsymbol{y}_{n_{l-1}+k} = \boldsymbol{x}_l(t_k^l)$ where t_k^l is the value of the parameter t for the last node, $t_k^l \in [t_l^-, t_l^+)$. Then we construct next node of the curve γ_l by 3 steps.

Step 1.

Denote by $s_l(t_k^l, t)$ the length of a part of the curve γ_l, corresponding to the values $t_k^l,\ t:$

$$s_l(t_k^l, t) = \int_{t_k^l}^{t} \mid \dot{\boldsymbol{x}}(t) \mid dt, \quad t \in [t_k^l, t_l^+],$$

where $\dot{\boldsymbol{x}}(t)$ is derivative of $\boldsymbol{x}(t)$ with respect to t.

A new node $\boldsymbol{y}_{n_{l-1}+k+1}$ is constructed as follows: the value of the function of steps $h(x,y)$ is calculated in the last constructed node $\boldsymbol{y}_{n_{l-1}+k}$ and the solution $\tilde{t}_{k+1}^l$ of equation

$$s_l(t_k^l, t) = h(\boldsymbol{y}_{n_{l-1}+k}), \quad t \in [t_k^l, t_l^+], \tag{2.2}$$

is looked for. Suppose that solution of this equation exists (the contrary is considered in step 2). In this case it is unique due to positiveness of $\mid \dot{\boldsymbol{x}}_l(t) \mid$

(may be, with exception of finite number of points which do not influence uniqueness). According to the obtained value $\tilde{t}^l_{k+1}$ we calculate $\boldsymbol{x}_l(\tilde{t}^l_{k+1})$ and look for the solution t^l_{k+1} of the equation

$$s_l(t^l_k, t) = \frac{1}{2}[h(\boldsymbol{y}_{n_{l-1}+k}) + h(\boldsymbol{x}_l(\tilde{t}^l_{k+1}))]. \tag{2.3}$$

Suppose that this equation has a solution as well (the contrary is considered in step 3). Consider the inequality

$$\left| \frac{d - s_l(t^l_k, t^l_{k+1})}{s_l(t^l_k, t^l_{k+1})} \right| \leq \varepsilon, \quad d = \mid \boldsymbol{y}_{n_{l-1}+k} - \boldsymbol{x}_l(t^l_{k+1}) \mid, \tag{2.4}$$

the left-hand side of which is the relative difference between the length of arc and the length d of segment corresponding to the arc. The inequality characterises deviation of the arc from segment of stright line. Value of the parameter ε is specified by the user (for instance, $\varepsilon = 0.01$). If the inequality (2.4) is satisfied, then we declare the point $\boldsymbol{x}_l(t^l_{k+1})$ as a new node $\boldsymbol{y}_{n_{l-1}+k+1}$ and turn to construction of the next node. The declaration of the constructed point as a new node is substantiated by the fact that due to validity of equation (2.4) the one-dimensional finite element $[\boldsymbol{y}_{n_{l-1}+k}, \boldsymbol{y}_{n_{l-1}+k+1}]$ approximates the corresponding arc of curve good enough, and its length d under $h(x, y)$ smooth enough correlates with the average value of the function of steps over the ends of this element (see right-hand side of equation (2.3)). Otherwise, if the inequality (2.4) is not true, we successively decrease the right part of equation (2.3) by certain value (for example, by one tenth of the right-hand side) till the inequality (2.4) would be true. This situation appears when the length d of one-dimensional element calculated in accordance with the function of steps is "large" enough for acceptable approximation by this element of the arc which corresponds to it. Therefore successive decrease of this length is performed down to the value required by inequality (2.4). After that, we turn to construction of the following node $\boldsymbol{y}_{n_{l-1}+k+2}$.

Completion of the procedure of construction of new nodes on l-th curve of the contour Γ (and, respectively, on the whole contour Γ) is connected with the absence of solution of equation (2.2) and is described in step 2.

Step 2.

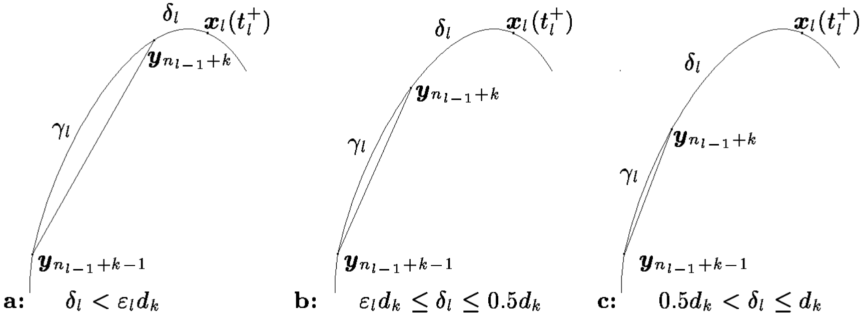
Now, consider the case when the equation (2.2) does not have solution, i.e.,

$$s_l(t^l_k, t^+_l) < h(\boldsymbol{y}_{n_{l-1}+k}).$$

This means that the last constructed node $\boldsymbol{y}_{n_{l-1}+k}$ is "close" to $\boldsymbol{x}_l(t^+_l)$, and construction of a new node by means of the function $h(x, y)$ is impossible.

Denote by δ_l the length of the remainder γ_l, and by d_k denote the length of the last constructed element:

$$\delta_l = s_l(t^l_k, t^+_l), \quad d_k =| \boldsymbol{y}_{n_{l-1}+k-1} - \boldsymbol{y}_{n_{l-1}+k} | \,.$$



δ_l
$\boldsymbol{x}_l(t^+_l)$
$\boldsymbol{y}_{n_{l-1}+k}$
γ_l
$\boldsymbol{y}_{n_{l-1}+k-1}$
a:
b:
c:

a: $\delta_l < \varepsilon_l d_k$ **b:** $\varepsilon_l d_k \leq \delta_l \leq 0.5 d_k$ **c:** $0.5 d_k < \delta_l \leq d_k$

Fig. 3. All possible situations when $\delta_l \leq d_k$.

In Fig. 3 the situations are shown when $\delta_l \leq d_k$ and displacement of the last constructed node takes place into the last point $\boldsymbol{x}_l(t^+_l)$ of the curve γ_l. In Fig. 3a, or a redistribution of δ_l occurs over all the previous one-dimensional elements approximating γ_l proportionally to the lengths (Fig. 3b). In Fig. 3c a new one-dimensional element with the length d_k is constructed, and new residual $\delta_l - d_k$ is introduced and redistributed as above, over all the elements proportionally to their lengths.

Thus, if the residual δ_l satisfies the inequality

$$\delta_l < \varepsilon_1 d_k \tag{2.5}$$

where ε_1 is small enough, for instance, 0.1, then we displace the last constructed node $\boldsymbol{y}_{n_{l-1}+k}$ into the last point of γ_l :

$$\boldsymbol{y}_{n_{l-1}+k} = \boldsymbol{x}_l(t^+_l)$$

and turn to construction of nodes on the next curve γ_{l+1}. If (2.5) is not valid, we consider the inequality

$$\delta_l < 0.5 d_k. \tag{2.6}$$

If (2.6) is true, then the number of nodes on γ_l remains the same, and the residual δ_l is redistributed over all arcs constructed on γ_l proportionally to their lengths. Denoting by $\tilde{s}_l$ the length of that part of the curve γ_l which was passed when constructing the nodes:

$$\tilde{s}_l = \sum_{i=0}^{k-1} s^l_{i,i+1}, \quad s^l_{i,i+1} \equiv s_l(t^l_i, t^l_{i+1}), \quad t^l_0 \equiv t^-_l \,. \tag{2.7}$$

Then the lengths of new arcs $s^{*l}_{i,i+1}$ are determined through the lengths of previous arcs $s^{l}_{i,i+1}$ according to the formulas

$$s^{*l}_{i,i+1} = s^{l}_{i,i+1}(1 + \delta_l/\tilde{s}_l), \quad i = 0, \dots, k-1. \tag{2.8}$$

Successively solving the equations

$$s_l(t^{*l}_i, t^{*l}_{i+1}) = s^{*l}_{i,i+1}, \quad t^{*l}_0 \equiv t^-_l, \quad i = 0, \dots, k-1, \tag{2.9}$$

we obtain new values t^{*l}_n and as well as new nodes on γ_l :

$$\begin{aligned} \boldsymbol{y}^*_{n_{l-1}+i} &= \boldsymbol{x}_l(t^{*l}_i), \quad i = 0, \dots, k-1, \\ \boldsymbol{y}^*_{n_{l-1}+k} &= \boldsymbol{x}_l(t^+_l). \end{aligned} \tag{2.10}$$

After that we turn to construction of nodes on the next curve γ_{l+1}.

If the residual δ_l does not satisfy the condition (2.6), then consider a new inequality

$$\delta_l \leq d_k. \tag{2.11}$$

If this inequality is true, i.e., the length of the remainder part of γ_l is less than the length of the last constructed one-dimensional finite element but exceeds its half-length due to violation of (2.6), then the number of nodes on γ_l is increased by one, new residial is introduced

$$\delta_l = s_l - \tilde{s}_l$$

where s_l is lenfth of γ_l, and

$$\tilde{s}_l = \sum_{i=0}^{k-1} s^l_{i,i+1} + s^l_{k-1,k},$$

Then we come to (2.8)-(2.10) with new $\delta_l, \tilde{s}_l$ and with addition of one more arc, the length $s^l_{k,k+1}$ of which is equal to the length of the last constructed arc $s^l_{k-1,k}$. At that, it is necessary to increase the value of k by one in (2.8)-(2.10).

In the case when (2.11) is not satisfied, we consider the equation

$$s_l(t^l_k, t) = d_k, \quad t \in [t^l_k, t^+_l], \tag{2.12}$$

which has a solution due to inequalities

$$s_l(t^l_k, t^+_l) \equiv \delta_l > d_k,$$

$$s_l(t^l_k, t^l_k) = 0 < d_k.$$

The value t^l_{k+1} obtained from (2.12) is tested for realization of the inequality (2.4) and further in accordance with the algorithm, with the difference that if (2.4) is not satisfied then not the right-hand side of (2.3) is successively decreased, but the right-hand side of (2.12).

Step 3.

Consider the case when the equation (2.3) have no solution. Then after obtaining $\tilde{t}^l_{k+1}$ from equation (2.2) we come to the inequality (2.4), in which t^l_{k+1} is substituted by $\tilde{t}^l_{k+1}$. If the inequality is true, we declare the point $\boldsymbol{x}_l(\tilde{t}^l_{k+1})$ as a new node $\boldsymbol{y}_{n_{l-1}+k+1}$ and turn to construction of new node on γ_l. Otherwise successively decrease the right-hand side of the equation (2.2) till (2.4) is satisfied, after that turn to construction of the next node.

The described algorithm allows to decompose the boundary of domain into one-dimensional finite elements (further called units), each of them being specified by a pair of integers n_1 and n_2 – numbers of its nodes – and their coordinates. The information on successive order of the units can be stored in two one-dimensional arrays K and M.

1) $k_i = K(i)$ is the number of units on the i-th contour of CGB, $i = 1, \dots, N$.

2) $m_j = M(j)$, $\quad m_{j+1} = M(j+1)$ are the numbers of the first and second nodes of j-th unit, respectively, if $j \neq \sum_{i=1}^{n} k_i$ for all $n = 1, \dots, N$. Otherwise, i.e., there exists such n_* that $j = \sum_{i=1}^{n_*} k_i$, then the number of the first node of such a unit is $M(j)$ as before, and the number of the second node is $M(\sum_{i=1}^{n_*-1} k_i + 1)$.

It is necessary to stress that the length of the array K changes in the process of triangulation, what is connected with change of the number of connectedness of the domain being not triangulated yet. The length of M also is not fixed, since either M is supplemented with new units, or the exhausted units from M are removed (the units which are not included in CGB on the next stage of construction of element).

3 Triangulation of a domain

It was noted above that the triangulation algorithm is based on successive filling of the domain with triangular elements. When filling the domain with the elements, CGB changes and in general case represents a number of closed broken contours. The number of connectivity of the domain being triangulated and the number of units of CGB change and can exceed the initial quantities. Therefore in the program one should watch that the length of the arrays K and M (see section 3) would not exceeded the given one.

Under coming together of different parts of CGB or in the domains of sharp changes of the function $h(x, y)$ the the triangular elements being constructed can be of an elongated form, and that can result in considerable errors when using this grid in finite element method. In order to avoid this defect, after construction of the grid an improvement is made which little changes compact triangles and significantly changes the elongated ones. The improvement of the grid is performed by means of the relations

$$\boldsymbol{x}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \boldsymbol{x}_{k_i}$$

where $\boldsymbol{x}_k$ is the node being corrected; n_k is the number of nodes surrounding the node $\boldsymbol{x}_k$; $\boldsymbol{x}_{k_i}$ are the surrounding nodes. The number n_k and nodes $\boldsymbol{x}_{k_i}$ are determined by means of the triangles possessing the common vertex $\boldsymbol{x}_k$.

The triangulation algorithm consists in the following.

Step 1.

Find an unit z_{min} of CGB which has the minimal length l_{min} and the nodes $\boldsymbol{x}^1_{min}, \boldsymbol{x}^2_{min}$. Denote by z^-_{min}, z^+_{min} the units preceeding and following z_{min}, respectively. Choose from z^-_{min}, z^+_{min} an unit z^*_{min} which forms with z_{min} the minimal angle β_{min} ($\beta_{min} = \min(\beta_1, \beta_2)$; the angles β_1 and β_2 are measured counterclockwise from z_{min} to z^-_{min} and from z^+_{min} to z_{min}, respectively). Denote the nodes of the choosen pair of nodes (they are either z^-_{min}, z_{min} or z_{min}, z^+_{min}) by $\boldsymbol{x}^{min}_1, \boldsymbol{x}^{min}_2, \boldsymbol{x}^{min}_3$. If $\beta_{min} \leq 80^o$ (otherwise we come to step 4), then go to step 2.

Step 2.

Make a test of getting into the triangle $\Delta(z_{min}, z^*_{min})$ (see Appendix 1) of the nodes of CGB, with exception of the nodes forming z_{min}, z^*_{min}. If there are no such nodes, come to step 3, otherwise from all the nodes got into $\Delta(z_{min}, z^*_{min})$ choose a node $\boldsymbol{y}_*$ closest to z_{min} (see Appendix 2) and go to step 12.

Step 3.

Consider a circle with radius $|\boldsymbol{x}^{min}_3 - \boldsymbol{x}^{min}_1|/2$ and the center $\boldsymbol{x}_c$ which is the midpoint of the third side z in $\Delta(z_{min}, z^*_{min})$, $\quad \boldsymbol{x}_c = (\boldsymbol{x}^{min}_1 + \boldsymbol{x}^{min}_3)/2$. If nodes of CGB do not get into the half of the circle external with respect to $\Delta(z_{min}, z^*_{min})$, then come to step 13. Otherwise from all the nodes choose the closest to z node $\boldsymbol{x}_m$ and divide the quadrangle $\Box(z_{min}, z^*_{min}, \boldsymbol{x}_m)$ so that the minimal angle of the resulting triangles would be maximal (the choice should be done from two variants of division of the quadrangle into two triangles). The consideration of the quadrangle is necessary in order to avoid constructing elongated triangle with the vertices $\boldsymbol{x}^{min}_1, \boldsymbol{x}^{min}_3, \boldsymbol{x}_m$, since

$\boldsymbol{x}_m$ can be located close enough to z. Further, declare both the obtained triangles as elements, remove z_{min}, z^*_{min} from CGB, determine connectivity of the domain, add two new units $[\boldsymbol{x}_1^{min}, \boldsymbol{x}_m], [\boldsymbol{x}_m, \boldsymbol{x}_3^{min}]$ (see Appendix 8) and go to step 1.

Step 4.

Construct the point

$$\boldsymbol{x}_* = \boldsymbol{x}^c_{min} + h_{cp}\boldsymbol{n}, \quad \boldsymbol{x}^c_{min} = \frac{1}{2}(\boldsymbol{x}^1_{min} + \boldsymbol{x}^2_{min}). \tag{3.1}$$

Here $\boldsymbol{n}$ is the normal to z_{min} directed inside the domain:

$$(\boldsymbol{x}^2_{min} - \boldsymbol{x}^1_{min}) \times \boldsymbol{n} = l_{min}\boldsymbol{e}_3, \quad \boldsymbol{e}_3 = (0,0,1);$$

h_{cp} is average value of the function of steps over the vertices of equilateral triangle constructed on the basis z_{min} :

$$h_{cp} = \frac{1}{3}\sum_{i=1}^{3} h(\boldsymbol{\xi}_i), \quad \boldsymbol{\xi}_j = \boldsymbol{x}^j_{min}, \quad j = 1,2; \quad \boldsymbol{\xi}_3 = \boldsymbol{x}^c_{min} + \frac{\sqrt{3}}{2} l_{min}\boldsymbol{n}.$$

Under certain conditions described below, the point $\boldsymbol{x}_*$ will be a new node, and the triangle $\Delta(z_{min}, \boldsymbol{x}_*)$ will be a new element.

On the basis of the unit z_{min} construct a rectangle Ω, one of which sides is z_{min} and the another is directed normally and its length equals $2H$. Here H is the altitude in $\Delta(z_{min}, \boldsymbol{x}_*)$ dropped on z_{min} :

$$2H = \sqrt{4|\boldsymbol{x}_* - \boldsymbol{x}^1_{min}|^2 - l^2_{min}}.$$

By means of the control domain Ω, let ascertain the criterions of proximity of the new node $\boldsymbol{x}_*$ to the previously constructed nodes and units. If in certain sense $\boldsymbol{x}_*$ is close to nodes or units, then we refuse to construct the new node and choose the best node from the close ones for construction of the new element.

Let define two sets M_0 and M_1 as follows.

M_0 is the set of numbers of nodes of CGB, which got into Ω, with exception of the numbers of nodes $\boldsymbol{x}^1_{min}$ and $\boldsymbol{x}^2_{min}$

$$M_0 = \left\{n : \ \boldsymbol{x}_n \in \Omega, \ \boldsymbol{x}_n \neq \boldsymbol{x}^i_{min}, \quad i = 1,2\right\}.$$

M_1 is the set of numbers of the units of CGB, which crosses $\partial\Omega$, with exception of the number of the minimal unit z_{min} :

$$M_1 = \left\{n : \ z_n \cap \partial\Omega \neq \emptyset, z_n \neq z_{min}\right\}.$$

Introduce two additional points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$:

$$\boldsymbol{z}_i = \boldsymbol{x}^{\,i}_{min} + (-1)^i \Delta l\, \boldsymbol{\tau}, \quad i = 1, 2,$$

where

$$\boldsymbol{\tau} = \frac{1}{l_{min}}(\boldsymbol{x}^{\,2}_{min} - \boldsymbol{x}^{\,1}_{min}),$$

$$\Delta l = \frac{1}{4}(h(\boldsymbol{x}^{\,c}_{min}) - l_{min}).$$

Determine the angle ξ at the vertex $\boldsymbol{z}_\xi$ in the triangle $\Delta(\boldsymbol{x}_*, \boldsymbol{z}_1, \boldsymbol{z}_2)$, where $\boldsymbol{z}_\xi = \boldsymbol{z}_1$, if $z^*_{min} = z^-_{min}$ and $\boldsymbol{z}_\xi = \boldsymbol{z}_2$, if $z^*_{min} = z^+_{min}$ (see step 1).

If one of the sets M_o or M_1 is not empty, then come to step 5.

Determine the angle α_1 at the vertex $\boldsymbol{x}_*$ in the triangle $\Delta(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{x}_*)$. If $\alpha_1 \geq 30^o$ and $\beta_{min} - \xi \geq 20^o$, then declare the point $\boldsymbol{x}_*$ a new node and come to step 2.

If $\alpha_1 < 30^o$, then redetermine the point $\boldsymbol{x}_*$ so that the new node has $\alpha_1 = 30^o$:

$$\boldsymbol{x}_* = \boldsymbol{x}^{\,c}_{min} + |\boldsymbol{z}_2 - \boldsymbol{x}^{\,c}_{min}| \cdot \text{tg } 75^o \cdot \boldsymbol{n}.$$

At that, if $\beta_{min} - 75^o < 20^o$, then come to step 2, else to step 14.

The introduction of the points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ is obliged to the fact that the unit z_{min} at one of the previous steps of construction of the element can be produced by different ways: through connection of two neighbouring units (step 13), through connection of two previously constructed nodes (step 12), through construction of a node (step 14). Therefore the length of z_{min} in the domains of larger gradients of the function of steps $h(x, y)$ can be 2-4 times less than the value $h(\boldsymbol{x}^c_{min})$. Since after construction of the grid an improvement is made which allows to extend z_{min} somewhat in such domains, it is better to estimate the quality of the element being constructed through the points $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ which are midpoints between $\boldsymbol{x}^1_{min}, \boldsymbol{x}^c_{min} - \frac{1}{2}h(\boldsymbol{x}^c_{min})\boldsymbol{\tau}$ and $\boldsymbol{x}^2_{min}, \boldsymbol{x}^c_{min} + \frac{1}{2}h(\boldsymbol{x}^c_{min})\boldsymbol{\tau}$, respectively. Thus, in the course of construction of a new node, in such domains we analyse not the elements constructed according to the new node but their possible transformations after inprovement of the grid. The estimation of value of the angle $\beta_{min} - \xi$ is performed in order to avoid acute angles between the units $z^-_{min}, [\boldsymbol{x}^1_{min}, \boldsymbol{x}^2_*]$ (if $z^*_{min} = z^-_{min}$) or $[\boldsymbol{x}_*, \boldsymbol{x}^2_{min}], z^+_{min}$ (if $z^*_{min} = z^+_{min}$).

In Fig. 4 a situation is shown when declaration of the triangle $\Delta(z_{min}, \boldsymbol{x}_*)$ as a new element results further in appearing the triangle $\Delta(z^+_{min}, \boldsymbol{x}_*)$ with acute angle. Therefore in such situations we will refuse to construct new node and (under favourable conditions) take as an element $\Delta(z_{min}, z^*_{min})$, i.e., come to step 2.

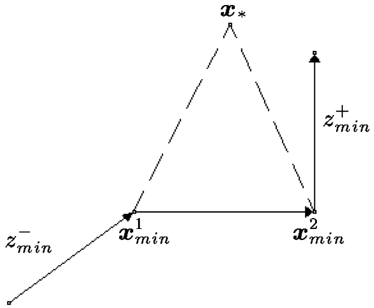


Fig. 4. In this situation there is no new node.

The analysis of the angles α_1 and $\beta_{min} - \xi$ was introduced into the initial variant of the algorithm after consideration of a large number of experimental calculations made in the domains of large gradients of $h(\boldsymbol{x})$. One of examples is shown in Fig. 5: a grid is shown before (Fig. 5a) and after (Fig. 5b) its improvement. From this, one can see that if the analysis of the elements being constructed is performed over the lengths of their sides (see Fig. 5a) but not over average values, then a sharp enough transition is possible from the elements of small sizes to elements of large sizes, what entails poor quality of elements in the domains in such vicinity.

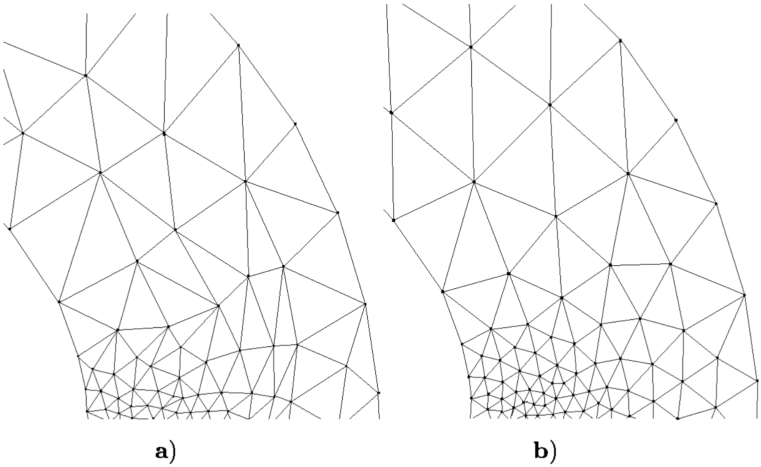


Fig. 5. The grid before **a**) and after **b**) improvement.

Step 5.

If $M_o \neq \emptyset$ (else go to step 9), then choose from M_0 a number m for which the corresponding node $\boldsymbol{x}_m$ is closest to z_{min}. To do this, determine the distances l_i (see Appendix 2) from the points $\boldsymbol{x}_i, i \in M_o$, to z_{min} and choose

$$l_m = \min_{i \in M_o} l_i \to m. \tag{3.2}$$

In the triangle $\Delta(\boldsymbol{x}_m, \boldsymbol{z}_1, \boldsymbol{z}_2)$ consider the angle α at the vertex $\boldsymbol{x}_m$ ($\boldsymbol{z}_i$ are determined in step 4). If $\alpha \geq 30^o$ (else come to step 8), then test an intersection of the segment $[\boldsymbol{x}_m, \boldsymbol{x}^1_{min}]$ with the units of CGB with numbers from M_1 without the nimbers of units neighbouring the nodes $\boldsymbol{x}_m, \boldsymbol{x}^1_{min}$ (see Appendix 3). For convenience, denote $\boldsymbol{x}_m$ by $\boldsymbol{y}_*$. Introduce an integer parameter IND of switching and set $IND = 0$.

Step 6.

If there are no intersections, then go to step 12 if $IND = 0$ or to step 14 if $IND = 1$.

Step 7.

There are intersections. With use of the nodes of the intersecting unit z_p, construct oriented triangles $\Delta(z_{min}; \boldsymbol{x}_k), \boldsymbol{x}_k$ are nodes of the unit z_p, $k \in \{k_1, k_2\}$ (see Appendix 4). When constructing these triangles, one should make sure of their existence (in Fig. 6 oriented triangle $\Delta(z_{min}, \boldsymbol{x}_{k_2})$ does not exist; $\boldsymbol{x}_m$ is the node closest to z_{min}).

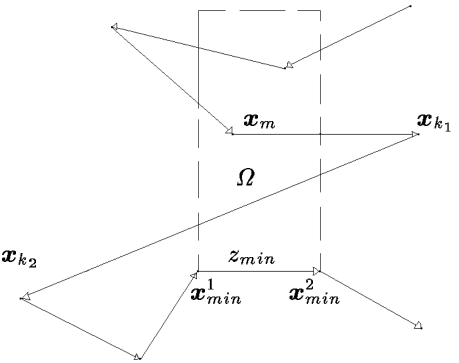


Fig. 6. Testing rectangle Ω.

From these triangles (if both exist) choose $\Delta(z_{min}, \boldsymbol{x}_{k_i})$, $k_i \in \{k_1, k_2\}$, whose minimal angle is larger (further, we will denote the minimal angle of any triangle $\Delta(z, \boldsymbol{x})$ by $\alpha(z, \boldsymbol{x})$). Having denoted the choosen node $\boldsymbol{x}_{k_i}$ by $\boldsymbol{y}_*$, test an intersection of both lateral sides of $\Delta(z_{min}, \boldsymbol{y}_*)$ with all the units of CGB except the minimal one and those adjoining it and node $\boldsymbol{y}_*$. Then come to step 6.

Step 8.

The closest node $\boldsymbol{x}_m$ is far enough from z_{min} (since $\alpha < 30^o$), therefore displace the constructed point $\boldsymbol{x}_*$ to z_{min} so that the distance between its new location (denote this point by $\boldsymbol{y}_*$) and z_{min} would be equal to $l_m/2$:

$$\boldsymbol{y}_* = \boldsymbol{x}^c_{min} + \frac{1}{2} l_m \boldsymbol{n},$$

and l_m is defined in (2).

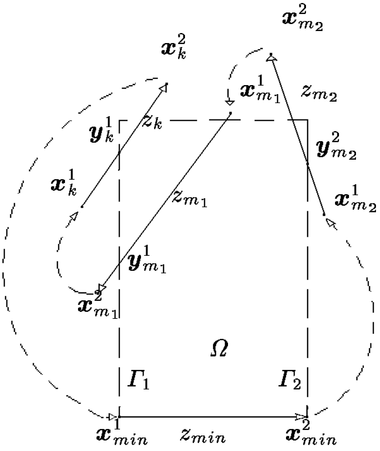


Fig. 7. Test of an intersection rectangle Ω with units.

In the triangle $\Delta(\boldsymbol{z}_1, \boldsymbol{z}_2, \boldsymbol{y}_*)$ determine the angle ξ at the vertex $\boldsymbol{z}_\xi$ defined in step 4. If $\beta_{min} - \xi < 20^o$, then come to step 2, else test an intersection of one of lateral sides of $\Delta(z_{min}, \boldsymbol{y}_*)$ with the units with numbers from M_1 and come to step 6, setting $IND = 1$.

Step 9.

If $\beta_{min} - \xi < 20^o$, then come to step 2, otherwise choose from all the units intersecting $\partial\Omega$ the units z_{m_1} and z_{m_2} which are "closest" to z_{min}. For this, consider all the points of intersection $\boldsymbol{y}^1_i$ of the units z_{k_i}, $k_i \in M_1$, with the lateral side Γ_1 of rectangle Ω, which comes through $\boldsymbol{x}^1_{min}$, and analogous points $\boldsymbol{y}^2_i$ for Γ_2. Then the numbers m_1 and m_2 are determined as

$$\min_{k_i \in M_1} |\boldsymbol{y}^j_i - \boldsymbol{x}^j_{min}| = m_j, \quad j = 1, \ldots, N_p.$$

If $\varGamma_1$ and $\varGamma_2$ intersect different "closest units ", then $N_p = 2$ (Fig. 7). If one of the lateral sides $\varGamma_j$ is not intersected by the units (Fig. 6), or if $\varGamma_1$ and $\varGamma_2$ are intersected by the same unit (Fig. 7), then $N_p = 1$, and consider only the number m_1 (if necessary, m_1 is specified as m_2, Fig. 8a).

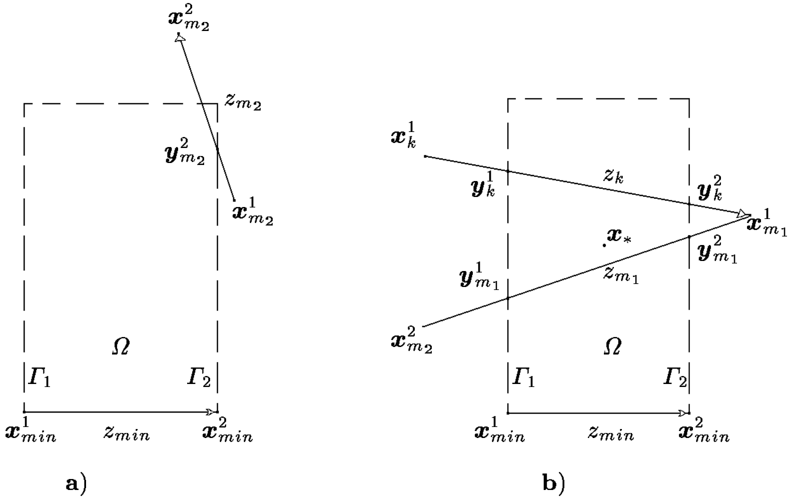


Fig. 8. All possible situations when lateral sides of $\varOmega$ intersect with units.

Consider oriented $\varDelta(z_{m_i}, \boldsymbol{x}_*)$, $\quad i = 1, \ldots, N_p$; $\quad \boldsymbol{x}_*$ is defined in (3.1). If such oriented triangles do not exist (that is possible only in the case, when the "closest" intersecting unit is unique and comes through $\varGamma_1$ and $\varGamma_2$ between z_{min} and $\boldsymbol{x}_*$, see Fig. 8b), then denote the unit z_{m_1} by z_p and come to step 7. Under existence of $\varDelta(z_{m_i}, \boldsymbol{x}_*)$ consider the angles α_i at the vertex $\boldsymbol{x}_*$ in these triangles, $i = 1, \ldots, N_p$. If $\alpha_i \le 90^o$, then go to step 14, otherwise choose $\alpha_{i_1} = \max\limits_{1 \le i \le N_p} \alpha_i$.

Step 10.

Construct oriented triangles $\varDelta(z_{min}, \boldsymbol{x}^j_{m_{i_1}})$, where $\boldsymbol{x}^j_{m_{i_1}}$ are nodes of the unit $z_{m_{i_1}}$. Possible values of the parameter j can be of the following list: $j \in \{1, 2\}$, if both the triangles exist. If only one triangle exists, then $j = 1$ if the node $\boldsymbol{x}^1_{m_{i_1}}$ is used, and $j = 2$ for the second node of the unit $z_{m_{i_1}}$.

Choose from $\varDelta(z_{min}, \boldsymbol{x}^j_{m_{i_1}})$ the triangle which has larger minimal angle:

$$\alpha(z_{min}, \boldsymbol{x}^{j_1}_{m_{i_1}}) \ge \alpha(z_{min}, \boldsymbol{x}^j_{m_{i_1}})$$

for all j from the list of values of this index: Then come to step 11.

Step 11.

If

$$\alpha(z_{min}, \boldsymbol{x}_{m_{i_1}}^{j_1}) \leq \alpha(z_{min}, \boldsymbol{x}_*),$$

then come to step 14, otherwise test an intersection of lateral sides of $\Delta(z_{min}, \boldsymbol{x}_{m_{i_1}}^{j_1})$ with all the units of CGB, except the nodes adjoining the node $\boldsymbol{x}_{m_{i_1}}^{j_1}$ and the unit z_{min}. Besides, test a getting the nodes of CGB into this triangle, except the node $\boldsymbol{x}_{m_{i_1}}^{j_1}$ and nodes $\boldsymbol{x}_{min}^1$, $\boldsymbol{x}_{min}^2$.

If there are no intersections and nodes inside the triangle, then redenote the node $\boldsymbol{x}_{mi_1}^{j_1}$ by $\boldsymbol{y}_*$ and go to step 12.

If there are intersections or if the triangle contains at least one node of CGB, then choose a new value j_2 from the list of values of parameter j (if it is not exhausted) and come to step 11, preliminary redenoting j_2 by j_1. If the list of parameter j is exhausted, then consider the second intersecting unit $z_{m_{i_2}}$ (under the condition that the list of parameter m_i is not exhausted, $i = 1, \ldots, N_p$), and if $\alpha_{i_2} > 90^o$, come to step 10, preliminarily redenoting i_2 by i_1. Otherwise (either $\alpha_{i_2} < 90^o$, or the list of parameter m_i is exhausted) come to step 14.

Step 12.

Declare the triangle $\Delta(z_{min}, \boldsymbol{y}_*)$ as an element, remove z_{min} from CGB, determine the number of connectivity of the domain, add two new units $[\boldsymbol{x}_{min}^1, \boldsymbol{y}_*], [\boldsymbol{y}_*, \boldsymbol{x}_{min}^2]$, and come to step 1.

The number of connectivity is increased by one, if $\boldsymbol{y}_*$ and $\boldsymbol{x}_{min}^1$ belong to one contour of CGB, and decreased by one, if these nodes belong to different contours (see Appendix 7).

Step 13.

Declare the triangle $\Delta(z_{min}, z_{min}^*)$ as an element, remove z_{min}, z_{min}^* from CBN, add one unit $[\boldsymbol{x}_1^{min}, \boldsymbol{x}_3^{min}]$ (see Appendix 5), and come to step 1.

Step 14.

Declare the triangle $\Delta(z_{min}, \boldsymbol{x}_*)$ as a new element, remove z_{min} from CBN, add two new units $[\boldsymbol{x}_{min}^1, \boldsymbol{x}_*], [\boldsymbol{x}_*, \boldsymbol{x}_{min}^2]$ (see Appendix 6), and come to step 1.

4 Conclusion

We illustrate of performance of the algorithms for different domains in figures below.

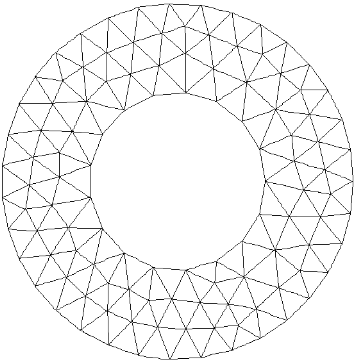


Fig. 9. A grid for a ring.

In Fig. 9 a grid is given for $h(\boldsymbol{x}) = 0.18$ for a ring with internal and external radii 0.5 and 1, respectively (all the values here and below are divided by dimensional unity). The equations of the contours of a ring were given in parametric form:

$$\boldsymbol{x}(t) = (\cos t,\ \sin t), \quad \boldsymbol{x}(t) = 0.5(\cos t,\ -\sin t), \quad 0 \le t \le 2\pi. \tag{4.1}$$

The value of the parameter ε (see inequality (2.4)) was set as 0.001. The grid has 178 elements and 115 nodes.

A ring with two circular cuts is shown in Fig. 10a . Two contours have parametrization (4.1), the other two are defined as follows:

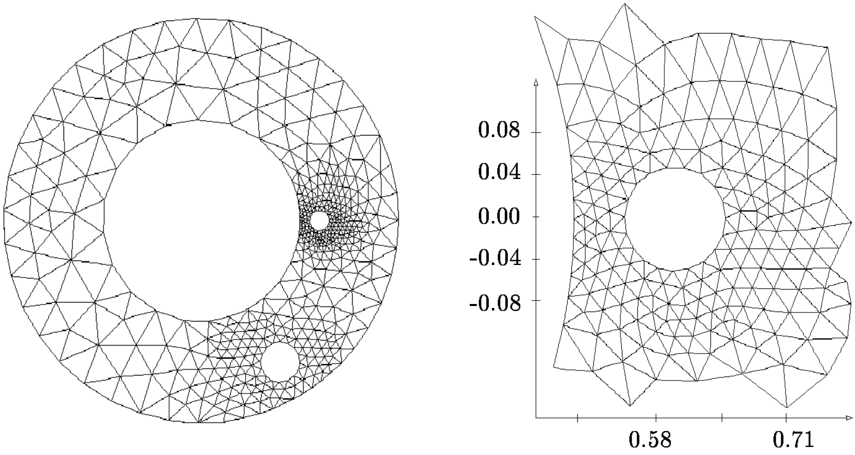
$$\begin{gathered}(x-0.6)^2 + y^2 = (0.05)^2, \quad \boldsymbol{x} = (0.6 + 0.05\cos t,\ -0.05\sin t);\\ (x-0.4)^2 + (y+0.7)^2 = (0.1)^2, \quad \boldsymbol{x} = (0.4 + 0.1\cos t,\ -0.7 - 0.1\sin t),\\ -2\pi \le t \le 0.\end{gathered} \tag{4.2}$$

The function of steps has two points of concentration in the centers of the circumferences (4.3):

$$h(x,y) = h_0 + (h_1 - h_0)/A(x,y) + (h_2 - h_0)/B(x,y),$$

$$A(x,y) = 1 + \left[\left(\frac{x-0.6}{0.3}\right)^2 + \left(\frac{y}{0.2}\right)^2\right]^2,$$

$$B(x,y) = 1 + \left[\left(\frac{x-0.4}{0.4}\right)^2 + \left(\frac{y+0.7}{0.3}\right)^2\right]^2,$$

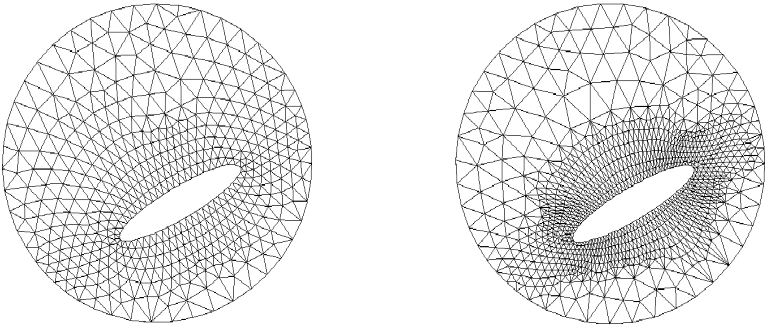


a) zoom, normal size. **b**) zoom, large size.

Fig. 10. A grid for a ring with two circular cuts.

$$h_0 = 0.168, \quad h_1 = 0.02, \quad h_2 = 0.04.$$

The grid has 864 elements and 483 nodes. The vicinity of the circumference with radius 0.05 is shown in Fig. 10b in larger scale.



a) 489 nodes. **b**) 954 nodes.

Fig. 11. Grids for a circular disk with an ellipsoidal cut.

Figures 11a and 11b demonstrate fragmentations of the same domain under diverse parameters of the function of steps. The external boundary of the domain is a circumference with radius 2, and its parametric equation is

$$\boldsymbol{x}(t) = 2(\cos t, \, \sin t), \quad 0 \le t \le 2\pi.$$

The internal boundary is an ellipse with center in the point $\boldsymbol{x}_c = (0.3; -0.5)$ and its major semiaxis is inclined at the angle $\alpha = 30^o$ to the axis Ox. The principal axes are $a = 0.9$ and $b = 0.2$. Parametric equation of the ellipse with the account of clockwise encircling of the boundary has the form

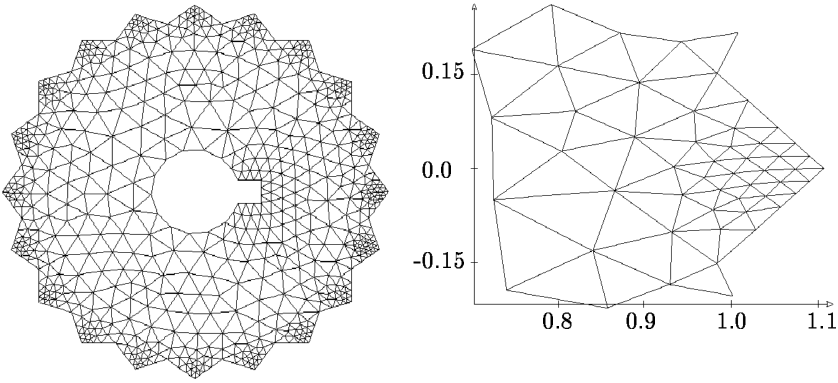
$$\begin{aligned} x(t) &= x_c + a\cos t\cos\alpha + b\sin t\sin\alpha, \\ y(t) &= y_c + a\cos t\sin\alpha - b\sin t\cos\alpha, \quad 0 \le t \le 2\pi. \end{aligned}$$

The center of the domain of concentration is the center of the ellipse. The function of steps for both the triangulations was taken in the form

$$h(x,y) = h_0 + (h_1 - h_0)\left\{1 + \left(\frac{\tilde{x}}{2a}\right)^m + \left(\frac{\tilde{y}}{2b}\right)^m\right\}^{-1},$$

$$\tilde{x} = (x - x_c)\cos\alpha + (y - y_c)\sin\alpha, \quad \tilde{y} = -(x - x_c)\sin\alpha + (y - y_c)\cos\alpha,$$

with $h_0 = 0.3$, $h_1 = 0.03$. In Fig. 11a we take $m = 2$, and in Fig. 11b we set $m = 4$. The grid in Fig. 11a has 880 elements and 489 nodes, and that in Fig. 11b has 1768 and 954, respectively.



a)a pinion: normal size.

b) one cog: large size.

Fig. 12. Grids for a pinion with 20 cogs.

In Fig. 12a the domain is a pinion with $N = 20$ cogs. A grid has 1219 elements and 702 nodes. The parameters of k–th cog are given in Fig. 13a, where

$$\alpha = 360^o/N, \quad \alpha_k = (k - 1.5)\alpha, \quad \alpha_k^* = \alpha_k + 0.5\alpha\,, 1 \le k \le N.$$

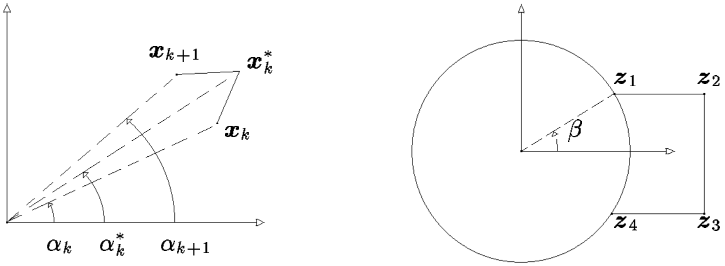
Then the points of bases and tops of the cogs are calculated as

$$\boldsymbol{x}_k = r(\cos\alpha_k; \sin\alpha_k), \quad \boldsymbol{x}_k^{\;*} = R(\cos\alpha_k^*; \sin\alpha_k^*), \quad k = 1,..,N; \;\; \boldsymbol{x}_{N+1} = \boldsymbol{x}_1 \;.$$

Parametrization of the external boundary is fulfilled for each side of a cog:

$$\boldsymbol{x} = \boldsymbol{x}_k + t(\boldsymbol{x}_{k} - \boldsymbol{x}_k), \quad \boldsymbol{x} = \boldsymbol{x}_{k} + t(\boldsymbol{x}_{k+1} - \boldsymbol{x}_{k}), \quad 0 \le t \le 1, \quad 1 \le k \le N.$$

Parametrization of the internal boundary is demonstrated in Fig. 13b:



a) the external boundary. **b)** the internal boundary.

Fig. 13. The boundary parametrization for a pinion.

$$\begin{aligned} &\boldsymbol{x}(t) = r_1(\cos t\;; -\sin t)\;, \quad r_1 = 0.25, \;\; \beta \le t \le 2\pi - \beta, \\ &\boldsymbol{x}(t) = \boldsymbol{z}_{i-1} + t(\boldsymbol{z}_i - \boldsymbol{z}_{i-1})\;, \quad 0 \le t \le 1, \;\; i = 2,3,4 \end{aligned}$$

Here

$$\begin{aligned} &\boldsymbol{z}_1 = r_1(\cos\beta;\; \sin\beta), \quad \boldsymbol{z}_2 = r_1(1.5;\; \sin\beta), \\ &\boldsymbol{z}_3 = r_1(1.5;\; -\sin\beta), \quad \boldsymbol{z}_4 = r_1(\cos\beta;\; -\sin\beta). \end{aligned}$$

The centers of the concentration domains for the cogs are located in the vertices $\boldsymbol{x}_k^{\;*}$; the axes of the domains lies on the rays $\beta = \alpha_k^*$ and in orthogonal directions. The value of step is the same and equals $h_1 = 0.25|\boldsymbol{x}_k - \boldsymbol{x}_k|$. The center of the concentration domain for internal cut is located in the point $(r_1; 0)$; the value of step is $h_2 = 0.05$; the concentration domain is stretched along the axis Ox.

So, the final form of the function of steps is

$$h(x,y) = h_0 + (h_1 - h_0)\sum_{i=1}^{N}\left\{1 + \left(\frac{\tilde{x}_k}{a}\right)^4 + \left(\frac{\tilde{y}_k}{b}\right)^4\right\}^{-1} +$$
$$(h_2 - h_0)\left\{1 + \left(\frac{x - r1}{cr_1}\right)^4 + \left(\frac{y}{dr_1}\right)^4\right\}^{-1},$$

where

$$a = 1.5(R - r), \quad b = 0.7\pi R/N, \quad c = 1.3, \quad d = 0.7,$$
$$\tilde{x}_k = (x - x_k{}^*)\cos\alpha_k^* + (y - y_k{}^*)\sin\alpha_k^*,$$
$$\tilde{y}_k = -(x - x_k{}^*)\sin\alpha_k^* + (y - y_k{}^*)\cos\alpha_k^*.$$

In order to demonstrate details of the grid on a cog, one of the cogs was cut out, and the enlarged grid is shown in Fig. 12b.

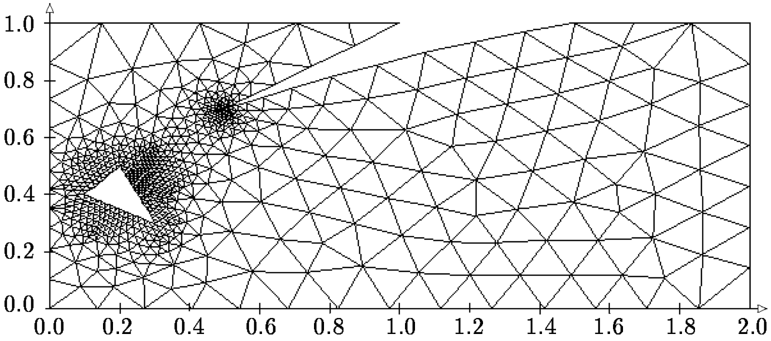


Fig. 14. A grid for the rectangle with wedge-shaped and triangle cuts.

In Fig. 14 a grid for a rectangle with wedge-shaped and triangle cuts is shown. The grid has 1348 elements and 726 nodes.

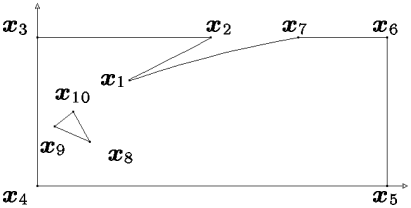


Fig. 15. The rectangle with wedge-shaped and triangle cuts.

In Fig. 15 the vertices x_i, $i = 1, ..., 10$, are shown which determine a domain where

$$x_1 = (0.5;\ 0.7),\ x_2 = (1;\ 1),\ x_3 = (0;\ 1),\quad x_4 = (0;\ 0),$$
$$x_5 = (2;\ 0),\qquad x_6 = (2;\ 1),\ x_7 = (1.5;\ 1),\ x_8 = (0.3;\ 0.3),$$
$$x_9 = (0.1;\ 0.4),\ x_{10} = (0.2;\ 0.5).$$

All the lines are straight except of the line from $\boldsymbol{x}_7$ into $\boldsymbol{x}_1$. T heir parametrization is

$$\boldsymbol{x}(t) = \boldsymbol{x}_* + t(\boldsymbol{x}_{**} - \boldsymbol{x}_*), \quad 0 \le t \le 1;$$

$\boldsymbol{x}_*$ is the beginning point of segment, $\boldsymbol{x}_{**}$ is the end point of segment. The line from $\boldsymbol{x}_7$ into $\boldsymbol{x}_1$ is the parabola $y = ax^2+bx+c$, where a, b, c are selected according to the conditions that the parabola comes through the points $\boldsymbol{x}_7$ and $\boldsymbol{x}_1$ and the value of derivative under $x = x_1$ is 0.4. Parametrization of this line with account of counterclockwise encircling of the external contour is

$$\boldsymbol{x}(t) = (-t;\ at^2 - bt + c), \quad -x_7 \le t \le -x_1.$$

The grid obtained has two concentration domains; the center of the first domain is the point $\boldsymbol{x}_c = \frac{1}{3}(\boldsymbol{x}_8 + \boldsymbol{x}_9 + \boldsymbol{x}_{10})$, the center of the second one is the point $\boldsymbol{x}_1$. Sizes of steps were chosen as

$$\begin{aligned} h_1 &= 0.1 \min(|\boldsymbol{x}_8 - \boldsymbol{x}_{10}|,\ |\boldsymbol{x}_9 - \boldsymbol{x}_{10}|,\ |\boldsymbol{x}_8 - \boldsymbol{x}_9|) \simeq 0.022, \\ h_2 &= 0.04\,|\boldsymbol{x}_1 - \boldsymbol{x}_7| \simeq 0.023. \end{aligned}$$

Each concentration domain was symmetric with respect to its center and the axes of a local coordinate system obtained by parallel transfer of the initial system into the center of concenteration

$$\begin{aligned} r_1 &= 1.6 \max(|\boldsymbol{x}_8 - \boldsymbol{x}_{10}|,\ |\boldsymbol{x}_9 - \boldsymbol{x}_{10}|,\ |\boldsymbol{x}_8 - \boldsymbol{x}_9|) \simeq 0.22, \\ r_2 &= 0.2\,|\boldsymbol{x}_1 - \boldsymbol{x}_7| \simeq 0.12\,. \end{aligned}$$

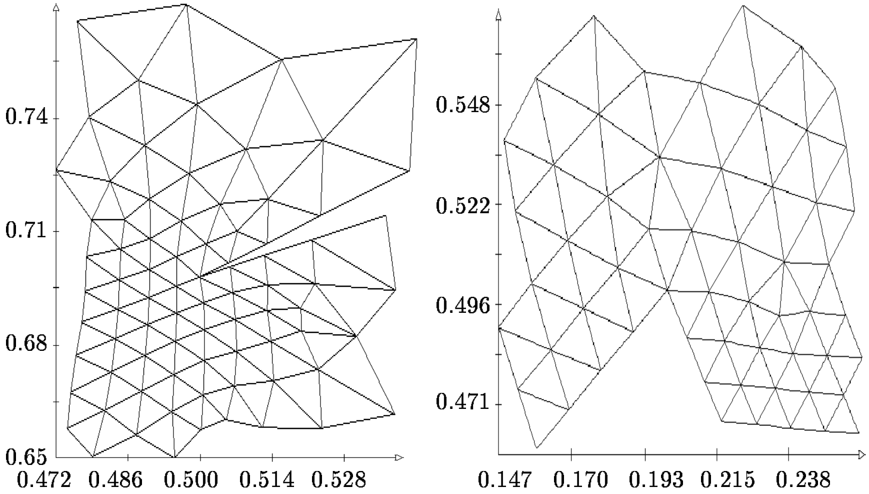
Major step of the grid is $h_0 = 0.15$. Then, with account of the form of the concentration domains and their centers, the function of steps was taken as

$$\begin{aligned} h(x,y) = h_0 + (h_1 - h_0)\left\{1 + \left(\frac{x - x_c}{r_1}\right)^4 + \left(\frac{y - y_c}{r_1}\right)^4\right\}^{-1} + \\ + (h_2 - h_0)\left\{1 + \left(\frac{x - x_7}{r_2}\right)^2 + \left(\frac{y - y_7}{r_2}\right)^2\right\}^{-1}. \end{aligned}$$

In figures 16a, 16b enlarged vicinities of the vertex $\boldsymbol{x}_1$ of wedge-shaped cut and of the vertex $\boldsymbol{x}_{10}$ of triangle cut are shown.

From explanations it is clear that for the construction of a grid it is necessary to input only the sufficient information: equations of contours and function of steps.

There are two disadvantages of the proposed algorithms. First, it is impossible to determine beforehand the length of the arrays storing the information about the grid. Second, numeration of nodes is not optimal. Since



a) the wedge-shaped cut.

b) the triangle cut.

Fig. 16. Zoom of cut areas, large size.

the length of arrays is not known beforehand, then in programs it is necessary to check the border of the array. If the ordered length of some array is less than it is necessary, then the programs halts and a message is displayed. Since the programs of fragmentation are used within the frames of more extensive computations, then for determination of lengths of the arrays it is recommended at first to run these programs for the given domain without the complementary programs.

More essential disadvantage is the non-optimality of numeration of nodes, what results in sparse stiffness matrix when using the finite element method. Storage of the whole stiffness matrix considerably increases the required resources of memory, therefore in the present case it is necessary either to use specific methods of storage and solution of large sparsed systems [6-18], or to avoid constructing global matrix and use some iterative methods. The reason for use of iterative methods is that they presuppose only calculation of products of matrix by vector, what can be done if we known the local siffness matrices. The array which stores the numbers of elements adjoining x_k can be filled immediately in the process of triangulation of the domain.

Thus, the existence of practically effective algorithms for sparse matrices and a possibility to solve a system of equations without formation of global matrix by some iterative method allow to eliminate the second shortcoming.

As a conclusion let note that these algorithms extremely convenient for use due to the possibility of elimination of the direct and indirect shortcomings of the algorithms of fragmentation together with the simplicity of handling, high degree of complexity of triangulated domains and a good quality of grid.

5 Appendix 1

Let the points $(x_1,\ y_1), (x_2,\ y_2), (x_3,\ y_3)$ be vertices of a triangle which are written down in the order of counterclockwise encircling. The point (x_*, y_*) lies outside the triangle if at least one of the following inequalities is valid:

$$v_i > 0, \qquad i = 1, 2, 3,$$

where

$$v_1 = (x_* - x_1)(y_2 - y_1) - (x_2 - x_1)(y_* - y_1),$$

$$v_2 = (x_* - x_2)(y_3 - y_2) - (x_3 - x_2)(y_* - y_2),$$

$$v_3 = (x_* - x_3)(y_1 - y_3) - (x_1 - x_3)(y_* - y_3).$$

6 Appendix 2

Let in the triangle $\Delta(\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2)$ the angles at the vertices $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are acute. Only such situations arise in the algorithm of triangulation. The distance l from the point $\boldsymbol{x}$ to the segment $[\boldsymbol{x}_1, \boldsymbol{x}_2]$ under the condition is

$$l = |\boldsymbol{x} - t(\boldsymbol{x}_2 - \boldsymbol{x}_1)|, \quad t = \frac{(\boldsymbol{x},\ \boldsymbol{x}_2 - \boldsymbol{x}_1)}{|\boldsymbol{x}_2 - \boldsymbol{x}_1|^2},$$

where (,) is the Euclidian scalar product; $|\ \cdot\ |$ is length of vector.

7 Appendix 3

Let two segments be determined by the points $\boldsymbol{x}_1 = (x_1^1,\ x_2^1), \boldsymbol{x}_2 = (x_1^2,\ x_2^2)$ and $\boldsymbol{y}_1 = (y_1^1, y_2^1)$, $\boldsymbol{y}_2 = (y_1^2,\ y_2^2)$. The test of intersection of two these segments can be performed as follows: points of the first segment are

$$\boldsymbol{x} = \boldsymbol{x}_1 + t_1(\boldsymbol{x}_2 - \boldsymbol{x}_1), \quad t_1 \in [0, 1],$$

and points of the second segment are

$$\boldsymbol{y} = \boldsymbol{y}_1 + t_2(\boldsymbol{y}_2 - \boldsymbol{y}_1), \quad t_2 \in [0, 1].$$

These segments do not intersect, if the system of two equations with respect to t_1 and t_2

$$\boldsymbol{x}_1 + t_1(\boldsymbol{x}_2 - \boldsymbol{x}_1) = \boldsymbol{y}_1 + t_2(\boldsymbol{y}_2 - \boldsymbol{y}_1)$$

does not have solution (the segments are parallel) or one of the solutions does not belong to the interval [0, 1].

More effective algorithm of testing is as follows: if at least one of the inequalities

$$v_i > 0 \qquad i = 1, 2,$$

is satisfied, then the segments do not intersect. Here

$$\begin{aligned} v_1 &= [(y_1^1 - x_1^1)(x_2^2 - x_2^1) - (x_1^2 - x_1^1)(y_2^1 - x_2^1)] \\ &\times [(y_1^2 - x_1^1)(x_1^2 - x_1^1) - (x_1^2 - x_1^1)(y_2^2 - x_2^1)], \\ v_2 &= [(x_1^1 - y_1^1)(y_2^2 - y_2^1) - (y_1^2 - y_1^1)(x_2^1 - y_2^1)] \\ &\times [(x_1^2 - y_1^1)(y_2^2 - y_2^1) - (y_1^2 - y_1^1)(x_2^2 - y_2^1)]. \end{aligned}$$

8 Appendix 4

Oriented triangle $\Delta(z_{min}, \boldsymbol{x}_k)$ is understood as a triangle with ordered list of vertices $\boldsymbol{x}_{min}^1 = (x_1^1,\ x_2^1)$, $\boldsymbol{x}_{min}^2 = (x_1^2,\ x_2^2)$, $\boldsymbol{x}_k = (x_1^k,\ x_2^k)$, its direction of encircling is determined counterclockwise and this encircling does not contradict to the list of vertices. Existence criterion of oriented triangle is the inequality

$$(x_1^2 - x_1^1)(x_2^k - x_2^1) - (x_1^k - x_1^1)(x_2^2 - x_2^1) > 0.$$

9 Appendix 5

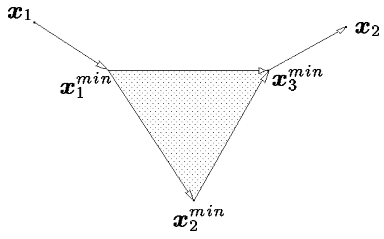


Fig. 17. Connection of two adjacent units creates the new triangular element.

When constructing triangular element by connection of two adjacent units on l-th contour of CGB, the value $K(l)$ of the array K is decreased by one, and the transformation of the array M is performed as shown in the figure 17. I.e. from the list of units of l-th contour the units

$$[\boldsymbol{x}_1^{min}, \boldsymbol{x}_2^{min}],\ [\boldsymbol{x}_2^{min}, \boldsymbol{x}_3^{min}]$$

are removed, and a new unit $[\boldsymbol{x}_1^{min}, \boldsymbol{x}_3^{min}]$ is included. In other words, the contour $\boldsymbol{x}_1$, $\boldsymbol{x}_1^{min}$, $\boldsymbol{x}_2^{min}$, $\boldsymbol{x}_3^{min}$, $\boldsymbol{x}_2$, $\ldots$, $\boldsymbol{x}_1$ is transformed into the contour $\boldsymbol{x}_1$, $\boldsymbol{x}_1^{min}$, $\boldsymbol{x}_3^{min}$, $\boldsymbol{x}_2$, $\ldots$, $\boldsymbol{x}_1$.

10 Appendix 6

When constructing triangle element (on the basis of unit z_{min} belonging to l-th contour of CGB) by construction of a new node $\boldsymbol{x}_*$, the value $K(l)$ of the array K is increased by one, and transformation of the array M is performed as shown in the figure 18.

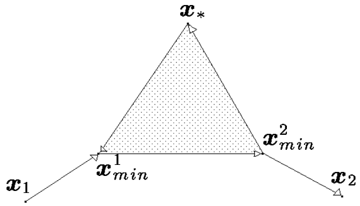


Fig. 18. A new node $\boldsymbol{x}_*$ creates the new triangular element.

I.e., from the list of units of l-th contour the unit $[\boldsymbol{x}^1_{min}, \boldsymbol{x}^2_{min}]$ is removed, and the units $[\boldsymbol{x}^1_{min}, \boldsymbol{x}_*]$, $[\boldsymbol{x}_*, \boldsymbol{x}^2_{min}]$ are included. In other words, the contour $\boldsymbol{x}_1$, $\boldsymbol{x}^1_{min}$, $\boldsymbol{x}^2_{min}$, $\boldsymbol{x}_2$, $\ldots$, $\boldsymbol{x}_1$ is transformed into the contour $\boldsymbol{x}_1$, $\boldsymbol{x}^1_{min}$, $\boldsymbol{x}_*$, $\boldsymbol{x}^2_{min}$, $\boldsymbol{x}_2 \ldots$, $\boldsymbol{x}_1$.

11 Appendix 7

When constructing a triangular element (on the basis of the unit z_{min} belonging to l_1 -th contour of CGB) through the previously constructed node $\boldsymbol{y}_*$ belonging to l_2-th contour, the number of connectivity of the domain is increased by one if $l_1 = l_2$, and two new contours of CGB are introduced due to fragmentation of the previous one; if $l_1 \neq l_2$, then the connectivity is decreased by one, to $K(l_1)$ the value $K(l_1) + K(l_2) + 1$ is assigned to $K(l_1)$ and $K(l_2)$ is set to be zero.

Transformation of the array M is performed as shown in the figures 19 and 20.

Under condition $l_1 = l_2$

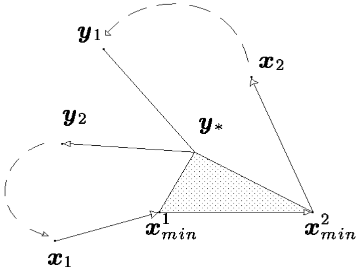


Fig. 19. The connectivity is increased.

the contour $\boldsymbol{x}_1,\ \boldsymbol{x}^1_{min},\ \boldsymbol{x}^2_{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{y}_1,\ \boldsymbol{y}_*,\ \boldsymbol{y}_2,\ \ldots,\ \boldsymbol{x}_1$ has divided into two contours $\boldsymbol{x}_1,\ \boldsymbol{x}^1_{min},\ \boldsymbol{y}_*,\ \boldsymbol{y}_2,\ \ldots,\ \boldsymbol{x}_1$ and $\boldsymbol{x}^2_{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{y}_1,\ \boldsymbol{y}_*,\ \boldsymbol{x}^2_{min}$. If $l_1 \neq l_2$

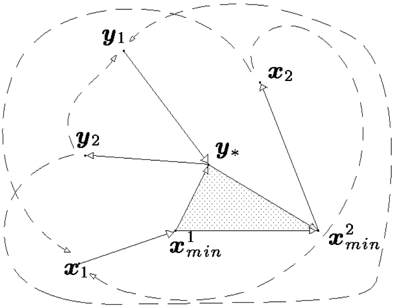


Fig. 20. The connectivity is decreased.

the contours $\boldsymbol{y}_1,\ \boldsymbol{y}_*,\ \boldsymbol{y}_2,\ \ldots,\ \boldsymbol{y}_1$ and $\boldsymbol{x}_1,\ \boldsymbol{x}^1_{min},\ \boldsymbol{x}^2_{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{x}_1$ have combined into one contour

$$\boldsymbol{y}_1,\ \boldsymbol{y}_*,\ \boldsymbol{x}^2_{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{x}_1,\ \boldsymbol{x}^1_{min},\ \boldsymbol{y}_*,\ \boldsymbol{y}_2,\ \ldots \boldsymbol{y}_1.$$

The possible variants of closure of l_1–th and l_2–th contours are shown with dotted lines, with exception of the case when the lines intersect.

12 Appendix 8

When constructing two triangle elements by division of a quadrangle into two triangles so that the minimal angle of the triangles would be maximal, two situations, as in App. 7, are possible.

1. The units z_{min} and z^*_{min} belong to l_1-th contour of CGB, and the node $\boldsymbol{x}_m$ which completes these units to a quadrangle belongs to l_2-th contour ($l_1 \neq l_2$). In this case the number of connectivity is decreased by one, because a junction of two contours into one takes place. The contours $\boldsymbol{y}_1$, $\boldsymbol{x}_m$, $\boldsymbol{y}_2$, $\ldots$, $\boldsymbol{y}_1$ and $\boldsymbol{x}_1$, $\boldsymbol{x}_1^{min}$, $\boldsymbol{x}_2^{min}$, $\boldsymbol{x}_3^{min}$, $\boldsymbol{x}_2$, $\ldots$, $\boldsymbol{x}_1$ have formed the contour

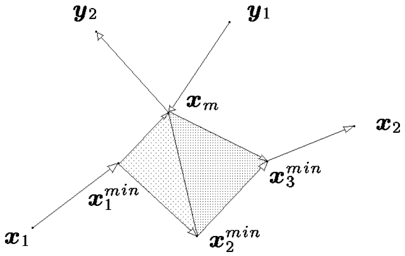


Fig. 21. The number of connectivity is decreased by 1.

$$\boldsymbol{x}_1,\ \boldsymbol{x}_1^{min},\ \boldsymbol{x}_m,\ \boldsymbol{y}_2,\ \ldots,\ \boldsymbol{y}_1,\ \boldsymbol{x}_m,\ \boldsymbol{x}_3^{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{x}_1.$$

2. The units z_{min} and z^*_{min} and the node $\boldsymbol{x}_m$ belong to the same contour of CGB. In this case the number of connectivity can increase by one or remain the same.

Increase of connectivity by one takes place if the node $\boldsymbol{x}_m$ does not form an unit of CGB with one of the nodes $\boldsymbol{x}_1^{min}$ or $\boldsymbol{x}_3^{min}$. Transformation of the array M is performed according to the figure 21 given above, but in this case the contour

$$\boldsymbol{x}_1,\ \boldsymbol{x}_1^{min},\ \boldsymbol{x}_2^{min},\ \boldsymbol{x}_3^{min},\ \boldsymbol{x}_2,\ \ldots,\ \boldsymbol{y}_1,\ \boldsymbol{x}_m,\ \boldsymbol{y}_2,\ \ldots,\ \boldsymbol{x}_1$$

is divided into two contours $\boldsymbol{x}_1$, $\boldsymbol{x}_1^{min}$, $\boldsymbol{x}_m$, $\boldsymbol{y}_2$, $\ldots$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_m$, $\boldsymbol{x}_3^{min}$, $\boldsymbol{x}_2$, $\ldots$, $\boldsymbol{y}_1$, $\boldsymbol{x}_m$.

The number of connectivity remains the same if the node $\boldsymbol{x}_m$ forms an unit with one of the nodes $\boldsymbol{x}_1^{min}$ or $\boldsymbol{x}_3^{min}$ (in the figure 22 it is $\boldsymbol{x}_3^{min}$). The contour

$$\boldsymbol{x}_1,\ \boldsymbol{x}_1^{min},\ \boldsymbol{x}_2^{min},\ \boldsymbol{x}_3^{min},\ \boldsymbol{x}_m,\ \boldsymbol{y}_1,\ \ldots,\ \boldsymbol{x}_1$$

turns into a new contour $\boldsymbol{x}_1$, $\boldsymbol{x}_1^{min}$, $\boldsymbol{x}_m$, $\boldsymbol{y}_1$, $\ldots$, $\boldsymbol{x}_1$.

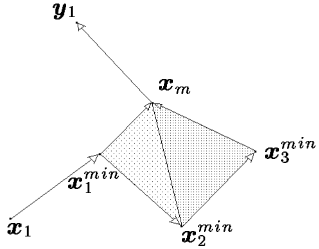


Fig. 22. The number of connectivity either increase by 1 or remain the same.

Note that in the previous Appendix similar case was not taken into consideration (when $\boldsymbol{y}_*$ forms an unit with $\boldsymbol{x}_{min}^1$ or $\boldsymbol{x}_{min}^2$ under $l_1 = l_2$), since it is eliminated by the algorithm.

References

1. Kamel Kh.A., Eisenshtein G.K.: *Automated construction of grid in two- and three- dimensional composite domains.* In: Calculation of elastic constructions by means of computer, Leningrad, 1974, pp. 21–35 (In Russian).
2. Kvitka A.L., Voroshko P.P., Bobritskaya S.D.: *Strained and deformed state of rotational bodies.* Kiev, 1977 (In Russian).
3. Umansky S.E.: *An algorithm and a program for triangulation of two-dimensional domain of arbitrary shape.* Problems of durability, 1978, № 6, pp. 83–87 (In Russian).
4. Milkova N.I.: *Peculiaries of discretization of a domain for solution of the problems of stress concentration by the method of finite elements.* Mashinovedenie, 1979, № 2, pp. 67–71 (In Russian).
5. Sakalo V.I., Shkurin A.A.: *An universal program for triangulation of dwo-dimensional domain of arbitrary shape with concentrations of the grid.* Problems of durability, 1985, № 1, pp. 106–108(In Russian).
6. George A., Lu G.: *Numerical solution of large sparse systems of equations.* Moscow, 1984 (In Russian).
7. Zlatnev Z., Esterbu O.: *Direct methods for sparse matrices.* Moscow, 1987 (In Russian).
8. Pissanetsky S.: *Technology of sparse matrices.* Moscow, 1988 (In Russian).
9. Tewarson R.P.: *Sparse Matrices.* Academic Press, New York, 1973.
10. Stewart G.W.: *Introduction to Matrix Computations.* Academic Press, New York, 1973.
11. Bunch J.R., Hopcroft J.E.: *Triangular factorization and inversion by fast matrix multiplication.* Math. Comput., 1974, 28, pp. 231–236.

12. Bunch J.R., Rose D.J.: *Sparse Matrix Computations.* Academic Press, New York, 1976.
13. Duff I.S.: *Sparse Matrices and their Uses.* Proceedings of the IMA Conference, University of Reading, 9-11 July, 1 980. Academic Press, London.
14. Gustavson F.G.: *An efficient algorithm to perform sparse matrix multiplication.* Argonne Conference on Sparse Matrix Computations, Argonne National Laboratory, 1976, Research Report RC-6176, IBM T.J. Watson Research Center, New York.
15. Munksgaard N.: *Fortran subroutines for direct solution of sets of sparse and symmetric linear equations.* Report 77.05, 1977, Numerical Institute Lyngby, Denmark.
16. Willoughby R.A.: *Proceedings of the symposium on sparse matrices and their applications.* Yorktown Heights, NY, IBM Report RAI, № 11707, 1969.
17. Willoughby R.A.: *A survey of sparse matrix technology.* IBM Research Report RC 3872, 1972.
18. Duff I.S.: *The solution of nearly symmetric sparse linear equations.* In: Computing Methods in Applied Sciences and Engineering VI, Proc. 6th Int. Symp. (Versailles, 1983), Amsterdam, North Holland, 1984, pp. 57–74.

A batch of applied programs for numerical solution of convection-diffusion boundary-value problem

Kireev I.V., Pyataev S.F., Shaidurov V.V.

Introduction

The work consists in development of an economical algorithm based on the classic variant of finite element method and intended for numerical solution of convection-diffusion boundary-value problem

$$-\varepsilon \triangle u + b_1 \frac{\partial u}{\partial x} + b_2 \frac{\partial u}{\partial y} = f \quad \text{in} \quad \Omega, \tag{1}$$

$$u = g \quad \text{on} \quad \Gamma. \tag{2}$$

Here two-dimensional domain Ω is limited by piecewise smooth boundary Γ; ε is a small positive number; b_1, b_2, f, g are smooth enough functions.

A good adaptation to the conditions of this problem is required from the algorithm, which would ensure high-accurate solution of boundary-value problem under linear approximation of the function $u(x, y)$ on each finite element. This means that an automatic division of the initial domain into finite elements oriented along the characteristics should be anticipated in the algorithm, and the requirement of economy indispensably leads to the use of the technology of embedded grids.

The idea of this algorithm is as follows: for construction of a new grid it is sufficient to analyze the behavior of piecewise linear and Hermitian cubic interpolations of the approximate solution obtained within the framework of standard finite element approach, and on the basis of this analysis to construct a partition of edges of finite elements, which automatically leads to construction of a new embedded grid accounting for more subtle peculiarities of the desired solution.

For realization of this idea, it is necessary to have an algorithm of determination of partial derivatives of the function $u(x, y)$ from its given node values. From a number of algorithms of determination of partial derivatives we have chosen a method described below, which, in our opinion, most organically matches this class of boundary-value problems. Unfortunately, theoretical substantiation of this statement is very problematic, but the approach being proposed has made a good showing in a great number of numerical experiments.

The testing of algorithms and programs has been performed for a simpler boundary- value problem; it was assumed that the domain $\Omega = [0,1] \times [0,1]$ is unit square, $b_1 \equiv 1$, $b_2 \equiv 0$, and $f \equiv c$ is constant, i.e., the following equation was considered

$$-\varepsilon \Delta u + \frac{\partial u}{\partial x} = c \quad \text{in} \quad (0,1) \times (0,1).$$

It is easy to verify that this equation admits solutions of the form

$$u(x, y) = \left(c_1 e^{\lambda_1 (x-1)} + c_2 e^{\lambda_2 x}\right) \sin n\pi y + ay + b + cx,$$

where a, b, c_1, c_2 are certain constants, and

$$\lambda_1 = \frac{1}{2\varepsilon}(1 + \sqrt{1 + (2n\varepsilon)^2}) > 0, \quad \lambda_2 = \frac{1}{2\varepsilon}(1 - \sqrt{1 + (2n\varepsilon)^2}) < 0,$$

at that $\lambda_1 \cong \varepsilon^{-1} + n^2\varepsilon$ and $\lambda_2 \cong -n^2\varepsilon$ under $\varepsilon \ll 1$.

If we are interested in solutions which do not depend on y, then, as it is easy to show, the solution of the equation is of the form

$$u(x) = c_1 \exp \frac{x-1}{\varepsilon} + c_2 + cx.$$

Just in this class of functions the debugging and testing of the algorithm of edge division have been carried out.

1 An algorithm of determination of partial derivatives

Let approximate the partial derivatives of a function $u(x, y)$ determined by numerical values u_J in nodes $M_J(x_J, y_J) : u(x_J, y_J) = u_J$ in the vicinity of the point M_0 as a linear combination

$$\frac{\partial u(x,y)}{\partial x} = u_x^0 + (x - x_0)u_{xx}^0 + (y - y_0)u_{xy}^0,$$

$$\frac{\partial u(x,y)}{\partial y} = u_y^0 + (x - x_0)u_{xy}^0 + (y - y_0)u_{yy}^0,$$

where $u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0$ are certain unknown constants to be determined for each node of the grid.

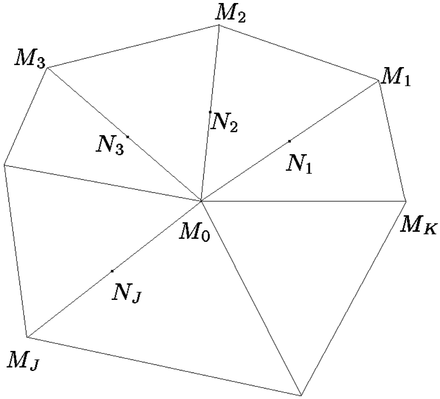


Fig. 1:
The nondegenerate case of d_{M_0}.

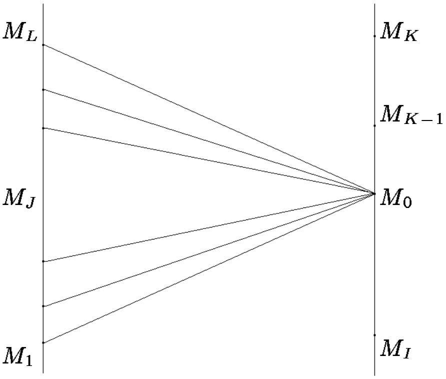


Fig. 2:
The degenerate case of d_{M_0}.

Then the central difference $u_J - u_0$ approximates the derivative of the function

$$u_J(t) = u(x_0 + t(x - x_0), y_0 + t(y - y_0)),\ t \in [0,1]$$

in the point N_J $(t = 0.5)$ with error $O(h_J^3)$, where

$$h_J = \sqrt{(x_J - x_0)^2 + (y_J - y_0)^2},$$

$u_J(0) = u_0$, and $u_J(1) = u_J$. Therefore for smooth enough function $u(x,y)$ the following relations should be valid:

$$d_J(u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0)/h_J^3 = O(1)$$

where

$$\begin{aligned} d_J(u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0) &= (u_J - u_0) \\ - (x_J - x_0)&[u_x^0 + 0.5(x_J - x_0)u_{xx}^0 + 0.5(y_J - y_0)u_{xy}^0] \\ - (y_J - y_0)&[u_y^0 + 0.5(x_J - x_0)u_{xy}^0 + 0.5(y_J - y_0)u_{yy}^0]. \end{aligned}$$

Grouping together neighbouring to M_0 nodes M_1, $M_2, \ldots, M_J, \ldots, M_K$, (see Fig. 1) construct the functional

$$d_{M_0}(u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0) = \sum_{J=1}^{J=K} \{d_J(u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0)\}^2 / h_J^6.$$

It is easy to verify that the components u_x^0, u_y^0 of solution of the least square problem

$$d_{M_0}(u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0) \underset{u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0}{\longrightarrow} \inf$$

give an approximation to partial derivatives

$$\frac{\partial u(x_0, y_0)}{\partial x}, \frac{\partial u(x_0, y_0)}{\partial y}$$

of a smooth enough function $u(x, y)$ to within $\max\{h_J^2\}$. At that, total number of nodes M_J necessary for calculation of partial derivatives in the point M_0 must be not less than 5 ($K \geq 5$).

However, arbitrary sequences of points close to M_0 cannot be used for such procedure. So, for instance, a sequence of points similar to that shown in Fig. 2 gives a functional d_{M_0} degenerate with respect to u_x^0, u_y^0, u_{xx}^0, u_{xy}^0, u_{yy}^0, whose minimization problem has infinite number of solutions. Therefore, if in the process of calculation it appears that quadratic functional d_{M_0} generates a linear system of algebraic equations with singular matrix, then additional nodes M_J immediately neighbouring the node M_0 are consequently taken into consideration, till the functional d_{M_0} will become nondegenerate. The highest accuracy of the derivatives calculated in such a way is reached in internal points of the domain.

2 Construction of a sequence of embedded grids

The described above algorithm of determination of partial derivatives in vertices of finite elements from calculated node values of numerical solution $u(x, y)$ has been used for construction of a sequence of embedded grids.

A new grid was constructed on the basis of analysis of behaviour on each edge of the initial grid of both linear and Hermitian cubic interpolations of function $u(x, y)$ constructed from node values of function u and values of partial derivatives u_x, u_y calculated in the nodes of the grid.

An edge was not divided, if on the edge the module of maximal difference of values between linear and cubic splines did not exceed ε_a. Otherwise, a new point was chosen inside the edge, proceeding from the following reasonings.

Denote by M'_J the desired point of division of the edge M_0M_J. Then, as it is shown in Fig. 3, on the edge M_0M_J a piecewise linear approximation of the function $u(x,y)$ appears; minimizing in some norm the residual between the latter and cubic approximations, we obtain an algorithm of construction of the point M'_J of the edge M_0M_J.

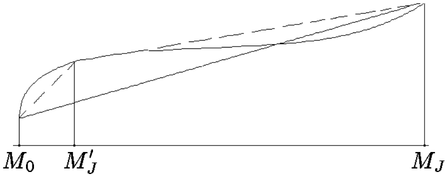


Fig. 3:
A piecewise linear approximation $u(x,y)$ on the edge M_0M_J.

Let give more details. Denote by $u(t)$ the cubic spline for the edge $[M_0, M_J]$; $t \in [0,1]$ and the values

$$u(0) = u_0,\ u(1) = u_J,\ \frac{du}{dt}(0) = u'_0,\ \frac{du}{dt}(1) = u'_J$$

are given. Then

$$u(t) = a_{00} + a_{01}t + a_{02}t^2 + a_{03}t^3,$$
$$u(t) = a_{10} + a_{11}(1-t) + a_{12}(1-t)^2 + a_{13}(1-t)^3,$$

where

$$\begin{aligned}
a_{00} &= u_0; \quad a_{01} = u'_0;\\
a_{02} &= 3(u_1 - u_0) - 2u'_0 - u'_1;\\
a_{03} &= -2(u_1 - u_0) + u'_0 + u'_1;\\
a_{10} &= u_1; \quad a_{11} = -u'_1;\\
a_{12} &= -3(u_1 - u_0) + u'_0 + 2u'_1;\\
a_{13} &= 2(u_1 - u_0) - u'_0 - u'_1.
\end{aligned}$$

Let $v_0(t)$ and $v_1(t)$ be linear approximations of the function $u(t)$ on the intervals $[M_0, M'_J]$ and $[M'_J, M_J]$, where $t \in [0, \tau]$ and $t \in [\tau, 1]$, respectively;

$0 < \tau < 1$. Then the functional

$$\Phi_{L_2} = \int\limits_0^\tau \left(u(t) - v_0(t)\right)^2 dt + \int\limits_\tau^1 \left(u(t) - v_1(t)\right)^2 dt$$

gives the square of norm of residual between $u(t)$ and its piecewise linear approximation $v_0(t)$, $v_1(t)$ in the space $L_2[0,1]$. Direct computations give the following expression for the functional:

$$\begin{aligned}\Phi_{L_2} &= (7a_{02}^2 + 21a_{02}a_{03}\tau + 16a_{03}^2\tau^2)\tau^5 \\ &+ (7a_{12}^2 + 21a_{12}a_{13}(1-\tau) + 16a_{13}^2(1-\tau)^2)(1-\tau)^5\end{aligned}$$

where the numbers a_{ij} are defined above. Minimizing this functional with respect to $\tau \in (0,1)$, we determine the coordinates of the point M'_J which is new node for the new grid; for this purpose it is necessary to solve an equation of the fifth power with respect to τ.

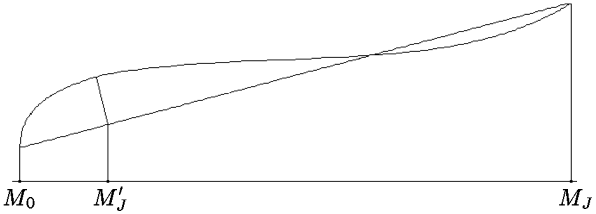


Fig. 4: The distance between graphs of the cubic polynomial and its linear interpolation.

For numerical solution of algebraic equation the Newton method was used, and as an initial approximation the point of the edge $[M_0, M_J]$ was taken, in which the distance between the graphs of the cubic polynomial and its linear interpolation is maximal, as shown in Fig. 4. As a rule, this approximation is rather good, and for correction of it with a reasonable accuracy it is sufficient to make only several iterations of the Newton method.

Besides Φ_{L_2}, other functionals have been considered. So, for instance, a number of functionals have been considered which approximate residual functional from $C[0,1]$. However, test computations have shown that the results differ insignificantly, but the time of computation when constructing the embedded grid increases greatly. Apparently, this is connected with the fact that the node values u_J themselves are results of computations and have the accuracy of the order $\max\limits_J\{h_J^2\}$ where h_J is the length of J-th edge.

3 Program realization of the algorithm

A complex of programs in language **C** for numerical solution of convection-diffusion boundary-value problem (1)–(2) has been designed on the basis of the algorithm described above.

For numerical solution of convection-diffusion boundary-value problem a scheme with the second order of accuracy has been used, which generates a system of equations with M-matrix, satisfying discrete maximum principle.

The solution of the obtained system of linear algebraic equations has been carried out by iterative Gauss-Seidel procedure under special ordering of equations and unknowns. The number of iterations on each embedded grid was fixed and did not exceed 15.

In Fig. 5 – 20 some results of operation of the procedure of construction of embedded grid for solution of boundary-value problem under $x \in [0,1]$, $y \in [0,1]$ and $u(x,0) = u(x,1) = u(0,y) = u(1,y) = 0$, $\varepsilon = 10^{-3}$ are shown. The initial triangulation was generated by uniform division of the sides of the square into 8 equal intervals with subsequent diagonal division of each elementary square into triangles. An edge was not divided, if the module of the maximal on the edge difference between linear and cubic splines did not exceed $\varepsilon_a = 10^{-3}$; for solution of finite-dimensional problem on each of the grids the Seidel method with fixed number of iterations (=50) was used.

Fig. 5 – 8 show the character of arising grids under $\varepsilon_a = \varepsilon$. In figures 9 – 20 the information on the sequence of grids arising under $\varepsilon_a = 50\varepsilon$ is reflected;

Fig. 9 – 12 show general dynamics of the sequence of grids;

Fig. 13 – 14 show the changes of grid at two last steps in the square [0.0, 0.125] × [0.0, 0.125], eightfold enlarged;

Fig. 15 –16 show the changes of grid at two last steps in the square [0.5, 0.625] × [0.0, 0.125], eightfold enlarged;

Fig. 17 – 18 show the changes of grid at two last steps in the square [0.875, 1.0] × [0.0, 0.125], eightfold enlarged;

Fig. 19 – 20 show the changes of grid at two last steps in the square [0.875, 1.0] × [0.375, 0.5], eightfold enlarged.

In the course of joint researches in Augsburg Technical University series of test computations on different computers and under several operation systems has been carried out. Main objective of these computations was to estimate real time necessary for solving the considered boundary-value problem.

In our opinion, some of these results are rather interesting; they are represented below in the form of a table.

HOSTNAME	TYPE	MHz	MB	SYSTEM	1	2	3
MALAGA	R4000	150	96	IRIX	58.9	3.75	187.0
SEVILLA	R8000	75	512	IRIX64	42.8	3.88	171.0
MARBELLA	PENTIUM PRO	200	128	LINUX	20.2	2.40	56.7
ALCALA	PENTIUM II	266	128	LINUX	13.8	1.68	71.7
ZARAGOZA	ALPHA	500	128	D.UNIX	9.7	1.10	36.6
LACORUNA	ALPHA	533	256	LINUX	11.5	1.30	47.6
BURGOS	ALPHA	533	256	LINUX	11.7	1.28	47.9

Here the first five columns contain general information about the computers which were used in the computational experiment; this information has been kindly granted to us by professor U. Rüde.

C-version of the program contains four main parts:
(I) procedures of construction of initial triangulation for Ω;
(II) procedures of formation of the global system of linear algebraic equations;
(III) procedures realizing the iterative Gauss-Seidel process with special ordering of equations and unknowns;
(IV) procedures which construct embedded grid by above method using the solution from (III).

The tests have shown that at the beginning of the computational process the time of execution of each of I-IV parts of the **C**-program is proportional to N_{point} which is the number of points of the initial grid. Therefore, the time of execution of each part of the program in the course of the test computations was divided by $N_{point} = 10^6$ after the statistical processing of several numerical experiments. Here "time of execution" means the user time obtained by the command "**time**" of UNIX operation system.

The column 1 contains the time of computations for the parts I, II on a regular grid. The column 2 represents the time spent on realization of one full iteration in the Gauss-Seidel method when solving the system of linear algebraic equations. And, finally, the column 3 contains total time of computation of the stages IV and II of **C**-program.

The computations have been performed for the case when

$$\Omega = [0,1] \times [0,1],\ \varepsilon = 0.001,\ b_1 \equiv 1,\ b_2 \equiv 0,\ f \equiv 1,\ g \equiv 0.$$

One can see from this table that the efficiency of a computational complex strongly depends on both the parameters of computer and the type of operation system.

Fig. 5: Changes of grid at step 1; $\varepsilon = 10^{-3}$.

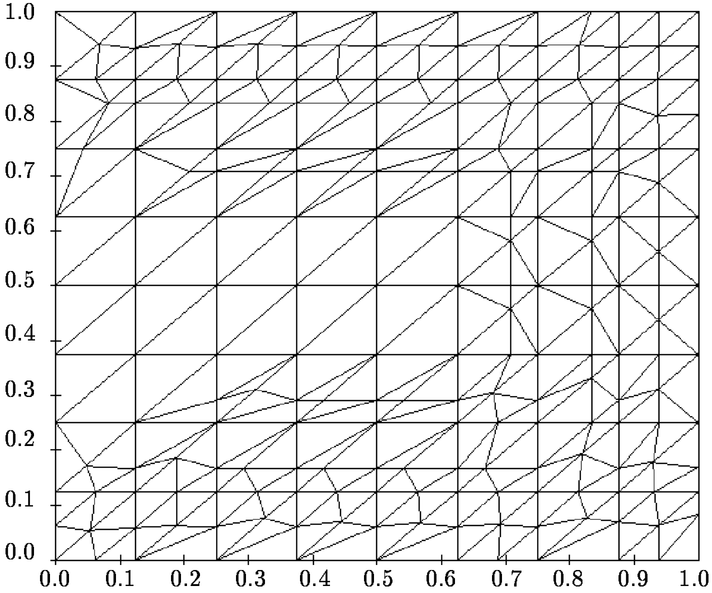


Fig. 6: Changes of grid at step 2; $\varepsilon = 10^{-3}$.

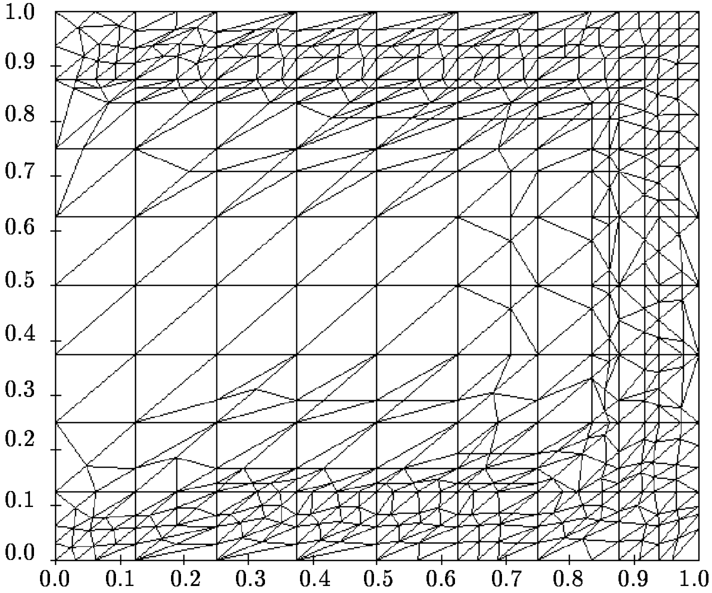


Fig. 7: Changes of grid at step 3; $\varepsilon = 10^{-3}$.

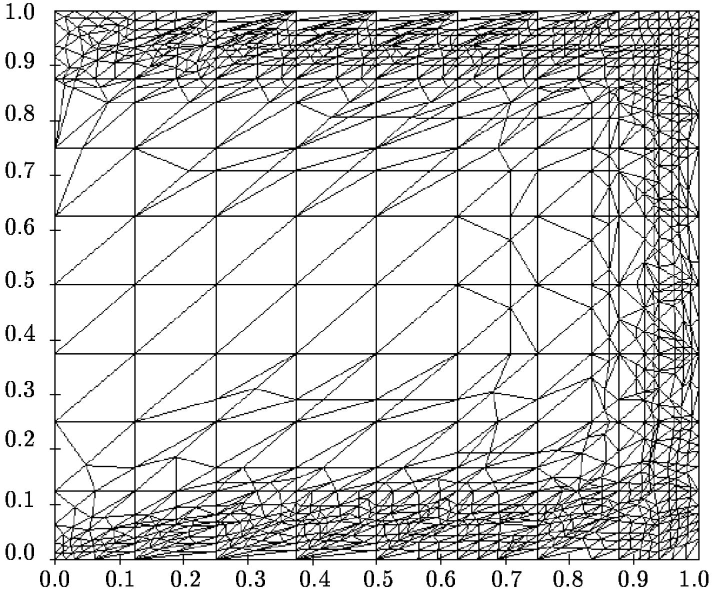


Fig. 8: Changes of grid at step 4; $\varepsilon = 10^{-3}$.

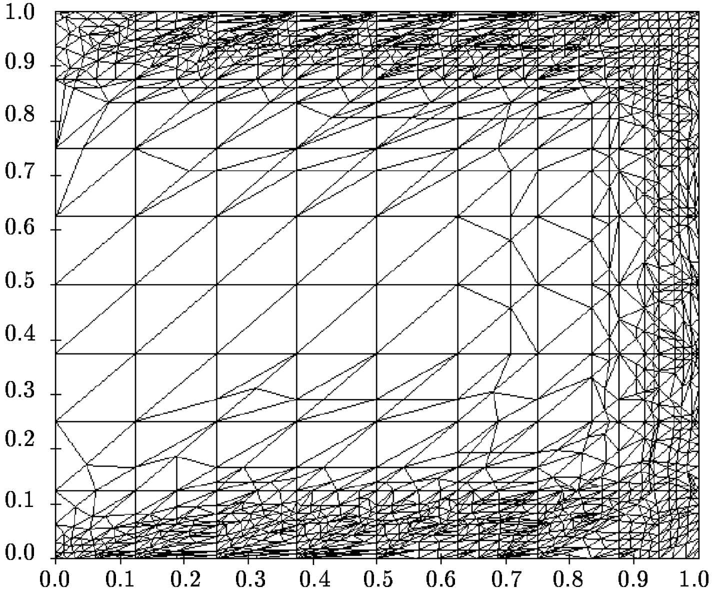


Fig. 9: Changes of grid at step 1; $\varepsilon = 0.05$.

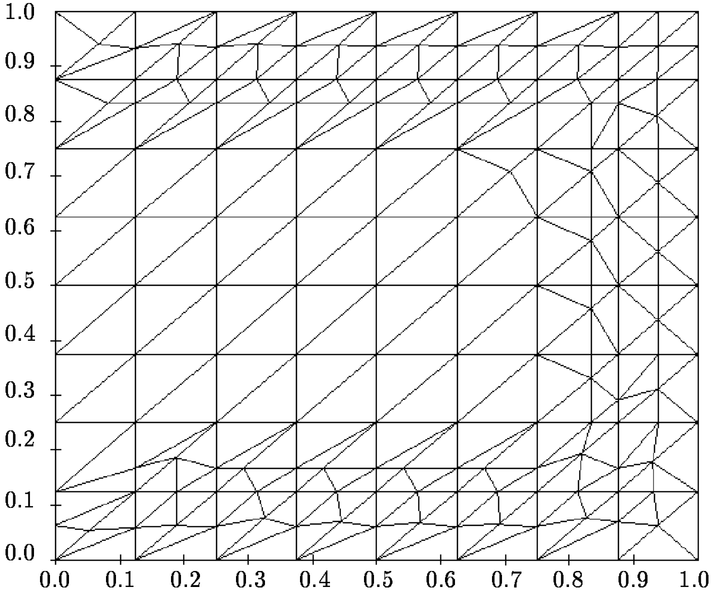


Fig. 10: Changes of grid at step 2; $\varepsilon = 0.05$.

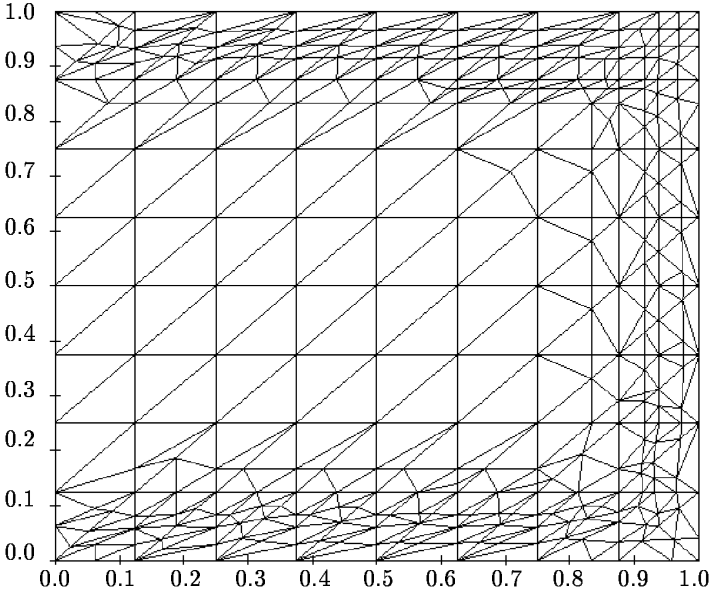


Fig. 11: Changes of grid at step 3; $\varepsilon = 0.05$.

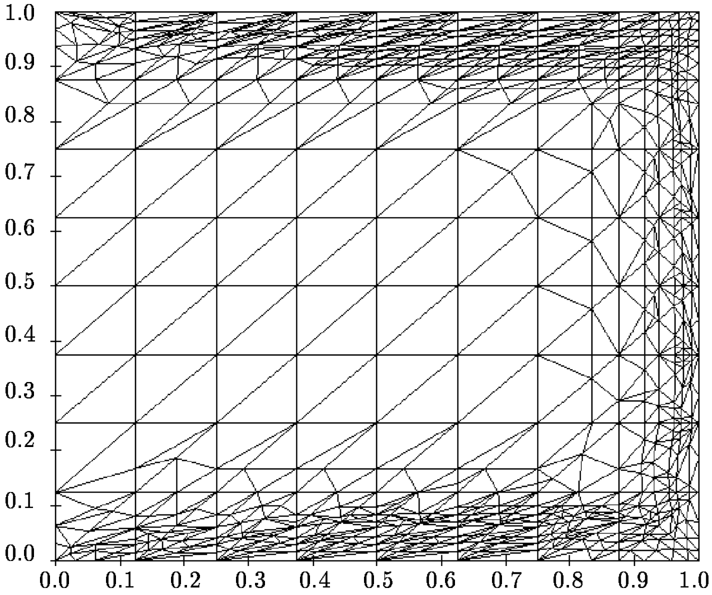


Fig. 12: Changes of grid at step 4; $\varepsilon = 0.05$.

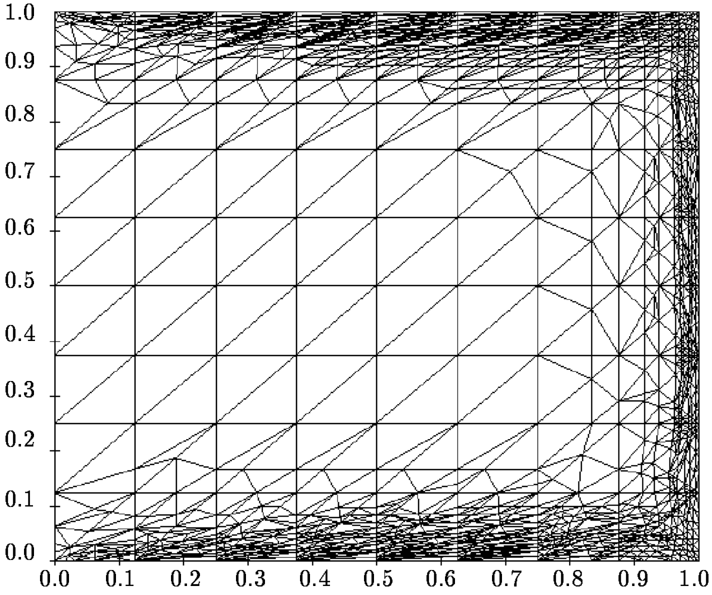


Fig. 13: Step 3; $\varepsilon = 0.05$ (vicinity of the origin).

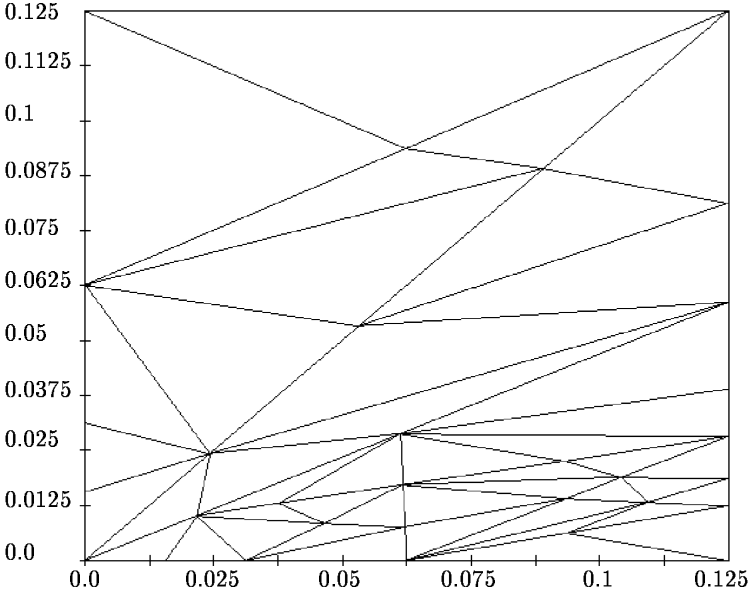


Fig. 14: Step 4; $\varepsilon = 0.05$ (vicinity of the origin).

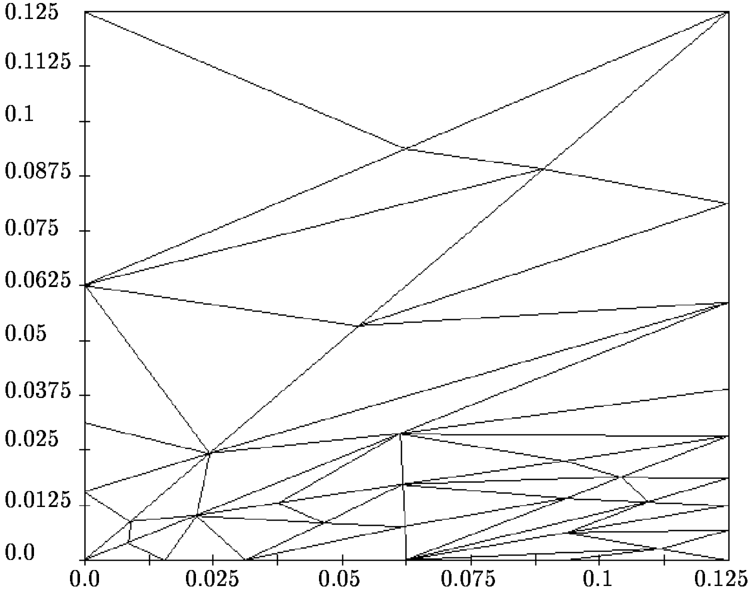


Fig. 15: Step 3; $\varepsilon = 0.05$ (at the bottom).

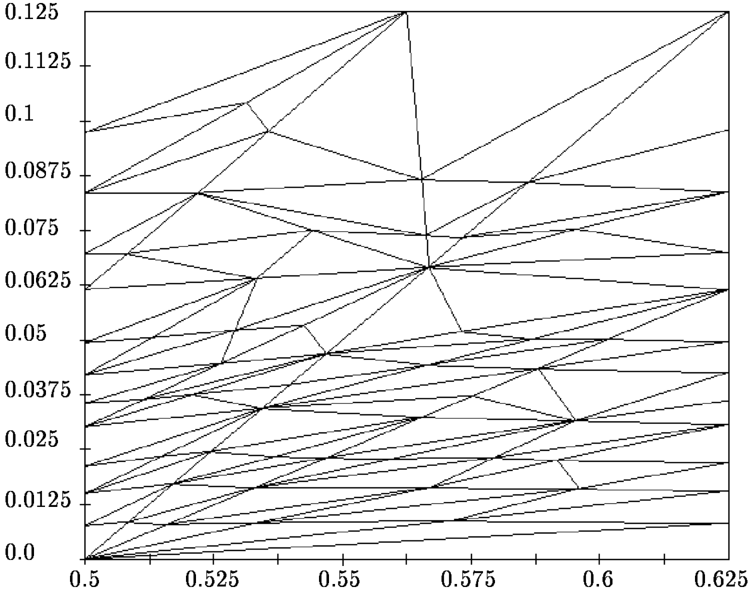


Fig. 16: Step 4; $\varepsilon = 0.05$ (at the bottom).

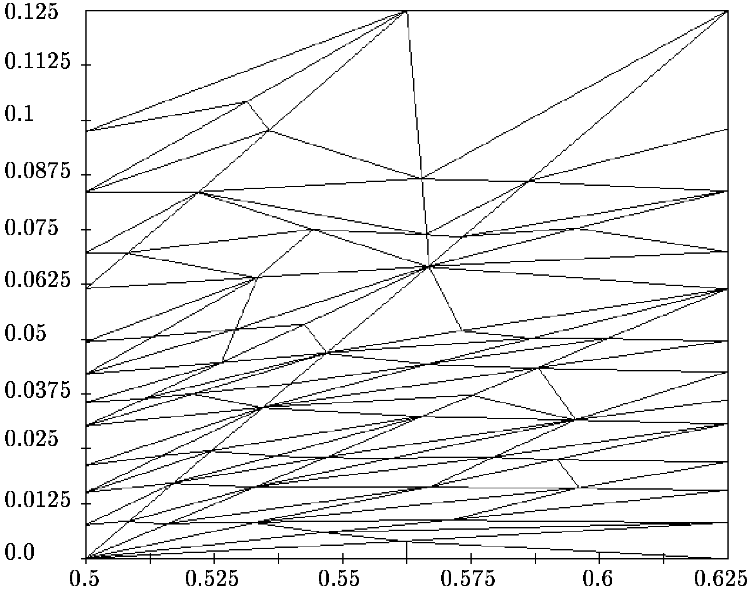


Fig. 17: Step 3; $\varepsilon = 0.05$ (the right-hand bottom corner).

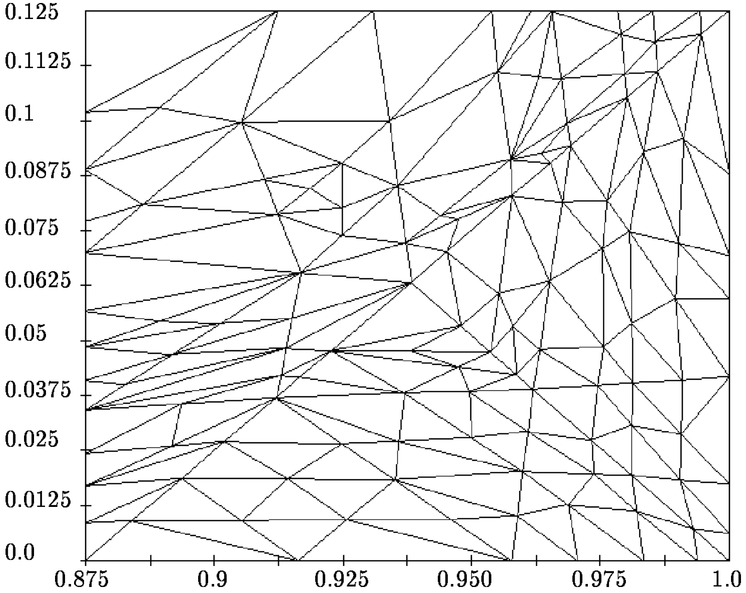


Fig. 18: Step 4; $\varepsilon = 0.05$ (the right-hand bottom corner).

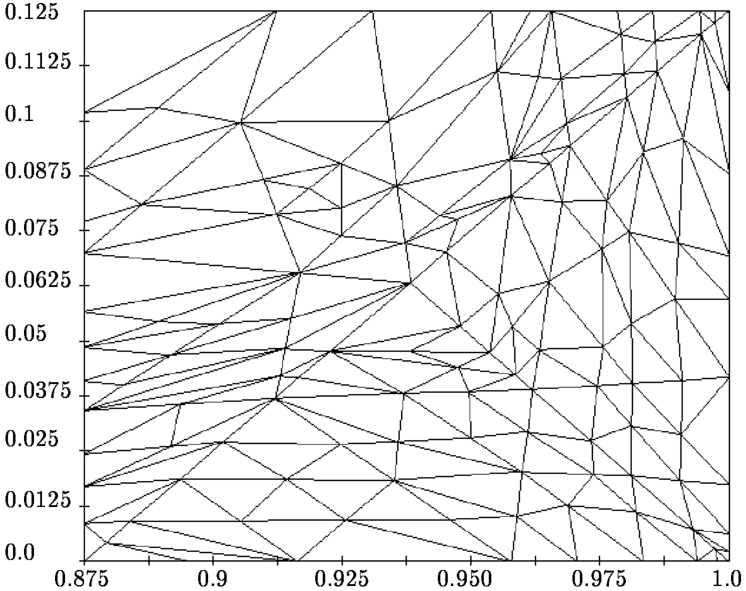


Fig. 19: Step 3; $\varepsilon = 0.05$ (the right-hand boundary).

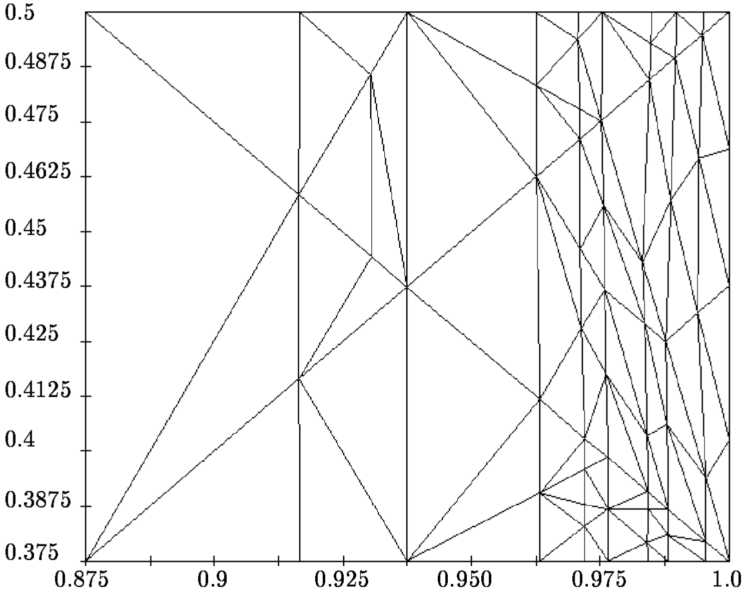
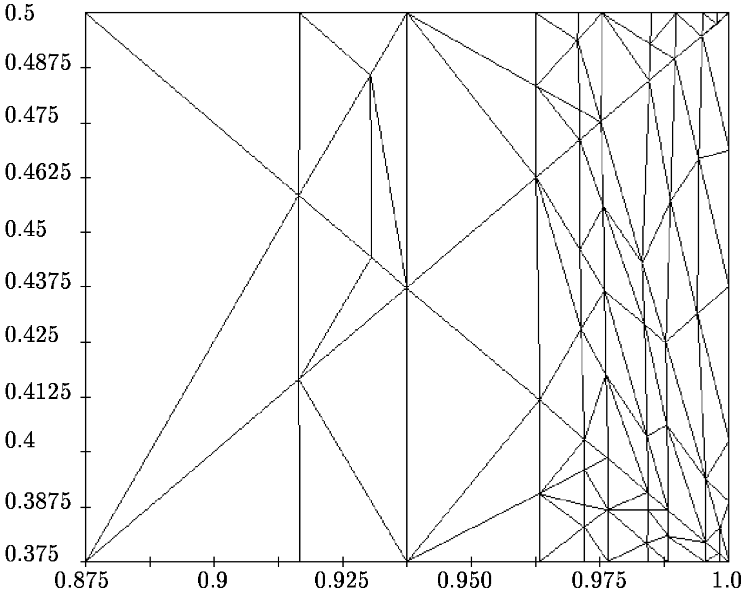


Fig. 20: Step 4; $\varepsilon = 0.05$ (the right-hand boundary).



A difference scheme for convection-diffusion problem on the oriented grid

Kalpush T.V., Shaidurov V.V.

Introduction

The work is devoted to a difference method for solving two-dimensional problem for convection-dominated convection-diffusion equation. This problem is related to the class of singular disturbed problems and it often has a solution of a boundary layer type with strong increase of derivatives in a vicinity of certain lines and points [1, 2, 3].

An application of the finite element method or difference methods for such problems has some specific features in comparison with the boundary value problem when convection and diffusion items have the same order. First, in zone of boundary layer it is necessary to take into consideration the boundary layer type of the solution [2] or to condense grid to compensate strong increase of derivatives [3]. Second, in zone of smoothness, when the influence of higher derivatives is low, we should take into account that the equation becomes the convection one (called here as reduced equation), while the area of solution dependence in points of this zone tends to a piece of reduced equation characterictic. Third, the standard difference schemes and the schemes of the finite element method with central differences lose a stability, while the schemes with directional differences possess computational diffusion which is essentially greater than the physical one and it disturbs even qualitative description of solution, not to mention the quantitative similarity. In contrast to the physical diffusion, the computational one differs both in various space points and in various directions in the same point. The role of the "longitudinal" computational diffusion, i.e., the diffusion along the convective flow, is already evidently seen in one-dimensional case, where it studied well and give the same consequences as in two-dimensional problems. In section 3, we define more precisely the influence of the "transversal" computational diffusion, which "washes-out" the difference solution in nontangent directions to convective flow. In certain difference schemes it essentially exceeds the physical diffusion, therefore to check it, we introduce the value, which is called the criterion of grid orientation along the convective flow.

In section 5, we state the algorithm of successive strengthening orientation for an arbitrary grid without new inner nodes addition and without node coordinates modification. In section 6, this algorithm is illustrated with an example of grids with uniform arrangement of nodes, but more and more oriented along the flow at the expence of changing stencil topology of difference scheme.

In section 4, we suggest the method of construction of inverse-monotone second-order finite-difference scheme. The combination of these properties is usually reached by special matching of the flow direction and the arrangement of grid nodes. Such, for example, is Crank-Nikolson scheme for convective term approximation with the arrangement of two nodes along the flow in the characterictic method. The use of the strengthening orientation algorithm provides this opportunity for arbitrary arrangement of the grid nodes.

1 The difference problem statement

Let us introduce Euclidean distance $|z - z'| = ((x - x')^2 + (y - y')^2)^{1/2}$ between two points $z = (x, y)$ and $z' = (x', y')$ in R^2. Let $\Omega = \{z = (x, y) : 0 < x < 1,\ 0 < y < 1\}$ be opened unit square with boundary Γ.

We shall use notation $C^k(\bar{D})$ in an arbitrary subdomain $D \subset \Omega$ for the class of functions having continuous k-th partial derivatives on closure $\bar{D}$ with the norm

$$||u||_{k,\bar{D}} = \max_{\alpha_1+\alpha_2 \leq k} \max_{\bar{D}} \left| \frac{\partial^{\alpha_1+\alpha_2} u}{\partial x^{\alpha_1} \partial y^{\alpha_2}} \right|$$

where α_1, α_2 are non-negative integers. Assume that $C^0(\bar{D}) = C(\bar{D})$.

Consider the problem

$$-\varepsilon \Delta u + b_1 \frac{\partial u}{\partial x} + b_2 \frac{\partial u}{\partial y} = f \quad \text{in} \quad \Omega, \tag{1.1}$$

$$u = g \quad \text{on} \quad \Gamma, \tag{1.2}$$

where $\varepsilon \ll 1$ is a small positive parameter; functions $b_1, b_2 \in C(\overline{\Omega})$ and the right-hand sides $f \in C(\overline{\Omega}), g \in C(\Gamma)$ are known. Thus, we have a solvable boundary value problem for the elliptic second-order equation [4].

In a subdomain, where the second derivatives are limited, their influence is low due to the small parameter ε. Therefore the equality (1.1) comes to the equation of first order, which characterictic system of ordinary differential equations corresponds

$$\frac{dx}{b_1(x,y)} = \frac{dy}{b_2(x,y)} = \frac{du}{f(x,y)}. \tag{1.3}$$

Its solution is the set of characteristic curves or simply the characteristic. In each points $z = (x, y) \in \overline{\Omega}$ the vector $t(z) = (b_1(z), b_2(z))$ touches the characteristic passing through this point. Therefore, we call it as characteristic vector, while the opposite vector as anticharacteristic one. We assume that a direction is a corresponding vector of unit length. In particular, the direction $(b_1^2 + b_2^2)^{-1/2}(b_1, b_2)$ with $b_1^2 + b_2^2 \neq 0$ in point (x, y) is called as characteristic direction, while any other direction, that does not coincide with it or with the opposite one is called as direction that "transversal" to characteristic one.

2 The difference approximation of convective item on an arbitrary trianqular stencil

Consider the triangle with vertices

$$z_t = (x_t, y_t), \quad z_s = (x_s, y_s), \quad z_r = (x_r, y_r),$$

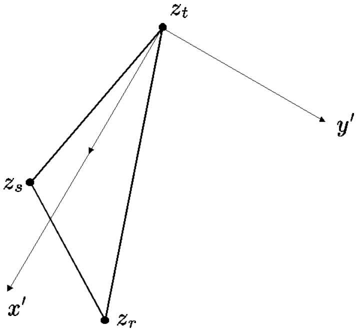


Fig. 1: The triangular stencil and the new local coordinates.

at a distance not greater that h from each other and not lying on one straight line. It is suppose that $b_1^2(z_t) + b_2^2(z_t) \neq 0$ and anticharacteristic vector $-\,t(z_t)$ lies in angle $\angle z_s z_t z_r$ (see Fig. 1).

Let construct the following approximation for this tree-point stencil with the help of indefinite coefficient method [12] :

$$b_1 \frac{\partial u}{\partial x} + b_2 \frac{\partial u}{\partial y} \approx \alpha u(z_t) + \beta u(z_s) + \gamma u(z_r) \tag{2.1}$$

in node z_t. Suppose that u belongs to $C^3(B(z_t,h))$ in the closed ball $B(z_t,h) = \{z : |z - z_t| \leq h\}$.

To simplify the problem, we introduce new local Cartesian coordinates (x', y') with the origin in z_t and with axis Ox' along $-\underline{t}(z_t)$ (see Fig. 1). The vector $\overline{b} = (b_1, b_2)$ in new coordinates comes to $\overline{b'} = (b'_1, b'_2)$ with coordinates $b'_1 = (b_1^2 + b_2^2)^{1/2}$, $b'_2 = 0$. Points $z_s = (x_s, y_s)$, $z_r = (x_r, y_r)$, $z_t = (x_t, y_t)$ comes to $z'_s = (x'_s, y'_s)$, $z'_r = (x'_r, y'_r)$, $z'_t = (0,0)$, and function $u(z)$ does in $\tilde{u}(z')$ respectively. The item in the right-hand side (2.1) comes to $(\,b_1^2 + b_2^2\,)^{1/2}\, \partial\tilde{u}/\partial x'$. Let take Taylor series with respect to $z'_t = (0,0)$ for function $\tilde{u}(z')$, summate them and $\tilde{u}(z'_t)$ with indefinite weights α, β, γ :

$$\begin{aligned} &\alpha u(z_t) + \beta u(z_s) + \gamma u(z_r) = (\alpha + \beta + \gamma)\, \tilde{u}(z'_t) \\ &+ \beta x'_s + \gamma x'_r)\, \frac{\partial \tilde{u}}{\partial x}(z'_t) + (\beta y'_s + \gamma y'_r)\, \frac{\partial \tilde{u}}{\partial y}(z'_t) \\ &+ c'_1\, \frac{\partial^2 \tilde{u}}{\partial x^2}(z'_t) + c'_2\, \frac{\partial^2 \tilde{u}}{\partial x \partial y}(z'_t) + c'_3\, \frac{\partial^2 \tilde{u}}{\partial y^2}(z'_t) + O(h^3) \end{aligned} \tag{2.2}$$

where

$$\begin{aligned} c'_1 &= \frac{1}{2}\beta\, {x'_s}^2 + \frac{1}{2}\gamma\, {x'_r}^2, \\ c'_2 &= \beta\, x'_s y'_s + \gamma\, x'_r y'_r, \\ c'_3 &= \frac{1}{2}\beta\, {y'_s}^2 + \frac{1}{2}\gamma\, {y'_r}^2. \end{aligned} \tag{2.3}$$

Here, we can distinctly see the computational diffusion $c'_1\, \partial^2\tilde{u}/\partial {x'}^2$ along the characteristic line, diffusion $c'_3\, \partial^2\tilde{u}/\partial {y'}^2$ in perpendicular direction, and diffusion $c'_2\, \partial^2\tilde{u}/\partial x' \partial y'$ in some intermediate directions.

Since h is small enough, in order to get at least the first order of approximation, we need the following equalities:

$$\begin{aligned} \alpha + \beta + \gamma &= 0, \\ \beta\, x'_s + \gamma\, x'_r &= b'_1, \\ \beta\, y'_s + \gamma\, y'_r &= 0. \end{aligned} \tag{2.4}$$

The matrix's determinant of this system equals double square of triangle $\triangle z_t z_s z_r$, which is denoted as S. Since the triangle does not degenerate into a line or a point, the determinant is not equal to zero and the system has the unique solution

$$\begin{aligned} \alpha &= (b_2(x_r - x_s) - b_1(y_r - y_s))/S, \\ \beta &= (b_1(y_r - y_t) - b_2(x_r - x_t))/S, \\ \gamma &= (b_2(x_s - x_t) - b_1(y_s - y_t))/S, \end{aligned} \tag{2.5}$$

where $S = (y_r - y_t)(x_s - x_t) - (x_r - x_t)(y_s - y_t) = y'_r x'_s - x'_r y'_s$.

Using β and γ in (2.3), we obtain:

$$\begin{aligned} c'_1 &= \frac{y'_r {x'}^2_s - y'_s {x'}^2_r}{y'_r x'_s - x'_r y'_s}, \\ c'_2 &= \frac{y'_s y'_r}{y'_r x'_s - x'_r y'_s}(x'_s - x'_r), \\ c'_3 &= \frac{y'_s y'_r}{y'_r x'_s - x'_r y'_s}(y'_s - y'_r). \end{aligned} \tag{2.6}$$

Hence, to decrease the coefficients c'_2 and c'_3, we need to minimize the following value

$$Kr(z_t) = y'_s y'_r / S \tag{2.7}$$

which is called *the index of triangle's orientation $\Delta z_t z_z z_r$ in point z_t.*

It should be noted that if y'_s or y'_r is zero, then the approximating transversal diffusion equals zero. That is the case, for example, in the method of characteristics.

3 Construction of inverse-monotone second-order finite-difference scheme

To construct the difference scheme, we first introduce the discrete set Ω_h of nodes in Ω and the discrete set Γ_h of nodes on Γ. Assume that $\bar{\Omega}_h = \Omega_h \cup \Gamma_h$. For each node $z \in \Omega_h$ we form the subset N_z of some nearer nodes of $\bar{\Omega}_h$. Denote by h_z the local radius of this subset:

$$h_z = \max_{z' \in N_z} |z - z'| \sim h.$$

Let us take an arbitrary inner node $\bar{z} \in \Omega_h$ and introduce local orthogonal coordinates ξ, η with origin in $\bar{z}$, with axis $O\xi$ along $t(\bar{z})$ and axis $O\eta$

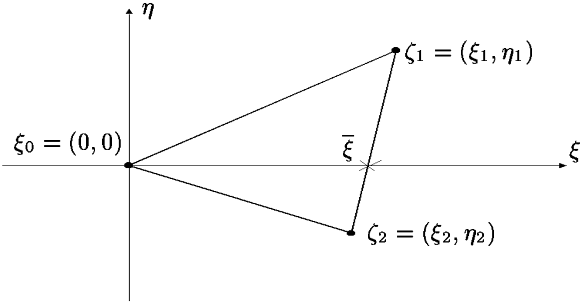


Fig. 2: The local coordinates (ξ, η) and the arrangement of nodes $\zeta_0, \zeta_1, \zeta_2$.

to the left of $t(\bar{z})$ (see Fig.2). In this coordinates equation (2.1) comes to another one:

$$-\varepsilon\tilde{\Delta}\tilde{u} - d\frac{\partial\tilde{u}}{\partial\xi} + \sigma\frac{\partial\tilde{u}}{\partial\eta} = \tilde{f} \tag{3.1}$$

in the h_z-vicinity of node z. Here for any function $w(x,y)$ we put

$$\tilde{w}(\xi,\eta) = w(x(\xi,\eta), y(\xi,\eta)) \tag{3.2}$$

and introduce new functions

$$d(\xi,\eta) = \frac{b_1(\bar{z})\tilde{b}_1(\xi,\eta) + \tilde{b}_2(\bar{z})b_2(\xi,\eta)}{|t(\tilde{z})|},$$

$$\zeta(\xi,\eta) = \frac{b_2(\bar{z})\tilde{b}_1(\xi,\eta) - b_1(\bar{z})\tilde{b}_2(\xi,\eta)}{|t(\tilde{z})|};$$

operator $\tilde{\Delta} = \partial^2/\partial\xi^2 + \partial^2/\partial\eta^2$ has the same form but in new coordinates.

Further we study two situations separately: $\varepsilon \leq c_0^{-2}h_{\bar{z}}^2$ and $c_1^{-2}h_{\bar{z}}^2 < \varepsilon$ with some constants c_0, c_1 independent of $\varepsilon, h_{\bar{z}}$. Let start with the first one.

3.1. Large $h_{\bar{z}}$. Tree-point stencil.

First situation means that

$$h_{\bar{z}} \geq c_0\sqrt{\varepsilon}. \tag{3.3}$$

Suppose that u belongs to $C^3(B(\bar{z}, h_{\bar{z}}))$ in the closed ball $B(\bar{z}, h_{\bar{z}}) = \{z : |z - \bar{z}| \leq h_{\bar{z}}\}$ and has bounded norm

$$||\tilde{u}||_{3,B(0,h_{\bar{z}})} = ||u||_{3,B(\bar{z},h_{\bar{z}})} \leq c_2 \tag{3.4}$$

with constant c_2 independent of $h_{\bar{z}}$ and ε.

Our goal is to derive an equality

$$\alpha_0\tilde{u}(\zeta_0)+\alpha_1\tilde{u}(\zeta_1)+\alpha_2\tilde{u}(\zeta_2)=\tilde{f}(\zeta_0)+\beta_1\frac{\partial\tilde{f}}{\partial\xi}(\zeta_0)+\beta_2\frac{\partial\tilde{f}}{\partial\eta}(\zeta_0)+O(h_{\bar z}^2). \quad (3.5)$$

We consider special arrangement of nodes. It is supposed that $\zeta_0=(0,0)$; node ζ_1 lies in first quadrant: $\xi_1>0$, $\eta_1\geq 0$; and node ζ_2 lies in fourth one: $\xi_2>0$, $\eta_2\leq 0$. Let us take Taylor series in nodes ζ_1,ζ_2 with respect to ζ_0 for function $\tilde{u}$, $\tilde{f}$:

$$\begin{aligned}\tilde{u}(\zeta_i)=\tilde{u}(\zeta_0)+\xi_i\frac{\partial\tilde{u}}{\partial\xi}(\zeta_0)+\eta\frac{\partial\tilde{u}}{\partial\eta}(\zeta_0)+\frac{\xi_i^2}{2}\frac{\partial^2\tilde{u}}{\partial\xi^2}(\zeta_0)\\+\xi_i\eta_i\frac{\partial^2\tilde{u}}{\partial\xi\partial\eta}(\zeta_0)+\frac{\eta_i^2}{2}\frac{\partial^2\tilde{u}}{\partial\eta^2}(\zeta_0)+O(h_{\bar z}^3)\end{aligned} \quad (3.6)$$

and

$$\tilde{f}(\zeta_0)=-d(\zeta_0)\frac{\partial\tilde{u}}{\partial\xi}(\zeta_0)+O(h_{\bar z}^2), \quad (3.7)$$

$$\frac{\partial\tilde{f}}{\partial\xi}(\zeta_0)=-\frac{\partial d}{\partial\xi}(\zeta_0)\frac{\partial\tilde{u}}{\partial\xi}(\zeta_0)-d(\zeta_0)\frac{\partial^2\tilde{u}}{\partial\xi^2}(\zeta_0)+\frac{\partial^2\sigma}{\partial\xi}(\zeta_0)\frac{\partial\tilde{u}}{\partial\eta}(\zeta_0)+O(h_{\bar z}), \quad (3.8)$$

$$\frac{\partial\tilde{f}}{\partial\eta}(\zeta_0)=-\frac{\partial d}{\partial\eta}(\zeta_0)\frac{\partial\tilde{u}}{\partial\xi}(\zeta_0)-d(\zeta_0)\frac{\partial^2\tilde{u}}{\partial\xi\partial\eta}(\zeta_0)+\frac{\partial^2\sigma}{\partial\eta}(\zeta_0)\frac{\partial\tilde{u}}{\partial\eta}(\zeta_0)+O(h_{\bar z}). \quad (3.9)$$

Now use these decompositions in both sides of (3.5). In order to get at least first order of approximation, we need to cancel terms $\tilde{u},\partial\tilde{u}/\partial\xi,\partial\tilde{u}/\partial\eta$:

$$\alpha_0+\alpha_1+\alpha_2=0, \quad (3.10)$$

$$\alpha_1\xi_1+\alpha_2\xi_2=-d(\zeta_0)-\beta_1\frac{\partial d}{\partial\xi}(\zeta_0)-\beta_2\frac{\partial d}{\partial\eta}(\zeta_0), \quad (3.11)$$

$$\alpha_1\eta_1+\alpha_2\eta_2=\beta_1\frac{\partial\sigma}{\partial\xi}(\zeta_0)+\beta_2\frac{\partial\sigma}{\partial\eta}(\zeta_0). \quad (3.12)$$

Two more equalities follow from elimination of $\partial^2\tilde{u}/\partial\xi^2$, $\partial^2\tilde{u}/\partial\xi\partial\eta$:

$$\frac{1}{2}(\alpha_1\xi_1^2+\alpha_2\xi_2^2)=-d(\zeta_0)\beta_1, \quad (3.13)$$

$$\alpha_1\xi_1\eta_1+\alpha_2\xi_2\eta_2=-d(\zeta_0)\beta_2. \quad (3.14)$$

In principle, we get 5 linear equations for 5 unknowns. Later we shall see that $|\beta_1|,|\beta_2|$ are small enough and

$$\tilde{\zeta}=(\beta_1,\beta_2)\in\Delta\zeta_0\zeta_2\zeta_1. \quad (3.15)$$

It gives us a possibility to change the right-hand side in (3.11) and $-d(\zeta_0)$ in (3.13), (3.14) by $-d(\tilde{\zeta})$ without violation of the second order of approximation:

$$\alpha_1\xi_1 + \alpha_2\xi_2 = -d(\tilde{\zeta}), \tag{3.16}$$

$$\frac{1}{2}(\alpha_1\xi_1^2 + \alpha_2\xi_2^2) = -d(\tilde{\zeta})\beta_1, \tag{3.17}$$

$$\alpha_1\xi_1\eta_1 + \alpha_2\xi_2\eta_2 = -d(\tilde{\zeta})\beta_2. \tag{3.18}$$

Since $\sigma(\zeta_0) = 0$, the same modification may be done in (3.12) without violation of the second order of approximation:

$$\alpha_1\eta_1 + \alpha_2\eta_2 = \sigma(\tilde{\zeta}). \tag{3.19}$$

Equalities (3.10), (3.16), (3.19) give the system with respect to α_i with unique solution

$$\begin{aligned} \alpha_0 &= ((\eta_1 - \eta_2)d(\tilde{\zeta}) + (\xi_1 - \xi_2)\sigma(\tilde{\zeta}))/(2s_{21}), \\ \alpha_1 &= (\eta_2 d(\tilde{\zeta}) + \xi_2\sigma(\tilde{\zeta}))/(2s_{21}), & (3.20) \\ \alpha_2 &= (\eta_1 d(\tilde{\zeta}) - \xi_1\sigma(\tilde{\zeta}))/(2s_{21}), & (3.21) \end{aligned}$$

where $s_{21} = (\xi_2\eta_1 - \xi_1\eta_2)/2$ is the area of triangle $\Delta\zeta_0\zeta_2\zeta_1$. Due to (3.15) and equality $\sigma(\zeta_0) = 0$, the inequalities

$$|\sigma(\tilde{\zeta})| \le ch_{\bar{z}} \ll d(\tilde{\zeta}) \tag{3.22}$$

hold. Therefore when η_1 is comparable with ξ_1, i.e., $\eta_1 \sim \xi_1$, we get

$$\alpha_2 \le 0; \tag{3.23}$$

analogously from comparability of $|\eta_2|$ with ξ_2 it follows that

$$\alpha_1 \le 0. \tag{3.24}$$

Both previous inequalities involve

$$\alpha_0 \le 0. \tag{3.25}$$

It would give M-property of the difference operator in the left-hand side of (4.5).

Now let us use (3.20) in (3.17) and (3.18):

$$\beta_1 = \frac{1}{4s_{21}}((\eta_1\xi_2^2 - \eta_2\xi_1^2) + \xi_1\xi_2(\xi_2 - \xi_1)\sigma(\tilde{\zeta})/d(\tilde{\zeta})), \tag{3.26}$$

$$\beta_2 = \frac{1}{2s_{21}}((\xi_2 - \xi_1)\eta_1\eta_2 + \xi_1\xi_2(\eta_2 - \eta_1)\sigma(\tilde{\zeta})/d(\tilde{\zeta})). \tag{3.27}$$

From arrangement of ζ_i it follows that

$$0 \le \beta_1 \le \frac{1}{2}\max\{\xi_1, \xi_2\} \le h_{\bar{z}}/2, \tag{3.28}$$

$$\eta_2 \le \beta_2 \le \eta_1, \ |\beta_2| \le h_{\bar{z}}. \tag{3.29}$$

So, you see that β_1, β_2 are small enough and was found by unique way with (3.26), (3.27). After that one can find α_i from (3.20) with the help of equality

$$d(\tilde{\zeta}) = d(\beta_1, \beta_2).$$

Finally, in order to get second order of approximation we need coefficient A_{22} before $\partial^2\tilde{u}/\partial\eta^2$ in (3.6) to be small enough:

$$|A_{22}| = \left|\alpha_1\frac{\eta_1^2}{2} + \alpha_2\frac{\eta_2^2}{2}\right| \le c_3 h_{\bar{z}}^2. \tag{3.30}$$

Since A_{22} is positive, we need only

$$A_{22} = -\eta_1\eta_2(\eta_1 - \eta_2)d(\tilde{\zeta})/(4s_{21}) \le c_3 h_{\bar{z}}^2. \tag{3.31}$$

Let $\tilde{\zeta}$ is cross-point of edge ζ_1, ζ_2 with axis $O\xi$, then

$$s_{21} = \frac{1}{2}(\eta_1 - \eta_2)\tilde{\xi}. \tag{3.32}$$

Combining it with (3.31) we get

$$-\frac{\eta_1\eta_2}{2\tilde{\xi}}d(\tilde{\zeta}) \le c_3 h_{\bar{z}}^2. \tag{3.33}$$

In principle, $\tilde{\xi} \sim h_{\bar{z}}$. Therefore we need

$$-\eta_1\eta_2 \sim c_4 h_{\bar{z}}^3. \tag{3.34}$$

From the first sight it seems to be unusual since the left-hand side has only second order of smallness. But in the next section we shall describe an algorithm of grid reorientation which gives this inequality and (3.31) by regular way. Therefore we consider inequality (3.31) to be valid.

3.2. Small $h_{\bar{z}}$. Five-point stencil.

Second situation means that

$$c_1 h_{\bar{z}} < \sqrt{\varepsilon}. \tag{3.35}$$

Let us again try to get (4.5). But this time we need to keep in consideration more terms because ε is not $O(h_{\bar{z}}^2)$ now. Therefore instead of (4.7) we have

$$\tilde{f}(\zeta_0) = -d(\zeta_0)\frac{\partial \tilde{u}}{\partial \xi}(\zeta_0) - \varepsilon\frac{\partial^2 \tilde{u}}{\partial \xi^2}(\zeta_0) - \varepsilon\frac{\partial^2 \tilde{u}}{\partial \eta^2}(\zeta_0). \tag{3.36}$$

It gives (3.10), (3.11), (3.12), (3.14) to be the same, and (3.13) comes to the following:

$$\frac{1}{2}(\alpha_1\xi_1^2 + \alpha_2\xi_2^2) = -d(\zeta_0)\beta_1 - \varepsilon. \tag{3.37}$$

Let us repeat considerations (3.15) – (3.27). We obtain the same α_i from (3.20) and β_2 from (3.27). But we get another β_1 and A_{22} :

$$\beta_1 = \frac{1}{8s_{21}}(\eta_1\xi_2^2 - \eta_2\xi_1^2) - \varepsilon/d(\zeta_0), \tag{3.38}$$

$$A_{22} = -\eta_1\eta_2(\eta_1 - \eta_2)d(\tilde{\zeta})/(8s_{21}) - \varepsilon. \tag{3.39}$$

From arrangement of ζ_i it follows that

$$-\varepsilon/d(\zeta_0) \leq \beta_1 \leq \frac{1}{2}\max\{\xi_1, \xi_2\} - \varepsilon/d(\zeta_0). \tag{3.40}$$

Due to (3.35)

$$|\beta_1| \leq h_{\bar{z}}/d(\zeta_0). \tag{3.41}$$

So, β_1, β_2 are of order $O(h_{\bar{z}})$ and are found by unique way from (3.27), (3.38). After that, one can find α_i from (3.20) with the help of equality $d(\tilde{\zeta}) = d(\beta_1, \beta_2)$. As a result, we obtain

$$\begin{aligned}\alpha_0\tilde{u}(\zeta_0) + \alpha_1\tilde{u}(\zeta_1) + \alpha_2\tilde{u}(\zeta_2) = \tilde{f}(\zeta_0) + \beta_1\frac{\partial \tilde{f}}{\partial \xi}(\zeta_0)& \\ + \beta_2\frac{\partial \tilde{f}}{\partial \eta}(\zeta_0) + A_{22}\frac{\partial^2 \tilde{u}}{\partial \eta^2}(\zeta_0) + O(h_{\bar{z}}^2).&\end{aligned} \tag{3.42}$$

In principle, we can make first item in the right-hand side of (3.39) to be small enough due to algorithm of reorientation. For example, let us demand that

$$-\eta_1\eta_2(\eta_1 - \eta_2)d(\tilde{\zeta})/(8s_{21}) \leq \varepsilon. \tag{3.43}$$

It implies

$$-\varepsilon \leq A_{22} \leq 0. \tag{3.44}$$

In order to cancel item $A_{22}\partial^2\tilde{u}/\partial\eta^2$ in the right-hand side of (3.42), let us introduce one more triangle with vertices $\zeta_0, \zeta_3, \zeta_4$ (see for Fig. 3); node ζ_3

lies in third quadrant: $\zeta_3 < 0$, $\eta_3 \leq 0$; node ζ_4 lies in second one: $\zeta_4 < 0$, $\eta_4 \geq 0$. Consideration like (3.36) – (3.42) gives one more equality

$$\begin{aligned}\alpha'_0\tilde{u}(\zeta_0) + \alpha'_4\tilde{u}(\zeta_4) + \alpha'_3\tilde{u}(\zeta_3) = \tilde{f}(\zeta_0) + \beta'_1\frac{\partial \tilde{f}}{\partial \xi}(\zeta_0) \\ + \beta'_2\frac{\partial \tilde{f}}{\partial \eta}(\zeta_0) + A'_{22}\frac{\partial^2 \tilde{u}}{\partial \eta^2}(\zeta_0) + O(h^2_{\bar{z}})\end{aligned} \tag{3.45}$$

with coefficients

$$\alpha'_0 = -(\eta_4 - \eta_3)d(\tilde{\zeta}')/(4s_{43}), \tag{3.46}$$
$$\alpha'_4 = -\eta_3 d(\tilde{\zeta}')/(4s_{43}), \tag{3.47}$$
$$\alpha'_3 = \eta_4 d(\tilde{\zeta}')/(4s_{43}), \tag{3.48}$$
$$\beta'_1 = -(\eta_4\xi_3^2 - \eta_3\xi_4^2) - \varepsilon d(\zeta_0)/(8s_{43}), \tag{3.49}$$
$$\beta'_2 = -\eta_4\eta_3(\xi_3 - \xi_4)/(4s_{43}), \tag{3.50}$$
$$A'_{22} = \eta_4\eta_3(\eta_4 - \eta_3)d(\tilde{\zeta}')/(8s_{43}) - \varepsilon, \tag{3.51}$$

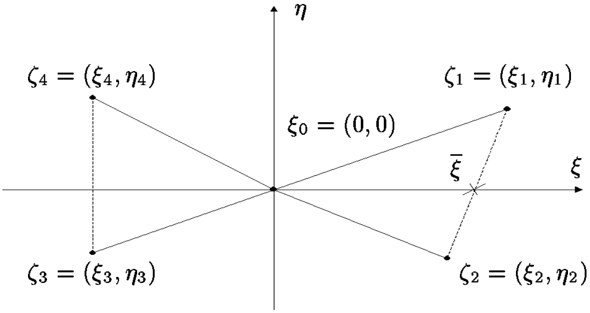

η
$\zeta_4 = (\xi_4, \eta_4)$
$\zeta_1 = (\xi_1, \eta_1)$
$\xi_0 = (0, 0)$
$\bar{\xi}$
ξ
$\zeta_3 = (\xi_3, \eta_3)$
$\zeta_2 = (\xi_2, \eta_2)$

Fig. 3: The local coordinates (ξ, η) and the arrangement of nodes $\zeta_0, ..., \zeta_4$.

where $s_{43} = (\eta_3\xi_4 - \eta_4\xi_3)/2$ is the area of triangle $\Delta\eta_0\eta_4\eta_3$. This time, coefficient A'_{22} consists of two negative items and

$$A'_{22} \leq -\varepsilon \tag{3.52}$$

due to arrangement of nodes ζ_4, ζ_3. Let us combine (3.42) and (3.45) with weights δ_1, δ_2 in order to cancel term $\partial^2\tilde{u}/\partial\eta^2(\zeta_0)$:

$$\delta_1 A_{22} + \delta_2 A'_{22} = 0. \tag{3.53}$$

For scaling we take also

$$\delta_1 + \delta_2 = 1. \tag{3.54}$$

This system gives unique solution

$$\begin{aligned} \delta_1 &= A'_{22}/(A'_{22} - A_{22}) > 0, \\ \delta_2 &= -A_{22}/(A'_{22} - A_{22}) \leq 0, \end{aligned} \tag{3.55}$$

when

$$A'_{22} \neq A_{22}. \tag{3.56}$$

The last is guaranteed when, for example,

$$\eta_3\eta_4 \neq 0 \quad \text{or} \quad \eta_1\eta_2 \neq 0. \tag{3.57}$$

Due to (3.52), (3.54) we get

$$\sum_{i=0}^{4} \alpha''_i \tilde{u}(\zeta_i) = \tilde{f}(\zeta_0) + \beta''_1 \frac{\partial \tilde{f}}{\partial \xi}(\zeta_0) + \beta''_2 \frac{\partial \tilde{f}}{\partial \eta}(\zeta_0) + O(h^2_{\bar{z}}) \tag{3.58}$$

where

$$\begin{aligned} \alpha''_0 = \delta_1\alpha_0 + \delta_2\alpha'_0 > 0,\ \alpha''_1 = \delta_1\alpha_1 \leq 0, \\ \alpha''_2 = \delta_1\alpha_2 \leq 0,\ \alpha''_3 = \delta_2\alpha'_3 \leq 0,\ \alpha''_4 = \delta_2\alpha'_4 \leq 0, \\ \beta''_1 = \delta_1\beta_1 + \delta_2\beta'_1,\ \beta''_2 = \delta_1\beta_2 + \delta_2\beta'_2. \end{aligned} \tag{3.59}$$

The signs of α''_i provide the inverse monotonicity of difference operator in the left-hand side of (3.58).

4 The algorithm for the orientation strengthening of the difference grid

Let us consider an arbitrary opened limited, and connected polygon $\Omega \in R^2$. We construct its triangulation $\mathcal{J}$, i.e., we cut this polygon into the finite number of opened triangles T_i, $i = 1, ..., m$, so that their closure $\overline{T}_i$ cover $\overline{\Omega}$:

$$\overline{\Omega} = \bigcup_{i=1}^{m} \overline{T}_i. \tag{4.1}$$

This triangulation should be consistent, i.e., any two different closed triangles $\overline{T_i}$ and $\overline{T_j}$ from $\mathcal{J}$, $i \neq j$, either have no common points, or only one common vertex, or have the whole common side.

Let us denote by $\overline{\Omega}_h$ a set of all vertices of triangulation triangles, which are called by nodes. Suppose that

$$\Omega_h = \overline{\Omega}_h \cap \Omega, \quad \Gamma_h = \overline{\Omega}_h \cap \Gamma. \tag{4.2}$$

Our goal is to describe the algorithm of triangles reconstruction in order to decrease the computational diffusion across characteristic lines of difference analogue to necessary limits. In section 3, we introduce a special local value to control it.

There are many ways of grid construction of different complexity. The grid are condensing in the required subdomains or oriented with some method. But all of them are connected either with the new nodes addition, or with the inner nodes coordinates modification.

We propose the algorithm that does not change the coordinates of inner nodes, but it makes better the desired quality of triangulation due to reconnection of the nodes among themselves.

Now, we consider the initial consistent triangulation $\mathcal{J}'$. One of the algorithm cycles consists of step-by-step sorting out of inner apexes $z_i \in \mathcal{J}' \cap \Omega$, $i = 1, ..., n$, by means of possible triangles reconstruction. Let us describe one step of this algorithm.

Let i be inner node $z_i = (x_i, y_i)$ of the consistent triangulation $\mathcal{J}'$ with anticharacteristic vector $-t(z_i)$, which we reconstruct to perform the inequality

$$\max_{z_i \in \mathcal{J}' \cap \Omega} Kr(z_i) \leq \delta \tag{4.3}$$

with some constant δ.

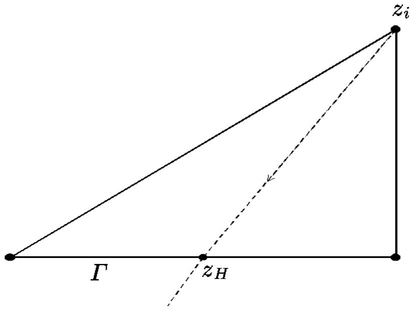


Fig. 4: The anticharacteristic direction crossing the boundary.

1. If this vector is directed along one of the triangle sides, with the origin in this vertex, then local criterion ($K_r(z_i) = 0$) is considered to be valid and we complete the step without changing the triangulation, i.e., the result $\mathcal{J}''$ of this step coincides with $\mathcal{J}'$.

2. If this coincidence (which is unlikely in real problems) does not take place, then there is triangle $T_k \in \mathcal{J}'$ with vertex z_i, for which vector $-t(z_i)$ is enclosed between its sides. Let construct a ray in this direction to cross a side of this triangle, which is opposite to vertex z_i.

Futher, there are two variants.

2.1) The triangle side crossed lies on the boundary Γ (Fig. 4). In this case we add a new node z_H to Γ_h , which is the intersection point of the constructed ray and boundary Γ. In this case we obtain the ideal situation, $Kr(z_i) = 0$.

2.2) The triangle side cross is the inner one (Fig. 5). Since triangulation is consistent, there exists one more triangle with the same side.

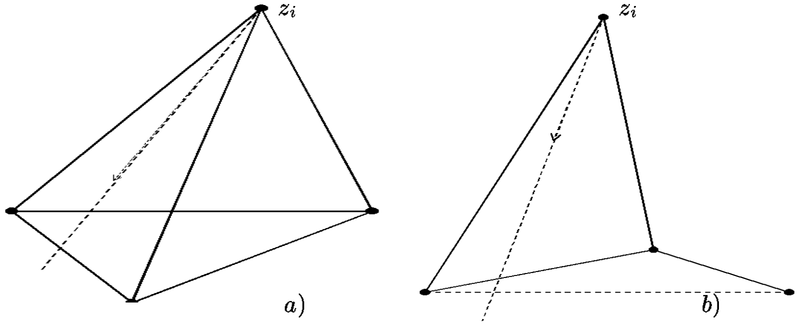


Fig. 5: The anticharacteristic direction crossing the inner side.

Further, there are two variant as well.

2.2.1) The obtained quadrangle is convex (Fig. 5.a). From two available variants we choose such that gives criterion $Kr(z_i)$ to be smaller.

2.2.2) The obtained quadrangle is not convex (Fig. 5.b). Then we complete the step without changing the triangulation.

In this way, the process is periodically repeated for all nodes z of Ω_h where $Kr(z) > \delta$. It should be pointed out that the effectiveness of algorithm will be better, if we move forward by front through inner nodes along the convective flow.

5 The numerical experiment

For the numerical experiment we considered problem (2.1) – (2.2) with coefficients $b_1 = -1$, $b_2 = 0.7$. The function g is equal to zero on the boundary Γ except for two sections

$$\Gamma_1 = \{(x, y) : x = 0,\ y \in [35/40, 39/40]\}$$

and

$$\Gamma_2 = \{(x, y) : x = 1,\ y \in [1/40, 5/40]\}$$

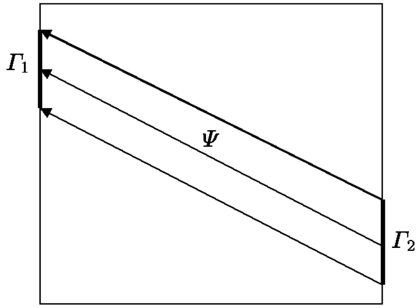


Fig. 6: The characteristics of the reduced equation.

(see Fig. 6), where g equals 1. The right-hand side is identically equal to zero on the Ω. For $\varepsilon = 0$ the exact solution of reduced problem is the function u_0 which is equal to 1 in band $\overline{\Psi}$ and 0 outside it (see Fig. 7, 8). The band $\overline{\Psi}$ represents the parallelogram with sides Γ_1 and Γ_2.

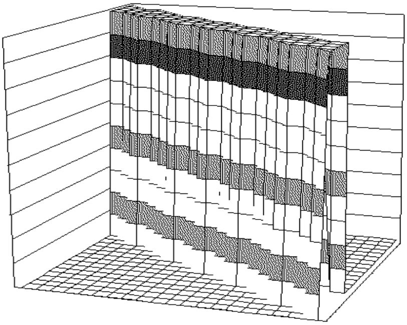


Fig. 7: The exact solution.

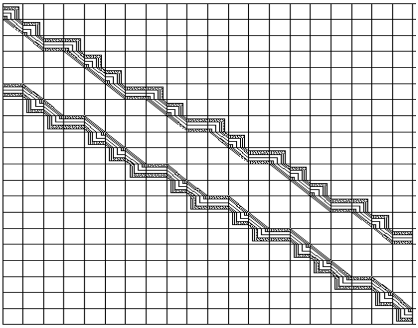


Fig. 8: The isolines of exact solution.

In square $\Omega = \{ (x, y) : 0 < x < 1, 0 < y < 1\}$ we build the uniform triangulation with the mesh-size $h = 1/n$ by means of two families of lines $x_i = ih$, $y_j = jh$, $i, j = 1, ..., n-1$, and then construct the diagonals in the obtained elementary squares with angle $\pi/4$ to axis Ox.

Then we build the grid aproximation in the following way. To aproximate items Δu we use on the uniform five-point stencil "cross". As the aproximation of item $b_1\, \partial u/\partial x + b_2\, \partial u/\partial y$, we realize it on the constructed triangulation.

This triangulation is unsuccessful (Fig. 9) in term of the value of orientation. Solving the problem for $n = 40$ with this triangulation, we do not

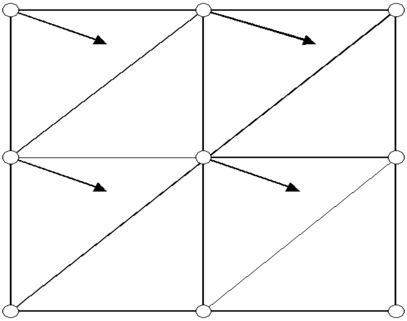


Fig. 9: The inital triangulation.

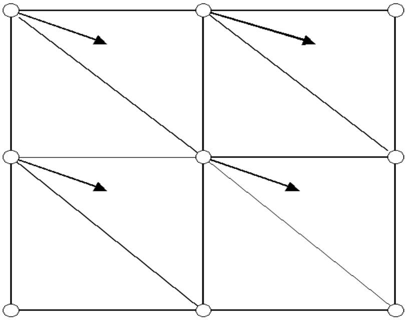


Fig. 10: The grid after the first reconstruction.

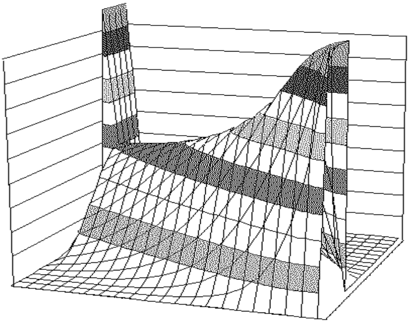


Fig. 11: The numerical solution on the inital grid.

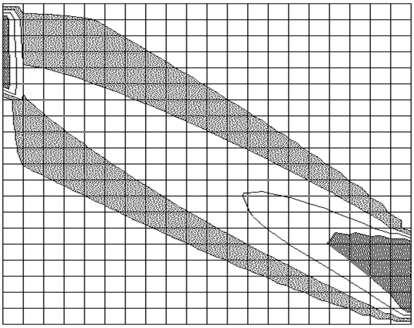


Fig. 12: The isolines of numerical solution on the inital grid.

obtain even the qualitative similarity solution. The considerable "transversal" calculation diffusion appeares which washes out the solution (Fig. 12), and obtained error equals 60% (Fig. 11).

Further the first reconstruction of grid is made, which we implement according to section 5. It only reorients some diagonals without coordinates modification of inner nodes (Fig. 10). Solving again the problem for $n = 40$ with this triangulation, we obtain the considerable improvement of the solution quality. The essential decrease of computing diffusion took place (Fig. 14), and the obtained error equals 20% (Fig. 13).

After the second application of the reorientation algorithm the new nodes on the boundary of domain appear, which do not involve the increase of the unknown values in consequence of known boundary conditions. Apart from that, the recombination of inner grid nodes with each other (Fig. 15) consequently decreases $Kr(z_i)$ in every inner nodes. The obtained

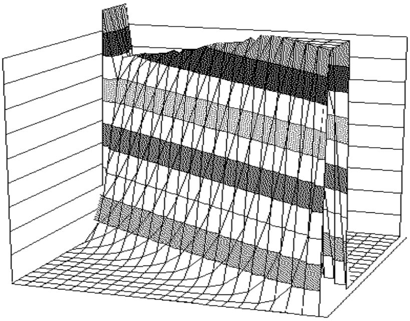


Fig. 13: The numerical solution after the first grid reconstruction.

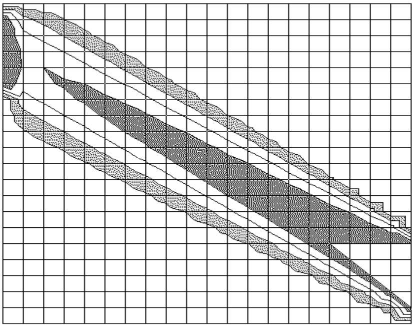


Fig. 14: The isolines of numerical solution after the first grid reconstruction.

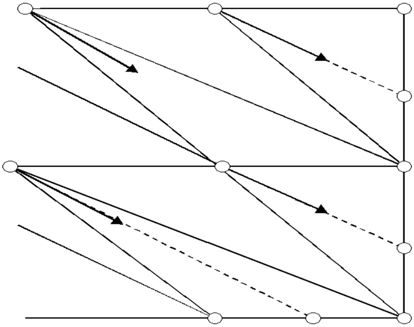


Fig. 15: The grid after the second reconstruction.

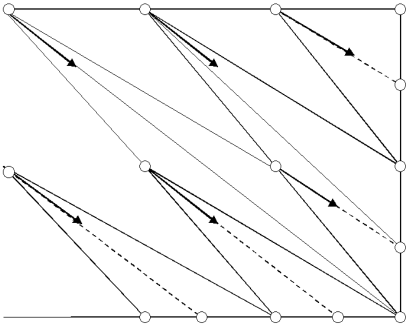


Fig. 16: The grid after the third reconstruction.

error is not greater than 10% (Fig. 17) and we note some decrease of the solution wash-out (Fig. 18).

After the third reconstruction of grid (Fig. 16), we obtain the considerable improvement of solution. Apart from similar qualitative behavior of the solution (Fig. 19), we also obtain good quantitative similarity.

Thus, this numerical experiment illustrates the successive improvement of numerical solution on first three stages of the grid reconstruction due to strengthening of the orientation along the characteristic curves.

References

1. Vishik M.I., Lyusternik L.A.: *Regular degeneration and boundary layer for linear differential equations with a small parameter.* AMS Translations, 1975.

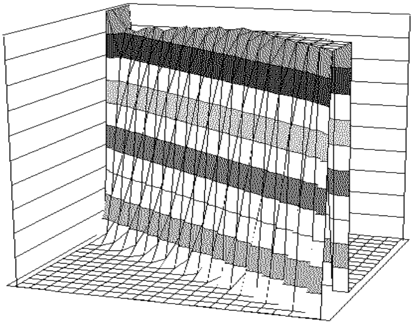


Fig. 17: The numerical solution after the second grid reconstruction.

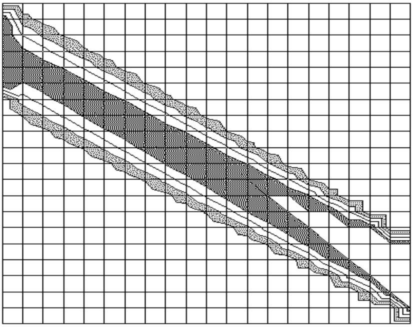


Fig. 18: The isolines of numerical solution after the second grid reconstruction.

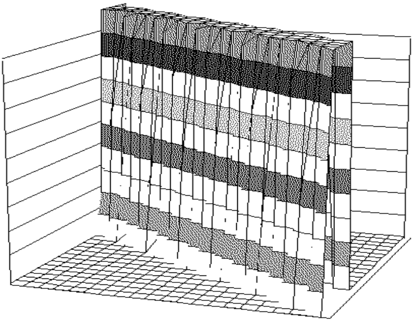


Fig. 19: The numerical solution after the third grid reconstruction.

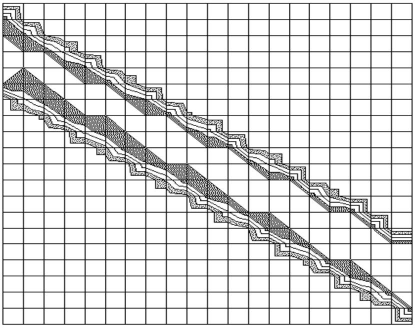


Fig. 20: The isolines of numerical solution after the third grid reconstruction.

2. Doolan E.P., Miller J.J.H., Schilders W.H.A.: *Uniform numerical methods for problem with inital and boundary layers.* Boole Press, Dublin, 1980.
3. Liseikin V.D., Petrenko V.E.: *Adaptive invariant method for numerical solution of problem with boundary and inner layers.* Novosibirsk, 1989 (in Russian).
4. Ladyzhenskaya O.A., Uraltseva N.N.: *Linear and quasilinear equations of elliptic type.* Moscow, Nauka, 1973 (in Russian).
5. Miller J.J.H., O'Riordan E., Shishkin G.I.: *Solution of singularly perturted problems with ε-uniform numerical methods – introduction to the theory of linear problems in one and two dimensions.* World Scientific, 1995.
6. Bagaev B.M., Shaidurov V.V.: *Variation-difference solution of equation with a small parameter.* In: Differential and Integral-Differential equations, Novosibirsk, Nauka, 1977, pp. 89–99 (in Russian).
7. Shaidurov V.V., Tobiska L.: *Special integration formulae for a convection-diffusion problem.* East-West J. Numer. Math., 1995, vol .3, №. 4, pp. 281–

299.

8. Voevodin V.V., Kuznetsov Yu.A.: *Matrices and computations.* Moscow, Nauka, 1984 (in Russian).
9. Samarskii A.A.: *Inroduction in difference schemes theory.* Moscow, Nauka, 1971 (in Russian).
10. Shokin Yu.I.: *Method of differential approximation.* Novosibirsk, Nauka, 1979 (in Russian).

A two-dimensional nonuniform difference scheme with higher order of accuracy

Bykova E.G., Shaidurov V.V.

Introduction

The present paper is devoted to construction and justification of *nonuniform* difference schemes of higher orders of accuracy for two-dimensional boundary-value problem for elliptic type equation on a rectangle. The general idea of construction of such scheme is similar to that in the paper [1], where it is stated for ordinary differential equation, but the increase of dimensionality has complicated both the scheme and the proof of its accuracy. Nevertheless, the fourth order of accuracy in uniform norm is proved for the constructed scheme, and this fact is illustrated with numerical examples.

As it is in one-dimensional case, the difference scheme is similar in structure to the system of the method of extrapolated equations by U. Rüde [2] for finite elements. However, the proof of accuracy of the constructed scheme differs from substantiation of U. Rüde method based on minimization of functional.

Let recall that the standard difference method with the second order of accuracy on a rectangle gives a system of linear algebraic equations with five-diagonal matrix under corresponding ordering of unknowns. The scheme constructed here results in a system of equations with nine-diagonal matrix preserving the basic properties: positive definiteness, symmetry and positive invertibility.

Let also recall that the term "nonuniform scheme" had appeared due to different rules of construction of grid equations in neighbouring nodes as distinct from uniform schemes [3], where the rule of construction is the same for all nodes of the grid.

1 Boundary-value problem and its nonuniform difference approximation

Let Ω be unit square $(0,1) \times (0,1)$ with boundary Γ. Consider a boundary-value problem

$$-\Delta u + du = f \ \text{ in } \ \Omega, \tag{1.1}$$

$$u = g \ \text{ on } \ \Gamma \tag{1.2}$$

with smooth enough given functions

$$d, f \in C^4(\overline{\Omega}), \tag{1.3}$$

$$d \geq 0 \quad \text{in} \quad \overline{\Omega}. \tag{1.4}$$

These conditions ensure unique solvability of the problem. Suppose the solution to be smooth enough:

$$u \in C^6(\overline{\Omega}). \tag{1.5}$$

For difference approximation of the problem (1.1) — (1.2) construct an uniform difference grid

$$\overline{\omega}_h = \{z_{i,j} = (x_i, y_j) : \ x_i = ih, \ y_j = jh, \ i = 0,1,\dots,n, \ j = 0,1,\dots,n\}$$

with the step $h = 1/n$ and *even* $n \geq 4$. Also, introduce the set of inner nodes

$$\omega_h = \{z_{i,j} \in \overline{\omega}_h \ : \ i = 1,2,\dots,n-1, \ j = 1,2,\dots,n-1\}$$

and divide it into the sets of nodes only with even indices, only with odd indices and with indices of different evenness (the first index is even and the second is odd, or vice versa):

$$\begin{aligned}
&\overline{\omega}_{00} = \{z_{i,j} \in \overline{\omega}_h : \ i = 0,2,\dots,n, \ j = 0,2,\dots,n\}, \ \omega_{00} = \overline{\omega}_{00} \setminus \Gamma,\\
&\omega_{11} = \{z_{i,j} \in \omega_h : \ i = 1,3,\dots,n-1, \ j = 1,3,\dots,n-1\},\\
&\overline{\omega}_{01} = \{z_{i,j} \in \overline{\omega}_h : \ i = 0,2,\dots,n, \ j = 1,3,\dots,n-1\}, \ \omega_{01} = \overline{\omega}_{01} \setminus \Gamma,\\
&\overline{\omega}_{10} = \overline{\omega}_h \setminus (\overline{\omega}_{00} \cup \omega_{11} \cup \overline{\omega}_{01}), \ \omega_{10} = \overline{\omega}_{10} \setminus \Gamma.
\end{aligned}$$

The standard finite difference approximation of the equation (1.1) consists in change of the second derivatives with respect to x and y with the second central differences

$$\begin{aligned}
u_{\overset{\circ}{x}\overset{\circ}{x}}(x,y) &= (u(x-h,y) - 2u(x,y) + u(x+h,y))/h^2,\\
u_{\overset{\circ}{y}\overset{\circ}{y}}(x,y) &= (u(x,y-h) - 2u(x,y) + u(x,y+h))/h^2.
\end{aligned} \tag{1.6}$$

As a result, the following grid problem is obtained:

$$\begin{aligned} L^h u^h &= f \quad \text{in} \quad \omega_h, \\ u^h &= g \quad \text{on} \quad \gamma_h = \Gamma \cap \overline{\omega}_h, \end{aligned} \tag{1.7}$$

with the difference operator

$$L^h v(z) = -v_{\overset{\circ}{x}\overset{\circ}{x}}(z) - v_{\overset{\circ}{y}\overset{\circ}{y}}(z) + d(z)v(z). \tag{1.8}$$

The second order of approximation is established by Taylor-series expansion of the solution u [3], and on the basis of difference maximum principle [3] the stability of solution in the grid norm

$$\|v\|_{\infty,\bar{\omega}_h} = \max_{z\in\overline{\omega}_h} |v(z)|$$

is proved. On the whole, this gives convergence of the approximate solution u^h of the problem (1.7) to the exact solution u of the problem (1.1) – (1.2) with the second order of accuracy:

$$\|u^h - u\|_{\infty,\overline{\omega}_h} \le c_1 h^2 \|u\|^{(4)}_{\infty,\overline{\Omega}} \quad {}^{*)} \tag{1.9}$$

here the following denotation is used:

$$\|u\|^{(k)}_{\infty,\overline{\Omega}} = \sum_{0\le i+j\le k} \left\| \frac{\partial^{i+j} u}{\partial x^i \partial y^j} \right\|_{\infty,\overline{\Omega}}$$

with integer $k \ge 0$ and

$$\|u\|_{\infty,\overline{\Omega}} = \sup_{\overline{\Omega}} |u|.$$

For construction of a scheme of the fourth order introduce an operator with doubled step

$$\begin{aligned} L^{2h} v(x,y) \;&= -(v(x-2h,y) + v(x,y-2h) - 4v(x,y) \\ &\quad + v(x+2h,y) + v(x,y+2h))/4h^2 + d(x,y)v(x,y) \end{aligned}$$

only in even nodes ω_{00}.

With the preceding notations consider the difference problem

$$L^h u^h = f \quad \text{in} \quad \omega_h \setminus \omega_{00}, \tag{1.10}$$

$$L^h u^h - L^{2h} u^h = 0 \quad \text{in} \quad \omega_{00}, \tag{1.11}$$

$$u^h = g \quad \text{on} \quad \gamma_h. \tag{1.12}$$

*) Here and below we denote by a symbol c_i with integer indices i various constants independent of x and h.

This grid problem as well as (1.7) contains $(n+1)^2$ unknowns and $(n+1)^2$ equations. In even nodes nine-point stencil is obtained (Fig. 1. b), and in other nodes the scheme has a standard five-point stencil (Fig. 1. a).

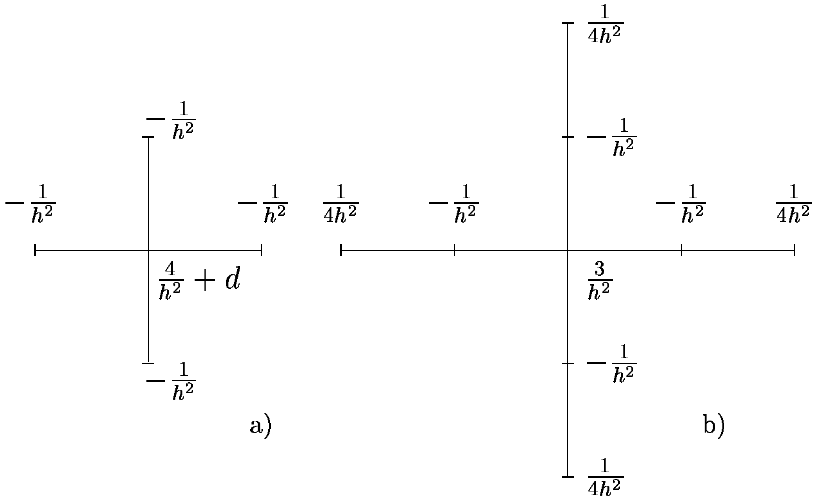


Fig. 1: Stencils of nonuniform difference scheme in even (b) and other (a) nodes.

For the functions defined on $\overline{\Omega}$ apply the denotation

$$v_{i,j} = v(x_i, y_j) = v(ih, jh).$$

In the equations (1.10) — (1.11) eliminate the boundary values (1.12). The remaining unknowns and equations number from 1 to $(n-1)^2$ in lexicographical order determined by the inner nodes $z_{1,1}$, $z_{1,2}, \ldots, z_{1,n-1}$, $z_{2,1}, \ldots, z_{n-1,n-1}$. As a result we obtain a system of linear algebraic equations with symmetric sparse matrix A^h

$$A^h U^h = F^h. \tag{1.13}$$

By way of illustration in Fig. 2 the structure of nonzero elements of the matrix A^h for the step h=1/8 is given.

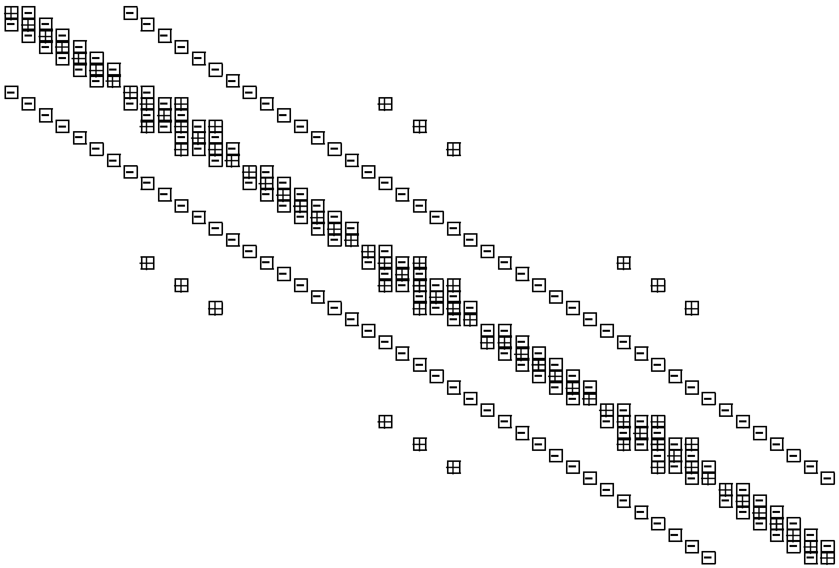


Fig. 2: Structure of nonzero elements of the matrix $A^{1/8}$.
The sign ⊞ marks a positive element,
the sign ⊟ marks negative one, and their absence implies zero element.

For theoretical consideration it is useful to write down the system (1.10)—(1.12) in vector form as well. To do this, number unknowns and equations from 1 to $(n+1)^2$ in lexicographical order determined by the nodes z_{00}, $z_{01}, \dots, z_{0n}, z_{10}, \dots, z_{nn}$. As a result, we obtain a system of linear algebraic equations with a matrix B^h

$$B^h V^h = G^h. \tag{1.14}$$

2 Stability and solvability of the grid problem

Let proof that matrix of the system (1.13) is positive definite.

Theorem 32. *If the condition* (1.4) *is satisfied, then the matrix* A^h *of the system* (1.13) *is positive definite.*

Proof. Multiply left part of each equation (1.10) and (1.11) by $hu^h(z)$ with corresponding z and sum over all $z \in \omega_h$:

$$h \sum_{z \in \omega_h} u^h(z) L^h u^h(z) - h \sum_{z \in \omega_{00}} u^h(z) L^{2h} u^h(z). \tag{2.1}$$

Set $u^h = 0$ on γ_h and for the obtained expression apply difference analog of the first Green function [3], going over to index notations:

$$\begin{aligned} h\sum_{z\in\omega_h} u^h(z)L^h u^h(z) &= h\sum_{i,j=1}^{n-1} d_{ij}(u^h_{ij})^2 \\ &+\frac{1}{h}\sum_{i,j=1}^{n}\left[(u^h_{ij}-u^h_{i-1,j})^2+(u^h_{ij}-u^h_{i,j-1})^2\right], \end{aligned} \tag{2.2}$$

$$\begin{aligned} 2h\sum_{z\in\omega_{00}} u^h(z)L^{2h}u^h(z) &= 2h\sum_{i,j=1}^{n/2-1} d_{2i,2j}(u^h_{2i,2j})^2 \\ &+\frac{1}{2h}\sum_{i,j=1}^{n/2}\left[(u^h_{2i,2j}-u^h_{2i-2,2j})^2+(u^h_{2i,2j}-u^h_{2i,2j-2})^2\right]. \end{aligned} \tag{2.3}$$

For real numbers a, b the equality $a^2+b^2 \geq (a+b)^2/2$ is true, from which follows that

$$\begin{aligned} &(u^h_{2i,2j}-u^h_{2i-1,2j})^2+(u^h_{2i,2j}-u^h_{2i,2j-1})^2 \\ &+(u^h_{2i-1,2j}-u^h_{2i-2,2j})^2+(u^h_{2i,2j-1}-u^h_{2i,2j-2})^2 \\ &\leq\frac{1}{2}\left[(u^h_{2i,2j}-u^h_{2i-2,2j})^2+(u^h_{2i,2j}-u^h_{2i,2j-2})^2\right]. \end{aligned} \tag{2.4}$$

With account of this inequality the expression (2.1) is estimated from below by the value

$$\begin{aligned} &\frac{3}{4h}\sum_{i,j=1}^{n}\left[(u^h_{i,j}-u^h_{i-1,j})^2+(u^h_{i,j}-u^h_{i,j-1})^2\right] \\ &+h\sum_{i,j=1}^{n-1} d_{i,j}(u^h_{i,j})^2 - h\sum_{i,j=1}^{n/2} d_{2i,2j}(u^h_{2i,2j})^2. \end{aligned} \tag{2.5}$$

The sum $h\sum_{i,j=1}^{n/2} d_{i,j}(u^h_{i,j})^2$ contains all the terms $h\sum_{i,j=1}^{n/2} d_{2i,2j}(u^h_{2i,2j})^2$. Therefore the difference

$$h\sum_{i,j=1}^{n/2} d_{i,j}(u^h_{i,j})^2 - h\sum_{i,j=1}^{n/2} d_{2i,2j}(u^h_{2i,2j})^2$$

is nonnegative. The first sum in (2.5) is estimated from below by means of the equation [3]

$$16h^2 \sum_{i,j=1}^{n} (u_{i,j}^h)^2 \leq \sum_{i,j=1}^{n} \left[(u_{i,j}^h - u_{i-1,j}^h)^2 + (u_{i,j}^h - u_{i,j-1}^h)^2 \right], \tag{2.6}$$

which is an analog of embedding of norms from $H_0^1(\Omega)$ into $L^2(\Omega)$. Finally, the expression (2.5) is estimated from below by the value

$$12h \sum_{i,j=1}^{n-1} (u_{ij}^h)^2 = 12h \sum_{z\in\omega_h} (u^h(z))^2. \tag{2.7}$$

Comparing it with (2.1) we arrive at the statement of Theorem. □

Symmetry and positive definiteness of the matrix A^h lead to two useful conclusions. First, the system (1.13) has unique solution u^h for any right part F^h, which follows from inadmissibility of zero eigenvalue of the matrix A^h. Second, for approximate solution of the system (1.13) an application of a number of various direct and iterative methods [4] becomes possible.

Now, let show that the system (1.14) satisfies comparison theorems despite that it is not M-matrix. For this purpose introduce a denotation $G^h \leq 0$ for the vector G^h with components G_j^h, $j = 1, \ldots, (n+1)^2$, which signifies component-wise comparison.

Theorem 33. *Let the condition* (1.4) *be satisfied and step h be small enough:*

$$h \leq 2/(5\|d\|_{\infty,\overline{\Omega}}). \tag{2.8}$$

Then for the system (1.14) *from* $G^h \geq 0$ *the inequality* $V^h \geq 0$ *follows.*

Proof. In order to use standard results on M-matrices it is necessary that diagonal elements would be positive and off-diagonal ones do nonnegative. This condition is satisfied for equations in the nodes ω_{11}, ω_{10} and ω_{01}, but not for equations in the nodes ω_{00} (see Fig. 2). Therefore slightly transform the system (1.14) or, what is the same, the system (1.10) — (1.12) so that to get rid of positive off-diagonal elements in the nodes ω_{00}. For that, to each equation corresponding to $(x,y) \in \omega_{00}$ add four equations corresponding to the nodes $(x \pm h, y \pm h) \in \omega_{11}$ with a weight a and four equations corresponding to the nodes $(x \pm h, y) \in \omega_{10}, (x, y \pm h) \in \omega_{01}$ with a weight b. As a result, in a node $(x,y) \in \omega_{00}$ we obtain an equation with the stencil

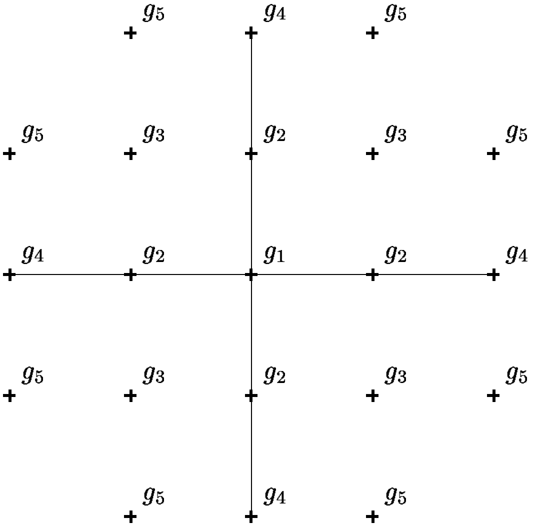


Fig. 3: 21-point stencil of the equation in a node $(x, y) \in \omega_{00}$ after transformation.

shown in Fig. 3., where

$$\begin{aligned}
g_1 &= \frac{3}{h^2} - \frac{4b}{h^2}, \\
g_2 &= -\frac{1}{h^2} + b\left(\frac{4}{h^2} + d\right) - \frac{2a}{h^2}, \\
g_3 &= a\left(\frac{4}{h^2} + d\right) - \frac{2b}{h^2}, \\
g_4 &= \frac{1}{4h^2} - \frac{b}{h^2}, \\
g_5 &= -\frac{a}{h^2}.
\end{aligned} \tag{2.9}$$

Let try to choose the weights a, b so that in the equation obtained after transformation the diagonal element would be positive and off-diagonal elements do nonnegative. This will be true if

$$g_1 \geq 0, \quad g_2 \leq 0, \quad g_3 \leq 0, \quad g_4 \leq 0, \quad g_5 \leq 0. \tag{2.10}$$

This results in the problem to determinate the admissible state. Let for a step h the condition (2.8) be satisfied. Then the problem (2.10) has a nonempty set of admissible values, from which we choose

$$a = 1/20, \ b = 1/4. \tag{2.11}$$

Finally, the following coefficients of the stencil in Fig. 3 are obtained:

$$\begin{gathered} g_1 = \frac{2}{h^2}, \quad g_2 = -\frac{1}{10h^2} + \frac{d}{4}, \\ g_3 = -\frac{3}{10h^2} + \frac{d}{20}, \quad g_4 = 0, \quad g_5 = -\frac{1}{20h^2}. \end{gathered}$$

It is easy to verify that under the condition (2.8) we arrive at the inequalities (2.10). Thus, instead of (1.14) we obtain a system

$$\overline{B}^h V^h = \overline{G}^h \tag{2.12}$$

with M-matrix $\overline{B}^h$ and the same solution V^h. Due to positiveness of the weights $a,\ b$ the inequality $\overline{G}^h \geq 0$ is true. Therefore on the basis of the properties of M-matrices [3]

$$V^h \geq 0. \quad \square$$

Prove an a priori estimate useful for further reasonings.

Theorem 34. *Let for the problem*

$$\begin{aligned} &L^h v^h = g^h \quad in \quad \omega_h \setminus \omega_{00}, \\ &L^h v^h - L^{2h} v^h = g^h \quad in \quad \omega_{00}, \\ &v^h = g^h \quad on \quad \gamma_h \end{aligned} \tag{2.13}$$

the estimates (1.4) *and* (2.8) *be fulfilled. Then*

$$\left\| v^h \right\|_{\infty, \overline{\omega}_h} \leq \frac{11}{48} \left\| g^h \right\|_{\infty, \omega_h} + \left\| g^h \right\|_{\infty, \gamma_h}. \tag{2.14}$$

Proof. Introduce a function

$$w = c_3 + c_4 x(1-x) \quad \text{in} \quad \overline{\Omega} \tag{2.15}$$

with the constants

$$c_3 = \left\| g^h \right\|_{\infty, \gamma_h}, \quad c_4 = \frac{11}{12} \left\| g^h \right\|_{\infty, \omega_h}. \tag{2.16}$$

Note that

$$L^h w = Lw = dw + 2c_4 \geq 2c_4 \quad \text{in } \omega_h, \tag{2.17}$$

$$L^{2h} w = Lw = dw + 2c_4 \quad \text{in } \omega_{00}. \tag{2.18}$$

Therefore for the nodes $(x, y) \in \omega_h \setminus \omega_{00}$ we have

$$L^h w \geq 2c_4 \geq \left\| g^h \right\|_{\infty, \omega_h} \geq \left| g^h \right|. \tag{2.19}$$

For the boundary nodes $(x, y) \in \gamma_h$ it is also evident that

$$w \geq \left\| g^h \right\|_{\infty, \gamma_h} \geq \left| g^h \right|. \tag{2.20}$$

Consider grid operator in a node $(x, y) \in \omega_{00}$, which is transformed according to the rule pointed out in Theorem 2:

$$\begin{aligned} & L^h w - L^{2h} w + a\left(L^h w(x+h, y+h) + L^h w(x-h, y+h)\right. \\ & \left.+L^h w(x+h, y-h) + L^h w(x-h, y-h)\right) + b\left(L^h w(x, y+h)\right. \\ & \left.+L^h w(x, y-h) + L^h w(x+h, y) + L^h w(x-h, y)\right) \\ & \geq 8ac_4 + 8bc_4 = \frac{12}{5} c_4 \geq \frac{11}{5} \left\| g^h \right\|_{\infty, \omega_h} \\ & \geq \left| g^h + a\left(g^h(x+h, y+h) + g^h(x-h, y+h)\right.\right. \\ & \left. +g^h(x+h, y-h) + g^h(x-h, y-h)\right) + b\left(g^h(x, y+h)\right. \\ & \left.\left. +g^h(x, y-h) + g^h(x+h, y) + g^h(x-h, y)\right)\right|. \end{aligned} \tag{2.21}$$

Introduce vectors V^h and W^h with the components

$$V^h = \left\{ v_{ij}^h \right\}_{i,j=0}^{n+1}, \quad W^h = \left\{ w_{ij} \right\}_{i,j=0}^{n+1},$$

which are ordered as in the system (1.14). Then from the inequalities (2.19) — (2.21) it follows that

$$\overline{B}^h W^h \geq \overline{B}^h V^h, \quad \text{i.e.} \quad \overline{B}^h \left(W^h - V^h\right) \geq 0.$$

From the properties of M-matrices it follows that

$$W^h - V^h \geq 0, \quad \text{i.e.} \quad w \geq v^h \quad \text{in} \quad \overline{\omega}_h.$$

Similarly, from (2.19) — (2.21) it follows that

$$w \geq -v^h \quad \text{in} \quad \overline{\omega}_h.$$

Therefore

$$\left|v^h\right| \le w \quad \text{in} \quad \overline{\omega}_h. \tag{2.22}$$

In the left-hand side take maximum over $\overline{\omega}_h$, and in the right-hand side do over $\overline{\Omega}$. Finally we obtain

$$\left\|v^h\right\|_{\infty,\overline{\omega}_h} \le c_3 + c_4/4,$$

that is equivalent to (2.14). □

3 Convergence of the nonuniform difference scheme

Theorem 35. *Let* u, u^h *be solutions of the problems* (1.1)–(1.2) *and* (1.10)–(1.12), *respectively, and the conditions* (1.3) — (1.5) *be satisfied. Then*

$$\|u - u^h\|_{\infty,\overline{\omega}_h} \le c_5 h^4. \tag{3.1}$$

Proof. We will establish a finer structure of the error. Let prove that the solution u^h can be represented as

$$u^h = u + h^4\rho^h \qquad \text{in} \quad \omega_{11}, \tag{3.2}$$

$$u^h = u + w_{01}h^4 + h^4\rho^h \quad \text{in} \quad \overline{\omega}_{01} \cup \overline{\omega}_{10}, \tag{3.3}$$

$$u^h = u + w_{00}h^4 + h^4\rho^h \qquad \text{in} \quad \overline{\omega}_{00}, \tag{3.4}$$

where the functions

$$w_{01} = -\frac{1}{48}\mu, \quad w_{00} = -\frac{1}{12}\mu, \quad \mu = \frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4} \tag{3.5}$$

does not depend on h, and the remainder term ρ^h is limited in the following way:

$$\left\|\rho^h\right\|_{\infty,\overline{\omega}_h} \le c_6. \tag{3.6}$$

In the expression (1.7), apply Taylor series expansion from the points $(x \pm h, y)$ and $(x, y \pm h)$ into the node (x, y). Further we omit the argument (x, y) if this does not arouse misunderstanding:

$$\begin{aligned} u_{\overset{\circ}{x}\overset{\circ}{x}} &= \frac{\partial^2 u}{\partial x^2} + \frac{h^2}{12}\frac{\partial^4 u}{\partial x^4} + h^4\mu^h_{1x}, \\ u_{\overset{\circ}{y}\overset{\circ}{y}} &= \frac{\partial^2 u}{\partial y^2} + \frac{h^2}{12}\frac{\partial^4 u}{\partial y^4} + h^4\mu^h_{1y}, \end{aligned} \tag{3.7}$$

where

$$|\mu_{1x}^h| \leq \frac{1}{360}\left\|\frac{\partial^6 u}{\partial x^6}\right\|_{\infty,\overline{\Omega}} \quad \text{in} \quad \omega_h,$$
$$|\mu_{1y}^h| \leq \frac{1}{360}\left\|\frac{\partial^6 u}{\partial y^6}\right\|_{\infty,\overline{\Omega}} \quad \text{in} \quad \omega_h. \tag{3.8}$$

With consideration of the expansions (3.2), (3.4) and (3.7) for odd nodes ω_{11} we obtain

$$\begin{aligned} L^h u^h = {} & L^h u + h^4 L^h \rho^h - h^2 (w_{01}(x+h,y) \\ & + w_{01}(x-h,y) + w_{01}(x,y+h) + w_{01}(x,y-h)). \end{aligned} \tag{3.9}$$

For the function w_{01} use Taylor series expansion from the points $(x \pm h, y)$ and $(x, y \pm h)$ into the node (x,y):

$$\begin{aligned} w_{01}(x+h,y) + w_{01}(x-h,y) + w_{01}(x,y+h) + w_{01}(x,y-h) \\ = 4w_{01} + 2h^2 \mu_{01}^h, \end{aligned} \tag{3.10}$$

where with account of (3.5) we have

$$2\left|\mu_{01}^h\right| \leq \left\|\frac{\partial^2 \omega_{01}}{\partial x^2} + \frac{\partial^2 \omega_{01}}{\partial y^2}\right\|_{\infty,\overline{\Omega}} = \frac{1}{124}\|u\|_{\infty,\overline{\Omega}}^{(6)}. \tag{3.11}$$

Taking into consideration the expansions (3.7), (3.10) into (3.9), we obtain the equality

$$\begin{aligned} L^h u^h &= (-\Delta u + du) - \frac{h^2}{12}\mu \\ -h^4(\mu_{1x}^h &+ \mu_{1y}^h) + h^4 L^h \rho^h - 4h^2 \omega_{01} - 2h^4 \mu_{01}. \end{aligned}$$

On the basis of equations (1.1), (1.10) and definitions (3.5) a cancellation of terms of the orders 1 and h^2 is performed. Divide the remaining terms by h^4. As a result we arrive at the inequality

$$L^h \rho^h = \mu_{1x}^h + \mu_{1y}^h + 2\mu_{01}^h \quad \text{in} \quad \omega_{11}. \tag{3.12}$$

Substitution of the expansions (3.3), (3.4), (3.10) into the grid operator (1.11) for even nodes ω_{00} gives the following:

$$\begin{aligned} L^h u^h - L^{2h} u^h = L^h u - L^{2h} u + h^4 (L^h \rho^h - L^{2h} \rho^h) \\ + h^4 \left(4w_{00}/h^2 + dw_{00}\right) - 4h^2 w_{01} - 2h^4 \mu_{01}^h. \end{aligned} \tag{3.13}$$

In odd nodes w_{00} expressions similar to (3.7) are valid, but with doubled step, which gives

$$L^{2h}u = (-\Delta u + du) - \frac{h^2}{3}\mu - h^4(\mu_{2x}^h + \mu_{2y}^h), \tag{3.14}$$

where

$$\left|\mu_{2x}^h\right| \leq \frac{4}{45}\left\|\frac{\partial^6 u}{\partial x^6}\right\|_{\infty,\overline{\Omega}}, \quad \left|\mu_{2y}^h\right| \leq \frac{4}{45}\left\|\frac{\partial^6 u}{\partial y^6}\right\|_{\infty,\overline{\Omega}} \quad \text{in} \quad \omega_{00}. \tag{3.15}$$

Taking account of (3.14) in (3.13), we obtain the equality

$$\begin{aligned} L^h u^h - L^{2h} u^h \ &= h^4(L^h\rho^h - L^{2h}\rho^h) + 4h^2 w_{00} + dw_{00}h^4 - 4h^2 w_{01} - 2h^4\mu_{01}^h \\ &-\frac{h^2}{12}\mu - h^4\mu_{1x}^h - h^4\mu_{1y}^h + \frac{h^2}{3}\mu + h^4\mu_{2x}^h + h^4\mu_{2y}^h. \end{aligned}$$

Again, on the basis of equations (1.1), (1.10) and definitions (3.5) the cancellation of terms of the orders 1 and h^2 takes place. In this case the cancellation of the terms of order h^2 is performed due to proper choice of multiplier at $L^{2h}u^h$. The remaining terms after division by h^4 give the equality

$$L^h\rho^h - L^{2h}\rho^h = \mu_{1x}^h + \mu_{1y}^h - \mu_{2x}^h - \mu_{2y}^h - dw_{00} \quad \text{in} \quad \omega_{00}. \tag{3.16}$$

Substitution of the expansions (3.2), (3.3), (3.4) into the grid operator (1.10) for nodes with alternating evenness of indices w_{10} gives the relation

$$\begin{aligned} &L^h u^h = L^h u + h^4 L^h \rho^h \\ &+h^4\left(\left(\frac{4w_{01}}{h^2} + dw_{01}\right) - \frac{w_{00}(x+h,y)}{h^2} - \frac{w_{00}(x-h,y)}{h^2}\right). \end{aligned} \tag{3.17}$$

For the function w_{00} apply Taylor series expansion from the points $(x \pm h, y)$ into the node (x, y) similarly to (3.10), (3.11). This yields the equality:

$$w_{00}(x+h,y) + w_{00}(x-h,y) = 2w_{00}(x,y) + h^2\mu_{00}^h(x,y), \tag{3.18}$$

where with account of (3.5) we have

$$\left|\mu_{00}^h\right| \leq \frac{1}{6}\|u\|_{\infty,\overline{\Omega}}^{(6)}. \tag{3.19}$$

Taking into account the expansions (3.7), (3.18) into (3.17), we obtain the equality

$$\begin{aligned} L^h u^h \ &= (-\Delta u + du) - \frac{h^2}{12}\mu + h^4 L^h\rho^h + 4h^2 w_{01} \\ &+h^4 dw_{01} - 2h^2 w_{00} - h^4\mu_{1x}^h - h^4\mu_{1y}^h - h^4\mu_{00}^h. \end{aligned}$$

Again, on the basis of equations (1.1), (1.10) and definitions (3.5) cancellation of terms of the orders 1 and h^2 takes place. Finally, after division by h^4 we arrive at the equality

$$L^h\rho^h = -dw_{01} + \mu^h_{1x} + \mu^h_{1y} + \mu^h_{00} \quad \text{in} \quad \omega_{10}. \tag{3.20}$$

Similarly for the nodes of another group of alternating evenness ω_{01} we obtain the equality

$$L^h\rho^h = -dw_{01} + \mu^h_{1x} + \mu^h_{1y} + \mu^h_{00} \quad \text{in} \quad \omega_{01} \tag{3.21}$$

with the same estimate (3.19) for the remainder term μ^h_{00}.

Taking into consideration (3.3), (3.4), (3.12), (3.16), (3.20), and (3.21), for ρ^h we obtain the problem

$$\begin{aligned}
&L^h\rho^h = \xi^h \quad \text{in} \quad \omega_h \setminus \omega_{00},\\
&L^h\rho^h - L^{2h}\rho^h = \xi^h \quad \text{in} \quad \omega_{00},\\
&\rho^h = -\omega_{01} \quad \text{in} \quad \gamma_h \cap (\overline{\omega}_{01} \cup \overline{\omega}_{10}),\\
&\rho^h = -\omega_{00} \quad \text{in} \quad \gamma_h \cap \overline{\omega}_{00}
\end{aligned} \tag{3.22}$$

with the right-hand side

$$\begin{aligned}
&\xi^h = \mu^h_{1x} + \mu^h_{1y} + 2\mu^h_{01} \quad \text{in} \quad \omega_{11},\\
&\xi^h = -dw_{01} + \mu^h_{1x} + \mu^h_{1y} + \mu^h_{00} \quad \text{in} \quad \omega_{01} \cup \omega_{10},\\
&\xi^h = \mu^h_{1x} + \mu^h_{1y} - \mu^h_{2x} - \mu^h_{2y} - dw_{00} \quad \text{in} \quad \omega_{00}.
\end{aligned}$$

Owing to the estimates (3.8), (3.11), (3.15), (3.19) and boundedness of functions d and μ from (3.5) the following inequality is valid

$$\left|\xi^h\right| \le c_7 \quad \text{in} \quad \omega_h. \tag{3.23}$$

Use the a priori estimate from Theorem 3. Then with account of (3.5) we have

$$\left\|\rho^h\right\|_{\infty,\overline{\omega}_h} \le \frac{11}{48}\left\|\xi^h\right\|_{\infty,\omega_h} + \frac{1}{12}\|\mu\|_{\infty,\gamma_h}. \tag{3.24}$$

Taking into account the estimate (3.23), we obtain (3.6) with the constant

$$c_6 = 11c_7/48 + \|\mu\|_{\infty,\gamma_h}/12.$$

From the representation (3.2) — (3.4) it follows that

$$\left\|u - u^h\right\|_{\infty,\overline{\omega}_h} \le h^4\left(\left\|\rho^h\right\|_{\infty,\overline{\omega}_h} + \|w_{01}\|_{\infty,\overline{\Omega}} + \|w_{00}\|_{\infty,\overline{\Omega}}\right).$$

With account of (3.24) this proves the estimate (3.1). □

4 Numerical examples

By analogy with the work [1] apply the constructed method to two problems of the form (1.1)–(1.2) with smooth and with oscillation solutions. The first problem is

$$\begin{aligned} -\Delta u = \; & 2\cos\left(\frac{\pi x}{2}\right) y(1-y)\cos\left(\frac{\pi y}{2}\right) \\ & +(1-x)\sin\left(\frac{\pi x}{2}\right)\pi y(1-y)\cos\left(\frac{\pi y}{2}\right) \\ & -x\sin\left(\frac{\pi x}{2}\right)\pi y(1-y)\cos\left(\frac{\pi y}{2}\right) \\ & +\frac{1}{2}x(1-x)\cos\left(\frac{\pi x}{2}\right)\pi^2 y(1-y)\cos\left(\frac{\pi y}{2}\right) \\ & +2x(1-x)\cos\left(\frac{\pi x}{2}\right)\cos\left(\frac{\pi y}{2}\right) \qquad (4.1) \\ & +x(1-x)\cos\left(\frac{\pi x}{2}\right)(1-y)\sin\left(\frac{\pi y}{2}\right)\pi \\ & -x(1-x)\cos\left(\frac{\pi x}{2}\right) y\sin\left(\frac{\pi y}{2}\right)\pi \quad \text{in} \quad \Omega, \\ u = \; & 0 \quad \text{in} \quad \Gamma. \end{aligned}$$

Its exact solution is

$$u(x,y) = x(1-x)\cos\left(\frac{\pi x}{2}\right) y(1-y)\cos\left(\frac{\pi y}{2}\right).$$

The second problem is

$$\begin{aligned} -\Delta u = \; & -32c(1-x)y(1-y) + 512sx(1-x)y(1-y) \\ & +32cxy(1-y) + 2sy(1-y) - 32cx(1-x)(1-y) \\ & +32cx(1-x)y + 2sx(1-x) \quad \text{in} \quad \Omega, \qquad (4.2) \\ u = \; & 0 \quad \text{in} \quad \Gamma, \end{aligned}$$

where the denotations $s = \sin(16x+16y)$ and $c = \cos(16x+16y)$ are used. Its exact solution is

$$u(x,y) = \sin(16x+16y)x(1-x)y(1-y).$$

In Tables 1,2 the errors $\delta_2 = \|u - u^h\|_{\infty,\overline{\omega}_h}$ and

$$\delta_1 = \|u - u^h\|_{2,\overline{\omega}_h} = \left(\sum_{z\in\overline{\omega}_h} \left(u(z) - u^h(z)\right)^2 \right)^{1/2}$$

of solutions of both the problems by standard method (1.7) of the second order of accuracy and by the proposed method (1.10)–(1.12) of the fourth order are presented.

Table 1: Error of approximate solutions for the problem with smooth solution.

N	Problem I			
	method (1.7)		method (1.10) — (1.12)	
	$2,\overline{\omega}_h$	$\infty,\overline{\omega}_h$	$2,\overline{\omega}_h$	$\infty,\overline{\omega}_h$
4	$1.18_{10}-03$	$2.24_{10}-03$	$6.73_{10}-04$	$2.20_{10}-03$
8	$2.92_{10}-04$	$6.11_{10}-04$	$4.30_{10}-05$	$1.64_{10}-04$
16	$7.27_{10}-05$	$1.52_{10}-04$	$2.68_{10}-06$	$1.02_{10}-05$
32	$1.82_{10}-05$	$3.82_{10}-05$	$1.68_{10}-07$	$6.46_{10}-07$
64	$4.54_{10}-06$	$9.54_{10}-06$	$1.06_{10}-08$	$4.08_{10}-08$

Table 2: Error of approximate solutions for the problem with oscillating solution.

N	Problem II			
	method (1.7)		method (1.10) — (1.12)	
	$2,\overline{\omega}_h$	$\infty,\overline{\omega}_h$	$2,\overline{\omega}_h$	$\infty,\overline{\omega}_h$
4	$1.38_{10}-01$	$2.70_{10}-01$	$1.53_{10}-01$	$3.64_{10}-01$
8	$1.18_{10}-02$	$2.67_{10}-02$	$4.70_{10}-02$	$1.41_{10}-01$
16	$2.42_{10}-03$	$5.76_{10}-03$	$2.57_{10}-03$	$1.04_{10}-02$
32	$5.78_{10}-04$	$1.39_{10}-03$	$1.42_{10}-04$	$6.00_{10}-04$
64	$1.43_{10}-04$	$3.52_{10}-04$	$8.45_{10}-06$	$3.61_{10}-05$

These data are represented on graphs (in logarithmic scale over both axes). In figures 4 and 5 the errors δ_1 and δ_2 of the method (1.7) are denoted by numbers 1, 2; The errors of the method (1.10) — (1.12) are denoted by numbers 3, 4; numbers 5 and 6 denote the lines with inclinations $tg(\varphi) = 2$ and $tg(\varphi) = 4$, characterizing the dependences $\delta = h^2$ and $\delta = h^4$, respectively.

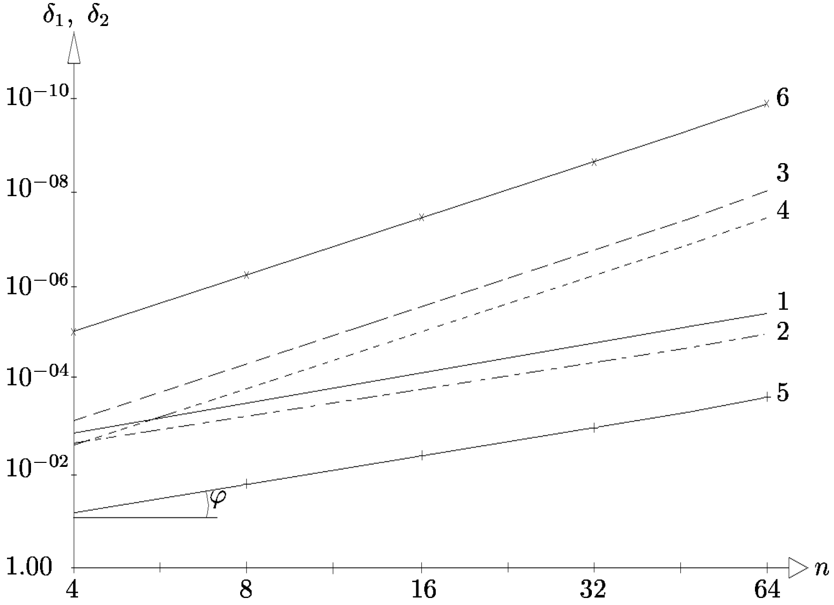


Fig. 4: Error of approximate solutions for the first problem.

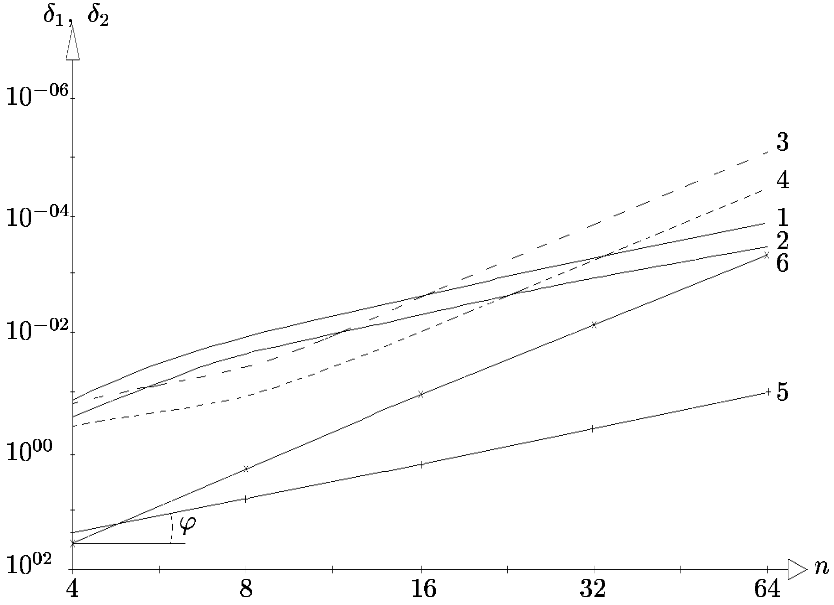


Fig. 5: Error of approximate solutions for the second problem.

Except that, in Fig. 6 a pointwise graph of the error $\delta_2' = u - u^h$ of the proposed method (1.10)–(1.12) on a grid $\overline{\omega}_h$ with the step $h = 1/32$ for the first problem is given.

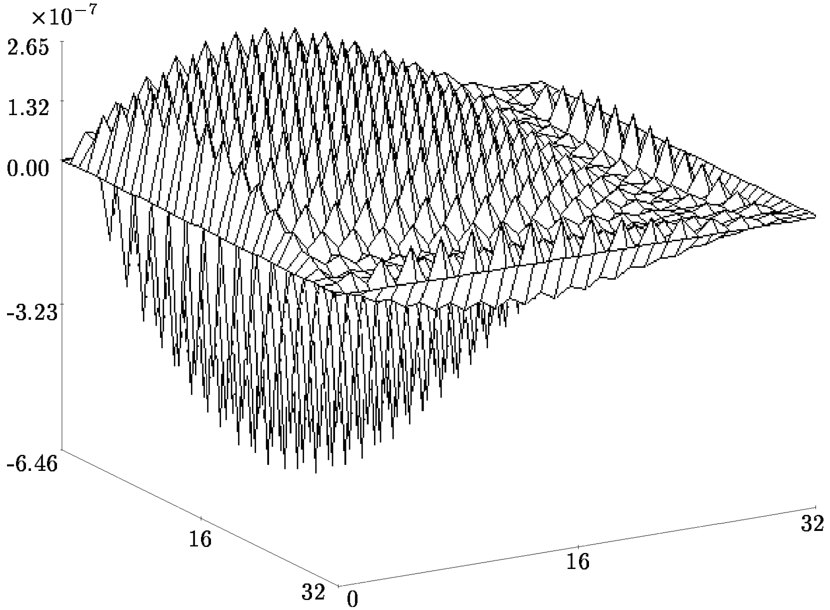


Fig. 6: Error δ_2' of the method (1.10)–(1.12) under $n = 32$. The first problem.

References

1. Bykova E.G., Shaidurov V.V.: *A Nonuniform Difference Scheme of Higher Order of Accuracy. One-dimensional Illustrative Example.* Preprint №17 of the Computing Center of SB RAS, Krasnoyarsk, 1996 (In Russian).
2. Rüde U.: *Extrapolation and Related Techniques for Solving Elliptic Equations.* Preprint №I–9135, München Technical University, 1991.
3. Samarsky A.A.: *Theory of Difference Schemes.* Moscow, Nauka, 1977 (In Russian).
4. Samarsky A.A., Nikolaev E.S.: *Methods of Solution of Grid Equations.* Moscow, Nauka, 1978 (In Russian).

A nonuniform difference scheme with fourth order of accuracy in a domain with smooth boundary

Bykova E.G., Shaidurov V.V.

Introduction

The present paper continues a series of works devoted to construction and justification *nonuniform* difference schemes of higher degrees of accuracy. Two-dimensional boundary-value problem for elliptic type equation in a domain with smooth curvilinear boundary is considered. The main idea of construction of such scheme is similar to that in the papers [1], [11], where it is stated for the same equation in a rectangle. The transition to curvilinear boundary required either to solve the question on special approximation of boundary values or to re-construct the grid equations on non-standard stencils near the boundary. Both approaches were used as applied to Richardson extrapolation in [2], [3] and [4], [5], [6], respectively. The first, although leads to required result, gives extensive stencils; the second, being more complicated in theoretical respect, gives more compact stencils of difference equations near the boundary, so it appeared to be more preferable.

As it is in one-dimensional case, the difference scheme inside the domain is similar in structure to the equations of the method of extrapolated equations by U. Rüde [7] for finite elements. But near the boundary the equations appear to be different. The justification of accuracy here is also different from [7] and based on the maximum principle for a system of linear algebraic equations equivalent to the difference scheme.

Let recall that the term nonuniform scheme was introduced in [8] and used in [1] due to two different rules of construction of grid equations in neighbouring nodes as distinct from uniform schemes [9], when the rule of construction is the same for all nodes of the grid, at least inside the domain.

1 Boundary-value problem

Let Ω be a limited domain in R^2 with smooth boundary Γ (i.e. of the class C^1). Consider a boundary-value problem

$$-\Delta u + du = f \quad \text{in} \quad \Omega, \tag{1.1}$$

$$u = g \quad \text{on} \quad \Gamma \tag{1.2}$$

with continuous on $\overline{\Omega}$ functions d, f and continuous on Γ function g, and

$$d \geq 0 \quad \text{on} \quad \overline{\Omega}. \tag{1.3}$$

These conditions ensure unique solvability of the problem. Suppose that the solution is smooth enough:

$$u \in C^6(\overline{\Omega}). \tag{1.4}$$

2 Construction of the difference grid and classification of its nodes

Likewise in [6], suppose that the domain Ω is located within the square $\{(x,y): \ 0 \leq x \leq 1, \ \ 0 \leq y \leq 1\}$. Cover it with a square grid with the step $h = 1/N$, formed by the lines $x_i = ih$ and $y_j = jh$, where $i, j = 0, 1, \ldots, N$ and N is integer. Let call *nodes* the points of intersection of these lines. A node z_{ij} is called *inner*, if $z_{ij} \in \Omega$. Denote the set of all inner nodes by ω_h.

Each line of the grid x_i or y_j which intersects Ω also intersects the boundary Γ. Due to smoothness of the boundary the intersection with the domain consists of certain number of intervals. Let call the end of these intervals *boundary nodes* in direction x (or y), if the line being considered is parallel to the coordinate axis Ox (respectively Oy). The set of all boundary nodes in direction x denote by $\gamma_{h,x}$, and the set of all boundary nodes in direction y denote by $\gamma_{h,y}$. Also denote

$$\gamma_h = \gamma_{h,x} \cup \gamma_{h,y} \quad \text{and} \quad \overline{\omega}_h = \omega_h \cup \gamma_h.$$

For convenience, let divide the set ω_h into four subsets

$$\begin{aligned}
\omega_{00} &= \{z_{ij}: \ z_{ij} \in \omega_h, \ i \text{ is even}, \ j \text{ is even}\};\\
\omega_{01} &= \{z_{ij}: \ z_{ij} \in \omega_h, \ i \text{ is even}, \ j \text{ is odd}\};\\
\omega_{10} &= \{z_{ij}: \ z_{ij} \in \omega_h, \ i \text{ is odd}, \ j \text{ is even}\};\\
\omega_{11} &= \{z_{ij}: \ z_{ij} \in \omega_h, \ i \text{ is odd}, \ j \text{ is odd}\}.
\end{aligned}$$

For each inner node $z_{ij} = (x_i, y_j)$ introduce two definitions of the distance to the boundary Γ which is parallel to two coordinate axes:

$$\rho_1(x_i, y_j) = \min_{(x, y_j)\in\Gamma} |x_i - x|,$$
$$\rho_2(x_i, y_j) = \min_{(x_i, y)\in\Gamma} |y_j - y|.$$

With the help of these definitions introduce a classification of the inner nodes ω_h. Geometric illustration of this classification is given at the end of the paper for a concrete numerical example. Denote by $\gamma^1_{1,h}$ *the set of inner irregular nodes of the first type*, for which only one of the distances ρ_1 or ρ_2 is less than h, and the another is greater than or equal to $2h$:

$$\gamma^1_{1,h} = \{z_{ij} : z_{ij} \in \omega_h,\ (\rho_1(z_{ij}) < h)\ \&\ (\rho_2(z_{ij}) \geq 2h)$$
$$\text{or } (\rho_1(z_{ij}) \geq 2h)\ \&\ (\rho_2(z_{ij}) < h)\}.$$

By $\gamma^2_{1,h}$ let denote *the set of inner irregular nodes of the second type*, for which both the distances ρ_1 and ρ_2 are less than h:

$$\gamma^2_{1,h} = \{z_{ij} :\ z_{ij} \in \omega_h,\ (\rho_1(z_{ij}) < h)\ \&\ (\rho_2(z_{ij}) < h)\}.$$

Respectively, by $\gamma^3_{1,h}$ let denote the set of *inner irregular nodes of the third type*, for which at least one adjacent node $z_{i,j\pm1}$, $z_{i\pm1,j}$ belongs to $\gamma^2_{1,h}$:

$$\gamma^3_{1,h} = \{z_{ij} : z_{ij} \in\ \omega_h,\ z_{i+1,j} \in \gamma^2_{1,h} \text{ or } z_{i-1,j} \in \gamma^2_{1,h}$$
$$\text{or } z_{i,j+1} \in \gamma^2_{1,h} \text{ or } z_{i,j-1} \in \gamma^2_{1,h}\}.\ ^{*)}$$

By $\gamma^{out}_{1,h}$ denote the set of external irregular nodes z_{ij}, for which at least one adjacent node $z_{i\pm1,j}$, $z_{i,j\pm1}$ belongs to $\gamma^1_{1,h} \cup \gamma^3_{1,h}$:

$$\gamma^{out}_{1,h} = \{z_{ij} : z_{ij} \notin\ \overline{\Omega},\ z_{i+1j} \in \gamma^1_{1,h} \cup \gamma^3_{1,h} \text{ or } z_{i-1j} \in \gamma^1_{1,h} \cup \gamma^3_{1,h}$$
$$\text{or } z_{ij+1} \in \gamma^1_{1,h} \cup \gamma^3_{1,h} \text{ or } z_{ij-1} \in \gamma^1_{1,h} \cup \gamma^3_{1,h}\}.$$

Now, let classify the nodes from ω_{00} near the boundary, which have not come into $\gamma^1_{1,h}$, $\gamma^2_{1,h}$ or $\gamma^3_{1,h}$. Denote by $\gamma_{2,h}$ the set of multiple nodes near the boundary, for which at least one of the distances ρ_1 or ρ_2 is less than $3h$, i.e.,

$$\gamma_{2,h} = \ \{z_{ij} : z_{ij} \in \omega_{00} \setminus (\gamma^1_{1,h} \cup \gamma^2_{1,h} \cup \gamma^3_{1,h}),$$
$$(\rho_1(z_{ij}) < 3h) \text{ or } (\rho_2(z_{ij}) < 3h)\}.$$

By $\gamma^1_{2,h}$ denote a subset of nodes from $\gamma_{2,h}$, for which at least one of the distances ρ_1 or ρ_2 is less than $2h$:

$$\gamma^1_{2,h} = \{z_{ij} : z_{ij} \in \gamma_{2,h},\ (\rho_1(z_{ij}) < 2h) \text{ or } (\rho_2(z_{ij}) < 2h)\}.$$

By $\gamma^2_{2,h}$ denote a subset of nodes from $\gamma_{2,h}$, for which both the distances ρ_1 and ρ_2 are greater than $2h$:

$$\gamma^2_{2,h} = \{z_{ij} : z_{ij} \in \gamma_{2,h},\ (\rho_1(z_{ij}) > 2h)\ \&\ (\rho_2(z_{ij}) > 2h)\}.$$

And, finally, by $\gamma^3_{2,h}$ let denote a subset of nodes from $\gamma_{2,h}$, for which only one of the distances ρ_1 or ρ_2 is greater than $2h$, and the another is greater than $3h$:

$$\gamma^3_{2,h} = \gamma_{2,h} \setminus (\gamma^1_{2,h} \cup \gamma^2_{2,h}).$$

For convenience of subsequent consideration, let divide $\gamma^3_{1,h}$ into three subsets:
1) $\gamma^{31}_{1,h}$ consists of nodes whose both adjacent nodes belong to $\gamma^2_{1,h}$;
2) $\gamma^{32}_{1,h}$ consists of nodes whose one adjacent node belongs to $\gamma^2_{1,h}$, and the other one belongs to $\gamma^{out}_{1,h}$;
3) $\gamma^{33}_{1,h} = \gamma^3_{1,h} \setminus (\gamma^{31}_{1,h} \cup \gamma^{32}_{1,h})$.

Make a classification of regular nodes. Let call a node *regular of the first kind,* if it belongs to $\omega_h \setminus \omega_{00}$ and is not included in $\gamma^1_{1,h}$, $\gamma^2_{1,h}$, $\gamma^3_{1,h}$; denote the set of such nodes by

$$\omega^r_{h,1} = \omega_h \setminus (\omega_{00} \cup \gamma^1_{1,h} \cup \gamma^2_{1,h} \cup \gamma^3_{1,h}).$$

Let call a node *regular of the second kind,* if it belongs to ω_{00}, but is not included in $\gamma_{2,h}$; denote the set of such nodes by

$$\omega^r_{h,2} = \omega_{00} \setminus (\gamma_{2,h} \cup \gamma^1_{1,h} \cup \gamma^2_{1,h} \cup \gamma^3_{1,h}).$$

The totality of regular nodes denote by $\omega^r_h = \omega^r_{h,1} \cup \omega^r_{h,2}$, and the nodes $\omega^{ir}_h = \omega_h \setminus \left\{\omega^r_h \cup \gamma^2_{1,h}\right\}$ let call irregular one.

For an arbitrary function v defined on a set D (finite or infinite) let introduce the denotation

$$||v||_{\infty,D} = \sup_D |v|.$$

3 Interpolation formula

For interpolation of boundary values we will use the interpolation formulas of two forms, selected in each case for ensuring stability (the stencils are shown in Fig. 1).

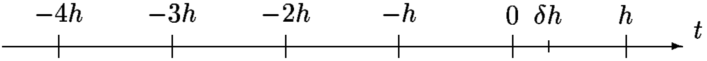


Fig. 1: Stencil of Lagrange interpolation formula; $0 < \delta \leq 1$.

Let the function $v(t) \in C^4[-3h, \delta h]$. The first of the formulas approximates the value $v(0)$ through the values of the same function in the points $-h$, $-2h$, $-3h$ and δh $(0 < \delta \leq 1)$:

$$\begin{aligned} v(0) &\approx \varphi_1(v(-h), v(-2h), v(-3h), v(\delta h)) \qquad (3.1)\\ &= \frac{3\delta}{\delta+1}v(-h) - \frac{3\delta}{\delta+2}v(-2h) + \frac{\delta}{\delta+3}v(-3h) + \varphi_{1f}(\delta)v(\delta h) \end{aligned}$$

where $\varphi_{1f}(\delta) = 6/((\delta+1)(\delta+2)(\delta+3))$.

Now, let supplement the definition of the function v to the right from δh with a segment of Taylor series with respect to δh up to the fourth derivative inclusive. Let keep the denotation $v(t)$ for the supplement, and note that it belongs to $C^4[-2h, h]$, and

$$||v^{(4)}||_{\infty,[-2h,h]} = ||v^{(4)}||_{\infty,[-2h,\delta h]}.$$

The second formula expresses $v(h)$ through four values in the points 0, $-h$, $-2h$, and δh:

$$\begin{aligned} v(h) &\approx \varphi_2(v(0), v(-h), v(-2h), v(\delta h)) \qquad (3.2)\\ &= -\frac{3(1-\delta)}{\delta}v(0) + \frac{3(1-\delta)}{\delta+1}v(-h) - \frac{1-\delta}{\delta+2}v(-2h) + \varphi_{2f}(\delta)v(\delta h) \end{aligned}$$

where $\varphi_{2f}(\delta) = 6/(\delta(\delta+1)(\delta+2))$.

Let recall [10] that interpolation over four nodes gives result with fourth order of accuracy in the following form:

$$\max\{|v(0) - \varphi_1|, |v(h) - \varphi_2|\} \leq ch^4||v^{(4)}||_{\infty,[-3h,\delta h]} \qquad (3.3)$$

with the constant c independent of h, $v(t)$, and $\delta \in (0, 1]$.

4 Construction of difference approximation

For difference approximation of the equation (1.1) introduce the following operators:

$$\begin{aligned} L^h v(x,y) =&(v(x-h,y) + v(x+h,y) + v(x,y-h) + v(x,y+h) -\\ &4v(x,y))/h^2 + d(x,y)v(x,y), \qquad (4.1)\\ L^{2h} v(x,y) =&(v(x-2h,y) + v(x+2h,y) + v(x,y-2h) + v(x,y+2h) -\\ &4v(x,y))/(4h^2) + d(x,y)v(x,y). \qquad (4.2) \end{aligned}$$

Start the construction of grid equations with the regular nodes. Let $z_{ij} \in \omega^r_{h,1}$. Then the difference approximation is performed as a standard five-point equation

$$L^h u^h(z_{ij}) = f(z_{ij}), \quad z_{ij} \in \omega^r_{h,1}, \tag{4.3}$$

on the stencil small cross (see Fig. 2.a)). But if z_{ij} is a regular node of the second kind, i.e., $z_{ij} \in \omega^r_{h,2}$, then the difference approximation is taken as nine-point equation on the stencil large cross (see. Fig. 2.b)):

$$L^h u^h(z_{ij}) - L^{2h} u^h(z_{ij}) = 0, \quad z_{ij} \in \omega^r_{h,2}. \tag{4.4}$$

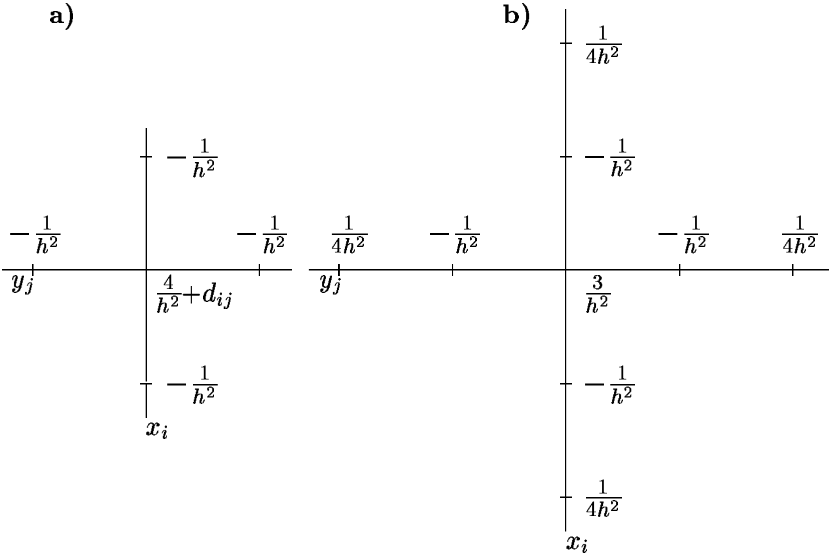


Fig. 2: Stencils **a)** small cross and **b)** large cross with the values of coefficients of the operators L^h and $L^h - L^{2h}$, respectively.

Consider the construction of grid equation in irregular nodes. Let $z_{ij} \in \gamma^1_{1,h}$ and one node of the stencil small cross, for instance, $z_{i,j+1}$ does not belong to Ω (see Fig. 3.a)). Denote by s_{ij} the point of intersection of the boundary Γ with the segment $[z_{ij}, z_{i,j+1}]$. At the beginning assume that the solution $u(x, y)$ is determined in the point $z_{i,j+1}$ and write down an ordinary five-point equation (4.3). Then construct interpolation formula (3.2) for the function $u(x,y)$ with respect to y coordinate, directing the axis $0t$ from z_{ij} into $z_{i,j+1}$ and assuming δ be equal to the distance δ_y from z_{ij} to s_{ij}, i.e., $\delta_y = \rho_2(z_{ij})$. As a result, we obtain five-point grid equation with *the first asymmetric T-shaped stencil* :

$$\begin{aligned}&\left(\frac{4}{h^2}+d(z_{ij})+\frac{3(1-\delta_y)}{\delta_y h^2}\right)u^h(z_{ij})-\left(\frac{1}{h^2}+\frac{3(1-\delta_y)}{(1+\delta_y)h^2}\right)u^h(z_{i,j-1})\\&+\frac{1-\delta_y}{(2+\delta_y)h^2}u^h(z_{i,j-2})-\frac{1}{h^2}u^h(z_{i+1,j})-\frac{1}{h^2}u^h(z_{i-1,j})\\&=f(z_{ij})+\frac{1}{h^2}\varphi_{2f}(\delta_y)g(s_{ij})\end{aligned}\tag{4.5}$$

where the boundary condition (2) is used in the form $u(s_{ij}) = g(s_{ij})$.

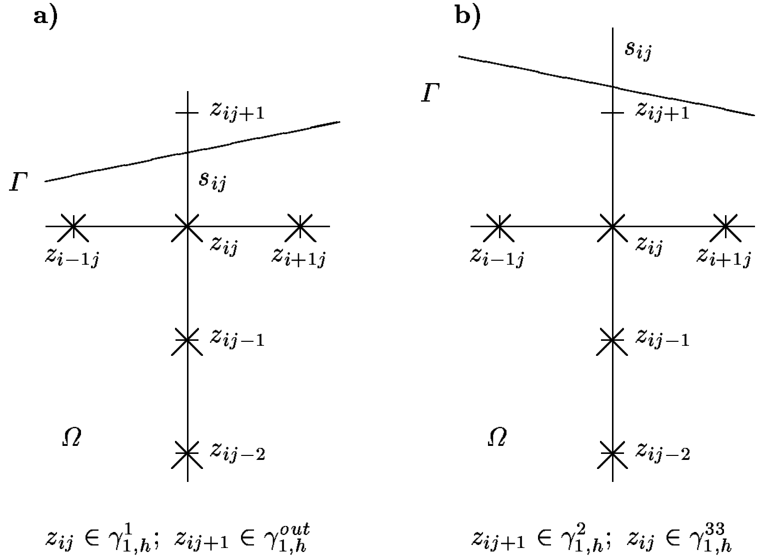

a)
b)
Γ
Γ
z_{ij+1}
s_{ij}
z_{i-1j}
z_{ij}
z_{i+1j}
z_{ij-1}
z_{ij-2}
Ω
Ω
$z_{ij} \in \gamma^1_{1,h};\ z_{ij+1} \in \gamma^{out}_{1,h}$
$z_{ij+1} \in \gamma^2_{1,h};\ z_{ij} \in \gamma^{33}_{1,h}$

Fig. 3: The first (a) and the second (b) T-shaped asymmetric stencils. Cross sign marks the nodes of corresponding stencils.

Remark. Construction of grid equations is not performed in the nodes $z_{ij} \in \gamma^2_{1,h}$. An attempt of double application of the above method in these nodes (elimination of external nodes by means of mean value formula (3.2)) gives grid equations which do not provide sufficient conditions for comparison theorems and the proof of stability. Therefore the authors refused from use of grid equations in nodes of the set $\gamma^2_{1,h}$. Accordingly, the resulting system of equations should not contain variables $u^h(z_{ij})$ with arguments from $\gamma^2_{1,h}$.

Taking into account the above remark, it is necessary to exclude the values in the nodes $\gamma^2_{1,h}$ from the other grid equations. Three variants are possible when one or two nodes of the stencil small cross belong to $\gamma^2_{1,h}$, and there are two variants when one or two nodes of the stencil large cross belong to $\gamma^2_{1,h}$.

Consider these variants.

1) Suppose that $z_{ij} \in \gamma^{33}_{1,h}$ and one node of the stencil small cross, for instance, $z_{i,j+1}$ belongs to $\gamma^2_{1,h}$ (see Fig. 3.b)). Denote by s_{ij} the point of intersection of the boundary Γ with the ray $[z_{ij}, z_{i,j+1})$. At the beginning assume that the solution $u(x,y)$ is determined in the point $z_{i,j+1}$ and write down an ordinary five-point equation (4.3). Then construct interpolation formula (3.1) for the function $u(x,y)$ with respect to y coordinate, directing the axis $0t$ from $z_{i,j+1}$ into s_{ij} and assuming δ be equal to the distance δ_y from $z_{i,j+1}$ to s_{ij}, i.e., $\delta_y = \rho_2(z_{i,j+1})$. As a result, we obtain five-point grid equation with *the second asymmetric T-shaped stencil*:

$$\begin{aligned}&\left(\frac{4}{h^2} + d(z_{ij}) - \frac{3\delta_y}{(\delta_y+1)h^2}\right) u^h(z_{ij}) - \left(\frac{1}{h^2} - \frac{3\delta_y}{(2+\delta_y)h^2}\right) u^h(z_{i,j-1})\\ &-\frac{\delta_y}{(3+\delta_y)h^2}u^h(z_{i,j-2}) - \frac{1}{h^2}u^h(z_{i+1,j}) - \frac{1}{h^2}u^h(z_{i-1,j})\\ &= f(z_{ij}) + \frac{1}{h^2}\varphi_{1f}(\delta_y)g(s_{ij}).\end{aligned} \tag{4.6}$$

2) Let $z_{ij} \in \gamma^{31}_{1,h}$ and two nodes of the stencil small cross, for instance, $z_{i,j+1}$ and $z_{i+1,j}$ belong to $\gamma^2_{1,h}$ (see Fig. 4.). Both the values $u^h(z_{i,j+1})$ and $u^h(z_{i+1,j})$ are eliminated by means of formula (3.1). As a result, we obtain five-point grid equation with *the first asymmetric Γ-shaped stencil* (see Fig. 4.):

$$\begin{aligned}&\left(\frac{4}{h^2} + d(z_{ij}) - \frac{3\delta_x}{(\delta_x+1)h^2} - \frac{3\delta_y}{(\delta_y+1)h^2}\right) u^h(z_{ij})\\ &- \left(\frac{1}{h^2} - \frac{3\delta_x}{(2+\delta_x)h^2}\right) u^h(z_{i-1,j}) - \left(\frac{1}{h^2} - \frac{3\delta_y}{(2+\delta_y)h^2}\right) u^h(z_{i,j-1})\\ &-\frac{\delta_x}{(3+\delta_x)h^2}u^h(z_{i-2,j}) - \frac{\delta_y}{(3+\delta_y)h^2}u^h(z_{i,j-2})\\ &= f(z_{ij}) + \frac{1}{h^2}\varphi_{1f}(\delta_x)g(s_{ij_x}) + \frac{1}{h^2}\varphi_{1f}(\delta_y)g(s_{ij_y})\end{aligned} \tag{4.7}$$

where $\delta_x = \rho_1(z_{i+1,j})$ and $\delta_y = \rho_2(z_{i,j+1})$.

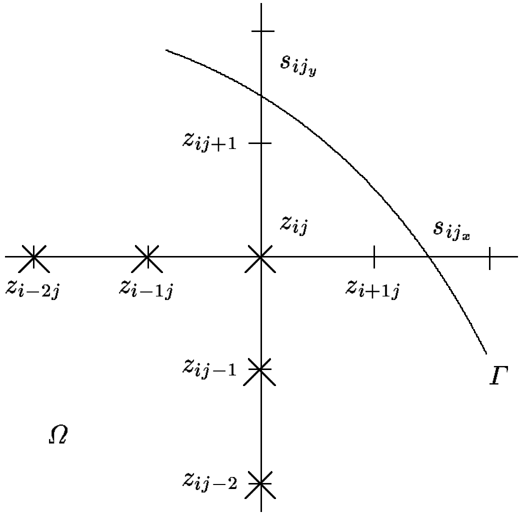


Fig. 4: The first Γ-shaped asymmetric stencil.
Cross sign marks the nodes of the stencil. $z_{i,j+1} \in \gamma^2_{1,h}$; $z_{i+1,j} \in \gamma^2_{1,h}$; $z_{ij} \in \gamma^{31}_{1,h}$.

3) Let $z_{ij} \in \gamma^{32}_{1,h}$, one node of the stencil small cross, for instance, $z_{i,j+1}$ belong to $\gamma^2_{1,h}$ and the second, for instance, $z_{i+1,j}$ belong to $\gamma^{out}_{1,h}$ (see Fig. 5.). Both the values $u^h(z_{i,j+1})$ and $u^h(z_{i+1,j})$ are eliminated by means of corresponding formula (3.1) or (3.2). As a result, we obtain five-point grid equation with *the second asymmetric Γ-shaped stencil* (see Fig. 5.):

$$\begin{aligned}&\left(\frac{4}{h^2} + d(z_{ij}) + \frac{3(1-\delta_x)}{\delta_x h^2} - \frac{3\delta_y}{(\delta_y+1)h^2}\right) u^h(z_{ij}) \\ &- \frac{\delta_y}{(3+\delta_y)h^2} u^h(z_{i,j-2}) - \left(\frac{1}{h^2} + \frac{3(1-\delta_x)}{(1+\delta_x)h^2}\right) u^h(z_{i-1,j}) \\ &- \left(\frac{1}{h^2} - \frac{3\delta_y}{(2+\delta_y)h^2}\right) u^h(z_{i,j-1}) + \frac{1-\delta_x}{(2+\delta_x)h^2} u^h(z_{i-2,j}) \\ &= f(z_{ij}) + \frac{1}{h^2}\varphi_{2f}(\delta_x) g(s_{ij_x}) + \frac{1}{h^2}\varphi_{1f}(\delta_y) g(s_{ij_y})\end{aligned} \tag{4.8}$$

where $\delta_x = \rho_1(z_{ij})$ and $\delta_y = \rho_2(z_{i,j+1})$.

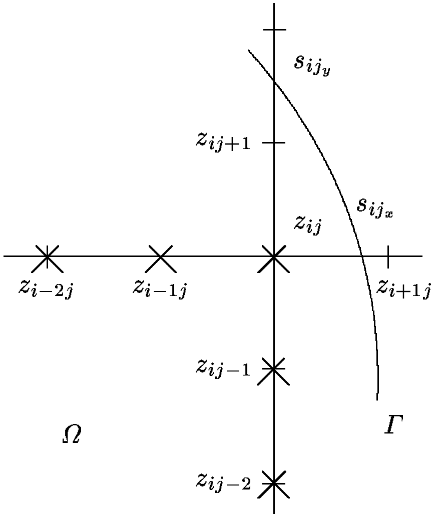


Fig. 5: The second Γ-shaped asymmetric stencil.

Cross sign marks the nodes of the stencil. $z_{i,j+1} \in \gamma^2_{1,h}$; $z_{i+1,j} \in \gamma^{out}_{1,h}$; $z_{ij} \in \gamma^{32}_{1,h}$.

Consider the equations in irregular nodes z_{ij} belonging to $\gamma_{2,h}$. There are two variants when one or two points of the stencil large cross belong to $\gamma^2_{1,h}$. Let $z_{ij} \in \gamma^3_{2,h}$ and $z_{i-2,j} \in \gamma^2_{1,h}$ (see Fig. 6.). Denote by s_{ij} the point of intersection of the boundary Γ and the ray $(z_{i-2,j}, z_{ij}]$. At the beginning assume that the solution $u(x,y)$ is determined in the point $z_{i-2,j}$ and write down five-point equation:

$$\begin{aligned}&\left(\frac{1}{h^2} + d(z_{ij})\right) u^h(z_{ij}) - \frac{1}{4h^2}u^h(z_{i-2,j}) - \frac{1}{4h^2}u^h(z_{i+2,j}) \\ &-\frac{1}{4h^2}u^h(z_{i,j-2}) - \frac{1}{4h^2}u^h(z_{i,j+2}) = f(z_{ij}).\end{aligned} \tag{4.9}$$

Then construct interpolation formula (3.1) for the function $u(x,y)$ with respect to y coordinate, directing the axis $0t$ from z_{ij} into $z_{i-2,j}$ and assuming δ be equal to the distance δ_y from $z_{i-2,j}$ to s_{ij}. As a result, we obtain six-point grid equation with *the first asymmetric X-shaped stencil*:

$$\begin{aligned}&\left(\frac{1}{h^2}+\frac{3\delta_y}{(\delta_y+2)4h^2}+d(z_{ij})\right)u^h(z_{ij})-\frac{3\delta_y}{(\delta_y+1)4h^2}u^h(z_{i-1,j})\\&-\frac{\delta_y}{(\delta_y+3)4h^2}u^h(z_{i+1,j})-\frac{1}{4h^2}u^h(z_{i+2,j})-\frac{1}{4h^2}u^h(z_{i,j+2})\\&-\frac{1}{4h^2}u^h(z_{i,j-2})=f(z_{ij})+\frac{1}{4h^2}\varphi_{1f}(\delta_y)g(s_{ij}).\end{aligned} \tag{4.10}$$

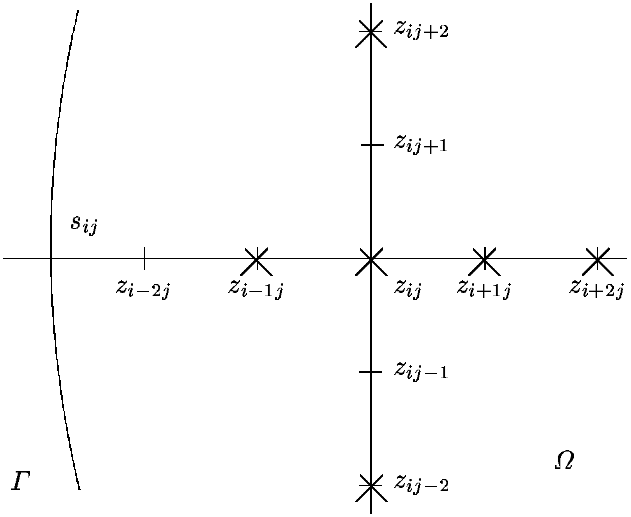


Fig. 6: The first X-shaped asymmetric stencil.

Cross-sign marks the nodes of the stencil. $z_{ij}\in\gamma^3_{2,h}$; $z_{i-2,j}\in\gamma^2_{1,h}$.

Consider the second variant. Let $z_{ij}\in\gamma^2_{2,h}$ and $z_{i-2,j}\in\gamma^2_{1,h}$, $z_{i,j+2}\in\gamma^2_{1,h}$ (see Fig. 7.). Denote by s_{ij_x} the point of intersection of the boundary Γ and the ray $(z_{i-2,j},z_{ij}]$, and by s_{ij_y} denote the point of intersection of the boundary Γ and the ray $(z_{i,j+2},z_{ij}]$. At the beginning assume that the solution $u(x,y)$ is determined in the points $z_{i-2,j}$, $z_{i,j+2}$ and write down five-point equation (4.9). Then by means of formula (3.1) eliminate the points belonging to $\gamma^2_{1,h}$. As a result, we obtain seven-point equation with *the second asymmetric X-shaped stencil:*

$$\begin{aligned}
&\left(\frac{1}{h^2}+\frac{3\delta_x}{(\delta_y+2)4h^2}+\frac{3\delta_y}{(\delta_y+2)4h^2}+d(z_{ij})\right)u^h(z_{ij})\\
&-\frac{3\delta_y}{(\delta_y+1)4h^2}u^h(z_{i,j+1})-\frac{3\delta_x}{(\delta_x+1)4h^2}u^h(z_{i-1,j})\\
&-\frac{\delta_y}{(\delta_y+3)4h^2}u^h(z_{i,j-1})-\frac{\delta_x}{(\delta_x+3)4h^2}u^h(z_{i+1,j})\\
&-\frac{1}{4h^2}u^h(z_{i,j-2})-\frac{1}{4h^2}u^h(z_{i,j+2})\\
&=f(z_{ij})+\frac{1}{4h^2}\varphi_{1f}(\delta_x)g(s_{ij_x})+\frac{1}{4h^2}\varphi_{1f}(\delta_y)g(s_{ij_y})
\end{aligned}\qquad(4.11)$$

where $\delta_x=\rho_1(z_{i-2,j})$, $\delta_y=\rho_2(z_{i,j+2})$.

In the rest of the nodes $z_{ij}\in\gamma^1_{2,h}$ the equations are constructed according to the following principle. Let $z_{ij}\in\gamma^1_{2,h}$, consequently, one adjacent node belongs to $\gamma^1_{1,h}$. Then in the point z_{ij} an equation similar to (4.6) can be constructed with elimination of the point belonging to $\gamma^1_{1,h}$.

Thus, in the result of these constructions a system of linear algebraic equations is obtained, which unites the equalities (4.3) – (4.8), (4.10), (4.11) taken in corresponding nodes. Write down this system in operator form

$$A^hu^h=f^h\quad\text{on}\quad\omega_h\setminus\gamma^2_{1,h}\qquad(4.12)$$

with sought for grid function $u^h(z_{ij})$ and known right hand side $f^h(z_{ij})$ with the argument $z_{ij}\in\omega_h\setminus\gamma^2_{1,h}$.

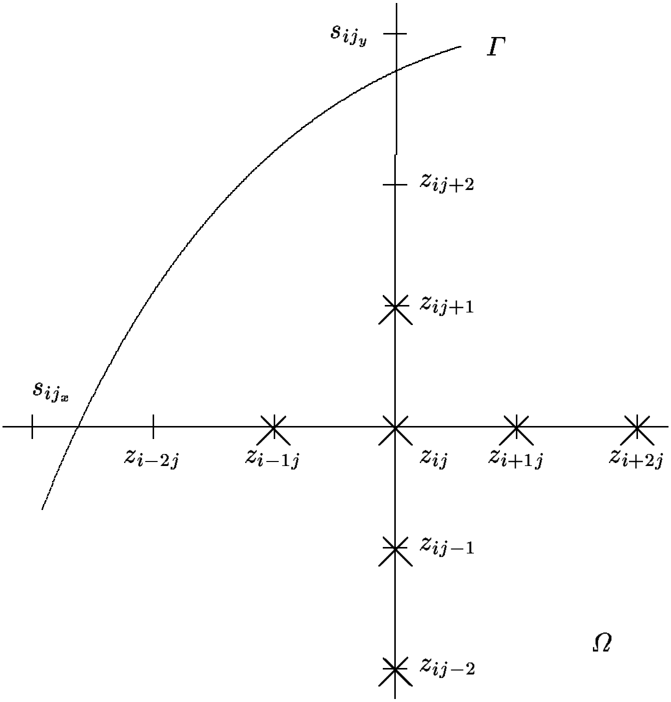


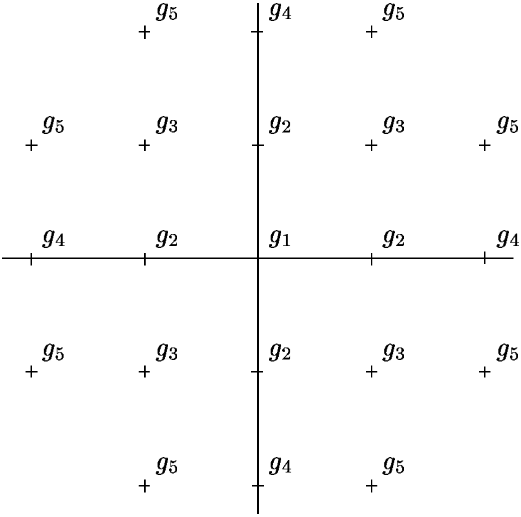
Fig. 7: The second X-shaped asymmetric stencil.

Cross sign marks the nodes of the stencil. $z_{ij} \in \gamma^2_{2,h}$; $z_{i-2,j} \in \gamma^2_{1,h}$; $z_{i,j+2} \in \gamma^2_{1,h}$.

5 Stability, solvability and convergence of the grid problem

Transform the system (4.12) so that its matrix would be M – matrix. At first, enumerate the nodes of the set $\omega_h \setminus \gamma^2_{1,h}$ from 1 to M and give corresponding numbers to the equations in the nodes $z_{ij} \in \omega_h \setminus \gamma^2_{1,h}$ and the variables $u^h(z_{ij})$. In order to utilize the standard results concerning M – matrices, it is necessary that diagonal elements would be positive and off-diagonal ones would be non-negative, and the sum of modules of off-diagonal elements would not exceed a diagonal element. For equations in the nodes $\omega^r_{h,1}$, $\gamma^{31}_{1,h}$,

$\gamma_{1,h}^{33}$, $\gamma_{2,h}$ these conditions are satisfied, but that is not true for equations in the nodes $\gamma_{1,h}^{1}$, $\gamma_{1,h}^{32}$, $\omega_{h,2}^{r}$.



$$g_1 = \frac{2}{h^2}, \quad g_2 = -\frac{1}{10h^2} + \frac{d}{4},$$
$$g_3 = -\frac{3}{10h^2} + \frac{d}{20}, \quad g_4 = 0,$$
$$g_5 = -\frac{1}{20h^2}.$$

Fig. 8: 21-point stencil of the equation in node $z_{ij} \in \omega_{h,2}^{r}$ after the transformation.

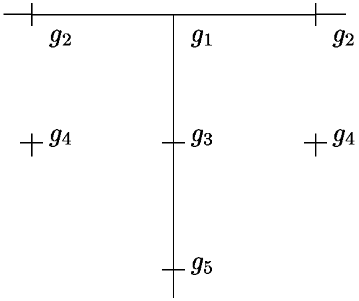
In order to eliminate positive off-diagonal elements, let add to each equation (4.4) in $z_{ij} \in \omega_{h,2}^{r}$ four equations in four regular nodes $z_{i\pm1,j\pm1} \in \omega_{h,1}^{r}$, with weight $a = 1/20$, and four equations in the nodes $z_{i\pm1,j} \in \omega_{h,1}^{r}$, $z_{i,j\pm1} \in \omega_{h,1}^{r}$, with weight $b = 1/4$ (for details see [11]). As a result, in the node $z_{ij} \in \omega_{h,2}^{r}$ we obtain an equation with the stencil shown in Fig. 8 (compare to Fig. 2.b). Then, in addition, require that the following inequality would be true:

$$h^2 \leq 2/\left(5\|d\|_{\infty,\overline{w}_h}\right). \tag{5.1}$$

It is easy to verify that under this condition we come to the following inequalities for coefficients of the new grid equation with extended stencil:

$$\begin{aligned}&g_1 \geq 0,\ g_2 \leq 0,\ g_3 \leq 0,\ g_4 = 0,\ g_5 \leq 0,\\&|g_1(z_{ij})| \geq |g_2(z_{i+1,j}) + g_2(z_{i-1,j}) + g_2(z_{i,j+1}) + g_2(z_{i,j-1})\\&+g_3(z_{i+1,j+1}) + g_3(z_{i-1,j+1}) + g_3(z_{i+1,j-1}) + g_3(z_{i-1,j-1}) + 8g_5|,\end{aligned}$$

which confirm both right signs of coefficients of the stencil and diagonal prevalence.

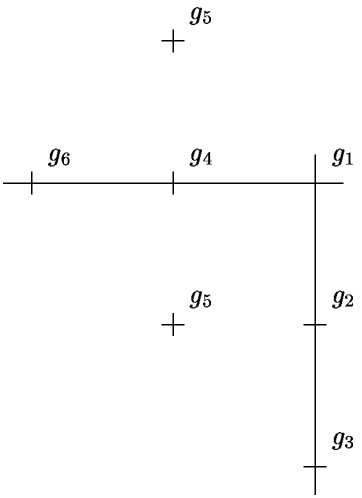


$$\begin{aligned}g_1 &= \frac{4}{h^2} + d + \frac{2(-\delta_y^2 + 2\delta_y + 3)}{\delta_y(2+\delta_y)h^2},\\g_2 &= -\frac{1}{h^2},\\g_3 &= -\frac{1}{h^2} - \frac{\delta_y^2 - 3\delta_y + 2}{(1+\delta_y)(2+\delta_y)h^2} + \frac{(1-\delta_y)d}{2+\delta_y},\\g_4 &= -\frac{1-\delta_y}{(2+\delta_y)h^2},\\g_5 &= 0.\end{aligned}$$

Fig. 9: 7-point stencil of the equation in the node $z_{ij} \in \gamma^1_{1,h}$ after the transformation.

Now, let $z_{ij} \in \gamma^1_{1,h}$ and $z_{i,j+1} \in \gamma^{out}_{1,h}$, and $z_{i,j-1} \in \omega^r_{1,h}$. Add to the equation (4.5) in the point z_{ij} one more equation (4.3) in the regular node $z_{i,j-1} \in \omega^r_{h,1}$, with weight $a = (1-\delta_y)/(2+\delta_y)$. As a result, in the point $z_{ij} \in \gamma^1_{1,h}$ we obtain an equation with seven-point stencil, as shown in Fig. 9 (compare to Fig. 3.a). It is easy to verify that under the condition (5.1) and taking into account that $\delta_y \in (0,1]$ we come to the inequalities

$$\begin{aligned}&g_1 \geq 0,\ g_2 \leq 0,\ g_3 \leq 0,\ g_4 \leq 0,\ g_5 \leq 0,\\&|g_1(z_{ij})| > |2g_2 + g_3(z_{i,j-1}) + 2g_4 + g_5|.\end{aligned}$$



$$
\begin{aligned}
g_1 &= \frac{4}{h^2} + d + \frac{2(-\delta_x^2 + 2\delta_x + 3)}{\delta_x(2+\delta_x)h^2} - \frac{3\delta_y}{(\delta_y+1)h^2},\\
g_2 &= -\left(\frac{1}{h^2} - \frac{3\delta_y}{(2+\delta_y)h^2}\right),\\
g_3 &= -\frac{\delta_y}{(3+\delta_y)h^2},\\
g_4 &= -\frac{1}{h^2} - \frac{\delta_x^2 - 3\delta_x + 2}{(1+\delta_x)(2+\delta_x)h^2} + \frac{(1-\delta_x)d}{2+\delta_x},\\
g_5 &= -\frac{1-\delta_x}{(2+\delta_x)h^2}, \quad g_6 = 0.
\end{aligned}
$$

Fig. 10: 7-point stencil of the equation in the node $z_{ij} \in \gamma_{1,h}^{32}$ after the transformation.

Let $z_{ij} \in \gamma_{1,h}^{32}$, $z_{i+1,j} \in \gamma_{1,h}^{out}$, and $z_{i-1,j} \in \omega_{h,1}^r$. Add to the equation (4.8) in the point z_{ij} one more equation (4.3) in the regular node $z_{i-1,j} \in \omega_{h,1}^r$ with weight $a = (1-\delta_x)/(2+\delta_x)$. As a result, in the point $z_{ij} \in \gamma_{1,h}^{32}$ we obtain an equation with seven-point stencil shown in Fig. 10 (compare to Fig. 5). Under the condition (5.1) and taking into account that $\delta_x \in (0,1]$ and $\delta_y \in (0,1]$ we obtain the inequalities

$$
\begin{aligned}
&g_1 \geq 0, \; g_2 \leq 0, \; g_3 \leq 0, \; g_4 \leq 0, \; g_5 \leq 0, \; g_6 \leq 0,\\
&|g_1(z_{ij})| > |g_2 + g_3 + g_4(z_{i-1,j}) + 2g_5 + g_6|.
\end{aligned}
$$

Thus, we obtain a system consisting of the equations corresponding to the points belonging to $\omega^r_{h,1}$, $\gamma^{31}_{1,h}$, $\gamma^{33}_{1,h}$ and $\gamma_{2,h}$, and transformed equations corresponding to the points belonging to $\omega^r_{h,2}$, $\gamma^1_{1,h}$ and $\gamma^{32}_{1,h}$. With the regard for the signs of diagonal and off-diagonal elements, diagonal prevalence and indecomposability [9], the matrix of the transformed system is M–matrix. The obtained equivalent grid problem can be written down as

$$B^h u^h = g^h \quad \text{on} \quad \omega_h \setminus \gamma^2_{1,h} \tag{5.2}$$

with the same unknown grid function u^h but with transformed right-hand side g^h.

Theorem 36. *Let the condition* (1.3) *be satisfied and the step h be small enough:*

$$h^2 \le 2/(5\|d\|_{\infty,\overline{\Omega}}). \tag{5.3}$$

Then for arbitrary right-hand side f^h the solution of the problem (4.12) *satisfies the estimate*

$$\|u^h\|_{\infty,\overline{\omega}_h \setminus \gamma^2_{1,h}} \le \frac{11}{48}\|f^h\|_{\infty,\omega^r_h} + \|f^h/S^h\|_{\infty,\omega^{ir}_h}, \tag{5.4}$$

where $S^h(z_{ij})$ is the sum of coefficients of the grid equation (4.12) *in the node z_{ij} and is strictly positive on ω^{ir}_h.*

Proof. Introduce a function

$$w_1(x,y) = c_1 x(1-x), \quad c_1 = \frac{11}{12}\|f^h\|_{\infty,\omega^r_h}. \tag{5.5}$$

Note that derivatives of the order 3 and higher of this function are equal to zero. Therefore the exact approximation of the difference operators L^h, L^{2h} and interpolation formulas is attained for this function. From this, under the condition (1.3) we have

$$L^h w_1 = L w_1 = d w_1 + 2c_1 \ge 2c_1 > 0 \quad \text{on} \quad \omega_h, \tag{5.6}$$

$$L^{2h} w_1 = L w_1 = d w_1 + 2c_1 \ge 2c_1 > 0 \quad \text{on} \quad \omega_{00}. \tag{5.7}$$

Taking into account (5.6), in regular nodes of the first kind we obtain

$$B^h w_1 = L w_1 \ge \frac{11}{6}\|f\|_{\infty,\omega^r_h} \ge A^h u^h = B^h u^h \quad \text{on} \quad \omega^r_{h,1}. \tag{5.8}$$

Similar expression for regular nodes of the second kind can be obtained by means of the rule of transformation of the operator A^h into B^h (detailed computations see in [11]):

$$B^h w_1 \geq \frac{12}{5} c_2 \geq \frac{11}{5} \|f\|_{\infty,\omega_h^r} \geq B^h u^h \quad \text{on} \quad \omega_{h,2}^r. \tag{5.9}$$

In irregular nodes, from the analysis of the rules of transformation of A^h into B^h (i.e., possible addition of a regular equation with weight $\leq 1/2$) with account of (5.6) or (5.7) it follows that

$$B^h w_1 \geq L w_1 \geq 2c_1 \quad \text{on} \quad \omega_h^{ir}. \tag{5.10}$$

Introduce a constant function

$$w_2(x,y) = c_2 \quad \text{where} \quad c_2 = \|f^h/S^h\|_{\infty,\omega_h^{ir}}. \tag{5.11}$$

After the substitution of it into the operator B^h, two possible situations take place: either coincidence or lack of coincidence of the equations for z_{ij} in (5.2) and (4.12). In the first case (when $z_{ij} \in \omega_{h,1}^r \cup \gamma_{1,h}^{31} \cup \gamma_{1,h}^{33} \cup \gamma_{2,h}$) we obtain

$$B^h w_2(z_{ij}) = A^h w_2(z_{ij}) = c_2 S^h(z_{ij}) \geq 0, \tag{5.12}$$

in irregular nodes being diagonal prevalence and the value of $S^h(z_{ij})$ being strictly positive. In the second case the equation in (5.2) is obtained from (4.12) by addition of regular equations with positive weights. Therefore (for $z_{ij} \in \omega_{h,2}^r \cup \gamma_{1,h}^1 \cup \gamma_{1,h}^{32}$) we have

$$B^h w_2(z_{ij}) \geq A^h w_2(z_{ij}) = c_2 S^h(z_{ij}) \geq 0, \tag{5.13}$$

the value of diagonal prevalence being not less in irregular nodes and $S^h(z_{ij})$ being strictly positive again. So, combining the inequalities (5.12) and (5.13) we obtain

$$B^h w_2(z_{ij}) \geq 0 \quad \text{on} \quad \omega_h^r, \tag{5.14}$$

$$B^h w_2(z_{ij}) \geq c_2 S^h(z_{ij}) \geq f^h(z_{ij}) \quad \text{on} \quad \omega_h^{ir}. \tag{5.15}$$

Summing up the inequalities (5.8) and (5.9) with (5.19), we come to the following expression in regular nodes

$$B^h(w_1 + w_2) \geq B^h u^h \quad \text{on} \quad \omega_h^r. \tag{5.16}$$

In irregular nodes this expression is obtained by summation of (5.10) with (5.15) and taking into account the rule of transformation of A^h into B^h:

$$B^h(w_1 + w_2) \geq 2c_1 + c_2 \geq 2c_1 + A^h u^h \geq B^h u^h \quad \text{on} \quad \omega_h^{ir}.$$

From two last inequalities on the basis of the comparison theorem [9] it follows that

$$w_1 + w_2 \geq u^h \quad \text{on} \quad \omega_h \setminus \gamma_{1,h}^2.$$

After the replacement of f^h with $-f^h$ the above reasonings give the evaluation

$$w_1 + w_2 \geq -u^h \quad \text{on} \quad \omega_h \setminus \gamma_{1,h}^2.$$

The two last evaluations can be combined into the inequality

$$|u^h| \leq w_1 + w_2 \quad \text{on} \quad \omega_h \setminus \gamma_{1,h}^2.$$

After taking maximum in the right-hand side over $[0,1] \times [0,1]$ we come to the evaluation (5.4). □

Theorem 1 conveys stability of the problem (4.12) with respect to the boundary values and right-hand side, and, besides that, from it naturally follows unique solvability, since corresponding to it uniform system admits only zero solution.

Lemma 1. *If the conditions* (1.3) *and* (5.3) *are satisfied, then there exists a constant c_3 independent from h and domain Ω, such that the value $c_3 S^h(z_{ij})$ in irregular node $z_{ij} \in \omega_h^{ir}$ majorizes the modules of all non-zero coefficients of the grid equation* (4.12) *corresponding to this node.*

Proof. Let consider in details only one variant, for instance, the equation (4.5) in the node $z_{ij} \in \gamma_{1,h}^1$. Computation of $S^h(z_{ij})$ with the account of (1.3) and (3.2) gives the evaluation

$$S^h(z_{ij}) = d(z_{ij}) + \frac{1}{h^2}\varphi_{2f}(\delta_y) \geq \frac{6}{\delta_y(1+\delta_y)(2+\delta_y)h^2}. \tag{5.17}$$

For any $\delta_y \in (0,1]$ we have

$$S^h(z_{ij}) \geq \frac{1}{\delta_y h^2} \geq \frac{1}{h^2}. \tag{5.18}$$

From this and (5.3) we obtain

$$\frac{2}{5}S^h(z_{ij}) \geq \frac{2}{5h^2} \geq d(z_{ij}). \tag{5.19}$$

Except that, from (5.18) follow the inequalities

$$3S^h(z_{ij}) \geq \frac{3}{\delta_y h^2} \geq \frac{3(1-\delta_y)}{\delta_y h^2}.$$

By summation of the three last inequalities (the first with the factor 4), we obtain

$$\frac{37}{5}S^h(z_{ij}) \geq \frac{4}{h^2} + d(z_{ij}) + \frac{3(1-\delta_y)}{\delta_y h^2}. \tag{5.20}$$

Thus, the expression in the left-hand side majorizes the positive diagonal coefficient. It is easy to verify that it majorizes the modules of the other four coefficients of the equation (4.5).

So, the statement of the lemma is proved for the nodes $\gamma^1_{1,h}$ with constant 37/5. Similarly to the reasonings in (5.17) – (5.20), the existence of such constants for other kinds of irregular nodes can be proved. Denoting the maximal of them by c_3, we complete the proof of the Lemma. □

Corollary 1. Looking through the equations (4.5) – (4.8), (4.10) and (4.11) one can make sure that each of them contains a coefficient with absolute value not less than $1/h^2$ (in (4.7), (4.8), (4.10) and (4.11) that is diagonal coefficient). Therefore from Lemma 1 it follows that

$$S^h(z_{ij}) \geq \frac{1}{c_3 h^2} \quad \text{for} \quad z_{ij} \in \omega_h^{ir}. \tag{5.21}$$

Theorem 37. *Let* u, u^h *be the solutions of the problems* (1.1) – (1.2) *and* (4.12), *respectively, and the conditions* (1.3), (1.4), (5.3) *be satisfied. Then*

$$\|u - u^h\|_{\infty,\bar{\omega}_h \setminus \gamma^2_{1,h}} \leq Ch^4 \tag{5.22}$$

where constant C *is independent of* h.

Proof. Let show that the solution u^h can be represented in the form:

$$u^h = u + h^4\rho^h \quad \text{on} \quad \omega_{11} \setminus \gamma^2_{1,h}, \tag{5.23}$$

$$u^h = u + h^4 w_{01} + h^4\rho^h \quad \text{on} \quad (\omega_{01} \cup \omega_{10}) \setminus \gamma^2_{1,h}, \tag{5.24}$$

$$u^h = u + h^4 w_{00} + h^4\rho^h \quad \text{on} \quad \omega_{00} \setminus \gamma^2_{1,h}, \tag{5.25}$$

$$\tag{5.26}$$

where the functions

$$w_{01} = -\frac{1}{48}\mu, \quad w_{00} = -\frac{1}{12}\mu, \quad \mu = \frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4} \tag{5.27}$$

do not depend on h, and the remainder term ρ^h is limited in the following way:

$$\left\|\rho^h\right\|_{\infty,\bar{\omega}_h \setminus \gamma^2_{1,h}} \leq c_4. \tag{5.28}$$

The proof of the representations (5.23) – (5.27) is obtained by complication of proof of the Theorem 4 from the work [11]. Indeed, on the basis of the computations given there one can obtain equalities in regular nodes ω_h^r

$$A^h \rho^h = L^h \rho^h = \xi^h \quad \text{on} \quad \omega_{h,1}^r, \tag{5.29}$$

$$A^h \rho^h = L^h \rho^h - L^{2h} \rho^h = \xi^h \quad \text{on} \quad \omega_{h,2}^r \tag{5.30}$$

with a grid function

$$|\xi^h| \le c_5 \quad \text{on} \quad \omega_h^r. \tag{5.31}$$

Let consider in details the situation in irregular nodes after the example of the grid equation (4.5) in the node $z_{ij} \in \gamma_{1,h}^1$. Substitute the expansions (5.23) – (5.25) into the expression $L^h u^h(z_{ij})$ and for the function u perform the expansion into Taylor series with respect to z_{ij} with the remainder term of the order h^4. In the node $z_{i,j+1}$ lying outside $\overline{\Omega}$ the value of $u^h(z_{i,j+1})$ is determined as (one-dimensional) Taylor series with respect to s_{ij} up to the derivative $\partial^4 u / \partial y^4$ inclusive. Then for the function u interpolation formula (3.2) with remainder term (3.3) and multiplied by $1/h^2$ is used. As a result, we obtain the equality

$$A^h u(z_{ij}) = f^h(z_{ij}) + h^2 \zeta(z_{ij}) \tag{5.32}$$

with the evaluation of the remainder term

$$|\zeta(z_{ij})| \le c_6. \tag{5.33}$$

Consider terms of the form $h^4 w_{01}$ and $h^4 w_{00}$ in the expansions (5.24), (5.25). On the basis of (5.27) they are evaluated as

$$h^4 \max_{\overline{\Omega}} \{|w_{00}|, |w_{01}|\} \le \frac{h^4}{12} \|\mu\|_{\infty,\overline{\Omega}}. \tag{5.34}$$

On the basis of Lemma 1, under any possible arrangement of these terms on the stencil of the equation (4.5) (see Fig. 3a) the result $\eta(z_{ij}))$ of linear combination of these terms with corresponding coefficients of the grid equation (4.5) can be evaluated as

$$|\eta(z_{ij})| \le 5 c_3 S^h(z_{ij}) \frac{h^4}{12} \|\mu\|_{\infty,\overline{\Omega}}. \tag{5.35}$$

Thus, after substitution of (5.23) – (5.25) into the expressions $A^h u^h(z_{ij})$ reduce a part of terms due to (4.12) and (5.32), divide the others terms by h^4 and group together the terms with $\zeta(z_{ij})$ and $\eta(z_{ij})$ into one remainder term ξ^h:

$$A^h \rho^h(z_{ij}) = \xi^h(z_{ij}), \quad z_{ij} \in \gamma_{1,h}^1. \tag{5.36}$$

Due to (5.33), (5.35), and Corollary 1 we have the evaluation

$$|\xi^h(z_{ij})| \le h^{-2}|\zeta(z_{ij})| + h^{-4}|\eta(z_{ij})| \le c_3(c_6 + 5/12\,\|\mu\|_{\infty,\overline{\Omega}})S^h(z_{ij}). \quad (5.37)$$

Similar expressions are obtained in other kinds of irregular nodes. Finally,

$$A^h\rho^h = \xi^h \quad \text{on} \quad \omega_h^{ir} \quad (5.38)$$

with a grid function ξ^h for which the following evaluation is valid:

$$|\xi^h(z_{ij})| \le c_7 S^h(z_{ij}), \quad z_{ij} \in \omega_h^{ir}, \quad (5.39)$$

where $c_7 = c_3(c_6 + 7/12\,\|\mu\|_{\infty.\overline{\Omega}})$.

In the end we arrive at the system of equations (5.29), (5.30), and (5.38), which uniquely determines the grid function ρ^h. On the basis of Theorem 1 we obtain the evaluation

$$\|\rho^h\|_{\infty,\overline{\omega}_h\setminus\gamma^2_{1,h}} \le \frac{11}{48}\|\xi^h\|_{\infty,\omega_h^r} + \|\xi^h/S^h\|_{\infty,\omega_h^{ir}}. \quad (5.40)$$

From it, due to (5.31) and (5.39), (5.28) follows with constant

$$c_4 = 11/48\; c_5 + c_7.$$

The final affirmation of (5.22) follows from (5.23) – (5.25) with use of (5.28) and (5.34). □

6 Numerical examples

As in [11], let apply the constructed method to two problems of the form (1.1) – (1.2) with improved smoothness and with oscillating solution. Let the domain Ω be bounded by a circumference Γ with center in point $(0.5, 0.5)$ and radius 0.49.

The problem I has the form

$$\begin{aligned}-\Delta u = {} & 2\cos\left(\frac{\pi x}{2}\right) y(1-y)\cos\left(\frac{\pi y}{2}\right)\\ & + (1-x)\sin\left(\frac{\pi x}{2}\right)\pi y(1-y)\cos\left(\frac{\pi y}{2}\right)\\ & - x\sin\left(\frac{\pi x}{2}\right)\pi y(1-y)\cos\left(\frac{\pi y}{2}\right)\\ & + \frac{1}{2}x(1-x)\cos\left(\frac{\pi x}{2}\right)\pi^2 y(1-y)\cos\left(\frac{\pi y}{2}\right)\\ & + 2x(1-x)\cos\left(\frac{\pi x}{2}\right)\cos\left(\frac{\pi y}{2}\right)\\ & + x(1-x)\cos\left(\frac{\pi x}{2}\right)(1-y)\sin\left(\frac{\pi y}{2}\right)\pi\\ & - x(1-x)\cos\left(\frac{\pi x}{2}\right) y\sin\left(\frac{\pi y}{2}\right)\pi \quad \text{in} \quad \Omega,\\ u = {} & g \quad \text{on} \quad \Gamma\end{aligned} \tag{I}$$

with a function g being equal on Γ to the exact solution

$$u(x,y) = x(1-x)\cos\left(\frac{\pi x}{2}\right) y(1-y)\cos\left(\frac{\pi y}{2}\right).$$

The problem II has the form

$$\begin{aligned}-\Delta u = {} & -32c_x(1-x)y(1-y) + 512s_x x(1-x)y(1-y)\\ & +32c_x xy(1-y) + 2s_x y(1-y) - 32c_x x(1-x)(1-y)\\ & +32c_x x(1-x)y + 2s_x x(1-x) \quad \text{in} \quad \Omega,\\ u = {} & g \quad \text{on} \quad \Gamma\end{aligned} \tag{II}$$

where $c_x = \cos(16x+16y)$ and $s_x = \sin(16x+16y)$. The function g on Γ is equal to the exact solution as well $u(x,y) = \sin(16x+16y)x(1-x)y(1-y)$.

In Fig. 11 a quarter of the domain Ω is shown for $N = 44$.

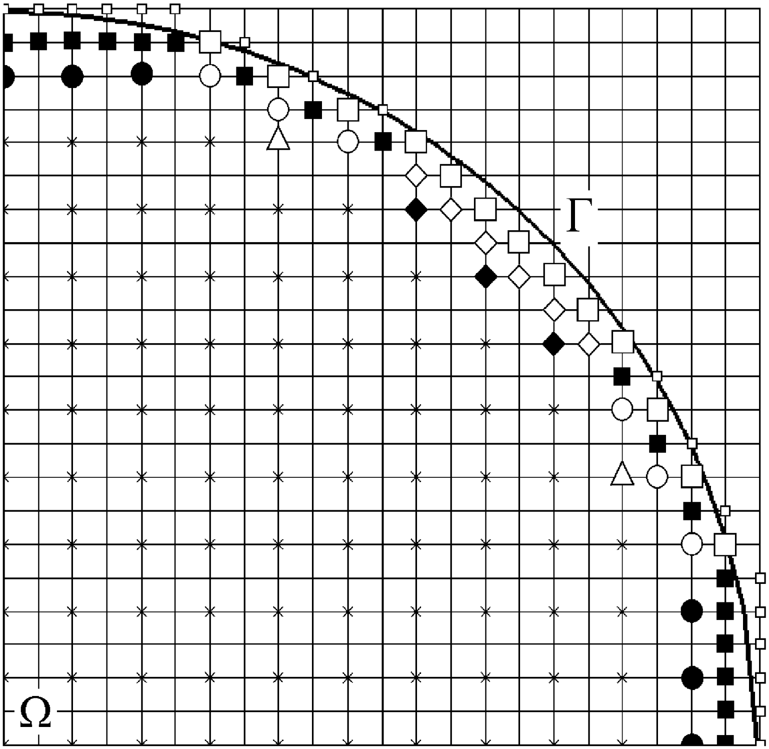


Fig. 11: Scheme of possible arrangement of kinds of nodes on the grid ω_h;

Here new symbols are introduced:

+ — $z_{ij} \in \omega^r_{h,1}$; ✳ — $z_{ij} \in \omega^r_{h,2}$; ⊞ — $z_{ij} \in \gamma^{out}_{1,h}$;
■ — $z_{ij} \in \gamma^1_{1,h}$; □ — $z_{ij} \in \gamma^2_{1,h}$; ● — $z_{ij} \in \gamma^1_{2,h}$;
○ — $z_{ij} \in \gamma^{32}_{1,h} \cup \gamma^{33}_{1,h}$; △ — $z_{ij} \in \gamma^3_{2,h}$; ◇ — $z_{ij} \in \gamma^{31}_{1,h}$;
◆ — $z_{ij} \in \gamma^2_{2,h}$.

The data of the numerical experiment are presented in Table 1.

Table 1: Error of approximate solutions of the problem with improved smoothness.

N	method of the fourth order		method of the second order	
	Ψ_1	Ψ_2	Ψ_1	Ψ_2
10	$5.84_{10}-04$	$1.68_{10}-04$	$4.39_{10}-04$	$2.04_{10}-04$
14	$6.32_{10}-05$	$1.50_{10}-05$	$1.84_{10}-04$	$8.40_{10}-05$
18	$4.24_{10}-05$	$9.22_{10}-06$	$1.07_{10}-04$	$4.79_{10}-05$
20	$1.72_{10}-05$	$6.09_{10}-06$	$8.64_{10}-05$	$3.86_{10}-05$
28	$4.32_{10}-06$	$1.31_{10}-06$	$4.37_{10}-05$	$1.95_{10}-05$
30	$2.18_{10}-06$	$4.79_{10}-07$	$3.80_{10}-05$	$1.70_{10}-05$
32	$4.31_{10}-06$	$9.11_{10}-07$	$3.35_{10}-05$	$1.49_{10}-05$
36	$1.55_{10}-06$	$3.85_{10}-07$	$2.64_{10}-05$	$1.18_{10}-05$
40	$1.02_{10}-06$	$2.77_{10}-07$	$2.13_{10}-05$	$9.48_{10}-06$
56	$2.52_{10}-07$	$5.29_{10}-08$	$1.08_{10}-05$	$4.82_{10}-06$
60	$2.88_{10}-07$	$4.87_{10}-08$	$9.17_{10}-06$	$4.07_{10}-06$
64	$2.59_{10}-07$	$4.29_{10}-08$	$7.09_{10}-06$	$3.10_{10}-06$

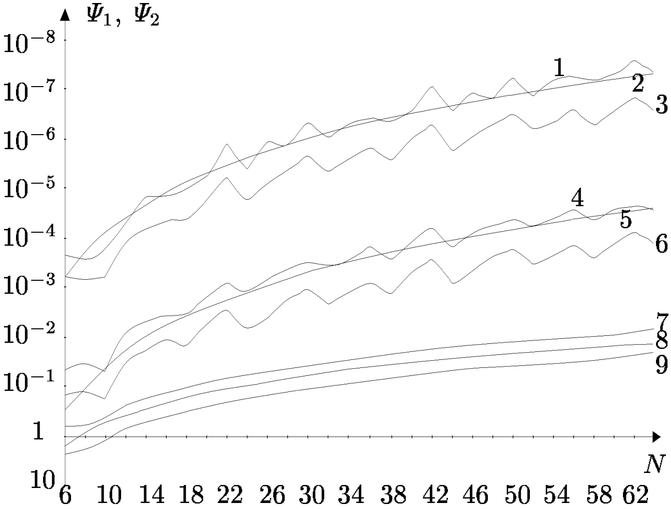


Fig. 12: Error of approximate solutions of the problems I and II.

In Fig. 12 the results of numerical experiments are shown in logarithmic scale over the $Y-$axis. The numbers 1, 4 and 7 mark mean square error

$$\Psi_1 = \|u-u^h\|_{2,\overline{\omega}_h\setminus\gamma^2_{1,h}} = \left(\sum_{z\in\overline{\omega}_h\setminus\gamma^2_{1,h}} \left(u(z)-u^h(z)\right)^2\right)^{1/2}$$

for the problems I and II, solved by the proposed in the present paper method, and for the problem I, solved by a standard method with the second order of accuracy, respectively [2], [6]. The numbers 3, 6 and 9 mark uniform errors

$$\Psi_2 = \|u - u^h\|_{\infty,\overline{\omega}_h \setminus \gamma^2_{1,h}}$$

for the problems I and II, solved by the method proposed in the present paper, and for the problem I, solved by a standard method with the second order of accuracy, respectively. The numbers 2, 5 and 8 mark diagrams of the curves $\delta = c_1 h^4$, $\delta = c_2 h^4$ and $\delta = h^2$, respectively.

References

1. Bykova E.G., Shaidurov V.V.: *Two-dimensional nonuniform difference scheme with higher order of accuracy.* Computational technologies, Novosibirsk, 1997, vol. 2, № 5, pp. 12–25 (In Russian).
2. Volkov E.A.: *A study of one method of improvement of accuracy of grid method when solving Poisson equation.* In: Computational mathematics, Moscow, 1957, № 1, pp. 62–80 (In Russian).
3. Volkov E.A.: *Solution of boundary-value problem by the method of more precise determination by the differences of higher orders, I.* Diff. equations, 1965, vol. 1, № 7, pp. 946–960 (In Russian).
4. Schortley G., Weller R.: *The numerical solution of the Laplace's equation.* J. Appl. Phys., 1938, vol. 9, № 5, pp. 334–348.
5. Mikeladze Sh.E.: *On numerical integration of elliptic and parabolic equations.* Izv. AN SSSR. Ser. matem., 1941, vol. 5, № 1, pp. 57-73 (In Russian).
6. Marchuk G.I., Shaidurov V.V.: *Improvement of the accuracy of solutions of difference schemes.* Moscow, Nauka, 1979 (In Russian).
7. Rüde U.: *Extrapolation and Related Techniques for Solving Elliptic Equations.* Preprint № I–9135, München Technical University, 1991.
8. Bykova E.G., Shaidurov V.V.: *Nonuniform difference scheme of higher order of accuracy. One-dimensional illustrative example.* Preprint № 17 of the Computing Center of SB RAS, Krasnoyarsk, 1996 (In Russian).
9. Samarsky A.A.: *The theory of difference schemes.* Moscov, Nauka, 1977 (In Russian).
10. Bakhvalov N.S.: *Numerical methods.* Moscow, Nauka, 1975 (In Russian).
11. Bykova E.G., Shaidurov V.V.: *A two-dimensional nonuniform difference scheme with higher order of accuracy.* In: This book.

Experimental analysis of fourth-order schemes for Poisson's equation

Bykova E.G., Rüde U., Shaidurov V.V

Introduction

This is not the first attempt to perform a comparison of numerical schemes for Poisson's equation [3]. However, during the last few years some new approach had been developed which was not studied experimentally in a comparison. Here, we consider several finite-difference schemes for Poisson's equation with Dirichlet boundary condition and evaluate them for three different types of solution: smooth, oscillatory and exponentially growing. The results are evaluated in the discrete L_2-, L_∞- and energy norms. In all computations, the problem is discretized on uniform square mesh (divided into triangles, if necessary). Of course, a uniform mesh does not permit to demonstrate the ability of some methods to adapt for an arbitrary (triangle or quadrangle) meshes. Moreover, different methods on a uniform mesh may result in same discrete algebraic systems if they are combined with appropriate quadrature rules for the right-hand side. Nevertheless, even these simple comparisons yield interesting insights.

1 Formulation of the differential problems

Let $\Omega = (0,1) \times (0,1)$ be the unit square with the boundary Γ. Consider the Dirichlet problem

$$-\Delta u = f \quad \text{in} \quad \Omega, \tag{I}$$

$$u = 0 \quad \text{on} \quad \Gamma. \tag{II}$$

We shall treat three examples with known exact solution. (The first and second examples are taken from [3]).

Example 1. Let

$$\begin{aligned} f(x,y) := f_1(x,y) &= c_x c_y (2y(1-y) + 2x(1-x) + \pi^2 x(1-x)y(1-y)/2) \\ &+ s_x c_y \pi(1-2x)y(1-y) + c_x s_y \pi(1-2y)x(1-x) \qquad \text{(III)} \end{aligned}$$

where

$$\begin{aligned} s_x &= \sin(\pi x/2),\ c_x = \cos(\pi x/2), \\ s_y &= \sin(\pi y/2),\ c_y = \cos(\pi y/2). \end{aligned}$$

This right-hand side gives rise to a comparatively smooth solution of problem (I)–(II):

$$u(x,y) := u_1(x,y) = x(1-x)\cos(\pi x/2)y(1-y)\cos(\pi y/2). \qquad \text{(IV)}$$

Example 2. Let

$$\begin{aligned} f(x,y) := f_2(x,y) &= -32c(1-2x)y(1-y) \\ &+ 512sx(1-x)y(1-y) + 2sy(1-y) \qquad \text{(V)} \\ &- 32cx(1-x)(1-2y) + 2sx(1-x) \end{aligned}$$

where

$$s = \sin(16x + 16y),\ c = \cos(16x + 16y).$$

With this right-hand side we obtain an oscillatory solution of problem (I)–(II):

$$u(x,y) := u_2(x,y) = \sin(16x + 16y)x(1-x)y(1-y). \qquad \text{(VI)}$$

Example 3. Let

$$\begin{aligned} f(x,y) := f_3(x,y) &= (x(1-x)y(y+3) \\ &+ x(x+3)y(1-y))e^{x+y}. \qquad \text{(VII)} \end{aligned}$$

For this right-hand side we obtain an exponentially growing but comparatively smooth solution of problem (I)–(II):

$$u(x,y) := u_3(x,y) = x(1-x)y(1-y)e^{x+y}. \qquad \text{(VIII)}$$

2 Tested methods

Let

$$\bar{\Omega}_h = \{z_{ij} : z_{ij} = (x_i, y_j);\ x_i = ih,\ i = 0, 1, \ldots, n;\ y_j = jh,\ j = 0, 1, \ldots, n\}$$

be uniform square grid with mesh-size $h = 1/n$. Let also

$$\Omega_h = \{z_{ij} : z_{ij} \in \bar{\Omega}_h \bigcap \Omega\}$$

and

$$\Gamma_h = \{z_{ij} : z_{ij} \in \bar{\Omega}_h \bigcap \Gamma\}.$$

To simplify the notation, we shall use the shortening

$$v_{ij} = v(z_{ij}) = v(x_i, y_j).$$

2.1 Five-point scheme and Richardson extrapolation

Here we use standard scheme

$$\begin{aligned} &\frac{4}{h^2}u^h_{i,j} - \frac{1}{h^2}u^h_{i+1,j} - \frac{1}{h^2}u^h_{i-1,j} - \frac{1}{h^2}u^h_{i,j+1} - \frac{1}{h^2}u^h_{i,j-1} = f_{ij}, && \text{(I)} \\ &i, j = 1, \ldots, n-1, \quad \text{i.e.,} \quad z_{ij} \in \Omega_h; \\ &u^h_{ij} = 0 \quad \text{if} \quad z_{ij} \in \Gamma_h. && \text{(II)} \end{aligned}$$

Of course, the solution of this problem has only second order of accuracy. However, using Richardson extrapolation the accuracy can be improved. For this purpose we assume n to be even and solve one more auxiliary problem (I)–(II) with mesh-size $2h$. Then we take both solutions u^h and u^{2h} and form a linear combination

$$u^{Rich}(z) = \frac{4}{3}u^h(z) - \frac{1}{3}u^{2h}(z) \quad \forall z \in \bar{\Omega}_{2h}. \qquad \text{(III)}$$

According to the theory, this combination has fourth order of accuracy in the discrete L_∞-norm [4].

2.2 Nonhomogeneous Bykova-Shaidurov scheme

This discretization uses different stencils at different grid points [5], [6]. Let again n be even. In the nodes (i, j) with both i and j even, this scheme has the form

$$\begin{aligned}&\frac{3}{h^2}u^h_{i,j} - \frac{1}{h^2}u^h_{i+1,j} - \frac{1}{h^2}u^h_{i-1,j} - \frac{1}{h^2}u^h_{i,j+1} - \frac{1}{h^2}u^h_{i,j-1}\\ &+\frac{1}{4h^2}u^h_{i+2,j} + \frac{1}{4h^2}u^h_{i-2,j} + \frac{1}{4h^2}u^h_{i,j+2} + \frac{1}{4h^2}u^h_{i,j-2} = 0, \qquad \text{(IV)}\\ &i,j = 2,4,\dots,n-2.\end{aligned}$$

At the rest nodes of Ω_h we use equations (I) and, finally, on the boundary nodes Γ_h we use equation (II). In [5] the fourth order of accuracy is proved in discrete L_∞-norm.

2.3 Khoromskij combination

The method is similar to Richardson extrapolation and uses solutions of two difference schemes [7]. But this time we perform the computation on the same grid and n is not necessarily even. The first scheme coincides with (I)–(II). The second one uses the oblique 5-point cross:

$$\begin{aligned}&\frac{1}{2h^2}(4\bar{u}^h_{i,j} - \bar{u}^h_{i+1,j+1} - \bar{u}^h_{i-1,j-1} - \bar{u}^h_{i-1,j+1} - \bar{u}^h_{i+1,j-1}) = f_{ij}, \qquad \text{(V)}\\ &i,j = 1,\dots,n-1;\\ &\bar{u}^h_{i,j} = 0 \quad \text{if} \quad z_{ij} \in \Gamma_h. \qquad \text{(VI)}\end{aligned}$$

Then we form the linear combination

$$u^{Khor}(z) = \frac{2}{3}u^h(z) + \frac{1}{3}\bar{u}^h(z) - \frac{h^2}{12}f(z) \quad \forall z \in \bar{\Omega}_h. \qquad \text{(VII)}$$

According to the proof in [7], this combination has fourth order of accuracy in discrete L_∞-norm.

2.4 Nine-point box scheme

This scheme uses only one grid and is homogeneous in the sense that it exploits only one 9-point stencil over all inner nodes of the grid [1], [2]. We apply it in the following form:

$$\begin{aligned}&\frac{1}{6h^2}(20u^h_{i,j} - 4u^h_{i+1,j} - 4u^h_{i-1,j} - 4u^h_{i,j+1} - 4u^h_{i,j-1}\\ &-u^h_{i+1,j+1} - u^h_{i-1,j-1} - u^h_{i-1,j+1} - u^h_{i+1,j-1}) = f_{i,j} + \frac{h^2}{12}(\Delta f)_{i,j}, \qquad \text{(VIII)}\\ &i,j = 1,\dots,n-1.\end{aligned}$$

For the boundary nodes Γ_h we again use equations (II). Note, that right-hand side in (VIII) is often used in the form

$$\frac{2}{3}f_{i,j} + \frac{1}{12}f_{i+1,j} + \frac{1}{12}f_{i-1,j} + \frac{1}{12}f_{i,j+1} + \frac{1}{12}f_{i,j-1}.$$

The difference between them is of fourth order of smallness and therefore they both give the same fourth order of accuracy for the difference solution in the discrete L_∞-norm [1], [2]. From practical point of view, the last value is preferable since does not involve an analytical modification of the right-hand side. But it contains difference differentiation in an implicit form. In order to eliminate the additional truncation error, we have used (VIII) with the exact analytical differentiation in all our problems.

3 Two ways to compare the computational cost

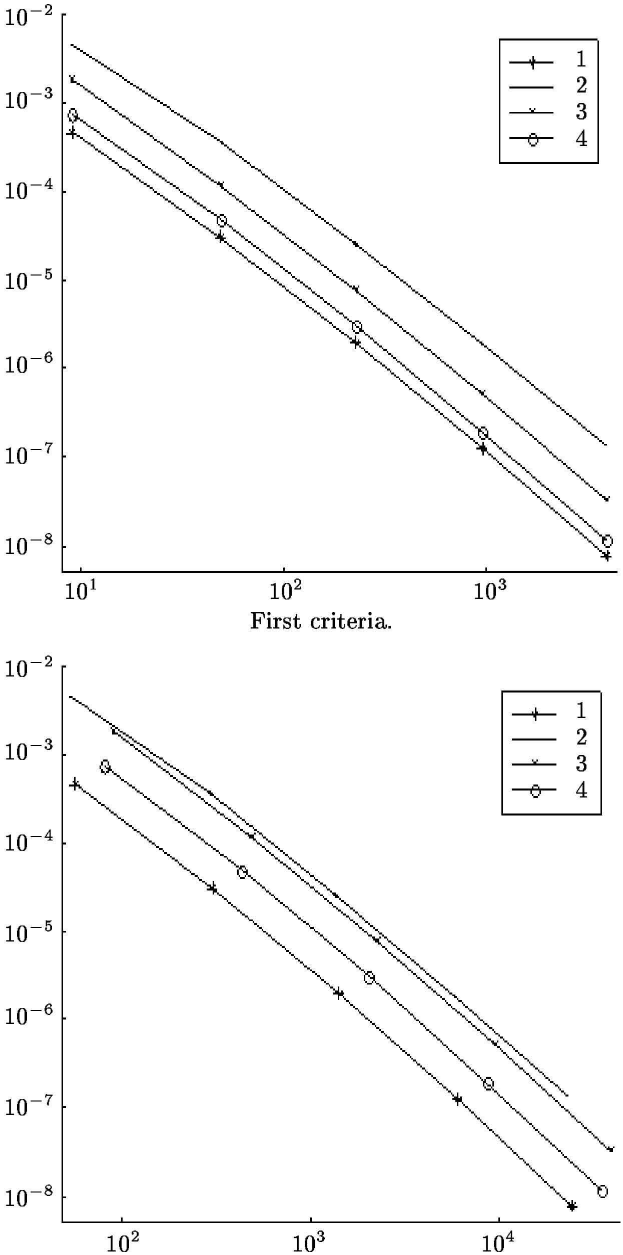
The traditional basis for a comparison is simply to use the number of unknowns as a measure of complexity. So we simply use the same grids with number of inner nodes $(n-1)^2$ for all example problems. Therefore, we performed the computation for $n = 2, 4, 8, 16, 32, 64$ and display the results for Example 1 in fig. 1 (top), 2 (top), and 3 (top) which correspond to the evaluated discrete energy-, L_∞-, and L_2-norms, respectively. The figures plot the error versus the number of mesh points. In each figure
line 1 (marked by asterisks) demonstrates Richardson extrapolation,
line 2 (marked by dots) corresponds to Bykova-Shaidurov scheme,
line 3 (marked by crosses) demonstrates Khoromskij combination,
line 4 (marked by circles) corresponds to nine-point box scheme.

The second comparison is based on the number of non-zero coefficients of the system matrices. This number is the amount of input data for the iterative process and should be useful for the evaluation of the complexity of smoother iterations (s.f. [8]). This point of view changes the situation, since the different methods on the same $(n-1) \times (n-1)$ grid result in the following number of coefficients:

in Richardson extrapolation	$6.25n^2$,
in Shaidurov-Bykova scheme	$6n^2$,
in Khoromskij combination	$10n^2$,
in nine-point box scheme	$9n^2$.

The result for this approach are displayed for Example 1 in figures 1 (bottom), 2 (bottom), and 3 (bottom) for the discrete energy-, L_∞-, and L_2-norms, respectively.

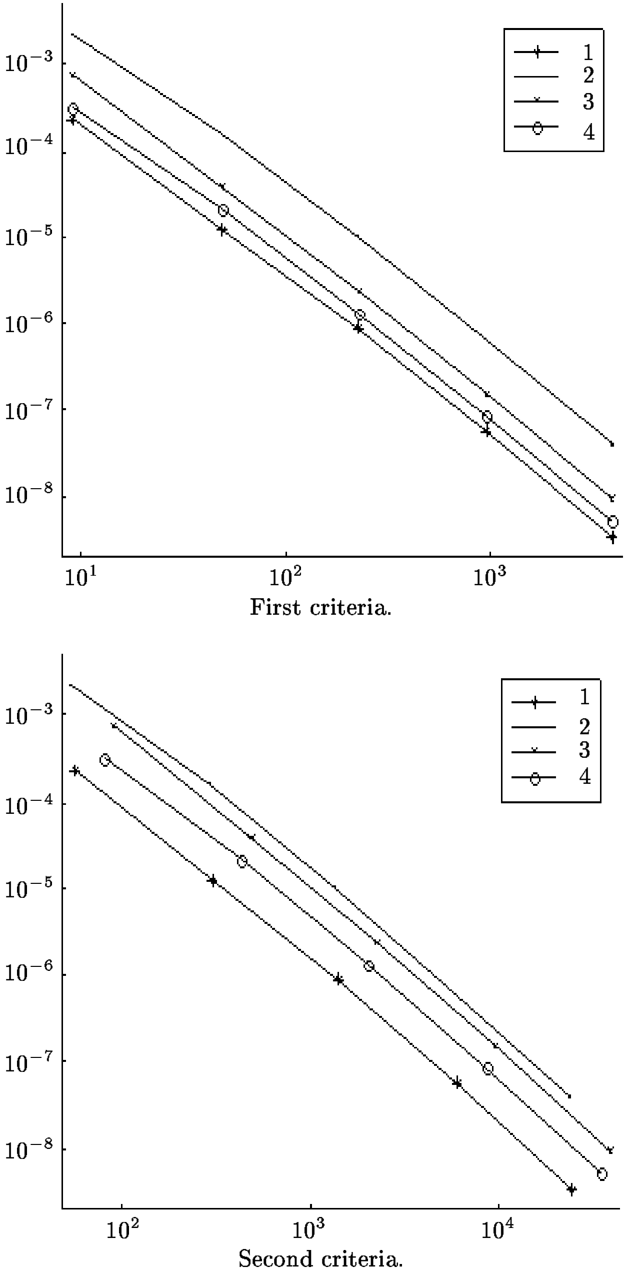
From the figures 1, 2, 3 one can see that the difference between first and second comparison criteria of is not significant for the relative ranking



First criteria.

Second criteria.

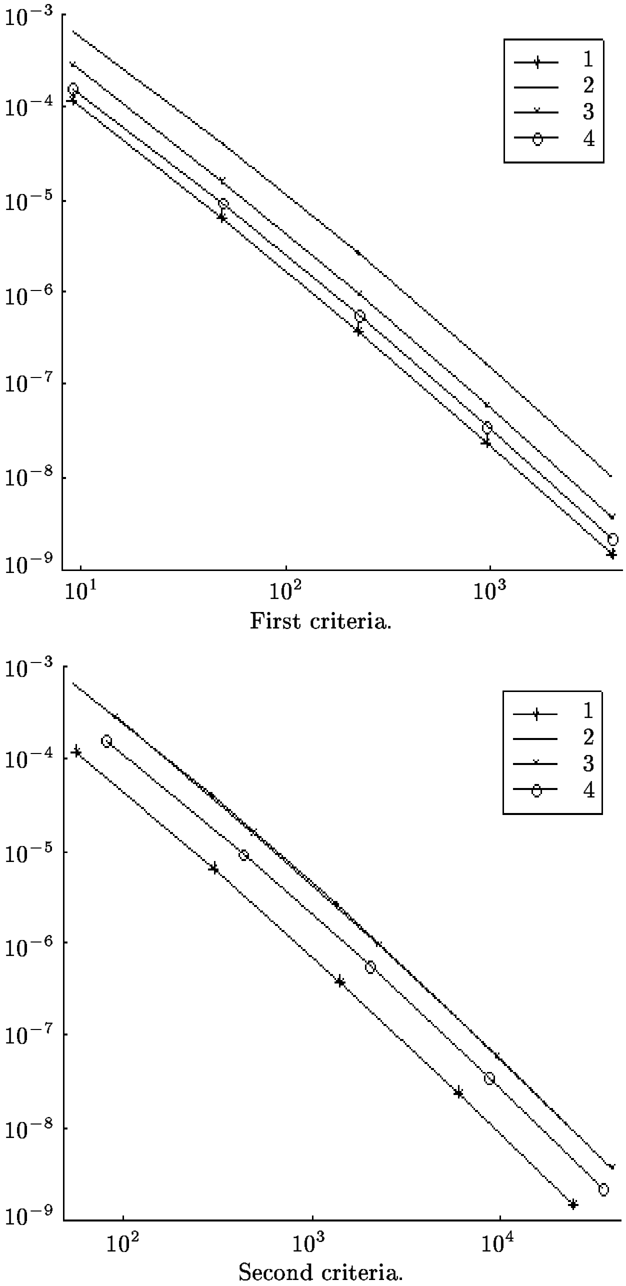
Fig. 1. Energy-norm of error in Example 1.



First criteria.

Second criteria.

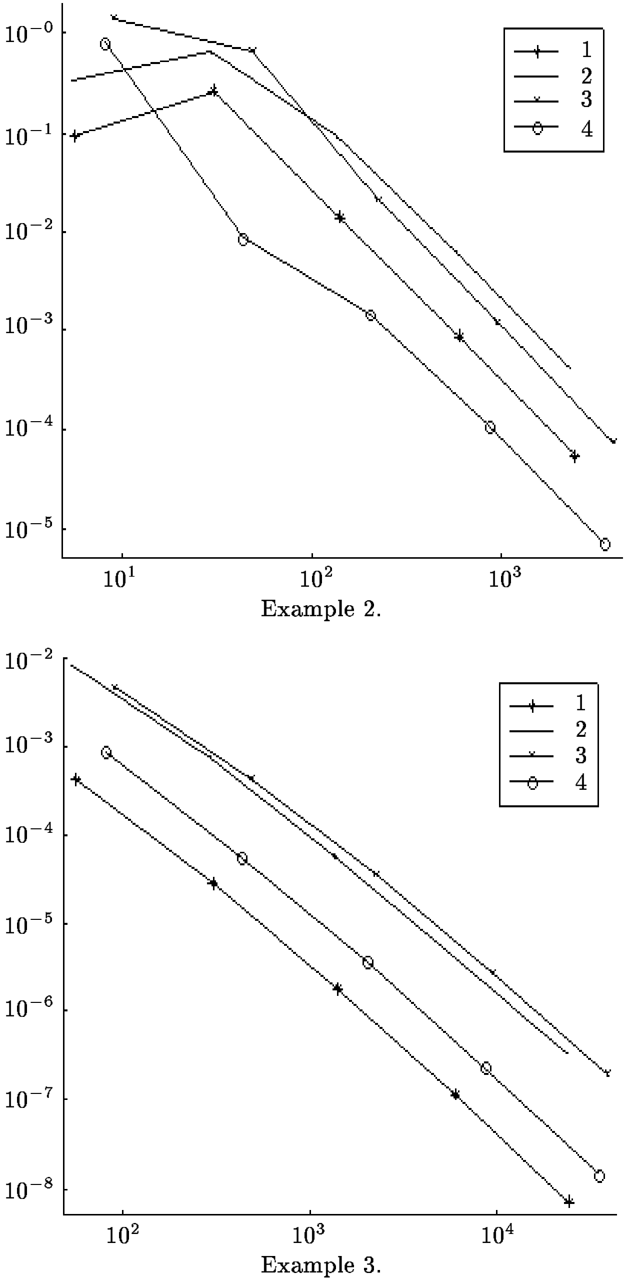
Fig. 2. L_∞−norm of error in Example 1.



First criteria.

Second criteria.

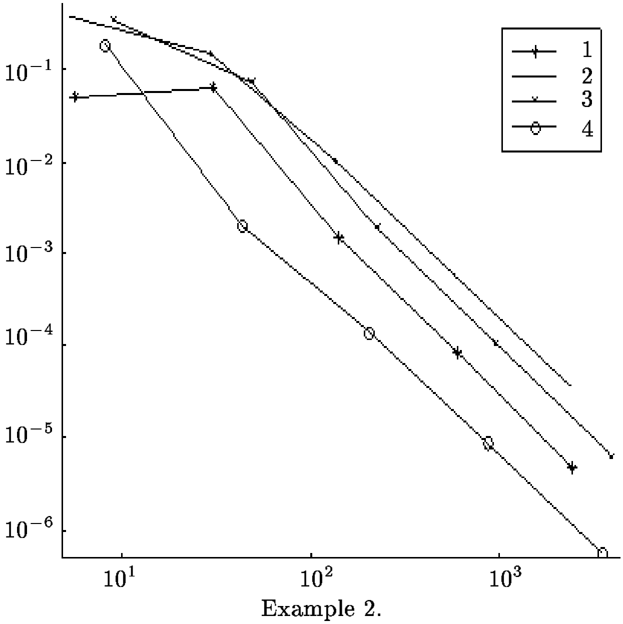
Fig. 3. L_2−norm of error in Example 1.



Example 2.

Example 3.

Fig. 4. Energy-norm of error for Examles 2 and 3. Second criterium.



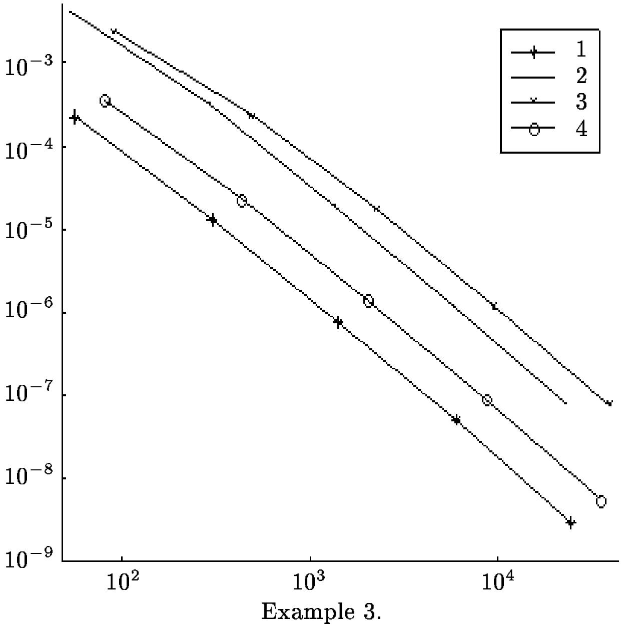
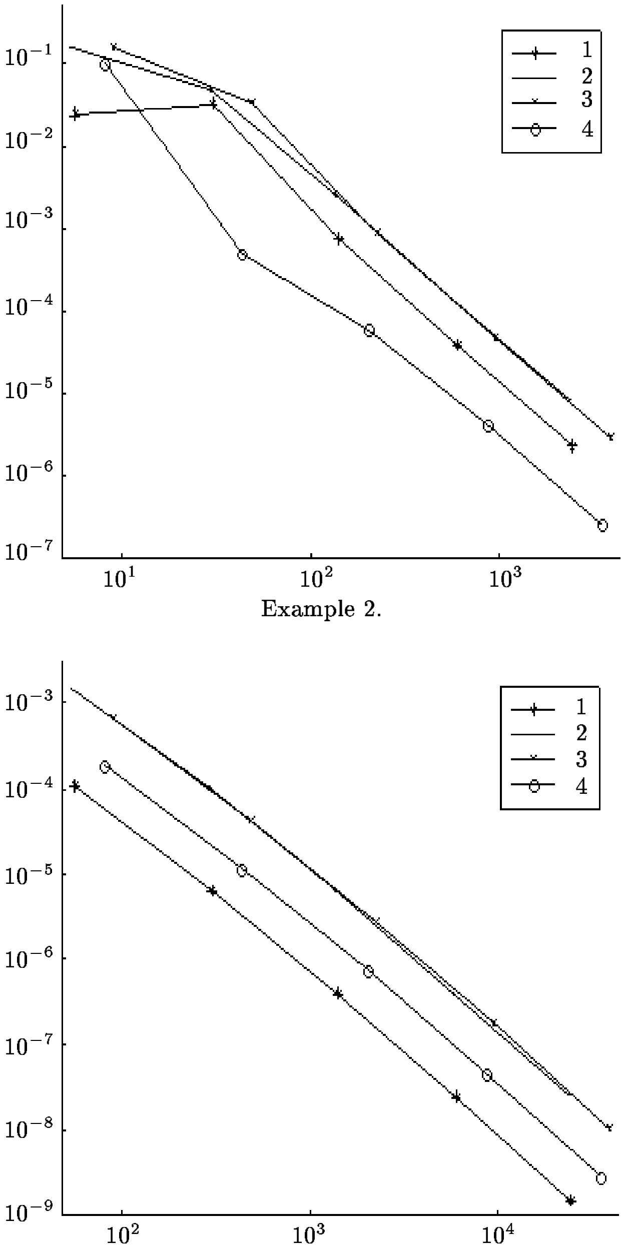


Fig. 5. L_∞−norm of error for Examples 2 and 3. Second criterium.



Example 2.

Example 3.

Fig. 6. L_2—norm of error for Examples 2 and 3. Second criterium.

of the methods. Therefore in Example 2 and 3 we present only the results for the second type of comparison where the complexity is evaluated with respect to number of nonzero coefficients of the matrices. In figures 4, 5, and 6 we show graphs of errors for both examples in the discrete energy norm, L_∞- and L_2-norms, respectively. In each figure the graphs for the L_∞- and L_2-norms are asymptotically lines with a slope that clearly indicates an $O(h^4)$-behavior.

Summarizing, in Examples 1 and 3, where the solution is smooth, Richardson extrapolation is most effective in all norms among the tested methods. For the oscillatitive solution of Example 2, the nine-point box scheme is most efficient, again in all three norms.

References

1. Collatz L.: *The Numerical Treatment of Differential Equations.* Berlin, 1966.
2. Samarskij A.A.: *Theorie der Differenzenverfahren.* Leipzig, 1984.
3. Rüde U.: *Extrapolation and Related Techniques for Solving Elliptic Equations.* Preprint № I-9135, München Technical University, 1991.
4. Marchuk G.I., Shaidurov V.V.: *Difference Methods and Their Extrapolations.* N.Y., Springer, 1983.
5. Bykova E.G., Shaidurov V.V.: *A two-dimensional nonuniform difference scheme with higher order of accuracy.* This book.
6. Bykova E.G., Shaidurov V.V.: *A nonuniform difference scheme with fourth order of accuracy in a domain with smooth boundary.* This book.
7. Khoromskij B.N.: *Method of increasing accuracy of discrete solutions of boundary-value problems with the operator invariant with respect to turning of coordinate system.* Preprint № P5-80-736, Institute of Nuclear resourches, Dubna, 1980 (in Russian).
8. Hackbusch W.: *Theorie und Numerik elliptischer Differentialgleichungen.* Stuttgart, 1996.

Научное издание

Быкова Е.Г., Калпуш Т.В., Карепова Е.Д., Киреев И.В.,
Пятаев С.Ф., Рюде У., Шайдуров В.В.

Уточнённые численные методы для задач конвекции-диффузии

(на англ. яз.) Том 1

Подписано в печать 17.07.2001. Формат 60 × 90 $^1/_{16}$.
Печать офсетная. Усл. печ. л. 16.0. Уч.-изд. л. 14.2.
Тираж 250 экз. Заказ № 17.

Лицензия ЛР № 065614 от 8 января 1998 г.
Издательство Института математики,
пр. Академика Коптюга, 4, 630090 Новосибирск, Россия.

Лицензия ПЛД № 57–43 от 22 апреля 1998 г.
Отпечатано на полиграфическом участке ИМ СО РАН,
пр. Академика Коптюга, 4, 630090 Новосибирск, Россия.